

## INTERAÇÃO HUMANO-COMPUTADOR: FUNDAMENTOS E APLICAÇÕES

JEAN CARLO HILGER<sup>1</sup>, JUNIOR VITOR RAMISCH<sup>2</sup>, FERNANDO BEVILACQUA<sup>3</sup>

### 1 INTRODUÇÃO

As emoções e sentimentos de um usuário submetendo-se a uma sessão de jogos digitais são informações de profunda utilidade. Tendo posse delas, indústria e pesquisadores podem projetar jogos destinados a produzir emoções específicas, otimizando o impacto na comunidade (BEVILACQUA, 2018a). Isto posto, fica evidente que detectar e monitorar o estado emocional de um jogador pode contribuir significativamente para o sucesso do jogo em questão.

No entanto, muitas das ferramentas empregadas na execução da tarefa supracitada podem ser consideradas intrusivas (BEVILACQUA, 2018a). Questionários aplicados durante uma sessão de jogos quebram a imersão, ao passo que sensores (como oxímetros) obstruem a movimentação do usuário, perturbando os resultados. Identifica-se que são necessárias técnicas não intrusivas para o monitoramento do estado emocional de um usuário de jogos eletrônicos.

No contexto de jogos digitais, entende-se como ideal o estado emocional que oscila entre estresse e tédio, mantendo um equilíbrio (BEVILACQUA, 2018a). Sinais psicofisiológicos são de extrema valia para detecção destas emoções, por serem relacionados ao sistema nervoso central são manifestados independente da pretensão do usuário. Dentre estes sinais, expressões faciais estão entre os mais poderosos (LI; DENG, 2019).

Avanços em áreas como visão computacional e computação afetiva tornaram possível que o rosto de uma pessoa seja monitorado, fazendo uso de uma câmera, com a finalidade de identificar suas emoções. Assim, Bevilacqua (2018a) propôs um método para aquisição do estado emocional de uma pessoa durante uma sessão de jogos eletrônicos, utilizando uma câmera para captar expressões faciais. O processo é composto de duas etapas: na primeira (calibração), o usuário é exposto a jogos planejados para causar tédio e estresse em momentos

<sup>1</sup>Discente de Ciência de Computação, Universidade Federal da Fronteira Sul, *campus* Chapecó, **Bolsista**, contato: hilgerjeancarlo@gmail.com

<sup>2</sup>Discente de Ciência de Computação, Universidade Federal da Fronteira Sul, *campus* Chapecó, **Colaborador**, contato: junior.ramisch@gmail.com

<sup>3</sup>Doutor, Universidade Federal da Fronteira Sul, *campus* Chapecó, **Orientador**, contato: [fernando.bevilacqua@uffs.edu.br](mailto:fernando.bevilacqua@uffs.edu.br)

Subprojeto: Detecção remota de estados emocionais de estresse e tédio baseados em modelos calibrados por jogos centrados no usuário

específicos. Nesta fase são obtidos dados acerca do perfil emocional do usuário analisado, utilizados no treinamento do modelo de *machine learning*. Na segunda etapa (estimação), o modelo treinado é utilizado para o monitoramento e aferição do estado emocional do usuário em um jogo comercial.

## 2 OBJETIVOS

De modo geral, o intuito do trabalho desenvolvido foi o de aperfeiçoar o método proposto por Bevilacqua (2018a). Dentre as diversas possibilidades existentes para se alcançar o objetivo, duas delas chamaram a atenção: melhorias no modelo de *machine learning* e melhorias na qualidade das características utilizadas para seu treinamento.

## 3 METODOLOGIA

Num primeiro momento, conduziu-se um breve estudo cujo intuito era identificar ferramentas e técnicas que auxiliassem na concretização do objetivo proposto. Deste estudo, identificou-se que as características faciais produziram resultados superiores se utilizássemos um *framework* mais requintado para a etapa de extração de marcos faciais. No método original são produzidos 68 pontos, ao passo que o *framework* empregado no método melhorado utiliza 468 pontos, possibilitando uma detecção mais precisa dos músculos faciais.

Adicionalmente, o estudo original mostrou evidências de que redes neurais apresentam resultados superiores aos demais algoritmos de *machine learning* - no contexto do estudo realizado. Portanto, coube o estudo e aplicação de técnicas de *deep learning* ainda não investigadas pelo trabalho original.

## 4 RESULTADOS E DISCUSSÃO

Originalmente, o modelo de *machine learning* consistia em uma rede neural, do tipo *feedforward*, com apenas uma camada escondida. Uma rede com esta arquitetura aprende analisando cada entrada do conjunto de dados de forma individual. Como já mencionado, o processo produz um modelo de *machine learning* para cada indivíduo. Na fase de calibragem, cada jogo é projetado para provocar tédio no início e estresse no fim, apresentando uma progressão sequencial.

O caráter sequencial dos dados evidencia que informações da dimensão temporal - não captada pela rede neural do tipo *feedforward* - podem corroborar para a eficácia do método. Também, na literatura há evidências sólidas do fato (HUANG; LIU; TAO, 2021). Aproveitando-se disso, neste trabalho, buscou-se implementar uma arquitetura de rede neural

passível de aproveitar a natureza sequencial do conjunto de dados.

O *dataset* utilizado nos experimentos é o mesmo do trabalho original: com 62 indivíduos, o conjunto de treino consiste dos dados coletados durante os jogos da fase de calibração, ao passo que os dados de validação foram coletados na etapa de estimação. As *features* em uso são as mesmas descritas em Bevilacqua (2018b). O treino da rede neural sucedeu da seguinte forma: inicialmente, foi pré-treinada utilizando todos os dados de treino de todos os indivíduos. Após, foi treinada novamente nos conjuntos de treino para cada indivíduo. A acurácia final é a média das acurácias de todos os conjuntos de validação aplicados na rede treinada para cada indivíduo.

Dois tipos de redes neurais foram explorados: RNN (*Recurrent Neural Network* - Rede Neural Recorrente) simples e RNN com LSTM (*Long-Short Term Memory*). A grosso modo, uma rede neural recorrente possui a capacidade de manter informações no espectro temporal devido ao fato de passarem para a camada subsequente, dados da camada atual (GOODFELLOW; BENGIO; COURVILLE, 2016). Células do tipo LSTM ajudam a sanar problemas com a arquitetura canônica como o *vanishing gradient* (GOODFELLOW; BENGIO; COURVILLE, 2016).

Em cada arquitetura foram experimentadas diferentes variações de número de camadas escondidas, de número de neurônios e de tipos de neurônios. Os melhores resultados foram obtidos com uma RNN utilizando células LSTM com três camadas escondidas, duas de 64 neurônios e uma de 16 neurônios, produzindo acurácia média de 69% - uma melhoria de quase 7% em comparação com os 61.2% do trabalho original.

Analisando os resultados de forma mais minuciosa, verifica-se que há bastante variação na acurácia para cada indivíduo, alguns apresentando valores inferiores a 30%, evidenciando falhas graves no método para casos específicos. Todavia, para verificar de forma precisa o motivo de valores tão baixos, é necessário avaliar o processo de extração de *features* e de aprendizagem em conjunto com os vídeos originais, que não estavam disponíveis à nós no decorrer desta pesquisa.

## 5 CONCLUSÃO

Neste trabalho, explorou-se o desafio de identificar emoções humanas durante sessões de jogos. Técnicas clássicas utilizadas nesta tarefa envolvem questionários e sensores. Embora possam parecer assertivas, elas produzem um inconveniente grave: atrapalham a experiência do usuário, causando perturbações nos resultados. Tendo em vista este problema, Bevilacqua (2018a) propõe um método, no intuito de proporcionar à comunidade de pesquisa em jogos,

uma ferramenta não intrusiva e de fácil aplicação para detecção de emoções. O aperfeiçoamento deste método foi o objetivo central desta pesquisa.

Para tal, o modelo de *machine learning* foi o principal ponto explorado. O simples uso de um modelo de *deep learning* capaz de captar dados sequenciais apresentou melhora nos resultados, em termos de acurácia. A progressão de um estado emocional para outro é uma informação valiosa, e redes do tipo recorrente podem captá-la e utilizá-la para promover resultados melhores.

Todavia, alguns indivíduos apresentaram resultados muito baixos. Para averiguar com precisão o motivo, faz-se necessário acesso aos vídeos originais, não disponíveis no curso da pesquisa. Também, algoritmos de *deep learning* requerem uma quantidade elevada de dados. Utilizando as mesmas características obtidas dos mesmos vídeos mas com intervalos de tempo menores poderia resultar em um *dataset* com mais instâncias. Mas novamente, seria preciso acesso aos vídeos.

Por fim, vale ressaltar que um fluxo novo de extração de características foi implementado. Nele, as *features*, mesmas expostas por Bevilacqua (2018b), são produzidas sobre 468 marcos faciais em contraste aos 68 do trabalho original. Com mais pontos no rosto, os músculos envolvidos nas expressões faciais podem ser monitorados com maior precisão. Embora esta simples mudança possa refinar os resultados, a ausência dos vídeos dos indivíduos jogando impossibilitou a aplicação do fluxo.

## REFERÊNCIAS BIBLIOGRÁFICAS

BEVILACQUA, Fernando. **Game-calibrated and user-tailored remote detection of emotions: a non-intrusive, multifactorial camera-based approach for detecting stress and boredom of players in games.** 2018a. 170 f. Tese (Doutorado) - Curso de Computer Science, Human Computer Interaction, University Of Skövde, Skövde, 2018.

BEVILACQUA, Fernando; ENGSTRÖM, Henrik; BACKLUND, Per. Automated Analysis of Facial Cues from Videos as a Potential Method for Differentiating Stress and Boredom of Players in Games. **International Journal Of Computer Games Technology**, [S.L.], v. 2018, p. 1-14, 2018b. Hindawi Limited. <http://dx.doi.org/10.1155/2018/8734540>.

CANEDO, Daniel; NEVES, António J. R.. Facial Expression Recognition Using Computer Vision: a systematic review. **Applied Sciences**, [S.L.], v. 9, n. 21, p. 4678, 2 nov. 2019. MDPI AG. <http://dx.doi.org/10.3390/app9214678>.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. [S.L.]: Mit Press, 2016.

HUANG, Jian; LIU, Bin; TAO, Jianhua. Learning long-term temporal contexts using skip

RNN for continuous emotion recognition. **Virtual Reality & Intelligent Hardware**, [S.L.], v. 3, n. 1, p. 55-64, fev. 2021. Elsevier BV. <http://dx.doi.org/10.1016/j.vrih.2020.11.005>.

LI, Shan; DENG, Weihong. Deep Facial Expression Recognition: a survey. **Ieee Transactions On Affective Computing**, [S.L.], p. 1-1, 2020. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/taffc.2020.2981446>.

**Palavras-chave:** Reconhecimento facial. Detecção de emoções. *Machine learning*. Jogos

**Nº de Registro no sistema Prisma:** PES-2020-0322.

**Financiamento:** UFFS.