

1.03.03 - Ciência da Computação / Metodologia e Técnicas da Computação

DESENVOLVIMENTO DE UM MODELO PARA RECONHECIMENTO DE EMOÇÕES NUM SISTEMA TUTOR INTELIGENTE

Felipe Zago Canal^{1*}, Eliane Pozzebon²,

1. Discente de Engenharia de Computação da Universidade Federal de Santa Catarina (UFSC)

2. Professora da UFSC - Departamento de Computação/Orientadora

Resumo

Para acompanhar a demanda evolutiva do Sistema Tutor Inteligente (STI) Mazk, é necessário não somente o aprimoramento das ferramentas que o compõem, mas também o desenvolvimento de novos recursos para o sistema. O movimento de cada músculo facial faz com que as características da face produzam configurações que podem ser interpretadas como uma emoção. Pela análise destas características, é possível obter o estado emocional da pessoa e, assim, ajustar um sistema para que apresente comportamento apropriado. Para isso, este trabalho tem o objetivo de desenvolver uma rede neural artificial que seja capaz de classificar a expressão facial dos estudantes durante o processo de ensino-aprendizagem. Esta informação pode ser utilizada pelo STI para aprimorar a tomada de decisão, adicionando o estado emocional ao modelo do aluno. O modelo desenvolvido neste projeto foi capaz de alcançar uma acurácia média de 98% para as expressões neutra, raiva, nojo, medo, felicidade, tristeza e surpresa.

Palavras-chave: Educação; Expressões Faciais; Redes Neurais Convolucionais.

Apoio financeiro: Bolsa PIBIC/CNPQ.

Trabalho selecionado para a JNIC: Universidade Federal de Santa Catarina (UFSC), Propeq.

Introdução

Nos últimos anos, com o evidente avanço e difusão das tecnologias de informação, têm-se cada vez mais espaço para o emprego de ferramentas digitais de auxílio para a educação, como os Sistemas Tutores Inteligentes (STIs) (HWANG, 2003). Esses sistemas caracterizam-se como ambientes virtuais de aprendizagem que incorporam modelos computacionais de áreas como ciências cognitivas, ciências da educação, inteligência artificial, entre outras (GRAESSER; CONLEY; OLNEY, 2012). Desta forma, STIs podem também fazer uso de alguma técnica de Inteligência Artificial (IA) com o objetivo de otimizar o processo de ensino-aprendizagem, e personalizar a interação humano-computador (GALAFASSI et al., 2020).

O reconhecimento de expressões pela face é uma tarefa relativamente simples e realizada de maneira natural por humanos, contudo ainda é uma área em aberto no meio computacional (BISWAS; SIL, 2015). Apesar de não se ter acesso a um método com efetividade comparada à humana para realização de tal tarefa, são vários os algoritmos e ferramentas aptos a categorizar emoções a partir de imagens faciais e obter alto índice de corretude. Alguns métodos de reconhecimento de expressão pela face são aplicados a STIs a certo tempo, todavia, devido ao seu custo computacional, são usualmente bem simplificados e superficiais, baseando-se em poucas métricas para realização das classificações (AMMAR et al., 2010; LIN et al., 2012; BALDASSARRI et al., 2015; WU; LIU; WANG, 2008).

Neste trabalho é proposta uma solução computacional para uso de reconhecimento de emoções em expressões faciais, com enfoque no STI MAZK. Tendo em vista as necessidades do sistema, bem como as limitações do ambiente de execução e os métodos disponíveis para extração das emoções, um modelo de Rede Neural Convolucional (CNN) é proposto e aplicado para o reconhecimento de emoções a partir de imagens no STI. O treinamento do modelo em questão é realizado de maneira supervisionada e com a inserção de dois datasets na rede: CK+ e MUG. As métricas de validação do modelo apontam uma precisão média quantitativa de 98% para a detecção das seguintes emoções: neutro, raiva, felicidade, nojo, medo, tristeza e surpresa.

Metodologia

Este projeto objetiva o reconhecimento de emoções por meio de expressões faciais e sua relação com o desempenho do aluno num STI. Para análise das expressões faciais foram utilizadas redes neurais artificiais, mais especificamente modelo CNN. Os resultados obtidos nesta captura foram incluídos no STI MAZK, desenvolvido no Laboratório de Tecnologias Computacionais (LabTeC) da UFSC, Campus Araranguá.

O processo de desenvolvimento do modelo de CNN iniciou-se com a obtenção das imagens de dois datasets: Extended Cohn-Kanade Dataset (CK+) e MUG Facial Expression Database (MUG). Levando em consideração as expressões convergentes dentre as bases, foi extraído um total de 19065 amostras. Após a seleção das imagens, um processo de recorte da região de interesse (ROI) foi aplicado a cada amostra. Nessa

mesma etapa, um redimensionamento foi aplicado nas imagens para que se pudesse padronizar o conjunto de dados. As imagens foram convertidas para uma resolução de 180x180 aplicando o método de interpolação por área.

Para evitar o reconhecimento de padrões sem relevância para o reconhecimento de expressões, uma etapa adicional de extração de landmarks foi aplicada. Com esse processo, foram identificados 68 pontos relevantes da face. Por fim, um processo de espelhamento foi aplicado nas imagens para ampliar o conjunto de treino, obtendo um total de 38130 imagens, sendo esse montante dividido entre treino, validação e teste com 26691, 5720 e 5720 amostras para cada etapa, respectivamente. A separação das imagens se deu de forma aleatória, onde os subconjuntos de cada expressão foram embaralhados e divididos conforme as proporções apresentadas acima.

Após a separação do dataset, o modelo desenvolvido foi estruturado com base nas principais camadas em CNNs: convolução, pooling e camada densa. Todavia, algumas camadas adicionais foram integradas ao modelo para aprimorar a capacidade de classificação, como normalização, linearização e dropout.

A etapa de treinamento da CNN foi executada pelo período de 1,5 horas, até que alcançou-se um ponto de convergência na medida de acurácia da rede, após 150 épocas de treinamento. Após o treinamento e validação do modelo, um micro serviço foi desenvolvido para receber as informações do STI MAZK, processar as imagem e retornar a emoção detectada.

Dentro do MAZK, ao ingressar em uma sala onde é efetuado o reconhecimento de expressões, é solicitado ao aluno a liberação do uso de sua câmera para dar início ao processo. A partir disso o sistema imediatamente inicia a captura de imagens da sua face de forma temporal e as submete ao micro serviço citado acima. O micro serviço, ao receber uma requisição, submete essa imagem a todas as etapas de pré-processamento e, posteriormente, avalia a expressão aplicando-a no modelo treinado. A partir disso, o resultado da classificação é retornado à aplicação principal para que essa informação seja repassada ao tutor humano que acompanha a atividade de forma síncrona.

Resultados e Discussão

Ao final da etapa de treinamento o modelo proposto foi capaz de classificar corretamente $\approx 98\%$ das amostras a ele apresentadas, ou seja, a acurácia alcançada pela rede neural convolucional foi de 0,9801. No entanto, a acurácia obtida pelo modelo muitas vezes não representa o alcance do objetivo geral de sua aplicação e, por conta disso, é importante a utilização de outras métricas. Sendo assim, as métricas de precision e recall foram aplicadas ao modelo, onde precision representa as predições feitas pelo modelo de forma correta e recall, por outro lado, representa a fração de eventos verdadeiros detectados. Com a obtenção de um modelo ideal, busca-se alcançar valores próximos a 1 para ambos precision e recall e, para podermos avaliar de forma mais justa a correlação entre esses dois métodos, a métrica f1-score é implementada da seguinte forma:

$$F1 = \frac{2*(precision*recall)}{(precision+recall)}$$

Assim, caso os valores de precision e recall sejam muito diferentes um do outro, o valor dessa métrica será baixo e, em um caso ideal, alcançar-se-á F1-score = 1.

A aplicação desta métrica na CNN produzida obteve um F1-score médio de 0,98 sendo o valor mais baixo 0,95 e o mais alto 0,99 para as expressões neutra e de tristeza, respectivamente.

Para que fosse possível aplicar o modelo de classificação no STI MAZK, foi preciso garantir que os recursos computacionais necessários para sua aplicação não excedesse o poder computacional disponível no ambiente de produção. Para isso, foram verificadas as utilizações dos principais recursos durante a execução da ferramenta: CPU e Memória RAM. Através do acompanhamento do servidor e de observações dos índices de utilização de memória, pode-se perceber que o uso de RAM pelo micro serviço aplicado é constante. Isso ocorre pelo fato de que o serviço funciona a partir do carregamento do modelo na RAM, e disso em diante, não necessita de recurso adicional referente a esse parâmetro. Contudo, a utilização de memória, durante os testes preliminares, atingiu um valor de $\approx 7\%$ do total disponível no servidor, o que corresponde a aproximadamente 420 MB. No momento do teste aplicado, com a adição desse valor ao uso total de memória, o servidor chegou a atingir uma utilização de $\approx 62,5\%$ do total de RAM disponível, o que leva a acreditar que tal recurso não é um problema para a aplicação do modelo de classificação.

Por outro lado, na utilização de recursos de processador (CPU) pode-se perceber um pico durante a avaliação de uma amostra, onde o algoritmo atinge $\approx 25\%$ de utilização de um núcleo de processamento do servidor. Contudo, pode-se também observar que o uso do recurso citado não inviabiliza a aplicação do modelo pois, mesmo utilizando uma quantidade considerável do poder de processamento disponível, a aplicação principal ainda possui a sua disposição cerca de 87% de CPU, sendo que sua utilização média sequer atinge 1% do total durante seu uso normal.

Conclusões

Em termos de métricas de desempenho, um resultado bem expressivo foi obtido para a classificação das seguintes expressões: neutro, raiva, nojo, medo, felicidade, tristeza e surpresa, atingindo uma média ponderada dentre as expressões de ≈98%.

As possibilidades de utilização do modelo construído são inúmeras e se mantém em aberto, tendo em vista que ele foi implantado como um micro serviço no servidor do STI, podendo ser utilizado de qualquer parte do sistema e até mesmo servir como uma ferramenta isolada ao sistema MAZK. Entretanto, algumas aplicações para o próprio tutor podem ser citadas. Uma delas é a integração da classificação de expressões ao modelo do aluno, conceito de inteligência artificial aplicado no MAZK. Essa aplicação possibilita à ferramenta a interpretação dos dados para utilização como base na tomada de decisão, aperfeiçoando o loop interno do tutor, por exemplo, através de um feedback visual, deixando a interação com o estudante mais robusta.

Além disso, o algoritmo deve ser testado com mais indivíduos e diferentes perfis de usuários dentro das ferramentas do MAZK, podendo ser integrado a outras técnicas de IA para obtenção de métricas relevantes para o processo de ensino-aprendizagem. De toda forma, uma análise qualitativa do método proposto deve ser conduzida em um trabalho futuro, mediante sua aplicação em situações reais de utilização do MAZK. Nessa aplicação poderá ser analisada a aceitação da funcionalidade por parte dos professores e alunos por meio de feedbacks solicitados aos mesmos.

Em resumo, o modelo apresentado neste trabalho foi capaz de alcançar elevados índices de desempenho, tanto em sua precisão, quanto no custo computacional envolvido na sua execução. Ao ser incorporado no modelo do aluno, o método desenvolvido pode fornecer base de informações ao STI, promovendo uma melhor tomada de decisão durante a interação do aluno com o tutor e, sendo assim, pode agregar valor à ferramenta MAZK por meio da sua contribuição nesse processo.

Referências bibliográficas

HWANG, Gwo-Jen. A conceptual map model for developing intelligent tutoring systems. *Computers & Education*, Elsevier, v. 40, n. 3, p. 217–235, 2003.

GRAESSER, Arthur C; CONLEY, Mark W; OLNEY, Andrew. Intelligent tutoring systems. *APA educational psychology handbook, Vol 3: Application to learning and teaching.*, American Psychological Association, p. 451–473, 2012.

GALAFASSI, Cristiano et al. EvoLogic: Sistema Tutor Inteligente para Ensino de Lógica. In: SBC. ANAIS do XLVII Seminário Integrado de Software e Hardware. [S.l.: s.n.], 2020. P. 222–233.

BISWAS, Suparna; SIL, Jaya. An efficient expression recognition method using contourlet transform. In: PROCEEDINGS of the 2nd International Conference on Perception and Machine Intelligence. [S.l.: s.n.], 2015. P. 167–174.

AMMAR, Mohamed Ben et al. The affective tutoring system. *Expert Systems with Applications*, Elsevier, v. 37, n. 4, p. 3013–3023, 2010.

LIN, Hao-Chiang Koong et al. Employing Textual and Facial Emotion Recognition to Design an Affective Tutoring System. *Turkish Online Journal of Educational Technology-TOJET*, ERIC, v. 11, n. 4, p. 418–426, 2012.

BALDASSARRI, Sandra et al. Affective-aware tutoring platform for interactive digital television. *Multimedia Tools and Applications*, Springer, v. 74, n. 9, p. 3183–3206, 2015.

WU, Yanwen; LIU, Wei; WANG, Jianbo. Application of emotional recognition in intelligent tutoring system. In: IEEE. FIRST International Workshop on Knowledge Discovery and Data Mining (WKDD 2008). [S.l.: s.n.], 2008. P. 449–452.