15º Congresso de Inovação, Ciência e Tecnologia do IFSP - 2024

Método de Reconhecimento Automático das Emoções Acadêmicas de Estudantes

JOÃO VICTOR S. PEREIRA [1], NELSON NASCIMENTO JUNIOR [2]

[1] IFSP, Câmpus Cubatão, pereira.joao2@aluno.ifsp.edu.br
[2] IFSP, Câmpus Cubatão, nelsonjr@ifsp.edu.br

Área de conhecimento (Tabela CNPq): 1.03.01.01-1 Computabilidade e Modelos de Computação

RESUMO:

O artigo aborda o desenvolvimento de modelos computacionais para detectar e reconhecer emoções acadêmicas, como engajamento, tédio, frustração e confusão, em ambientes virtuais de aprendizado. O estudo utiliza a base de dados Daisee, composta por vídeos das expressões faciais de estudantes, para treinar modelos de aprendizado profundo que identificam essas emoções. A metodologia incluiu a extração e normalização de imagens, balanceamento dos dados e treinamento com Redes Neurais Convolucionais (CNN). Os resultados indicam que os modelos que empregaram técnicas de balanceamento por pesos tiveram um desempenho superior em comparação aos modelos com balanceamento por aumento de dados, que sofreram de *overfitting*.

PALAVRAS-CHAVE: dinâmica das emoções; aprendizado de máquina; aprendizado profundo.

Automatic Recognition Method for Students' Academic Emotions

ABSTRACT: The article discusses the development of computational models to detect and recognize academic emotions, such as engagement, boredom, frustration, and confusion, in virtual learning environments. The study utilizes the Daisee dataset, composed of videos of students' facial expressions, to train deep learning models that identify these emotions. The methodology included image extraction and normalization, data balancing, and training with Convolutional Neural Networks (CNNs). The results indicate that models employing weight balancing techniques performed better compared to models using data augmentation, which suffered from overfitting.

KEYWORDS: emotion dynamics; academic emotions; machine learning; deep learning

INTRODUÇÃO

A partir dos conhecimentos descobertos por Paul Ekman (1994) sobre as emoções básicas, como raiva, nojo, medo, felicidade, tristeza e o estado neutro, consideradas emoções universais, entende-se que qualquer pessoa, independentemente de conceitos culturais, apresenta essas emoções da mesma forma. A comunidade científica começou a estudar outras emoções esboçadas por um ser humano no seu dia a dia; assim, iniciaram-se os estudos sobre o que ficaram conhecidas como emoções acadêmicas, ou seja, as emoções que as pessoas sentem enquanto estão em um ambiente de aprendizado.

As emoções acadêmicas, que são o engajamento, o tédio, a frustração e a confusão, mostraram-se fatores importantes relacionado ao desempenho de estudantes em sala de aula. Sendo assim, um estudante engajado tende a se manter mais emocionalmente conectado com o ambiente de aprendizado

do que se ele estivesse entediado, resultando em maior foco, na maioria das vezes (D'Mello e Calvo, 2013).

Foram realizadas pesquisas ao redor do mundo, e foi constatado que, diferente das emoções básicas, as emoções acadêmicas não são universais, ou seja, são consequências de contextos sociais, do ambiente e da experiência pessoal dos estudantes (Pekrun *et al*; 2007). Logo, uma pessoa no hemisfério sul do planeta tende a demonstrar-se engajada de uma forma diferente de uma pessoa do hemisfério norte, ou até mesmo pessoas que frequentam o mesmo ambiente acadêmico podem se sentir engajados com tarefas diferentes, implicando em uma demonstração de engajamento divergente.

Atualmente, com os avanços no entendimento sobre as emoções acadêmicas, pode-se constatar que, por mais contrastantes que elas sejam entre si, tendem a se interconectar, criando o que é conhecido como dinâmica das emoções. Isso ocorre porque as emoções são temporais. A temporalidade das emoções significa que um estudante, quando inserido em um contexto educacional, tende a transitar entre o engajamento, o tédio, a frustração e a confusão em um curto período de tempo, sendo que, em um minuto, o mesmo pode experienciar todas as quatro emoções acadêmicas, por exemplo (F. Morais et al., 2023).

Nossa pesquisa busca compreender e interpretar de diferentes formas a dinâmica das emoções acadêmicas e como elas se correlacionam. Sendo assim, serão desenvolvidos quatro modelos de aprendizado profundo capazes de reconhecer automaticamente o quanto um estudante está engajado, frustrado, confuso e entediado em um ambiente virtual de aprendizado.

Os conceitos de dinâmica das emoções são a base da Métrica L. A Métrica L é uma abordagem quantitativa usada para avaliar a transição de emoções acadêmicas em um conjunto de dados. Ela considera a intensidade, a frequência e a direção das mudanças emocionais, como de engajamento para tédio ou de confusão para frustração. Para calcular essa métrica, utiliza-se uma matriz de transição que registra a frequência com que essas mudanças ocorrem (F. Morais et al., 2023). Isso ajuda a entender como as emoções dos estudantes variam durante as aulas, permitindo avaliar a eficácia das aulas, desenvolver intervenções para minimizar emoções negativas e personalizar o ensino para melhorar a experiência de aprendizagem.

MATERIAL E MÉTODOS

Primeiramente, a fim de criar os modelos de aprendizado profundo, foi selecionada uma base de dados eficiente. O dataset utilizado foi o Daisee, que serviu para o modelo de predição das emoções, no qual foi aplicado aprendizado profundo. O conjunto de dados Daisee, desenvolvido por Gupta et al. (2018), compreende imagens que capturam emoções acadêmicas, incluindo engajamento, tédio, confusão e frustração. Este conjunto é composto por 9.068 vídeos, cada um com 10 segundos de duração, envolvendo a participação de 112 indivíduos (32 mulheres e 80 homens). A rotulagem das emoções neste conjunto foi realizada considerando a intensidade de cada uma delas, variando de 0 a 3. O valor "0" indica um baixo nível de intensidade na emoção específica, enquanto o valor "3" representa um grau elevado daquela emoção. Esses rótulos foram atribuídos aos participantes com base na intensidade percebida de suas expressões emocionais acadêmicas.

A fim de possibilitar os treinamentos, realizou-se a extração dos dados do Daisee. Cada vídeo da base foi extraído a uma taxa de 5 quadros por segundo (os vídeos, por padrão, possuem 30 quadros por segundo). Cada um desses frames foi submetido a algoritmos de extração da face, utilizando a biblioteca MediaPipe do Python. Após o rastreamento da face, as imagens foram recortadas a 64x64 pixels e transformadas em preto e branco. Cada um desses frames, depois de serem normalizados, foram armazenados em um banco de dados (foi usada a biblioteca TinyDB do Python para criá-lo) junto às informações de nível de engajamento, tédio, confusão e frustração.

emoção	quantidade
Muito engajado	177.001
Engajado	182.319
Desengajado	14.439
Muito desengajado	1.521
total:	375.280

emoção	quantidade
Muito entediado	9.235
Entediado	72.120
Pouco entediado	120.237
Pouquíssimo entediado	173.688
total:	375.280

emoção	quantidade
Muito confuso	4.331
Confuso	29.815
Pouco confuso	87.356
Pouquíssimo confuso	253.778
total:	375.280

emoção	quantidade
Muito frustrado	3.264
Frustrado	12.940
Pouco frustrado	66.235
Pouquíssimo frustrado	292.841
total:	375.280

TABELAS 1, 2 3 e 4. Quantidade de frames por classe.

As tabelas acima demonstram a quantidade de frames original, representando cada classe na base de dados do Daisee.

O Daisee, por padrão, é uma base de dados desbalanceada, sendo assim, foi realizada uma tentativa de balanceamento dos dados. Por conta dos vídeos do *dataset* estarem rotulados entre níveis de cada uma das emoções, foi preciso realizar um tipo de balanceamento diferente para cada treinamento, pois, por exemplo, no caso do engajamento, a quantidade de pessoas "Muito desengajadas" (0) é menor do que a quantidade de pessoas "Muito engajadas" (3), mas ao analisar o nível de confusão, foi possível constatar que uma pessoa "Muito confusa" (3) é minoria, enquanto há muito mais exemplos de pessoas "Pouquíssimo confusas" (0). Entretanto, uma pessoa que está "Muito desengajada" (0) não necessariamente está "Muito confusa" (3), mesmo ambos sendo exemplos de minorias na base de dados.

O Daisee possui as pastas de treinamento, teste e validação; cada uma delas possui 5.484, 1.864 e 1.720 vídeos, respectivamente. Primeiramente, foi feito com que a pasta de treinamento assumisse o papel de teste e validação, logo os dados seriam posteriormente divididos na fase de treinamento. Os dados da pasta de treinamento estão apresentados abaixo, baseados em cada uma das emoções.

A fim de diminuir a disparidade entre os níveis de cada emoção, foram selecionados os vídeos das pastas de teste e validação em que os estudantes apresentavam níveis de determinadas emoções que representavam menos de um quinto (⅕) da pasta de treinamento. Por exemplo, no caso do engajamento, foram selecionados os níveis “Desengajado” (1) e “Muito desengajado” (0). Esses vídeos foram

normalizados e adicionados ao banco de dados. Esses novos frames extraídos dos vídeos das pastas de teste e validação representaram um aumento de cerca de 80% na quantidade de dados.

Outro tipo de estratégia de balanceamento da base de dados usada foi o *Data augmentation*. *Data augmentation* é uma técnica usada para aumentar a quantidade e a variedade de dados disponíveis para treinamento de modelos de aprendizado de máquina, sem precisar coletar novos dados. Isso é feito aplicando transformações e alterações aos dados existentes para gerar novas amostras. Essas transformações podem incluir rotações, translações, escalonamentos, alterações de cor e outras manipulações que preservam as características relevantes dos dados originais. Cada um dos frames relacionados às menores classes passou por um processo de *Data augmentation* e também foi adicionado ao banco de dados.

Além do balanceamento por aumento de dados, foi proposto o balanceamento por pesos. O balanceamento por pesos é uma técnica usada para lidar com classes desbalanceadas em um conjunto de dados de treinamento. Em problemas de classificação, quando algumas classes têm muitos exemplos e outras têm poucos, os modelos tendem a ser enviesados em favor das classes mais frequentes. Sendo assim, no balanceamento, cada classe no conjunto de dados recebe um peso. Classes menos frequentes recebem pesos maiores, enquanto classes mais frequentes recebem pesos menores. Esses pesos são usados para ajustar a contribuição de cada classe no cálculo da função de perda. Quando o modelo calcula a perda (ou erro) durante o treinamento, ele multiplica a perda para cada exemplo pelo peso da classe a que pertence.

Após os tipos de balanceamento, inicia-se o processo de treinamento dos modelos, no qual cada um dos conjuntos de frames foi submetido a um algoritmo de Rede Neural Convolucional (CNN). Uma CNN é um tipo de rede neural artificial especializada no processamento de dados estruturados, como imagens, vídeos e áudios. Foram realizados dois tipos de treinamento para cada uma das emoções acadêmicas: o primeiro utilizando a técnica de balanceamento por aumento dos dados e o segundo usando o balanceamento por pesos.

RESULTADOS E DISCUSSÃO

Este trabalho envolveu o desenvolvimento de modelos computacionais para detectar e reconhecer emoções durante situações de aprendizagem, como engajamento, confusão, frustração e tédio. Para classificar essas emoções não básicas em situações de aprendizagem, foi usada uma base de dados com vídeos das expressões faciais de estudantes assistindo a conteúdos educacionais, as características das faces das pessoas foram extraídas e correlacionadas por Redes Neurais Convolucionais (CNN's). Ao total foram treinados oito modelos usando CNN's, com os dados a 48x48 pixels.

- Treinamentos relativos ao engajamento
    - Treinamento usando balanceamento por aumento de dados
        - Resultado: *Overfitting*
    - Treinamento usando balanceamento por pesos
        - Resultado: 77% de acurácia

- Treinamentos relativos ao tédio
    - Treinamento usando balanceamento por aumento de dados
        - Resultado: 73% de acurácia
    - Treinamento usando balanceamento por pesos
        - Resultado: 81% de acurácia

- Treinamentos relativos ao Confusão
    - Treinamento usando balanceamento por aumento de dados

        - Resultado: *Overfitting*
    - Treinamento usando balanceamento por pesos
        - Resultado: 73% de acurácia

- Treinamentos relativos ao Frustração
    - Treinamento usando balanceamento por aumento de dados
        - Resultado: *Overfitting*
    - Treinamento usando balanceamento por pesos
        - Resultado: 74% de acurácia

Em geral, os melhores resultados foram dos modelos que usaram a técnica de balanceamento por pesos, visto que dos quatro treinamentos que usaram o aumento de dados resultaram em overfitting. O overfitting é um problema no aprendizado de máquina que ocorre quando um modelo se ajusta de forma desregulada ao conjunto de dados em que está sendo treinado, como se o modelo tivesse "decorado" o que foi aprendido, mas não consegue se adaptar a outros casos, causando um vício nas respostas.

Os modelos que obtiveram sucesso em seu treinamento foram testados e selecionados, por fim foram utilizados apenas os modelos com o balanceamento por pesos. Eles foram testados em dois vídeos de um estudante enquanto em processo de aprendizagem. O primeiro vídeo mostra a face de um estudante assistindo aulas de matemática, e no segundo, o mesmo realiza uma atividade do tema que foi estudado. Em ambos os vídeos, o engajamento foi a emoção que apareceu mais vezes em maior intensidade, enquanto o tédio foi a segunda em número de aparições. A confusão apareceu em alta intensidade apenas no vídeo em que o estudante estava realizando o exercício, mas em momento algum, em ambos os vídeos, o estudante se mostrou frustrado, de acordo com os resultados dos modelos.

Fonte: Próprio autor
FIGURA 1. Exemplo de saída mostrando o resultado dos modelos.

Conforme podemos observar, a FIGURA 1 demonstra os resultados dos modelos a partir de um frame do vídeo em que o estudante está assistindo uma aula, na imagem mostra o nível de engajamento (3), tédio (1), confusão (1) e frustração (1).

CONCLUSÕES

Este trabalho focou no desenvolvimento de modelos computacionais para a detecção e reconhecimento das emoções acadêmicas (engajamento, confusão, frustração e tédio) em ambientes virtuais de aprendizado. Ademais, mesmo que as emoções acadêmicas estejam relacionadas a contextos culturais e da experiência pessoal dos estudantes, mostrou-se possível criar ferramentas para identificar o nível de cada emoção por características da face.

Os resultados mostraram que os modelos que usaram a técnica de balanceamento por pesos tiveram um desempenho superior. Enquanto os treinamentos baseados no aumento de dados sofreram de *overfitting*, evidenciado pela incapacidade dos modelos de generalizar para novos dados, os modelos ajustados com pesos atingiram uma acurácia consistente: 77% para engajamento, 81% para tédio, 73% para confusão e 74% para frustração. O balanceamento por pesos demonstrou ser mais eficaz ao evitar o ajuste excessivo aos dados de treinamento e ao melhorar a capacidade de generalização dos modelos.

Embora o trabalho tenha avançado na classificação das emoções acadêmicas, há espaço para melhorias. A ampliação da base de dados e a inclusão de faces culturalmente diversas poderiam aprimorar a generalização e a precisão dos modelos. Futuras pesquisas devem focar na coleta de dados adicionais para balancear a base de dados e na exploração de técnicas para otimizar ainda mais o desempenho dos modelos, atendendo melhor às necessidades e contextos variados dos ambientes de aprendizado.

CONTRIBUIÇÕES DOS AUTORES

Todos os autores contribuíram com o texto e a revisão do trabalho e aprovaram a versão submetida.

REFERÊNCIAS

D'MELLO, Sidney; CALVO, Rafael A. Beyond the basic emotions: what should affective computing compute?. In: **CHI'13 extended abstracts on human factors in computing systems**. 2013. p. 2287-2294.

D'MELLO, Sidney; GRAESSER, Art. Dynamics of affective states during complex learning. Learning and Instruction, v. 22, n. 2, p. 145-157, 2012.

EKMAN, Paul. All emotions are basic. **The nature of emotion: Fundamental questions**, p. 15-19, 1994.

MORAIS, Felipe de; JAQUES, Patricia A. The dynamics of Brazilian students' emotions in digital learning systems. **International Journal of Artificial Intelligence in Education**, v. 34, n. 2, p. 519-544, 2024.

GUPTA, Abhay et al. Daisee: Towards user engagement recognition in the wild. **arXiv preprint arXiv:1609.01885**, 2016.

PEKRUN, R. The control-value theory of achievement emotions: An integrative approach to emotions in education. **Emotion in education/Academic Press**, 2007.