

Considerando personalidade e transições de emoções na detecção de emoções baseada em mineração de dados

Title: Considering personality and the transition of emotions to improve data mining-based emotion detection

Felipe de Moraes
PPG em Computação Aplicada (PPGCA)
Universidade do Vale do Rio dos Sinos (UNISINOS)
felipmoraes@edu.unisinos.br

Patrícia A. Jaques
PPG em Computação Aplicada (PPGCA)
Universidade do Vale do Rio dos Sinos (UNISINOS)
pjaques@unisinos.br

Resumo

Sistemas Tutores Inteligentes (STI) baseados em passos são capazes de auxiliar os alunos na resolução de tarefas passo a passo, gerando uma grande quantidade de dados de interação, chamados logs, entre o sistema e o aluno. Este trabalho aplica técnicas de mineração de dados sobre os logs gerados de um STI de matemática baseado em passos para detectar quatro emoções de aprendizagem: confusão, engajamento, frustração e tédio. O objetivo deste trabalho é verificar se a personalidade dos alunos pode proporcionar uma melhoria na precisão da detecção destas quatro emoções. Para realizar o treinamento dos detectores, foram utilizados rótulos de emoções dos alunos, obtidos por meio de um protocolo de anotação de emoções, baseado em análise de vídeos, que permite também a captura das transições de emoções. Como resultado, foi possível identificar que a personalidade impactou somente na detecção do engajamento. Embora a diferença na precisão tenha sido pequena, foi possível verificar que, dentre 348 características disponíveis durante o treinamento, a personalidade foi considerada uma das dez características mais representativas. Com a combinação dos dados de personalidade, transições de emoções e logs capturados de um STI baseado em passos, foi possível atingir um índice $K = 0,633$ e $A' = 0,846$ na detecção de engajamento, que são valores superiores aos mínimos exigidos de codificadores humanos em protocolos de anotação de emoções.

Palavras-chave: Detecção de Emoções Livre de Sensores; Personalidade; Transição de Emoções; Mineração de Dados Educacionais; Sistemas Tutores Inteligentes Baseados em Passos.

Abstract

This paper uses data mining, with data coming from the interaction of the students in a step-based ITS, to detect four learning emotions: confusion, engagement, frustration, and boredom. Unlike related works, our model aims to verify whether the students' personality can improve the precision of the detection. Besides personality, we have also considered data from the transition of emotions. For that, the emotion labels were obtained through an annotation protocol which allows the capture of transitions of the students' emotions. As results, we have identified that only the engagement detector, trained with personality data of the students, obtained a small precision improvement in the detection. However, with the use of a feature selection algorithm, it was possible to verify that among 348 available features, only ten were selected, including personality data. With the combination of personality data, emotion transitions, and logs captured from a step-based ITS, it was possible to reach an accuracy of $K = .633$ and $A' = .846$ in detecting engagement. This being the minimum value required for human encoders in emotion detection protocols.

Keywords: Sensor-free Affect Detection; Personality; Transition of Emotions; Educational Data Mining; Step-Based Intelligent Tutoring Systems.

1 Introdução

Estar presente em sala de aula não está diretamente relacionado ao ato de aprender. Vários fatores cognitivos e afetivos podem fazer com que o aluno se desvie de seu objetivo. Afeto, motivação e meta-cognição podem influenciar a cognição, interferindo no processo de aprendizagem (R. Azevedo & Alevén, 2013). Mais especificamente, emoções, um tipo de estado afetivo (Scherer, 2005), desempenham um papel fundamental para os alunos em sala de aula. Elas podem afetar o aprendizado de maneira positiva ou negativa, pois são capazes de controlar a atenção, motivação, cognição, modelagem de estratégias e a autorregulação da aprendizagem (Pekrun, 2014). Emoções positivas como alegria, engajamento e curiosidade podem afetar aspectos físicos, sociais, intelectuais e criativos, influenciando positivamente no desenvolvimento dessas habilidades (Fredrickson, 1998). Por outro lado, emoções negativas, como frustração e tédio, podem causar um desequilíbrio cognitivo, afetando negativamente no processo de aprendizagem (Graesser & D’Mello, 2011). Confusão tem sido vista como um caso particular porque ela pode melhorar a aprendizagem quando regulada e resolvida apropriadamente, mas também pode prejudicar caso contrário (D’Mello, Lehman, Pekrun, & Graesser, 2014). Portanto, é importante que os ambientes computacionais de aprendizagem sejam capazes de detectar e adaptar suas estratégias pedagógicas de acordo com as emoções dos alunos (Arroyo et al., 2014).

Para a realização da detecção das emoções, diferentes métodos têm sido utilizados pela comunidade científica, incluindo expressões faciais, síntese de voz, comportamento observável, entre outros (Jaques & Nunes, 2019). Porém, alguns desses métodos utilizam equipamentos específicos para realizar a detecção, como câmeras ou sensores, adicionando um custo e uma intromissão em aplicações no mundo real (Calvo & D’Mello, 2010). Em ambientes computacionais de aprendizagem, por exemplo, a necessidade de câmeras ou sensores para a utilização desses métodos pode inviabilizar a aplicação desses sistemas em larga escala e por um longo período (Baker et al., 2012), podendo, inclusive, deixar os alunos desconfortáveis (Shanabrook, Arroyo, & Woolf, 2012).

Uma fonte de dados mais natural e menos invasiva que vem sendo considerada para a detecção de emoções em ambientes computacionais de aprendizagem pode ser obtida através da interação dos alunos com o computador por meio de sua interface gráfica com o usuário (GUI) (Salmeron-Majadas, Santos, & Boticario, 2014; Harley, 2016). Durante o uso destes ambientes, vários registros das ações dos alunos, chamados *logs*, são gerados e armazenados, contendo informações sobre todos os tipos de interação dos alunos com o sistema. Ao usar um Sistema Tutor Inteligente (STI), por exemplo, o aluno interage com o conteúdo por meio da interface gráfica, resolvendo tarefas e obtendo *feedback* do STI. Assim, cada uma dessas interações é considerada um *log* e contém informações sobre passos de resolução corretos ou errados, dicas solicitadas, *feedback* de erros, entre outros eventos da interface gráfica. Mais especificamente, os STIs baseados em passos (Vanlehn, 2006) podem ajudar o aluno em todas as etapas de resolução de uma determinada tarefa, fornecendo dicas, *feedback* mínimo e *feedback* de erro. Assim, esse tipo de ambiente computacional de aprendizagem gera muitos dados de interação entre o aluno e

o sistema.

Pesquisas mostram que é possível detectar as emoções dos alunos nestes ambientes através da mineração desses dados de *log* (Paquette et al., 2016). Esse tipo de detecção de emoção não requer sensores específicos e foi chamado de detecção de afeto livre de sensores, do inglês *sensor-free affect detection* (Baker et al., 2012). No entanto, os resultados obtidos até o momento, com base na detecção de emoções livre de sensores, ainda não são confiáveis para a detecção de emoções em tempo real (Botelho, Baker, & Heffernan, 2017). Uma vez que a tomada de decisão de estratégias pedagógicas seria realizada com base no resultado dessa detecção e, caso errônea, poderia interferir negativamente no processo de aprendizagem do aluno.

O presente trabalho utiliza técnicas de mineração de dados sobre os *logs*, gerados pela interação dos alunos com um STI de matemática baseado em passos, para detectar as quatro emoções de aprendizagem que vêm sendo encontradas com mais frequência nesses tipos de ambientes (Calvo & D’Mello, 2010; D’Mello & Calvo, 2013): confusão, frustração, tédio e engajamento. Diferente das pesquisas mais recentes, este trabalho tem como objetivo verificar se os dados da personalidade dos alunos, considerando também os dados das transições de emoções, podem impactar de forma positiva no desempenho da detecção das quatro emoções consideradas. Para representar as transições das emoções dos alunos, os rótulos de emoções, utilizados para o treinamento supervisionado dos algoritmos de classificação, foram coletados por meio de um protocolo para anotação de emoções e comportamentos, baseado em análise de vídeo, que é capaz de capturar as transições das emoções dos alunos (Morais, Kautzmann, Bittencourt, & Jaques, 2019). Além disso, a coleta dos dados de *logs* foi realizada em um STI baseado em passos, que é capaz de gerar informações muito mais precisas e detalhadas sobre o aluno/sistema em um determinado momento. Nenhum dos trabalhos relacionados reportou a coleta de dados de um STI baseado em passos, seguindo a classificação de (Vanlehn, 2006).

A hipótese de pesquisa, relacionada à adição de dados da personalidade, é que a personalidade dos alunos pode proporcionar uma melhoria na precisão da detecção das emoções de aprendizagem, pois sabe-se que a personalidade pode influenciar as emoções (Rusting, 1998; Canli et al., 2001; Tong, 2010) e que a duração das emoções negativas depende da personalidade (Reis, Alvares, Jaques, & Isotani, 2018). Outra diferença está na adição de informações sobre as transições de emoções dos alunos, pois estas podem impactar positivamente na precisão da detecção de emoções. Esta hipótese foi baseada no trabalho de (D’Mello, 2012), que identificou um conjunto de transições esperadas entre as emoções dos alunos usando um STI. Dessa forma, se existe uma probabilidade maior que estando em uma emoção x o aluno passe para a emoção y , os algoritmos de aprendizagem de máquina podem considerar essa informação, que se encontra de forma implícita nos dados, para melhorar a classificação.

Este trabalho apresenta uma extensão do artigo publicado nos anais do Simpósio Brasileiro de Informática na Educação de 2019 (Morais & Jaques, 2019). Essa versão do trabalho estende a versão anterior ao utilizar dados das transições das emoções dos alunos no treinamento dos detectores de emoções. Também, foi elaborado um fluxo que apresenta a probabilidade de cada possível transição entre emoções, bem como a representatividade de cada transição em relação à

todas que foram capturadas. Ainda, foi realizada uma comparação dos rótulos de emoções com a personalidade dos alunos e com o nível de dificuldade das equações. Além disso, no presente artigo é apresentada a listagem completa das características (do inglês *features*) selecionadas automaticamente durante a fase de treinamento destes detectores. Também foi realizada uma revisão da literatura e uma atualização em relação ao estado da arte, onde foi possível identificar que a comunidade científica passou a utilizar técnicas de *deep learning* para o treinamento dos detectores de emoções. Porém, também de acordo com os trabalhos mais recentes, este trabalho apresenta o porquê esta técnica não foi considerada. Por fim, a seção de discussões e de conclusões foram expandidas, trazendo implicações das abordagens adotadas neste trabalho e apresentações de novos trabalhos futuros.

2 Estado da arte

Esta seção tem como objetivo apresentar uma visão geral do estado da arte em detecção e predição de emoções em ambientes de aprendizagem com o uso de mineração de dados. Para isso, foi realizada uma busca, que teve início com base em um conjunto de palavras-chaves, incluindo variações e combinações de *emotion*, *detect* e *learning environment*, 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 a novos trabalhos com base nos trabalhos previamente selecionados. Um total de 359 artigos foram analisados. Porém, apenas os trabalhos que visavam detectar emoções em ambientes de aprendizagem por meio de dados de *logs*, publicados mais recentemente, foram selecionados, sendo estes descritos a seguir.

O trabalho de (Pardos, Baker, San Pedro, Gowda, & Gowda, 2014) (**Artigo [1]**¹) apresenta como as emoções de tédio, engajamento, confusão e frustração dos alunos no STI ASSISTments, tutor de matemática baseado em respostas, de acordo com a classificação de (Vanlehn, 2006), podem ajudar a prever a performance dos alunos nos exames de final de ano. Para a coleta de dados, os alunos utilizaram o STI e tiveram seus *logs* de interação capturados enquanto estavam sendo analisados por codificadores humanos, seguindo o protocolo BROMP. Com base nestes dados coletados, foi desenvolvido um detector para cada emoção considerada e essa informação foi utilizada para predizer a performance dos alunos. Para o desenvolvimento dos detectores, um conjunto de algoritmos foi testado, onde o que apresentasse o melhor resultado por emoção era selecionado². Além disso, foi utilizado o algoritmo *forward selection* para seleção de características. Como conclusões os autores relatam que é possível prever efetivamente a performance dos alunos nos exames. Ainda, ressaltam que os alunos entediados ou confusos obtiveram resultados inferiores nos testes.

Também utilizando o STI ASSISTments, outro estudo foi conduzido para verificar a

¹Número utilizado como identificador do artigo na Tabela 1.

²Somente os algoritmos que obtiveram os melhores resultados são apresentados na Tabela 1.

influência de diferentes populações na precisão dos detectores de emoções (Ocumpaugh, Baker, Gowda, Heffernan, & Heffernan, 2014) (**Artigo [2]**). 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 das emoções, e a seleção dos algoritmos para o desenvolvimento dos detectores foram realizadas da mesma forma que o trabalho de (Pardos et al., 2014). A principal diferença é que o treinamento dos detectores foi dividido em modelos específicos para cada tipo de população e emoção. Por fim, os autores desenvolveram um detector com base nos dados obtidos das três populações, que apresentou um resultado quase tão efetivo quanto o detector mais customizado para cada grupo.

No trabalho de (Paquette et al., 2014) (**Artigo [3]**), os autores apresentam a detecção automática do afeto dos estudantes no ambiente virtual Inq-ITS. Inq-ITS é um STI no qual os alunos utilizam de interações simuladas para aprender conteúdos de ciências físicas, sendo esse um ambiente menos restrito do que os de resolução de problemas, permitindo mais comportamentos exploratórios. A coleta de dados, de *logs* e afeto, e o desenvolvimento dos detectores também seguiram a mesma metodologia utilizada por (Pardos et al., 2014). Novamente, quatro emoções de aprendizagem foram consideradas: tédio, engajamento, frustração e confusão. 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.

(Paquette et al., 2016) (**Artigo [4]**) realizaram um estudo comparando a detecção de emoções baseada em dados de *logs* (*sensor-free*) e na postura dos alunos, utilizando sensores específicos. Os dados analisados foram coletados do ambiente vMedic, um ambiente de aprendizagem baseado em jogo que possibilita alunos a interagir com um mundo virtual, guiando-os na administração de cuidados médicos. Para a coleta de rótulos de emoções, os autores também seguiram o protocolo BROMP. Após a análise dos resultados, os autores identificaram que os detectores baseados em *logs* obtiveram melhores resultados do que os detectores baseados em sensores. Porém, os autores ressaltam que os resultados apontam para uma continuidade na pesquisa e que ainda há muitas questões em aberto relacionadas à área.

Mais recentemente, (Jiang et al., 2018) (**Artigo [5]**) realizaram um estudo comparando a estratégia de engenharia de características com *deep learning*. O estudo foi realizado com o ambiente Betty's Brain, um ambiente computacional de aprendizagem aberto que ajuda os alunos a aprender ciências construindo mapas causais para descrever fenômenos científicos. Com base nos *logs* de interação dos alunos com o sistema e dos rótulos de emoções coletados seguindo o protocolo BROMP, os autores desenvolveram detectores para as emoções de aprendizagem engajamento, confusão, frustração, tédio e surpresa. Na engenharia de características, os autores calcularam as características manualmente, enquanto com *deep learning* os autores selecionaram os dados e tornaram classes em colunas binárias. Como resultados, a estratégia de engenharia de características apresentou melhor acurácia na detecção de todas as emoções, exceto para confusão, onde as duas estratégias alcançaram o mesmo resultado. Baseado no trabalho de (Wang, Heffernan, & Heffernan, 2015), (Botelho et al., 2017) também aplicou técnicas de *deep learning*, buscando promover uma melhora na precisão dos detectores de emoções. Porém, os autores

relatam resultados inferiores comparados ao método tradicional de seleção de características (*feature selection*). Assim, com base nos resultados relatados por (Jiang et al., 2018) e (Botelho et al., 2017), não foi considerada nenhuma abordagem de *deep learning* neste trabalho.

Todos os trabalhos relacionados tinham como objetivo realizar a detecção das emoções de aprendizagem confusão, tédio, engajamento e frustração. Além disso, os cinco trabalhos compartilham de métodos em comum, sendo eles: *i*) os rótulos de emoções foram coletados por meio do BROMP, que é um protocolo para anotação *online* das emoções dos alunos. Neste protocolo, um codificador humano analisa os alunos enquanto eles usam o ambiente computacional de aprendizagem, anotando uma única emoção de um aluno a cada 20 segundos, observando um aluno por vez até ter analisado todos os alunos da turma (Ocumpaugh, 2015). Esse método impõe um intervalo de tempo nas anotações de emoções do mesmo aluno, o que pode atingir um intervalo de minutos entre um rótulo e o próximo. *ii*) utilizam um conjunto de algoritmos, onde todos são testados para cada uma das emoções consideradas e somente o melhor algoritmo para cada emoção é selecionado. *iii*) aplicam a técnica de validação cruzada, do inglês *cross-validation*, variando entre 5 ou 10 grupos, a nível do aluno. Por fim, *iv*) aplicam as métricas de avaliação Cohen's Kappa e A'. Neste artigo, todas essas estratégias foram consideradas, com exceção da utilização do protocolo BROMP para anotação dos rótulos de emoções e da inserção dos dados da personalidade do aluno. Portanto, podem-se destacar três diferenças principais entre este trabalho e os demais relacionados: o uso de um protocolo de anotação de emoções que permite a captura da transição das emoções do aluno, a inserção de dados de personalidade e a coleta de dados em um STI de matemática baseado em passos.

A Tabela 1 apresenta uma síntese dos cinco trabalhos selecionados, destacando o número de anotações realizadas, anotadores, alunos que participaram da coleta, características consideradas no desenvolvimento dos detectores e *logs* gerados pelos alunos, que foram considerados no treinamento dos detectores³. Também é identificado se os trabalhos consideram: a transição de emoções, ou seja, se eles anotam as emoções do mesmo aluno de maneira sequencial, permitindo analisar a transição de emoções, e a personalidade do aluno durante o treinamento dos detectores de emoções. Os tipos de sistemas em que as emoções são analisadas também são identificados, sendo eles: STIs baseado em respostas (que apenas avaliam a resposta final, sem considerar a resolução), simulações, jogos e aberto (termo utilizado para traduzir do inglês *open-ended*). Além disso, os resultados obtidos em cada trabalho são apresentados, destacando os algoritmos⁴ que obtiveram os melhores resultados, baseados nas métricas Cohen's Kappa K e A'. 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 visa apresentar até onde as pesquisas têm chegado, destacando a complexidade na obtenção de um bom desempenho dos detectores, justificando a necessidade de métodos mais eficientes. As características empregadas por cada trabalho não foram listadas,

³Alguns artigos não apresentam certas informações, como número de *logs* considerados ou de anotações realizadas. Assim, quando tal informação não foi encontrada nos trabalhos, esta foi representada com “-”.

⁴Para reduzir o espaçamento, os nomes dos algoritmos foram abreviados na Tabela 1: JRip (JR), Logistic Regression (LR), J48 (J4), Naïve Bayes (NB), REP-Trees (RT), KStar (KS) e Step Regression (SR).

Tabela 1: Parâmetros e resultados dos trabalhos relacionados. Abreviações: JRip (JR), Logistic Regression (LR), J48 (J4), Naïve Bayes (NB), REP-Trees (RT), KStar (KS) e Step Regression (SR).

Artigo	[1]		[2]		[3]		[4]		[5]		
Nº de anotações	-		-		1241		755		5212		
Nº de anotadores	2		2		4		2		2		
Nº de alunos	1622		-		326		119		93		
Nº de características	172		69		127		38		249		
Nº de logs	810.000		-		-		-		146.141		
Transição de emoções	Não		Não		Não		Não		Não		
Personalidade	Não		Não		Não		Não		Não		
Tipo de sistema	Repostas		Repostas		Simulações		Jogos		Aberto		
Resultados	Valor	Alg.	Valor	Alg.	Valor	Alg.	Valor	Alg.	Valor	Alg.	
Frustração	K	0,324	NB	0,150	JR	0,389	J4	0,105	LR	0,056	LR
	A'	0,682		0,600		0,726		0,692		0,634	
Tédio	K	0,229	JR	0,240	J4	0,332	J4	0,469	LR	0,278	LR
	A'	0,632		0,660		0,723		0,848		0,682	
Confusão	K	0,274	J4	0,150	RT	0,334	JR	0,056	NB	0,091	LR
	A'	0,736		0,630		0,735		0,552		0,568	
Engajamento	K	0,358	KS	0,420	JR	0,360	SR	0,156	SR	0,142	LR
	A'	0,678		0,730		0,697		0,590		0,624	

pois seria uma lista extensa. Porém, nenhum dos trabalhos utiliza características relacionadas à transição de emoções e personalidade dos alunos, nem realizaram a coleta dos dados de *logs* de um STI baseado em passos. Após a obtenção dos resultados deste trabalho, foi realizada uma comparação destes resultados com os resultados apresentados pelos trabalhos relacionados, descrita com detalhes na Seção 4.3.

3 Trabalho Desenvolvido

Esse trabalho realiza a detecção das emoções de aprendizagem frustração, confusão, tédio e engajamento dos alunos por meio da mineração dos dados de *logs*. A detecção é realizada com base em algoritmos de aprendizagem de máquina supervisionada, seguindo uma abordagem de classificação. Embora a pesquisa relacionada à computação afetiva no Brasil, em sua maioria, ainda utiliza as emoções básicas (Morais, da Silva, Reis, Isotani, & Jaques, 2017), estas quatro emoções foram escolhidas pois diferentes pesquisas vêm mostrando que elas ocorrem com mais frequência em ambientes computacionais de aprendizagem (Calvo & D'Mello, 2010; D'Mello, 2013). Como objetivo de pesquisa, este trabalho visa verificar se os dados da personalidade do aluno podem impactar de forma positiva na detecção destas emoções de alunos utilizando um STI baseado em passos. Esta seção descreve o trabalho desenvolvido, incluindo detalhes do ambiente computacional de aprendizagem utilizado para a realização da coleta de dados, da sessão de coleta de dados realizada com alunos, do procedimento para a coleta de dados de *logs*, modelo de aluno, personalidade e emoções e, por fim, do desenvolvimento dos detectores de emoções.

Sabendo que personalidade pode influenciar as emoções diretamente (Rusting & Larsen, 1997; Rusting, 1998; Canli et al., 2001; Tong, 2010), pesquisas têm mostrado que a personalidade

influencia também na duração das emoções (Reis et al., 2018) e quais emoções são experimentadas (Nunes & Cazella, 2011). Desta forma, a hipótese de pesquisa é que inserindo características da personalidade dos alunos no desenvolvimento dos detectores de emoções, pode resultar em uma melhora nesse processo de detecção. Neste trabalho também foram considerados os dados das transições de emoções dos alunos, pois existe um conjunto de transições esperadas entre emoções durante a aprendizagem dos alunos (D’Mello & Graesser, 2012; D’Mello et al., 2014; Bosch & D’Mello, 2017; Sinclair et al., 2018; Andres et al., 2019). Assim, a transição das emoções pode levar a uma melhora na detecção de emoções baseada na mineração de dados. Portanto, as emoções dos alunos foram rotuladas em ordem sequencial, com dados refinados, com tamanho do clipe de 5 segundos. Além disso, (D’Mello & Graesser, 2012) sugere que as transições são desencadeadas por algum desequilíbrio cognitivo, que ocorre quando os alunos são confrontados com contradições ou quando não sabem como prosseguir. Os STIs baseados em passos têm uma pista mais precisa sobre quando o aluno enfrentará esse desequilíbrio cognitivo, pois eles acompanham a resolução de cada passo de uma determinada tarefa pelo aluno, identificando quando ele está errando ou acertando.

Para detectar as emoções de aprendizagem dos alunos, foram desenvolvidos 24 detectores de emoções baseados na mineração de dados de *logs*. Esses detectores foram treinados com os cinco algoritmos mais bem-sucedidos, descritos na seção de trabalhos relacionados, sendo estes algoritmos de classificação que seguem uma abordagem de aprendizagem de máquina com treinamento supervisionado. Os dados usados para treinar esses detectores foram coletados dos registros de interação dos alunos com um STI de matemática baseado em passos e foram sincronizados com rótulos de emoções, contendo informações sobre a transição das emoções e a personalidade de cada aluno.

Ao todo, foram consideradas 348 características no desenvolvimento dos detectores, calculadas para cada *log* gerado pelo aluno, que juntas têm o objetivo de representar o estado do aluno/sistema em um determinado momento. As características foram calculadas com informações de quatro fontes de dados, provenientes dos *i) logs* de interação, *ii) do modelo do aluno*, *iii) da personalidade* e *iv) das emoções e comportamentos dos alunos*. Todos estes dados foram coletados em 10 sessões, com média de 40 minutos de duração cada, de 55 alunos utilizando o STI PAT2Math, com idade entre 12 e 13 anos (Média = 12,15 e Desvio Padrão = 0,36), sendo 26 meninos e 29 meninas. Os alunos eram de duas turmas do sétimo ano de uma escola privada da região metropolitana do estado do Rio Grande do Sul.

Para verificar a validade da hipótese deste trabalho, foram desenvolvidos cinco detectores para cada emoção, sendo cada um deles com um algoritmo de treinamento diferente e treinados sem dados de personalidade. Após, apenas os quatro melhores detectores, um para cada emoção, foram selecionados para uma comparação. Esta comparação contou com mais quatro detectores que, além das características já utilizadas, foram retreinados com características da personalidade dos alunos. Assim, foi possível comparar os melhores detectores, treinados com e sem dados de personalidade.

3.1 Sistema Tutor Inteligente PAT2Math

Neste trabalho, o ambiente educacional utilizado para coleta de dados é o PAT2Math, um STI *web* que auxilia os alunos na aprendizagem de matemática (Jaques et al., 2013). Por ser um STI baseado em passos, PAT2Math é capaz de fornecer *feedback* e dicas especializadas aos alunos em cada passo de resolução de equações de primeiro grau. Ao utilizar o sistema, todas as correções de passos de resolução ou solicitações de dicas do aluno, entre outras, são armazenadas no modelo do aluno. Desta forma, 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 detectores.

D: Feedback de erro fornecido ao aluno

Você viu que tem frações nesta equação? Como você deve fazer a equivalência delas?

A: Equação Inicial

$$\frac{x}{3} + \frac{x}{2} = 15$$

B: Passos de resolução inseridas pelo aluno

$$2x + 3x = 15$$

$$6$$

$$5x = 15$$

$$6$$

C: Feedback fornecido a cada passo de resolução

E: Botão de dica e quadro com a dica fornecida pelo sistema

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.

F: Conteúdos

Pontuação total: 36621
Pontuação no nível atual: 35661

Erros gratuitos disponíveis: 9

20 de 150 pontos
Equações concluídas: 1 / 10
19 dicas gratuitas disponíveis

Dica

Figura 1: Interface gráfica do STI PAT2Math durante a resolução de uma equação.

A Figura 1 exibe uma captura de tela da interface gráfica *web* do STI PAT2Math. Na figura, foram destacadas, de forma sequencial, as etapas da trajetória que um aluno pode percorrer para solucionar uma equação de primeiro grau utilizando o PAT2Math. A partir de uma equação inicial (Figura 1.A), o aluno pode inserir passos de resolução desta equação (Figura 1.B). Para cada passo inserido pelo aluno, o sistema fornece um *feedback*, que pode ser positivo, caso o passo esteja correto, ou negativo, caso o passo esteja incorreto (Figura 1.C). No caso do aluno ter submetido

um passo incorreto, o sistema também fornece um *feedback* de erro, informando o erro cometido e dando uma dica de como o aluno pode prosseguir (Figura 1.D). Ainda, o aluno tem a possibilidade de solicitar uma dica ao sistema, que é capaz de fornecer uma ajuda especializada de acordo com o passo de resolução em que o aluno se encontra (Figura 1.E) (Seffrin et al., 2012). Por fim, ao solucionar corretamente a equação atual, o aluno pode escolher, dentre uma lista de conteúdos, qual a próxima equação será resolvida (Figura 1.F). A interface gráfica do PAT2Math também possui uma versão para inserção das equações através da escrita à mão dos alunos (Morais & Jaques, 2017). Porém, para a coleta de dados deste trabalho, todos os alunos utilizaram a versão que utiliza o teclado e *mouse* do computador para inserir as equações no sistema.

Cada passo de resolução inserido pelo aluno no PAT2Math pode disparar uma nova emoção, ou seja, uma transição de uma emoção x para uma emoção y . Isso porque em cada passo o aluno precisa avaliar se tem conhecimento para resolvê-lo, pedir ajuda, e ainda obter *feedback* positivo ou negativo sobre a corretude do seu passo, o que pode levar o estudante a experimentar diferentes emoções em um curto espaço de tempo. Para isso, destaca-se a importância da captura de todas as ações do aluno com a interface gráfica do sistema.

3.2 Coleta dos Dados de Logs

O STI PAT2Math é um sistema *web* (Morais & Jaques, 2013), assim foi desenvolvida uma ferramenta, utilizando a linguagem JavaScript, que captura todos os comandos de entrada e saída do navegador do aluno. Desta forma, toda vez que o aluno realizasse qualquer movimento com o *mouse* ou com o teclado, um registro, *log*, era gerado. Um *log* representa qualquer ação do aluno com a interface gráfica do sistema, incluindo cliques do *mouse*, teclas pressionadas etc. Além dessas informações, cada *log* contempla todo estado atual do aluno no sistema, pois informações do modelo do aluno são sincronizadas com cada *log* e armazenadas no banco de dados do PAT2Math. A coleta dos *logs* foi realizada por meio da utilização do ambiente por alunos no laboratório da escola, chamada de coleta de dados. Ao todo, os 55 alunos utilizaram o STI PAT2Math por mais de 360 horas, resultando na geração e armazenamento de um total de 386.015 *logs* de interação dos alunos com a interface do PAT2Math. Destes, foram utilizados 5.525 *logs* para o treinamento dos modelos, após a sincronização com os rótulos de emoções. Ainda, foram geradas 31 características desta fonte de dados, utilizadas no treinamento dos detectores. Todos os alunos que participaram da coleta de dados entregaram um Termo de Consentimento Livre e Esclarecido (TCLE), assinado por eles e seus responsáveis, autorizando os alunos a participarem da coleta de dados.

3.3 Coleta dos Dados do Modelo de Aluno

O modelo do aluno do STI PAT2Math armazena todas as operações realizadas pelo estudante no sistema, ou seja, o sistema armazena em sua base de dados cada equação resolvida, passo verificado, erro cometido, dica fornecida, operações aritméticas utilizadas, *misconceptions*, entre outras informações obtidas a cada passo de resolução de uma equação realizada por cada aluno

(Seffrin, Rubi, & Jaques, 2011). Foi realizada uma etapa de engenharia de características, conforme apresentado por (Jiang et al., 2018), na qual além das características descritas acima, também foram calculadas características para diferentes intervalos de tempos e também com informações da turma. Assim, para cada *log*, três intervalos de tempo foram considerados para a seleção dos dados: 1) dados dos últimos cinco segundos⁵, chamado clipe; 2) dados gerados desde o início do dia até o *log* atual, chamado diário; e, 3) dados gerados pelo aluno desde o primeiro dia de coleta até o *log* atual, chamado total. Desta forma, além das características do *log* atual, características do clipe, diário e total foram calculadas para a média do tempo gasto para resolver as equações, a média do tempo gasto em cada passo de resolução, o número de passos realizados, número de passos corretos, número de passos errados, número de equações resolvidas e efetividade dos passos⁶. Ainda, todas essas características foram recalculadas considerando os dados gerados pelos alunos de toda a turma, também considerando a seleção dos dados nos três intervalos. Dos dados do modelo do aluno, foram calculadas 298 características, utilizadas no desenvolvimento dos detectores.

3.4 Coleta dos Dados de Personalidade

Após a última sessão de coleta de dados, os alunos responderam um questionário de personalidade. Esse questionário identifica os traços de personalidade dos alunos com base na teoria dos cinco grande fatores (amabilidade, extroversão, conscienciosidade, abertura à experiência, neuroticismo), foi 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 variando de 1 (discordo totalmente) a 5 (concordo totalmente). As questões perguntavam para o aluno relatar, por exemplo, o quanto ele se considerava conversador, comunicativo. Havia quatro questões para cada um dos cinco traços de personalidade. A soma das respostas dessas quatro questões resultava no índice que representava um traço de personalidade, sendo esse índice variando de 4 a 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. Os dados de um único aluno, que não respondeu o questionário, foram descartados. Dos dados da personalidade do aluno, foram utilizadas 10 características no treinamento dos detectores, sendo o índice (valor normalizado entre 0 e 1) e a presença ou ausência (valor 0 ou 1) de cada traço.

3.5 Coleta dos Rótulos de Emoções

A detecção de emoções livre de sensores é desenvolvida com base em algoritmos de aprendizagem de máquina supervisionados, treinados com rótulos alvos. Assim, é necessário ter um conjunto de rótulos de emoções, sincronizados com os dados de *logs*, para realizar o treinamento dos detectores. Analisando os trabalhos relacionados, foi possível identificar que todos eles usam

⁵Cinco segundos é o tamanho de uma clipe para anotação das emoções do protocolo EmAP-ML.

⁶A efetividade dos passos foi calculada pela divisão entre o número de passos corretos pelo número total de passos inseridos pelo aluno em um determinado intervalo.

o protocolo BROMP para anotar as emoções e comportamentos dos alunos. No entanto, esse protocolo não é capaz de capturar as transições de emoções de um único aluno, porque a estratégia de análise dos alunos impõe um intervalo de tempo nos rótulos das emoções para o mesmo aluno. Essa lacuna pode chegar a minutos sem rotular as emoções do mesmo aluno. Uma transição de emoções é apenas considerada no protocolo BROMP se esse intervalo for menor que cinco minutos (Ocumpaugh, 2015). A anotação de emoções realizada dessa maneira não permite a identificação confiável das transições das emoções dos alunos devido a esse intervalo de 5 minutos, pois não se tem informações sobre todas as emoções que possam ter ocorridas nesse período. Além disso, a janela de anotação tem uma duração de 20 segundos, permitindo a gravação de apenas uma única emoção.

O STI PAT2Math tem sido usado por centenas de estudantes em diferentes escolas no Brasil. Assim, foi possível observar empiricamente que os alunos levam, em média, entre 5 segundos e mais de um minuto para inserir uma etapa de resolução no sistema. De acordo com (D’Mello & Graesser, 2012), qualquer desequilíbrio cognitivo pode desencadear uma transição de emoções. Ao resolver uma etapa de resolução no PAT2Math, o aluno pode enfrentar uma contradição ou pode ter dúvidas sobre como proceder. Essas situações podem levar o aluno a um desequilíbrio cognitivo, desencadeando uma transição de emoções. Desta forma, é importante que o método ou protocolo utilizado para realizar as anotações das emoções dos alunos seja capaz de capturar as emoções dentro deste intervalo.

Para a obtenção dos rótulos de emoções deste artigo, foi utilizado um protocolo capaz de capturar as transições de emoções dos alunos. Esse protocolo, chamado EmAP-ML (Morais et al., 2019), permite a identificação das emoções e dos comportamentos dos alunos, com base na análise de vídeos do rosto com áudio ambiente e da tela do computador do aluno. Possui fases de treinamento e teste para codificadores, tornando-os capazes de anotar emoções apenas quando atingem um coeficiente de concordância Cohen’s Kappa maior ou igual que 0,6. Nesse protocolo, os codificadores podem anotar as emoções e os comportamentos de um único aluno por um período prolongado, capturando as transições de emoções em janelas de cinco segundos (clipes), possibilitando também anotar mais de uma emoção por clipe, quando isto ocorrer. Após passarem nas fases de treinamento e teste, três codificadores analisaram 30 vídeos de diferentes alunos, obtidos durante as sessões de coleta de dados.

Durante esta fase de coleta de rótulos de emoções, foi gerado um total de 2099 rótulos de emoções (53,3% engajamento, 22,1% confusão, 4,3% tédio, 2,4% frustração e 18% outros/desconhecido, identificado por “?” na Figura 2)⁷ e 2059 rótulos de comportamentos (57% trabalhando na tarefa (*on task*), 14,1% realizando tarefas não relacionadas (*off task*), 11,9% no sistema sem resolver a tarefa (*on system*), 8,6% resolvendo a tarefa no caderno (*on task out*), 7,7% recebendo ajuda de um colega ou professor (*on task conversation*) e 0,6% outros/desconhecido). A partir dessa fonte de dados, foram utilizadas nove características no desenvolvimento dos detectores, sendo elas a presença ou ausência de cada uma das quatro emoções e dos cinco

⁷Por ser um STI baseado em passos e que utiliza de componentes de gamificação (O. Azevedo, de Morais, & Jaques, 2018), acredita-se que os alunos não sentem tédio e frustração com muita frequência ao utilizar o PAT2Math.

comportamentos.

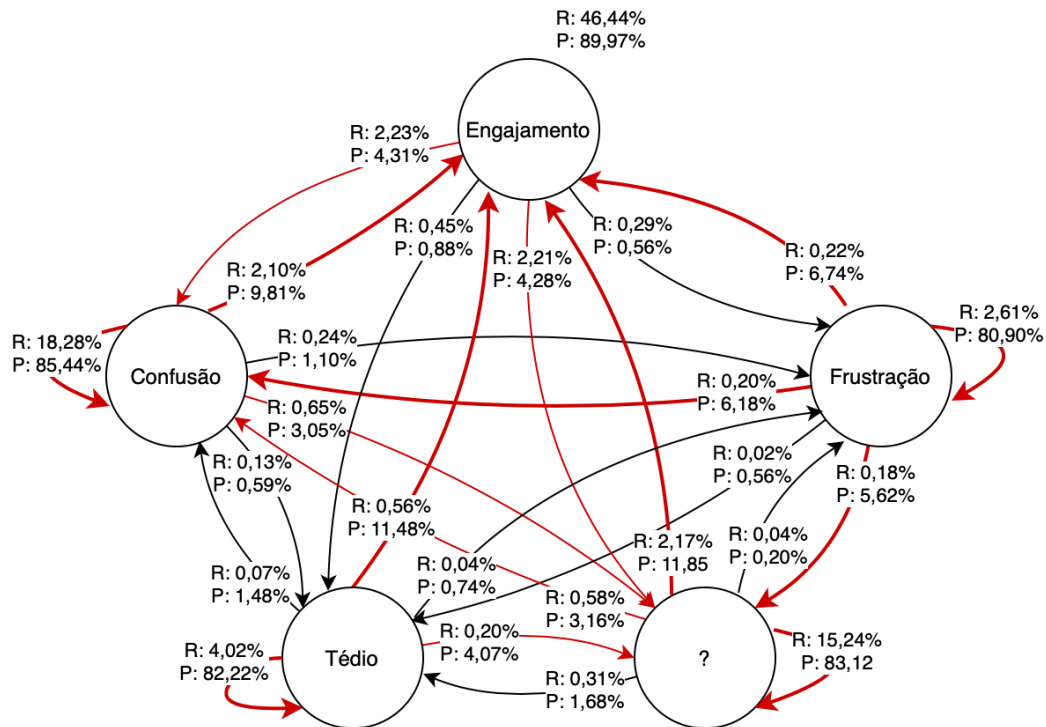


Figura 2: Fluxo de transições de emoções.

A Figura 2 apresenta o fluxo de transição entre as emoções dos alunos enquanto utilizavam o STI PAT2Math. A letra R ao lado de cada transição indica a representatividade daquela transição em relação ao número total de transições. Diferentemente, a letra P ao lado de cada transição indica a probabilidade de, estando na emoção x , ir para a emoção y . Ainda, foram destacadas em vermelho as transições que apresentam probabilidade maior do que 3% e com uma espessura maior as transição com probabilidade maior do que 5%. Para gerar este fluxo, os logs de cada aluno, já rotulados com as emoções, foram ordenados de acordo com a data e hora da sua ocorrência. Assim, foi possível percorrer essa lista de emoções, por aluno, computando cada transição de emoção de um log para outro.

Além das probabilidades e representatividade das transições das emoções, também foi calculada a distribuição das ocorrências das emoções de acordo com o nível de dificuldade das equações resolvidas pelos alunos. O nível de dificuldade de cada equação foi computado de acordo com o número de operações necessárias para resolver a equação. Assim, quanto mais operações são necessárias, mais difícil a equação se torna. De acordo com a Figura 3, é possível identificar que quanto mais difíceis as equações ficam, menos engajados os alunos se apresentam. Também, foi possível identificar que os alunos que resolveram as equações com maior grau de dificuldade apresentaram uma maior ocorrência de confusão. Essa distribuição é proveniente dos rótulos de emoções de todos os alunos dos 30 vídeos. Desta forma, não foi realizada uma análise

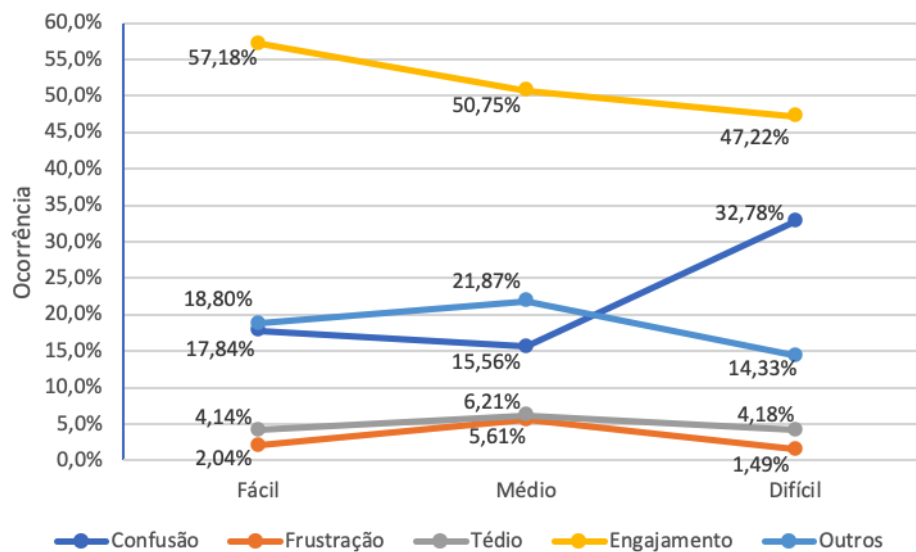


Figura 3: Ocorrência das emoções de acordo com a dificuldade das equações.

horizontal de um único aluno, visto que a coleta dos rótulos ocorreu de acordo com a seleção dos vídeos, que foi realizada de forma aleatória.

Outra informação que foi possível calcular de acordo com os rótulos de emoções coletados, foi a ocorrência das emoções de acordo com a personalidade do aluno. A Figura 4 exibe um gráfico com a representatividade das emoções por traço de personalidade do aluno. É importante ressaltar que para obter essas distribuições, apenas o traço de personalidade mais forte foi considerado. Caso o aluno apresentasse dois ou mais traços como os mais fortes, todos eram considerados. Assim, a distribuição dos 30 alunos com seus respectivos traços de personalidade mais fortes foi de amabilidade 43,3%, extroversão 33,3%, conscienciosidade 10,0%, abertura à experiência 26,7% e neuroticismo 26,7%. O somatório destes valores não fecha 100% pois um aluno pode apresentar dois ou mais traços de maior intensidade.

3.6 Desenvolvimento dos Detectores

Para o desenvolvimento dos detectores, foram selecionados cinco algoritmos de aprendizagem de máquina, que vêm sendo utilizados pela comunidade, conforme apresentado na seção de trabalhos relacionados. Esses algoritmos foram escolhidos por terem atingidos os melhores resultados na detecção das emoções de aprendizagem de acordo com os trabalhos relacionados. Os cinco algoritmos utilizados foram JRip (JR), Logistic Regression (LR), J48 (J4), Naïve Bayes (NB) e KStar (KS). Estes cinco algoritmos foram utilizados individualmente para o desenvolvimento do detector de cada emoção. Desta forma, foi possível comparar qual algoritmo obteve o melhor desempenho na detecção de cada emoção. Com os quatro melhores algoritmos, um para cada emoção, os detectores foram retreinados com dados de personalidade. As seguintes decisões

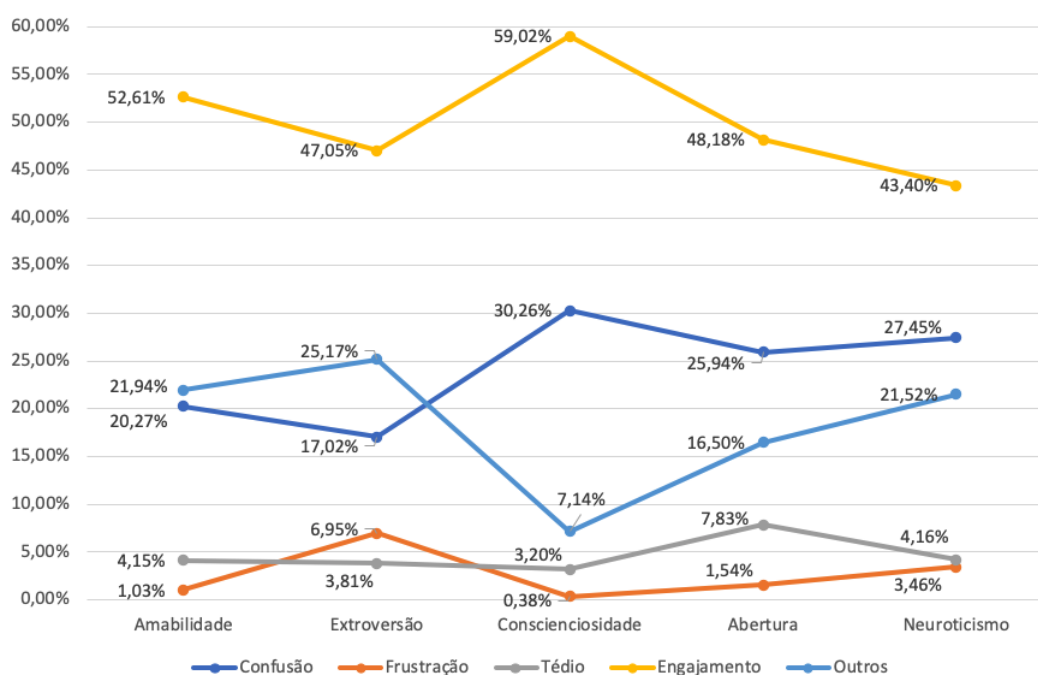


Figura 4: Ocorrência das emoções de acordo com a personalidade dos alunos.

sobre o desenvolvimento dos detectores foram tomadas com base nos trabalhos relacionados: 1) a ferramenta utilizada para o desenvolvimento dos detectores foi o RapidMiner Studio; 2) para a validação dos detectores, foi aplicada a técnica de validação cruzada com 10 grupos a nível do aluno⁸, do inglês *10 fold student-level cross-validation*; 3) as métricas de avaliação dos detectores foram realizadas com os cálculos de concordância Cohen's Kappa K e A' ⁹; 4) o algoritmo de seleção progressiva, do inglês *forward selection*, foi aplicado para selecionar automaticamente as melhores características. O objetivo desta seleção é descobrir quais as características que melhor representam o conjunto de dados, de acordo com o rótulo de treinamento. Assim, das 348 características disponíveis, apenas as 10 (no máximo) mais representativas eram selecionadas para o treinamento dos modelos¹⁰; 5) cada detector foi treinado para identificar a presença ou ausência de uma única emoção.

⁸A validação a nível do aluno (*student-level*) envolve a seleção de dados considerando todos os registros (*logs*) de um aluno, ao invés de um registro por vez, durante a fase de treinamento dos algoritmos de detecção. Assim, os detectores se tornam mais aptos a generalizar os resultados de detecção para novos alunos.

⁹A métrica Cohen's Kappa tem o objetivo de medir o quanto o detector é melhor do que a seleção ao acaso das emoções (seleção aleatória), onde 0 indica seleção ao caso e 1 indica performance perfeita. A métrica A' , por sua vez, indica a probabilidade de que o detector tenha identificado corretamente a presença ou ausência de uma emoção, por exemplo, estar confuso ou não estar confuso, onde 0,5 indica probabilidade ao acaso e 1 indica performance perfeita.

¹⁰A seleção de características não foi aplicada durante o treinamento do algoritmo KStar, pois o tempo de treinamento de uma única iteração, incluindo seleção de características, requerido por esse algoritmo foi o mesmo tempo do treinamento completo dos outros algoritmos.

Para verificar se os dados de personalidade proporcionam uma melhora na detecção das emoções, foram treinados 20 detectores sem dados de personalidade: cinco detectores para cada uma das quatro emoções, variando apenas o algoritmo de treinamento utilizado na detecção de cada uma delas. Após a análise da performance destes 20 detectores, apenas os quatro melhores foram selecionados, sendo estes os que obtiveram o melhor índice Kappa na detecção de cada emoção, e retreinados com dados de personalidade. Desta forma, foi possível realizar uma comparação entre os detectores que melhor detectaram as emoções treinados sem dados de personalidade, com os mesmos detectores treinados com dados de personalidade.

4 Resultados e Discussões

Esta seção apresenta os resultados obtidos e traz discussões de acordo com cada resultado. Assim, a seção foi dividida em subseções específicas para (i) apresentar a precisão obtida na detecção de cada emoção, considerando ou não a personalidade, e seus respectivos melhores algoritmos de classificação; (ii) listar as características (*features*) selecionadas de acordo com a seleção progressiva das características na detecção de cada emoção; e (iii) a comparação dos resultados obtidos com os trabalhos relacionados.

4.1 Precisão na Detecção das Emoções

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 uma das quatro emoções e os quatro últimos, treinados utilizando os dados de personalidade dos alunos com base nos melhores detectores encontrados para cada emoção, serviram para comparar os detectores treinados com e sem dados da personalidade dos alunos. A Tabela 2 ilustra o resultado dos 24 detectores e seus respectivos algoritmos utilizados para detectar cada emoção. Foram destacados em verde os melhores resultados obtidos no treinamento sem dados de personalidade e em azul os mesmos detectores retreinados com dados de personalidade, utilizando o valor Kappa como métrica primária.

Tabela 2: Resultados dos detectores com diferentes algoritmos e dos melhores algoritmos treinados com personalidade na detecção de cada emoção.

	J48		JRip		Logistic Regression		Naïve 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
							0,309	0,688		
Tédio	-0,002	0,470	0,025	0,542	0,000	0,452	0,040	0,608	0,005	0,560
							0,000	0,558		
Frustração	-0,004	0,535	0,042	0,533	-0,001	0,600	0,022	0,421	0,098	0,543
									0,094	0,537
Engajamento	0,630	0,844	0,617	0,821	0,592	0,847	0,564	0,799	0,104	0,530
	0,633	0,846								

Conforme apresentado na Tabela 2, o melhor algoritmo para a detecção de confusão foi o

Naïve Bayes, com $K = 0,342$ e $A' = 0,690$. Na detecção de tédio, o melhor algoritmo também foi o Naïve Bayes, com $K = 0,04$ e $A' = 0,608$. Para a detecção de frustração, o melhor algoritmo foi o KStar, mesmo sem a seleção de características, com $K = 0,098$ e $A' = 0,543$. E, na detecção de engajamento, o melhor algoritmo foi o J48, com $K = 0,630$ e $A' = 0,844$. Depois de identificar qual o melhor algoritmo para cada emoção, sem considerar os dados de personalidade, foi possível treinar novamente os detectores. Porém, desta vez as características de personalidade foram inseridas no treinamento dos detectores. Após, estes quatro detectores foram retreinados com dados de personalidade. Ao comparar o desempenho dos detectores treinados com e sem dados de personalidade, foi identificado que houve uma redução na precisão dos detectores de confusão, tédio e frustração que consideram a personalidade. Porém, houve uma melhora na precisão do detector de engajamento, onde o detector que considera a personalidade obteve $K = 0,633$ e $A' = 0,846$, comparado com $K = 0,630$ e $A' = 0,844$ do mesmo detector sem dados de personalidade.

4.2 Características Seleccionadas por Emoção

Todos os resultados de treinamento dos detectores foram armazenados para a realização de uma análise. Assim, foi possível listar quais foram as características seleccionadas para detectar cada emoção, de acordo com cada algoritmo. A seguir são listadas as características dos melhores algoritmos para cada emoção, com exceção do estado de frustração, que apresentou seu melhor resultado com o algoritmo KStar, o qual não teve seleção de características. A Tabela 3 descreve as características seleccionadas para a detecção de confusão, a Tabela 4 lista as características seleccionadas para a detecção de tédio e a Tabela 5 enumera as características seleccionadas para a detecção de engajamento. Somente a Tabela 5 apresenta as diferentes características seleccionadas pelos detectores que consideram e que não consideram a personalidade do aluno na seleção do engajamento, pois os detectores de confusão e tédio não apresentaram nenhuma alteração na seleção das características.

Ao verificar as características dos detectores treinados com dados da personalidade do aluno, foi possível identificar que não houve mudanças na seleção de características das emoções de confusão e tédio, indicando que a personalidade não influenciou negativamente nos resultados. Porém, para o engajamento, duas características foram substituídas, sendo que uma das novas características seleccionadas é relacionada à personalidade dos alunos. Assim, as próximas discussões irão descrever as características seleccionadas para confusão e tédio sem considerar os detectores treinados com dados da personalidade do aluno. Apenas a discussão sobre as características seleccionadas para a detecção do engajamento contemplam os detectores treinados com e sem os dados de personalidade do aluno. É importante ressaltar que essas foram as características resultantes da seleção progressiva de características, realizada de forma automática.

Especificamente sobre a detecção de confusão, a Tabela 3 apresenta as características consideradas como mais relevantes para a predição da confusão. A escolha destas dez características ocorreu de forma automática, por meio da execução do algoritmo de seleção progressiva de características, especificamente para a confusão e para o algoritmo Naive Bayes,

que apresentou o melhor resultado na detecção da confusão. Ao analisar as características selecionadas, é possível identificar que o comportamento do aluno, especificamente os comportamentos *off task* e *on task conversation*, impactam positivamente na detecção da confusão. Além disso, as outras oito características que mais representam a 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 (no clipe), apontando que a confusão tem uma relação com as ações mais recentes do aluno.

Tabela 3: Características selecionadas para a detecção da 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.

Sobre a detecção de tédio, ao analisar a Tabela 4, que apresenta as 10 características que melhor representam o tédio, é possível identificar que o comportamento *on system* impacta positivamente na detecção do 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. Outro fator interessante é em relação à seleção de características relacionadas aos dados dos alunos de toda a turma. Das dez características selecionadas, três estavam relacionadas com informações da turma. Esta seleção pode significar que o tédio talvez seja uma emoção mais generalizável aos indivíduos de uma mesma situação.

Para a detecção do estado de frustração, o algoritmo que apresentou o melhor desempenho foi o KStar. Conforme descrito anteriormente, não foi possível aplicar o algoritmo de seleção progressiva de características para o algoritmo KStar, pois, o tempo gasto para o treinamento de somente uma rodada deste algoritmo foi superior a sete horas, que foi o mesmo tempo necessário para executar todas as rodadas de treinamento dos demais algoritmos. Estima-se que para a

Tabela 4: Características selecionadas para a detecção 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.

aplicação do algoritmo de seleção progressiva das características no detector com algoritmo KStar seriam necessárias mais de 600 horas de processamento. Desta forma, foram consideradas as 348 características nas fases de treinamento do algoritmo KStar, sem nenhuma fase de seleção automática das características.

A Tabela 5 descreve as características que melhor representam o engajamento, de acordo com o resultado do algoritmo de seleção progressiva das características. Novamente, o comportamento do aluno está presente nas características selecionadas, 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. Essas características apontam que os dados de *logs* podem ser utilizados de forma efetiva para a detecção do engajamento. As duas características que foram substituídas, em vermelho, estão relacionadas à posição do clique na interface e ao número de cliques seguidos. Foram destacadas, em verde, as características selecionadas pela seleção automática quando os dados da personalidade do aluno foram considerados, substituindo as características em vermelho. Nestas duas novas características foi considerado o acerto na operação nos últimos cinco segundos e a identificação do lado do traço de personalidade abertura. Assim, a seleção automática das características aponta que a personalidade foi selecionada como uma das características mais relevantes para a detecção do engajamento do aluno.

Ao analisar os resultados obtidos, é possível identificar que não houve uma melhora na detecção das emoções confusão, tédio e frustração, mas houve uma melhora na detecção de engajamento ao considerar os dados da personalidade dos alunos. Mais especificamente, por meio da seleção de características, foi possível notar que a característica relacionada ao traço de personalidade abertura foi considerada uma das mais preditivas do engajamento. Esta relação tem sido encontrada também na literatura da psicologia na educação. No trabalho de (Komarraju & Karau, 2005), os autores relatam que os alunos com níveis mais altos de abertura e extroversão apresentam mais engajamento na aprendizagem, com o traço de abertura explicando a maior

Tabela 5: Características selecionadas para a detecção de engajamento no detector que considera e que não 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 (considera personalidade).
9	Identificação do lado do traço abertura, ou seja, 0 se o traço for menor que a metade ou 1 caso contrário (considera personalidade).
10	Clique no grupo resolução (não considera personalidade).
11	Número de cliques seguidos (não considera personalidade).

variação. Em outro estudo, descrito em (Douglas, Bore, & Munro, 2016), os autores apontam que os alunos com alto nível dos traços de consciência e abertura têm maior probabilidade de se engajar e usam estratégias eficazes para estudar com o mínimo apoio dos educadores. E, além disso, os alunos com baixos níveis de abertura ou consciência são menos propensos a se engajar.

Este trabalho tinha como objetivo verificar se a personalidade dos alunos poderia proporcionar uma melhora na detecção das emoções destes alunos. Embora a diferença entre os dois detectores, treinados com e sem dados de personalidade, seja pequena e tenha sido encontrada somente para o engajamento, ela traz indício de que a hipótese de pesquisa deste trabalho é verdadeira. Por meio da seleção de características 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 dez mais importantes na detecção do engajamento.

4.3 Comparação com Trabalhos Relacionados

Ao comparar os resultados obtidos neste trabalho com trabalhos relacionados, tanto considerando quanto não considerando os dados da personalidade, pode-se notar que a detecção de tédio e frustração atingiram uma baixa precisão. Nota-se também uma tendência na baixa ocorrência dessas emoções, conforme apresentado pelos trabalhos relacionados (Wang et al., 2015; Botelho et al., 2017; Botelho, Baker, Ocumpaugh, & Heffernan, 2018). No entanto, a detecção de confusão e engajamento obtiveram melhores índices de precisão do que os trabalhos relacionados, utilizando Cohen's Kappa como métrica primária. Além disso, o detector de engajamento alcançou o Kappa mínimo exigido de codificadores humanos em alguns protocolos de anotação de emoções

(Ocumpaugh, 2015; Morais et al., 2019).

Embora a transição de emoções seja um diferencial deste trabalho, não foi realizado um comparativo com detectores que não consideram essa informação, uma vez que nenhum trabalho relacionado apresenta uma coleta de dados de um STI baseado em passos. Assim, é difícil comparar os bons resultados alcançados na detecção de confusão e, principalmente, no engajamento sem trabalhos anteriores que considere o mesmo tipo de STI, mas sem os dados de transição de emoções. No entanto, acredita-se que esses resultados estão associados à riqueza nos dados de *logs* sincronizados com rótulos de emoções e dados da personalidade dos alunos obtidos a partir da interação dos alunos com STI PAT2Math baseado em passos. Depois de assistir a todos os vídeos dos alunos interagindo com o sistema, adquiridos durante a coleta de dados, foi observado empiricamente que os alunos mudam suas emoções de uma maneira muito dinâmica, e essas mudanças dependem muito da etapa de resolução dos problemas que o aluno está trabalhando. Essa observação está alinhada com o trabalho de (D’Mello & Graesser, 2012), sobre gatilhos de transições de emoções com base no desequilíbrio cognitivo.

5 Conclusões

Este trabalho teve como objetivo verificar se a utilização de dados da personalidade dos alunos no treinamento de detectores de emoções de aprendizagem pode proporcionar uma melhora na precisão da detecção das emoções frustração, confusão, tédio e engajamento. Além disso, dados das transições de emoções dos alunos foram inseridos durante o treinamento dos detectores. Foram treinados 24 detectores de emoções, com dados obtidos de 55 alunos que utilizaram o STI PAT2Math. Ao todo, foram calculadas 348 características, provenientes dos *logs* de interação, modelo do aluno, personalidade, emoções e comportamentos dos alunos. Destas, somente as dez características mais representativas de cada emoção foram selecionadas automaticamente e utilizadas no treinamento dos detectores. Ao comparar os modelos treinados com e sem personalidade, foi possível identificar que a personalidade impactou somente na escolha das características do engajamento.

De acordo com a literatura de psicologia na educação (Komarraju & Karau, 2005; Douglas et al., 2016), nota-se que o traço de personalidade abertura, o mesmo selecionado pelo detector de engajamento, é capaz de explicar parte da variância do engajamento dos alunos, sendo visto também como um possível preditor para o mesmo. Desta forma, a característica selecionada automaticamente pelo detector de engajamento, que considera os dados da personalidade, vai de encontro à literatura. Portanto, verifica-se que a personalidade dos alunos é relevante para o treinamento de classificadores das emoções de alunos utilizando STIs baseados em passos. Acredita-se que quanto mais informações do aluno os detectores possuem, mais precisa será a detecção das emoções.

Além dos dados da personalidade, também foram utilizadas informações das transições das emoções dos alunos. Embora não tenha sido realizada uma comparação específica em relação

a esta fonte de dados, acredita-se que a combinação de dados das transições das emoções, juntamente com os *logs* de um STI baseado em passos e com a personalidade dos alunos, permitiu que a detecção do engajamento alcançasse um índice de detecção maior do que o mínimo exigido de anotadores humanos, na identificação das emoções. Este pode ser considerado um passo promissor no desenvolvimento de STIs baseados em passos capazes de detectar e predizer as emoções dos alunos em tempo real livre de sensores. Possuir tal habilidade iria permitir que estratégias pedagógicas fossem adotadas antes mesmo que o aluno entrasse em estados emocionais prejudiciais à aprendizagem, como frustração e tédio.

5.1 Limitações

Conforme apresentado na seção de coleta dos rótulos de emoções, seção 3.5, a representatividade de rótulos das emoções de engajamento (53,3%) e confusão (22,1%) é superior à de rótulos das emoções de tédio (4,3%) e frustração (2,4%). Essa quantidade de rótulos está diretamente relacionada com a quantidade de *logs* utilizados durante as fases de treinamento dos detectores de cada uma dessas emoções. Portanto, acredita-se que o desbalanceamento na quantidade dos rótulos de emoções pode ter impactado na baixa precisão da detecção de tédio e frustração e, também, pode ter inserido uma ameaça à validade nos resultados da detecção de engajamento e confusão. Outras pesquisas também vêm mostrando a existência desta tendência neste tipo de ambiente de aprendizagem (Wang et al., 2015; Botelho et al., 2017, 2018), visto que existe uma interação muito grande entre o aluno e o sistema, evitando que o aluno entre em tais emoções negativas. Em (Botelho et al., 2017), os autores testaram um método de reamostragem para superar este desbalanceamento de dados nos rótulos das emoções. No entanto, os detectores treinados com os dados reamostrados alcançaram resultados inferiores do que aqueles treinados com o conjunto de dados originais. Isso acontece porque um mesmo *log* pode ser igual para diferentes emoções. Assim, não há uma maneira simples de reproduzir ou gerar *logs* de interação rotulados com as emoções de baixa ocorrência dos alunos.

Outra limitação está relacionada com a generalização dos resultados alcançados. Para treinar os detectores de emoções, foi utilizado um conjunto de dados gerados por alunos, com idade entre 12 e 13 anos, de duas turmas do sétimo ano de uma escola privada da região metropolitana do RS. Conforme apresentado por (Ocumpaugh et al., 2014), não é possível generalizar os resultados de detectores de emoções em ambientes educacionais para diferentes populações sem uma coleta de dados que considere amostras das mais diversas populações. Portanto, não é possível generalizar os resultados obtidos para diferentes alunos. Além disso, estes dados foram gerados enquanto os alunos interagem com um STI baseado em passos, resolvendo equações de primeiro grau. Acredita-se que os *logs* gerados nestes tipos de ambiente possuem informações mais precisas sobre o aluno, uma vez que esses ambientes são capazes de auxiliar os alunos na resolução de tarefas passo a passo. Desta forma, acredita-se que existe uma forte dependência dos resultados com esse tipo de ambiente de aprendizagem.

5.2 Trabalhos Futuros

Em relação à utilização de dados de transição das emoções, não foi realizada uma comparação dos resultados de detectores que consideram e que não consideram a transição das emoções durante as fases de treinamento. Como não havia outro estudo coletando dados de um STI baseado em passos, nem levando em consideração a transição de emoções para detectar as emoções de aprendizagem dos alunos, essa pesquisa pode ser vista como pioneira nestes quesitos. Assim, como trabalho futuro, planeja-se gerar rótulos de emoções usando um intervalo de tempo específico entre eles, a fim de levar em consideração apenas a riqueza nos dados de *logs* de interação gerados em um STI baseado em passos.

Também foram analisadas pesquisas relacionadas que vêm aplicando abordagens de *deep learning* para treinar os detectores de emoção (Botelho et al., 2017; Jiang et al., 2018). Essa abordagem ainda não trouxe nenhuma melhora na precisão da detecção de emoções. No entanto, acredita-se que a aplicação de algoritmos como o *Long Short-Term Memory* (LSTM) poderia levar em consideração as transições de emoções de forma mais adequada, pois é um algoritmo capaz de considerar sequências de dados, que neste caso seriam as sequências ou transições de emoções, levando a uma melhoria na precisão da detecção. Desta forma, o objetivo será estudar essas duas abordagens em trabalhos futuros. Ainda, a detecção dos comportamentos e da personalidade dos alunos foi realizada de forma manual. Assim, a automatização destas tarefas também são vistas como trabalhos futuros.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, da FAPERGS (Processo 17/2551-0001203-8) e do CNPq (processo 309218/2017-9).

References

- Andres, J. M. A. L., Ocumpaugh, J., Baker, R. S., Slater, S., Paquette, L., Jiang, Y., ... others (2019). Affect sequences and learning in betty's brain. In *International learning analytics and knowledge conference* (pp. 383–390). doi: <http://dx.doi.org/10.1145/3303772.3303807> [GS Search]
- Arroyo, I., Woolf, B. P., Burelson, W., Muldner, K., Rai, D., & Tai, M. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal of Artificial Intelligence in Education*, 24(4), 387–426. doi: <http://dx.doi.org/10.1007/s40593-014-0023-y> [GS Search]
- Azevedo, O., de Moraes, F., & Jaques, P. A. (2018). Exploring gamification to prevent gaming

- the system and help refusal in tutoring systems. In *European conference on technology enhanced learning* (pp. 231–244). doi: http://dx.doi.org/10.1007/978-3-319-98572-5_18 [GS Search]
- Azevedo, R., & Aleven, V. (2013). *International handbook of metacognition and learning technologies* (Vol. 26). Springer. [GS Search]
- Baker, R. S., Gowda, S. M., Wixon, M., Kalka, J., Wagner, A. Z., Salvi, A., ... Rossi, L. (2012). Towards sensor-free affect detection in cognitive tutor algebra. *International Conference on Educational Data Mining*. [GS Search]
- Barbosa, A. A. G. (2009). *Modelo hierárquico de fobias infanto-juvenis: testagem e relação com os estilos maternos*. Unpublished doctoral dissertation, PhD thesis, Psicologia Social (UFRN), João Pessoa.
- Bosch, N., & D’Mello, S. (2017). The affective experience of novice computer programmers. *International Journal of Artificial Intelligence in Education*, 27(1), 181–206. doi: <http://dx.doi.org/10.1007/s40593-015-0069-5> [GS Search]
- Botelho, A. F., Baker, R. S., & Heffernan, N. T. (2017). Improving sensor-free affect detection using deep learning. In *International conference on artificial intelligence in education* (pp. 40–51). doi: http://dx.doi.org/10.1007/978-3-319-61425-0_4 [GS Search]
- Botelho, A. F., Baker, R. S., Ocumpaugh, J., & Heffernan, N. T. (2018). Studying affect dynamics and chronometry using sensor-free detectors. In *International conference on educational data mining*. ERIC. [GS Search]
- Calvo, R. A., & D’Mello, S. K. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1(1), 18–37. doi: <http://dx.doi.org/10.1109/T-AFFC.2010.1> [GS Search]
- Canli, T., Zhao, Z., Desmond, J. E., Kang, E., Gross, J., & Gabrieli, J. D. (2001). An fmri study of personality influences on brain reactivity to emotional stimuli. *Behavioral neuroscience*, 115(1), 33. doi: <http://dx.doi.org/10.1037/0735-7044.115.1.33> [GS Search]
- D’Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology*, 105(4), 1082. doi: <http://dx.doi.org/10.1037/a0032674> [GS Search]
- D’Mello, S., & Calvo, R. A. (2013). Beyond the basic emotions: what should affective computing compute? In *Chi’13 extended abstracts on human factors in computing systems* (pp. 2287–2294). doi: <http://dx.doi.org/10.1145/2468356.2468751> [GS Search]
- Douglas, H. E., Bore, M., & Munro, D. (2016). Coping with university education: The relationships of time management behaviour and work engagement with the five factor model aspects. *Learning and individual Differences*, 45, 268–274. doi: <http://dx.doi.org/10.1016/j.lindif.2015.12.004> [GS Search]
- D’Mello, S. (2012). Monitoring affective trajectories during complex learning. In *Encyclopedia of the sciences of learning* (pp. 2325–2328). Springer. [GS Search]
- D’Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, 22(2), 145–157. doi: <http://dx.doi.org/10.1016/j.learninstruc.2011.10.001> [GS Search]

- D’Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, 29, 153 - 170. doi: <http://dx.doi.org/10.1016/j.learninstruc.2012.05.003> [GS Search]
- Fredrickson, B. L. (1998). What good are positive emotions? *Review of General Psychology*, 2(3), 300-319. doi: <http://dx.doi.org/10.1037/1089-2680.2.3.300> [GS Search]
- Graesser, A., & D’Mello, S. K. (2011). Theoretical perspectives on affect and deep learning. In *New perspectives on affect and learning technologies* (pp. 11–21). Springer. doi: http://dx.doi.org/10.1007/978-1-4419-9625-1_2 [GS Search]
- Harley, J. M. (2016). Measuring emotions: a survey of cutting edge methodologies used in computer-based learning environment research. In *Emotions, technology, design, and learning* (pp. 89–114). Elsevier. doi: <http://dx.doi.org/10.1016/B978-0-12-801856-9.00005-0> [GS Search]
- Jaques, P. A., & Nunes, M. A. S. N. (2019). Computação afetiva aplicada à educação. *Informática na Educação: técnicas e tecnologias computacionais*, 3. Retrieved from <https://ieducacao.ceie-br.org//computacaoafetiva> [GS Search]
- Jaques, P. A., Seffrin, H., Rubi, G., Morais, F., Ghilardi, C., Bittencourt, I. I., & Isotani, S. (2013). Rule-based expert systems to support step-by-step guidance in algebraic problem solving: The case of the tutor PAT2math. *Expert Systems with Applications*, 40(14), 5456–5465. doi: <http://dx.doi.org/10.1016/j.eswa.2013.04.004> [GS Search]
- Jiang, Y., Bosch, N., Baker, R. S., Paquette, L., Ocumpaugh, J., Andres, J. M. A. L., ... Biswas, G. (2018). Expert feature-engineering vs. deep neural networks: which is better for sensor-free affect detection? In *International conference on artificial intelligence in education* (pp. 198–211). doi: http://dx.doi.org/10.1007/978-3-319-93843-1_15 [GS Search]
- Komarraju, M., & Karau, S. J. (2005). The relationship between the big five personality traits and academic motivation. *Personality and individual differences*, 39(3), 557–567. doi: <http://dx.doi.org/10.1016/j.paid.2005.02.013> [GS Search]
- Morais, F., da Silva, J., Reis, H., Isotani, S., & Jaques, P. (2017). Computação afetiva aplicada à educação: uma revisão sistemática das pesquisas publicadas no brasil. In *Brazilian symposium on computers in education (simpósio brasileiro de informática na educação-sbie)* (Vol. 28, p. 163). doi: <http://dx.doi.org/10.5753/cbie.sbie.2017.163> [GS Search]
- Morais, F., & Jaques, P. (2013). Avaliação de usabilidade do sistema tutor inteligente PAT2Math. *RENTE-Revista Novas Tecnologias na Educação*, 11(3). doi: <http://dx.doi.org/10.22456/1679-1916.44929> [GS Search]
- Morais, F., & Jaques, P. A. (2017). PAT2Math + Handwriting: Evoluindo Sistemas Tutores de Matemática com reconhecimento da escrita à mão. In *Simpósio brasileiro de informática na educação (sbie)* (Vol. 28). doi: <http://dx.doi.org/10.5753/cbie.sbie.2017.1237> [GS Search]
- Morais, F., & Jaques, P. A. (2019). Predição de emoções baseada em mineração de dados: considerando a personalidade para melhorar a detecção. In *Simpósio brasileiro de informática na educação (sbie)* (Vol. 30). doi: <http://dx.doi.org/10.5753/cbie.sbie.2019.1521> [GS Search]
- Morais, F., Kautzmann, T. R., Bittencourt, I. I., & Jaques, P. A. (2019). Emap-ml: A

- protocol of emotions and behaviors annotation for machine learning labels. In *European conference on technology enhanced learning* (pp. 495–509). Netherlands: Springer. doi: http://dx.doi.org/10.1007/978-3-030-29736-7_37 [GS Search]
- Nunes, M., & Cazella, S. (2011). O que sua personalidade revela? fidelizando clientes web através de sistemas de recomendação e traços de personalidade. *WebMedia, 1*. [GS Search]
- Ocuppaugh, J. (2015). Baker rodrigo ocuppaugh 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*. [GS Search]
- Ocuppaugh, J., Baker, R., Gowda, S., Heffernan, N., & Heffernan, C. (2014). Population validity for educational data mining models: A case study in affect detection. *British Journal of Educational Technology, 45*(3), 487–501. doi: <http://dx.doi.org/10.1111/bjjet.12156> [GS Search]
- Paquette, L., Baker, R. S., Sao Pedro, M. A., Gobert, J. D., Rossi, L., Nakama, A., & Kauffman-Rogoff, Z. (2014). Sensor-free affect detection for a simulation-based science inquiry learning environment. In *International conference on intelligent tutoring systems* (pp. 1–10). doi: http://dx.doi.org/10.1007/978-3-319-07221-0_1 [GS Search]
- Paquette, L., Rowe, J., Baker, R., Mott, B., Lester, J., DeFalco, J., ... Georgoulas, V. (2016). Sensor-free or sensor-full: A comparison of data modalities in multi-channel affect detection. In *International conference on educational data mining*. ERIC. [GS Search]
- Pardos, Z. A., Baker, R. S., San Pedro, M. O., Gowda, S. M., & Gowda, S. M. (2014). Affective states and state tests: Investigating how affect and engagement during the school year predict end-of-year learning outcomes. *Journal of Learning Analytics, 1*(1), 107–128. [GS Search]
- Pekrun, R. (2014). Emotions and learning. *Educational practices series, 24*. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.691.9950&rep=rep1&type=pdf> [GS Search]
- Reis, H., Alvares, D., Jaques, P., & Isotani, S. (2018). Analysis of permanence time in emotional states: A case study using educational software. In *International conference on intelligent tutoring systems* (pp. 180–190). doi: http://dx.doi.org/10.1007/978-3-319-91464-0_18 [GS Search]
- Rusting, C. L. (1998). Personality, mood, and cognitive processing of emotional information: three conceptual frameworks. *Psychological bulletin, 124*(2), 165. doi: <http://dx.doi.org/10.1037/0033-2909.124.2.165> [GS Search]
- Rusting, C. L., & Larsen, R. J. (1997). Extraversion, neuroticism, and susceptibility to positive and negative affect: A test of two theoretical models. *Personality and Individual Differences, 22*(5), 607 - 612. doi: [http://dx.doi.org/10.1016/S0191-8869\(96\)00246-2](http://dx.doi.org/10.1016/S0191-8869(96)00246-2) [GS Search]
- Salmeron-Majadas, S., Santos, O. C., & Boticario, J. G. (2014). An evaluation of mouse and keyboard interaction indicators towards non-intrusive and low cost affective modeling in an educational context. *Procedia Computer Science, 35*, 691–700. doi: <http://dx.doi.org/10.1016/j.procs.2014.08.151> [GS Search]
- Scherer, K. R. (2005). What are emotions? and how can they be measured? *Social Science*

- Information*, 44(4), 695–729. doi: <http://dx.doi.org/10.1177/0539018405058216> [GS Search]
- Seffrin, H., Rubi, G., Ghilardi, C., Morais, F., Jaques, P., Isotani, S., & Bittencourt, I. I. (2012). Dicas inteligentes no sistema tutor inteligente pat2math. In *Brazilian symposium on computers in education (simpósio brasileiro de informática na educação-sbie)* (Vol. 23). doi: <http://dx.doi.org/10.5753/cbie.sbie.2012.%25p> [GS Search]
- Seffrin, H., Rubi, G., & Jaques, P. (2011). O modelo cognitivo do sistema tutor inteligente pat2math. In *Sbie* (Vol. 1). doi: <http://dx.doi.org/10.5753/cbie.sbie.2011.%25p> [GS Search]
- Shanabrook, D. H., Arroyo, I., & Woolf, B. P. (2012). Using touch as a predictor of effort: what the ipad can tell us about user affective state. In *International conference on user modeling, adaptation, and personalization* (pp. 322–327). doi: http://dx.doi.org/10.1007/978-3-642-31454-4_29 [GS Search]
- Sinclair, J., Jang, E. E., Azevedo, R., Lau, C., Taub, M., & Mudrick, N. V. (2018). Changes in emotion and their relationship with learning gains in the context of metatutor. In *International conference on intelligent tutoring systems* (pp. 202–211). doi: http://dx.doi.org/10.1007/978-3-319-91464-0_20 [GS Search]
- Tong, E. M. (2010). Personality influences in appraisal–emotion relationships: The role of neuroticism. *Journal of Personality*, 78(2), 393–417. doi: <http://dx.doi.org/10.1111/j.1467-6494.2010.00620.x> [GS Search]
- Vanlehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, 16(3), 227–265. [GS Search]
- Wang, Y., Heffernan, N. T., & Heffernan, C. (2015). Towards better affect detectors: effect of missing skills, class features and common wrong answers. In *International learning analytics and knowledge conference* (pp. 31–35). doi: <http://dx.doi.org/10.1145/2723576.2723618> [GS Search]