



Programa de Pós-Graduação em

Computação Aplicada

Mestrado Acadêmico

Felipe de Moraes

Detecção e predição de estados afetivos baseadas em mineração de dados educacionais: considerando a personalidade do aluno para aumentar a precisão da detecção

São Leopoldo, 2019

UNIVERSIDADE DO VALE DO RIO DOS SINOS — UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA
NÍVEL MESTRADO

FELIPE DE MORAIS

**DETECÇÃO E PREDIÇÃO DE ESTADOS AFETIVOS BASEADAS EM MINERAÇÃO
DE DADOS EDUCACIONAIS:
CONSIDERANDO A PERSONALIDADE DO ALUNO PARA AUMENTAR A
PRECISÃO DA DETECÇÃO**

São Leopoldo
2019

Felipe de Moraes

**DETECÇÃO E PREDIÇÃO DE ESTADOS AFETIVOS BASEADAS EM MINERAÇÃO
DE DADOS EDUCACIONAIS:
considerando a personalidade do aluno para aumentar a precisão da detecção**

Dissertação apresentada como requisito parcial
para a obtenção do título de Mestre pelo
Programa de Pós-Graduação em Computação
Aplicada da Universidade do Vale do Rio dos
Sinos — UNISINOS

Orientador:
Profa. Dra. Patrícia Augustin Jaques Maillard

São Leopoldo
2019

M827d

Morais, Felipe de.

Detecção e predição de estados afetivos baseadas em mineração de dados educacionais : considerando a personalidade do aluno para aumentar a precisão da detecção / Felipe de Moraes. – 2019.

127 f. : il. ; 30 cm.

Dissertação (mestrado) – Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Computação Aplicada, 2019.

“Orientador: Profa. Dra. Patrícia Jaques.”

1. Mineração de dados educacionais. 2. Estados afetivos. 3. Sistemas tutores inteligentes. 4. Detecção de afeto livre de sensores. 5. Personalidade. I. Título.

CDU 004

Dados Internacionais de Catalogação na Publicação (CIP)
(Bibliotecária: Amanda Schuster – CRB 10/2517)

Felipe de Moraes

Detecção e predição de estados afetivos baseadas em mineração de dados educacionais:
considerando a personalidade do aluno para aumentar a precisão da detecção

Dissertação apresentada à Universidade do Vale do Rio dos Sinos – Unisinos, como requisito parcial para obtenção do título de Mestre em Computação Aplicada.

Aprovado em 01 de março de 2019

BANCA EXAMINADORA

Profa. Dra. Patrícia Augustin Jaques Maillard – Universidade do Vale do Rio dos Sinos (UNISINOS)

Prof. Dr. Sandro José Rigo – Universidade do Vale do Rio dos Sinos (UNISINOS)

Prof. Dr. Ig Ibert Bittencourt – Universidade Federal de Alagoas (UFAL)

Profa. Dra. Patrícia Augustin Jaques Maillard (Orientadora)

Visto e permitida a impressão
São Leopoldo,

Prof. Dr. Rodrigo da Rosa Righi
Coordenador PPG em Computação Aplicada

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES) - Código de Financiamento 001 / This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

Dedico esta dissertação à minha família e amigos por
sempre acreditarem no meu trabalho.

*Tenha em mente que tudo que você aprende na escola
é trabalho de muitas gerações. Receba essa herança,
honre-a, acrescente a ela e, um dia, fielmente, deposite-a
nas mãos de seus filhos. — ALBERT EINSTEIN*

AGRADECIMENTOS

Gostaria de agradecer primeiramente à Deus e à minha família, meu pai Osmar, minha mãe Jucilene, meu irmão Neimar e minha namorada Karen, que nunca mediram esforços para me apoiar e estar ao meu lado em todos os momentos. Incluindo a paciência em entender a minha ausência em eventos e encontros de família e amigos que não pude estar presente devido aos compromissos com o mestrado.

Também gostaria de agradecer intensamente à minha orientadora, Patrícia Jaques, que jamais deixou de me ajudar e me apoiar durante os três anos de iniciação científica, um ano e meio de intercâmbio, um ano de trabalho de conclusão no curso de ciência da computação e dois anos de mestrado. A dedicação com seu trabalho e com todos que trabalham com ela é admirável. Parabéns pela incrível competência e pela pessoa excepcional que és. Também, aos colegas Otávio, Anderson, Lucas e Henrique do grupo de pesquisa de informática na educação do PPGCA, que me auxiliaram durante as sessões de coletas de dados e com dúvidas durante a implementação com o STI PAT2Math. Em especial ao Tiago Kautzmann, que participou do desenvolvimento do protocolo de anotação de estados afetivos e comportamentos utilizado nesta dissertação, participou das sessões de coleta de dados e também me auxiliou com aspectos de implementação e com sugestões de melhoria do trabalho.

Também fica um agradecimento especial ao colégio São Luís, de São Leopoldo, por aceitarem a realização da coleta dos dados utilizados neste trabalho. Principalmente à professora de matemática Mariza, à supervisora Gisele e aos laboratoristas Marcos e Patricia que sempre foram muito prestativos em auxiliar no que fosse preciso para a realização desta pesquisa.

Queria agradecer também aos professores que me deram aula durante o mestrado e que não negaram esforços para me auxiliar durante o desenvolvimento deste trabalho, especialmente aos professores Sandro Rigo e Gustavo Pessin, que participaram do seminário de andamento deste trabalho, trazendo contribuições muito relevantes para a continuidade desta dissertação. Também aos colegas do mestrado, que sempre estavam dispostos a trocar ideias e dar sugestões de melhoria, em especial aos colegas e amigos Eduardo, Ismael e Gustavo pelas conversas, trocas de ideias e parceria durante todo mestrado. E, principalmente, ao Juarez Machado da Silva, que esteve ao meu lado durante todos os dias do mestrado no PPGCA, no qual o apoio e as inúmeras conversas foram essenciais para a conclusão deste trabalho e que além de um colega se tornou um grande amigo.

Por fim, mas não menos importante, gostaria de agradecer a Luciana Aquino e a Bruna Severo, secretárias do PPGCA, que sempre estiveram dispostas a tirar todas as dúvidas e me ajudar com qualquer questão que fosse necessária para a conclusão do mestrado.

“Inteligência é a capacidade de se adaptar à mudança”.
(Stephen Hawking)

RESUMO

As emoções, um tipo de estado efetivo, interferem no processo de aprendizagem, bem como no engajamento dos alunos. Assim, é importante que ambientes educacionais de aprendizagem que visam proporcionar uma melhora no processo de aprendizagem dos alunos, como os Sistemas Tutores Inteligentes (STIs), também possuam a habilidade de reconhecer as emoções dos estudantes e responder adequadamente. Sabe-se que vários ambientes educacionais têm apresentado diferentes formas para realizar o reconhecimento do afeto por meio de sensores ou *hardwares* específicos. Porém, tal estratégia se torna inviável para a utilização desses ambientes em massa. Neste contexto, a estratégia de detecção livre de sensores, por meio da análise dos dados de interação dos alunos com os sistemas, tem se tornado uma solução interessante. Esse trabalho tem como objetivo detectar os estados de frustração, confusão, tédio e engajamento concentrado dos alunos por meio da mineração dos dados em STIs baseados em passos. A hipótese de pesquisa deste trabalho é que a inserção de características da personalidade dos alunos no treinamento dos detectores pode resultar em uma melhora na precisão da detecção destes quatro estados, pois sabe-se que a personalidade influencia os estados afetivos e sua duração. O método utilizado foi o desenvolvimento de detectores treinados com e sem dados de personalidade. Para a obtenção dos dados de treinamento destes detectores, foi realizada uma coleta de dados com 55 alunos de uma escola privada, que utilizaram o STI PAT2Math durante 10 sessões. Em todas estas sessões os alunos tiveram seus rostos gravados, juntamente com áudio ambiente e a tela do computador enquanto utilizavam o sistema. Dos dados obtidos destas sessões de coleta, foi selecionado um total de 5525 *logs* de interação entre o aluno e o sistema. Para cada um desses *logs*, 348 características foram calculadas, contendo informações provenientes (i) das interações do aluno com a interface do PAT2Math, (ii) do módulo do aluno, (iii) dos traços de personalidade e (iv) dos estados afetivos e comportamentos dos alunos. Para a obtenção dos rótulos de estados afetivos, foi desenvolvido um novo protocolo para anotação de estados afetivos e comportamentos baseado na análise dos vídeos gerados durante as sessões de coleta de dados. Esse protocolo segue um conjunto de fases, incluindo treino e teste dos codificadores, e é capaz de ser flexível e generalizável para diferentes aplicações e escalável por não ter a necessidade da presença de anotadores durante a coleta de dados, sendo esta uma contribuição adicional deste trabalho. Por meio deste protocolo, foram coletados 2099 rótulos de estados afetivos e 2059 de comportamentos. Foram desenvolvidos dois detectores para cada estado afetivo, sendo uma versão treinada com e outra treinada sem os dados de personalidade, permitindo verificar o impacto da personalidade na detecção dos estados afetivos. A métrica Cohen's Kappa foi aplicada para identificar a concordância entre os rótulos de estados afetivos gerados pelos codificadores e as saídas dos detectores de estados afetivos desenvolvidos. Como resultados, foi possível identificar que apenas o detector de engajamento concentrado, treinado com dados da personalidade dos alunos, obteve uma pequena melhora na precisão da detecção. Mas, uma característica de personalidade foi selecionada automaticamente durante o treinamento deste detector. Deste modo, este trabalho aponta indícios de que a personalidade pode impactar de forma positiva na detecção dos estados afetivos dos alunos em ambientes de aprendizagem. Destacam-se como contribuições uma possível nova fonte de dados e um novo protocolo para anotação dos estados afetivos. Ambas as contribuições vão em direção ao objetivo de realizar uma detecção automática e em tempo real dos estados afetivos dos alunos, permitindo uma adaptação instantânea dos STIs de acordo com as emoções dos alunos.

Palavras-chave: Mineração de Dados Educacionais. Estados Afetivos. Sistemas Tutores Inteligentes. Detecção de Afeto Livre de Sensores. Personalidade.

ABSTRACT

It has already been shown in the literature that emotions, an affective state type, interfere in the learning process, as well as in the students' engagement. Thus, it is important that learning educational environments that aim to provide an improvement in the student learning process, such as Intelligent Tutoring Systems (ITSs), also have this ability. It is known that several educational environments have presented different ways to realize the recognition of affect through specific sensors or hardware. However, such a strategy becomes unfeasible when it comes to the use of these environments in mass, by hundreds or even thousands of students. Thus, the strategy of sensor-free detection, through the use of interaction data of students with educational environments, has become an interesting solution. This work aims to detect students' frustration, confusion, boredom, and engaged concentration through data mining in step-based ITSs. The research hypothesis of this work is that the addition of personality features of the students in the detectors training of these states can result an accuracy improvement in the detection of these four states since it is known that the personality influences the affective states. The method we used was the development of detectors trained with and without personality features. To collect the training data of these detectors, we carried out a data collection with 55 students from a private school, who used the PAT2Math ITS during ten sessions. During the sessions, students had their faces recorded along with ambient audio and the computer screen while using the system. From the data obtained from these sessions, a total of 5525 interaction logs between the student and the system was selected. For each of these logs, 348 features were calculated, containing information from the (i) student interactions with the PAT2Math interface, (ii) the student module, (iii) the personality traits, and (iv) the affective states and behaviors of the students. We developed a new annotation protocol of affective states and behaviors based on the analysis of the videos generated during the data collection. This protocol follows a set of phases, including training and testing of the coders, and it can be flexible and generalizable for different applications and scalable because it does not require expert coders during data collection. Thus, it is considered an additional contribution to this work. Through this protocol, 2099 labels of affective states and 2059 labels of behaviors were collected. We developed two detectors for each affective state, a trained version with and another without the personality features, allowing to verify the impact of the personality in the detection of the affective states. We applied Cohen's Kappa metric to identify the agreement between the affective state labels generated by the coders and the outputs of the developed affective state detectors. As results, it was possible to identify that only the engaged concentration detector, trained with personality data of the students, obtained a small improvement in the precision of the detection. But, one personality feature was selected automatically during detectors training. Thus, this work points to the evidence that personality can positively impact the detection of students' affective states in learning environments. We highlighted the possible viability of a new data source and a new annotation protocol for affective states as the contributions of this work. Both contributions go towards the goal of performing automatic and real-time detection of students' affective states, allowing instantaneous adaptation of the learning environments according to the students' emotions.

Keywords: Educational Data Mining. Affective States. Intelligent Tutoring Systems. Sensor-free Affect Detection. Personality.

LISTA DE FIGURAS

Figura 1 – As seis emoções básicas.	24
Figura 2 – Arquitetura do STI PAT2Math.	36
Figura 3 – Ferramenta PATequation.	37
Figura 4 – Técnicas de vários domínios.	41
Figura 5 – Ciclos da aplicação de mineração de dados em ambientes educacionais . . .	42
Figura 6 – Etapas de realização deste trabalho.	58
Figura 7 – Classificação dos elementos da interface em grupos.	61
Figura 8 – Diagrama entidade relacionamento de parte da base de dados do módulo do aluno.	65
Figura 9 – Fluxo de troca de fases do protocolo de anotação dos estados afetivos e comportamentos.	78
Figura 10 – Ferramenta desenvolvida utilizada para coletar as anotações dos estados afetivos e comportamentos.	80
Figura 11 – Exemplo de uma sequência de <i>logs</i> gerada por um aluno utilizando o STI PAT2Math, com seus rótulos de estados afetivos e comportamentos.	89
Figura 12 – Exemplo de uma rodada dos ciclos de teste e treino e seleção das características no treinamento dos detectores.	94

LISTA DE TABELAS

Tabela 1 – Adjetivos relacionados a cada um dos cinco fatores.	26
Tabela 2 – Abreviações dos algoritmos utilizados nos trabalhos relacionados.	53
Tabela 3 – Comparativo entre os trabalhos relacionados.	54
Tabela 4 – Comparativo dos resultados entre os trabalhos relacionados.	55
Tabela 5 – Ações dos alunos e as combinações que foram utilizadas.	75
Tabela 6 – Exemplo de cinco anotações geradas pelos três codificadores para uma mesma sessão de anotação.	82
Tabela 7 – Resultados das anotações de estados afetivos e comportamentos.	85
Tabela 8 – Resultados dos detectores com diferentes algoritmos na detecção de cada estado afetivo.	97
Tabela 9 – Características selecionadas para o estado de confusão.	98
Tabela 10 – Características selecionadas para o estado de tédio.	98
Tabela 11 – Características selecionadas para o estado de engajamento concentrado.	99
Tabela 12 – Comparação dos detectores que consideram contra os detectores que não consideram características de personalidade.	99
Tabela 13 – Características selecionadas para o estado de engajamento concentrado no detector que considera a personalidade do aluno.	100
Tabela 14 – Comparação da quantidade de rótulos com os resultados obtidos nos detectores.	104
Tabela 15 – Comparação dos resultados apresentados pelos trabalhos relacionados com o trabalho desenvolvido.	105

LISTA DE SIGLAS

APA	Agente Pedagógico Animado
AU	<i>Action Unit</i>
AVA	Ambiente Virtual de Aprendizagem
BROMP	<i>Baker, Rodrigo, Ocumpaugh Monitoring Protocol</i>
CGF	Cinco Grandes Fatores
EAD	Ensino À Distância
ECG	Eletroencefalograma
EDM	<i>Educational Data Mining</i>
ER	Entidade Relacionamento
FACS	<i>Facial Action Coding System</i>
GRU	<i>Gated Recurrent Unit</i>
GUI	<i>Graphical User Interface</i>
HAC	<i>Hierarchical Agglomerative Clustering</i>
HD	<i>High Definition</i>
ITS	Inteligência Artificial
ITS	<i>Intelligent Tutoring System</i>
KDD	<i>Knowledge Discovery in Databases</i>
LA	<i>Learning Analytics</i>
LSTM	<i>Long-Short Term Memory</i>
PAT2Math	<i>Personal Affective Tutor to Math</i>
PPGCA	Programa de Pós-Graduação em Computação Aplicada
RMSE	<i>Root Mean Square Error</i>
RNN	<i>Recurrent Neural Network</i>
SoLAR	<i>Society for Learning Analytics</i>
STI	Sistema Tutor Inteligente
TCLE	Termo de Consentimento Livre e Esclarecido

LISTA DE SÍMBOLOS

AB	AdaBoost
AD	Adição
AF	Adição e Subtração de Frações
ALR	<i>Additive Logistic Regression</i>
BLR	<i>Bayesian Logistic Regression</i>
BP	<i>Bagging Predictors</i>
C4	<i>C4.5 Decision Trees</i>
DM	Propriedade Distributiva
DT	<i>Decision Trees</i>
DTA	<i>Decision Tables</i>
DV	Divisão
ID	Identificador
JR	<i>JRip</i>
J4	<i>J48 Decision Trees</i>
K	Kappa
KS	<i>KStar ou K*</i>
LIR	<i>Linear Regression</i>
LMT	<i>Logistic Model Trees</i>
LR	<i>Logistic Regression</i>
LWL	<i>Locally Weighted Learning</i>
M	Média
MF	Multiplicação de Frações
MM	Mínimo Múltiplo Comum
MP	<i>Multilayer Perceptron</i>
MT	Multiplicação
NB	<i>Naive Bayes</i>
NBU	<i>Naive Bayes Updatable</i>
NN	<i>Nearest Neighbor</i>
NNG	<i>Nearest Neighbor Generalization</i>
OI	Operação Inversa
PA	<i>PART</i>
RI	<i>RIPPER</i>

RT *REP-Trees*
SB Subtração
SP Simplificação
SR *Step Regression*
SVM *Support Vector Machines*

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Objetivos	18
1.2	Hipótese de Pesquisa	19
1.3	Escopo	19
1.4	Relevância	20
1.5	Contribuições	20
1.6	Organização da Dissertação	21
2	AMBIENTES DE APRENDIZAGEM E ESTADOS AFETIVOS	23
2.1	Estados afetivos	23
2.1.1	Emoções	24
2.1.2	Personalidade	25
2.2	Estados Afetivos em Ambientes de Aprendizagem	26
2.2.1	Emoções	27
2.2.2	Personalidade	29
2.2.3	Estado da arte	30
3	SISTEMAS TUTORES INTELIGENTES	32
3.1	Comportamento dos STIs	32
3.1.1	Laço Externo	32
3.1.2	Laço Interno	33
3.2	Arquitetura Clássica em STIs	33
3.2.1	Módulo Especialista	34
3.2.2	Módulo do Aluno	34
3.2.3	Módulo Tutor	34
3.2.4	Módulo de Comunicação	34
3.3	STI PAT2Math	35
3.3.1	Comportamento do STI PAT2Math	35
3.3.2	Arquitetura do STI PAT2Math	36
4	MINERAÇÃO DE DADOS EDUCACIONAIS	39
4.1	<i>Big Data</i>	39
4.2	Mineração de Dados	39
4.3	Mineração de Dados na Educação	41
4.3.1	Métodos de Predição	43
4.3.2	Mineração de Relacionamento	44
4.3.3	Descoberta de Estrutura	45
4.3.4	Descoberta com Modelos	45
4.4	<i>Learning Analytics</i> e Mineração de Dados Educacionais	46
5	TRABALHOS RELACIONADOS	47
6	TRABALHO DESENVOLVIDO	57
6.1	Etapas do Trabalho	57
6.2	Coleta dos Dados de <i>Logs</i>	59
6.2.1	Desenvolvimento do Coletor de <i>Logs</i>	59
6.2.2	Coleta dos <i>Logs</i>	63

6.3 Coleta dos Dados do Módulo de Aluno	64
6.4 Coleta dos Dados de Personalidade	67
6.5 Coleta dos Rótulos de Estados Afetivos	67
6.5.1 Estado da Arte	67
6.5.2 Protocolo Proposto	71
6.5.3 Definições do Protocolo de Anotação dos Estados Afetivos	72
6.5.4 Fluxo de Troca de Fases do Protocolo	77
6.5.5 Fase dos Materiais	78
6.5.6 Fase de Treinamento dos Codificadores	80
6.5.7 Fase de Teste dos Codificadores	83
6.5.8 Fase de Anotação	84
6.6 Síntese das Características	85
6.7 Desenvolvimento dos Detectores	90
6.7.1 Ferramenta de Desenvolvimento	91
6.7.2 Seleção dos Algoritmos de Treinamento	91
6.7.3 Método de Validação dos Detectores	92
6.7.4 Nível de Validação dos Detectores	92
6.7.5 Seleção das Características	93
6.7.6 Métricas de Avaliação dos Detectores	93
6.7.7 Treinamento dos Detectores	94
7 RESULTADOS E DISCUSSÕES	97
7.1 Resultados	97
7.1.1 Detectores Treinados Sem Considerar a Personalidade	97
7.1.2 Detectores Treinados Considerando a Personalidade	99
7.2 Análise dos Resultados	100
7.2.1 Resultados dos Detectores Sem Personalidade	101
7.2.2 Resultados dos Detectores Com Personalidade	103
7.2.3 Resultados Obtidos Comparados Com os Trabalhos Relacionados	105
8 CONCLUSÕES	107
8.1 Ameaças à Validade dos Resultados	110
8.2 Trabalhos Futuros	110
REFERÊNCIAS	112

1 INTRODUÇÃO

“Embora ninguém possa voltar atrás e fazer um novo começo, qualquer um pode começar agora e fazer um novo fim.”

Chico Xavier

Estar presente em sala de aula não está diretamente relacionado ao ato de aprender. Vários fatores cognitivos e afetivos podem fazer o aluno divergir de seu objetivo. Afeto, motivação e metacognição podem influenciar a cognição, interferindo no processo de aprendizagem. (AZEVEDO; ALEVEN, 2013). Mais especificamente, as emoções, um tipo específico de estado afetivo (SCHERER, 2005), têm um papel fundamental nos alunos em sala de aula. Elas podem afetar a aprendizagem positiva ou negativamente, pois são capazes de controlar atenção, motivação, cognição, modelagem de estratégias e autorregulação da aprendizagem. (PEKRUN, 2014). Emoções positivas, como alegria, engajamento e curiosidade podem afetar aspectos físicos, sociais, intelectuais e criativos, influenciando positivamente no desenvolvimento destas capacidades. (FREDRICKSON, 1998). Embora emoções positivas sejam mais frequentes que as negativas em ambientes de aprendizagem (PEKRUN et al., 2002), as emoções negativas, como frustração, confusão e tédio, podem ocasionar um desequilíbrio cognitivo, afetando também o processo de aprendizagem, quando ocorridas repetidamente e sem intervenção. (GRAESSER; D’MELLO, 2011).

É importante que *softwares* educacionais possuam habilidades afetivas, tornando esses ambientes mais interativos e credíveis aos alunos. Neste contexto se enquadra a computação afetiva. Computação afetiva é definida como o uso de sistemas computacionais capazes de detectar, representar e expressar o afeto humano. (PICARD, 1997). Na área de informática na educação, para que um ambiente de aprendizagem possua habilidades afetivas, primeiramente é necessário que esse sistema seja capaz de detectar os estados afetivos dos alunos para então adaptar suas estratégias pedagógicas de acordo com os estados detectados. (WOOLF et al., 2009; CONATI, 2011; ARROYO et al., 2014).

Os Sistemas Tutores Inteligentes (STI), um tipo de ambiente educacional, são sistemas computacionais especialistas em domínios específicos, que utilizam técnicas de Inteligência Artificial (IA) para proporcionar ensino e assistência individualizados ao aluno. (SHUTE; PSOTKA, 1994; CORBETT; KOEDINGER; ANDERSON, 1997; MURRAY, 1999). A ideia principal de um STI é melhorar a aprendizagem de quem está usando o sistema, sendo essa melhora relacionada com efetividade, material abordado, tempo e custo. (PSOTKA; MASSEY; MUTTER, 1988). Os STIs estão se tornando tão efetivos quanto tutores humanos. (VANLEHN, 2011). Com a disseminação da internet, esses sistemas passaram a ser disponibilizados a todos os tipos de usuários através da *web*. (BRUSILOVSKY; PEYLO, 2003). Assim, eles vêm proporcionando uma melhoria na aprendizagem em várias áreas de conhecimento, como Fís-

sica (LITMAN; SILLIMAN, 2004; VANLEHN et al., 2005), Computação (BRUSILOVSKY; PEYLO, 2003; GRAESSER et al., 2005), treinamento na indústria (FRASSON; AIMEUR, 1998), Matemática (KOEDINGER et al., 1997; JAQUES et al., 2013), entre outros. Os ambientes de aprendizagem, principalmente os STIs, são capazes de gerar muitas informações de interação entre o aluno e o sistema, sendo capazes, inclusive, de inferir o conhecimento do aluno (SEFFRIN; JAQUES, 2015).

Os principais meios de detecção dos estados afetivos identificados na literatura são por expressões faciais, síntese de voz, comportamento observável, sinais fisiológicos, monitoramento cerebral, análise de textos ou pela união de mais de uma estratégia, definida como multimodal. (CALVO; D'MELLO, 2010). Porém, alguns desses métodos utilizam de *hardwares* específicos para realizar a detecção, como sensores e câmeras, adicionando um custo e uma intromissão em aplicações no mundo real. (CALVO; D'MELLO, 2010). Especificamente em ambientes de aprendizagem, a necessidade de câmeras ou sensores para a utilização de alguns métodos pode inviabilizar a utilização desses sistemas em larga escala e por um longo período. (BAKER et al., 2012). Durante a utilização desses ambientes de aprendizagem pelos alunos, vários registros, chamados de *logs*, são gerados e armazenados para uma futura utilização. Esses *logs* contém informações sobre todo tipo de interação do aluno com o sistema.

A mineração de dados educacionais, do inglês *Educational Data Mining* (EDM), é uma área voltada para a descoberta de conhecimento de dados provenientes de ambientes educacionais, visando uma melhor compreensão dos alunos em seus ambientes de aprendizagem. (BAKER et al., 2010a). É uma área recente, que teve um de seus primeiros estudos publicados em 2008. (D'MELLO et al., 2008). Desde então, EDM vem ganhando ênfase na comunidade de informática na educação. (BAKER; YACEF, 2009). Mais recentemente, pesquisas apontam que é possível realizar a detecção dos estados afetivos dos alunos por meio da mineração dos dados de *logs* capturados de ambientes educacionais, como é o caso do trabalho de Paquette et al. (2016), que realizaram um estudo comparativo entre a detecção de estados afetivos livre de sensores contra a detecção baseada em sensores específicos de postura. Os autores apontam resultados favoráveis à detecção livre de sensores. Ainda, Paquette et al. (2014) desenvolveram detectores de estados afetivos para diferentes tarefas dentro de um ambiente para o ensino de física, relatando uma das maiores médias do índice Kappa já obtidas por detectores. Por fim, Jiang et al. (2018) realizaram uma comparação entre duas estratégias de treinamento dos detectores, por meio de uma seleção inteligente das características de treinamento ou simplesmente utilizando técnicas de *deep learning*. Como resultados, os autores apresentaram diferenças significativas a favor da seleção de características.

Atualmente, não existe uma formalização que defina qual o nível mínimo de detecção exigido para a utilização desses detectores na prática. Assim, assume-se que um índice Kappa maior ou igual a 0,6 seria o ideal, pois este valor é o mesmo índice de concordância exigido por codificadores humanos, durante as fases de teste de anotadores em protocolos de anotação de estados afetivos. (OCUMPAUGH, 2015; MORAIS et al., 2019). Ao analisar os trabalhos

relacionados, é possível identificar que nenhum conseguiu alcançar um nível de detecção ótimo (o maior índice reportado foi de $Kappa = 0,469$), para que esses detectores saiam dos testes e possam ser implementados em ambientes educacionais de aprendizagem de forma efetiva. É importante que os detectores apresentem um bom nível de detecção dos estados afetivos, pois se forem implementados na prática, em ambientes de aprendizagem que se adaptam automaticamente aos estados afetivos do aluno, e cometerem muitos erros, os alunos podem vir a desacreditar no sistema ou até mesmo se frustrar ou se irritar com ele, prejudicando diretamente na aprendizagem destes alunos. Com este problema em mente, esta dissertação visa apresentar uma estratégia para melhorar a precisão da detecção dos estados afetivos do aluno. Para isso, serão inseridos dados pessoais de cada aluno no treinamento desses detectores, sendo estes dados referentes à personalidade do aluno. Acredita-se que este tipo de dado pode proporcionar uma melhora na detecção dos estados afetivos, uma vez que a personalidade é capaz de interferir na permanência dos alunos em certos estados afetivos. (REIS et al., 2018).

1.1 Objetivos

Esse trabalho tem como objetivo geral verificar se através da mineração dos dados de *logs*, gerados pela interação dos alunos com um STI de matemática, é possível detectar e prever os estados afetivos destes alunos. Mais especificamente, este trabalho visa identificar os quatro estados afetivos que vêm sendo encontrados com mais frequência nesses tipos de ambientes (CALVO; D'MELLO, 2010): confusão, frustração, tédio e engajamento concentrado. Como objetivos específicos destacam-se (i) a identificação da melhor forma para a coleta dos rótulos de estados afetivos, (ii) a utilização de dados de personalidade para melhorar a precisão dos detectores de estados afetivos e (iii) a comparação dos detectores treinados com e sem dados de personalidade.

Sobre a identificação da melhor forma para a coleta dos rótulos de estados afetivos, objetivo específico (i), para que se possa treinar detectores de estados afetivos baseados em dados de interação entre o usuário e o ambiente educacional, *logs*, é necessário que sejam apresentadas amostras do que seriam os dados representativos de cada um dos estados afetivos durante o processo de treinamento dos detectores. Porém, primeiramente é necessário realizar a coleta destas amostras, contendo rótulos de cada um dos tipos de estados afetivos que visa-se identificar. Desta forma, o objetivo é identificar no estado da arte quais as abordagens que as pesquisas vêm adotando para a realização desta coleta e qual é a melhor forma para ser utilizada nesta dissertação.

Sobre a utilização de dados de personalidade para melhorar a precisão dos detectores de estados afetivos, objetivo específico (ii), para que se possa considerar a personalidade dos alunos, primeiramente deve-se identificar essa personalidade. Uma vez que esses dados forem identificados e estiverem disponíveis, eles serão sincronizados com os *logs* e os rótulos de estados afetivos dos alunos e serão inseridos na fase de treinamento dos detectores. Desta forma, se os

algoritmos de *machine learning*, utilizados para o treinamento dos detectores de estados afetivos, considerarem alguma característica da personalidade, significa que esta característica é capaz de influenciar ou melhorar a detecção desses estados afetivos.

Por fim, sobre a comparação dos detectores treinados com e sem dados de personalidade, objetivo específico (iii), uma vez que se tenha os detectores de estados afetivos treinados e validados, por meio da análise de concordância do cálculo Cohen's Kappa, o objetivo é retreinar os detectores incluindo características da personalidade. Tendo os mesmos modelos treinados com e sem dados de personalidade, pode-se realizar uma comparação dos melhores resultados dos detectores sem dados de personalidade com os detectores que foram treinados com dados de personalidade. Tal comparação permite verificar se os detectores treinados com características de personalidade atingem uma precisão melhor em comparação aos detectores treinados sem dados de personalidade, indicando que a personalidade pode impactar na precisão da detecção dos estados afetivos.

1.2 Hipótese de Pesquisa

A principal diferença desta pesquisa, em relação ao estado da arte, se dá devido à hipótese de pesquisa deste trabalho. Essa hipótese aponta que a personalidade dos alunos pode proporcionar uma melhora na precisão da detecção dos estados afetivos. A justificativa desta hipótese é baseada em pesquisas que apontam que a personalidade pode direcionar ou influenciar as emoções em ambientes computacionais. Sabe-se que a personalidade pode direcionar as emoções de usuários em sistemas *web* (NUNES; CAZELLA, 2011), influenciar o comprometimento de funcionários (PANACCIO; VANDENBERGHE, 2012) e que a duração das emoções negativas depende da personalidade. (REIS et al., 2018). Assim, para validar essa hipótese, dois tipos de detectores foram desenvolvidos, um que não considera e outro que considera a personalidade dos alunos durante a fase de treinamento dos detectores. Desta forma, foi possível realizar uma comparação por meio do cálculo de concordância Cohen's Kappa dos detectores desenvolvidos com e sem dados de personalidade em contraste com os rótulos de estados afetivos gerados pelos codificadores.

1.3 Escopo

Este trabalho limita-se à detecção e predição dos estados afetivos de confusão, frustração, tédio e engajamento concentrado. Esta escolha foi realizada com base em trabalhos publicados nos principais meios de divulgação científica, onde os autores vêm apontando que esses quatro são os estados afetivos que mais ocorrem em ambientes de aprendizagem. (CALVO; D'MELLO, 2010). Além disso, a aplicação desta detecção limita-se à sistemas tutores baseados em passos, pois estes sistemas possuem uma representação maior sobre o conhecimento do aluno, visto que são capazes de auxiliar os alunos passo a passo na resolução das tarefas e ge-

ram uma grande quantidade de *logs* de interação entre o aluno e o sistema. Ainda, este trabalho apresenta uma comparação entre modelos de detecção treinados com e sem dados de personalidade. Desta forma, não é realizado uma implantação destes detectores para detecção dos estados afetivos de forma *online*, ou seja, enquanto o aluno utiliza o ambiente de aprendizagem. (GRAESSER; D'MELLO, 2011).

1.4 Relevância

É importante ressaltar que este trabalho é considerado uma etapa intermediária entre ambientes de aprendizagem que não consideram à ambientes que consideram os estados afetivos de forma automática e livre de sensores. Assim, o resultado final desta trajetória seria a detecção dos estados afetivos em tempo real, ocorrendo no mesmo instante em que os alunos estivessem utilizando o ambiente de aprendizagem. Desta forma, tendo o conhecimento de qual estado afetivo o aluno está sentido, o ambiente de aprendizagem poderia se adaptar de forma automática a este estado e, por sua vez, tomar estratégias pedagógicas adaptativas e individuais, evitando que o aluno entre em estados que possam ser prejudiciais a aprendizagem quando não regulados. (GRAESSER; D'MELLO, 2011). Este trabalho se enquadra como uma etapa intermediária pois visa identificar qual a melhor maneira de realizar essa detecção de forma mais precisa, ao se basear na hipótese de que a personalidade pode influenciar e direcionar os estados afetivos dos alunos, assumindo um papel fundamental para alcançar o objetivo final desta trajetória. Sem uma detecção precisa dos estados afetivos, os ambientes de aprendizagem tomariam decisões com base em dados incorretos, falsos rótulos de estados afetivos, podendo fazer com que os alunos se desmotivem e percam o interesse de continuar utilizado o ambiente de aprendizagem.

1.5 Contribuições

Esse trabalho apresenta duas principais contribuições, que podem ser aplicadas em diferentes áreas. A primeira contribuição está na proposta de um novo protocolo para anotação e identificação dos estados afetivos e comportamento dos alunos enquanto utilizam um ambiente de aprendizagem. O objetivo deste protocolo é a geração de rótulos de estados afetivos e comportamentos para serem utilizados no treinamento de algoritmos de aprendizagem de máquina supervisionado. Os diferenciais deste protocolo, em relação aos outros métodos que vêm sendo utilizados atualmente, se dá por este (i) permitir a anotação de múltiplos estados afetivos em cada clipe, (ii) por possuir um tamanho de clipe menor, que pode variar de acordo com a realidade de cada ambiente de aprendizagem, (iii) por permitir uma tomada de decisão mais precisa pelos anotadores, uma vez que os vídeos dos alunos podem ser vistos quantas vezes cada anotador julgar necessário, (iv) por ser mais escalável, uma vez que os anotadores não precisam estar presentes na sala de aula durante a coleta dos dados e (v) por não exigir um grande número de anotadores, comparado com protocolos em que alunos devem rotular os comportamentos e

estados de colegas. Assim, visto que esse protocolo pode ser expansível e ajustável de acordo com cada aplicação e ambiente de aprendizagem, essa contribuição não se dá apenas na área de sistemas tutores baseados em passos, mas sim a qualquer sistema que possa ser utilizado por aprendizes que possam ter seus rostos e tarefas gravados de forma simultânea. Como o objetivo da elaboração deste protocolo era a geração de rótulos alvo para o treinamento de algoritmos de aprendizagem de máquina, esse protocolo apresenta uma alta qualidade na fase de anotação. Assim, tais rótulos também podem ser utilizados por diferentes pesquisas que visam identificar padrões ou analisar a relação entre estados afetivos e comportamentos com a aprendizagem como um todo.

A segunda contribuição deste trabalho está relacionada diretamente com o objetivo deste trabalho, que era verificar se a personalidade dos alunos pode proporcionar uma melhora na detecção dos seus estados afetivos em ambientes de aprendizagem. Após o treinamento dos algoritmos para a predição dos estados afetivos, sem considerar os dados de personalidade, foi possível identificar um limiar mínimo que poderia ser atingido com os dados, algoritmos e métodos de treinamento utilizados. A partir deste conjunto de detectores treinados sem dados de personalidade, os que obtiveram um melhor índice de concordância entre os rótulos gerados pelos algoritmos e os rótulos obtidos pelo protocolo desenvolvido, indicando o nível de precisão dos detectores, foram selecionados para serem retreinados com os dados de personalidade. Desta forma, foi possível fazer uma comparação onde a única diferença no desenvolvimento e treinamento dos detectores foi a utilização de características específicas sobre a personalidade dos alunos. Ao final, foi possível fazer uma comparação do nível de precisão obtido pelos detectores que consideram e que não consideram a personalidade dos alunos. Como resultado, foi possível identificar que a diferença entre os algoritmos, comparando cada estado afetivo individualmente, foi muito pequena ou não houve diferença. Porém, uma das estratégias de treinamento foi a utilização de métodos de seleção de características que, dentre as 348 possíveis, selecionava no máximo 10 para serem utilizadas durante o treinamento. O resultado positivo em direção à veracidade da hipótese deste trabalho é que para o detector de engajamento concentrado o método de seleção de características selecionou uma característica relacionada a personalidade dos alunos. Tal resultado não afirma que utilização de dados de personalidade impacta na detecção dos estados afetivos de forma positiva. Porém, este resultado traz indícios de que é possível obter uma precisão melhor ao utilizar os dados da personalidade, uma vez que uma característica relacionada a personalidade foi selecionada entre as 10 mais representativas, dentre 348 possíveis.

1.6 Organização da Dissertação

Essa dissertação está organizada da seguinte forma: primeiramente, o capítulo 2 apresenta uma conceitualização do que são os estados afetivos, como estes interferem no processo de ensino aprendizagem do aluno e apresenta uma tendência mais recente da comunidade científica

na utilização dos quatro estados afetivos utilizados neste trabalho. Após, o capítulo 3 descreve o que são os sistemas tutores inteligentes, a arquitetura, como eles funcionando e apresenta com mais detalhes o STI PAT2Math, utilizado como ambiente de aprendizagem de obtenção dos dados deste trabalho. Depois, o capítulo 4 fundamenta o conceito de mineração de dados aplicada em dados educacionais, área denominada como mineração de dados educacionais. Posteriormente, o capítulo 5 faz um levantamento do estado da arte sobre detecção de estados afetivos em ambientes de aprendizagem com base na mineração de dados de interação dos usuários com os sistemas e, ao final, destaca as principais diferenças do trabalho desenvolvido em relação aos trabalhos relacionados. Após, o capítulo 6 apresenta o trabalho desenvolvido, detalhando cada uma das etapas do desenvolvimento, bem como todas as coletas de dados e o protocolo de anotação de estados afetivos e comportamentos desenvolvido neste trabalho. Ainda, também são apresentadas uma síntese detalha das características utilizadas para o desenvolvimento dos detectores e os métodos utilizados para o desenvolvimento destes detectores. Depois, o capítulo 7 apresenta os resultados e realiza uma análise detalhada destes resultados, comparando os detectores treinados com e sem dados de personalidade dos alunos. E, por fim, o capítulo 8 apresenta as conclusões, ameaças à validade dos resultados desta pesquisa e os trabalhos vistos como futuros.

2 AMBIENTES DE APRENDIZAGEM E ESTADOS AFETIVOS

Esse capítulo tem como objetivo apresentar uma visão geral de ambientes de aprendizagem que utilizam os estados afetivos para proporcionar uma melhor aprendizagem. Assim, o foco desse capítulo é apresentar o que são e quais são os estados afetivos abordados, bem como ressaltar a importância destes na aprendizagem e apontar como os pesquisadores da área vêm considerando estes estados para proporcionar uma adaptação e, por consequência, uma melhor aprendizagem aos alunos. Por fim, um levantamento do estado da arte é apresentado, ressaltando um redirecionamento dos trabalhos.

2.1 Estados afetivos

Os estados afetivos representam estados mentais afetivos e têm sido classificados em diferentes categorias, como emoções, humor e traços de personalidade. (SCHERER, 2000; FRIJDA et al., 1994).

As **emoções** são intensas, possuem uma curta duração, são focadas em um evento e, por isso, são disparadas pela avaliação desse evento. (SCHERER, 2005). As emoções variam entre níveis de intensidade alto e baixo (IZARD, 2009), possuem valência positiva ou negativa (ORTONY; CLORE; COLLINS, 1990) e podem mudar rapidamente. (SCHERER, 2000). Por serem direcionadas a um objeto, objeto intencional, considera-se uma emoção um estado mental intencional. (FRIJDA et al., 1994). Um exemplo de emoção é um aluno que se sente confuso com o novo conteúdo de matemática que o professor está explicando.

O **humor**, ou estado de ânimo, ao contrário das emoções, é um estado de baixa intensidade, mas que possui uma duração maior, podendo permanecer por dias. (SLOMAN, 2002). Ainda, o humor não tem uma causa aparente, não necessita de um objeto intencional como as emoções, e consiste em mudanças no sentimento subjetivo. Segundo Scherer (2005), o humor é classificado em animado, desanimado, indiferente, ansioso e irritado. Um exemplo prático é quando alguém diz que não está em um dia bom, ou que está mal humorada, sem um motivo específico.

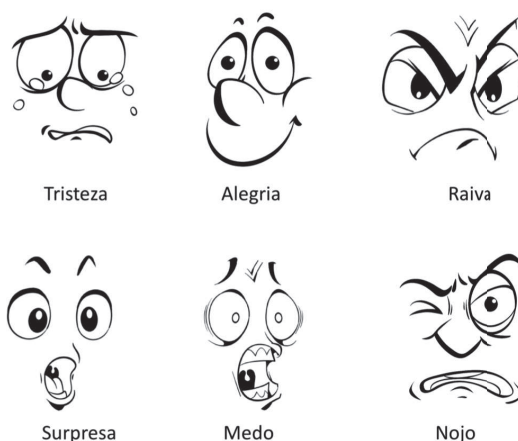
Os **traços de personalidade** representam a tendência das pessoas experimentarem certo humor ou de reagir de alguma maneira emocional. Alguns exemplos são introvertido e extrovertido.

Esse trabalho foca em dois principais estados afetivos, as emoções e a personalidade. Assim, as próximas seções têm o objetivo de apresentar uma visão mais detalhada de cada um desses estados, bem como descrever a relação deles com ambientes de aprendizagem e com esta dissertação.

2.1.1 Emoções

Vários pesquisadores vêm construindo o conhecimento de emoções que se tem atualmente. Esses estudos tiveram seu início por volta de 430 a.c. com Platão, sugerindo que a alma é uma estrutura dividida em cognição, emoção e motivação. Quinze anos depois, Aristóteles sugere que esses componentes se interagem. No século 17, Descartes afirma que a cognição não tinha relação com as emoções. Depois, James (1884), psicólogo e filósofo norte-americano, define emoções como reações corporais a eventos. Ainda, Darwin e Prodger (1998) identificaram padrões em expressões de emoções, mesmo em diferentes culturas. Tal teoria deu origem ao que se conhece por teoria das emoções básicas, sendo elas raiva, alegria, tristeza, nojo, medo e surpresa (EKMAN, 1994; IZARD, 1984), ilustradas na Figura 1.

Figura 1 – As seis emoções básicas.



Fonte: (JAQUES; NUNES, 2019).

Atualmente, uma definição de emoções, bem aceita na comunidade científica, é a de Scherer (2005). Segundo esse pesquisador, emoção é “um episódio de mudanças sincronizadas e inter-relacionadas nos estados de todos ou da maioria dos cinco componentes do subsistema do organismo em resposta à avaliação de um evento de estímulo interno ou externo como sendo de grande importância para um organismo”. Essa definição vê a emoção como resultado de vários componentes. Esses componentes são divididos em: (i) cognitivo (*appraisal*), responsável pela avaliação de objetos e eventos; (ii) neurofisiológicos (mudanças corporais), responsável pela regulação do sistema; (iii) motivacional (tendências de ação), que é a preparação e direção da ação; (iv) expressão motora (expressão facial e vocal), responsável pela comunicação da intenção comportamental com a reação; e, por fim, (v) sentimento subjetivo (experiência emocional), responsável pelo monitoramento dos estados internos e pela interação organismo ambiente.

Cada componente desses abre uma gama de possibilidades, dependendo da linha teórica dos pesquisadores. (JAQUES; NUNES, 2019). Pesquisadores da área de psicologia comportamental, por exemplo, se aproveitam principalmente do componente de expressão motora. Como é

o caso de Ekman (1994) e Izard (1984), que definiram a teoria das emoções básicas, influenciando estudos envolvendo a detecção das emoções por meio da captura de expressões faciais em imagens ou de *webcams*. (OLIVEIRA; JAQUES, 2008).

Em outro exemplo, pesquisadores da área de psicologia cognitiva, utilizam o componente cognitivo das emoções, responsável pelo *appraisal*. (SCHERER, 1999; ARNOLD, 1960; LAZARUS, 1991). O *appraisal* é a avaliação cognitiva de situações e eventos que precedem uma emoção. (SCHERER, 1999). Diante de uma determinada situação, o cérebro realiza uma avaliação cognitiva, identificando se as ações são favoráveis ou não, gerando assim uma emoção. A psicologia cognitiva tem influenciado trabalhos na computação afetiva. (JAQUES; VICARI, 2007; CONATI; ZHOU, 2002). Além disso, também baseado na abordagem cognitiva das emoções, foram propostos modelos psicológicos que descrevem as emoções e os processos cognitivos associados a elas, como o modelo OCC (ORTONY; CLORE; COLLINS, 1990), o modelo estrutural (LAZARUS, 1991) e o modelo “*Multi-level Sequential Check*”. (SCHERER, 2001).

2.1.2 Personalidade

A personalidade é definida como um conjunto de características presentes nas ações e interações das pessoas que ocorrem cotidianamente e que são permanentes e exclusivas de cada um. (SCHULTZ, 1990). Ainda, outras definições também foram descritas, como a de Burger (2000) afirmando que personalidade é “um padrão de comportamento consistente e processo intrapessoal que é originado internamente no indivíduo” e a de Soldz e Vaillant (1999), dizendo que a personalidade humana, embora estável e previsível, não é permanente, conforme o que Schultz (1990) afirmou. Ainda, Soldz e Vaillant (1999) afirmaram que normalmente a personalidade se mantém estável por um período de 45 anos, tendo seu início já na fase adulta do ser humano.

Allport e Allport (1921) definiram que a personalidade das pessoas poderia ser distinguida através de seus traços de personalidade, que seriam características comuns e individuais expressadas em diferentes intensidades. Tais traços são decorrentes de influências externas e ambientais e do histórico de vida de cada pessoa. Os autores criaram cerca de 18 mil traços diferentes, o que depois de vários anos de pesquisa, divergiram para cinco principais traços, conhecidos atualmente como *Big Five* (JOHN; SRIVASTAVA, 1999), do português modelo dos Cinco Grandes Fatores (CGF).

Os Cinco Grandes Fatores são extroversão, socialização, realização, neuroticismo, abertura (abertura à experiências). Cada fator representa características únicas do comportamento humano: (i) extroversão identifica pessoas ativas, sensíveis, impulsivas e assertivas; (ii) socialização aponta pessoas úteis, carinhosas, gentis e confiantes; (iii) realização mostra pessoas eficientes, organizadas, confiáveis e com pensamento de responsabilidade social; (iv) neuroticismo reflete pessoas ansiosas, mal-humoradas e autopunitivas; e, por fim, (v) abertura caracteriza pessoas curiosas, criativas, com facilidade intelectual e aberta a novas experiências. (BERGER,

2003; LOEHLIN, 1992). A Tabela 1 descreve alguns adjetivos relacionados a cada um dos cinco fatores, bem como o oposto de cada um dos polos (adaptado de (LOEHLIN, 1992)).

Tabela 1 – Adjetivos relacionados a cada um dos cinco fatores.

	Extroversão	Socialização	Realização	Neuroticismo	Abertura
Polo do rótulo	Ativo	Altruísta	Confiável	Ansioso	Artístico
	Aventureiro	Amigável	Consciente	Apreensivo	Curioso
	Barulhento	Carinhoso	Eficiente	Emotivo	Engenhoso
	Energético	Confiante	Minucioso	Instável	Esperto
	Entusiástico	Cooperativo	Organizado	Nervoso	Imaginativo
	Exibido	Gentil	Prático	Preocupado	Inteligente
	Sociável	Sensível	Preciso	Temeroso	Original
	Tagarela	Simpático	Responsável	Tenso	Sofisticado
Polo oposto	Acanhado	Antipático	Desatento	Calmo	Comum
	Introvertido	Brigão	Descuidado	Contido	Simples
	Quieto	Bruto	Desorganizado	Estável	Superficial
	Reservado	Crítico	Distraído	Indiferente	Tolo
	Silencioso	Frio	Imprudente	Sereno	Trivial
	Tímido	Insensível	Irresponsável	Tranquilo	Vulgar

Fonte: (NUNES, 2012).

O comportamento cotidiano das pessoas deixa pistas sobre os traços de personalidade humana, diferindo uma das outras. (GOSLING, 2009). Existem algumas ferramentas, em forma de escalas ou questionários que podem ser utilizados para identificar e inferir esses traços. Embora hajam questionários e escalas reconhecidos e validados internacionalmente (JOHNSON, 2000, 2005; GOSLING; RENTFROW; SWANN JR, 2003), esses normalmente são elaborados na língua inglesa. Assim, não é possível utilizá-los no Brasil sem uma tradução para o português, pois apenas traduzir esse tipo de documento pode alterar o real sentido das questões. Por isso, é necessário que, depois de traduzido, essas ferramentas passem por um processo de validação. Andrade (2008), em sua tese de doutorado, apresentou uma dessas ferramentas validada para a língua portuguesa, a qual foi utilizada nessa dissertação como forma de captura dos traços de personalidade dos alunos.

2.2 Estados Afetivos em Ambientes de Aprendizagem

Vários pesquisadores vêm destacando a importância dos estados afetivos para a aprendizagem. (VYGOTSKY, 1994; VAIL, 1994; JOHN-STEINER, 2000). Primeiramente, acredita-se que a parte cognitiva não existe sem a parte afetiva e que o afeto aumenta a atividade intelectual. (PIAGET, 1989). Depois, o intelecto está completamente relacionado com a aprendizagem. As emoções positivas podem afetar de forma construtiva na aprendizagem. (IZARD, 1984). Por outro lado, alunos com emoções negativas podem ter uma maior dificuldade em aprender, prejudicando a aprendizagem. (GOLEMAN, 1995). Mais especificamente na aprendizagem do conteúdo de matemática, Piaget (1989) sugere que boa parte dos alunos que apresentam difi-

culdades, se deve a um bloqueio afetivo. Assim, sabendo que os estados afetivos interferem na aprendizagem, é importante que ambientes de aprendizagem, que tem como objetivo oferecer uma melhor aprendizagem aos alunos, também sejam capazes de detectar e lidar com esses estados de forma adaptativa.

Nessa seção, serão descritos trabalhos na área de computação afetiva aplicada à educação, que têm como objetivo geral proporcionar uma melhora no processo de aprendizagem dos alunos, baseando-se nas teorias e trabalhos que relacionam afeto e cognição. Mais precisamente, essa seção descreve relações de emoção e personalidade com aprendizagem em ambientes educacionais.

2.2.1 Emoções

Mais precisamente sobre emoções, em seu trabalho, Jaques e Nunes (2019) classificam os trabalhos em computação afetiva na educação em três grupos: (i) reconhecimento de emoções, (ii) expressão de emoções e (iii) síntese de emoções. As subseções seguintes apresentam uma visão geral de cada um desses grupos, dando uma ênfase maior no reconhecimento das emoções, que é onde esse trabalho está inserido.

2.2.1.1 Reconhecimento de Emoções

Jaques e Vicari (2005) descrevem quatro principais fontes de dados para o reconhecimento das emoções de alunos em ambientes de aprendizagem.

- i) Voz (prosódia), que apresenta informações como entonação, volume ou até mesmo vocalizações não-linguísticas, como risos, gritos. (KOPECEK, 2000; TCHERKASSOF, 1999);
- ii) Comportamento observável, que representa as ações dos alunos na interface de um ambiente de aprendizagem, por exemplo, número de cliques, pedidos de ajuda, tempo ocioso, entre outros. (BERCHT; VICCARI, 2000; MARTINHO; MACHADO; PAIVA, 2000; VICENTE; PAIN, 1999; JAQUES; VICCARI, 2004; JAQUES et al., 2004; CONATI, 2002; CONATI; ZHOU, 2002; CONATI; MACLARE, 2004).
- iii) Expressões faciais, obtidas tanto por imagem estática, como fotos, quanto por vídeos. (EKMAN, 1999; WEHRLE; KAISER, 2000).
- iv) Sinais fisiológicos, que representam dados de batimentos cardíacos, tensão muscular, condutividade da pele (eletromiograma) e respiração, emitidos por equipamentos específicos. (PICARD; VYZAS; HEALEY, 2001).

Mais recentemente, Jaques e Nunes (2019) destacam mais três fontes de dados:

- v) Texto, que representa os sistemas que possuem algum tipo de diálogo entre os alunos e o sistema. Quando em conjunto com a voz, o conteúdo de texto pode trazer bons resultados. (FORBES-RILEY; LITMAN, 2004);
- vi) Ondas cerebrais, que representam os dados obtidos por eletroencefalograma (ECG). Alguns trabalhos envolvendo a detecção de emoções por meio de expressões faciais juntamente com ondas cerebrais (sinais fisiológicos) vêm sendo realizados. (DAVIDSON et al., 1990; PETRANTONAKIS; HADJILEONTIADIS, 2010);
- vii) Expressão corporal, que representa os dados obtidos pela postura do corpo do aluno. Trabalhos vêm utilizando as expressões faciais combinados com a expressão corporal para um reconhecimento mais preciso das emoções. (BURLESON; PICARD, 2007; GUNES; PICCARDI, 2007).

Quando mais de uma fonte de dados é utilizada, utiliza-se o termo de reconhecimento multimodal de emoções. (ARROYO et al., 2009; WOOLF et al., 2009). Sebe et al. (2005) apresenta uma visão mais abrangente do tema e aponta que é um desafio utilizar várias abordagens em conjunto. Porém, autores têm afirmado que essa é a forma mais efetiva de detecção. (CALVO; D'MELLO, 2010).

Diferentes pesquisas têm utilizado diferentes métodos de detecção de estados afetivos em ambientes educacionais. Litman e Forbes-Riley (2006) utilizaram a voz, D'Mello e Graesser (2010) utilizaram expressões faciais e corporais, Arroyo et al. (2009) e Muldner, Burleson e VanLehn (2010) utilizaram sinais fisiológicos, entre outros. Porém, tais métodos se tornam restritos, limitando a pesquisa apenas a escolas que possuam suporte para essas tecnologias. O custo de sensores, câmeras e microfones pode tornar a utilização desses ambientes inviável. Além disso, os dados providos de sensores e *hardwares* específicos podem restringir a aplicabilidade de novos métodos. Com base nessas limitações, uma opção que vem se apresentando muito efetiva é a detecção de estados afetivos por meio da mineração de dados de *logs*, na utilização dos sistemas educacionais pelos alunos, que representam o comportamento observável do aluno com o sistema, descrito com mais detalhes no capítulo 4.

2.2.1.2 Expressão de Emoções

A tarefa de reconhecer as emoções dos alunos tem como objetivo principal melhorar a aprendizagem. Porém, apenas detectar essas emoções não é suficiente para atingir esse objetivo. São necessárias abordagens ou estratégias adaptativas. Uma delas é por meio da expressão de emoções. Assim, o sistema pode encorajar o aluno e se mostrar mais empático, no sentido de se importar com as emoções deste (JAQUES; LEHMANN; PESTY, 2009; JAQUES et al., 2011; ARROYO; et al., 2011), ou até mesmo realizar essa expressão de forma reflexiva, como se o sistema estivesse imitando os estados afetivos do aluno. (LEITE et al., 2010, 2012; BURLESON; PICARD, 2007).

Uma forma bastante utilizada para realizar a expressão de emoções é com o uso de Agentes Pedagógicos Animados (APA). Estes agentes simulam um personagem animado que interagem com o aluno, introduzindo uma comunicação mais social e antropomórfica. (JAQUES et al., 2001). Várias pesquisas têm sido realizadas com agentes animados em ambientes educacionais (BURLESON et al., 2004; BURLESON; PICARD, 2007; JAQUES; JAQUES; LEHMANN, 2008; JAQUES; LEHMANN; PESTY, 2009; JAQUES et al., 2011), até mesmo apresentando impacto nas atitudes dos alunos em relação à matemática. (WOOLF et al., 2009; ARROYO; et al., 2011). Além de agentes pedagógicos animados, alguns trabalhos também têm utilizado robôs para expressar emoções. (LEITE et al., 2010, 2012).

2.2.1.3 Síntese de Emoções

A síntese de emoções engloba sistemas computacionais que possuem emoções. (PICARD, 1997). Também utiliza de agentes pedagógicos animados, como por exemplo, o trabalho de Elliott, Rickel e Lester (1999) que apresenta um agente que possui emoções geradas através de um conjunto de regras baseado no modelo OCC. (ORTONY; CLORE; COLLINS, 1990). Assim como na expressão de emoções, a síntese de emoções também tem como objetivo aumentar a credibilidade do sistema, normalmente representado com agentes animados ou robôs. (EL-NASR et al., 1999; ELLIOTT, 1997; REILLY; BATES, 1992; LOYALL; BATES, 1997; BATES, 1994). Como exemplo, pode-se citar o trabalho de Elliott, Rickel e Lester (1999), onde o agente Steve fica deprimido se o aluno estiver entediado com suas tarefas.

2.2.2 Personalidade

A abordagem de traços é considerada a melhor forma para representar a personalidade em sistemas computacionais. (NUNES, 2009). Embora a técnica mais comum para a identificação da personalidade das pessoas ainda seja por meio de inventários de personalidade (DUNN et al., 2009), alguns trabalhos vêm pesquisando maneiras de realizar essa identificação de forma automática e menos intrusiva, como por exemplo, através da mineração de textos (MACHADO, 2016), ou até mesmo através da forma com que o usuário digita um texto no teclado do computador. (PORTO et al., 2011).

Na área de informática na educação, mais especificamente, pesquisadores têm utilizado a personalidade dos alunos de várias formas, como por exemplo, para formação de grupos em ambientes de aprendizagem (ROMERO et al., 2009), para tornar agentes animados mais críveis (LOYALL; BATES, 1997), mais empáticos (COOPER, 2003) e para promover auto-eficácia. (BANDURA; RAMACHAUDRAN, 1994; JAQUES; LEHMANN; PESTY, 2009).

2.2.3 Estado da arte

Atualmente, já se sabe que os estados afetivos interferem na motivação e na aprendizagem em ambientes educacionais. (DRAGON et al., 2008; ROWE et al., 2009; BAKER et al., 2010b, 2011; D’MELLO, 2012). Conforme Calvo e D’Mello (2010) apontam, muitos trabalhos na área de computação afetiva ainda se baseiam na teoria das emoções básicas, proposta por Ekman (1994). Porém, já existem vários estudos relatando que as emoções básicas têm uma relevância muito pequena em sessões de alunos interagindo com ambientes de aprendizagem. (D’MELLO; PICARD; GRAESSER, 2007; LEHMAN et al., 2008; LEHMAN; D’MELLO; PERSON, 2008; BAKER et al., 2010b; D’MELLO; CALVO, 2013).

Assim, as pesquisas passaram a buscar outros tipos de estados afetivos, mais comuns nesses tipos de ambientes de aprendizagem. Como exemplo, no trabalho de Lagud e Rodrigo (2010), que buscam identificar tédio, confusão, frustração, surpresa e *flow*. Conhecido na língua inglesa como *flow*, esse é um estado afetivo que envolve um forte interesse em determinada tarefa sendo realizada com alta concentração. Outro exemplo é o trabalho de Guia et al. (2012), que visa detectar tédio, confusão e concentração em um STI. Nesse mesmo contexto, Baker et al. (2010b) observou a utilização de três ambientes educacionais diferentes pelos alunos. O objetivo era verificar como os estados afetivos detectados, tédio, confusão e concentração, interferem no modo que os alunos interagem com os sistemas.

Calvo e D’Mello (2010) sugerem que as emoções confusão, frustração, tédio, *flow*, curiosidade e ansiedade são mais aplicáveis em ambientes educacionais que reconhecem, expressam ou sintetizam emoções. Os autores também reafirmam que “o afeto está sempre influenciando ativamente a cognição e o comportamento e o desafio é modelar essas manifestações de emoção perenemente presentes, mas de certa forma sutis”.

Autores têm apontado que frustração e tédio podem ter um impacto negativo na aprendizagem, enquanto que confusão e *flow* podem impactar positivamente. (GRAESSER et al., 2005; D’MELLO et al., 2014). Porém, se persistir por um longo período, a confusão pode se tornar frustração ou tédio. Assim, para apresentar esse impacto positivo, a confusão precisa ser regulada de acordo com a personalidade e conhecimentos anteriores do aluno. (CRAIG et al., 2004; D’MELLO; PICARD; GRAESSER, 2007). Por exemplo, o trabalho de Nunes e Cazella (2011) apresenta um sistema de recomendação *web* baseado na personalidade dos usuários. Eles identificaram que a personalidade direciona as emoções do usuário e utilizaram essas informações para desenvolver o sistema que obteve resultados muito satisfatórios. D’Mello (2012) apresenta um conjunto de transições esperadas entre esses quatro estados mencionados anteriormente. Com base neste trabalho, Reis et al. (2018) realizaram um estudo sobre o tempo de permanência dos alunos no estado de confusão até este se tornar um estado de frustração ou tédio, com base em conhecimentos algébricos e nos traços de personalidade dos alunos em um ambiente educacional. Ainda sobre o trabalho de Reis et al. (2018), os autores apontam que os traços de personalidade impactam de forma significativa no tempo de permanência dos alunos no estado

de confusão. Esse resultado pode ser utilizado na construção de modelos para a regulação emocional dos alunos, evitando que estes fiquem frustrados ou entediados e venham a desistir de realizar suas tarefas.

3 SISTEMAS TUTORES INTELIGENTES

Este capítulo tem como objetivo apresentar uma visão geral do conceito de sistemas tutores inteligentes, bem como apresentar principais definições, arquitetura e comportamento desse tipo de sistema. Por fim, este capítulo também apresenta o PAT2Math, que é um sistema tutor inteligente voltado para o ensino de matemática que será utilizado para a obtenção dos dados de interação dos alunos que serão analisados neste trabalho.

Atividades que proporcionam uma maior interatividade ao aluno, possibilitando-o de fazer ao invés de apenas ver, resultam em uma melhora na aprendizagem. (KOEDINGER et al., 2015). Os Sistemas Tutores Inteligentes (STI), do inglês *Intelligent Tutoring Systems (ITS)*, são *softwares* educacionais que têm a capacidade de proporcionar essa interatividade, pois são capazes de auxiliar o aluno individualmente em suas tarefas. Diferentemente de outros *softwares* educacionais, os STIs são especialistas em uma determinada área e possuem o conhecimento para fornecer uma tutoria individualizada ao aluno. (CORBETT; KOEDINGER; ANDERSON, 1997; GRAESSER; CONLEY; OLNEY, 2012). Sistemas tutores inteligentes estão sendo desenvolvidos em várias áreas (por exemplo Matemática, Física, Leitura, Computação, Indústria, Medicina, entre outras) e estão atingindo bons resultados no processo de aprendizagem. Os STIs têm se mostrado quase tão efetivos quanto tutores humanos. (VANLEHN, 2011; STEENBERGEN-HU; COOPER, 2013; MA et al., 2014).

Independente da área de aplicação, o comportamento dos STIs é similar. Basicamente, os STIs possuem dois laços (*loops*), o laço externo (*outer loop*) e o laço interno (*inner loop*) (VANLEHN, 2006), que são descritos com mais detalhes na seção 3.1. A arquitetura dos STIs também possui uma similaridade, conhecida por arquitetura clássica, a qual divide o STI em uma estrutura modular. (DEVEDZIC; HARRER, 2005). Essa arquitetura clássica de um STI é descrita na seção 3.2.

3.1 Comportamento dos STIs

Os STIs são responsáveis por selecionar a próxima tarefa a ser apresentada ao aluno e, diferentemente de outros *softwares* educacionais, também são responsáveis por auxiliar o aluno em cada passo da resolução de uma tarefa. Essa seção tem como objetivo descrever o comportamento dos STIs. Tal comportamento é dividido em dois laços: o externo, que é responsável pela seleção de tarefas, e o interno, que é responsável por auxiliar o aluno na resolução passo-a-passo de cada tarefa.

3.1.1 Laço Externo

O laço externo, do inglês *outer loop*, é responsável pela escolha do próximo conteúdo ou exercício que será apresentado ao aluno. Essa escolha pode ser realizada de uma maneira inte-

ligente pelo sistema ou de uma maneira já pré-definida, seguindo uma sequência.

Vanlehn (2006) define quatro estratégias adotadas por STIs. A primeira estratégia é deixar o próprio aluno selecionar a próxima tarefa. A segunda estratégia é selecionar a próxima tarefa a partir de uma sequência fixa de tarefas. A terceira estratégia é baseada no domínio do conteúdo, do inglês *mastery learning*, na qual há uma sequência definida de tarefas, porém o aluno somente pode passar para a próxima tarefa quando ele atingir um determinado nível de conhecimento. A quarta estratégia, e mais complexa, é a macro-adaptação, do inglês *macroadaptation*, na qual o sistema seleciona a próxima tarefa baseado nas estimativas de conhecimento que o aluno possui. O laço externo utiliza o módulo tutor, descrito na seção 3.2, para a seleção das tarefas.

3.1.2 Laço Interno

O laço interno, do inglês *inner loop*, é responsável por auxiliar o aluno em cada passo da resolução de uma tarefa, diferentemente do laço externo, que é responsável pela seleção de tarefas. Um passo de resolução de uma tarefa representa uma interação do usuário com o sistema que tem o objetivo de completar uma tarefa.

Alguns serviços são destacados como mais utilizados no laço interno dos STIs, segundo Vanlehn (2006):

- i) um *feedback* mínimo, referente ao passo inserido pelo aluno, indicando se o passo do aluno está correto ou não;
- ii) um *feedback* específico para o erro do aluno, caso o passo esteja incorreto;
- iii) um sistema de dicas, no qual o aluno pode solicitar uma dica, referente ao passo em que ele se encontra, quando não souber como prosseguir;
- iv) um acesso do aluno ao seu conhecimento, através de estimativas obtidas pelo sistema;
- v) e, por último, uma revisão da solução da tarefa, na qual o aluno pode visualizar todos os passos inseridos para a resolução de cada tarefa.

Para fornecer tais serviços, o laço interno utiliza o módulo especialista, descrito na seção 3.2, que é responsável por verificar e fornecer um *feedback* de cada passo inserido pelo aluno no sistema.

3.2 Arquitetura Clássica em STIs

A arquitetura dos STIs varia de sistema para sistema, dependendo principalmente do seu domínio. No entanto, observa-se que a maioria dos STIs segue uma arquitetura básica em sua implementação. Essa arquitetura foi classificada como arquitetura clássica em STIs e é dividida em quatro módulos: especialista, do aluno, tutor e o de comunicação. (Nwana, 1990).

3.2.1 Módulo Especialista

O módulo especialista é onde o conhecimento do STI é armazenado. Ele é considerado a coluna principal de qualquer STI, pois contém o conhecimento do domínio. Tal conhecimento é “ensinado” ao STI a partir do conhecimento de especialistas, que normalmente são os professores de uma determinada área de conhecimento. (ANDERSON, 1988).

O módulo especialista pode ser dividido em dois componentes, o modelo cognitivo e o componente declarativo. O componente declarativo contém o conteúdo declarativo a ser exibido ao aluno. Por exemplo, enunciados de exercícios, explicações, exemplos etc. O modelo cognitivo ou procedural representa a expertise do STI. Tal conhecimento habilita o STI a resolver tarefas e problemas em sua área de domínio.

3.2.2 Módulo do Aluno

O módulo do aluno é responsável por armazenar as informações sobre o estudante, conforme ele interage com o STI. Tais informações estão relacionadas à qualquer interação do aluno com o sistema ou com as preferências do aluno, tais como planos ou caminhos para a solução utilizados na resolução de uma tarefa, a avaliação do desempenho ou habilidades do aluno na resolução de problemas ou até mesmo restrições que o aluno tenha violado durante a utilização do sistema, como pedir muitas dicas, preferência por cores na interface gráfica, preferência por métodos de ensino, entre outras informações. (ZHOU; EVENS, 1999).

3.2.3 Módulo Tutor

O módulo tutor é responsável por analisar as informações do aluno, contidas no módulo do aluno. Com base nesta análise, o módulo tutor é capaz de selecionar uma estratégia de ensino adequada para este aluno em específico. O módulo tutor deve ter acesso ao módulo especialista para identificar a relação entre os conteúdos a serem ensinados. Desta forma, o módulo tutor pode selecionar uma estratégia adequada com o conhecimento que o aluno possui. Por exemplo, caso um determinado conteúdo esteja relacionado com outros cinco e o aluno possui conhecimento em apenas um desses, o módulo tutor identifica que este conteúdo ainda não é adequado para o aluno e seleciona algum conteúdo mais adequado. Essa é uma característica importante em STIs, pois a partir dela, o aluno recebe um ensino individualizado.

3.2.4 Módulo de Comunicação

O módulo de comunicação é representado pela interface gráfica do usuário, do inglês *Graphical User Interface (GUI)*. A interface gráfica é o meio de comunicação entre o aluno e o STI. Todas as interações do aluno com a interface gráfica são enviadas ao módulo tutor, que por

sua vez armazena as informações relevantes no módulo do aluno. Tais informações podem ser utilizadas nas tomadas de decisão do módulo tutor. Essa comunicação entre o STI e o aluno deve ser simples e intuitiva, para que o aluno não tenha que perder tempo aprendendo a utilizar a interface e sim aprendendo com o STI.

3.3 STI PAT2Math

STI PAT2Math, do inglês *Personal Affective Tutor to Math*, é um sistema tutor inteligente para auxiliar alunos na aprendizagem de álgebra, mais especificamente em equações de primeiro grau. Através de uma ferramenta, chamada PATEquation, o aluno pode, a partir de uma equação inicial, entrar com passos de resolução da equação utilizando o teclado e o *mouse* do computador. O STI PAT2Math possui o conhecimento para verificar se cada passo de resolução de uma equação, inserido pelo aluno, está correto ou não. Ao final de cada interação entre o aluno e o sistema, o aluno recebe um *feedback* sobre sua interação. (SEFFRIN et al., 2010).

O STI PAT2Math possui um sistema de dicas inteligente. Ao solicitar uma dica ou ao cometer um erro na resolução de uma equação, o sistema provê uma dica específica para ajudar o aluno a corrigir ou solucionar um determinado passo de resolução. (SEFFRIN et al., 2012). Um experimento com quarenta e três alunos, da sétima série do ensino fundamental, apontou um aumento no desempenho dos alunos que utilizaram o STI PAT2Math. (JAQUES et al., 2013).

3.3.1 Comportamento do STI PAT2Math

Nessa seção é descrito o comportamento do STI PAT2Math, conforme seção 3.1. O laço externo do STI PAT2Math utiliza uma sequência fixa de planos. Cada plano contém equações do mesmo nível de dificuldade. Dentro de um plano, o aluno pode escolher a equação que deseja resolver. Porém, para um aluno ir de um plano para outro, ele deve concluir todas as tarefas do plano atual e o próximo plano será desbloqueado para o aluno.

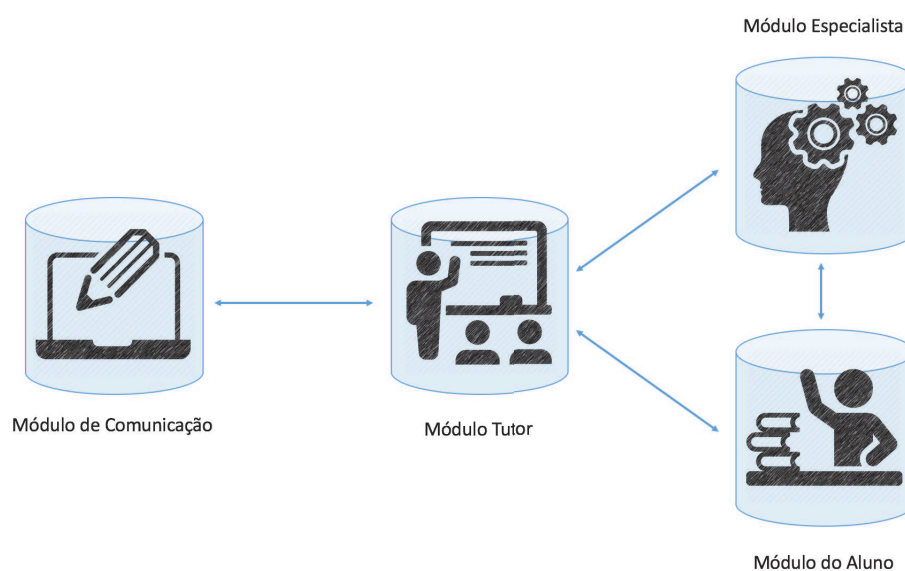
O laço interno do STI PAT2Math é representado pela ferramenta PATEquation. O PATEquation é um editor inteligente de equações, no qual o aluno pode resolver as equações passo a passo, inserindo-as no sistema com o uso do teclado e *mouse* do computador ou com o uso da escrita à mão em dispositivos *touchscreen*. (MORAIS; JAQUES, 2017). Nessa, o aluno recebe uma equação inicial e entra com passos de resolução para solucionar esta equação. A cada passo que o aluno inserir, o sistema irá verificar se o passo está correto ou não e fornecer um *feedback* ao aluno. Caso o passo esteja incorreto, o sistema irá informar o aluno o porquê do passo estar errado. Outra característica do laço interno é de fornecer dicas para resolução do próximo passo. A cada iteração do laço interno, o módulo do aluno é atualizado com novas estimativas do conhecimento do aluno na unidade de conhecimento em que o passo foi dado. (SEFFRIN; JAQUES, 2015).

Por exemplo, caso o aluno tenha inserido o passo correto em uma equação que envolvia a subtração, o sistema identifica que o aluno está aumentando seu conhecimento em subtração. Caso contrário, o sistema infere que o aluno talvez precise realizar mais exercícios relacionados à subtração para reforçar esse conteúdo.

3.3.2 Arquitetura do STI PAT2Math

O STI PAT2Math também segue a arquitetura clássica dos STIs, descrita na seção 3.2. (JAQUES et al., 2013). A Figura 2 ilustra a arquitetura do STI PAT2Math.

Figura 2 – Arquitetura do STI PAT2Math.



Fonte: Adaptado de Jaques et al. (2013).

No STI PAT2Math, o módulo especialista armazena o conhecimento necessário para resolver equações de primeiro grau, representadas através de regras de produção. Além disso, ele também é capaz de verificar se um passo de resolução inserido pelo aluno no sistema está correto ou não, dada uma equação inicial. (SEFFRIN; RUBI; JAQUES, 2011; JAQUES et al., 2013). O módulo especialista se comunica com o módulo do aluno para obter as informações do aluno e para salvar novas informações sobre este aluno, conforme ilustrado na Figura 2. O módulo especialista também se comunica com o módulo tutor, que envia os passos de resolução do aluno para a avaliação do módulo especialista e recebe um conjunto de informações, que pode incluir as operações utilizadas para a resolução de um determinado passo de resolução, a informação se o passo está correto ou não e uma possível dica ou mensagem.

O módulo do aluno do STI PAT2Math é responsável por armazenar as informações do aluno, que inclui todas as interações do aluno com o sistema. Essas interações podem ser o tempo que o aluno levou para resolver uma determinada equação, quantas vezes o aluno acertou ou errou equações que envolviam determinado conhecimento, em quais conteúdos o aluno pediu dicas

ao sistema etc. Em cada passo de resolução de uma equação, o aluno aplica vários conceitos matemáticos. O módulo do aluno armazena as estimativas do conhecimento do aluno em cada um desses conceitos utilizados para a resolução das equações. Essas estimativas são obtidas através do modelo cognitivo, inserido no módulo especialista, pois este informa quais foram as regras de produção utilizadas pelo aluno na resolução dos exercícios. A partir do módulo do aluno, o módulo tutor é capaz de inferir quais os próximos conteúdos que devem ser apresentados ao aluno. Tal ligação é ilustrada na Figura 2.

O módulo tutor é responsável por receber os passos de resolução das equações, enviados pelo módulo de comunicação, e retornar um *feedback* para ele. Cada um desses passos é repassado ao módulo especialista, que verifica o passo e fornece um *feedback* ao módulo tutor, conforme ilustrado na Figura 2.

O módulo de comunicação do STI PAT2Math é representado pela ferramenta PATequation e pela interface de exibição dos planos ao aluno. A versão atual do PATequation foi desenvolvida como uma ferramenta *web*. Tal estratégia foi utilizada para solucionar uma série de problemas de usabilidade na antiga versão do PATequation, que utilizava de um applet. (MORAIS; JAQUES, 2013). A Figura 3 ilustra a última versão da ferramenta PATequation.

Figura 3 – Ferramenta PATequation.

D: Feedback de erro fornecido ao aluno

The screenshot shows the PAT2Math interface. On the left is a sidebar with a navigation menu: Básico, Intermediário, Avançado, Expert, and Fase 30: Frações... (highlighted). Below the menu is a list of math problems with checkmarks. The main area displays a math problem: $x + x = 15$ with denominators 3 and 2. Below it, the student's steps are shown: $2x + 3x = 15$, $5x = 15$, and $5x = 15/6$. A feedback box (D) asks: "Você viu que tem frações nesta equação? Como você deve fazer a equivalência delas?". A hint box (E) says: "Observe que as frações a serem operadas possuem denominadores diferentes. Para realizar as operações nos numeradores, (parte de cima da fração), você precisa transformar as frações de forma". A score box shows "20 de 150 pontos" and "Equações concluídas: 1 / 10". The interface also shows "Erros gratuitos disponíveis: 9" and "19 dicas gratuitas disponíveis".

E: Botão de dica e quadro com a dica, possibilitando o aluno a solicitar dicas

C: Feedback fornecido a cada passo de resolução

B: Passos de resolução inseridas pelo aluno

A: Equação Inicial

F: Conteúdos

Fonte: Elaborado pelo autor.

Na Figura 3, o quadro A ilustra uma equação inicial. Tal equação é fornecida pelo sistema

ao aluno, sendo que ele não pode alterá-la. A partir da equação inicial, o aluno pode entrar com passos de resolução, conforme ele identificar adequado, para encontrar a solução da equação. O quadro B ilustra os passos de resolução inseridos pelo aluno. O sistema identifica os passos corretos e permite a inserção de um próximo passo. Para cada passo da resolução que o aluno insere no sistema, um *feedback* mínimo é fornecido ao aluno. O quadro C ilustra o caso em que este *feedback* é positivo, indicando que o passo inserido pelo aluno está correto. Caso o passo inserido pelo aluno esteja incorreto, o sistema fornece esse *feedback* juntamente com uma mensagem de ajuda, conhecido como *error-feedback*. O quadro D ilustra uma mensagem de ajuda fornecida pelo sistema, referente ao último passo de resolução inserido pelo aluno e que está incorreto. Caso o aluno desejar dicas para resolver o próximo passo, se em algum determinado momento ele não souber como prosseguir na resolução de uma equação, ele pode solicitar uma dica ao sistema. O quadro E ilustra o botão de dicas, no qual o aluno pode clicar a qualquer momento que lhe parecer adequado solicitar uma dica, e o quadro em forma de “*post it*”, no qual a dica é exibida ao aluno. E, por último, o quadro F ilustra os conteúdos a serem ou já vistos pelo aluno. Neste exemplo todos os planos já estão desbloqueados. Mas, eles vão sendo desbloqueados conforme o aluno vai progredindo e concluindo todas as equações de cada plano. Nesta versão mais recente da interface *web* do PAT2Math, foram inseridos componentes de gamificação, como pontuação, *ranking* da turma, erros e dicas gratuitas. Tal estratégia tem sido utilizada para evitar que os alunos abusem do sistema de dicas e de *feedbacks*, comportamento conhecido como *gaming the system*, e ao mesmo tempo evitar que os alunos evitem de pedir ajuda, comportamento conhecido como *help refusal*. Mais informações podem ser encontradas em Azevedo, Morais e Jaques (2018).

4 MINERAÇÃO DE DADOS EDUCACIONAIS

Este capítulo tem como objetivo descrever os conceitos gerais sobre mineração de dados educacionais. Para isso, algumas definições mais abrangentes devem ser levadas em consideração, como *big data* e mineração de dados. Depois, um *overview* sobre os principais métodos e algoritmos de mineração de dados educacionais e a diferença deste com o conceito de *Learning Analytics* é apresentado.

4.1 *Big Data*

Big data tem sido definido como três dimensões de desafios em gerenciamento de dados, também conhecido como os três Vs, que são Volume, Variedade e Velocidade. (LANEY, 2001). Definições de grandes empresas, como a Gartner, também são muito aceitas: “*Big data* é um conjunto de informações de alto volume, alta velocidade e alta variedade, que demanda formas inovadoras e econômicas de processamento de informações para melhorar percepção e tomada de decisões.” (GARTNER, 2013).

Conforme apresentado por Gandomi e Haider (2015), volume se refere à magnitude dos dados, ou seja, a quantidade de informações. Variedade se refere à heterogeneidade estrutural em uma base de dados, uma vez que pode-se utilizar tipos de dados estruturados, semi-estruturados e não estruturados. Atualmente, a maior parte dos dados é proveniente de textos, imagens, áudios, vídeos ou até mesmo de sensores, os quais são exemplos de dados não estruturados. E, por fim, velocidade se refere à taxa em que os dados são gerados e à velocidade em que esses dados devem ser analisados e utilizados. Por exemplo, aplicações *mobile* podem, atualmente, fazer *streaming* de vídeo em tempo real, enviando e recebendo informações de milhares de pessoas ao redor do mundo ao mesmo tempo. Bancos de dados tradicionais não são capazes de lidar com esse volume de dados não estruturados, muito menos para aplicações de tempo real. Nesse contexto, encontra-se a área chamada de *big data*.

Wu et al. (2014) apresentam o teorema HACE, que define *big data* como sendo dados de alto volume Heterogêneos, de fontes Autônomas com controle distribuído e descentralizado e visa explorar dados Complexos e que Evoluem a relação entre os dados. Assim, com essas características, a descoberta de conhecimento se tornam um desafio, do inglês *knowledge discovery*. Autores têm enfatizado a necessidade de novas abordagens de processamento para manipular essa grande quantidade de dados (KWON; SIM, 2013), sendo essa uma das premissas da área de mineração de dados.

4.2 Mineração de Dados

Segundo Hand (2007), “a mineração de dados é a análise de conjuntos de dados observacionais (geralmente grandes) para encontrar relacionamentos insuspeitos e para resumir os dados

de maneiras novas que são compreensíveis e úteis para o proprietário dos dados.” A mineração de dados, do inglês *data mining*, geralmente é vista como uma subárea da descoberta de conhecimento em bases de dados, do inglês *Knowledge Discovery from Database (KDD)*. (HAN; PEI; KAMBER, 2011). Ainda, segundo esses autores, esse processo de descoberta (KDD) possui uma sequência iterativa dos seguintes passos:

1. limpar os dados, consiste em remover ruídos e dados inconsistentes;
2. integrar os dados, quando mais de uma fonte de dados é utilizada é necessário fazer a integração de forma consistente;
3. seleção de dados, onde os dados relevantes para a tarefa de análise são recuperados da base de dados;
4. transformação dos dados, onde os dados são transformados e consolidados, de maneiras apropriadas para a mineração, realizando operações de agregação e redução;
5. mineração dos dados, processo essencial onde métodos inteligentes são utilizados para a extração de padrões de dados;
6. avaliação do padrão, identificar os padrões realmente interessantes de acordo com as métricas;
7. apresentação do conhecimento, onde técnicas de visualização e representação do conhecimento são utilizadas para apresentar o conhecimento minerado para os usuários.

A mineração de dados herda técnicas de vários domínios, conforme ilustrado na Figura 4. Além disso, Han, Pei e Kamber (2011) também descrevem as principais funcionalidades da mineração de dados, sendo elas: categorização e discriminação, padrões frequentes, associações e correlações, classificação e regressão, clusterização e análise de *outlier*. Essas funcionalidades ou métodos serão descritos com mais detalhes na seção 4.3, tendo seu foco voltado para a educação.

A mineração de dados tem influenciado na tomada de decisões de organizações, com base na descoberta de novas informações úteis, criando uma grande lacuna entre empresas que usam e não usam a mineração de dados para a tomada de decisões. (KIRON et al., 2012). Assim, as metodologias de análise e mineração de dados, que extraem informações úteis e relevantes de grandes bases de dados, têm transformado e influenciado diversas áreas de conhecimento. (SUMMERS; COLLABORATION et al., 2000; COLLINS; MORGAN; PATRINOS, 2003). Quando essas técnicas de mineração são aplicadas em contextos educacionais, estes levam os nomes de *Learning Analytics (LA)* e mineração de dados educacionais, do inglês *Educational Data Mining (EDM)*. (BAKER; INVENTADO, 2014). O foco principal desse capítulo é apresentar o conteúdo de mineração de dados educacionais, onde esta dissertação está inserida. Assim, a próxima seção apresenta os principais conceitos, métodos, algoritmos e definições da área. Por fim, um comparativo entre mineração de dados e *learning analytics* é apresentado.

Figura 4 – Técnicas de vários domínios.



Fonte: Adaptado de Han, Pei e Kamber (2011).

4.3 Mineração de Dados na Educação

A mineração de dados educacionais, do inglês *Educational Data Mining* (EDM), “é uma disciplina emergente, preocupada com o desenvolvimento de métodos para explorar os tipos exclusivos de dados provenientes de ambientes educacionais e o uso desses métodos para entender melhor os alunos e as configurações em que eles aprendem.” (BAKER; YACEF, 2009).

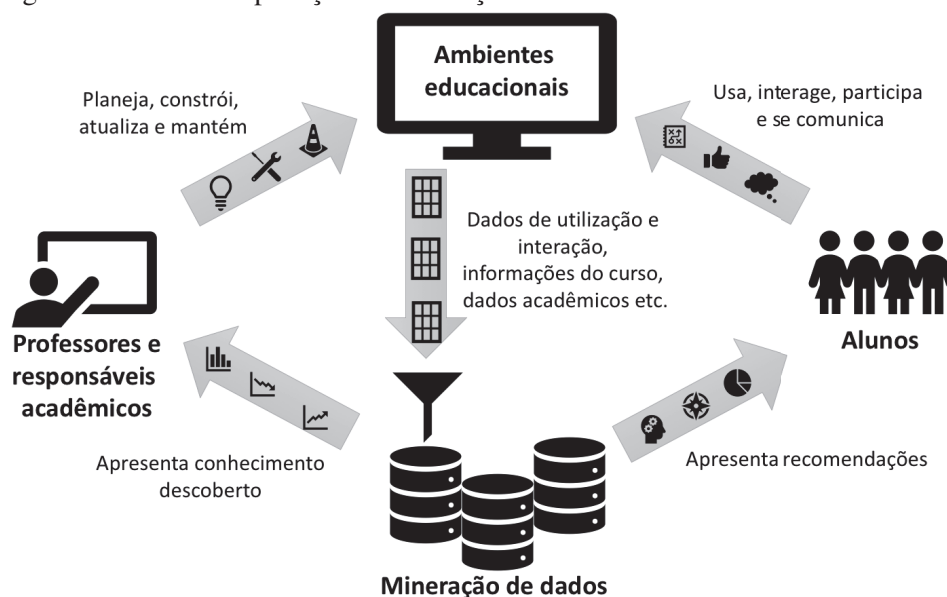
Vários trabalhos têm sido apresentados na comunidade científica ressaltando aspectos importantes da mineração de dados na pesquisa educacional, como o de Romero e Ventura (2007) que apresentam uma revisão sobre aplicações de técnicas de mineração de dados em ambientes educacionais entre 1995 e 2005. Outro exemplo é o trabalho de Castro et al. (2007), que apresenta uma revisão sobre técnicas de mineração que visam avaliar o material de aprendizagem, bem como acessar a performance de aprendizagem e o comportamento dos alunos, especificamente na área de ensino a distância baseado em computador, do inglês *e-learning*. Não apenas em *e-learning*, mas revisões em ambientes de aprendizagem virtual foram apresentadas. (MOHAMAD; TASIR, 2013). Também, Baker e Yacef (2009) apresentam uma revisão ressaltando as tendências e deslocamentos de tópicos de pesquisa em EDM, as quais são descritas no capítulo 5.

Romero e Ventura (2007) enfatizam que, embora os métodos de descoberta de conhecimento em áreas como comércio eletrônico e *e-learning* são similares, há diferenças importantes para serem levadas em consideração, como o domínio, os dados obtidos do sistema (*logs*), o objetivo

(aumentar os lucros vs. melhorar a aprendizagem) e as técnicas, onde os dados educacionais precisam de cuidados especiais, conforme apresentado por Donnellan e Pahl (2002) e Li e Zaiane (2004).

Ainda sobre o trabalho de Romero e Ventura (2007), os autores apontam que a aplicação e utilização de mineração de dados em ambientes educacionais é um ciclo, conforme ilustrado na Figura 5, onde a descoberta de novas informações deve entrar nesse ciclo, facilitando e proporcionando uma melhora na aprendizagem como um todo, e não apenas transformando dados em conhecimento. A aplicação da mineração de dados em ambientes educacionais pode ser direcionada a diferentes atores (ZORRILLA et al., 2005): orientado aos estudantes (HERAUD; FRANCE; MILLE, 2004; LU, 2004; TANG; MCCALLA, 2005), orientado aos professores (MERCERON; YACEF, 2004; MOR; MINGUILLÓN, 2004; ROMERO; VENTURA; DE BRA, 2004) e orientado à administração acadêmica das instituições. (LUAN, 2002; URBANCIC; SKRJANC; FLACH, 2002; GROB; BENSBERG; KADERALI, 2004).

Figura 5 – Ciclos da aplicação de mineração de dados em ambientes educacionais



Fonte: Adaptado de Romero e Ventura (2007).

Embora os autores ressaltem que não apresentam uma exaustiva seleção de métodos de mineração, como outros trabalhos (BAKER; YACEF, 2009; ROMERO; VENTURA, 2007, 2010; SCHEUER; MCLAREN, 2012), em seu trabalho Baker e Inventado (2014) apresentam os métodos que estão sendo mais utilizados pela comunidade EDM. Esses são divididos em quatro principais métodos: métodos de predição, mineração de relacionamento, descoberta de estrutura e descoberta com modelos, que serão descritos com mais detalhes nas sub-seções seguintes.

4.3.1 Métodos de Predição

Os métodos de predição, do inglês *prediction methods*, têm como objetivo inferir o valor de uma variável, baseado na combinação de várias informações. Para isso, esses tipos de métodos utilizam dados rotulados, ou seja, um conjunto de dados que possui o valor de saída desejado. Esses métodos são divididos em (i) classificação, (ii) regressão e (iii) estimativa de conhecimento latente (a versão mais recente do método de regressão), descritos abaixo com base em Baker e Inventado (2014). Pesquisadores têm usado esses métodos em vários trabalhos, para prever o conhecimento dos alunos (CORBETT; ANDERSON, 1995), os estados afetivos (D'MELLO et al., 2008; BAKER et al., 2012), para prever os futuros resultados do aluno (DEKKER; PECHENIZKIY; VLEESHOUWERS, 2009; PEDRO et al., 2013) etc.

- i) Nos **métodos de classificação**, do inglês *classification*, a variável a ser inferida pode assumir valores binários ou categóricos. Os algoritmos mais usados para esse tipo de método de classificação são as árvores de decisão, *random forests*, regras de decisão, regressão passo-a-passo e logística. Para a validação destes testes, normalmente é utilizada a técnica de *cross-validation* (validação cruzada em português), que consiste em utilizar parte do conjunto de dados para treinamento e parte para teste, de forma repetida e sistemática. É comum em pesquisas de EDM, onde a fonte de dados são os alunos ou usuários de ambientes educacionais, que essa validação seja realizada a nível do aluno, ou seja, que os dados de cada aluno sejam mantidos juntos e a validação seja baseada no conjunto inteiro de dados do aluno. As métricas mais comuns para esses classificadores são A'/AUC (HANLEY; MCNEIL, 1982), Kappa (COHEN, 1960), *precision e recall*. (DAVIS; GOADRICH, 2006).
- ii) Nos **métodos de regressão**, do inglês *regression*, a variável a ser inferida é contínua. Os métodos mais populares são a regressão linear e as árvores de regressão. O alto grau de ruídos nos dados, assim como os múltiplos fatores explicativos, fazem com que seja mais comum o uso desses algoritmos mais conservadores, deixando as redes neurais e *support vector machines* de lado. As métricas de avaliação dos métodos de regressão podem ser as mesmas dos classificadores, mas também utilizam métricas de correlação linear ou a raiz do erro quadrático médio, do inglês *root mean squared error* (RMSE).
- iii) Outro método de predição de grande importância em EDM é a **estimativa de conhecimento latente**, do inglês *latent knowledge estimation*. O uso da palavra latente refere ao fato de que o conhecimento não pode ser medido de forma direta, apenas inferido a partir do desempenho do aluno, ou seja, da forma com que o aluno vai aprendendo e progredindo dentro do sistema. A inferência do conhecimento do aluno pode ser utilizada para ajustar as estratégias curriculares, ou ainda realizar outros tipos de intervenções (ROLL et al., 2007) e também pode ser útil para os professores. (FENG; HEFFERNAN, 2007). Os algoritmos

mais usados nesses tipos de métodos têm sido redes Bayesianas, *Bayesian Knowledge Tracing*, regressão logística e ainda combinando múltiplas abordagens em grandes conjuntos de dados.

4.3.2 Mineração de Relacionamento

A mineração de relacionamento, do inglês *relationship mining*, conforme o nome sugere, tem como objetivo identificar relações entre duas ou mais variáveis dentro de um conjunto de dados. Há quatro tipos de mineração de relacionamento comumente vistos em EDM: (i) mineração de regras de associação, (ii) mineração de padrões sequenciais, (iii) mineração de correlação e (iv) mineração de dados causais.

- i) A **mineração de regras de associação**, do inglês *association rule mining*, tem como objetivo descobrir regras do tipo se-então. Assim, a ideia é que ao identificar um determinado conjunto de valores, quando uma regra se-então for identificada, pode-se presumir que um outro conjunto de valores será encontrado também. Pesquisas têm usado esse tipo de método para predição de performance satisfatória do aluno (BEN-NAIM; BAIN; MARCUS, 2009) e para identificar tendências em conjuntos de erros. (MERCERON; YACEF, 2005).
- ii) A **mineração de padrões sequenciais**, do inglês *sequential pattern mining*, tem como objetivo identificar associações temporais entre eventos. Pesquisas vêm utilizando tal abordagem para descobrir caminhos no comportamento colaborativo dos colegas de um grupo (PERERA et al., 2009) e para identificar o padrão de comportamento de procura de ajuda durante o tempo. (SHANABROOK et al., 2010).
- iii) A **mineração de correlação**, do inglês *correlation mining*, tem como objetivo identificar correlações lineares, tanto negativas quanto positivas, entre diferentes variáveis. Herda principalmente conceitos da estatística. Tem sido utilizada na identificação de correlação entre atitudes e comportamento dos alunos (ARROYO; WOOLF, 2005; BAKER et al., 2008) e na relação entre a usabilidade da interface gráfica de um STI com determinados comportamentos do aluno. (BAKER et al., 2009).
- iv) A **mineração de dados causais**, do inglês *causal data mining*, tem como objetivo encontrar se um evento foi causado por outro evento; causalidade. Essa tarefa é realizada por meio da análise de covariância entre variáveis do conjunto de dados. Alguns trabalhos têm utilizado para identificar fatores que podem fazer um aluno ir mal nas aulas (FANC-SALI, 2012) ou ainda para analisar se diferentes condições de estudo podem interferir na aprendizagem. (RAU; SCHEINES, 2012).

4.3.3 Descoberta de Estrutura

A descoberta de estrutura, do inglês *structure discovery*, tem como objetivo identificar algum tipo de estrutura sem a obrigatoriedade de se ter dados rotulados. Os principais algoritmos são: (i) agrupamento, (ii) análise de fator e (iii) descoberta de estrutura de domínio.

- i) O **agrupamento**, do inglês *clustering*, como o nome já diz, visa agrupar os dados conforme algum valor em comum, normalmente dividindo o conjunto de dados em *clusters*, ou grupos. O tamanho e a quantidade desses grupos pode variar de acordo com a necessidade e objetivo da pesquisa. Os principais algoritmos são os de abordagem hierárquica, como o *Hierarchical Agglomerative Clustering* (HAC), e os não-hierárquicos, como *k-means*, *gaussian mixture modeling* e agrupamento espectral. Exemplos de trabalhos que utilizam essa abordagem são Amershi e Conati (2009) e Beal, Qu e Lee (2006).
- ii) A **análise de fatores**, do inglês *factor analysis*, tem como objetivo identificar variáveis que se agrupem naturalmente, dividindo o conjunto de variáveis em um conjunto de fatores. Normalmente, é utilizado em EDM para a redução da dimensão do problema. Os principais algoritmos são análise de componentes principais e análise de componentes principais de famílias exponenciais. Um exemplo de pesquisa utilizando essa abordagem é o de Minaei-Bidgoli et al. (2003).
- iii) A **descoberta de estrutura de domínio**, do inglês *domain structure discovery*, visa identificar quais características mapeiam certas habilidades específicas entre os alunos. A abordagem mais conhecida é o *Q-Matrix*. (BARNES, 2005). Porém, outros algoritmos podem ser usados, como algoritmos automatizados e abordagens de descoberta de processos baseado no julgamento humano. Exemplos de trabalhos nessa abordagem são Desmarais et al. (2011); Koedinger, McLaughlin e Stamper (2012).

4.3.4 Descoberta com Modelos

A descoberta com modelos, do inglês *discovery with models*, tem como objetivo utilizar modelos desenvolvidos com base em outras abordagens, como predição ou agrupamento, como componente em um segundo modelo ou análise. Esse tipo de abordagem pode generalizar um modelo em diferentes contextos. Por exemplo, Baker e Gowda (2010) usaram modelos preditivos de comportamentos (*off-task*, descuido e *gaming the system*) durante um ano letivo para analisar a diferença dos comportamentos em escolas urbanas, rurais e suburbanas. Aqui não houve o desenvolvimento de novos modelos com base nos modelos avaliados. Porém, uma análise de diferentes populações foi realizado com base em modelos preditivos já desenvolvidos, o que gerou uma descoberta de novos conhecimentos com base em modelos previamente desenvolvidos e testados.

Essa abordagem tem sido muito utilizada na comunidade EDM. Porém, não é uma estratégia de mineração de dados utilizada amplamente, em diferentes áreas. Um uso comum dessa abordagem é quando um modelo inicial de predição passa a ser a variável preditora de um novo modelo preditivo.

4.4 *Learning Analytics* e Mineração de Dados Educacionais

Conforme proposto pela *Society for Learning Analytics* (SoLAR), “*Learning Analytics* (LA) é a medição, coleta, análise e relato de dados sobre os alunos e seus contextos, com o objetivo de compreender e otimizar o aprendizado e os ambientes em que ele ocorre.” (SOCIETY FOR LEARNING ANALYTICS RESEARCH, SOLAR). As áreas de LA e EDM compartilham de várias características e possuem interesses e objetivos em comum, porém ambas as áreas possuem orientações metodológicas, ideologias e tecnologias distintas, conforme apontado por Baker e Inventado (2014); Papamitsiou e Economides (2014); Ihantola et al. (2015); Calvet Liñán e Juan Pérez (2015). Segundo Siemens e Baker (2012), LA e EDM se diferem em alguns aspectos bem específicos. EDM tende à utilização de paradigmas automatizados para a análise de dados, enquanto que LA tem seu foco em deixar o julgamento final para os humanos. Além disso, a EDM tem um foco mais reducionista, gerando modelos e relações entre eles, enquanto LA tem um foco mais holístico, visando entender o sistema como um todo. Por fim, EDM possui uma ênfase em adaptação automatizada, contra um suporte à intervenção humana de LA.

Embora haja diferenças, Baker e Inventado (2014) concluem que ter múltiplas comunidades discutindo e descobrindo como desenvolver e aplicar novos métodos pode contribuir positivamente e mais efetivamente na pesquisa educacional baseada em dados. Ainda, os autores afirmam que não é uma questão de qual método é o melhor, mas sim quais ou qual o método é o mais útil para determinada aplicação e contexto.

5 TRABALHOS RELACIONADOS

Este capítulo tem como objetivo apresentar uma visão geral do estado da arte em detecção e predição de estados afetivos em ambientes de aprendizagem com o uso de mineração de dados. Para isso, foi realizado uma busca, iniciada por um conjunto de palavras chaves nos principais veículos de divulgação da área e após seguindo uma metodologia *snowballing*, por trabalhos relacionados. O seguinte conjunto de palavras chaves foi utilizado para a busca inicial de trabalhos relacionados: *educational data mining, data mining AND education, sensor-free, affect, emotion, affective state, detect, recognize, infer, predict, learning environment, tutoring system, educational software, educational system*. Essa busca foi realizada nas bases da ACM, IEEE, Web of Science, Springer, Scopus, Science Direct e Journal of Educational Data Mining. Após identificar um conjunto inicial de trabalhos, a metodologia *snowballing* foi aplicada, verificando a referência à novos trabalhos com base nos trabalhos previamente selecionados. Um total de 327 artigos foram analisados. Porém, apenas os trabalhos que visavam detectar estados afetivos em ambientes educacionais de aprendizagem por meio de dados de *logs* foram selecionados. Visando apresentar o estado da arte mais recente, apenas trabalhos mais atuais foram selecionados para essa análise, publicados dentro dos últimos cinco anos. Com exceção do primeiro trabalho, que foi publicado em 2008, mas que foi o trabalho pioneiro da área, e do segundo, que foi publicado em 2012, mas que apresenta mais características em comum com essa dissertação.

(Artigo [1]) O trabalho considerado como pioneiro na área de EDM aplicado à detecção de estados afetivos é o de D’mello et al. (2008). Nesse trabalho, os autores modelam os estados afetivos dos alunos ao utilizarem o STI AutoTutor em um laboratório. Os autores relataram que obtiveram resultados positivos na detecção de frustração, tédio, confusão e *flow*, quando comparados com os rótulos gerados pela análise manual de vídeos. Porém, nesse estudo os autores focam a análise dos resultados em diferenciar cada estado afetivo detectado (tédio, confusão, frustração e *flow*) do estado neutro. Assim, os resultados obtidos ao distinguir frustração de neutro foi aproximadamente $Kappa = 0,4$ e para tédio, confusão e *flow* aproximadamente $Kappa = 0,2$. Porém, quando submetidos a distinguir um estado afetivo do outro, seu desempenho caiu para uma média de $Kappa = 0,163$. Além disso, a validação foi realizada a nível de observação e não a nível de estudante, o que provê menos informação para os detectores generalizarem os resultados para novos alunos.

(Artigo [2]) Baker et al. (2012) coletaram *logs* de utilização de 89 alunos durante dois dias de coleta de dados com o Cognitive Tutor Algebra I, um STI de matemática, voltado para o ensino de equações algébricas. Durante a coleta, dois especialistas coletaram dados de afeto dos estudantes por meio de observação (estando na mesma sala que o aluno) e com o auxílio de um aplicativo *mobile* (protocolo BROMP), resultando em um total de 408,51 minutos de observações, provenientes de 763 observações ($M = 8,57$ observações por estudante). As observações tinham duração de até 20 segundos, sendo que se o especialista detectasse um estado afetivo ou comportamento antes disso, ele ia para a próxima observação. Os especialistas usaram a visão

periférica para codificar os estados afetivos dos alunos, capturando informações de ações, falas, expressões faciais, linguagem corporal e interação com professores ou colegas. Os estados afetivos eram classificados em cinco categorias: tédio, confusão, *flow*, frustração e “?” (empregado quando qualquer outro afeto, diferente dos quatro definidos, era identificado ou quando não era possível codificar afeto). Uma informação relevante é que quando dois estados afetivos fossem detectados em uma mesma observação, apenas a primeira era considerada. A partir dos dados coletados de *logs* e dos especialistas, os autores geraram 232 características (incluindo número de passos corretos ou incorretos, pedidos de ajuda, duração de ações no sistema, tempo na resolução de passos e na equação, sequências de ações, estimativa de conhecimento etc.), as quais foram utilizadas no desenvolvimento de um detector automático de estados afetivos. Foram desenvolvidos quatro detectores, um para cada estado afetivo. Cada um dos detectores foi avaliado usando a técnica de validação cruzada (*k-fold cross-validation*), com $k = 6$, utilizado a validação a nível de aluno. No desenvolvimento dos detectores, os autores utilizaram os algoritmos de classificação mais utilizados pela comunidade, incluindo J48 decision trees, step regression, Naive Bayes e REP-Trees. Além disso, os autores também realizaram uma seleção das características, identificando as que mais se relacionavam com cada um dos estados afetivos. Para a verificação da qualidade dos detectores, os autores utilizaram Cohen’s Kappa (COHEN, 1960) e A' . (HANLEY; MCNEIL, 1982). Como resultados, os autores apresentam que os melhores algoritmos para detecção de cada um dos estados afetivos foram, respectivamente: (i) detecção do *flow* o K^* ($A' = 0,71$ e $Kappa = 0,31$); (ii) para confusão o JRip ($A' = 0,99$ e $Kappa = 0,40$); (iii) para frustração o REPTree ($A' = 0,99$ e $Kappa = 0,23$), e, por fim, (iv) para o tédio o Naive Bayes ($A' = 0,69$ e $Kappa = 0,28$). Como conclusões, os autores apontam que os detectores obtiveram resultados melhores que outros trabalhos relacionados. (D’MELLO et al., 2008; CONATI; MACLAREN, 2009; SABOURIN; MOTT; LESTER, 2011; LEE et al., 2011). Eles acreditam que a seleção de variáveis pode ter impactado positivamente nos resultados. Porém, apenas 30% de precisão foi alcançado, quando comparados a detecção perfeita, a qual é provavelmente impossível.

(Artigo [3]) Em outro trabalho, também envolvendo a detecção do afeto dos alunos em um STI de matemática, Pardos et al. (2014) estimam a probabilidade do aluno estar em um estado de tédio, *flow*, confusão e frustração e estimam também a probabilidade dos comportamentos “*off-task*” e “*gaming the system*” a cada ação dele no sistema. Com base nos estados afetivos detectados, os autores também desenvolveram um modelo para prever o resultado que os alunos alcançariam nos testes, baseado na detecção do afeto, comportamento e performance dos alunos utilizando o STI. Para a construção dos detectores, dois conjuntos de dados foram utilizados. Um contendo dados de 229 estudantes de uma única escola e outro contendo dados de 1393 alunos de várias escolas da mesma cidade. O STI ASSISTments, um sistema tutor *web* baseado em passos, foi utilizado por turmas de matemática do quarto ao décimo ano, durante duas horas, duas vezes por semana, durante um ano letivo. Tais dados totalizaram 810 mil *logs* de utilização. Para a detecção do afeto e dos comportamentos dos alunos, os auto-

res utilizaram o mesmo processo do trabalho de Baker et al. (2012), ou seja, 20 segundos de observação onde dois especialistas analisavam informações de ações, falas, expressões faciais, linguagem corporal e interação com professores ou colegas. Novamente, apenas o primeiro estado era considerado quando dois fossem detectados na mesma observação. Ao final, os dados de afeto e comportamento eram sincronizados com os dados de *log* do sistema. Um total de 172 características foram utilizadas no desenvolvimento dos detectores (incluindo ações anteriores, número de tentativas em cada problema, número de pedidos de ajuda, número de ações incorretas, assim como a soma total, mínimo, máximo e média desses valores também foram inclusos). Novamente, foi desenvolvido um detector para cada estado afetivo e validado por meio de validação cruzada com alunos. Os mesmos oito algoritmos foram utilizados para testes (J48 decision trees, step regression, Naive Bayes e REP-Trees). Além disso, os autores também realizaram uma seleção das características, identificando as que mais se relacionavam com cada um dos estados afetivos. Uma diferença das características utilizadas nesse trabalho em relação ao de Baker et al. (2012), é que neste os autores não utilizaram informações sobre a estimativa do conhecimento dos alunos. Para a verificação da qualidade dos detectores, os autores utilizaram Cohen's Kappa (COHEN, 1960), A' (HANLEY; MCNEIL, 1982) e uma terceira métrica F-measure (VAN RIJSBERGEN, 1974) e validaram os modelos utilizando *5-fold cross-validation* a nível de aluno. Como resultados, os autores apresentam que o melhor algoritmo para detecção de *flow* foi o K* ($A' = 0,678$ e $Kappa = 0,358$). Para detecção de confusão foi o J48 ($A' = 0,736$ e $Kappa = 0,274$). Na detecção de frustração foi o Naive Bayes ($A' = 0,682$ e $Kappa = 0,324$). E, por fim, para o tédio foi o JRip ($A' = 0,632$ e $Kappa = 0,229$). Os autores também relatam os preditores de comportamento, porém esses não são relevantes nessa dissertação. Como conclusões os autores ressaltam que os alunos entediados ou confusos tendem a irem mal nos testes. Porém, tédio e confusão em *scaffolding problems* foram associados positivamente, indicando que esses dois estados são dependentes de contexto. Um resultado inesperado foi a relação positiva entre frustração e aprendizagem. Os autores ainda argumentam que, embora os detectores de afeto e comportamento podem prever efetivamente a performance dos alunos nos exames, os resultados ainda não são conclusivos, pois representam apenas um ambiente de aprendizagem.

(Artigo [4]) Ocumpaugh et al. (2014) também utilizaram o STI ASSISTments para conduzir outro estudo. Neste estudo o objetivo era verificar se diferentes populações poderiam influenciar os detectores de estados afetivos. Para isso, três grupos de diferentes populações utilizaram o sistema, gerando dados de *logs*: urbano, rural e suburbano. A coleta de dados, tanto de *logs*, quanto dos estados afetivos, e os algoritmos utilizados para o desenvolvimento dos detectores foi feita da mesma forma que Baker et al. (2012) e Pardos et al. (2014) descreveram. Dos dados obtidos, os autores utilizaram 69 características para o treinamento dos detectores (incluindo informações temporais, características baseadas em habilidades, em número de erros, número de respostas corretas e número de dicas solicitadas). A principal diferença entre esses relatos é que o treinamento dos detectores foi dividido em modelos específicos para cada tipo de população

e estado afetivo. Ainda, cada um dos detectores desenvolvidos utilizou um conjunto específico de características, as quais não foram relatadas no artigo. Como resultados, os autores apontam que não é possível generalizar um modelo, construído com base em uma única população, para outros tipos de população. Mais especificamente, os autores sugerem que, da forma com que os estados afetivos dos alunos é refletido na forma com que eles interagem com o sistema, alunos de populações urbanas e suburbanas são mais similares do que alunos de população rural. Assim, foi desenvolvido um detector com base nos dados obtidos das três populações, que apresentou um resultado quase tão efetivo quanto o detector mais customizado para um determinado grupo. A vantagem é de que esse novo modelo é aplicável em uma escala muito maior.

(Artigo [5]) Wang, Heffernan e Heffernan (2015) apresentam três tentativas de melhorar os detectores de afeto do STI ASSISTments, desenvolvido por Ocumpaugh et al. (2014). Na primeira, os autores analisam o efeito de rótulos de habilidades que estavam faltando na base de dados. A segunda está relacionada com a seleção de características, adicionando quatro características novas sobre as turmas de alunos, e a terceira eles adicionam duas características que descrevem sobre os erros comuns dos alunos. A coleta de dados, tanto de *logs*, quanto de afeto, seguiu a mesma metodologia utilizada no trabalho de Ocumpaugh et al. (2014). Porém, diferentemente do trabalho de Ocumpaugh et al. (2014), nesse trabalho os autores utilizaram 58 características (incluindo características temporais, informações de habilidades, número de erros, número de respostas corretas e o número de pedidos de ajudas). Além disso, a média, mínimo, máximo e somatório dessas 58 características foram calculados, gerando um total de 232 características. Assim como os outros trabalhos, quatro estados afetivos foram considerados: tédio, *flow*, frustração e confusão; os quais foram considerados separadamente nos modelos de detecção. Os algoritmos também foram os mesmos dos outros trabalhos e foram aplicados por meio da utilização da ferramenta RapidMiner 5. Novamente, os quatro detectores foram avaliados e testados separadamente, utilizando as métricas Kappa e A'. Na primeira tentativa de melhora na detecção os autores inseriram rótulos faltantes de forma manual na base de dados e retreinaram os detectores. Como resultados, os autores relataram que houve uma melhora de apenas 3%, indicando que rótulos faltantes não interferem tanto nos resultados. Na segunda tentativa, os autores adicionaram quatro características que envolvem informações sobre a turma em que o aluno está presente: porcentagem de questões corretas pela turma, a porcentagem do aluno relativo a média de performance da turma, número de dados anteriores da turma, número de alunos da turma. Como resultado, a adição de informações sobre a turma melhorou os detectores de *flow* e tédio em 3,5% no valor médio de Kappa. E na terceira, duas características sobre os erros comuns dos alunos: a porcentagem da resposta entre os *logs* que responderam o problema e uma característica binária, dizendo se uma resposta é da classe de erros comuns. Como resultados, os autores relataram uma melhora de 5% na média Kappa geral dos detectores.

(Artigo [6]) No trabalho de Paquette et al. (2014) os autores realizam um estudo de como detectar automaticamente o afeto dos estudantes no ambiente virtual Inq-ITS. Neste ambiente, os alunos utilizam de interações simuladas para aprender conteúdos de ciências físicas, da vida

e da Terra, sendo esse um ambiente menos restrito do que os de resolução de problemas, permitindo mais comportamentos exploratórios. Os dados para esse estudo foram coletados de 326 alunos, de 11 turmas de oitavo ano, de três escolas diferentes. Para a coleta dos estados afetivos dos alunos, o mesmo método utilizado por Baker et al. (2012) e Pardos et al. (2014) foi utilizado, diferenciando no número de especialistas, que nesse estudo foram quatro. Um total de 4155 observações foram realizadas utilizando a visão periférica dos especialistas, conforme descrito anteriormente, seguindo o protocolo BROMP. Porém, após a validação das observações, apenas 1241 foram utilizadas na análise final. As observações foram integradas com os dados de *logs*, gerando um total de 127 características (incluindo mudanças nas variáveis de hipóteses, alterações nas variáveis de simulação, pausas de simulação, testes incompletos, execução de testes completos, todos os ensaios executados, interações durante o estágio de análise, tempo gasto em cada ação, número de ações, tempo de inatividade e todas as ações relevantes, probabilidade de conhecimento, entre outras). Foram desenvolvidos 12 detectores, sendo eles divididos em três estágios de investigação dentro do ambiente: hipótese, experimento e análise, sendo que para cada um desses, quatro detectores foram desenvolvidos, um para cada estado afetivo. A validação dos detectores foi realizada por meio de validação cruzada a nível do aluno. No desenvolvimento dos detectores, foram utilizados três algoritmos de classificação: *J48 decision tree*, *JRip* e *step regression*. Novamente, as métricas utilizadas foram *Kappa* e A' . Como resultados, os autores apresentam que a média dos detectores pela validação cruzada foi $Kappa = 0,354$ e $A' = 0,720$, o qual é um resultado melhor do que o apresentado por Pardos et al. (2014) com $Kappa = 0,296$ e $A' = 0,682$. Como conclusões, os autores relatam que as características foram utilizadas de forma diferente por cada um dos detectores, sendo que poucas características foram reusadas em diferentes detectores.

(Artigo [7]) Paquette et al. (2016) realizaram uma comparação entre diferentes fontes de dados para a detecção de estados afetivos. Basicamente, os autores realizaram um estudo comparando a detecção baseado em dados de *logs* (*sensor-free*) e na postura dos alunos, utilizando sensores específicos. Após a análise dos resultados, os autores identificaram que os detectores baseados puramente em dados de interação (*logs*) obtiveram melhores resultados do que os detectores baseados em dados de sensores na detecção dos estados afetivos. Porém, os autores ressaltam que os resultados apontam para uma continuidade na pesquisa e que ainda há muitas questões em aberto relacionadas à área.

(Artigo [8]) Mais recentemente, Jiang et al. (2018) realizaram um estudo comparando duas estratégias de treinamento de detectores: *expert feature engineering* e *deep neural networks*. Primeiramente, os autores realizaram um estudo para a coleta de dados de *logs*. Nesta coleta, 93 alunos da sexta série de quatro turmas de ciências utilizaram o ambiente Betty's Brain para aprender ciências e completar desafios construindo mapas causais para descrever fenômenos científicos. Com base nos *logs* de interação dos alunos com o sistema, os autores desenvolveram detectores para os estados afetivos concentração engajada, confusão, frustração, tédio e surpresa. Para a coleta dos rótulos de treinamento os autores utilizaram o protocolo BROMP,

conforme mencionado nos trabalhos relacionados anteriores. Um total de 5212 anotações foram realizadas, levando em consideração o comportamento dos alunos também (*off-task*, *on task conversation* e *on task*). Nas duas estratégias os autores desenvolveram um detector para cada estado afetivo. A diferença é que na primeira delas, *expert feature engineering* que significa uma seleção inteligente de características, os autores utilizaram um conjunto de 249 características que foram inseridas no treinamento dos modelos. Dentre essas características haviam as (i) básicas, como tempo que o aluno ficou visualizando os mapas, número de cliques, frequência em que o aluno realizou cada ação, proporção de efetividade de ações, as (ii) de sequência, que contemplavam a frequência de sequências de três ações consecutivas, como ler um recurso -> adicionar um conceito -> adicionar um link no mapa, e as (iii) de limites, que envolviam selecionar um limite ideal para determinadas ações, como número total de pausas depois de criar um link. Visto o número consideravelmente alto de características, os autores realizaram uma etapa de seleção de características, visando selecionar as que melhor representassem cada um dos detectores. Após, os autores utilizaram um conjunto de algoritmos de classificação, que vêm sendo utilizados com frequência pela comunidade, como C4.5, RIPPER, Step Regression, Logistic Regression e Naïve Bayes. Apenas o algoritmo que obteve o melhor resultado foi considerado para a análise final. As métricas de avaliação também seguiram o que os outros trabalhos relacionados utilizam: Cohen's Kappa e A' . A validação dos modelos também utilizou a técnica de *k-fold cross-validation* ($k = 10$) com validação a nível de estudante. Por outro lado, a segunda estratégia utilizou métodos de *deep learning* para o treinamento dos modelos. Nessa estratégia, os autores não utilizaram o conjunto de características da outra estratégia, apenas treinaram modelos neurais com muitas camadas escondidas e neurônios. Também utilizaram vários algoritmos de teste, como *Recurrent Neural Network* (RNN), *Long-Short Term Memory* (LSTM), *Gated Recurrent Unit* (GRU) e *temporal Convolution*. Como resultados, a estratégia de *feature engineering* foi melhor em todos os estados afetivos, exceto para confusão, onde as duas estratégias empataram. O algoritmo que obteve a melhor performance foi *Logistic Regression*, alcançando para tédio ($A' = 0,682$, $Kappa = 0,278$), para confusão ($A' = 0,568$, $Kappa = 0,091$), para frustração ($A' = 0,634$, $Kappa = 0,056$), para surpresa ($A' = 0,570$, $Kappa = 0,070$) e, por fim, para engajamento concentrado ($A' = 0,624$, $Kappa = 0,142$) obtendo uma média final de ($A' = 0,634$, $Kappa = 0,168$).

A Tabela 3 apresenta uma síntese dos trabalhos relacionados, focando em quais estados afetivos foram detectados, quais os métodos de anotação das emoções foram utilizados, número de anotações realizadas, número de anotadores, número de alunos que participaram da coleta, número de características consideradas no desenvolvimento dos modelos, número de *logs* gerados pelos alunos e considerados no treinamento dos modelos, algoritmos empregados, métodos de avaliação dos modelos e a média dos resultados. Alguns artigos não apresentam certas informações, como número de *logs* considerados ou número de anotações realizadas. Assim, quando tal informação não foi encontrada nos trabalhos, esta foi representada com o símbolo “-”. Além disso, devido a uma limitação de espaço, para uma melhor visualização da Tabela 3, os nomes

de todos os algoritmos identificados foram abreviados e o estado de concentração engajada também foi abreviado para “concentr. engajada”. A lista de abreviações dos nomes dos algoritmos é apresentado na Tabela 2.

Tabela 2 – Abreviações dos algoritmos utilizados nos trabalhos relacionados.

AB	AdaBoost	LR	Logistic Regression
ALR	Additive Logistic Regression	LWL	Locally Weighted Learning
BLR	Bayesian Logistic Regression	MP	Multilayer Perceptron
BP	Bagging Predictors	NB	Naive Bayes
C4	C4.5 Decision Trees	NBU	Naive Bayes Updatable
DT	Decision Trees	NN	Nearest Neighbor
DTA	Decision Tables	NNG	Nearest Neighbor Generalization
J4	J48 Decision Trees	PA	PART
JR	JRip	RI	RIPPER
KS	KStar ou K*	RT	REP-Trees
LIR	Linear Regression	SR	Step Regression
LMT	Logistic Model Trees	SVM	Support Vector Machines

Fonte: Elaborado pelo autor.

A Tabela 3 apresenta uma síntese dos métodos e estratégias utilizadas por cada trabalho relacionado. Porém, nenhum resultado foi apresentado para não dificultar a visualização do comparativo. Assim, a Tabela 4 apresenta essa síntese apenas dos resultados e dos melhores algoritmos apresentados pelos trabalhos relacionados. Para isso, apenas os resultados para os estados de frustração, confusão, tédio e engajamento concentrado/*flow* foram selecionados. Por fim, a média dos resultados dos quatro modelos foi calculada. É importante ressaltar que esses resultados foram obtidos de diferentes ambientes de aprendizagem, em diferentes contextos e com diferentes características no desenvolvimento dos detectores. Assim, a comparação desses resultados tem como objetivo apresentar até onde os trabalhos relacionados têm chegado, destacando a complexidade na obtenção de uma boa acurácia na detecção, chamado de desempenho do detector.

Dos oito trabalhos relacionados, todos eles utilizaram as métricas de avaliação de desempenho baseado nos valores de Kappa e A'. Porém, apenas o primeiro artigo que apresenta somente o valor Kappa e não o A' como todos os outros. Além disso, o primeiro artigo também não apresenta de forma explícita os melhores algoritmos para cada estado afetivo, os autores descrevem que de todos os algoritmos testados os melhores foram os algoritmos baseado em funções (LR, MP e SVM), meta (AB, BP e ALR) e em árvores de decisão (C4, LMT e RT). Novamente, o símbolo “-” foi utilizado para demarcar as informações que não foram encontradas nos trabalhos. Nesta tabela, K foi utilizado como abreviação de Kappa e A1 como abreviação de Algoritmo.

Com base no comparativo apresentado na Tabela 3, é possível notar que todos os oito trabalhos detectam os estados de tédio, confusão, frustração e *flow*/concentração engajada. Alguns dos trabalhos ([1], [7] e [8]) ainda detectam os estados de prazer, surpresa, ansiedade e neutro. Porém, após realizar as anotações destes estados, os autores identificaram que a quantidade de

Tabela 3 – Comparativo entre os trabalhos relacionados.

Artigo	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Estados afetivos detectados	Tédio, Confusão, Frustração, Flow, Prazer, Surpresa, Neutro	Tédio, Confusão, Frustração, Flow, ?	Tédio, Confusão, Frustração, Flow, ?	Tédio, Confusão, Frustração, Flow, ?	Tédio, Confusão, Frustração, Flow, ?	Tédio, Confusão, Frustração, Flow, ?	Tédio, Confusão, Frustração, Concentr. Engajada, Surpresa, Ansiedade, ?	Tédio, Confusão, Frustração, Concentr. Engajada, Prazer, ?
Método de anotação	Análise de Vídeo	BROMP	BROMP	BROMP	BROMP	BROMP	BROMP	BROMP
Número de anotações	3380	763	-	-	-	1241	755	5212
Número de anotadores	4	2	2	2	2	4	2	2
Número de alunos	28	89	1622	-	-	326	119	93
Número de características consideradas	17	232	172	69	232	127	38	249
Número de logs considerados	-	-	810.000	-	-	-	-	146.141
Algoritmos empregados	NB, NBU, LR, MP, SVM, NN, KS, LWL, AB, BP, ALR, C4, LMT, RT, DTA, NNG, PA	J4, SR, NB, RT	J4, SR, JR, NB, KS, RT	J4, RT, JR, KS	LIR, DT, SR, NB, JR, J4, RT, BLR, KS	J4, JR, SR	J4, JR, NB, SR, LR, KS	C4, RI, SR, LR, NB
Método de avaliação	10-fold cross-validation	6-fold cross-validation	5-fold cross-validation	5-fold cross-validation	5-fold cross-validation	Leave-one-out cross-validation	10-fold cross-validation	10-fold cross-validation
Nível de avaliação	Observação	Aluno	Aluno	Aluno	Aluno	Aluno	Aluno	Aluno

Fonte: Elaborado pelo autor.

anotações era muito menor, quando comparado aos quatro primeiros, indicando que ocorrem com menor frequência nestes tipos de ambientes. Ainda, o estado “?” está associado com o protocolo BROMP, onde os anotadores devem anotar quando não tiverem certeza de um estado em um determinado momento, representando a falha na identificação do estado afetivo do aluno em um dado momento.

Os trabalhos mais recentes têm utilizado o termo concentração engajada, engajamento concentrado ou somente engajamento ao invés do termo *flow*. Segundo Baker et al. (2010b), isso se deve ao fato de que o *flow* é um estado de concentração plena, em que o aluno esteja com seu foco plenamente na tarefa que está realizando, com objetivos claros, envolvendo inclusive a

Tabela 4 – Comparativo dos resultados entre os trabalhos relacionados.

		[1]		[2]		[3]		[4]		[5]		[6]		[7]		[8]	
		Valor	AI	Valor	AI	Valor	AI	Valor	AI	Valor	AI	Valor	AI	Valor	AI	Valor	AI
Frustração	K	0,390	-	0,230	RT	0,324	NB	0,150	JR	0,157	JR	0,389	J4	0,105	LR	0,056	LR
	A'	-	-	0,990	RT	0,682	NB	0,600	JR	0,602	JR	0,726	J4	0,692	LR	0,634	LR
Tédio	K	0,207	-	0,280	NB	0,229	JR	0,240	J4	0,264	J4	0,332	J4	0,469	LR	0,278	LR
	A'	-	-	0,690	NB	0,632	JR	0,660	J4	0,671	J4	0,723	J4	0,848	LR	0,682	LR
Confusão	K	0,182	-	0,400	JR	0,274	J4	0,150	RT	0,148	RT	0,334	JR	0,056	NB	0,091	LR
	A'	-	-	0,990	JR	0,736	J4	0,630	RT	0,627	RT	0,735	JR	0,552	NB	0,568	LR
Flow/ Engajamento	K	0,193	-	0,310	KS	0,358	KS	0,420	JR	0,423	JR	0,360	SR	0,156	SR	0,142	LR
	A'	-	-	0,710	KS	0,678	KS	0,730	JR	0,743	JR	0,697	SR	0,590	SR	0,624	LR
Média	K	0,163	-	0,305		0,296		0,240		0,248		0,354		0,196		0,141	
	A'	-	-	0,845		0,682		0,650		0,660		0,720		0,670		0,627	

Fonte: Elaborado pelo autor.

perda de noção de tempo e de auto-consciência, conforme definido por Csikszentmihalyi (1990). Por outro lado, concentração engajada, considerado um estado afetivo-cognitivo, se refere ao aluno estar trabalhando na sua tarefa de forma focada e intensa e com envolvimento completo do aluno, mas sem se restringir aos aspectos e conceitualizações do *flow*.

Nota-se também que não há um padrão nos números de anotações, números de alunos que participaram da coleta, número de características consideradas e número de *logs* considerados. Isso indica que ainda não se sabe ao certo o que seria o ideal em cada um dos casos. O mesmo ocorre com a seleção dos algoritmos. Todos os trabalhos selecionaram uma gama de algoritmos e realizaram testes em cima de todos eles. O que obteve o melhor desempenho foi utilizado para o desenvolvimento dos modelos finais. Porém, com base na Tabela 4, é possível notar que os algoritmos que têm apresentado os melhores desempenho foram JR (7 ocorrências), LR (6 ocorrências), J4 (5 ocorrências), NB (3 ocorrências), RT (3 ocorrências), KS (2 ocorrências) e SR (2 ocorrências).

Referente ao método de avaliação dos resultados, todos os trabalhos utilizaram validação cruzada, *cross-validation*. A diferença foi no valor *K*, onde a maioria utilizou 5 ou 10. Ainda, o nível de validação da maioria dos trabalhos foi pelos dados do aluno. Ou seja, ao invés de selecionar os dados de teste e treino, para o desenvolvimento dos detectores, aleatoriamente do conjunto de dados, os dados de cada aluno foram selecionados em uma espécie de lote, do inglês *batch*. Essa estratégia é empregada com o objetivo de que os detectores sejam capazes de generalizar os estados afetivos de *logs* gerados por alunos que nunca haviam utilizado o sistema. Assim, ao invés de separar o conjunto integral de dados em grupos de *logs*, esse conjunto é separado em grupos de *logs* de alunos.

Ao analisar os trabalhos relacionados é possível destacar duas principais diferenças com o trabalho proposto nessa dissertação. A primeira delas está relacionada com a maneira de obtenção dos dados de afeto dos estudantes. Enquanto os trabalhos relacionados utilizam de normalmente dois especialistas anotando os estados dos alunos dentro da sala de aula, protocolo BROMP, durante a utilização dos ambientes educacionais, em curtos períodos de tempo (20 segundos por observação), o trabalho proposto utiliza da análise de vídeos para a geração dos rótulos dos estados afetivos dos alunos. Nesse vídeo, é possível visualizar o rosto dos es-

tudantes, obtidos com uso de *webcam*, e a tela do computador, obtidos por meio de *softwares* de captura de tela, *screen recorder*. Diferentemente do primeiro trabalho, que também utiliza a estratégia da análise de vídeo, este trabalho apresenta um protocolo para a anotação de estados afetivos para rótulos de treinamento de algoritmos de aprendizagem de máquina, que é descrito com detalhes na seção 6.2.2.

A segunda diferença é que no trabalho desenvolvido, diferentemente de todos os trabalhos relacionados, foram utilizadas informações da personalidade dos alunos como características no desenvolvimento dos detectores. Não foi possível descrever todas as características que foram utilizadas no desenvolvimento dos modelos dos trabalhos relacionados, pois seria uma lista extensa com mais de 1000 características, e, além disso, as características são restritas à área de aplicação do ambiente de aprendizagem. Porém, nenhum dos trabalhos utiliza informações da personalidade dos estudantes, mas sabe-se que a personalidade influencia na duração das emoções. (REIS et al., 2018).

6 TRABALHO DESENVOLVIDO

Esse trabalho tem como objetivo detectar os estados afetivos dos alunos por meio da mineração dos dados de *logs*. Mais especificamente, foram detectados os estados de frustração, confusão, tédio e engajamento concentrado. Esses estados foram escolhidos porque várias pesquisas vêm mostrando que eles ocorrem mais frequentemente em ambientes educacionais. (CALVO; D’MELLO, 2010). Conforme apresentado na seção de trabalhos relacionados (Seção 5), a maioria das pesquisas levam em consideração esses estados, divergindo em alguns trabalhos nos estados de prazer, surpresa, ansiedade e neutro. Porém, a incidência desses últimos é muito menor em ambientes educacionais (D’MELLO et al., 2008; PAQUETTE et al., 2016; JIANG et al., 2018), quando comparado aos quatro escolhidos. Assim, optou-se por trabalhar apenas com os mais proeminentes.

O ambiente educacional utilizado como caso de uso neste trabalho é o STI PAT2Math, descrito na Seção 3.3. Por ser um STI baseado em passos, PAT2Math é capaz de proporcionar uma ajuda aos alunos durante as etapas de resolução de equações de primeiro grau e não apenas para a resposta final. Além disso, o sistema pode fornecer dicas e verificar se cada passo de resolução de uma equação inserido pelo aluno está correto ou não. Caso incorreto, PAT2Math irá fornecer uma ajuda personalizada ao aluno, de forma que ele consiga prosseguir na resolução da equação. Assim, juntamente com os *logs* de interação dos alunos com a interface gráfica do STI, dados do modelo do aluno também foram utilizados como características no desenvolvimento dos modelos, que serão descritos nas seções seguintes.

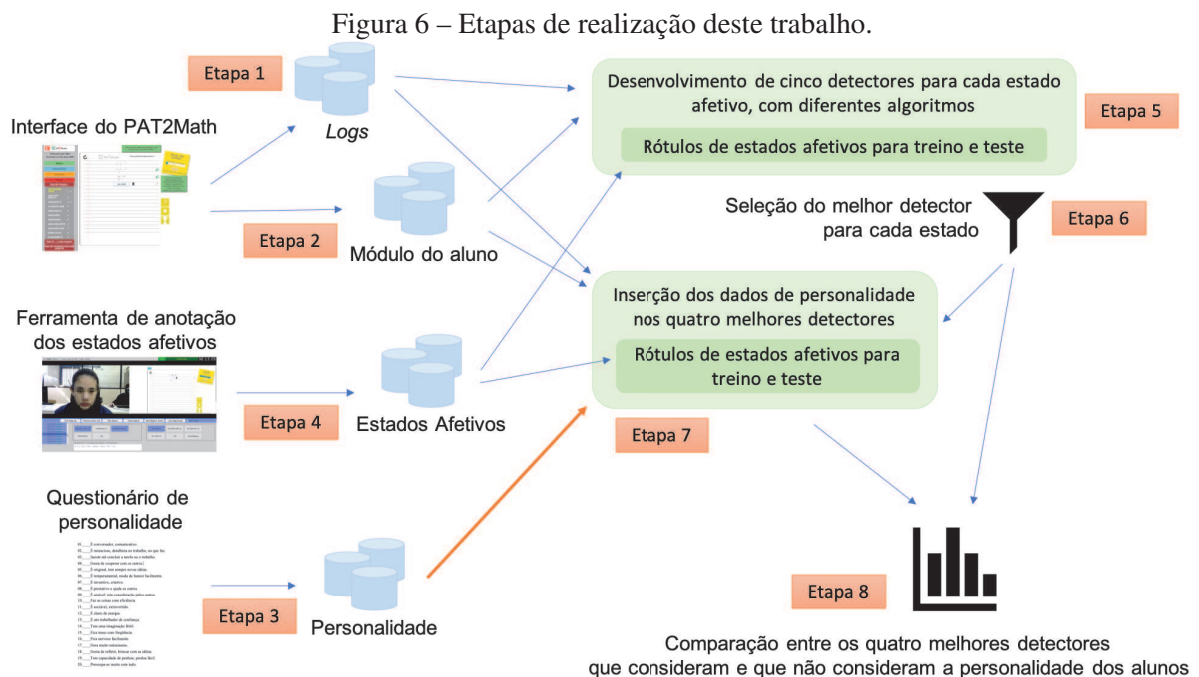
O diferencial deste trabalho está na forma de captura dos estados afetivos para criação dos rótulos dos dados e na adição de características de personalidade no desenvolvimento dos preditores. Sabendo que a personalidade influencia a duração das emoções (REIS et al., 2018) e quais emoções são experimentadas (NUNES; CAZELLA, 2011), a hipótese de pesquisa é que inserindo características da personalidade dos alunos no desenvolvimento dos preditores de estados afetivos, pode resultar em uma melhora nesse processo de detecção.

6.1 Etapas do Trabalho

Para atingir o objetivo proposto neste trabalho, foi necessário uma série de etapas para a coleta de dados e para o desenvolvimento e comparação dos detectores. Todos os detectores foram treinados com os *logs* de interação e com os dados do módulo do aluno. Os dados de interação dos alunos com a interface do PAT2Math foram coletados de 55 alunos utilizando o ambiente, em 10 sessões com média de 40 minutos. Dessa forma, todo tipo de ação do aluno com o sistema foi capturado. Os dados do módulo do aluno também são provenientes desta coleta de dados com alunos. Porém, estes dados foram capturados do banco de dados do PAT2Math. Eles contém todas as informações dos alunos geradas durante a utilização do sistema, incluindo equações resolvidas, passos de resolução corretos, errados, dicas solicitadas,

operações aplicadas etc. Os dados de personalidade dos alunos, utilizados somente para o treinamento dos detectores que consideram a personalidade, também foram obtidos destes mesmos alunos, porém por meio da aplicação de um questionário. E, por fim, a última fonte de dados, estados afetivos e comportamentos dos alunos, é proveniente de um protocolo elaborado no decorrer deste trabalho. Esses dados de estados afetivos foram utilizados como rótulos alvo para o treinamento dos detectores. Para a coleta desses dados, os alunos que utilizaram o PAT2Math tiveram seus rostos, áudio ambiente e tela do computador gravados em formato de vídeos. Com esses vídeos, codificadores treinados e testados por meio do protocolo desenvolvido, geraram rótulos de estados afetivos e comportamento dos alunos utilizando o sistema.

Para verificar a validade da hipótese de pesquisa deste trabalho, foram desenvolvidos cinco detectores para cada estado afetivo, sendo cada um deles com um algoritmo de treinamento diferente. Todos esses detectores foram desenvolvidos com base nos dados de *logs* e nos dados do módulo do aluno do STI PAT2Math. Após, apenas os quatro melhores detectores, um para cada estado, foram selecionados para uma comparação. Esta comparação contou com mais quatro detectores, que utilizaram como base esses quatro melhores detectores, mas que, além dos dados já utilizados, foram treinados com dados de personalidade dos alunos. Assim, foi possível comparar os melhores detectores treinados com e sem dados de personalidade. Ao final são apresentados os resultados desta comparação. Além disso, os melhores detectores desenvolvidos neste trabalho também são comparados com o estado da arte. As etapas realizadas neste trabalho são ilustradas na Figura 6.



Fonte: Elaborado pelo autor.

Além de ilustrar as etapas realizadas neste trabalho, a Figura 6 também exibe a ordem em que elas ocorrem. Assim, é possível identificar que as quatro primeiras etapas são relacionadas

a coleta dos dados de (etapa 1) *logs*, (etapa 2) módulo do aluno, (etapa 3) personalidade e (etapa 4) estados afetivos. Após, (etapa 5) cinco detectores para cada estado afetivo são desenvolvidos, cada um utilizando um algoritmo de aprendizagem de máquina diferente, destes (etapa 6) apenas o melhor detector de cada estado afetivo é selecionado. Esses quatro detectores são treinados novamente com dados de personalidade (etapa 7). E, por fim, é realizada uma comparação entre os detectores treinados com e sem os dados de personalidade (etapa 8). Cada uma dessas etapas serão descritas com mais detalhes nas próximas seções.

6.2 Coleta dos Dados de *Logs*

Esta seção tem como objetivo descrever como foi realizada a coleta dos dados de *logs*. Essa tarefa foi dividida em duas sub-tarefas. A primeira delas é o desenvolvimento de um coletor dos *logs* de interação do aluno com a interface do STI. A segunda é a coleta dos *logs*, realizada por meio da utilização do ambiente por alunos no laboratório de uma escola, chamada de coleta de dados.

6.2.1 Desenvolvimento do Coletor de *Logs*

Inicialmente, foi desenvolvido um *script*, utilizando a linguagem Java Script, para coletar os dados de *logs* enquanto os alunos utilizam o sistema. Esse *script* foi integrado à interface gráfica *web* do STI PAT2Math de forma que ao abrir o resolvedor de equações, o *script* era inicializado. Após sua inicialização, o *script* tinha acesso a todos os comandos de entrada e saída do aluno com o navegador. Assim, toda vez que o aluno realizasse qualquer movimento com o *mouse* ou com o teclado, uma função específica para cada ação era executada, gerando um registro (*log*). Um *log* representa toda e qualquer ação do aluno com a interface gráfica do sistema, incluindo cliques do *mouse*, teclas pressionadas etc. Além disso, o fato do estudante não realizar nenhum movimento também era considerado, uma vez que isso poderia significar que o aluno estivesse realizando outra atividade fora do ambiente.

O *script* integrado ao PAT2Math é capaz de capturar as seguintes informações, enquanto um aluno utiliza o ambiente:

1. **cliques:** o sistema captura todos os cliques do *mouse* do usuário e identifica em qual componente da interface esse clique foi realizado. Além disso, um contador de cliques seguidos é contabilizado em cada *log*, juntamente com um contador para cliques seguidos em uma equação e um contador para cliques seguidos em cada passo de resolução de uma equação.
2. **Teclas:** Qualquer tecla do teclado do computador que o aluno pressionar será salva como um *log*. A única exceção para essa regra é quando o aluno está com a caixa de entrada de equações selecionada, indicando que o aluno está digitando um passo de resolução de

equação, explicado no item 3.

3. **Passos de resolução:** Toda vez que o aluno entrar com um passo de resolução no sistema, um *log* é gerado. Esse *log* contém informações de como o mecanismo de verificação foi acionado, por clique no botão de verificar ou por pressionar a tecla *enter*. Além disso, esse *log* não é salvo imediatamente. Ele fica em um estado de espera, para que o servidor do PAT2Math possa retornar um *feedback* sobre o passo inserido pelo aluno. Assim, esse *log* contém a equação inicial, o passo anterior, o passo atual que o aluno acabou de inserir, o *feedback* mínimo retornado pelo servidor (certo ou errado) e a mensagem de *feedback* de erro, caso o passo esteja incorreto. Além disso, um contador de passos incorretos e um contador de passos incorretos seguidos foi adicionado a cada *log*.
4. **Tempo ocioso:** O sistema contabiliza o tempo em que o aluno não interage com a interface. Foi considerado como interação qualquer movimento do *mouse* ou tecla pressionada do teclado. Assim, a cada dois segundos que o aluno não interagir de alguma dessas maneiras com o sistema, um *log* de tempo ocioso é gerado. Esse intervalo de dois segundos está relacionado com os rótulos de estados afetivos, pois estes podem ser anotados em um tempo mínimo de dois segundos, de acordo com o protocolo de anotação dos estados afetivos elaborado, descrito na seção 6.5. Este valor está relacionado com os rótulos de estados afetivos de forma que se em uma determinada janela de tempo vários rótulos do aluno foram anotados, mas este aluno não realizou nenhuma ação no sistema, não haverá *logs* para sincronizar com esses rótulos anotados. Assim, com uma janela de tempo ocioso com o mesmo tempo mínimo para a anotação de cada rótulo, foi possível que todos os rótulos anotados pudessem ser sincronizados com algum *log*. Também, um contador foi inserido para contabilizar o tempo acumulado de ociosidade do aluno.
5. **Foco no navegador:** Como o STI PAT2Math é um sistema *web*, esse tipo de *log* é gerado quando o aluno faz qualquer ação de fechamento ou troca de abas no navegador. Assim, para cada vez que a aba em que o STI estiver sendo executada perder ou ganhar o foco, um desses *logs* é gerado.
6. **Redimensionamento do navegador:** Assim como o *log* de foco no navegador, esse *log* é gerado cada vez que o aluno altera o tamanho da janela em que o STI estiver sendo executado.
7. **Carregamento da página:** Cada vez que o aluno carregar a página principal do STI PAT2Math, um desses *logs* é gerado.
8. **Fechamento da página:** Caso o aluno simplesmente feche o navegador ou faça seu *logout* do sistema, esse *log* é gerado. Também é válido se o aluno realizar um recarregamento da página.

9. **Movimento rápido do mouse:** O sistema fica monitorando todo tipo de movimento que o aluno realiza com o *mouse* do computador. Um cálculo de velocidade é feito a cada evento de início e fim do movimento do *mouse*. Caso essa velocidade seja considerada como rápida, um *log* é gerado, salvando inclusive posições de início e fim do movimento na tela.

Conforme descrito no item 1, todos os cliques que o aluno realiza na interface gráfica são capturados. Porém, há mais de 100 elementos com identificadores únicos na interface do sistema. De modo a reduzir essa quantidade de possibilidades, os elementos foram agrupados em nove grupos distintos. Cada grupo engloba os elementos referentes a certas ações na interface do sistema. Os grupos foram definidos em (i) resolução, (ii) planos, (iii) atalhos, (iv) bugs, (v) gamificação, (vi) dicas, (vii) erros, (viii) notificações e (ix) sistema. A Figura 7 ilustra sete dos nove grupos, ficando de fora apenas os grupos de notificações e sistema.

Figura 7 – Classificação dos elementos da interface em grupos.

The image shows a screenshot of a math application interface. The interface is divided into several sections, each with a label indicating its group classification:

- Gamificação:** Located at the top left, showing total score (18655) and current level score (17695).
- Erros:** A blue box at the top center.
- Feedback de erro:** A green box at the top right.
- Planos:** A vertical sidebar on the left with levels: Básico, Intermediário, Avançado, Expert, and Fase 30: Frações... (with a list of equations).
- Resolução:** The main workspace in the center, showing a math problem being solved step-by-step: $\frac{x}{5} + \frac{x}{7} = 12$, $7x + 5x = 420$, $12x = 420$, $x = 420$, $\frac{12}{6}$, $x = 105$, $\frac{12}{3}$, $x = 35$. A green checkmark is at the bottom.
- Dica:** A yellow box on the right showing a tip: "150 de 150 pontos", "Equações concluídas: 17/10", and "20 dicas gratuitas disponíveis".
- Atalhos:** A yellow box with a question mark icon.
- Bugs:** A blue box with a bug icon.
- Gamificação:** A blue box with a trophy icon.
- Texto da dica solicitada pelo aluno, específico para cada passo:** A green box with a text input field.

Fonte: Elaborado pelo autor.

O grupo **resolução** contempla todos os cliques do aluno realizados na área de resolução das equações, incluindo cliques nos passos de resolução, nas linhas e no botão de verificação. O grupo **planos** agrupa elementos clicados na seleção das equações ou planos. O grupo **atalhos** engloba os cliques no botão que abre uma janela *pop-up* e os cliques dentro da janela, onde

os alunos podem verificar os atalhos de teclado disponíveis no ambiente. O grupo **bugs**, assim como o grupo atalhos, engloba o botão e a janela *pop-up* que abre para o aluno descrever algum erro que ocorreu no sistema durante sua utilização. O grupo **gamificação** agrupa os cliques em qualquer elemento relacionado à gamificação, como pontos, barra de progresso, erros gratuitos ou *ranking*. O grupo **dicas** contempla os cliques no botão de dicas e no texto fornecido pelo sistema quando o aluno solicita a dica. O grupo **erros** abrange os cliques na mensagem de *feedback* de erro, fornecido pelo sistema quando o aluno comete algum erro na resolução das equações. Os grupos **notificações** e **sistema** não aparecem na Figura 7, mas englobam, respectivamente, os cliques em janelas *pop-ups* de alerta e os cliques em qualquer outro componente ou local da interface em que não esteja englobado em um grupo específico.

Além dessas informações, para cada *log* criado, uma série de informações é anexada a esse *log*:

1. **Identificador (ID) do aluno:** o identificador único que representa o aluno no banco de dados do STI.
2. **ID do log:** o identificador único que representa o *log* salvo na base de dados do STI.
3. **Sequência do log:** a ordem em que o *log* foi gerado em cada dia de utilização do sistema pelo aluno.
4. **Sequência do dia:** a ordem, em dias, em que o aluno está utilizando o STI.
5. **Equação inicial:** a equação em que o aluno está trabalhando no momento em que o *log* foi gerado.
6. **Último passo correto:** o último passo que o aluno inseriu no sistema e foi verificado como correto.
7. **Passo corrente:** o passo de resolução que o aluno inseriu e que foi verificado, utilizado somente quando o aluno realiza uma requisição de verificação de passo ao servidor do PAT2Math.
8. **Pontos:** a pontuação que o aluno tem até o momento.
9. **Pontos de erro:** quantidade de pontos que o aluno perdeu por cometer algum erro.
10. **Contador de passos:** número de passos que o aluno inseriu no sistema para a equação atual.
11. **Pedido de ajuda:** valor booleano que identifica se o aluno solicitou alguma ajuda antes de verificar se o passo estava correto ou não.
12. **Dica:** dica que o sistema forneceu quando o aluno solicitou ajuda.

13. **Dicas por operação:** número de dicas solicitadas por operação.
14. **Dicas no último minuto:** número de dicas que o aluno solicitou no último minuto de utilização do sistema, se houve alguma.
15. **Tempo:** momento exato em que o *log* foi gerado, incluindo data e hora.
16. **Tempo entre logs:** a diferença de tempo entre cada *log* gerado pelo sistema.
17. **Tempo entre requisições:** a diferença de tempo entre cada requisição que o aluno fez ao servidor do PAT2Math, seja por pedido de ajuda ou verificação de passo.

No momento em que a página é carregada pela primeira vez no navegador, o sistema realiza uma sincronização de data e hora entre o servidor e o computador do aluno. Assim, todos os dados gerados no navegador do aluno foram armazenados baseado na data e hora do servidor, facilitando a integração com outros tipos de dados.

Todos os *logs* são gerados no navegador do aluno e enviados para o servidor do PAT2Math. Para que o servidor não fosse sobrecarregado com inúmeras requisições de *logs* por segundo, foi implementado no *script* um sistema de conjunto *logs*. A ideia é que cada *log* gerado seja armazenado localmente, ou seja, no navegador do aluno. Esse armazenamento funciona como se fosse um *buffer*. Os *logs* são anexados nesse *buffer*, sendo que a cada 60 segundos o sistema *web* faz uma requisição do tipo *POST* para o servidor, enviando todo conteúdo do *buffer*. O servidor, por sua vez, recebe esse conjunto de *logs* e armazena cada um deles no banco de dados do PAT2Math. Após enviar os dados do *buffer*, este é esvaziado para receber novos *logs*.

6.2.2 Coleta dos Logs

Após testar e validar o funcionamento do *script* de coleta dos *logs*, a coleta de dados foi realizada em uma escola privada na cidade de São Leopoldo, região metropolitana do Rio Grande do Sul. Duas turmas do sétimo ano do ensino fundamental utilizaram o STI PAT2Math como ferramenta de auxílio às aulas de matemática, proposto pela própria professora por já conhecer o STI. A professora e o tempo de utilização da ferramenta também foi o mesmo para as duas turmas. Ao todo, participaram 55 alunos, com idade entre 12 e 13 (média de 12,15 anos), sendo 29 do sexo feminino e 26 do sexo masculino. Os alunos utilizaram o STI durante 10 sessões, uma vez por semana. Cada sessão era definida como um período de aula, que nesta escola é de 50 minutos. Assim, levando em consideração o tempo de deslocamento dos alunos até o laboratório da escola, as sessões de utilização do STI eram em média de 40 minutos. Todas as sessões ocorreram no turno da manhã. As duas turmas sempre utilizaram o sistema no mesmo dia de coleta de dados. As sessões ocorreram entre os meses de maio e outubro de 2018, sendo duas sessões em maio, somente uma em junho e nenhuma em julho devido ao recesso escolar, quatro em agosto, duas em setembro e uma em outubro. Durante as sessões haviam no laboratório, junto com os alunos, a professora de matemática, a responsável pelo laboratório e um bolsista

do projeto PAT2Math. Assim, os alunos podiam tirar suas dúvidas relacionadas ao ambiente e à matéria, caso necessário.

A utilização do STI pelos 55 alunos em 10 sessões, com uma média de 40 minutos cada, totalizou uma média de aproximadamente 360 horas de utilização do sistema. Isso resultou na geração e armazenamento de um total de 386.015 *logs* de interação dos alunos com a interface do STI PAT2Math.

Um termo de consentimento (Termo de Consentimento Livre e Esclarecido - TCLE) foi elaborado, com o intuito de comunicar aos responsáveis dos alunos sobre a coleta que estava sendo realizado. Embora a utilização da ferramenta tenha sido uma iniciativa da professora, era importante que ficasse bem claro o que seria feito com os dados gerados pelos alunos. Esse termo foi submetido e aceito ao comitê de ética da UNISINOS. O TCLE foi entregue a todos os alunos, sem a obrigatoriedade da assinatura dos responsáveis para a participação. Caso algum responsável se negasse a deixar seu filho participar, os dados daquele aluno seriam descartados. No caso do presente estudo, todos os alunos entregaram o TCLE assinado por pelo menos um responsável.

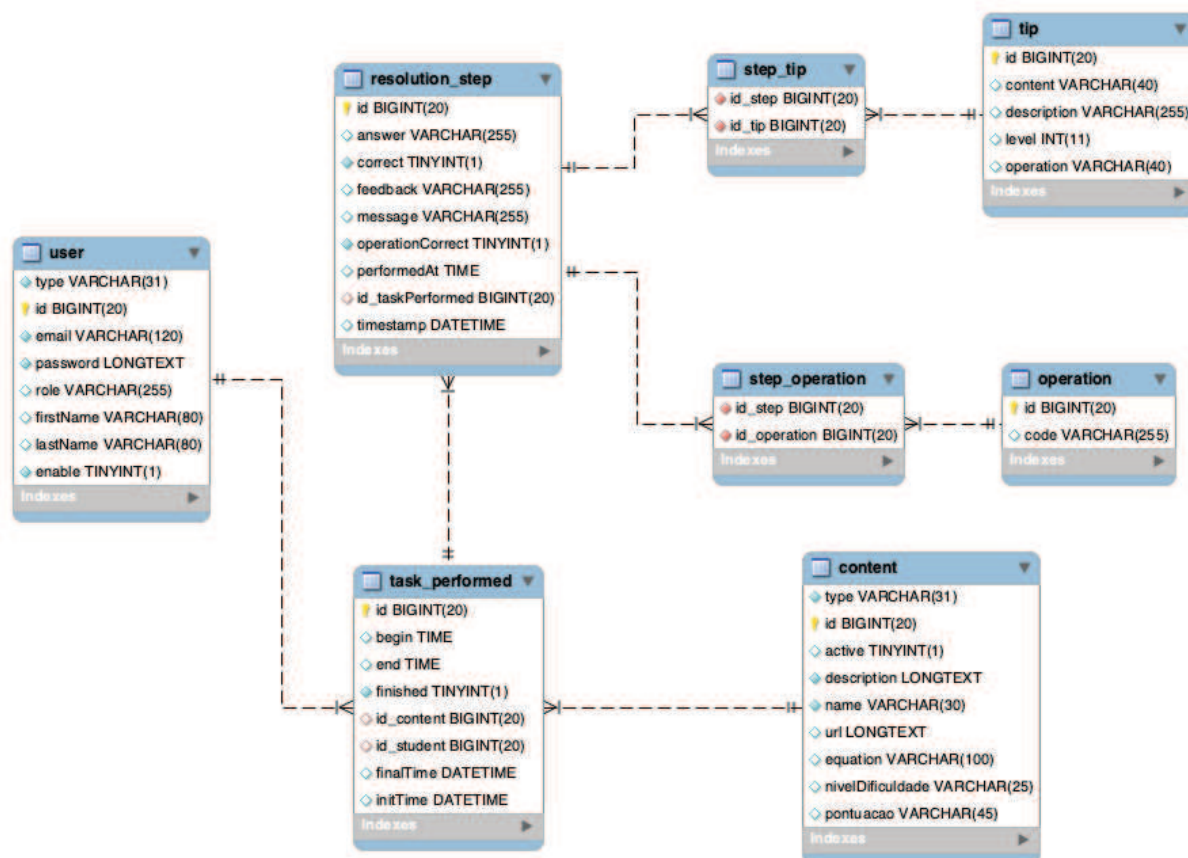
6.3 Coleta dos Dados do Módulo de Aluno

O módulo do aluno do STI PAT2Math, conforme descrito na seção 3.3.2, armazena todas as interações do aluno com o sistema. Assim, o sistema armazena em sua base de dados cada equação resolvida, passo verificado, erro cometido, dica fornecida entre outras informações para cada aluno. Essas informações são geradas pelo módulo especialista, que, ao receber um passo de resolução do aluno, é capaz de identificar quais as operações foram realizadas e se houve algum erro, qual foi esse erro.

Como apresentado pelo último trabalho relacionado (JIANG et al., 2018), uma seleção de características mais elaborada tende a apresentar melhores resultados nos modelos de detecção. Assim, os dados de *logs* da interação do aluno com a interface do STI foram integrados com os dados do módulo de aluno, após a coleta de dados, fornecendo uma gama ainda maior de características. Os *logs* e os dados do módulo do aluno foram armazenados na mesma base de dados, que utiliza do banco MySQL, o que facilitou a integração dessas duas fontes de informações. Como cada *log* possui a informação do ID do aluno, a equação sendo resolvida e qual é o passo atual, além de informações de data e hora, essas informações foram utilizadas para unir o *log* com o que já se havia salvo na base de dados. A Figura 8 ilustra o diagrama Entidade Relacionamento (ER) de parte do modelo relacional do módulo do aluno. Essa parte representa as tabelas onde ficam armazenadas todas as equações, passos de resolução, operações corretas e erros do aluno.

A tabela *user*, usuário ou aluno, armazena as informações pessoais do aluno, como nome, e-mail, senha etc. A tabela *task_performed*, tarefa realizada, armazena todas as equações que o aluno já finalizou ou ainda está resolvendo, incluindo data e hora de início e fim e o ID do

Figura 8 – Diagrama entidade relacionamento de parte da base de dados do módulo do aluno.



Fonte: Elaborado pelo autor.

conteúdo. A tabela **content**, conteúdo, armazena a equação em si. Assim, vários alunos podem resolver uma mesma equação. Nessa tabela também são armazenados o nível de dificuldade da equação e a pontuação máxima que o aluno pode obter ao resolver a equação. A tabela **resolution_step**, passos de resolução, armazena todos os passos que o aluno inseriu para cada equação, esteja ele correto ou errado. Caso o passo esteja correto, a tabela **operation**, operação, armazena qual foi a operação que o aluno aplicou corretamente para obter o resultado. Mas, como um passo de resolução pode englobar várias operações, a tabela **step_operation**, operação no passo, armazena todas as operações corretas para um determinado passo (tabela muitos para muitos). Por outro lado, se o passo estiver errado, a tabela **tip**, dica, armazena qual foi a operação que o aluno errou no passo fornecido. O STI PAT2Math também é capaz de identificar as *misconceptions* dos alunos. Segundo Ojose (2015), as “*misconceptions* são mal-entendidos e interpretações erradas com base em significado incorreto”. Por exemplo, o aluno que esquece de inverter o sinal da operação que está aplicando em uma operação inversa de adição. Assim, essa tabela armazena as operações erradas, *misconceptions* e dicas para fornecer ao aluno quando este cometer algum dos dois. Por fim, a tabela **step_tip**, dica no passo, tem o mesmo papel da tabela **step_operation**, armazenando mais de uma dica para cada passo, pois o aluno pode cometer mais de um erro em um único passo.

Ao integrar os *logs* com o módulo do aluno, novas características foram extraídas da base de dados. Sendo elas nível de dificuldade da equação (12 níveis), pontuação, número de passos na equação, número de passos errados na equação, número de passos errados seguidos na equação, efetividade dos passos na equação (calculada pela divisão entre o número de passos corretos pelo número total de passos), tempo que o aluno está resolvendo a equação e tempo que o aluno está no passo.

Além dessas informações específicas de cada equação, também foram calculadas características das equações resolvidas pelo aluno “até o momento atual”. Ou seja, dado um *log* de um aluno no meio de uma sessão de coleta, dois cálculos são realizados, o primeiro em cima das equações resolvidas pelo aluno desde o início do dia até o *log* atual, chamado de diário, e o segundo em cima das equações resolvidas pelo aluno desde o primeiro dia de coleta até o *log* atual, chamado de total. O resultado desses cálculos foi a média de tempo gasto para resolver as equações (diário e total), a média do tempo gasto em cada passo de resolução (diário e total), o número de passos realizados (diário e total), número de passos corretos (diário e total), número de passos errados (diário e total), número de equações resolvidas (diário e total) e efetividade dos passos (diário e total).

Além dos dois cálculos realizados para cada aluno (diário e total), os mesmos cálculos foram realizados com informações da turma. Assim, foi possível inserir características da turma sobre a média de tempo para resolver as equações (diário e total), média de tempo em cada passo (diário e total), média de número de passos (diário e total), média de passos corretos (diário e total), média de passos errados (diário e total), média do número de equações resolvidas (diário e total) e média da efetividade dos passos.

Referente às operações algébricas aplicadas pelos alunos durante a resolução das equações, o módulo especialista trabalha com dez operações distintas, sendo elas adição (AD), adição e subtração de frações (AF), propriedade distributiva (DM), divisão (DV), multiplicação de frações (MF), mínimo múltiplo comum (MM), multiplicação (MT), operação inversa (OI), subtração (SB) e simplificação (SP). Para cada operação considerada pelo PAT2Math, foram inferidas e calculadas as operações necessárias para resolver cada equação, quais e quantas operações o aluno acertou nos últimos cinco segundos¹, quais e quantas operações o aluno acertou na sessão (diário), quais e quantas operações o aluno acertou no total (total). A média das operações também foi calculada para a turma (diário e total). Todos esses cálculos se replicaram para as operações erradas. Mais uma vez, todas essas operações foram replicadas (aluno diário e total, turma diário e total) mas agora para verificar a efetividade das operações.

Além das operações de passos corretos e errados, características sobre as *misconceptions* também foram calculadas. Ao todo o PAT2Math é capaz de identificar 30 *misconceptions*. Para cada uma delas foram calculados quais e quantas ocorreram nos últimos cinco segundos (diário e total).

¹cinco segundos é o tamanho da janela para anotação das emoções no protocolo utilizado neste trabalho.

6.4 Coleta dos Dados de Personalidade

Após a última sessão de coleta de dados, conforme descrito na seção anterior, os alunos responderam um questionário de personalidade. Esse questionário identifica os traços de personalidade dos alunos com base na teoria dos grande cinco fatores, sendo este escrito em português e validado na tese de doutorado de Barbosa (2009) para o Brasil. Esse questionário é constituído de 20 questões, onde as respostas seguiam uma escala Likert com valores de 1 a 5. As questões perguntavam para o aluno relatar, por exemplo, o quanto ele se considerava conversador, comunicativo. Então, o aluno selecionava as opções discordo totalmente (1), discordo (2), nem concordo nem discordo (3), concordo (4) e concordo totalmente (5). Haviam quatro questões para cada um dos cinco traços de personalidade. A soma das respostas dessas quatro questões resultavam no índice que representava um traço de personalidade, sendo esse índice variando de 4 à 20. Assim, quanto mais próximo de 20, mais forte é a presença do traço e, por outro lado, quanto mais próximo de 4, mais fraco é a presença do traço. Apenas um aluno dos 55 não respondeu o questionário de personalidade, por não ter ido na aula no dia em que o questionário foi aplicado. Assim, seus dados não foram utilizados para o treinamento dos detectores.

6.5 Coleta dos Rótulos de Estados Afetivos

Para que os ambientes de aprendizagem possam detectar e inferir os estados afetivos dos alunos de forma automática e livre de sensores, primeiramente é necessário que esses sistemas possuam modelos de detecção, ou detectores de estados afetivos. Esses detectores utilizam algoritmos de aprendizagem de máquina e são treinados com base em algum rótulo alvo, chamado de treinamento supervisionado. Nesse caso, onde o objetivo é que esses ambientes possuam habilidades afetivas, o rótulo alvo de treinamento são os próprios estados afetivos. Assim, é necessário que se tenha um conjunto de rótulos, estados afetivos, para realizar o treinamento desses detectores. O ato de registrar os estados afetivos dos alunos é conhecido como anotação dos estados afetivos. Essa seção tem como objetivo apresentar (i) os métodos de anotação que vêm sendo utilizado pela comunidade e, devido a restrições que esses métodos apresentam, (ii) um novo protocolo para anotação dos estados afetivos e comportamentos dos alunos com base na análise de vídeos.

6.5.1 Estado da Arte

A análise de métodos para anotação dos estados afetivos deste trabalho foi realizada em cima dos trabalhos relacionados, descritos no capítulo 5, desta dissertação, uma vez que para alcançar o objetivo de detectar e inferir os estados afetivos dos alunos essa seja uma tarefa necessária para todos.

O primeiro trabalho relacionado (artigo [1]) utiliza de análise de vídeos para realizar uma

anotação após a coleta de dados, chamado também de detecção *offline*. Os autores filmaram o rosto dos estudantes com áudio e a tela do computador enquanto o aluno utilizava o ambiente de aprendizagem. Porém, sua estratégia utilizou de quatro anotadores, sendo o primeiro deles o próprio aluno, estratégia de autorrelato das emoções, o segundo era outro aluno que relatava as emoções de um colega e, por fim, os últimos dois eram dois codificadores treinados em detectar unidades de ações faciais (Action Units - AUs), do modelo Facial Action Coding System (FACS) de Paul Ekman. A anotação das emoções era realizada em uma janela de tempo de 20 segundos, onde quando mais de uma emoção era identificada, apenas a que era considerada a mais proeminente devia ser anotada.

Algumas limitações foram identificadas no emprego desse método de anotação das emoções. A primeira delas é que esse não é um protocolo validado. Os autores apenas relatam o que fizeram e como fizeram. Nenhum tipo de avaliação de concordância entre os anotadores foi previamente realizada. Segunda, como a coleta de dados realizada nesta dissertação foi em uma escola privada, que já tem seu cronograma bem definido, não era possível solicitar mais um espaço para que os alunos pudessem rever os vídeos e autorrelatar suas emoções. Terceira, embora exista um mapeamento de AUs para os estados de frustração, confusão e tédio (CRAIG et al., 2008), sendo estes apenas três dos quatro estados afetivos considerados nesta dissertação, a utilização deste protocolo exige dois codificadores treinados pelo método FACS para detectar unidades de ações faciais. Tendo em vista todas essas limitações, esse método não era uma opção viável para as anotações dos estados afetivos deste trabalho.

De todos os trabalhos relacionados, apenas o primeiro, o qual foi listado acima, não utiliza o protocolo BROMP, que é um protocolo para anotação de estados afetivos e é o protocolo que mais vêm sendo usado pelos trabalhos publicados mais recentemente. BROMP (Baker, Rodrigo, Ocumpaugh Monitoring Protocol) já está em sua versão 2.0. (OCUMPAUGH, 2015). Nesse protocolo, um codificador ou anotador fica dentro da sala de aula com um *smartphone*. Neste dispositivo, o codificador utiliza um aplicativo para realizar as anotações dos estados afetivos de cada aluno, seguindo uma sequência pré definida de alunos. As anotações constituem estados afetivos e os comportamentos dos alunos. Cada anotação tem uma duração máxima de 20 segundos e, se o codificador identificar um estado afetivo ou comportamento antes do final deste tempo, ele pode realizar a anotação e ir para o próximo aluno. Quando mais de um estado afetivo ou comportamento é identificado, apenas o primeiro é anotado. Para usar o protocolo BROMP, é necessário passar por uma fase de treinamento, obtendo uma certificação BROMP. Para isso, novos codificadores devem ler o manual do protocolo, participar de sessões de treinamento de anotações e, por fim, alcançar um nível de concordância compatível com alguma pessoa já certificada. A métrica de comparação usada é o Cohen's Kappa maior ou igual a 0,6.

Ao analisar o protocolo BROMP foi possível identificar algumas limitações quando comparado ao objetivo deste trabalho. Primeira limitação, este trabalho tem como objetivo verificar se a personalidade pode impactar na precisão da detecção dos estados afetivos. Esta hipótese partiu da justificativa de que a personalidade impacta no tempo de permanência dos alunos no

estado de confusão, que utilizou como base a transição dos estados afetivos dos alunos. A inconsistência encontrada aqui foi de que o protocolo BROMP identifica um estado afetivo por vez por aluno. Ao identificar e anotar um estado afetivo para um determinado aluno, o codificador passa a analisar e anotar um estado afetivo de um segundo aluno e assim sucessivamente até analisar todos os alunos da turma. Ao final, o codificar começa a rodada novamente, porém há uma diferença muito grande de tempo entre anotações de um mesmo aluno. Desta forma, não seria possível capturar a transição dos estados afetivos de um único aluno utilizando o protocolo BROMP. Outra limitação, também relacionada com o objetivo de capturar a transição dos estados afetivos dos alunos, é em relação ao tamanho da janela de anotação. O protocolo BROMP define uma janela de anotação de 20 segundos, sendo esta janela variável se o codificador identificar um estado afetivo antes desse período. Além disso, se o codificador identificar mais de um estado afetivo, ele deve anotar somente o primeiro estado identificado. Novamente, esta estratégia diverge do objetivo de identificar as transições dos estados afetivos, pois dentro desse período de 20 segundos vários estados podem ocorrer. Pode-se utilizar o STI PAT2Math como exemplo desse fenômeno. O PAT2Math é capaz de corrigir e fornecer um *feedback* para cada passo de resolução que o aluno insere no sistema ao resolver uma equação de primeiro grau provida pelo sistema. Desta forma, ao receber um *feedback* indesejado, o aluno pode mudar seu estado, passando de engajado para frustrado, por exemplo. Foi realizado um levantamento do tempo que os alunos levam para resolver cada passo de resolução. Este tempo varia de cinco segundos até mais de um minuto. Desta forma, seguindo a janela de 20 segundos do protocolo BROMP, um aluno poderia entrar com até quatro passos de resolução, tendo seu estado afetivo afetado há qualquer momento dentro dessa janela de 20 segundos. Outra limitação encontrada foi em relação à obrigatoriedade de se ter um anotador em sala de aula durante a coleta dos dados. Essa limitação pode fazer com a coleta de dados fique restrita à locais mais próximos dos anotadores, adicionando um bias negativo à generalização dos dados para diferentes populações.

Outros trabalhos também foram pesquisados, seguindo uma metodologia de *snowballing* com os trabalhos relacionados sendo o conjunto inicial de artigos, com o intuito de verificar a existência de um protocolo ou método que pudesse ser usado. No trabalho de Mota e Picard (2003), professores assistiram vídeos de seus estudantes e realizaram a anotação dos estados afetivos, mas nenhum método ou métrica de concordância foi utilizada para o treinamento dos anotadores. Craig et al. (2008) também realizaram a anotação das emoções pela análise de vídeos e pelo *think aloud* dos estudantes, ou seja, seus autorrelatos de estados afetivos. Neste trabalho, os autores utilizaram uma janela de 10 segundos para anotação das emoções. Para a verificação de concordância foi utilizada a medida Cohen's Kappa. Porém, nenhum protocolo de treinamento dos anotadores foi seguido. No trabalho de Woolf et al. (2009), os autores utilizaram uma metodologia de anotação similar ao BROMP, com anotadores em sala de aula e com uma janela de anotação de 15 a 20 segundos. Porém, neste trabalho também não foi apresentado um cálculo de concordância entre os anotadores. Por fim, o trabalho de Bixler e

D'Mello (2013) realizou a anotação dos estados afetivos com base em análise de vídeo, com uma janela de 15 segundos para a anotação das emoções. Novamente, nenhum protocolo ou métrica de concordância entre os anotadores foi apresentado.

Visto tais limitações, deu-se à necessidade da elaboração de um protocolo que pudesse ser utilizado para a anotação dos estados afetivos desta dissertação. Este protocolo traz características de todos os protocolos utilizados nos trabalhos citados acima. Como os de Mota e Picard (2003); Craig et al. (2008); Bixler e D'Mello (2013), este protocolo usa o que é chamado de detecção *offline*, no qual a análise dos estados afetivos e comportamentos ocorre depois que a coleta foi realizada pela análise de vídeos. Também é influenciado pelo trabalho de Woolf et al. (2009) na ideia de anotar um aluno por vez em uma janela de tempo e pelo de Ocumpaugh (2015) que propõem uma fase de treinamento dos anotadores, bem como um processo de verificação de concordância entre eles antes da anotação de estados afetivos e comportamentos para serem usados como rótulos de treinamento. A diferença é que com o protocolo proposto, os anotadores, além de poder visualizar os vídeos quantas vezes julgarem necessário antes de tomarem uma decisão, podem fazer múltiplas anotações de estados afetivos ou comportamentos em cada clipe (um segmento do vídeo) do vídeo total, onde a sequência em que elas ocorrem também é gravada. Ao contrário de outros métodos e protocolos que sugerem que quando mais de um estado ou comportamento ocorre o anotador deve escolher apenas a primeira ocorrência para anotar, no protocolo proposto os codificadores podem anotar tantos estados afetivos e comportamentos que desejarem somente se cada um durar pelo menos dois segundos. Além disso, o anotador visualizará o mesmo aluno por um período de tempo, em vez de analisar um aluno e pular para o próximo, como ocorre no protocolo BROMP. Assim, os dados coletados por este protocolo permitem identificar as transições entre estados afetivos e comportamentos dos alunos. Nesta dissertação a transição dos estados afetivos é muito importante, pois sabe-se que a personalidade pode influenciar no tempo em que um aluno fica em um determinado estado e, também, reconhecer a transição entre estados afetivos permite determinar o melhor momento para fazer uma intervenção pedagógica. (D'MELLO, 2012; REIS et al., 2018).

Destacam-se como vantagens deste protocolo (i) o fato dele ser escalável, visto que os codificadores não precisam estar presentes durante a obtenção dos dados, enquanto os alunos utilizam o ambiente de aprendizagem e têm seus rostos e telas dos computadores gravadas. Assim, esse método permite que os dados sejam coletados de diferentes amostras populacionais, incluindo até mesmo diferentes países, trazendo uma generalização maior para os resultados. Ainda, (ii) este protocolo foi desenvolvido com o intuito de obter rótulos alvos para o treinamento de algoritmos de aprendizagem de máquina supervisionado para a detecção de estados afetivos. Porém, isso não limita o protocolo para somente essa aplicação. Diferentes pesquisas podem se aproveitar deste protocolo com diferentes objetivos, devido as fases de treinamento e teste dos codificadores. Assim, esse protocolo também é flexível e generalizável, podendo ser utilizado em diferentes ambientes de aprendizagem e com objetivos diferentes do objetivo visado neste trabalho. Além disso, (iii) este protocolo permite que os codificadores realizem

múltiplas anotações de estados afetivos e comportamentos dentro de um único clipe e (iv) também possui um tamanho do clipe menor, que pode variar de acordo com o objetivo de cada pesquisa. Tais características, *iii* e *iv*, permitem que a transição dos estados afetivos e comportamentos dos alunos possa ser capturada, trazendo informações mais precisas e detalhadas, gerando uma espécie de histórico afetivo de cada aluno. Por fim, (v) o protocolo não exige um grande número de anotadores, comparado com protocolos em que alunos são utilizados como codificadores.

6.5.2 Protocolo Proposto

Foi elaborado um protocolo para a anotação dos estados afetivos e os comportamentos dos estudantes com base em observações em vídeos do aluno enquanto ele utiliza o STI PAT2Math no contexto de uso real de ambientes de aprendizagem nas escolas. Mais especificamente, esses vídeos contêm três tipos de informação: rosto dos alunos, áudio ambiente e gravações de tela do ambiente de aprendizado, todos gravados simultaneamente. Com essa informação, é possível ao codificador analisar um único aluno inúmeras vezes para tomar uma decisão sobre qual estado ou comportamento anotar, permitindo também a identificação do estado afetivo e das transições de comportamento dos alunos. O objetivo deste protocolo é a geração de rótulos para treinamento dos detectores dos estados afetivos e comportamentos dos alunos, para serem usados em algoritmos de aprendizado de máquina supervisionados para detectar e inferir tais construtos (estados afetivos e comportamentos) automaticamente. Este protocolo foi elaborado e executado por um grupo de três codificadores (Felipe de Moraes (autor), Tiago Kautzman (doutorando no PPGCA) e Patrícia A. Jaques (orientadora)), que realizaram 10 sessões de treinamento, seguindo as fases e definições propostas, com base em vídeos obtidos da coleta de dados realizada e descrita na seção 6.2.2. Um artigo relatando este protocolo foi aceito para publicação, como trabalho completo, na *Fourteen European Conference on Technology Enhanced Learning (EC-TEL)*, 2019, como uma proposta de protocolo de anotação de estados afetivos para geração de rótulos de algoritmos de aprendizagem de máquina supervisionada (MORAIS et al., 2019). Neste artigo, intitulado “*EmAP-ML: A Protocol of Emotions and Behaviors Annotation for Machine Learning Labels*”, além do protocolo, também foi descrito uma avaliação sobre o protocolo proposto, onde duas pessoas seguiram a fase de treino e, na fase de teste, foram capazes de alcançar o nível mínimo de concordância entre anotadores exigido para a realização individual das anotações de emoções e comportamentos.

O protocolo foi dividido em quatro fases, visando simplificar o entendimento do processo de anotação proposto. A primeira fase é a coleta e o desenvolvimento dos materiais necessários para realizar o protocolo, chamada de fase de materiais. A segunda fase é onde os codificadores discutem e praticam o protocolo, chamado fase de treinamento dos codificadores. A terceira fase é onde os codificadores colocam seu treinamento em teste, chamado fase de teste dos codificadores. E, finalmente, a quarta fase é onde os codificadores colocam seu treinamento

e esforço para gerar os rótulos finais de estados afetivos e dos comportamentos na prática, chamada fase de anotação. As quatro fases serão descritas em detalhes nas próximas subseções, bem como o fluxo de mudança de fase. Porém, primeiramente, é necessário esclarecer uma série de definições adotadas para a execução deste protocolo.

6.5.3 Definições do Protocolo de Anotação dos Estados Afetivos

Esta seção apresenta as definições de termos e nomenclaturas usadas na descrição deste protocolo. Primeiramente, é necessário definir alguns termos gerais usados no protocolo.

Uma **anotação** consiste em identificar um ou mais estados afetivos e um ou mais comportamentos em um clipe durante uma sessão de uma gravação, também chamados de rótulos. Um **clipe** é um segmento de uma sessão a ser anotada. Os cliques são anotados por um **anotador** ou **codificador** em uma sessão, que é a pessoa responsável por fazer as anotações. Neste trabalho, participaram três anotadores, o que ajuda a nivelar a compreensão de cada um dos estados afetivos e comportamentos durante a fase de treinamento. Para a fase de anotação, uma vez que os codificadores são treinados, os rótulos podem ser registrados por apenas um anotador.

A **sessão** representa uma parte do vídeo completo que foi escolhido para ser anotado pelos codificadores. Uma sessão tem hora de início (o ponto inicial do vídeo a ser analisado), duração (o tamanho da sessão) e um conjunto de cliques. O **vídeo completo** contém o rosto do aluno e o áudio ambiente em um lado e a tela do computador em outro. Um conjunto de sessões é chamado de **estudo**, por exemplo, as sessões de vídeo para a fase de treinamento. Dois tipos de dados são anotados pelos codificadores em cada sessão: estados afetivos e comportamentos, chamados de construtos. Para cada estado afetivo ou comportamento a ser anotado, ele deve ter uma duração de pelo menos dois segundos.

Este protocolo considera os quatro estados afetivos que vêm sendo utilizados pelos trabalhos relacionados e considerados pela comunidade como os estados mais frequentes em ambientes de aprendizagem (D'MELLO; CALVO, 2013), que são confusão, frustração, tédio e engajamento. Porém, o protocolo não se restringe à anotação apenas destes estados, devido ao seu protocolo de treinamento e avaliação dos anotadores. Além destes estados, o estado “?” também foi adicionado para quando nenhum dos outros quatro for identificado.

- **Confusão:** A confusão acontece quando os alunos parecem ter dificuldade em entender os materiais da aula ou a tarefa que estão tentando resolver. Já foram relatadas evidências de que a confusão surge de uma avaliação cognitiva, *appraisal*, de uma incompatibilidade entre o conhecimento necessário para resolver a tarefa atual e o conhecimento prévio do aluno. (D'MELLO et al., 2014). Quando os alunos estão interessados, mas não têm interação com o sistema, isso pode indicar que eles estão confusos porque não sabem como proceder para resolver a tarefa. No entanto, quando o codificador está em dúvida, este deve classificar o estado afetivo como “?”. Os sinais de confusão não precisam ser expressos apenas na forma facial, como franzir a testa; os codificadores também podem

usar sinais verbais (por exemplo, estudante pedindo ajuda) ou outros sinais convencionais de confusão (por exemplo, um aluno levantando as mãos ou inclinando a cabeça pro lado).

- **Frustração:** De acordo com modelos da teoria de avaliação cognitiva, *appraisal*, a frustração surge porque uma consequência desejável esperada para um dado evento (situação) não aconteceu. (ORTONY; CLORE; COLLINS, 1990). Geralmente, ocorre um estado de frustração quando os alunos esperam estar corretos sobre uma etapa da tarefa, demonstrando seu aborrecimento pelo *feedback* negativo do sistema. Alguns alunos podem mostrar um sorriso sem graça quando estão frustrados.
- **Tédio:** Estados afetivos como a frustração são orientados por eventos, como quando os alunos estão certos de que sua solução para uma tarefa estaria correta, recebem *feedback* de que erraram a solução da tarefa. O tédio, no entanto, não é causado por um evento, mas pela falta subjetiva de valor em uma determinada situação ou atividade. (PEKRUN et al., 2010). O tédio implica sentimentos desagradáveis, percepção alterada do tempo, excitação reduzida, motivação para mudar de tarefa ou abandono da tarefa e pode ser expresso no rosto, na voz (mais baixa do que o normal) e na postura. (PEKRUN et al., 2010).
- **Concentração engajada ou engajamento:** Neste estado afetivo-cognitivo, os alunos estão focados e prestando atenção à tarefa atual ou estão executando várias tarefas enquanto continuam focados e prestando atenção a todas essas tarefas. Este estado afetivo não deve ser confundido com o constructo de *flow* descrito por Csikszentmihalyi (1990), em que uma pessoa está totalmente imersa na tarefa. Porém, para ser classificado como engajamento o aluno deve demonstrar algum esforço cognitivo. Aqueles momentos em que os alunos estão um pouco relaxados, olhando ao redor, geralmente na troca de planos, devem ser considerados como “?”, se durarem mais de dois segundos. Quando mais curto, eles não devem ser anotados.
- **Estado afetivo “?”:** Não se refere a um estado afetivo, mas a uma categoria na qual o vídeo não apresenta informações suficientes para o codificador tomar a decisão correta. Por exemplo, quando os alunos saem do campo de visão da câmera. Esta categoria também deve ser escolhida quando o codificador tiver identificado outros estados afetivos além daqueles previstos no protocolo. Por exemplo, quando os alunos mostram surpresa, que é uma emoção não considerada neste protocolo.

Em relação aos **comportamentos**, seis categorias também são anotadas para este protocolo: *on task* (na tarefa), *on task conversation* (conversação na tarefa), *on task out* (na tarefa fora do ambiente), *on system* (no ambiente), *off task* (fora da tarefa) e “?”. Esses comportamentos são os mesmos usados no protocolo BROMP (OCUMPAUGH, 2015), exceto o comportamento *on system*.

- **On task:** Os alunos estão *on task* quando estão concentrados em resolver a tarefa atual, que no caso do trabalho proposto, é a resolução de uma equação no PAT2Math. No entanto, uma pequena viradinha pro lado ou uma breve resposta a um colega de classe, sem tirar a atenção dos alunos da tarefa, ainda é considerado *on task*, se durar menos de dois segundos.
- **On system:** O aluno está utilizando o ambiente de aprendizagem, mas não está focado na resolução da sua tarefa (equação), por exemplo quando o aluno está olhando o *ranking*, verificando quais os planos já estão desbloqueados ou ainda escolhendo uma equação do plano atual para resolver.
- **On task conversation:** Refere-se à um comportamento quando os alunos estão trabalhando em uma tarefa enquanto conversam com o professor ou outro aluno especificamente sobre a tarefa que estão realizando. Basicamente, esse comportamento designa quando um aluno está recebendo ajuda na tarefa sendo realizada. Quando os alunos estão falando sobre outras coisas além da tarefa que estão fazendo, por exemplo, o sistema (desempenho do sistema, pontos de gamificação, mensagens, conteúdos) ou qualquer coisa relacionada ao aprendizado (meta cognitivo, habilidades históricas, avaliação de um erro, entre outros), isso deve ser classificado como comportamento *off task*. Da mesma forma, quando os alunos estão ajudando um colega a resolver outra tarefa, ele também deve ser classificado como *off task*. Embora essa classificação possa ser considerada não muito intuitiva (porque os alunos também aprendem quando explicam), é importante ter em mente que as anotações serão usadas por um classificador que visa inferir com base nos *logs* das ações dos alunos quando eles estão resolvendo sua própria tarefa (ou engajados na resolução da sua tarefa).
- **On task out:** Os alunos estão *on task out* quando estão trabalhando em suas tarefas, mas não estão usando o ambiente de aprendizagem. Por exemplo, quando os alunos usam cadernos para realizar cálculos intermediários ou anotações. Quando os alunos estão *on task out* e falam sobre sua tarefa ou desempenho de aprendizagem, isso ainda é considerado *on task out*. A palavra “*out*” significa que o aluno não está usando o sistema, mas ainda está trabalhando em sua tarefa; ou seja, *fora* do sistema.
- **Off task:** O aluno não está trabalhando na tarefa atribuída pelo sistema. Por exemplo, quando o aluno está conversando com colegas, está ouvindo e prestando atenção em conversas paralelas ou quando está ajudando um colega com a resolução de uma tarefa ou com a utilização do ambiente.
- **Comportamento “?”:** Esta categoria inclui qualquer comportamento que não esteja na lista de comportamentos deste protocolo. Por exemplo, quando os alunos saem das suas cadeiras para irem ao banheiro ou se abaixam para pegar algum material que caiu. Assim, uma regra que se aplica é que este comportamento deve ser usado quando o aluno

não está no vídeo ou em algum momento quando o codificador tem dúvidas sobre qual comportamento codificar.

A combinação dos estados afetivos e dos comportamentos também podem proporcionar um significado diferente. Assim, foi elaborado uma lista com algumas possíveis ações dos alunos e as combinações que foram utilizadas. Essa lista é descrita na Tabela 5.

Tabela 5 – Ações dos alunos e as combinações que foram utilizadas.

Ação	Estado Afetivo	Comportamento
<i>Gaming the system</i>	Qualquer estado menos engajamento	<i>On task</i>
Conversando sobre sua própria tarefa (alguém ajudado o aluno)	Engajado (se houver) esforço cognitivo	<i>On task conversation</i>
Em caso de dúvida do anotador (por não ter indícios suficientes)	?	?
Qualquer situação em que o aluno não está usando o sistema de forma focada, com exceção do <i>on task out</i>	Qualquer estado	<i>Off task</i>
Rosto totalmente escondido	?	Qualquer comportamento
Está no ambiente mas não está resolvendo uma tarefa (ver <i>ranking</i> , acessar plano)	Qualquer estado menos engajado	<i>On system</i>
<i>Bug</i> no sistema	Qualquer estado	<i>On system</i>
Lendo dica	Qualquer estado	<i>On task</i>
Emoção na transição	Anotar no 2º clipe	-
Ajudar colega de forma concentrada	Engajamento concentrado	<i>Off task</i>

Fonte: Elaborado pelo autor.

Baseado nas definições dos termos descritos anteriormente, algumas variáveis foram definidas com base na necessidade deste trabalho. A primeira definição é em relação ao tamanho do **clipe**, que foi definido em cinco segundos. Apontam-se duas justificativas para esta definição, a primeira delas é que as emoções possuem uma curta duração (SCHERER, 2005). Segunda, inicialmente foi considerado um tamanho de 10 segundos, porém por meio de observações empíricas foi possível identificar que esse tamanho era inadequado. Assim, passou-se a utilizar o clipe com a duração de cinco segundos. Para chegar neste valor, mais de 20 tarefas de anotações foram realizadas, onde notou-se que quando o tamanho do clipe era maior do que cinco segundos, os alunos experienciavam mais do que um estado por clipe. Desta forma, os codificadores eram forçados a escolher os estados mais representativos. O problema é que esta estratégia de escolher um estado mais representativo causa uma perda de dados, fazendo com que os rótulos não representem realmente as transições de estados afetivos do aluno. Ainda, por ser um STI baseado em passos, PAT2Math oferece uma grande quantidade de *feedback* ao aluno. A cada

passo que o aluno insere, o sistema informa o aluno se esse passo está correto e, se estiver incorreto, também apresenta um dica para o aluno prosseguir. Ao analisar os dados do módulo do aluno do PAT2Math, foi possível identificar que essas interações, do aluno inserindo passos no sistema, ocorriam com uma frequência entre cinco segundos a mais de um minuto. Como o sistema fornece um *feedback* mínimo por interação, esse *feedback* pode elicitare diferentes estados afetivos no aluno. Desta forma, o tamanho do clipe em cinco segundos representa melhor a duração dos estados afetivos que os alunos experimentam durante a utilização do STI PAT2Math.

Outras definições em relação ao tempo foram feitas para a **sessão**. O tempo de duração de uma sessão de anotação foi definido em cinco minutos. A escolha das sessões de cinco minutos deu-se porque foi considerado que era tempo suficiente para o codificador estar ciente do contexto do aluno sem deixar o codificador entediado ou cansado com uma tarefa longa. Além disso, também foi definido o tempo de início da sessão em relação ao vídeo. Os vídeos obtidos possuíam duração média de 40 minutos (duração da aula de matemática). Desta forma, foi definido que a sessão teria início 10 minutos após o início do vídeo. Esses 10 minutos foram desconsiderados dos vídeos para descartar a fase inicial, quando os alunos estão logando no sistema, se acomodando na cadeira e no computador e mais conscientes da câmera. Portanto, se a duração da sessão é igual a cinco minutos e o tamanho do clipe é cinco segundos, para cada sessão, cada codificador gera 60 anotações.

Este protocolo tem uma compreensão incremental dos conceitos de estados e comportamentos afetivos. Assim, foi definido o número de **codificadores** igual a três, que neste caso são três membros do grupo de pesquisa em informática na educação do Programa de Pós-Graduação em Computação Aplicada (PPGCA) da UNISINOS. Dessa forma, é possível que dois codificadores identifiquem um estado ou comportamento enquanto um terceiro discorda, gerando assim um ponto de discussão durante a fase de treinamento. Para as **anotações**, foi definido que cada anotação representa um ou mais estados afetivos e um ou mais comportamentos do aluno. Assim, o protocolo permite várias anotações de estados e comportamentos dentro de cada clipe. Para fazer uma anotação de estados afetivos e comportamentos, foi definido que a manifestação deve ter uma duração de pelo menos dois segundos para ser anotada. Assim, por exemplo, quando o aluno dá uma rápida olhada em algo fora da tela do computador por um período menor que dois segundos, o codificador não deve anotá-lo como comportamento *off task* devido ao tempo de duração do evento. Neste caso, como a duração dos cliques é de cinco segundos, os codificadores são capazes de anotar pelo menos um estado afetivo e um comportamento e no máximo dois estados afetivos e dois comportamentos em cada clipe. Para que o codificador possa observar a duração dos eventos, foi desenvolvido uma ferramenta *web* para anotação que possui um relógio marcando o tempo do clipe, descrito com mais detalhes na próxima seção.

Foram definidas algumas regras para a identificação de estados afetivos e comportamentos. Como cada codificador pode anotar mais de um estado afetivo ou comportamento por clipe, é importante ressaltar que a ordem em que esses estados ou comportamentos são codificados tam-

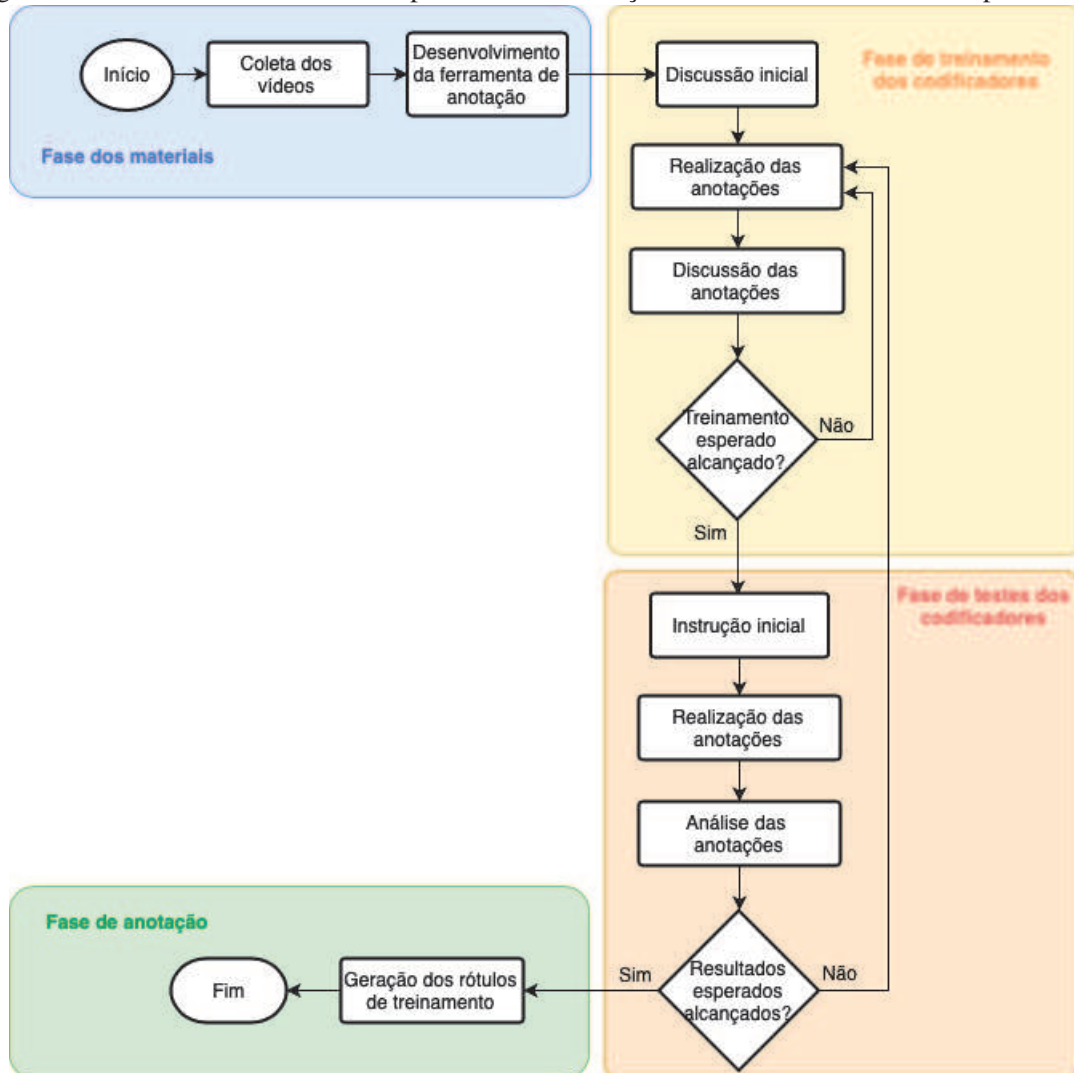
bém é armazenada. Ou seja, é importante anotar os estados e comportamentos conforme eles ocorrem ao longo do clipe. Para ajudar o codificador, um *feedback* visual na interface gráfica da ferramenta desenvolvida mostra ao codificador qual foi a ordem na qual os estados ou comportamentos foram codificados. Quando apenas um estado ou comportamento é identificado em um clipe, o codificador deve selecionar apenas esse estado/comportamento, o que significa que não há ordem de estados e comportamentos nesse clipe. Além disso, um requisito ao identificar um estado ou comportamento é que ele deve estar presente por pelo menos dois segundos do clipe, conforme indicado anteriormente. Um clipe deve conter pelo menos um estado afetivo e um comportamento anotado. Os estados e comportamentos afetivos dependem do clipe. Ou seja, se um codificador anota um estado afetivo e um comportamento, esse comportamento é associado ao tempo integral do estado afetivo anotado (tempo relacionado ao tamanho do clipe). Se um codificador anotar dois estados afetivos e apenas um comportamento, esse comportamento é associado aos dois estados afetivos anotados. Se o codificador anotar dois estados afetivos e dois comportamentos, isso significa que o primeiro estado afetivo está relacionado ao primeiro comportamento e o segundo estado ao segundo comportamento. Portanto, a ordem das anotações é importante. Quando o codificador não tem certeza de qual estado/comportamento afetivo escolher, ele deve escolher anotar a categoria “?” para o estado afetivo ou comportamento.

6.5.4 Fluxo de Troca de Fases do Protocolo

Conforme mencionado no início desta seção, o protocolo desenvolvido contém quatro fases. A fase inicial define a coleta e desenvolvimento de materiais e ferramenta de anotação. Uma vez que os vídeos foram coletados e a ferramenta de anotação foi desenvolvida, os codificadores iniciaram a discussão inicial sobre as definições do protocolo, bem como a configuração da ferramenta de anotação. Após essa discussão e com a ferramenta de anotação configurada, os codificadores realizaram uma etapa de anotações e a discussão sobre as anotações. Se eles identificassem que haviam muitas divergências nos resultados, significava que eles precisavam melhorar sua compreensão das definições e retomar a tarefa de realização das anotações e a discussão sobre as anotações. Para cada iteração, ciclo, de realização e discussão das anotações, foram comparados 180 anotações de estados afetivos e 180 de comportamentos, sendo 60 destas anotações provenientes de cada codificador. É importante destacar que um ciclo é um conjunto de realização das anotações e discussão sobre as anotações. Uma vez que a divergência dos resultados fosse muito pequena, os codificadores estavam aptos a ir para a fase de teste dos codificadores. Esta fase é bastante semelhante à fase de treinamento. No início, os codificadores têm uma instrução inicial para decidir os parâmetros da fase de teste, para então realizarem as anotações. No entanto, não há discussão após a tarefa de fazer anotações. Em vez disso, os codificadores analisam os resultados por meio de medições estatísticas. Esta estratégia de não incluir a análise estatística durante a fase de treinamento foi escolhida para não liderar o treinamento à obtenção de um limiar, mas para uma maior e verdadeira compreensão dos

conceitos de estados afetivos e comportamentos adotados pelo protocolo. Se os codificadores obtiverem um bom resultado na fase de teste, isto é, um Randolph Kappa igual ou maior que 0.6, então eles são considerados capazes de gerar os rótulos de estados afetivos e comportamentos sozinhos, isto é, a fase de anotação. Se não, os codificadores devem voltar à fase de treinamento e reforçarem o entendimento sobre os conceitos, realizando novos ciclos de treinamento. A Figura 9 ilustra o fluxo de troca de fases, conforme descrito anteriormente.

Figura 9 – Fluxo de troca de fases do protocolo de anotação dos estados afetivos e comportamentos.



Fonte: Elaborado pelo autor.

6.5.5 Fase dos Materiais

A fase dos materiais descreve a coleta dos vídeos para serem analisados e o desenvolvimento da ferramenta necessária para a coleta das anotações. Assim, essa fase é dividida nessas duas tarefas, que serão descritas com mais detalhes.

- **Coleta dos vídeos:** Esta fase descreve a coleta dos vídeos utilizados neste protocolo

para a obtenção dos rótulos de estados afetivos e comportamentos dos alunos. Os vídeos foram coletados durante a coleta de dados descrito na seção 6.2.2. O tempo médio dos vídeos é de aproximadamente 40 minutos. Dentro dos 10 dias de coleta de dados, mais de 230 vídeos completos foram gravados. Durante a coleta, foram gravados dois vídeos por aluno, um com rosto e áudio e outro contendo a tela do computador do aluno, enquanto ele utiliza o STI PAT2Math. Para capturar o rosto e o áudio ambiente, foi usado uma **webcam** com microfone e este vídeo foi gravado em Alta Definição (HD). Para gravar os vídeos, foi utilizado o *software* online Wistia Soapbox (<https://soapbox.wistia.com>), que funciona como um *plugin* para o navegador Google Chrome e grava vídeos (tela do computador e rosto do aluno) de forma simultânea. Uma característica interessante deste *plugin* é que ele não mostra na tela do computador que o aluno está sendo gravado, mesmo para a **webcam**. Assim, os alunos não conseguiam ver o seu próprio rosto enquanto estavam sendo gravados.

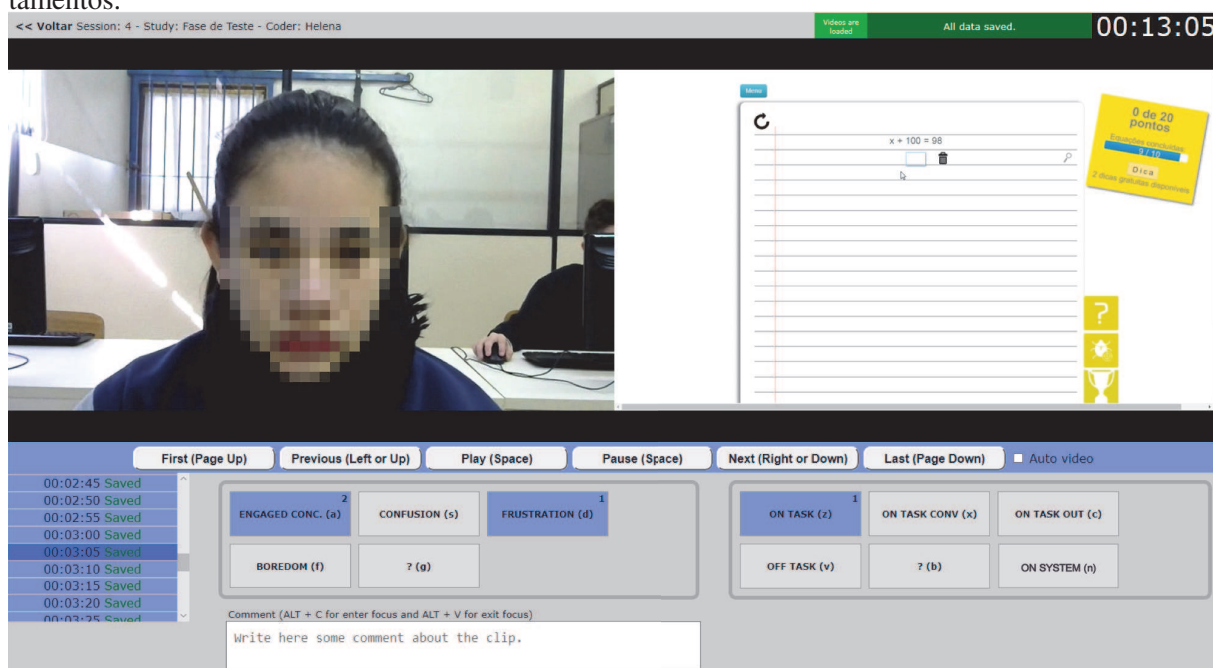
Algumas precauções foram tomadas para que os vídeos fossem gravados. Todos os alunos usaram o sistema ao mesmo tempo. Cada aluno usou um único computador. Como o uso do ambiente educacional foi uma atividade proposta pela escola, os alunos não puderam realizar nenhum tipo de atividade paralela ao uso do ambiente. Assim, os alunos colocaram o sistema em tela cheia e não podiam usar outro ambiente ou sistema de computador. Antes de iniciar a gravação dos vídeos, os alunos foram instruídos a permanecer em uma postura confortável, para que as *webcams* pudessem ser devidamente ajustadas de acordo com cada aluno.

- **Ferramenta de anotação:** Para a anotação dos estados afetivos, foi desenvolvida uma ferramenta computacional *web*, que recebe os dois vídeos (face com áudio e tela) capturados simultaneamente (ver Figura 10). A ferramenta exibe os dois vídeos simultaneamente e de forma síncrona, ou seja, para a anotação, o codificador escuta e vê o aluno e o que ele está fazendo no sistema ao mesmo tempo, lado a lado. A anotação de estados afetivos e comportamentos é realizada ao visualizar os dois vídeos simultaneamente. Para completar o processo de anotação, o codificador deve selecionar o estudo e a sessão a ser anotada.

As sessões têm um vídeo completo, hora de início e tamanho da sessão, conforme descrito na seção de definições. Ao acessar uma sessão, as informações predefinidas são carregadas e configuradas automaticamente. Assim, quando os codificadores reproduzirem o vídeo, eles verão o primeiro clipe para inserir sua anotação, que possui um tamanho de cinco segundos. O sistema reproduz o vídeo durante o tempo do clipe e para automaticamente quando esse tempo acabar. Assim, os anotadores podem refletir sobre sua anotação e inseri-la no sistema, que será armazenada automaticamente no banco de dados, com base na ordem de seleção do anotador, para estados afetivos e comportamentos. Se necessário, os codificadores podem rever o clipe quantas vezes quiserem para terem uma

maior precisão em seu julgamento. Depois de realizar a anotação do clipe, o codificador simplesmente pressiona o botão próximo, que reproduzirá o próximo clipe e limpará os dados da última anotação. A ferramenta armazena localmente as anotações executadas pelo codificador para que possam ser revisadas e modificadas se os codificadores considerarem necessário. Sabendo que o codificador pode selecionar mais de um estado afetivo ou comportamento por clipe, a interface gráfica da ferramenta exibe um número, indicando a ordem da anotação para que o codificador tenha certeza sobre a sequência das anotações realizadas. Após ter realizado o processo de anotação, a ferramenta indica ao codificador o término da sessão e salva todas as anotações no banco de dados.

Figura 10 – Ferramenta desenvolvida utilizada para coletar as anotações dos estados afetivos e comportamentos.



Fonte: Elaborado pelo autor.

6.5.6 Fase de Treinamento dos Codificadores

Esta fase tem como objetivo descrever o procedimento realizado para o treinamento de codificadores na detecção de estados afetivos e comportamentos. A fase de treinamento dos codificadores foi dividida em três tarefas, chamada (i) discussão inicial, (ii) realização das anotações e (iii) discussão das anotações. As tarefas ocorreram em sequência, mas a fase de treinamento como um todo representa vários ciclos de repetição de (ii) realização e (iii) discussão das anotações até que um nível de treinamento desejado seja alcançado.

1. **Discussão inicial:** Nesta tarefa foram esclarecidos os conceitos sobre os estados afetivos e os comportamentos a serem anotados com este protocolo, visando alcançar um entendimento mútuo. Para isso, foram estudadas teorias já conhecidas como estado da arte,

para que o entendimento compartilhado esteja correto e de acordo com a literatura. Além disso, nesta discussão também foram realizadas algumas definições, como o número de sessões para essa fase (inicialmente definida como uma única), para que os anotadores pudessem sentir como era usar a ferramenta e como estava a compreensão dos conceitos. A duração de uma sessão foi definida em cinco minutos e a hora de início foi definida em 10 minutos. O tamanho do clipe foi estabelecido em 10 segundos, com a anotação de um único estado afetivo e comportamento, nesta primeira discussão.

2. **Realização das anotações:** Esta tarefa descreve o processo de anotação dos estados afetivos, na qual a ferramenta desenvolvida foi utilizada com os vídeos coletados e com base nas definições e conceitos vistos na discussão inicial. Os três codificadores anotam a mesma sessão, permitindo a sincronização das anotações geradas para que uma comparação possa ser realizada na tarefa de discussão das anotações. Nesta tarefa, o codificador assiste o vídeo completo a partir do horário de início da sessão e anota clipe por clipe até o final da sessão, definido pela duração da sessão. Para cada clipe, o codificador seleciona um, pelo menos, ou dois estados afetivos e um, pelo menos, ou dois comportamentos. Se os codificadores tiverem dúvidas sobre qual estado ou comportamento anotar, eles podem reproduzir o clipe quantas vezes acharem necessário. No final do tempo da sessão e com todos os cliques contendo suas respectivas anotações (estado(s) afetivo(s) + comportamento(s)), a realização das anotações estava concluída.
3. **Discussão das anotações:** Esta tarefa consiste em coletar todas as anotações feitas por cada codificador em uma sessão. Depois que esses dados foram recuperados, essas anotações foram analisadas e comparadas entre os anotadores para a mesma sessão. O objetivo desta discussão era avaliar os resultados, de forma qualitativa, obtidos a partir da anotação do ciclo atual e identificar os pontos fortes e fracos dos codificadores. Assim, se fosse observado que os codificadores ainda não possuíam um entendimento mútuo para algum comportamento ou estado afetivo, um novo ciclo era realizado com maior eficiência e precisão dos codificadores, voltando à tarefa de realização das anotações.

A Tabela 6 mostra um exemplo de comparação de anotações dos três anotadores para cinco cliques da mesma sessão. A partir da análise uma a uma dessas anotações, foi possível identificar inconsistências nas anotações, por exemplo, nos cliques 2, 3 e 5. Assim, os codificadores revisaram o clipe e conversaram entre si para entender o ponto de vista de cada um. Desta forma, em grupo, foi possível refletir e tirar conclusões sobre qual estado afetivo ou comportamento representava melhor aquele clipe.

Após a discussão inicial, foi realizada o primeiro ciclo da fase de treino dos codificadores. Foi a primeira vez que todos os codificadores usaram a ferramenta desenvolvida. Após a realização das anotações, foi realizada a discussão das anotações. Ao analisar as anotações, conforme exemplificado na Tabela 6, foi possível notar algumas divergências na compreensão

Tabela 6 – Exemplo de cinco anotações geradas pelos três codificadores para uma mesma sessão de anotação.

Clipe	Codificador 1	Codificador 2	Codificador 3
1	Confusão	Confusão	Confusão
2	Tédio	Tédio	Frustração
3	Engajado	Tédio	Engajado
4	Engajado	Engajado	Engajado
5	Frustração	Confusão	Confusão

Fonte: Elaborado pelo autor.

dos conceitos. Para cada divergência encontrada, o respectivo clipe era visto novamente pelos três codificadores para resolver os conflitos e aprimorar as definições dos estados afetivos e comportamentos. Neste primeiro ciclo de treinamento, o tamanho do clipe era de 10 segundos e os codificadores podiam anotar apenas um estado afetivo e um comportamento por clipe. Assim, foi possível identificar que um clipe de 10 segundos era muito longo e gerou muitas dúvidas sobre qual estado ou comportamento representava melhor o clipe, quando mais de um acontecia.

Para o segundo ciclo da fase de treinamento, o tempo de clipe foi reduzido para cinco segundos. Assim, a segunda tarefa de realização de anotações e discussão das anotações foram realizadas, comparando anotações e vendo os cliques novamente. Ainda foi identificado que, mesmo com cliques de cinco segundos, as anotações não representavam fielmente as transições entre estados afetivos e comportamentos apresentados pelos alunos. Assim, a ferramenta foi modificada para que em cada clipe os codificadores pudessem observar uma sequência de estados afetivos e comportamentos, ao invés de um único, permitindo uma reprodução mais fiel das transições dos estados afetivos e dos comportamentos dos alunos. Além disso, foi definido que um mínimo de dois segundos é necessário para a tomada de decisão de qual estado ou comportamento deve ser selecionado. Portanto, com um clipe de cinco segundos, o número máximo de estados e comportamentos que podem ser selecionados por clipe é igual a dois.

Após este segundo ciclo, foi iniciado um novo ciclo de realização das anotações e discussão. Neste momento, todos os três codificadores concordaram que um clipe de cinco segundos com múltiplos estados e comportamentos era mais adequado. Foram utilizadas diferentes sessões para cada um dos ciclos de treino, mas todos os codificadores estavam codificando a mesma sessão por ciclo. Como foi definido, todos os codificadores também concordaram que o tempo de início da sessão de 10 minutos e a duração de cinco minutos estavam apropriados. Assim, mais sete ciclos da fase de treinamento foram realizados pelos três codificadores. Totalizando 10 ciclos de treinamento. No décimo ciclo, todos os codificadores concordaram que havia um entendimento mais consolidado de cada um dos estados afetivos e comportamentos e que as configurações do protocolo estavam de acordo. Isso permitiu passar para a fase de testes dos codificadores.

6.5.7 Fase de Teste dos Codificadores

Esta fase começou com uma conversa de instrução inicial, em que os codificadores definiram os parâmetros para a fase de teste. Nesse caso, os valores foram os mesmos da fase de treinamento. Durante as sessões de treinamento, que possuíam duração de cinco minutos, a tarefa de anotação levava cerca de 20 minutos para ser concluída por cada codificador. Isso ocorre porque o codificador pode reproduzir o clipe várias vezes para tomar a decisão correta. Foi concordado que essa tarefa precisa de atenção total e, durante longos períodos, pode causar algum cansaço aos codificadores. Assim, optou-se manter a duração da sessão de cinco minutos. Além disso, os valores para o tempo de início foram mantidos em 10 minutos e o tamanho do clipe em cinco segundos, com a possibilidade de múltiplos estados afetivos e comportamentos para cada anotação. Também foi definido três sessões para essa fase, com vídeos que nunca haviam sido vistos durante a fase de treinamento.

Depois de concluir a instrução inicial, foi iniciada a realização da anotação para a fase de testes. Neste momento, nenhuma discussão entre as sessões foi realizada. No final, as anotações foram agrupadas para a realização de análises estatísticas. Foi conduzido o Randolph Kappa separadamente para anotações de estados afetivos e de comportamentos. Antes de realizar as medições, foi necessário calcular o número de codificadores que selecionaram cada construto para cada clipe. Quando múltiplos estados afetivos ou comportamentos eram anotados no mesmo clipe, a comparação era realizada indiferente da ordem em que foram anotadas, chamado de acordo parcial, se os três anotadores tivesse anotado o mesmo construto. Ou seja, se todos os codificadores anotassem o mesmo estado afetivo ou comportamento para uma das duas anotações possíveis, seria considerado como um acordo parcial, independentemente da ordem. Em seguida, foi identificado qual era o estado ou comportamento parcial selecionado e seu valor foi alterado para três (o número de codificadores). Caso contrário, cada rótulo recebe o número de codificadores que o anotou. Cada linha representa a anotação de um clipe. Cada coluna representa um estado afetivo para o cálculo do estado afetivo e cada coluna representa um comportamento para o cálculo do comportamento. Assim, a soma de cada linha deve ser igual a três. Foi decidido não remover as anotações do cálculo quando um codificador anotou a categoria "?", tanto para estados afetivos quanto para comportamentos. Essa estratégia é diferente de outros protocolos, como o BROMP, que remove essa linha inteira. No entanto, acredita-se que se o codificador tenha anotado essa categoria, é porque ele/a tem certeza disso. No caso da anotação *online*, o codificador pode ter perdido parte da informação para a tomada de decisão. Porém, neste protocolo os anotadores podem rever o clipe quantas vezes julgarem necessário.

Atualmente, os trabalhos que realizam este tipo de comparação utilizam o cálculo do valor de Kappa, proposto por Cohen's. (COHEN, 1960). O coeficiente Kappa K é uma medida estatística de concordância entre avaliadores para valores categóricos nominais. Considera-se mais robusto do que um simples cálculo de percentual de concordância, já que K leva em conta

a concordância que ocorre por acaso. De acordo com o protocolo BROMP (OCUMPAUGH, 2015), um valor de Kappa considerado adequado para este tipo de análise é maior ou igual que 0.6, onde o valor de K varia de -1 a 1 , sendo 1 representando um acordo perfeito. A grande questão aqui é que o Cohen's Kappa é uma medida estatística para o acordo entre dois codificadores. Como este protocolo utiliza três codificadores, para que a fase de treinamento se torne mais confiável para discussões, essa medida não pôde ser usada.

Embora existam várias medidas de concordância para três ou mais codificadores, como Light (LIGHT, 1971), Fleiss (FLEISS, 1971) e Hubert (HUBERT, 1977), a medida de Randolph (RANDOLPH, 2005) foi adotada. Esta escolha foi feita com base nas recomendações de Randolph. (RANDOLPH, 2005). Em seu trabalho, o autor sugere que o kappa de Fleiss para múltiplos codificadores, a medida mais usada neste caso, não é apropriado quando os marginais não são fixos. Ou seja, Randolph é recomendado quando os codificadores não conhecem a priori a distribuição dos casos em cada categoria. Assim, o valor do Kappa não varia em função da simetria das distribuições marginais, em vez disso, varia em função do número de categorias de anotação utilizadas. Para este protocolo, não há como saber a distribuição de probabilidade para cada categoria, uma vez que os vídeos analisados são de diferentes alunos e cada aluno pode apresentar diferentes estados afetivos e padrões de comportamento. Além disso, uma recomendação de Randolph (2005) é que as categorias devem ser bem especificadas, porque podem inflar os resultados. No caso deste protocolo, todas as categorias de estados afetivos e comportamentos também foram utilizadas em outras pesquisas. Com base nas definições sugeridas por Ocumpaugh (2015), para este protocolo, foi assumido que um limite de aceitação para os codificadores serem capazes de trabalhar sozinhos na geração dos rótulos é um Randolph Kappa igual ou maior à 0.6.

Com o objetivo de comparar o nível de concordância entre os três anotadores, foi aplicado o teste estatístico Randolph Kappa. Um total de 180 rótulos (por codificador) de estados afetivos e comportamentos, provenientes das anotações de três sessões, foram considerados. Foram obtidos como resultados um total Randolph Kappa de 0,72 para estados afetivos, com 77,53% de concordância, considerando as cinco categorias de estados afetivos. Para a anotação de comportamentos, foi obtido uma Randolph Kappa de 0,78, com 81,58% de concordância, considerando as seis categorias de comportamentos. Desta forma, os codificadores apresentaram resultados satisfatórios na concordância entre as anotações, tornando-os aptos a irem para a próxima fase do protocolo. Para o cálculo do Randolph Kappa, foi utilizado uma ferramenta *web*, desenvolvido pelo próprio autor do método. (RANDOLPH, 2008).

6.5.8 Fase de Anotação

Depois de concluídas as fases de treinamento e teste, os codificadores são capazes de realizar o processo de anotação independentemente um do outro, ou seja, cada codificador pode anotar diferentes sessões, sem fazer algum tipo de comparação entre elas, uma vez que já foram

devidamente treinados e avaliados. Essa estratégia também é assumida por outros protocolos, como o BROMP. (OCUMPAUGH, 2015). O principal objetivo desta fase é gerar um conjunto de rótulos que serão utilizados no desenvolvimento dos detectores de estados afetivos desta dissertação.

Com base no protocolo proposto, um total de 30 vídeos foram selecionados para serem analisados, sendo estes provenientes das dez sessões de coleta de dados. Ainda, para uma seleção equilibrada dos vídeos entre sessões, três vídeos foram selecionados por cada sessão de coleta, resultando nos 30 vídeos a serem analisados. Os vídeos foram selecionados aleatoriamente entre os alunos, mas com o cuidado de não selecionar o mesmo aluno mais de uma vez. Assim, os 30 vídeos são de 30 alunos diferentes. Como definido na fase de treino dos codificadores, as sessões de anotação eram de cinco minutos, resultando em um total de 150 minutos vídeos anotados. Cada sessão leva em torno de 20 minutos para ser anotada, totalizando uma média de 600 minutos para realizar todas as anotações. Como o tamanho do clipe é de cinco segundos, foram realizadas 1800 anotações. Porém, como os codificadores podem anotar mais de um estado afetivo ou comportamento por anotação, o total de rótulos de estados afetivos foi de 2099 e 2059 para comportamentos, exibidos com mais detalhes na Tabela 7.

Tabela 7 – Resultados das anotações de estados afetivos e comportamentos.

		Quantidade	Percentual	Total
Estados afetivos	Engajamento	1118	53,3%	2099
	Confusão	463	22,1%	
	Tédio	90	4,3%	
	Frustração	51	2,4%	
	?	377	18,0%	
Comportamentos	<i>On task</i>	1174	57,0%	2059
	<i>Off task</i>	291	14,1%	
	<i>On system</i>	245	11,9%	
	<i>On task out</i>	177	8,6%	
	<i>On task conversation</i>	159	7,7%	
	?	13	0,6%	

Fonte: Elaborado pelo autor.

6.6 Síntese das Características

As características, do inglês *features*, são informações calculadas para cada *log* gerado pelo aluno, que juntas tem o objetivo de representar o estado do aluno/sistema em um determinado momento. Esta sessão tem como objetivo descrever todas as características utilizadas para o desenvolvimento dos detectores. Conforme descrito na seção 6.2.2, foram coletados 386.015 *logs* durante a coleta de dados realizada com alunos. Destes dados, foram selecionados somente aqueles que foram gerados no mesmo instante em que os vídeos, do protocolo de coleta dos estados afetivos e comportamentos, foram analisados. Esta seleção foi realizada por um processo de sincronização entre os *logs* e os rótulos de estados afetivos e comportamentos gerados.

Assim, um total de 5.525 *logs* foram utilizados para o treinamento dos modelos.

Cada um dos *logs* selecionados possuía um total de 348 características, provenientes de quatro fontes de dados: (i) interação do aluno com a interface gráfica (*logs*), descrito na seção 6.2, (ii) módulo do aluno, descrito na seção 6.3, (iii) questionário de personalidade, descrito na seção 6.4, e (iv) rótulos de estados afetivos e comportamentos, descrito na seção 6.5. Para os dados categóricos, foi aplicada a técnica de coluna binária, do inglês conhecida como *one-hot encoding*. Assim, para cada categoria possível de uma variável categórica, foi criada uma variável que representava essa categoria. Dessa forma, essa variável assumia somente valores binários, onde “0” indica a ausência desta categoria e “1” indica a presença desta categoria. Essa estratégia foi adotada pois alguns dos algoritmos de aprendizagem de máquina, como modelos de regressão, não são capazes de lidar com dados categóricos. (BECK; WOOLF, 2000; FEURER et al., 2015; CHEN; GUESTRIN, 2016).

Dos **dados de Logs de interação** do aluno com o sistema, foram utilizadas 31 características, as quais são listadas abaixo, contendo também o tipo da característica.

1. (Inteiro) Número do dia, variando de 1 até 10.
2. (Inteiro) Número do *log*, variando de 1 até 175, em média.
3. (Binário) Último passo de resolução estava correto.
4. (Binário) Passo de resolução foi verificado com o *mouse* do computador.
5. (Binário) Passo de resolução foi verificado com o teclado do computador (pressionando a tecla “enter”).
6. (Inteiro) Tempo ocioso acumulado (aluno sem interação com o sistema)
7. (Inteiro) Número de cliques seguidos.
8. (Inteiro) Número de cliques seguidos no passo atual de resolução.
9. (Inteiro) Número de cliques seguidos na equação atual.
10. ao 22. (Binário) Uma característica binária para um dos 13 possíveis tipos de *logs*: verificação de passo, clique, tecla pressionada, parada do *mouse*, *mouse* rápido, pedido de ajuda, aluno ocioso, navegador ganhou foco, navegador perdeu foco, navegador maximizado/minimizado, aluno saiu da página, aluno voltou para página e aluno fez *log out*.
23. ao 31. (Binário) Uma característica binária para cada um dos nove grupos de cliques: resolução, planos, atalhos, *bugs*, gamificação, dicas, erros, notificações e sistema.

Dos **dados do módulo do aluno**, foram utilizadas 298 características:

32. (Inteiro) Pontuação máxima da equação atual.

33. ao 44. (Binário) Uma característica binária para cada um dos 12 níveis de dificuldade da equação: muito fácil, fácil 1, fácil 2, fácil 3, fácil 4, médio 1, médio 2, médio 3, difícil 1, difícil 2, difícil 3, difícil 4.
45. (Inteiro) Número de passos dados na equação atual.
46. (Inteiro) Número de passos errados na equação atual.
47. (Inteiro) Número de passos errados seguidos na equação atual.
48. (Inteiro) Número de passos corretos na equação atual.
49. (Decimal) Efetividade dos passos na equação atual (número de passos corretos dividido pelo número total de passos).
50. (Inteiro) Tempo na resolução da equação atual, em segundos.
51. (Inteiro) Tempo no passo atual, em segundos.
52. (Decimal) Tempo médio gasto na resolução das equações resolvidas desde o início do dia até o *log* atual (diário), em segundos.
53. (Decimal) Tempo médio gasto nos passos de resolução das equações diário, em segundos.
54. (Inteiro) Número de equações resolvidas diário.
55. (Inteiro) Número de passos de resolução diário.
56. (Inteiro) Número de passos corretos diário.
57. (Inteiro) Número de passos incorretos diário.
58. (Decimal) Efetividade dos passos diário.
59. ao 65. As mesmas características dos itens 52 ao 58, porém utilizando os dados de toda a turma. Assim, foi calculado a média da turma para cada um dos itens especificados.
66. (Decimal) Tempo médio gasto na resolução das equações resolvidas desde o primeiro dia de coleta até o *log* atual (total), em segundos.
67. (Decimal) Tempo médio gasto nos passos de resolução das equações total, em segundos.
68. (Inteiro) Número de equações resolvidas total.
69. (Inteiro) Número de passos de resolução total.
70. (Inteiro) Número de passos corretos total.
71. (Inteiro) Número de passos incorretos total.

72. (Decimal) Efetividade dos passos total.
73. ao 79. As mesmas características dos itens 66 ao 72, porém utilizando os dados de toda a turma. Assim, foi calculada a média da turma para cada um dos itens especificados.
80. ao 89. (Binário) Uma característica binária para cada uma das 10 possíveis operações que a equação atual precisa para ser resolvida: adição (AD), adição e subtração de frações (AF), propriedade distributiva (DM), divisão (DV), multiplicação de frações (MF), mínimo múltiplo comum (MM), multiplicação (MT), operação inversa (OI), subtração (SB) e simplificação (SP).
90. ao 99. (Inteiro) Quantidade de operações aplicadas corretamente nos últimos cinco segundos, sendo uma característica para cada uma das dez operações.
100. ao 109. Os mesmos dos itens 90 ao 99, porém calculadas para as operações corretas diárias.
110. ao 119. Os mesmos dos itens 100 ao 109, porém calculadas para as operações corretas diárias da turma, ou seja, a média de acerto por operação da turma.
120. ao 129. Os mesmos dos itens 90 ao 99 porém calculadas para as operações corretas total.
130. ao 139. Os mesmos dos itens 100 ao 109, porém calculadas para as operações corretas totais da turma, ou seja, a média de acerto por operação da turma.
140. ao 189. Os mesmos dos itens 90 ao 139, porém calculadas para os erros por operação.
190. ao 239. Efetividade das operações, nos últimos cinco segundos, diário do aluno e da turma e total do aluno e da turma, com base nos itens 90 ao 189.
240. ao 329. Quantidade de *misconceptions* do aluno nos últimos cinco segundos, diário e total. Ao total, o PAT2Math é capaz de detectar 30 *misconceptions* distintas, assim, cada uma delas é calculada individualmente.

Dos **dados da personalidade** do aluno, foram utilizadas 10 características:

330. ao 334. (Decimal) Uma característica decimal para cada um dos cinco traços de personalidade do aluno: amabilidade, extroversão, conscienciosidade, abertura e neuroticismo. O valor de cada um dos traços foi calculado de acordo com o cálculo proposto pelo autor do questionário. Porém, foi realizada uma normalização dos dados. Esta foi feita com base no valor de cada traço dividido pela soma dos valores dos cinco traços. Assim, cada fator assume um valor entre 0 e 1. É importante ressaltar que essas características foram utilizadas somente no desenvolvimento dos últimos quatro detectores.

335. ao 339. (Binário) Além do fator de cada traço de personalidade, também foi calculado se o fator estava mais para o lado positivo ou negativo de cada traço. Assim, o valor 0 era atribuído ao traço se o fator fosse menor que a metade e o valor 1 era atribuído se o fator era maior que a metade do valor máximo do traço. Novamente, é importante ressaltar que essas características foram utilizadas somente no desenvolvimento dos últimos quatro detectores.

E, por fim, dos **dados dos estados afetivos e comportamentos** dos alunos, foram utilizadas nove características.

340. ao 343. (Binário) Uma característica binária para cada um dos estados afetivos do aluno: engajamento, confusão, frustração e tédio. Essas características indicam a presença ou ausência de cada um dos quatro estados afetivos. Porém, essas características nunca foram usadas em conjunto, somente uma por vez como rótulo de treinamento do seu respectivo detector.

344. ao 348. (Binário) Uma característica binária para cada um dos comportamentos dos alunos: *on task*, *on task conversation*, *on task out*, *on system* e *off task*. Novamente, essas características indicam a presença ou ausência de cada um dos cinco comportamentos. Ao contrário dos estados afetivos, essas características foram usadas no desenvolvimento de todos os detectores.

O que é chamado de *log* neste trabalho, quando referido à mineração de dados de *logs*, representa não apenas os dados de interação do aluno com o sistema, mas sim todas as características que são calculadas a partir dessas interações, detalhadas anteriormente. Por exemplo, quando o aluno realiza um clique na interface do sistema, todas as informações do estado atual do aluno no sistema, em relação à conteúdos e conhecimentos, são levados em consideração para o cálculo das 348 características definidas anteriormente. A Figura 11 apresenta o exemplo de uma sequência de *logs* gerada por um aluno, enquanto utilizava o STI PAT2Math.

Figura 11 – Exemplo de uma sequência de *logs* gerada por um aluno utilizando o STI PAT2Math, com seus rótulos de estados afetivos e comportamentos.

aluno	tipo do log	passo correto	último passo correto	tempo idle acumulado	num cliques passo	tempo na equacao	tempo no passo	tempo medio na equacao - diario	tempo medio no passo - diario	efetividade nos passos - diario	...	estado afetivo	comportamento
:	:	:	:	:	:	:	:	:	:	:	:	:	:
17	parou mouse		1	0	0	8	0	29,50	13,963	0,962963	...	engajado	on task
17	clique		1	0	1	9	1	29,50	13,963	0,962963	...	engajado	on task
17	digitou		1	0	1	9	1	29,50	13,963	0,962963	...	engajado	on task
17	idle		1	2	1	11	3	29,50	13,963	0,962963	...	engajado	on task
17	idle		1	4	1	13	5	29,50	13,963	0,962963	...	engajado	on task
17	verificou passo	0	1	0	1	14	6	29,50	13,963	0,962963	...	engajado	on task
17	parou mouse		0	0	0	15	0	29,00	13,752	0,961538	...	confuso	on task
17	pediu dica		0	0	0	16	1	29,00	13,752	0,961538	...	confuso	on task
17	idle		0	2	0	18	3	29,00	13,752	0,961538	...	engajado	on task
17	idle		0	4	0	20	5	29,00	13,752	0,961538	...	engajado	on task
17	idle		0	6	0	22	7	29,00	13,752	0,961538	...	frustrado	on task
17	verificou passo	1	0	0	0	23	8	29,00	13,752	0,961538	...	engajado	on task
17	clique		1	0	1	24	0	29,00	13,511	0,961659	...	engajado	on system
17	clique		1	0	2	24	0	29,00	13,511	0,961659	...	engajado	on system
:	:	:	:	:	:	:	:	:	:	:	:	:	:

Fonte: Elaborado pelo autor.

A Figura 11 apresenta uma sequência de *logs* gerados por um aluno, na resolução parcial de uma equação. Nestes *logs*, é possível identificar cada interação do aluno com a interface do sistema durante alguns segundos. É importante ressaltar que esse exemplo apresenta apenas uma janela de 15 *logs* e de 12 características, sendo o restante omitido para uma melhor visualização. Os três pontos na horizontal representam que existem mais características, totalizando 348, que não estão exibidas nesta imagem e os três pontos na vertical indicam que existem mais *logs* para este mesmo aluno, tanto antes quanto depois desta janela que está sendo exibida.

Ao observar este conjunto, apresentado na Figura 11, é possível identificar que o aluno acertou o passo anterior (*logs* que foram gerados anteriormente e não foram exibidos) e começou a resolução de um novo passo. Porém, neste novo passo o aluno cometeu um erro. Depois, o aluno solicitou uma dica ao sistema e ficou em estado de *idle*, inativo, por seis segundos. Provavelmente, durante este tempo sem interagir com o sistema o aluno estava lendo a dica fornecida pelo STI. Depois, o aluno entrou com um novo passo, desta vez correto. É possível notar que os valores das características vão sendo atualizados e recalculadas a cada novo *log*, de acordo com as novas informações inseridas pelo aluno. O tempo apresentado nas características foi calculado sempre em segundos. Assim, é possível identificar que o aluno pode gerar mais de um *log* por segundo. Por fim, as últimas duas colunas foram anotadas pelos codificadores, durante as sessões de anotação dos estados afetivos. Desta forma, é possível ver que o aluno estava engajado na resolução do passo. Porém, ficou confuso quando cometeu um erro. Após ver a dica solicitada, o aluno percebeu que cometeu um erro por falta de atenção, por isso ficou frustrado por um curto período de tempo. Mas, depois voltou a ficar engajado e entrou com o passo correto. Ainda, este aluno apresentou comportamento *on task* durante todo esse período, ficando *on system* apenas nos últimos dois *logs*, onde ele passou a verificar seu *hanking* na interface do sistema (os cliques realizados foram no grupo de gamificação).

6.7 Desenvolvimento dos Detectores

Esta seção tem como objetivo descrever o desenvolvimento de quatro detectores para a detecção de cada um dos estados afetivos: engajamento, frustração, tédio e confusão. O objetivo principal desta dissertação é verificar se a personalidade dos alunos pode proporcionar uma melhor precisão na detecção destes estados afetivos. Assim, primeiramente foram selecionados cinco algoritmos de aprendizado de máquina, que vêm sendo utilizado pela comunidade, conforme apresentado no capítulo 5, de trabalhos relacionados. Os cinco algoritmos foram utilizados para o desenvolvimento dos detectores de cada estado afetivo. Desta forma, foi possível comparar qual algoritmo obteve o melhor desempenho em cada estado. Com os quatro melhores algoritmos, os detectores foram retreinados com dados de personalidade. Todas as tomadas de decisão para o desenvolvimento dos modelos de detecção, chamados de detectores, foram realizadas com base nos trabalhos relacionados. Estas decisões serão descritas nas próximas seções.

6.7.1 Ferramenta de Desenvolvimento

A primeira decisão é em relação a ferramenta utilizada para o desenvolvimento dos detectores. Todos os trabalhos relacionados utilizam o ambiente RapidMiner Studio, que atualmente se encontra na versão 9.1. Esta ferramenta possui um ambiente gráfico, onde é possível arrastar componentes e interligá-los para a execução de operações. No final da execução das operações, o ambiente apresenta uma lista de resultados, com base nas métricas de avaliação selecionadas. Assim, este ambiente também foi utilizado para desenvolver e avaliar os modelos de detecção desta dissertação.

6.7.2 Seleção dos Algoritmos de Treinamento

A segunda decisão tomada foi em relação à escolha dos algoritmos de aprendizagem de máquina supervisionada. Cinco algoritmos foram selecionados para o desenvolvimento dos detectores desta dissertação. A escolha dos algoritmos foi realizada com base nos algoritmos que obtiveram o melhor desempenho na detecção dos estados afetivos dos oito trabalhos relacionados (ver Seção 5). Com base nos resultados apresentados na Tabela 4, os três melhores algoritmos foram JRip (JR) (sete ocorrências), Logistic Regression (LR) (6 ocorrências) e J48 (J4) (cinco ocorrências). Após, houve um empate entre Naive Bayes (NB) (três ocorrências) e REP-Trees (RT) (três ocorrências) e entre KStar (KS) (duas ocorrências) e Step Regression (SR) (duas ocorrências). Como critério de desempate, foram selecionados os algoritmos que utilizam estratégias diferentes dos que já haviam sido selecionados. RT é da mesma classe de algoritmos de J4, algoritmos baseados em árvores de decisão. SR é da mesma classe de algoritmos do LR, algoritmos baseado em regressão. Assim, foram selecionados NB e KS.

JRip é um algoritmo de classificação baseado em regras, no qual as características são interpretadas como classes. Assim, o objetivo do algoritmo é identificar regras do tipo *if-else* que cubram todos os membros das classes. (RAJPUT et al., 2011). Logistic Regression é um algoritmo de classificação baseado em regressão, utilizado para a classificação de características binárias, ou seja, apenas duas classes de valores. Assim, objetivo desse algoritmo é a geração de um modelo que calcula a probabilidade de associação das características às classes. (DREI-SEITL; OHNO-MACHADO, 2002). J48 é um algoritmo de classificação baseado em árvores de decisão, que tem como objetivo descobrir a forma que as características se comportam de acordo com os dados de treinamento. Assim, o algoritmo gera as regras de predição de acordo com o rótulo alvo. J48 é uma extensão do algoritmo ID3, que é capaz de trabalhar com valores ausentes, *missing values*, redução ou poda da árvore, *tree pruning*, características contínuas, derivação de regras, entre outras. (KAUR; CHHABRA, 2014). Naive Bayes é um algoritmo de classificação baseado em probabilidade, mais especificamente baseado na estatística Bayesiana. A palavra *naive* indica que as características são independentes entre si. Assim, o algoritmo assume que a presença ou ausência de uma característica não é relacionada com a presença ou

ausência de qualquer outra característica. (YUAN, 2010). KStar é um algoritmo de classificação baseado em instâncias. Assim, uma classe é classificada pelo voto majoritário dos dados de treinamento similares a ele. Ainda, a similaridade das instâncias é determinada usando uma medida de distância entrópica. (CLEARY; TRIGG, 1995).

6.7.3 Método de Validação dos Detectores

A terceira decisão também foi tomada de acordo com os trabalhos relacionados. Nesta, todos os trabalhos utilizaram o método de validação cruzada para treinamento dos modelos, do inglês *k-fold cross-validation*. Essa estratégia tem como objetivo dividir toda a base de dados em k conjuntos mutualmente exclusivos de treino e teste. Os detectores são treinados e testados k rodadas. Em cada rodada, os dados de $k - 1$ grupos são utilizados para treinar os detectores e os dados do grupo restante para testar o detector. (KOHAVI et al., 1995). Ao analisar os trabalhos relacionados, verificou-se que houve um empate: três utilizaram o valor de $k = 5$ e três utilizaram o valor de $k = 10$. Assim, o critério de escolha deu-se com base nos dois trabalhos mais recentes, que utilizaram $k = 10$. A estimativa de precisão dos detectores que são treinados e avaliados com a validação cruzada é realizada com base na média dos resultados de todas as k rodadas. A validação cruzada traz resultados mais realistas aos modelos desenvolvidos. Normalmente, algoritmos de aprendizagem de máquina são treinados com uma parte dos dados e testados com o restante dos dados. Em métodos de validação tradicionais, a escolha dos dados que formam esses dois conjuntos é realizada de forma aleatória. Porém, mesmo de forma aleatória, não é garantido que os dados de treino e de teste sejam distribuídos uniformemente. Diferentemente, na validação cruzada os dados são aleatoriamente distribuídos em grupos (*folds*). Sendo que, o treinamento contempla todas as permutações de treinamento e teste incluindo todos os grupos. Desta forma, a utilização da validação cruzada elimina possíveis ganhos ou perdas nos resultados por causa da má distribuição dos dados de treino e teste. É importante ressaltar que independente do tipo de validação utilizada, os dados de teste nunca são utilizados no desenvolvimento dos modelos, apenas para validação, indicando o quanto o modelo é capaz de generalizar os resultados sobre novos dados.

6.7.4 Nível de Validação dos Detectores

Uma outra escolha a ser realizada é o nível da validação dos modelos. O nível da validação significa como os dados são selecionados para irem para os conjuntos de treino e de teste e pode ser de dois tipos: a nível do aluno e a nível da observação. A nível de observação significa que os dados são selecionados um a um e a nível do aluno os dados são selecionados com base nos dados de cada aluno. Assim, os modelos treinados e avaliados a nível do aluno conseguem generalizar os resultados de novos alunos com uma maior precisão. Essa estratégia também é conhecida como divisão em lotes de dados, do inglês *split on batch*. Sete dos oito trabalhos

relacionados utilizaram a validação a nível do aluno. Dessa forma, esse trabalho também utiliza a validação a nível de aluno. Para isso, os 30 alunos foram divididos em 10 grupos, sendo estes selecionados de forma aleatória. O número dos grupos igual a 10 foi definido de acordo com o valor k , que foi escolhido de acordo com os trabalhos relacionados mais recentes, conforme descrito na seção 6.7.3. Assim, cada grupo contém dados de 3 alunos diferentes, o que representa 10% do total dos dados.

6.7.5 Seleção das Características

Todos os trabalhos relacionados realizam uma seleção das características, do inglês *feature selection*. O objetivo desta seleção é descobrir quais as características (*features*) que melhor representam o conjunto de dados de acordo com o rótulo de treinamento, ou seja, o valor objetivo que os detectores visam prever. Embora os trabalhos relacionados descrevam que utilizam a técnica de seleção progressiva, do inglês *forward selection*, os autores não deixam claro como as realizaram. Assim, não se sabe ao certo como (parâmetros e métodos de validação) os trabalhos realizaram essa seleção de característica. A técnica de seleção progressiva começa com um conjunto vazio de características e tem o objetivo de adicionar as características progressivamente. Para cada característica adicionada, o modelo é testado usando alguma operação de validação. Apenas as características que apresentarem as melhores performances são mantidas no conjunto final. (SUTTER; KALIVAS, 1993). Assim, como nos trabalhos relacionados, este trabalho também implementa a seleção progressiva das características, utilizando a validação cruzada como método de avaliação das características selecionadas.

6.7.6 Métricas de Avaliação dos Detectores

De acordo com todos os trabalhos relacionados, a principal métrica de avaliação dos detectores é pelo cálculo de concordância Cohen's Kappa. Segundo Cohen (1960), Cohen's Kappa, ou somente Kappa, tem o objetivo de medir o quanto o detector é melhor do que a seleção ao acaso dos estados afetivos. O valor de Kappa igual a 0 indica que o detector alcança a mesma precisão da seleção por acaso. O valor de Kappa igual a 1 indica uma performance perfeita. A segunda métrica, utilizada por sete dos oito trabalhos relacionados, é a A' , calculada como a estatística de Wilcoxon. (HANLEY; MCNEIL, 1982). A' indica a probabilidade de que o detector tenha identificado corretamente a presença ou ausência de um estado afetivo, por exemplo estar confuso ou não estar confuso. Um detector com A' igual a 0,5 indica performance igual à probabilidade ao acaso e 1 indica performance perfeita. Assim, essas duas métricas foram utilizadas para avaliação dos detectores desenvolvidos. Para a aplicação das métricas, os resultados dos detectores são armazenados e comparados com os rótulos de estados afetivos gerados pelos anotadores, na sessão de anotação dos estados afetivos.

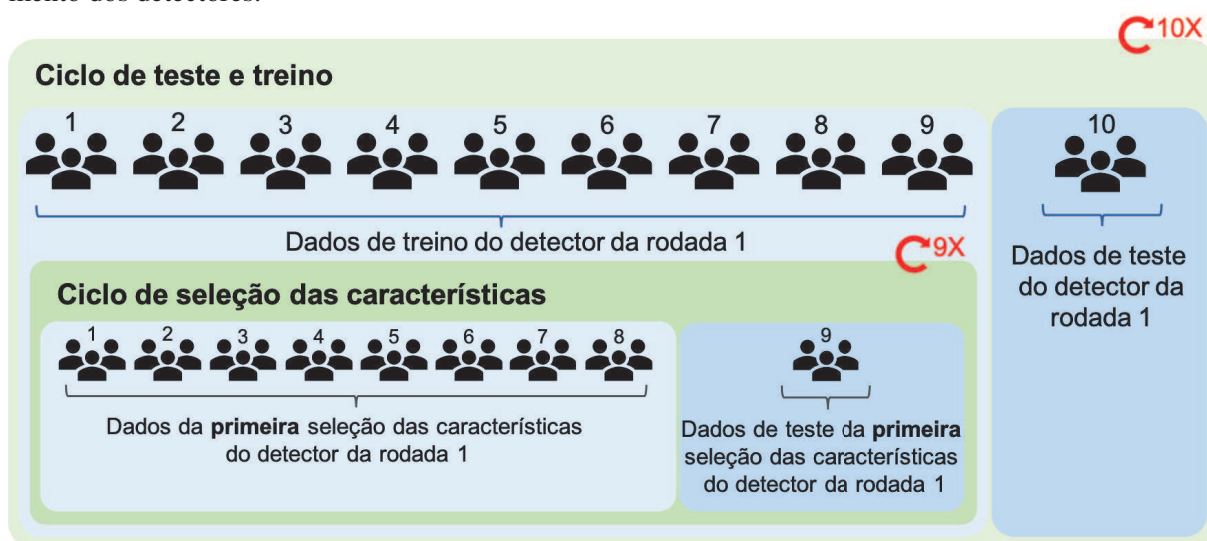
6.7.7 Treinamento dos Detectores

De acordo com as decisões descritas acima, todas baseadas nos trabalhos relacionados, foram utilizados cinco algoritmos para o desenvolvimento dos detectores de cada estado afetivo. Validação cruzada a nível de aluno, com $k = 10$, foi utilizado como método de avaliação dos modelos. Como métricas de avaliação, foram utilizados os valores de Kappa e A' . Além disso, seleção progressiva das características foi aplicado para seleção das características que melhor representam os estados afetivos.

Cada detector é treinado para identificar um único estado afetivo. Assim, os detectores foram treinados para identificar a presença ou ausência do seu respectivo estado afetivo. Desta forma, como característica alvo de cada detector, foi utilizada uma única característica binária, representando a ausência ou presença do estado afetivo em cada *log*. O objetivo é que o detector aprenda a generalizar os *logs* de novos alunos com base nos *logs* utilizados durante a fase de treinamento de cada detector.

A seleção progressiva das características também realiza uma validação para identificar quais as características que melhor representam o conjunto de dados que está sendo analisado. Essa análise é realizada com uma característica alvo, ou rótulo. No caso desse trabalho, os rótulos são a presença ou ausência do estado afetivo, conforme descrito anteriormente. Para essa validação da seleção progressiva, foi utilizada outra validação cruzada a nível de aluno. Assim, para cada conjunto de dados de treinamento da validação cruzada do detector, uma nova validação cruzada foi aplicada para validar a seleção progressiva das características. A Figura 12 tem como objetivo simplificar essa descrição com base em um exemplo.

Figura 12 – Exemplo de uma rodada dos ciclos de teste e treino e seleção das características no treinamento dos detectores.



Fonte: Elaborado pelo autor.

A Figura 12 exibe o exemplo de uma rodada do ciclo de teste e treino do detector e, somente

no conjunto de dados de treino, uma rodada do ciclo de seleção das características. Neste exemplo, os dados dos grupos de 1 a 9 foram selecionados para serem utilizados no treinamento do detector. Ao final desse treinamento, esse detector é submetido a um teste com dados nunca vistos antes, que neste exemplo são do grupo 10. Assim, os valores Kappa e A' são calculados nessa etapa de teste. O resultado indica o quanto esse detector é capaz de prever os estados afetivos de novos alunos. Porém, como essa é uma validação cruzada, esse exemplo ilustra apenas uma das 10 rodadas do ciclo de teste e treino da validação. Assim, um novo modelo é treinado e testado para cada permutação de 1 a 10, sempre separando os dados de nove grupos para treino e os dados do grupo restante para teste. Ao final das dez execuções, a precisão na detecção do estado afetivo de cada detector é calculada com base na média das dez rodadas.

Ainda conforme o exemplo da Figura 12, para cada rodada do ciclo de teste e treino, um ciclo de seleção das características é executado por completo, este contendo nove rodadas. Cada rodada do ciclo de teste e treino contém um conjunto de treino com os dados de nove grupos de alunos. Esse conjunto é dividido novamente em grupos de teste e treino, mas desta vez com o objetivo de validar a seleção das características. Dos nove grupos, oito são selecionados para seleção das características e um é reservado para testar esta seleção. No exemplo da Figura 12, os dados dos grupos de 1 a 8 foram selecionados como conjunto de treino e os dados do grupo 9 foram utilizados para o teste. Assim, novamente, uma nova seleção de características é realizada e testada para cada permutação de 1 a 9, sempre separando os dados de oito grupos para seleção e os dados do grupo restante para teste da seleção. Ao final das nove execuções, as características que apresentaram a melhor performance com base na média das nove rodadas são selecionadas.

Essa estratégia de validação cruzada para o desenvolvimento do detector e para a seleção de características, resultou em um total de 90 rodadas de treino e teste. Sendo estas 10 do ciclo de teste e treino e, para cada rodada do ciclo de teste e treino, mais nove rodadas do ciclo de seleção das características. Em média, o treinamento de cada detector, utilizando esta estratégia, levou mais de sete horas para ser processado. Além disso, para cada estado afetivo foram desenvolvidos cinco modelos, utilizando os cinco melhores algoritmos, resultando na construção de 20 modelos e de 140 horas de processamento. Para o processamento dos detectores, foi utilizado um MacBook Pro com processador Intel Core i5 de 2,6GHz com quatro núcleos de processamento e 8GB de memória RAM.

A estratégia apresentada acima foi utilizada na construção de todos os detectores, exceto para os detectores que utilizaram o algoritmo KStar. Não foi possível aplicar a mesma estratégia de seleção de características com este algoritmo devido ao fato de que o processamento deste algoritmo sem a seleção de característica levou o mesmo tempo dos outros algoritmos com seleção de característica, ou seja, mais de sete horas. Assim, se somente para 10 rodadas foram necessárias sete horas, para 90 rodadas estimou-se que seriam necessárias mais de 600 horas de processamento. Neste caso, optou-se por não aplicar a seleção de características no desenvolvimento dos detectores que utilizaram o algoritmo KStar, somente a validação cruzada

do ciclo de teste e treino.

Além disso, foram testados os valores 10 e 20 para o número de características a serem selecionadas na seleção progressiva de características. Porém, o melhor resultado foi encontrado quando 10 características eram selecionadas. Quando os algoritmos foram processados para selecionarem 20 características, o tempo de processamento foi quase o dobro e não houve melhora nos resultados. Assim, optou-se por permanecer com a seleção das 10 melhores características. Sendo este o limite máximo de características a serem selecionadas. Outra estratégia de parada é quando, mesmo adicionando novas características, não há mais melhora na performance, fazendo com que a seleção seja terminada.

O desenvolvimento de todos os detectores seguiu o mesmo padrão dentro da ferramenta RapidMiner. Primeiramente os dados foram carregados, sendo estes 5.525 *logs*, cada um com 348 características. Depois, para cada detector foi realizada uma operação de seleção de atributos, onde os estados afetivos que não fossem o rótulo alvo daquele detector eram removidos. Após, foi aplicado uma operação de definições de papéis das características. Nesta operação, a característica que representava o estado afetivo à ser detectado recebia o papel de *label* e a característica que definia em qual grupo de alunos cada *log* pertencia recebia o papel de *batch*. Após, o processo era o mesmo para todos os detectores, mudando apenas o algoritmo utilizado, conforme descrito anteriormente.

7 RESULTADOS E DISCUSSÕES

Este capítulo tem como objetivo descrever os resultados encontrados nos testes dos detectores de estados afetivos desenvolvidos e apresentar uma análise desses resultados. Ao todo, foram desenvolvidos 24 detectores. Destes, os 20 primeiros serviram para testar qual dos cinco algoritmos apresentava o melhor resultado na detecção de cada um dos quatro estados afetivos e os quatro últimos foram treinados utilizando os dados de personalidade dos alunos, com base nos melhores detectores encontrados para cada estado.

7.1 Resultados

A hipótese de pesquisa deste trabalho é que os dados de personalidade dos alunos podem influenciar de forma positiva nos resultados da detecção automática dos estados afetivos dos alunos. Desta forma, para validar esta hipótese, dois tipos de detectores de estados afetivos foram desenvolvidos, um detector que não leva em conta a personalidade e outro que considera os traços de personalidade dos alunos durante as suas fases de treinamento. Assim, esta seção tem como objetivo descrever os resultados encontrados nos dois detectores desenvolvidos para que a comparação desses resultados permita verificar se os dados de traço de personalidade melhoram a detecção das emoções.

7.1.1 Detectores Treinados Sem Considerar a Personalidade

Esta seção descreve somente os resultados dos detectores desenvolvidos sem considerar os dados de personalidade dos alunos. A Tabela 8 ilustra o resultado dos 20 detectores e seus respectivos algoritmos utilizados para cada estado afetivo. Foram destacados os melhores resultados, utilizando o valor de Kappa como métrica primária.

Tabela 8 – Resultados dos detectores com diferentes algoritmos na detecção de cada estado afetivo.

	J48		JRip		Logistic Regression		Naive Bayes		KStar	
	Kappa	A'	Kappa	A'	Kappa	A'	Kappa	A'	Kappa	A'
Confusão	0,250	0,624	0,288	0,636	0,307	0,728	0,342	0,690	0,083	0,552
Tédio	-0,002	0,470	0,025	0,542	0,000	0,452	0,040	0,608	0,005	0,560
Frustração	-0,004	0,535	0,042	0,533	-0,001	0,600	0,022	0,421	0,098	0,543
Engajamento	0,630	0,844	0,617	0,821	0,592	0,847	0,564	0,799	0,104	0,530

Fonte: Elaborado pelo autor.

Conforme apresentado na Tabela 8, o melhor algoritmo para a detecção de confusão foi o Naive Bayes, obtendo um índice Kappa = 0,342 e A' = 0,690. Na detecção do estado de tédio, o melhor algoritmo também foi o Naive Bayes, com um valor Kappa = 0,04 e A' = 0,608. Para a detecção do estado de frustração, o melhor algoritmo foi o KStar, mesmo sem a seleção de características, obtendo um valor Kappa = 0,098 e A' = 0,543. E, por fim, na detecção

do estado de engajamento concentrado o melhor algoritmo foi o J48, com um índice Kappa = 0,630 e A' = 0,844. Também foram calculadas as médias dos melhores resultados, sendo Kappa = 0,278 e A' = 0,671.

Todos os resultados de treinamento dos detectores foram armazenados para a realização de uma análise. Assim, é possível listar quais foram as características selecionadas para cada estado afetivo, de acordo com cada algoritmo. A seguir são listadas as 10 características dos melhores algoritmos para cada estado afetivo, com exceção do estado de frustração, que apresentou seu melhor resultado com o algoritmo KStar, o qual não teve seleção de características. A Tabela 9 descreve as características selecionadas para o estado de confusão, a Tabela 10 lista as características selecionadas para o estado de tédio e a Tabela 11 enumera as características selecionadas para o estado de engajamento concentrado.

Tabela 9 – Características selecionadas para o estado de confusão.

Número	Característica
1	Comportamento <i>off task</i> , indicando que o aluno não está trabalhando no ambiente de aprendizagem.
2	Comportamento <i>on task conversation</i> , indicando que o aluno está no ambiente de aprendizagem resolvendo as tarefas, mas pediu ajuda ao professor ou colega.
3	Número de erros na operação de adição nos últimos cinco segundos.
4	A equação contém a operação propriedade distributiva na sua resolução.
5	Número de acertos na operação de multiplicação nos últimos cinco segundos.
6	Número de erros na operação de multiplicação nos últimos cinco segundos.
7	Efetividade na operação de multiplicação nos últimos cinco segundos.
8	Efetividade na operação inversa diário (desde o início do dia atual até o <i>log</i> atual).
9	Número de passos errados seguidos na equação.
10	Quanto tempo que o aluno está no passo atual.

Fonte: Elaborado pelo autor.

Tabela 10 – Características selecionadas para o estado de tédio.

Número	Característica
1	Comportamento <i>on system</i> , indicando que o aluno está no sistema mas não está trabalhando na resolução das equações.
2	Média de acertos da turma na operação de multiplicação de frações diário.
3	A equação contém a operação mínimo múltiplo comum na sua resolução.
4	Média de erros da turma na operação mínimo múltiplo comum diário.
5	Média de erros da turma na operação de subtração diário.
6	Número de cliques seguidos.
7	Número de passos corretos na equação.
8	Número de passos corretos total (desde o primeiro dia de coleta até o <i>log</i> atual).
9	Número de passos total.
10	Tempo gasto no passo atual.

Fonte: Elaborado pelo autor.

Tabela 11 – Características selecionadas para o estado de engajamento concentrado.

Número	Característica
1	Comportamento <i>on task</i> .
2	Comportamento <i>on task out</i> .
3	A equação contém a operação inversa na sua resolução.
4	A equação é de categoria difícil 3.
5	Número de passos errados seguidos na equação.
6	O tipo do <i>log</i> é igual a parada do mouse.
7	Clique no grupo resolução.
8	Média da efetividade da turma na operação de mínimo múltiplo comum total.
9	Número de cliques seguidos.

Fonte: Elaborado pelo autor.

7.1.2 Detectores Treinados Considerando a Personalidade

Esta seção descreve os resultados obtidos nos detectores de estados afetivos que consideram os dados da personalidade dos alunos durante a fase de treinamento desses detectores. Após identificar qual o melhor algoritmo para cada estado afetivo, sem considerar os dados de personalidade, foi possível treinar novamente os detectores. Porém, desta vez as características de personalidade foram inseridas no treinamento destes detectores. Assim, os quatro detectores foram retreinados, mas sem alterar nada além da adição das características de personalidade. A Tabela 12 apresenta os resultados obtidos nos testes dos quatro últimos detectores e realiza uma comparação com os detectores que foram treinados sem os dados de personalidade.

Tabela 12 – Comparação dos detectores que consideram contra os detectores que não consideram características de personalidade.

		Kappa	A'
Confusão	Sem personalidade	0,342	0,690
	Com personalidade	0,309	0,688
Tédio	Sem personalidade	0,040	0,608
	Com personalidade	0,000	0,558
Frustração	Sem personalidade	0,098	0,543
	Com personalidade	0,094	0,537
Engajamento	Sem personalidade	0,630	0,844
	Com personalidade	0,633	0,846
Média	Sem personalidade	0,278	0,671
	Com personalidade	0,259	0,657

Fonte: Elaborado pelo autor.

Ao analisar os resultados obtidos nos testes dos detectores que consideram a personalidade do aluno, comparando com os detectores que não consideram a personalidade, conforme apresentado na Tabela 12, o único resultado em que houve uma melhora na precisão do detector foi sobre o estado de engajamento, destacado na cor verde na tabela. O detector que considera a

personalidade obteve os valores de $Kappa = 0,633$ e $A' = 0,846$, enquanto o detector que não considera a personalidade obteve os valores de $Kappa = 0,630$ e $A' = 0,844$.

Ao verificar as características dos detectores treinados com dados da personalidade do aluno, foi possível identificar que não houve mudanças na seleção de características dos estados de confusão e tédio. Porém, para o estado de engajamento concentrado, duas características foram substituídas em relação à seleção de características do detector que não considera a personalidade dos alunos, conforme apresentado na Tabela 13.

Tabela 13 – Características selecionadas para o estado de engajamento concentrado no detector que considera a personalidade do aluno.

Número	Característica
1	Comportamento <i>on task</i> (igual).
2	Comportamento <i>on task out</i> (igual).
3	A equação contém a operação inversa na sua resolução (igual).
4	A equação é de categoria difícil 3 (igual).
5	Número de passos errados seguidos na equação (igual).
6	O tipo do <i>log</i> é igual a parada do mouse (igual).
7	Média da efetividade da turma na operação de mínimo múltiplo comum total (igual).
8	Número de acertos na operação de subtração nos últimos cinco segundos (adicionado).
9	Identificação do lado do traço abertura, ou seja, 0 se o traço for menor que a metade ou 1 caso contrário (adicionado).
10	Clique no grupo resolução (removido).
11	Número de cliques seguidos (removido).

Fonte: Elaborado pelo autor.

As características “clique no grupo de resolução” (item 10) e “número de cliques seguidos” (item 11) foram substituídas pelas características “número de acertos na operação de subtração nos últimos cinco segundos” (item 8) e pela característica “identificação do lado do traço abertura” (item 9). É importante ressaltar que essas foram as características resultantes da seleção progressiva de características aplicadas somente na detecção do estado de engajamento e nenhuma alteração no desenvolvimento do modelo foi realizada, apenas a adição dos dados de personalidade.

7.2 Análise dos Resultados

Esta seção apresenta uma análise dos resultados obtidos no desenvolvimento dos dois tipos de detectores, os que consideram e os que não consideram a personalidade dos alunos. Além disso, os resultados obtidos foram comparados com os resultados apresentados pelos trabalhos relacionados. Desta forma, esta seção está subdividida de acordo com as análises realizadas.

7.2.1 Resultados dos Detectores Sem Personalidade

A Tabela 8 apresenta os resultados obtidos por cada um dos detectores treinados sem dados de personalidades, sendo estes provenientes de cinco algoritmos diferentes e quatro estados afetivos. Além disso, a Tabela 8 também destaca, na cor verde, os algoritmos que obtiveram o melhor desempenho na detecção de cada um dos estados afetivos.

7.2.1.1 Estado de Confusão

Na detecção do estado de confusão, o melhor resultado apresentado foi pelo algoritmo Naive Bayes, com Kappa de 0,342 e A' de 0,690. Embora não exista um limiar mínimo para considerar um modelo de detecção de estados afetivos ideal, considera-se como comparação o mesmo valor que é exigido por anotadores humanos dos estados afetivos. Assim, o resultado obtido no estado de confusão pode ser considerado intermediário, quando comparado ao nível mínimo exigido para a concordância de codificadores humanos, que é um Kappa maior ou igual à 0,6. Porém, ele indica que o detector desenvolvido foi capaz de prever o estado de confusão com uma precisão maior do que um algoritmo que leva em consideração apenas a probabilidade de ocorrência do estado de confusão.

Ainda sobre a detecção do estado de confusão, a Tabela 9 apresenta as características consideradas como mais relevantes para a predição do estado de confusão. A escolha destas dez características ocorreu de forma automática, por meio da execução do algoritmo de seleção progressiva de características, especificamente para o estado de confusão e para o algoritmo Naive Bayes, que obteve o melhor resultado na detecção da confusão. Ao analisar as características selecionadas, é possível identificar que o comportamento do aluno, principalmente os comportamentos *off task* e *on task conversation*, impactam na detecção do estado de confusão. Além disso, as outras oito características que mais representam o estado de confusão estão relacionadas ao número de acertos, erros e efetividade dos passos dos alunos. É possível notar que quatro destas características representam os dados gerados pelo aluno nos últimos cinco segundos, apontando que a confusão tem uma relação com *logs* mais recentes.

7.2.1.2 Estado de Tédio

Na detecção do estado de tédio, o melhor algoritmo também foi o Naive Bayes, com Kappa de 0,040 e A' de 0,608. Embora esse tenha sido o melhor resultado entre os cinco algoritmos utilizados para a detecção do estado de tédio, esse resultado ainda é muito baixo comparado com o mínimo exigido para a concordância de codificadores humanos, Kappa maior ou igual à 0,6. Esse resultado, $Kappa = 0,040$, indica que a detecção do estado de tédio é um pouco melhor do que um algoritmo probabilístico, que escolhe o estado de acordo com o probabilidade de ocorrência que ele ocorre.

Ao analisar a Tabela 10, que apresenta as 10 características que melhor representam o estado de tédio, é possível identificar que o comportamento *on system* está relacionado com o tédio, assim como o número de cliques seguidos que o aluno realizou, número de passos corretos, número de passos total e o tempo gasto no passo atual. Novamente, essas características foram selecionadas automaticamente pelo algoritmo de seleção progressiva de características. Além disso, foi possível identificar que informações mais antigas, calculadas com dados obtidos desde o início do dia (diário) e desde o início das sessões de coleta (total), foram as mais aparentes (cinco ocorrências). A escolha dessas características aponta que informações mais históricas do aluno pode refletir melhor na detecção do estado de tédio.

7.2.1.3 Estado de Frustração

Para a detecção do estado de frustração, o algoritmo que obteve o melhor desempenho foi o KStar, com um índice Kappa de 0,098 e A' de 0,543. Novamente, este resultado ainda é muito baixo, quando comparado ao mínimo exigido por anotadores humanos. Porém, esse resultado é um pouco melhor do que a detecção realizada por algoritmos baseado em probabilidade. Conforme descrito na seção 6.7.7, treinamento dos detectores, não foi possível aplicar o algoritmo de seleção progressiva de características para o algoritmo KStar. Desta forma, foram consideradas as 348 características nas fases de treinamento do algoritmo KStar.

7.2.1.4 Estado de Engajamento Concentrado

Por fim, na detecção do estado de engajamento concentrado, o algoritmo que apresentou o melhor resultado foi o J48, com Kappa de 0,630 e A' de 0,844. Esse resultado é considerado acima do mínimo exigido por codificadores humanos. Desta maneira, o detector do estado de engajamento concentrado pode ser considerado apto à ser testado em ambientes na identificação de engajamento em tempo real. Este foi o melhor resultado obtido na detecção dos estados afetivos, quando comparado aos melhores resultados obtidos na detecção dos demais estados.

A Tabela 11 descreve as características que melhor representam o estado de engajamento concentrado, de acordo com o resultado do algoritmo de seleção progressiva das características. Novamente, o comportamento do aluno está presente nas características que melhor representam o estado afetivo, mas, desta vez, os comportamentos *on task* e *on task out* são os que melhor representam o engajamento. Além disso, a maioria das demais características estão relacionadas à equação atual que o aluno está resolvendo ou à interação do aluno com a interface, como posição do clique, número de cliques ou movimento do mouse. Essas características apontam que os dados de *logs* podem ser utilizados de forma efetiva para a detecção do estado de engajamento concentrado.

7.2.2 Resultados dos Detectores Com Personalidade

Ao analisar os resultados apresentados na Tabela 12, é possível verificar que não houve uma melhora na detecção dos estados de confusão, tédio e frustração. Porém, no estado de engajamento houve uma melhora na detecção. Este trabalho tinha como objetivo verificar se os dados da personalidade do aluno poderiam resultar em uma melhora na detecção dos estados afetivos destes alunos. Embora a diferença entre os dois detectores, com e sem dados de personalidade, seja pequena, ela traz indícios de que a hipótese de pesquisa deste trabalho é verdadeira, conforme apresentado nos resultados destacados em verde na Tabela 12. A diferença de 0,003 no valor Kappa e 0,002 no valor de A' entre o detector treinado com e sem os dados de personalidade indica que o detector de engajamento com dados da personalidade foi 0,3% melhor que o mesmo detector treinado sem os dados de personalidade. Acredita-se que, embora essa diferença seja pequena, esse é o primeiro trabalho que utiliza os dados de personalidade do aluno no desenvolvimento dos detectores de estados afetivos. Mesmo assim, já é possível identificar que a personalidade pode trazer melhores resultados na detecção.

Ainda sobre a diferença entre os detectores de engajamento que foram treinados com e sem os dados de personalidade, foi possível identificar uma diferença nas características selecionadas para o desenvolvimento de ambos os detectores. O detector que considera a personalidade, dentre as 348 características possíveis, considerou a característica relacionada ao traço de personalidade abertura como sendo uma das mais importantes, juntamente com as outras oito selecionadas. Conforme descrito na seção 6.7.6, as características foram selecionadas e validadas utilizando a validação cruzada, que oferece mais segurança para os resultados encontrados. Acredita-se que a seleção desta característica de personalidade é um forte indício de que a personalidade pode influenciar os estados afetivos e suas transições.

Referente aos demais estados afetivos, não houve uma mudança no resultado de detecção pois nenhuma característica relacionada à personalidade dos alunos foi selecionada. Ao verificar as características dos detectores treinados com dados da personalidade do aluno, foi possível identificar que não houve mudanças na seleção de características dos estados de confusão e tédio, ou seja, as mesmas características foram selecionadas pelo detector com e sem dados da personalidade. O algoritmo de seleção progressiva de características selecionou apenas as dez características mais representativas de cada estado, de acordo com cada algoritmo, que neste caso não incluíram a personalidade. Ainda, como não houve seleção de características no desenvolvimento do detector de frustração, devido ao tempo de processamento do algoritmo KStar, não se sabe quais características foram utilizadas no treinamento do modelo.

Foram levantadas algumas hipóteses sobre o porquê das características da personalidade não terem sido selecionadas na identificação dos estados de tédio e confusão. A primeira delas é relacionado ao limitante de escolher no máximo dez características para representar o estado afetivo. Um teste foi realizado para verificar se 20 características iriam representar melhor cada um dos estados, mas com base nesse teste o tempo de treinamento do detector quase dobrou e

não houve uma melhora no resultado. Assim, acredita-se que é necessário realizar um estudo correlacional mais elaborado, visando identificar com maior precisão qual o conjunto de características que melhor representa cada estado afetivo. Ainda sobre o algoritmo de seleção de características, a segunda hipótese é que o algoritmo de seleção progressiva (*forward selection*) talvez não seja o mais apropriado para este tipo de problema, pois deve lidar com uma grande quantidade de características. Assim, é necessário realizar novos testes visando identificar algoritmos de seleção de características que possam levar em consideração esse grande número de características, 348, que foram calculadas. Por fim, a terceira hipótese é de que haviam poucas amostras (*logs* rotulados) para o treinamento dos detectores destes estados afetivos, principalmente para os estados de tédio e frustração, quando comparado com a quantidade de amostras do estado de engajamento.

Ao analisar o número de rótulos gerados na fase de anotações de estados afetivos, apresentados na Tabela 7 (seção 6.5.8), é possível identificar uma forte relação entre os resultados obtidos na detecção de cada estado afetivo (apresentados na Tabela 12), Conforme apresentado na Tabela 14.

Tabela 14 – Comparação da quantidade de rótulos com os resultados obtidos nos detectores.

	Rótulo coletado		Resultado detector	
	Quantidade	Representatividade dos Rótulos	Kappa	A'
Engajamento	1118	53,3%	0,633	0,846
Confusão	463	22,1%	0,342	0,690
Tédio	90	4,3%	0,040	0,608
Frustração	51	2,4%	0,098	0,543

Fonte: Elaborado pelo autor.

O propósito da Tabela 14 é apresentar a relação entre o número de rótulos coletados para cada estado afetivo contra o resultado obtido por cada um de seus respectivos detectores. É possível identificar que quanto mais rótulos foram utilizados no treinamento de cada detector, melhor os resultados obtidos na detecção do estado afetivo. Os trabalhos relacionados utilizam uma estratégia de duplicação dos dados que possuem uma menor quantidade, do inglês *resampling*. Porém, essa estratégia pode fazer com que os detectores decorem certos padrões, reduzindo a capacidade de generalização de novos dados. Os autores dos trabalhos que utilizaram esta estratégia relatam que a análise dos resultados é feita em cima do conjunto real de dados, sem a re-amostragem. Porém, neste trabalho foi escolhido não aplicar essa técnica, pois o número de rótulos dos estados de tédio e frustração eram muito pequenos comparados aos de engajamento e confusão. Assim, um grande número de duplicações deveria ser realizado, o que possivelmente poderia resultar em *overtraining*¹ dos algoritmos. Acredita-se que a pequena quantidade de rótulos dos estados de confusão e, principalmente, tédio e frustração, pode ter afetado na escolha das características. Assim, ainda são necessárias novas sessões de coleta de

¹Comportamento apresentado quando os algoritmos de aprendizagem de máquina são treinados demais ou com muitos dados, principalmente replicados, ocasionando um baixo nível de generalização dos resultados quando aplicados a novos dados.

dados para obtenção de um maior número desses rótulos e, por sua vez, o treinamento de novos detectores.

7.2.3 Resultados Obtidos Comparados Com os Trabalhos Relacionados

Além da análise interna dos resultados, comparando os melhores detectores que não consideram a personalidade com os melhores que consideram a personalidade, descritos anteriormente, também foi realizada uma análise dos melhores resultados obtidos com os resultados apresentados pelos trabalhos relacionados. Assim, a Tabela 15 apresenta os dados da Tabela 4, apresentada no capítulo 5 (trabalhos relacionados), juntamente com os resultados obtidos pelos melhores detectores de cada estado afetivo. Nesta tabela, K foi utilizado como abreviação de Kappa e TD como abreviação de Trabalho Desenvolvido.

Tabela 15 – Comparação dos resultados apresentados pelos trabalhos relacionados com o trabalho desenvolvido.

		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	TD
Frustração	K	0,390	0,230	0,324	0,150	0,157	0,389	0,105	0,056	0,098
	A'	-	0,990	0,682	0,600	0,602	0,726	0,692	0,634	0,543
Tédio	K	0,207	0,280	0,229	0,240	0,264	0,332	0,469	0,278	0,040
	A'	-	0,690	0,632	0,660	0,671	0,723	0,848	0,682	0,608
Confusão	K	0,182	0,400	0,274	0,150	0,148	0,334	0,056	0,091	0,342
	A'	-	0,990	0,736	0,630	0,627	0,735	0,552	0,568	0,690
Flow/ Engajamento	K	0,193	0,310	0,358	0,420	0,423	0,360	0,156	0,142	0,633
	A'	-	0,710	0,678	0,730	0,743	0,697	0,590	0,624	0,846
Média	K	0,163	0,305	0,296	0,240	0,248	0,354	0,196	0,141	0,278
	A'	-	0,845	0,682	0,650	0,660	0,720	0,670	0,627	0,672

Fonte: Elaborado pelo autor.

Conforme os resultados apresentados na Tabela 15, é possível identificar que os detectores de confusão e engajamento alcançaram resultados positivos, quando comparados ao estado da arte. O detector de confusão, destacado na cor azul, apresentou um resultado mais similar aos resultados apresentados nos trabalhos relacionados. Por outro lado, o detector de engajamento obteve o melhor desempenho, comparando com os trabalhos relacionados. É a primeira vez que um detector atinge um valor Kappa maior que o mínimo exigido por codificadores humanos, para a geração de rótulos de estados afetivos, como ocorre no protocolo BROMP. Esse resultado indica que os rótulos gerados pelo detector do estado de engajamento que considera a personalidade do aluno podem ser considerados semelhantes a de um codificador humano. Acredita-se que este resultado na detecção do estado de engajamento concentrado seja em (i) decorrência do maior número de rótulos, em relação aos rótulos dos outros estados afetivos, (ii) à quantidade de características utilizadas, levando em consideração todas as informações de interação do aluno com a interface, modelo do aluno, comportamentos e personalidade do aluno, (iii) à forma de obtenção dos rótulos, na qual somente um aluno era analisado por vez e os estados

afetivos codificados representavam de forma mais fiel o estado do aluno e (iv) a quantidade de *logs* gerados por cada aluno, nos quais eram possível identificar todas as interações do aluno com o sistema.

Diferentemente dos resultados dos detectores dos estados de confusão e engajamento, os resultados dos detectores de frustração e tédio não foram tão bons, quando comparados com o estado da arte. O detector do estado de tédio obteve um resultado muito abaixo do que os trabalhos relacionados vêm apresentando. O detector do estado de frustração também apresentou um resultado baixo, embora outros trabalhos também tenham apresentados resultados muito abaixo do mínimo exigido. O menor índice Kappa na detecção da frustração foi obtido pelo trabalho mais atual, indicando que a detecção do estado de frustração ainda é uma tarefa complexa. Acredita-se que o baixo desempenho destes detectores deve-se principalmente à quantidade de rótulos associados a cada um deles. Com tão poucos dados de treinamento, é difícil que os detectores sejam capazes de generalizar dados novos, de acordo com os testes aplicados. Além disso, não foi aplicada nenhuma técnica de duplicação dos dados, a qual poderia superestimar os resultados.

A média dos resultados dos quatro detectores foi calculada, resultando em um valor Kappa = 0,278 e $A' = 0,672$. Essa média não foi calculada com o intuito de fazer uma comparação com os trabalhos relacionados, uma vez que estes divergem em áreas de aplicação e em tipos de ambientes. Porém, ao calcular essa média e analisar com a média apresentada pelos trabalhos relacionados, é possível identificar que mesmo a melhor média ainda é muito baixa quando comparada ao índice Kappa mínimo exigido por codificadores humanos. Um fato interessante de se destacar é que a média dos dois últimos trabalhos relacionados, os mais recentes, foram menores que a média dos trabalhos anteriores. Embora várias pesquisas e estratégias para melhorar os detectores automáticos de estados afetivos vêm sendo publicadas, o nível de detecção dos detectores automáticos ainda é muito inferior aos resultados alcançados por detectores humanos, indicando que ainda há muito a ser aperfeiçoado para tornar esses detectores viáveis de serem utilizados para a detecção de estados afetivos em ambientes de aprendizagem em tempo real.

8 CONCLUSÕES

As emoções, um tipo de estado afetivo, podem influenciar no processo de aprendizagem dos alunos. (AZEVEDO; ALEVEN, 2013). Emoções como confusão podem afetar a aprendizagem de forma positiva, quando se tornam engajamento (FREDRICKSON, 1998), ou de forma negativa, quando se tornam frustração ou tédio. (GRAESSER; D’MELLO, 2011). Assim, é importante identificar os estados afetivos dos alunos para que se possa adotar estratégias pedagógicas, evitando que os alunos entrem em estados que afetem a aprendizagem negativamente.

Os sistemas tutores inteligentes são ambientes de aprendizagem capazes de guiar os alunos passo a passo na resolução de tarefas, em uma determinada área de conhecimento. PAT2Math é um STI voltado para o ensino de álgebra elementar, equações de primeiro grau, no qual o aluno pode inserir passos de resolução de uma equação inicial e, para cada passo inserido pelo aluno, o sistema fornece um *feedback* notificando o acerto ou erro, e no último caso acompanhado de um *feedback* de erro específico para o passo fornecido.

O principal objetivo dos STIs é melhorar o processo de aprendizagem dos alunos, fornecendo um auxílio individualizado e adaptado, de acordo com suas necessidades. Assim, sabendo que os estados afetivos podem influenciar de forma positiva ou negativa na aprendizagem, é importante que os STIs também levem em consideração a informação emocional sobre o estudante. Deste modo, estes ambientes poderão adaptar suas estratégias pedagógicas de acordo com o estados afetivos do aluno e saberão qual o melhor momento para realizar uma intervenção, proporcionando uma melhor aprendizagem. Para isso, é necessário que esses ambientes sejam capazes de detectar e prever os estados afetivos dos alunos.

O STI PAT2Math, assim como outros STIs, armazena todas as interações do aluno com o ambiente no módulo do aluno. Além de interações, também são armazenados estimativas de conhecimento, tarefas resolvidas, erros cometidos e dicas solicitadas ao sistema. Assim, um grande conjunto de dados individual de cada aluno é armazenado no banco de dados do sistema. Mais recentemente, a comunidade de mineração de dados educacionais tem usado dados provenientes da interação dos alunos com ambientes de aprendizagem, como os STIs, para a descoberta de novos conhecimentos, bem como a detecção e predição de determinados comportamentos ou estados afetivos dos alunos. Porém, nenhum trabalho conseguiu atingir um nível de precisão adequado para a utilização desses detectores em tempo real. O nível de detecção que os trabalhos relacionados tem reportado ainda é baixo, comparado com o mínimo exigido por codificadores humanos (Kappa maior ou igual à 0,6). Desta forma, esses detectores poderiam inferir os estados afetivos de forma incorreta, fazendo com que o sistema perca a credibilidade e o aluno pare de utiliza-lo, prejudicando sua aprendizagem.

Este trabalho teve como objetivo verificar se a utilização de dados de personalidade no treinamento de detectores de estados afetivos pode proporcionar uma melhora na precisão da detecção dos estados de frustração, confusão, tédio e engajamento concentrado. A escolha desta abordagem deu-se devido ao fato de que a personalidade pode direcionar as emoções (NUNES;

CAZELLA, 2011) e a duração das emoções parecem depender do traço de personalidade dos alunos. (REIS et al., 2018). Atualmente, já se sabe que trabalhos têm relatado um alto nível de concordância com os rótulos gerados pelos detectores com os rótulos gerados por codificadores humanos.

Outros trabalhos já vêm apresentando diferentes técnicas e métodos para a detecção desses quatro estados afetivos, identificados como mais frequentes nestes tipos de ambientes. Oito trabalhos foram classificados como relacionados ao trabalho proposto aqui. (D’MELLO et al., 2008; BAKER et al., 2012; PARDOS et al., 2014; OCUMPAUGH et al., 2014; WANG; HEFFERNAN; HEFFERNAN, 2015; PAQUETTE et al., 2014, 2016; JIANG et al., 2018). Porém, nenhum deles utilizou dados da personalidade do aluno para o treinamento dos detectores. Assim, as estratégias adotadas para o desenvolvimento destes detectores foram de acordo com o que os trabalhos relacionados vêm adotando.

Além dos dados de personalidade dos alunos, este trabalho também utilizou outro método para a obtenção dos rótulos de estados afetivos. Atualmente, a maioria dos trabalhos relacionados utilizam do protocolo BROMP para a coleta dos estados afetivos durante a utilização de ambientes de aprendizagem pelos alunos em sala de aula. Porém, este protocolo não é capaz de capturar as transições de estados afetivos de um único aluno, além de considerar apenas um estado afetivo a cada 20 segundos. Assim, um novo protocolo para a coleta dos rótulos de estados afetivos, com base em fases de treinamento e teste dos codificadores, foi proposto neste trabalho.

Este protocolo foi proposto para a coleta de rótulos de estados afetivos, com base em análise de vídeo com informações do rosto do aluno com áudio e tela do computador, para serem utilizados como rótulos de treinamento em algoritmos de aprendizagem de máquina. Este protocolo possui um conjunto de definições, fases de coleta dos vídeos e desenvolvimento da ferramenta de anotações, fase de treinamento dos codificadores, fase de testes dos codificadores e a fase de anotações, onde os codificadores podem gerar os rótulos de estados afetivos individualmente, depois de terem sido treinados e obterem um nível de concordância medidos com Randolph Kappa maior do que 0,6. Um artigo completo sobre este protocolo foi publicado na European Conference On Technology Enhanced Learning (EC-TEL) 2019, descrevendo as definições, fases, fluxo de execução e uma avaliação com diferentes anotadores utilizando uma versão aprimorada deste protocolo. (MORAIS et al., 2019).

Uma coleta dos dados de *logs*, módulo do aluno, personalidade e estados afetivos foi realizada em uma escola privada do RS. Nesse estudo, 55 alunos do sétimo ano utilizaram o STI PAT2Math em 10 sessões, contendo em média 40 minutos de duração. Deste estudo, foram utilizados para o desenvolvimento dos detectores 5.525 *logs*, cada um contendo 348 características. Além disso, 30 vídeos foram analisados, gerando um total de 2099 rótulos de estados afetivos e 2059 rótulos de comportamentos.

Os cinco melhores algoritmos apresentados nos trabalhos relacionados foram utilizados para o desenvolvimento do detector de cada estado afetivo. Assim, ao todo foram desenvolvidos 20

detectores. Porém, apenas o melhor detector de cada estado afetivo foi selecionado para ser retreinado com os dados de personalidade, permitindo a realização de um comparativo dos detectores que consideram e dos que não consideram a personalidade dos alunos. No desenvolvimento dos detectores foram utilizadas as técnicas de seleção de características e validação cruzada a nível de aluno, sendo $k = 10$, sendo estes 10 grupos contendo dados de três alunos selecionados aleatoriamente. As métricas de avaliação dos detectores foram as mesmas utilizadas pelos trabalhos relacionados, sendo estas Cohen's Kappa e A' .

Ao comparar os resultados dos detectores treinados com e sem os dados de personalidade dos alunos, apenas o detector do estado de engajamento apresentou uma melhora na precisão da detecção. Embora a melhora na detecção tenha sido muito pequena, com uma diferença de $Kappa = 0,003$ e $A' = 0,002$, foi possível identificar que uma característica relacionada à personalidade dos alunos foi selecionada como uma das mais representativas do estado de engajamento. Essa seleção de características ocorreu de modo automático, por meio da aplicação do algoritmo de seleção progressiva de características. Esse algoritmo selecionou, das 348 características disponíveis, somente nove características que melhor representaram o estado de engajamento concentrado, sendo uma delas relacionada à personalidade do aluno. Assim, este é um indício de que a hipótese de pesquisa deste trabalho é verdadeira.

Quando comparados os melhores resultados obtidos neste trabalho com os resultados apresentados pelos trabalhos relacionados, o detector de engajamento superou todos, apresentando um $Kappa = 0,633$ e $A' = 0,846$, que é maior que o nível de concordância exigido de codificadores humanos. Acredita-se que esse resultado é devido a utilização de rótulos de estados afetivos mais representativos dos estados dos alunos, provenientes do novo protocolo de anotação de estados afetivos e comportamentos proposto nesta dissertação. Assim, este trabalho apresenta o primeiro detector que é capaz de atingir um nível de concordância tão alto quanto o de um codificador humano na detecção do estado de engajamento concentrado. Quanto aos resultados dos outros detectores, o detector de confusão obteve um resultado intermediário, assim como os resultados apresentados pelos outros trabalhos. Porém, os detectores de frustração e tédio obtiveram resultados inferiores da maioria dos trabalhos relacionados, fazendo com que essa seja uma limitação e uma proposta de trabalho futuro desta dissertação.

Conforme apresentado na Tabela 15, que apresenta uma comparação dos resultados deste trabalho com os resultados apresentados nos trabalhos relacionados, ainda não se conhece um método padrão para o desenvolvimento de cada um dos detectores, nem para a coleta de dados, seja dos *logs* ou dos rótulos de treinamento. A área de EDM vem crescendo cada vez mais, mas ainda é uma área recente e que tem muito a ser descoberta. Desta forma, pesquisas exploratórias como essa podem apresentar direcionamentos e sugerir novas diretrizes, como o protocolo de anotação dos estados afetivos e uma possível fonte de dados que pode para melhorar a detecção destes estados com base em mineração de dados, descritos neste trabalho.

8.1 Ameaças à Validade dos Resultados

É importante salientar que os resultados obtidos nesta dissertação são provenientes da análise dos dados obtidos de apenas uma escola particular, situada na região metropolitana do Rio Grande do Sul, região sul do Brasil. Embora os dados tenham sido gerados por 55 alunos de duas turmas, durante 10 sessões que ocorreram em dias diferentes e que a seleção dos 30 vídeos analisados tenha sido realizada de forma aleatória, a amostra da sessão de coleta ainda é pequena. Por mais que os resultados obtidos aqui indiquem que os dados de personalidade dos alunos podem proporcionar uma melhora na detecção automática dos estados afetivos, ainda não é possível generalizar esses resultados para toda a população. Além disso, os dados de *logs* foram obtidos de apenas um ambiente de aprendizagem na área da matemática, o STI PAT2Math. Assim, também não é possível generalizar esses resultados para ambientes de outras áreas, uma vez que muitas das características utilizadas no desenvolvimento dos detectores finais estejam relacionadas com as operações que os alunos aplicaram para resolver as equações.

Outra ameaça à validade dos resultados está relacionada com o número de rótulos de estados afetivos. Da mesma forma que a pequena quantidade de rótulos dos estados de frustração e tédio tenha impactado negativamente na detecção destes estados, pode ser que esta pequena quantidade também tenha influenciado positivamente no resultado da detecção do estado de engajamento concentrado. Desta forma, os algoritmos de aprendizagem de máquina talvez tenham considerado essa quantidade de rótulos como um tipo de bias durante o treinamento dos detectores. Ainda sobre os detectores, embora a estratégia de validação cruzada tenha sido utilizada como método de validação, proporcionando resultados mais fidedignos na detecção dos estados afetivos, ao utilizar essa estratégia de validação juntamente com poucos dados de treinamento (5525 *logs*), os detectores podem ter decorado alguns resultados, do inglês *overtraining*, tornando-os muito dependentes de características específicas do STI PAT2Math, reduzindo assim a generalização dos resultados para novos ambientes de aprendizagem.

8.2 Trabalhos Futuros

Como trabalhos futuros, espera-se estudar maneiras de como melhorar a detecção dos estados de frustração e tédio. Uma das estratégias será retrainar estes detectores com uma maior quantidade de rótulos. Para isso, será necessário realizar a análise de mais vídeos para a geração de mais rótulos de estados afetivos. Além disso, conforme apontado na seção de ameaças à validade dos resultados, pretende-se realizar sessões de coleta para a obtenção dos dados em outras escolas. Outro trabalho futuro será na detecção dos comportamentos dos alunos. Estas informações dos comportamentos foram utilizadas no desenvolvimento dos detectores de estados afetivos, porém foram obtidas pelo protocolo de detecção dos estados afetivos e comportamentos descrito neste trabalho. Assim, o objetivo será desenvolver detectores para a detecção

automática desses comportamentos. Ainda, outro trabalho futuro é em relação a detecção automática da personalidade do aluno. Novamente, esses dados foram utilizados para o treinamento dos detectores apresentados nessa dissertação. Porém, o método de coleta desses dados foi por meio de um questionário que identificava os cinco traços de personalidade de cada aluno. Assim, o objetivo será verificar se é possível detectar e prever a personalidade dos alunos de forma automática.

REFERÊNCIAS

- ALLPORT, F. H.; ALLPORT, G. W. Personality traits: their classification and measurement. **The Journal of Abnormal Psychology and Social Psychology**, [S.l.], v. 16, n. 1, p. 6, 1921.
- AMERSHI, S.; CONATI, C. Combining unsupervised and supervised classification to build user models for exploratory. **JEDM| Journal of Educational Data Mining**, [S.l.], v. 1, n. 1, p. 18–71, 2009.
- ANDERSON, J. R. The expert module. **Foundations of intelligent tutoring systems**, [S.l.], p. 21–53, 1988.
- ANDRADE, J. M. d. **Evidências de validade do inventário dos cinco grandes fatores de personalidade para o brasil**. 2008. Tese (Doutorado em Ciência da Computação) — Tese (Doutorado em Psicologia Social e do Trabalho)—Universidade de Brasília, Distrito Federal, 2008.
- ARNOLD, M. B. **Emotion and personality**. [S.l.]: Columbia University Press, 1960.
- ARROYO, I. et al. Emotion sensors go to school. In: AIED, 2009. **Anais...** [S.l.: s.n.], 2009. v. 200, p. 17–24.
- ARROYO, I.; et al. The impact of animated pedagogical agents on girls’ and boys’ emotions, attitudes, behaviors and learning. In: ADVANCED LEARNING TECHNOLOGIES (ICALT), 2011 11TH IEEE INTERNATIONAL CONFERENCE ON, 2011. **Anais...** [S.l.: s.n.], 2011. p. 506–510.
- ARROYO, I. et al. A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. **International Journal of Artificial Intelligence in Education**, [S.l.], v. 24, n. 4, p. 387–426, 2014.
- ARROYO, I.; WOOLF, B. P. Inferring learning and attitudes from a bayesian network of log file data. In: AIED, 2005. **Anais...** [S.l.: s.n.], 2005. p. 33–40.
- AZEVEDO, O.; MORAIS, F. de; JAQUES, P. A. Exploring gamification to prevent gaming the system and help refusal in tutoring systems. In: EUROPEAN CONFERENCE ON TECHNOLOGY ENHANCED LEARNING, 2018. **Anais...** [S.l.: s.n.], 2018. p. 231–244.
- AZEVEDO, R.; ALEVEN, V. **International handbook of metacognition and learning technologies**. [S.l.]: Springer, 2013. v. 26.
- BAKER, R. et al. Why students engage in “gaming the system” behavior in interactive learning environments. **Journal of Interactive Learning Research**, Chesapeake, VA, v. 19, n. 2, p. 185–224, April 2008.
- BAKER, R. et al. Data mining for education. **International encyclopedia of education**, [S.l.], v. 7, n. 3, p. 112–118, 2010.
- BAKER, R. S. et al. Educational software features that encourage and discourage “gaming the system”. In: OF THE 14TH INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE IN EDUCATION, 2009. **Proceedings...** [S.l.: s.n.], 2009. p. 475–482.

- BAKER, R. S. et al. Better to be frustrated than bored: the incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. **International Journal of Human-Computer Studies**, [S.l.], v. 68, n. 4, p. 223–241, 2010.
- BAKER, R. S. et al. The dynamics between student affect and behavior occurring outside of educational software. In: INTERNATIONAL CONFERENCE ON AFFECTIVE COMPUTING AND INTELLIGENT INTERACTION, 2011. **Anais...** [S.l.: s.n.], 2011. p. 14–24.
- BAKER, R. S. et al. Towards sensor-free affect detection in cognitive tutor algebra. **International Educational Data Mining Society**, [S.l.], 2012.
- BAKER, R. S.; GOWDA, S. M. An analysis of the differences in the frequency of students' disengagement in urban, rural, and suburban high schools. In: EDUCATIONAL DATA MINING 2010, 2010. **Anais...** [S.l.: s.n.], 2010.
- BAKER, R. S.; INVENTADO, P. S. Educational data mining and learning analytics. In: **Learning analytics**. New York, NY: Springer New York, 2014. v. 18, n. 1, p. 61–75.
- BAKER, R. S.; YACEF, K. The state of educational data mining in 2009: a review and future visions. **JEDMI Journal of Educational Data Mining**, [S.l.], v. 1, n. 1, p. 3–17, 2009.
- BANDURA, A.; RAMACHAUDRAN, V. S. Self-efficacy. **Encyclopedia of human behavior**, [S.l.], 1994.
- BARBOSA, A. d. A. G. **Modelo hierárquico de fobias infanto-juvenis: testagem e relação com os estilos maternos**. 2009. Tese (Doutorado em Ciência da Computação) — Tese de doutorado apresentada ao Programa de Pós-Graduação Programa Programa Integrado de Doutorado em Psicologia Social (SFPB/UFRN), 2009.
- BARNES, T. The q-matrix method: mining student response data for knowledge. In: AMERICAN ASSOCIATION FOR ARTIFICIAL INTELLIGENCE 2005 EDUCATIONAL DATA MINING WORKSHOP, 2005. **Anais...** [S.l.: s.n.], 2005. p. 39–46.
- BATES, J. The role of emotion in believable agents. **Communications of the ACM**, New York, v. 37, n. 7, p. 122–125, 1994.
- BEAL, C. R.; QU, L.; LEE, H. Classifying learner engagement through integration of multiple data sources. In: NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE, 2006. **Proceedings...** [S.l.: s.n.], 2006. v. 21, n. 1, p. 151.
- BECK, J. E.; WOOLF, B. P. High-level student modeling with machine learning. In: INTERNATIONAL CONFERENCE ON INTELLIGENT TUTORING SYSTEMS, 2000. **Anais...** [S.l.: s.n.], 2000. p. 584–593.
- BEN-NAIM, D.; BAIN, M.; MARCUS, N. A user-driven and data-driven approach for supporting teachers in reflection and adaptation of adaptive tutorials. **International Working Group on Educational Data Mining**, [S.l.], 2009.
- BERCHT, M.; VICCARI, R. Pedagogical agents with affective and cognitive dimensions. In: V CONGRESO IBEROAMERICANO DE INFORMÁTICA EDUCATIVA, 2000. **Anais...** [S.l.: s.n.], 2000. p. 4–6.

- BERGER, K. S. **The developing person through the life span.** [S.l.: s.n.], 2003.
- BIXLER, R.; D'MELLO, S. Detecting boredom and engagement during writing with keystroke analysis, task appraisals, and stable traits. In: **INTELLIGENT USER INTERFACES**, 2013., 2013. **Proceedings...** [S.l.: s.n.], 2013. p. 225–234.
- BRUSILOVSKY, P.; PEYLO, C. Adaptive and intelligent web-based educational systems. **International Journal of Artificial Intelligence in Education (IJAIED)**, [S.l.], v. 13, p. 159–172, 2003.
- BURGER, J. M. **Personality.** [S.l.: s.n.], 2000.
- BURLESON, W. et al. A platform for affective agent research. In: **WORKSHOP ON EMPATHETIC AGENTS, INTERNATIONAL CONFERENCE ON AUTONOMOUS AGENTS AND MULTIAGENT SYSTEMS, COLUMBIA UNIVERSITY, NEW YORK, NY**, 2004. **Anais...** [S.l.: s.n.], 2004. v. 2, n. 4.
- BURLESON, W.; PICARD, R. Evidence for Gender Specific Approaches to the Development of Emotionally Intelligent Learning Companions. **IEEE Intelligent Systems, Special issue on Intelligent Educational Systems**, [S.l.], v. 22, n. 4, p. 62–69, 2007.
- Calvet Liñán, L.; Juan Pérez, Á. A. Educational data mining and learning analytics: differences, similarities, and time evolution. **RUSC. Universities and Knowledge Society Journal**, [S.l.], v. 12, n. 3, p. 98, 2015.
- CALVO, R. A.; D'MELLO, S. Affect detection: an interdisciplinary review of models, methods, and their applications. **IEEE Transactions on affective computing**, [S.l.], v. 1, n. 1, p. 18–37, 2010.
- CASTRO, F. et al. Applying data mining techniques to e-learning problems. In: **Evolution of teaching and learning paradigms in intelligent environment.** [S.l.]: Springer, 2007. p. 183–221.
- CHEN, T.; GUESTRIN, C. Xgboost: a scalable tree boosting system. In: **OF THE 22ND ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING**, 2016. **Proceedings...** [S.l.: s.n.], 2016. p. 785–794.
- CLEARY, J. G.; TRIGG, L. E. K*: an instance-based learner using an entropic distance measure. In: **Machine learning proceedings 1995.** [S.l.]: Elsevier, 1995. p. 108–114.
- COHEN, J. A coefficient of agreement for nominal scales. **Educational and psychological measurement**, [S.l.], v. 20, n. 1, p. 37–46, 1960.
- COLLINS, F. S.; MORGAN, M.; PATRINOS, A. The human genome project: lessons from large-scale biology. **Science**, [S.l.], v. 300, n. 5617, p. 286–290, 2003.
- CONATI, C. Probabilistic assessment of user's emotions in educational games. **Applied artificial intelligence**, [S.l.], v. 16, n. 7-8, p. 555–575, 2002.
- CONATI, C. Combining cognitive appraisal and sensors for affect detection in a framework for modeling user affect. In: **New perspectives on affect and learning technologies.** [S.l.]: Springer, 2011. p. 71–84.

CONATI, C.; MACLARE, H. Evaluating a probabilistic model of student affect. In: INTERNATIONAL CONFERENCE ON INTELLIGENT TUTORING SYSTEMS, 2004. **Anais...** [S.l.: s.n.], 2004. p. 55–66.

CONATI, C.; MACLAREN, H. Empirically building and evaluating a probabilistic model of user affect. **User Modeling and User-Adapted Interaction**, [S.l.], v. 19, n. 3, p. 267–303, 2009.

CONATI, C.; ZHOU, X. Modeling students' emotions from cognitive appraisal in educational games. In: INTERNATIONAL CONFERENCE ON INTELLIGENT TUTORING SYSTEMS, 2002. **Anais...** [S.l.: s.n.], 2002. p. 944–954.

COOPER, B. Care-making the affective leap: more than a concerned interest in a learner's cognitive abilities. **International Journal of Artificial Intelligence in Education**, [S.l.], v. 13, n. 1, p. 3–9, 2003.

CORBETT, A.; KOEDINGER, K.; ANDERSON, J. Intelligent tutoring systems. **Handbook of humancomputer interaction**, [S.l.], p. 849–874, 1997.

CORBETT, A. T.; ANDERSON, J. R. Knowledge tracing: modeling the acquisition of procedural knowledge. **User Modeling and User-Adapted Interaction**, [S.l.], v. 4, p. 253–278, 1995.

CRAIG, S. D. et al. Emote aloud during learning with autotutor: applying the facial action coding system to cognitive–affective states during learning. **Cognition and Emotion**, [S.l.], v. 22, n. 5, p. 777–788, 2008.

CRAIG, S. et al. Affect and learning: an exploratory look into the role of affect in learning with autotutor. **Journal of educational media**, [S.l.], v. 29, n. 3, p. 241–250, 2004.

CSIKSZENTMIHALYI, M. **Flow: the psychology of optimal performance**. [S.l.]: New York: Harper and Row, 1990.

DARWIN, C.; PRODGER, P. **The expression of the emotions in man and animals**. [S.l.]: Oxford University Press, USA, 1998.

DAVIDSON, R. J. et al. Approach-withdrawal and cerebral asymmetry: emotional expression and brain physiology: i. **Journal of personality and social psychology**, [S.l.], v. 58, n. 2, p. 330, 1990.

DAVIS, J.; GOADRICH, M. The relationship between precision-recall and roc curves. In: MACHINE LEARNING, 23., 2006. **Proceedings...** [S.l.: s.n.], 2006. p. 233–240.

DEKKER, G. W.; PECHENIZKIY, M.; VLEESHOUWERS, J. M. Predicting students drop out: a case study. **International Working Group on Educational Data Mining**, [S.l.], 2009.

DESMARAIS, M. et al. Conditions for effectively deriving a q-matrix from data with non-negative matrix factorization. In: EDM, 4., 2011. **Anais...** [S.l.: s.n.], 2011. p. 41–50.

DEVEDZIC, V.; HARRER, A. Software patterns in its architectures. **IJAIED**, [S.l.], v. 15, n. 2, p. 63–94, 2005.

- D'MELLO, S.; CALVO, R. A. Beyond the basic emotions: what should affective computing compute? In: CHI'13 EXTENDED ABSTRACTS ON HUMAN FACTORS IN COMPUTING SYSTEMS, 2013. **Anais...** [S.l.: s.n.], 2013. p. 2287–2294.
- D'MELLO, S.; PICARD, R. W.; GRAESSER, A. Toward an affect-sensitive autotutor. **IEEE Intelligent Systems**, [S.l.], v. 22, n. 4, 2007.
- DONNELLAN, D.; PAHL, C. Data mining technology for the evaluation of web-based teaching and learning systems. In: E-LEARN: WORLD CONFERENCE ON E-LEARNING IN CORPORATE, GOVERNMENT, HEALTHCARE, AND HIGHER EDUCATION, 2002. **Anais...** [S.l.: s.n.], 2002. p. 747–752.
- DRAGON, T. et al. Viewing student affect and learning through classroom observation and physical sensors. In: INTERNATIONAL CONFERENCE ON INTELLIGENT TUTORING SYSTEMS, 2008. **Anais...** [S.l.: s.n.], 2008. p. 29–39.
- DREISEITL, S.; OHNO-MACHADO, L. Logistic regression and artificial neural network classification models: a methodology review. **Journal of biomedical informatics**, [S.l.], v. 35, n. 5-6, p. 352–359, 2002.
- DUNN, G. et al. Evaluating interface variants on personality acquisition for recommender systems. In: INTERNATIONAL CONFERENCE ON USER MODELING, ADAPTATION, AND PERSONALIZATION, 2009. **Anais...** [S.l.: s.n.], 2009. p. 259–270.
- D'MELLO, S. Monitoring affective trajectories during complex learning. In: **Encyclopedia of the sciences of learning**. [S.l.]: Springer, 2012. p. 2325–2328.
- D'MELLO, S. et al. Confusion can be beneficial for learning. **Learning and Instruction**, [S.l.], v. 29, p. 153–170, 2014.
- D'MELLO, S. K. et al. Automatic detection of learner's affect from conversational cues. **User modeling and user-adapted interaction**, [S.l.], v. 18, n. 1-2, p. 45–80, 2008.
- D'MELLO, S. K.; GRAESSER, A. Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. **User Modeling and User-Adapted Interaction**, [S.l.], v. 20, n. 2, p. 147–187, 2010.
- EKMAN, P. All Emotions are Basic. In: DAVIDSON, R. (Ed.). **The nature of emotions: fundamental questions**. Oxford: Oxford University Press, 1994. p. 15–19.
- EKMAN, P. Facial expressions. **Handbook of cognition and emotion**, [S.l.], v. 16, p. 301–320, 1999.
- EL-NASR, M. S. et al. Emotionally expressive agents. In: COMPUTER ANIMATION, 1999. PROCEEDINGS, 1999. **Anais...** [S.l.: s.n.], 1999. p. 48–57.
- ELLIOTT, C. Affective reasoner personality models for automated tutoring systems. In: AI IN EDUCATION'97 WORKSHOP ON PEDAGOGICAL AGENTS, 1997. **Proceedings...** [S.l.: s.n.], 1997.
- ELLIOTT, C.; RICKEL, J.; LESTER, J. Lifelike pedagogical agents and affective computing: an exploratory synthesis. In: **Artificial intelligence today**. [S.l.]: Springer, 1999. p. 195–211.

- FANCSALI, S. Variable construction and causal discovery for cognitive tutor log data: initial results. In: EDUCATIONAL DATA MINING 2012, 2012. **Anais...** [S.l.: s.n.], 2012.
- FENG, M.; HEFFERNAN, N. T. Towards live informing and automatic analyzing of student learning: reporting in assistent system. **Journal of Interactive Learning Research**, [S.l.], v. 18, n. 2, p. 207, 2007.
- FEURER, M. et al. Efficient and robust automated machine learning. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 2015. **Anais...** [S.l.: s.n.], 2015. p. 2962–2970.
- FLEISS, J. L. Measuring nominal scale agreement among many raters. **Psychological bulletin**, [S.l.], v. 76, n. 5, p. 378, 1971.
- FORBES-RILEY, K.; LITMAN, D. Predicting emotion in spoken dialogue from multiple knowledge sources. In: HUMAN LANGUAGE TECHNOLOGY CONFERENCE OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: HLT-NAACL 2004, 2004. **Proceedings...** [S.l.: s.n.], 2004.
- FRASSON, C.; AIMEUR, E. Designing a multi-strategic intelligent tutoring system for training in industry. **Computers in industry**, [S.l.], v. 37, n. 2, p. 153–167, 1998.
- FREDRICKSON, B. L. What good are positive emotions? **Review of general psychology**, [S.l.], v. 2, n. 3, p. 300, 1998.
- FRIJDA, N. H. et al. **Varieties of affect: emotions and episodes, moods, and sentiments**. [S.l.]: New York: Oxford University Press., 1994.
- GANDOMI, A.; HAIDER, M. Beyond the hype: big data concepts, methods, and analytics. **International Journal of Information Management**, [S.l.], v. 35, n. 2, p. 137–144, 2015.
- GARTNER, I. Gartner it glossary. **Technology Research**, [S.l.], 2013.
- GOLEMAN, D. **Emotional intelligence**. New York: Bantam Books, 1995.
- GOSLING, S. **Snoop: what your stuff says about you**. [S.l.]: Hachette UK, 2009.
- GOSLING, S. D.; RENTFROW, P. J.; SWANN JR, W. B. A very brief measure of the big-five personality domains. **Journal of Research in personality**, [S.l.], v. 37, n. 6, p. 504–528, 2003.
- GRAESSER, A. C.; CONLEY, M. W.; OLNEY, A. Intelligent tutoring systems. In: **Apa educational psychology handbook, vol 3: application to learning and teaching**. Washington, DC, US: American Psychological Association, 2012. p. 451–473.
- GRAESSER, A. C. et al. Autotutor: an intelligent tutoring system with mixed-initiative dialogue. **IEEE Transactions on Education**, [S.l.], v. 48, n. 4, p. 612–618, 2005.
- GRAESSER, A.; D’MELLO, S. K. Theoretical perspectives on affect and deep learning. In: **New perspectives on affect and learning technologies**. [S.l.]: Springer, 2011. p. 11–21.
- GROB, H. L.; BENSBERG, F.; KADERALI, F. Controlling open source intermediaries—a web log mining approach. In: INFORMATION TECHNOLOGY INTERFACES, 2004. 26TH INTERNATIONAL CONFERENCE ON, 2004. **Anais...** [S.l.: s.n.], 2004. p. 233–242.

- GUIA, T. F. G. et al. Modeling the affective states of students using sql-tutor. In: INTERNATIONAL CONFERENCE ON INTELLIGENT TUTORING SYSTEMS, 2012. **Anais...** [S.l.: s.n.], 2012. p. 634–635.
- GUNES, H.; PICCARDI, M. Bi-modal emotion recognition from expressive face and body gestures. **Journal of Network and Computer Applications**, [S.l.], v. 30, n. 4, p. 1334–1345, 2007.
- HAN, J.; PEI, J.; KAMBER, M. **Data mining**: concepts and techniques. [S.l.]: Elsevier, 2011.
- HAND, D. J. Principles of data mining. **Drug safety**, [S.l.], v. 30, n. 7, p. 621–622, 2007.
- HANLEY, J. A.; MCNEIL, B. J. The meaning and use of the area under a receiver operating characteristic (roc) curve. **Radiology**, [S.l.], v. 143, n. 1, p. 29–36, 1982.
- HERAUD, J.-M.; FRANCE, L.; MILLE, A. Pixed: an its that guides students with the help of learners' interaction log. In: INTERNATIONAL CONFERENCE ON INTELLIGENT TUTORING SYSTEMS, WORKSHOP ANALYZING STUDENT TUTOR INTERACTION LOGS TO IMPROVE EDUCATIONAL OUTCOMES. MACEIO, BRAZIL, 2004. **Anais...** [S.l.: s.n.], 2004. p. 57–64.
- HUBERT, L. Kappa revisited. **Psychological Bulletin**, [S.l.], v. 84, n. 2, p. 289, 1977.
- IHANTOLA, P. et al. Educational data mining and learning analytics in programming : literature review and case studies. **ITiCSE WGR'16**, [S.l.], p. 41–63, 2015.
- IZARD, C. E. **Emotion-cognition relationships and human development**. [S.l.: s.n.], 1984. 17–37 p.
- IZARD, C. E. Emotion theory and research: highlights, unanswered questions, and emerging issues. **Annual review of psychology**, [S.l.], v. 60, p. 1–25, 2009.
- JAMES, W. What is an emotion? **Mind**, [S.l.], v. 9, n. 34, p. 188–205, 1884.
- JAQUES, P. A. et al. Applying affective tactics for a better learning. In: ECAI, 2004. **Anais...** [S.l.: s.n.], 2004. v. 16, p. 109.
- JAQUES, P. A. et al. Evaluating a Cognitive-Based Affective Student Model. In: INTERNATIONAL CONFERENCE ON AFFECTIVE COMPUTING AND INTELLIGENT INTERACTION (ACII), 2011. **Anais...** Springer, 2011. p. 599–608. (Lecture Notes in Computer Science, v. 6974).
- JAQUES, P. A. et al. Rule-based expert systems to support step-by-step guidance in algebraic problem solving: The case of the tutor PAT2math. **Expert Systems with Applications**, [S.l.], v. 40, n. 14, p. 5456–5465, 2013.
- JAQUES, P. A.; JAQUES, K. S. F.; LEHMANN, M. Avaliando a efetividade de um agente pedagógico animado emocional. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 2008, Fortaleza. **Anais...** SBC, 2008. p. 145–154.
- JAQUES, P. A.; LEHMANN, M.; PESTY, S. Evaluating the Affective Tactics of an Emotional Pedagogical Agent. In: ACM SYMPOSIUM ON APPLIED COMPUTING, 2009, Hawaii. **Anais...** ACM, 2009. v. 1, p. 104–109.

JAQUES, P. A.; NUNES, M. A. S. N. Computação afetiva aplicada à educação. **PIMENTEL, Mariano; SAMPAIO, Fábio F.; SANTOS, Edméa O. (Org.). Informática na Educação: técnicas e tecnologias computacionais**, Porto Alegre, v. 3, 2019. Disponível em: <<http://ieducacao.ceie-br.org/computacaoafetiva>>.

JAQUES, P. A.; VICARI, R. A BDI approach to infer student's emotions in an intelligent learning environment. **Computers & Education**, Puebla, Mexico, v. 49, n. 2, p. 360–384, sep 2007.

JAQUES, P. A.; VICARI, R. M. Estado da arte em ambientes inteligentes de aprendizagem que consideram a afetividade do aluno. **Revista informática na educação: teoria & prática**, [S.l.], v. 8, n. 1, p. 15–38, 2005.

JAQUES, P. A.; VICCARI, R. M. A bdi approach to infer student's emotions. In: IBERO-AMERICAN CONFERENCE ON ARTIFICIAL INTELLIGENCE, 2004. **Anais...** [S.l.: s.n.], 2004. p. 901–911.

JAQUES, P. et al. Interação com agentes pedagógicos animados: um estudo comparativo. In: WORKSHOP SOBRE FATORES HUMANOS EM SISTEMAS COMPUTACIONAIS, IHC, 2001. **Anais...** [S.l.: s.n.], 2001. v. 4.

JIANG, Y. et al. Expert feature-engineering vs. deep neural networks: which is better for sensor-free affect detection? In: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE IN EDUCATION, 2018. **Anais...** [S.l.: s.n.], 2018. p. 198–211.

JOHN, O. P.; SRIVASTAVA, S. The big five trait taxonomy: history, measurement, and theoretical perspectives. **Handbook of personality: Theory and research**, [S.l.], v. 2, n. 1999, p. 102–138, 1999.

JOHN-STEINER, V. **Creative collaboration**. [S.l.]: Oxford University Press, 2000.

JOHNSON, J. A. Web-based personality assessment. In: BALTIMORE, MD, 71., 2000. **Anais...** [S.l.: s.n.], 2000.

JOHNSON, J. A. Ascertaining the validity of individual protocols from web-based personality inventories. **Journal of research in personality**, [S.l.], v. 39, n. 1, p. 103–129, 2005.

KAUR, G.; CHHABRA, A. Improved j48 classification algorithm for the prediction of diabetes. **International Journal of Computer Applications**, [S.l.], v. 98, n. 22, 2014.

KIRON, D. et al. Analytics: the widening divide. **MIT Sloan Management Review**, [S.l.], v. 53, n. 2, p. 1, 2012.

KOEDINGER, K. et al. Intelligent tutoring goes to school in the big city. **IJAIED**, [S.l.], v. 8, n. 1, p. 30–43, 1997.

KOEDINGER, K. R. et al. Learning is not a spectator sport. In: ACM CONF. ON LEARNING SCALE, 2015. **Anais...** [S.l.: s.n.], 2015. p. 111–120.

KOEDINGER, K. R.; MCLAUGHLIN, E. A.; STAMPER, J. C. Automated student model improvement. In: 2012. **Anais...** [S.l.: s.n.], 2012. p. 17–24.

KOHAVI, R. et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: IJCAI, 1995. **Anais...** [S.l.: s.n.], 1995. v. 14, n. 2, p. 1137–1145.

- KOPECEK, I. Emotions and prosody in dialogues: an algebraic approach based on user modelling. In: ISCA TUTORIAL AND RESEARCH WORKSHOP (ITRW) ON SPEECH AND EMOTION, 2000. **Anais...** [S.l.: s.n.], 2000.
- KWON, O.; SIM, J. M. Effects of data set features on the performances of classification algorithms. **Expert Systems with Applications**, [S.l.], v. 40, n. 5, p. 1847–1857, 2013.
- LAGUD, M. C. V.; RODRIGO, M. M. T. The affective and learning profiles of students using an intelligent tutoring system for algebra. In: INTERNATIONAL CONFERENCE ON INTELLIGENT TUTORING SYSTEMS, 2010. **Anais...** [S.l.: s.n.], 2010. p. 255–263.
- LANEY, D. 3d data management: controlling data volume, velocity and variety. **META Group Research Note**, [S.l.], v. 6, n. 70, 2001.
- LAZARUS, R. S. Progress on a cognitive-motivational-relational theory of emotion. **American psychologist**, [S.l.], v. 46, n. 8, p. 819, 1991.
- LEE, D. M. C. et al. Exploring the relationship between novice programmer confusion and achievement. In: INTERNATIONAL CONFERENCE ON AFFECTIVE COMPUTING AND INTELLIGENT INTERACTION, 2011. **Anais...** [S.l.: s.n.], 2011. p. 175–184.
- LEHMAN, B.; D’MELLO, S.; PERSON, N. All alone with your emotions: an analysis of student emotions during effortful problem solving activities. In: WORKSHOP ON EMOTIONAL AND COGNITIVE ISSUES IN ITS AT THE NINTH INTERNATIONAL CONFERENCE ON INTELLIGENT TUTORING SYSTEMS, 2008. **Anais...** [S.l.: s.n.], 2008.
- LEHMAN, B. et al. What are you feeling? investigating student affective states during expert human tutoring sessions. In: INTERNATIONAL CONFERENCE ON INTELLIGENT TUTORING SYSTEMS, 2008. **Anais...** [S.l.: s.n.], 2008. p. 50–59.
- LEITE, I. et al. ” why can’t we be friends?” an empathic game companion for long-term interaction. In: INTERNATIONAL CONFERENCE ON INTELLIGENT VIRTUAL AGENTS, 2010. **Anais...** [S.l.: s.n.], 2010. p. 315–321.
- LEITE, I. et al. Modelling empathic behaviour in a robotic game companion for children: an ethnographic study in real-world settings. In: ACM/IEEE INTERNATIONAL CONFERENCE ON HUMAN-ROBOT INTERACTION, 2012. **Proceedings...** [S.l.: s.n.], 2012. p. 367–374.
- LI, J.; ZAIĀNE, O. R. Combining usage, content, and structure data to improve web site recommendation. In: INTERNATIONAL CONFERENCE ON ELECTRONIC COMMERCE AND WEB TECHNOLOGIES, 2004. **Anais...** [S.l.: s.n.], 2004. p. 305–315.
- LIGHT, R. J. Measures of response agreement for qualitative data: some generalizations and alternatives. **Psychological bulletin**, [S.l.], v. 76, n. 5, p. 365, 1971.
- LITMAN, D. J.; FORBES-RILEY, K. Recognizing student emotions and attitudes on the basis of utterances in spoken tutoring dialogues with both human and computer tutors. **Speech communication**, [S.l.], v. 48, n. 5, p. 559–590, 2006.
- LITMAN, D. J.; SILLIMAN, S. ITSPOKE: An intelligent tutoring spoken dialogue system. In: DEMONSTRATION PAPERS AT HLT-NAACL 2004, 2004. **Anais...** Association for Computational Linguistics, 2004. p. 5–8.

- LOEHLIN, J. C. **Genes and environment in personality development**. [S.l.]: Sage Publications, Inc, 1992.
- LOYALL, A. B.; BATES, J. Personality-rich believable agents that use language. In: AUTONOMOUS AGENTS, 1997. **Proceedings...** [S.l.: s.n.], 1997. p. 106–113.
- LU, J. A personalized e-learning material recommender system. In: INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY AND APPLICATIONS, 2004. **Anais...** [S.l.: s.n.], 2004.
- LUAN, J. **Data mining and knowledge management in higher education-potential applications**. [S.l.]: ERIC, 2002.
- MA, W. et al. Intelligent tutoring systems and learning outcomes: a meta-analysis. **Journal of Educational Psychology**, [S.l.], v. 106, n. 4, p. 901, 2014.
- MACHADO, A. A. d. A. **Inferência de personalidade a partir de textos de rede social utilizando um léxico afetivo em português brasileiro. (dissertação de mestrado em ciência da computação)**. [S.l.]: Universidade Federal de Sergipe, 2016.
- MARTINHO, C.; MACHADO, I.; PAIVA, A. A cognitive approach to affective user modeling. In: **Affective interactions - towards a new generation of computer interfaces**. [S.l.]: Springer, 2000. p. 64–75.
- MERCERON, A.; YACEF, K. Mining student data captured from a web-based tutoring tool: initial exploration and results. **Journal of Interactive Learning Research**, [S.l.], v. 15, n. 4, p. 319, 2004.
- MERCERON, A.; YACEF, K. Educational data mining: a case study. In: AIED, 2005. **Anais...** [S.l.: s.n.], 2005. p. 467–474.
- MINAEI-BIDGOLI, B. et al. Predicting student performance: an application of data mining methods with an educational web-based system. In: FRONTIERS IN EDUCATION, 2003. FIE 2003 33RD ANNUAL, 2003. **Anais...** [S.l.: s.n.], 2003. v. 1, p. T2A–13.
- MOHAMAD, S. K.; TASIR, Z. Educational data mining: a review. **Procedia - Social and Behavioral Sciences**, [S.l.], v. 97, p. 320–324, 2013.
- MOR, E.; MINGUILLÓN, J. E-learning personalization based on itineraries and long-term navigational behavior. In: WORLD WIDE WEB CONFERENCE ON ALTERNATE TRACK PAPERS & POSTERS, 13., 2004. **Proceedings...** [S.l.: s.n.], 2004. p. 264–265.
- MORAIS, F. de et al. Emap-ml: a protocol of emotions and behaviors annotation for machine learning labels. In: EC-TEL, 2019, Netherlands. **Anais...** Springer, 2019.
- MORAIS, F.; JAQUES, P. Avaliação de usabilidade do sistema tutor inteligente pat2math. **RENOTE**, [S.l.], v. 11, n. 3, 2013.
- MORAIS, F.; JAQUES, P. Pat2math + handwriting: evoluindo sistemas tutores de matemática com reconhecimento da escrita à mão. In: BRAZILIAN SYMPOSIUM ON COMPUTERS IN EDUCATION (SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO-SBIE), 2017. **Anais...** [S.l.: s.n.], 2017. v. 28, n. 1, p. 1237.

- MOTA, S.; PICARD, R. W. Automated posture analysis for detecting learner's interest level. In: COMPUTER VISION AND PATTERN RECOGNITION WORKSHOP, 2003. CVPRW'03. CONFERENCE ON, 2003. **Anais...** [S.l.: s.n.], 2003. v. 5, p. 49–49.
- MULDNER, K.; BURLESON, W.; VANLEHN, K. “yes!”: using tutor and sensor data to predict moments of delight during instructional activities. In: INTERNATIONAL CONFERENCE ON USER MODELING, ADAPTATION, AND PERSONALIZATION, 2010. **Anais...** [S.l.: s.n.], 2010. p. 159–170.
- MURRAY, T. Authoring intelligent tutoring systems: An analysis of the state of the art. **IJAIED**, [S.l.], v. 10, p. 98–129, 1999.
- NUNES, M. A. S. N. **Recommender systems based on personality traits: could human psychological aspects influence the computer decision-making process?** [S.l.]: VDM-Verlag Müller, 2009.
- NUNES, M. Computação afetiva personalizando interfaces, interações e recomendações de produtos, serviços e pessoas em ambientes computacionais. **DCOMP e PROCC: Pesquisas e Editora UFS: São Cristóvão**, [S.l.], p. 115–151, 2012.
- NUNES, M.; CAZELLA, S. O que sua personalidade revela? fidelizando clientes web através de sistemas de recomendação e traços de personalidade. **Tópicos em Banco de Dados e Multimídia e Web. Porto Alegre: SBC**, [S.l.], v. 1, p. 91–122, 2011.
- NWANA, H. S. Intelligent tutoring systems: an overview. **Artificial Intelligence Review**, [S.l.], v. 4, n. 4, p. 251–277, 1990.
- OCUMPAUGH, J. Baker rodrigo ocumpaugh monitoring protocol (bromp) 2.0 technical and training manual. **New York, NY and Manila, Philippines: Teachers College, Columbia University and Ateneo Laboratory for the Learning Sciences**, [S.l.], 2015.
- OCUMPAUGH, J. et al. Population validity for educational data mining models: a case study in affect detection. **British Journal of Educational Technology**, [S.l.], v. 45, n. 3, p. 487–501, 2014.
- OJOSE, B. Students' misconceptions in mathematics: analysis of remedies and what research says. **Ohio Journal of School Mathematics**, [S.l.], v. 72, p. 30–34, 2015.
- OLIVEIRA, E. de; JAQUES, P. A. Inferindo as emoções do usuário pela face através de um sistema psicológico de codificação facial. In: VIII BRAZILIAN SYMPOSIUM ON HUMAN FACTORS IN COMPUTING SYSTEMS, 2008. **Proceedings...** [S.l.: s.n.], 2008. p. 156–165.
- ORTONY, A.; CLORE, G. L.; COLLINS, A. **The cognitive structure of emotions**. [S.l.]: Cambridge university press, 1990.
- PANACCIO, A.; VANDENBERGHE, C. Five-factor model of personality and organizational commitment: the mediating role of positive and negative affective states. **Journal of vocational behavior**, [S.l.], v. 80, n. 3, p. 647–658, 2012.
- PAPAMITSIOU, Z.; ECONOMIDES, A. A. Learning analytics and educational data mining in practice: a systematic literature review of empirical evidence. **Educational Technology & Society**, [S.l.], v. 17, n. 4, p. 49–64, 2014.

PAQUETTE, L. et al. Sensor-free affect detection for a simulation-based science inquiry learning environment. In: INTERNATIONAL CONFERENCE ON INTELLIGENT TUTORING SYSTEMS, 2014. **Anais...** [S.l.: s.n.], 2014. p. 1–10.

PAQUETTE, L. et al. Sensor-free or sensor-full: a comparison of data modalities in multi-channel affect detection. **International Educational Data Mining Society**, [S.l.], 2016.

PARDOS, Z. A. et al. Affective states and state tests: investigating how affect and engagement during the school year predict end-of-year learning outcomes. **Journal of Learning Analytics**, [S.l.], v. 1, n. 1, p. 107–128, 2014.

PEDRO, M. O. et al. Predicting college enrollment from student interaction with an intelligent tutoring system in middle school. In: EDUCATIONAL DATA MINING 2013, 2013. **Anais...** [S.l.: s.n.], 2013.

PEKRUN, R. Emotions and learning. **International Academy of Education/International Bureau of Education, IEA (<http://www.iaoed.org>), IBE (<http://www.ibe.unesco.org/publications.htm>)**, [S.l.], 2014.

PEKRUN, R. et al. **Positive emotions in education**. 2002.

PEKRUN, R. et al. Boredom in achievement settings: exploring control–value antecedents and performance outcomes of a neglected emotion. **Journal of Educational Psychology**, [S.l.], v. 102, n. 3, p. 531, 2010.

PERERA, D. et al. Clustering and sequential pattern mining of online collaborative learning data. **IEEE Transactions on Knowledge and Data Engineering**, [S.l.], v. 21, n. 6, p. 759–772, 2009.

PETRANTONAKIS, P. C.; HADJILEONTIADIS, L. J. Emotion recognition from eeg using higher order crossings. **IEEE Transactions on Information Technology in Biomedicine**, [S.l.], v. 14, n. 2, p. 186–197, 2010.

PIAGET, J. Les relations entre l'intelligence et l'affectivité dans le development de l'enfant in: rimé. **B., Scherer, K (eds.) Les Émotions Textes de base en psychologie**, Paris: Delachaux et Niestlé, p. 75–95, 1989.

PICARD, R. W. **Affective computing**. [S.l.]: MIT Press, 1997. 292 p.

PICARD, R. W.; VYZAS, E.; HEALEY, J. Toward machine emotional intelligence: analysis of affective physiological state. **IEEE transactions on pattern analysis and machine intelligence**, [S.l.], v. 23, n. 10, p. 1175–1191, 2001.

PORTO, S. M. et al. Como a extração de personalidade através do teclado pode beneficiar a personalização na educação. **Anais do XXII SBIE-XVII WIE**, [S.l.], 2011.

PSOTKA, J.; MASSEY, L. D.; MUTTER, S. A. **Intelligent tutoring systems: Lessons learned**. [S.l.]: Psychology Press, 1988.

RAJPUT, A. et al. J48 and jrip rules for e-governance data. **International Journal of Computer Science and Security (IJCSS)**, [S.l.], v. 5, n. 2, p. 201, 2011.

RANDOLPH, J. **Online kappa calculator**. <http://justus.randolph.name/kappa>, Computer Software.

RANDOLPH, J. J. Free-marginal multirater kappa (multirater k [free]): an alternative to fleiss' fixed-marginal multirater kappa. **Online submission**, [S.l.], 2005.

RAU, M. A.; SCHEINES, R. Searching for variables and models to investigate mediators of learning from multiple representations. In: PROCEEDINGS OF THE 5TH INTERNATIONAL CONFERENCE ON EDUCATIONAL DATA MINING, 2012. **Anais...** [S.l.: s.n.], 2012. p. 110–117.

REILLY, W. S.; BATES, J. Building emotional agents. **Technical Report CMU-CS-92-143**, School of Computer Science, Carnegie Mellon University, Pittsburgh, 1992.

REIS, H. et al. Analysis of permanence time in emotional states: a case study using educational software. In: INTERNATIONAL CONFERENCE ON INTELLIGENT TUTORING SYSTEMS, 2018. **Anais...** [S.l.: s.n.], 2018. p. 180–190.

ROLL, I. et al. Can help seeking be tutored? searching for the secret sauce of metacognitive tutoring. In: AIED, 2007. **Anais...** [S.l.: s.n.], 2007. v. 2007, p. 203–10.

ROMERO, C. et al. Evolutionary algorithms for subgroup discovery in e-learning: a practical application using moodle data. **Expert Systems with Applications**, [S.l.], v. 36, n. 2, p. 1632–1644, 2009.

ROMERO, C.; VENTURA, S. Educational data mining: a survey from 1995 to 2005. **Expert Systems with Applications**, [S.l.], v. 33, n. 1, p. 135–146, 2007.

ROMERO, C.; VENTURA, S. Educational data mining: a review of the state of the art. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**, [S.l.], v. 40, n. 6, p. 601–618, 2010.

ROMERO, C.; VENTURA, S.; DE BRA, P. Knowledge discovery with genetic programming for providing feedback to courseware authors. **User Modeling and User-Adapted Interaction**, [S.l.], v. 14, n. 5, p. 425–464, 2004.

ROWE, J. P. et al. Off-task behavior in narrative-centered learning environments. In: AIED, 2009. **Anais...** [S.l.: s.n.], 2009. p. 99–106.

SABOURIN, J.; MOTT, B.; LESTER, J. C. Modeling learner affect with theoretically grounded dynamic bayesian networks. In: INTERNATIONAL CONFERENCE ON AFFECTIVE COMPUTING AND INTELLIGENT INTERACTION, 2011. **Anais...** [S.l.: s.n.], 2011. p. 286–295.

SCHERER, K. R. **Appraisal theory**. [S.l.]: John Wiley & Sons Ltd, 1999.

SCHERER, K. R. Psychological models of emotion. **The neuropsychology of emotion**, [S.l.], v. 137, n. 3, p. 137–162, 2000.

SCHERER, K. R. Appraisal considered as a process of multilevel sequential checking. **Appraisal processes in emotion: Theory, methods, research**, [S.l.], v. 92, n. 120, p. 57, 2001.

SCHERER, K. R. What are emotions? and how can they be measured? **Social Science Information**, [S.l.], v. 44, n. 4, p. 695–729, 2005.

- SCHEUER, O.; MCLAREN, B. M. Educational data mining. In: **Encyclopedia of the sciences of learning**. [S.l.]: Springer, 2012. p. 1075–1079.
- SCHULTZ, D. P. **Theories of personality**. [S.l.]: Pacific Grove: Brooks/Cole, 1990.
- SEBE, N. et al. Multimodal approaches for emotion recognition: a survey. In: INTERNET IMAGING VI, 2005. **Anais...** [S.l.: s.n.], 2005. v. 5670, p. 56–68.
- SEFFRIN, H. et al. Resolvendo equações algébricas no sti pat2math. In: SBIE, 2010. **Anais...** [S.l.: s.n.], 2010.
- SEFFRIN, H. et al. Dicas inteligentes no Sistema Tutor Inteligente PAT2math. In: SBIE, 2012. **Anais...** [S.l.: s.n.], 2012.
- SEFFRIN, H.; JAQUES, P. Avaliando o conhecimento algébrico dos estudantes através de redes bayesianas dinâmicas. In: BRAZILIAN SYMPOSIUM ON COMPUTERS IN EDUCATION (SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO-SBIE), 2015. **Anais...** [S.l.: s.n.], 2015. v. 26, n. 1, p. 987.
- SEFFRIN, H.; RUBI, G.; JAQUES, P. O modelo cognitivo do sistema tutor inteligente pat2math. In: SBIE, 2011. **Anais...** [S.l.: s.n.], 2011. v. 1, n. 1.
- SHANABROOK, D. H. et al. Identifying high-level student behavior using sequence-based motif discovery. In: EDUCATIONAL DATA MINING 2010, 2010. **Anais...** [S.l.: s.n.], 2010.
- SHUTE, V.; PSOTKA, J. **Intelligent Tutoring Systems: Past, Present, and Future**. [S.l.]: DTIC Document, 1994.
- SIEMENS, G.; BAKER, R. S. d. Learning analytics and educational data mining: towards communication and collaboration. In: OF THE 2ND INTERNATIONAL CONFERENCE ON LEARNING ANALYTICS AND KNOWLEDGE, 2012. **Proceedings...** [S.l.: s.n.], 2012. p. 252–254.
- SLOMAN, A. How many separately evolved emotional beasts live within us. **Emotions in humans and artifacts**, [S.l.], p. 35–114, 2002.
- SOCIETY for learning analytics research (solar). (Acessado em 27/06/2018), <https://solaresearch.org/>.
- SOLDZ, S.; VAILLANT, G. E. The big five personality traits and the life course: a 45-year longitudinal study. **Journal of Research in Personality**, [S.l.], v. 33, n. 2, p. 208–232, 1999.
- STEENBERGEN-HU, S.; COOPER, H. A meta-analysis of the effectiveness of intelligent tutoring systems on k–12 students' mathematical learning. **Journal of Educational Psychology**, [S.l.], v. 105, n. 4, p. 970, 2013.
- SUMMERS, D.; COLLABORATION, F. E. et al. Charm physics at fermilab e791. **arXiv preprint hep-ex/0009015**, [S.l.], 2000.
- SUTTER, J. M.; KALIVAS, J. H. Comparison of forward selection, backward elimination, and generalized simulated annealing for variable selection. **Microchemical journal**, [S.l.], v. 47, n. 1-2, p. 60–66, 1993.

- TANG, T. Y.; MCCALLA, G. Smart recommendation for an evolving e-learning system: architecture and experiment. **International Journal on elearning**, [S.l.], v. 4, n. 1, p. 105, 2005.
- TCHERKASSOF, A. Les indices de préparation à l'action et la reconnaissance des expressions émotionnelles faciales. **European Review of Applied Psychology/Revue Européenne de Psychologie Appliquée**, [S.l.], 1999.
- URBANCIC, T.; SKRJANC, M.; FLACH, P. Web-based analysis of data mining and decision support education. **AI Communications**, [S.l.], v. 15, n. 4, p. 199–204, 2002.
- VAIL, P. L. **Emotion: the on/off switch for learning**. Rosemont, NJ: Modern Learning Press, 1994.
- VAN RIJSBERGEN, C. J. Foundation of evaluation. **Journal of Documentation**, [S.l.], v. 30, n. 4, p. 365–373, 1974.
- VANLEHN, K. The behavior of tutoring systems. **IJAIED**, [S.l.], v. 16, n. 3, p. 227–265, 2006.
- VANLEHN, K. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. **Educational Psychologist**, [S.l.], v. 46, n. 4, p. 197–221, 2011.
- VANLEHN, K. et al. The Andes physics tutoring system: Lessons learned. **IJAIED**, [S.l.], v. 15, n. 3, p. 147–204, 2005.
- VICENTE, A. de; PAIN, H. Motivation self-report in its. In: NINTH WORLD CONFERENCE ON ARTIFICIAL INTELLIGENCE IN EDUCATION, 1999. **Proceedings...** [S.l.: s.n.], 1999. p. 648–650.
- VYGOTSKY, L. S. The problem of the environment. **The vygotsky reader**, Cambridge, MA, p. 338–354, 1994.
- WANG, Y.; HEFFERNAN, N. T.; HEFFERNAN, C. Towards better affect detectors: effect of missing skills, class features and common wrong answers. In: FIFTH INTERNATIONAL CONFERENCE ON LEARNING ANALYTICS AND KNOWLEDGE, 2015. **Proceedings...** [S.l.: s.n.], 2015. p. 31–35.
- WEHRLE, T.; KAISER, S. Emotion and facial expression. In: **Affective interactions**. [S.l.]: Springer, 2000. p. 49–63.
- WOOLF, B. et al. Affect-aware tutors: recognising and responding to student affect. **International Journal of Learning Technology**, [S.l.], v. 4, n. 3/4, p. 129–164, 2009.
- WU, X. et al. Data mining with big data. **IEEE transactions on knowledge and data engineering**, [S.l.], v. 26, n. 1, p. 97–107, 2014.
- YUAN, L. An improved naive bayes text classification algorithm in chinese information processing. **Science**, [S.l.], p. 267–269, 2010.
- ZHOU, Y.; EVENS, M. W. A practical student model in an intelligent tutoring system. In: INT. CONF. ON TOOLS WITH ARTIFICIAL INTELLIGENCE, 1999. **Anais...** [S.l.: s.n.], 1999. p. 13–18.

ZORRILLA, M. E. et al. Web usage mining project for improving web-based learning sites.
In: INTERNATIONAL CONFERENCE ON COMPUTER AIDED SYSTEMS THEORY,
2005. **Anais...** [S.l.: s.n.], 2005. p. 205–210.