

Predição de emoções baseada em mineração de dados: considerando a personalidade para melhorar a detecção

Felipe de Moraes¹, Patrícia A. Jaques¹

¹Programa de Pós Graduação em Computação Aplicada (PPGCA)
Universidade do Vale do Rio dos Sinos (UNISINOS), São Leopoldo – RS – Brasil

`felipmorais@edu.unisinos.br, pjaques@unisinos.br`

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 impact the precision of the detection positively. Also, the emotion labels were obtained through an annotation protocol developed by our group, which allows the capture of transitions of the students' emotions. As results, it was possible to identify that only the engagement detector, trained with personality data of the students, obtained a small improvement in the precision of 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.*

Resumo. *Este trabalho utiliza a mineração de dados, provenientes da interação dos alunos com um STI baseado em passos, para detectar quatro emoções de aprendizagem: confusão, engajamento, frustração e tédio. Diferente de outros trabalhos, este tem como objetivo verificar se os dados da personalidade dos alunos pode impactar de forma positiva na precisão da detecção. Além disso, os rótulos de emoções foram obtidos por meio de um novo protocolo de anotação, que permite a captura das transições das emoções dos alunos. Como resultado, foi possível identificar que apenas o detector de engajamento, treinado com dados da personalidade dos alunos, obteve uma pequena melhora na precisão da detecção. Porém, com a utilização de um algoritmo de seleção de características, foi possível verificar que entre 348 características disponíveis, apenas dez foram selecionadas, incluindo dados da personalidade.*

1. Introdução

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 and Alevén 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]. Por outro lado, 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 and D’Mello 2011]. Desta forma, é importante que ambientes educacionais, baseados em computador, sejam capazes de detectar e adaptar suas estratégias pedagógicas de acordo com as emoções dos alunos [Arroyo and 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 and Nunes 2019]. Porém, alguns desses métodos utilizam de equipamentos específicos para realizar a detecção, como câmeras ou sensores, adicionando um custo e uma intromissão em aplicações no mundo real [Calvo and D’Mello 2010]. Em ambientes de aprendizagem, por exemplo, a necessidade de câmeras ou sensores para a utilização desses métodos pode inviabilizar a utilização desses sistemas em larga escala e por um longo período [Baker and 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. Por exemplo, os Sistemas Tutores Inteligentes (STI), um tipo de ambiente educacional, geram muitos *logs* de interação, pois são sistemas computacionais especialistas em domínios específicos, que utilizam técnicas de Inteligência Artificial (IA) para proporcionar ensino e assistência individualizados ao aluno [Corbett et al. 1997]. Mais especificamente, os STIs baseados em passos são capazes de auxiliar o aluno a cada passo de resolução de uma determinada tarefa, fornecendo *feedbacks*, dicas etc. Por isso, ambientes deste tipo geram muitos dados de interação entre o aluno e o sistema.

Pesquisas vêm mostrando que é possível detectar as emoções dos alunos em ambientes de aprendizagem por meio da mineração de dados [Paquette and et al. 2016]. O presente trabalho tem como objetivo geral verificar se através da mineração dos dados de *logs*, gerados pela interação dos alunos com um STI de matemática baseado em passos, é possível detectar e prever as quatro emoções de aprendizagem que vêm sendo encontradas com mais frequência nesses tipos de ambientes [Calvo and D’Mello 2010]: confusão, frustração, tédio e engajamento. Diferente das pesquisas mais recentes, este trabalho visa verificar se os dados da personalidade dos alunos podem impactar de forma positiva no desempenho da detecção. Além disso, os rótulos de emoções, utilizados para o treinamento supervisionado dos detectores, foram coletados por meio de um novo protocolo para anotação de emoções e comportamentos, baseado em análise de vídeo, que é capaz de capturar as transições das emoções dos alunos [de Moraes and et al. 2019].

A hipótese de pesquisa deste trabalho é que a personalidade dos alunos pode proporcionar uma melhora na precisão da detecção das emoções de aprendizagem, pois sabe-se que a personalidade pode direcionar as emoções de usuários em sistemas *web* [Nunes and Cazella 2011], influenciar o comprometimento de funcionários [Panaccio and 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 considera e outro que não considera a personalidade dos alunos durante a fase de treinamento dos detectores.

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, iniciada por 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. Um total de 327 artigos foram analisados. Porém, apenas os trabalhos que visavam detectar emoções em ambientes de aprendizagem por meio de dados de *logs*, publicados nos últimos cinco anos, foram selecionados, sendo estes descritos a seguir.

O trabalho de [Pardos and et al. 2014] (**Artigo [1]**)¹ apresenta como as emoções de tédio, engajamento, confusão e frustração dos alunos no STI ASSISTments, tutor de matemática baseado em passos, podem ajudar a prever a performance dos alunos. Segundo os autores, foi desenvolvido um detector para cada emoção considerada e essa informação foi utilizada para prever 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 se diferentes populações poderiam influenciar os detectores de emoções [Ocumpaugh and et al. 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 and 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, 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 and et al. 2014] (**Artigo [3]**), os autores apresentam a detecção automática do afeto dos estudantes no ambiente virtual Inq-ITS. A coleta de dados, de *logs* e afeto, e o desenvolvimento dos detectores, também seguiram a mesma metodologia utilizada por [Pardos and 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 and 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. 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.

¹Número utilizado como identificador do artigo na Tabela 1.

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

Mais recentemente, [Jiang and et al. 2018] (**Artigo [5]**) realizaram um estudo comparando engenharia de características e *deep learning*. Com base nos *logs* de interação dos alunos com o sistema, 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 foi melhor na detecção de todas as emoções, exceto para confusão, onde as duas estratégias alcançaram o mesmo resultado.

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 de aprendizagem, anotando uma única emoção de um aluno a cada 20 segundos. *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 *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'. A Tabela 1 apresenta os parâmetros e os resultados obtidos nos cinco trabalhos relacionados.

Tabela 1. Parâmetros e resultados dos trabalhos relacionados.

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					
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	

A Tabela 1 apresenta uma síntese dos cinco trabalhos, destacando o número de anotações realizadas, anotadores, alunos que participaram da coleta, características consideradas no desenvolvimento dos modelos e *logs* gerados pelos alunos considerados no treinamento dos modelos³. Além disso, os resultados 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

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).

empregadas por cada trabalho não foram listadas, pois seria uma lista extensa. Porém, nenhum dos trabalhos utiliza características relacionadas à personalidade dos alunos.

3. Trabalho Desenvolvido

Esse trabalho tem como objetivo detectar as emoções de aprendizagem frustração, confusão, tédio e engajamento dos alunos por meio da mineração dos dados de *logs*. Essas emoções foram escolhidas pois várias pesquisas vêm mostrando que elas ocorrem mais frequentemente em ambientes educacionais [Calvo and D’Mello 2010]. O ambiente educacional utilizado como caso de uso neste trabalho é o PAT2Math, um STI *web* para auxiliar os alunos no ensino de matemática. Por ser um STI baseado em passos, PAT2Math é capaz de proporcionar *feedbacks* e dicas especializadas aos alunos em cada passo de resolução de equações de primeiro grau. Todas as interações do aluno com o sistema são armazenados 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. O diferencial deste trabalho está na forma de captura das emoções para geração dos rótulos alvos e na adição de características de personalidade no desenvolvimento dos detectores. Sabendo que a personalidade influencia a duração das emoções [Reis et al. 2018] e quais emoções são experimentadas [Nunes and Cazella 2011], 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.

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 55 alunos utilizando o PAT2Math. Os alunos são de duas turmas do sétimo ano de uma escola privada do sul do Brasil. 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. Coleta dos Dados de Logs

PAT2Math é um sistema *web*, assim foi desenvolvido 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. Em seguida, 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 PAT2Math 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, apenas 5.525 *logs* foram utilizados 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 seus responsáveis, autorizando os alunos a utilizarem a ferramenta.

3.2. Coleta dos Dados do Modelo de Aluno

O modelo do aluno do 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 para cada aluno. Foi realizada uma etapa de engenharia de características, conforme apresentado por [Jiang and et al. 2018]. 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 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 o 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 de toda a turma, também considerando a seleção dos dados nos três intervalos. Dos dados do modelo do aluno, foram utilizadas 298 características no desenvolvimento dos detectores.

3.3. 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 grandes fatores, 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 seguem uma escala Likert com valores variando de 1 (discordo totalmente) à 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 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. 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 entre 0 e 1) e a presença ou ausência (valor 0 ou 1) de cada traço.

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.

3.4. Coleta dos Rótulos de Emoções

Os detectores automáticos de emoções utilizam algoritmos de aprendizagem de máquina e são treinados com base em rótulos alvos, chamado de treinamento supervisionado. Assim, é necessário que se tenha um conjunto de rótulos de emoções, sincronizados aos dados, para realizar o treinamento dos detectores. Ao analisar os trabalhos relacionados, todos utilizam o protocolo BROMP para anotação das emoções e comportamentos. Porém, este protocolo não é capaz de capturar as transições das emoções de um único aluno. Ainda, a janela de anotação possui uma duração de 20 segundos, permitindo a anotação de uma única emoção. Desta forma, foi desenvolvido o protocolo EmAP-ML [de Moraes and et al. 2019] para a anotação das emoções e comportamentos dos estudantes, com base em vídeos do rosto com áudio ambiente e da tela do computador do aluno. Esse protocolo possui fases de treinamento e teste dos codificadores, tornando-os aptos a anotar emoções somente quando atingirem um coeficiente Cohen's Kappa maior ou igual à 0,6. Neste protocolo, os codificadores podem anotar as emoções e comportamentos de um único aluno por vez, capturando as transições de emoções em janelas (clipes) de cinco segundos, sendo ainda possível a anotação de mais de uma emoção por clipe. Após treinados, três codificadores analisaram 30 vídeos de alunos diferentes, obtidos durante as sessões de coleta de dados. Na fase de geração dos rótulos do protocolo, um total de 2099 rótulos de emoções e 2059 rótulos de comportamentos foram gerados. Desta fonte de dados, foram utilizadas 9 características no desenvolvimento dos detectores, sendo a presença ou ausência de cada uma das quatro emoções e de cinco comportamentos.

3.5. Desenvolvimento dos Detectores

Para o desenvolvimento dos detectores, foram selecionados cinco algoritmos de aprendizado 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. Os cinco algoritmos utilizados foram JRip (JR), Logistic Regression (LR), J48 (J4), Naïve Bayes (NB) e KStar (KS). Os 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 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 aluno, do inglês *10 fold student-level cross-validation*; 3) a principal métrica de avaliação dos detectores é pelo cálculo de concordância Cohen's Kappa K , seguida por A' ; 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 mais representativas eram selecionadas para o treinamento dos modelos⁷; 5) cada detector foi treinado para identificar a presença ou ausência de uma única emoção.

⁷A 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. Após a análise da performance destes 20 modelos, 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 modelos que melhor detectaram as emoções treinados sem dados de personalidade, com os mesmos modelos treinados com dados de personalidade.

4. Resultados e Discussões

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 foram treinados utilizando os dados de personalidade dos alunos, com base nos melhores detectores encontrados para cada emoção. A Tabela 2 ilustra o resultado dos 24 detectores e seus respectivos algoritmos utilizados para cada emoção. Foram destacados em verde os melhores resultados obtidos no treinamento sem dados de personalidade e em azul os mesmos modelos 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 $K = 0,342$. Na detecção de tédio, o melhor algoritmo também foi o Naïve Bayes, com $K = 0,04$. Para a detecção de frustração, o melhor algoritmo foi o KStar, mesmo sem a seleção de características, com $K = 0,098$. E, na detecção de engajamento, o melhor algoritmo foi o J48, com $K = 0,630$. Após, estes quatro detectores foram retreinados com dados de personalidade. Ao analisar 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 $K = 0,633$.

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 selecionadas era relacionada a personalidade dos alunos. É importante ressaltar que essas foram as características resultantes da seleção progressiva de características, aplicada na detecção do engajamento.

Ao analisar os resultados obtidos, é possível verificar que não houve uma melhora na detecção das emoções confusão, tédio e frustração, mas houve uma melhora na detecção de engajamento. 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.

5. Conclusões

Este trabalho teve como objetivo verificar se a utilização de dados de personalidade no treinamento de detectores de emoções de aprendizagem pode proporcionar uma melhora na precisão da detecção das emoções frustração, confusão, tédio e engajamento. As estratégias adotadas para o desenvolvimento destes detectores foram as mesmas empregadas pelos trabalhos relacionados, com exceção da inserção dos dados de personalidade e da forma de anotação das emoções, capaz de detectar as transições de emoções de um único aluno. Uma coleta dos dados de *logs*, modelo do aluno, personalidade e emoções foi realizada com 55 alunos que utilizaram o PAT2Math durante 10 sessões. Deste estudo, foram utilizados para o desenvolvimento dos detectores 5.525 *logs*, cada um contendo 348 características. Além disso, 30 vídeos foram analisados e anotados seguindo o protocolo EmAP-ML, gerando um total de 2099 rótulos de emoções e 2059 rótulos de comportamentos. Os cinco melhores algoritmos apresentados nos trabalhos relacionados foram utilizados para o desenvolvimento do detector de cada emoção. Dos 20 detectores desenvolvidos, apenas o melhor detector de cada emoção foi selecionado para ser retreinado com dados de personalidade, permitindo a comparação dos detectores que consideram com os detectores que não consideram a personalidade dos alunos.

Ao comparar os resultados dos detectores treinados com e sem os dados de personalidade dos alunos, apenas o detector de engajamento apresentou uma melhora na precisão da detecção. Embora essa melhora tenha sido muito pequena, com uma diferença de $K = 0,003$, das 348 características disponíveis, apenas 10 foram selecionadas automaticamente com um algoritmo de seleção de características, incluindo uma característica da personalidade do aluno, isto é um indício de que a hipótese de pesquisa deste trabalho é verdadeira. Como limitações, a detecção dos comportamentos e da personalidade dos alunos foi realizada de forma manual. Assim, a automatização destas tarefas são vistas como trabalhos futuros. Ainda, está previsto uma replicação dos experimentos realizados, considerando um número maior de participantes e diferentes anotadores, visando verificar se os resultados encontrados são generalizáveis para diferentes perfis de aluno.

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).

Referências

- Arroyo, I. and et al. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *IJAIED*, 24(4):387–426.
- Azevedo, R. and Aleven, V. (2013). *International handbook of metacognition and learning technologies*, volume 26. Springer.
- Baker, R. and et al. (2012). Towards Sensor-free automated detection of affect in a Cognitive Tutor for Algebra. In *5th Int. Conf. on Educational Data Mining*, pages 126–133.
- Barbosa, A. A. G. (2009). *Modelo hierárquico de fobias infanto-juvenis: testagem e relação com os estilos maternos*. PhD thesis, Psicologia Social (UFRN).
- Calvo, R. A. and D’Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Trans. on Affec. Comput.*, 1(1):18–37.
- Corbett, A., Koedinger, K., and Anderson, J. (1997). Intelligent tutoring systems. *Handbook of human computer interaction*, pages 849–874.
- de Moraes, F. and et al. (2019). EmAP-ML: A Protocol of Emotions and Behaviors Annotation for Machine Learning Labels. In *EC-TEL*, Netherlands. Springer.
- Fredrickson, B. L. (1998). What good are positive emotions? *Review of Gen. Psyc.*, 2(3).
- Graesser, A. and D’Mello, S. K. (2011). Theoretical perspectives on affect and deep learning. In *New perspectives on affect and learning technologies*, pages 11–21. Springer.
- Jaques, P. A. and Nunes, M. A. S. N. (2019). Computação afetiva aplicada à educação. *Informática na Educação: técnicas e tecnologias computacionais*, 3.
- Jiang, Y. and et al. (2018). Expert feature-engineering vs. deep neural networks: Which is better for sensor-free affect detection? In *AIED*, pages 198–211. Springer.
- Nunes, M. and Cazella, S. (2011). O que sua personalidade revela? fidelizando clientes web através de sistemas de recomendação e traços de personalidade. *WebMedia*, 1.
- Ocuppaugh, J. and et al. (2014). Population validity for educational data mining models: A case study in affect detection. *BJET*, 45(3):487–501.
- Panaccio, A. and Vandenberghe, C. (2012). Five-factor model of personality and organizational commitment. *Journal of vocational behavior*, 80(3):647–658.
- Paquette, L. and et al. (2014). Sensor-free affect detection for a simulation-based science inquiry learning environment. In *ITS*, pages 1–10. Springer.
- Paquette, L. and et al. (2016). Sensor-free or sensor-full: A comparison of data modalities in multi-channel affect detection. In *Int. Conf. on Educational Data Mining*. ERIC.
- Pardos, Z. and et al. (2014). Affective states and state tests: investigating how affect and engagement during the school year predict end-of-year learning outcomes. *JLA*, 1(1).
- Pekrun, R. (2014). Emotions and learning. In *Educational practices series*. IEA, IBE.
- Reis, H., Alvares, D., Jaques, P., and Isotani, S. (2018). Analysis of permanence time in emotional states: A case study using educational software. In *ITS*, pages 180–190.
- Scherer, K. R. (2005). What are emotions? and how can they be measured? *Social Science Information*, 44(4):695–729.