

AValiação DE REPRESENTAÇÕES EMOCIONAIS FACIAIS COMO FORMA DE COMUNICAÇÃO NÃO VERBAL

JOÃO BAPTISTA CARDIA NETO¹; CRISTINA ZAPATA SPERDUTTI²; KARINE BOBADILHA²

¹Fatec Marília - Desenvolvimento de Software Multiplataforma

²FATEC Catanduva – Automação Industrial

joao.cardia@fatec.sp.gov.br

Evaluation of facial emotional representations as a form of nonverbal communication

Eixo Tecnológico: *Informação e Comunicação*

Resumo

Esta pesquisa avalia as representações emocionais faciais como uma forma de comunicação não verbal, destacando sua ambiguidade inerente e o controle intencional ou não intencional das expressões. Os objetivos principais deste estudo são: criticar a categorização tradicional das seis emoções universais, devido à sua negligência quanto às influências culturais e sociais, e propor e explorar uma abordagem dimensional para o entendimento das representações emocionais, baseada em valência e excitação. O estudo utilizou o conjunto de dados AffectNet, juntamente com Python e bibliotecas como sklearn para Modelos de Mistura Gaussiana (GMMs) e Pytorch para a implementação de redes neurais. A metodologia envolveu o cálculo de discrepâncias entre anotações categóricas e medidas dimensionais utilizando a distância euclidiana. O espaço de estados emocionais foi expandido com a incorporação de novos termos, e um GMM foi criado para avaliar divergências nas anotações existentes. Além disso, uma rede ResNet18 foi treinada utilizando rótulos probabilísticos derivados da análise dimensional e avaliada com métricas como Brier Score, Divergência de Kullback-Leibler e Probabilidade de Classificação Ordenada (*Ranked Score Probability*). Os resultados indicaram uma discrepância significativa entre os rótulos emocionais categóricos e as medidas dimensionais, sugerindo um possível viés e incerteza dos anotadores na abordagem tradicional de rotulagem. No entanto, apesar dessas discrepâncias, os estados emocionais mais prováveis identificados pela abordagem dimensional permaneceram relativamente consistentes. Em conclusão, este trabalho demonstra que o método proposto aumenta a flexibilidade das representações emocionais ao ir além dos rótulos categóricos rígidos. Pesquisas futuras devem focar no refinamento da função de perda, explorar diferentes configurações de hiperparâmetros e, potencialmente, incorporar mecanismos de atenção e modelos de linguagem para uma geração mais sofisticada de rótulos emocionais.

Palavras-chave: *Aprendizado de máquina, visão computacional, reconhecimento de expressão facial.*

Abstract

This research evaluates facial emotional representations as a form of nonverbal communication, highlighting their inherent ambiguity and the intentional or unintentional control of expressions. The primary objectives of this study are: to critique the traditional categorization of the six universal emotions due to its neglect of cultural and social influences, and to propose and explore a dimensional approach to understanding emotional representation based on valence and arousal. The study utilized the AffectNet dataset, along with Python and libraries such as sklearn for Gaussian Mixture Models (GMMs) and Pytorch for neural network implementation. The methodology involved calculating discrepancies between categorical annotations and dimensional measures using Euclidean distance. New terms were added, and a GMM was developed to evaluate differences in current annotations within the emotional state space. Furthermore, a ResNet18 network was trained using probabilistic labels derived from the dimensional analysis and evaluated using metrics like Brier Score, Kullback-Leibler Divergence, and Ranked Score Probability. The results indicated a significant discrepancy between the categorical emotional labels and the dimensional measures, suggesting potential annotator bias and uncertainty in the traditional labeling approach. However, despite these discrepancies, the most probable emotional states identified by the dimensional approach remained relatively consistent. In conclusion, this work demonstrates that the proposed method enhances the flexibility of emotional representations by moving beyond rigid categorical labels. Future research should focus on refining the loss function, exploring different hyperparameter settings, and potentially incorporating attention mechanisms and language models for more nuanced emotional label generation.

Key-words: Machine learning, computer vision, facial expression recognition.

1. Introdução

A comunicação entre diferentes pessoas é marcada por uma série de interações, que vão desde a fala até indicativos não verbais. Parte das informações que são trocadas durante uma interação é dividida entre o que é dito e o que é demonstrado, sendo essa segunda categoria conhecida como comunicação não-verbal. Dentro da comunicação não-verbal é importante destacar que muito dela acontece por expressões faciais [1] e que parte da percepção emocional durante uma interação é dada justamente pela demonstração emocional dada pela face [2], que não só é capaz de demonstrar um conjunto básico emocional mas estados complexos que lidam com a junção de diferentes estados emocionais [3].

O aspecto não-verbal da comunicação é de suma importância em diferentes cenários onde a comunicação humana tem papel central, mas é constantemente ignorado. Um exemplo claro é dentro da sala de aula onde expressões faciais e olhares podem ser uma ferramenta ímpar para a gestão dos alunos, mas nem sempre é levada em consideração [4].

Importante frisar que as demonstrações não-verbais são ambíguas e é impossível dizer que existe a possibilidade de ler as reais intenções de uma pessoa utilizando esse tipo de informação [5], as pessoas decidem quais são as representações emocionais que desejam passar durante uma interação e isso pode ser intencional ou não, construindo assim a dialética durante uma interação.

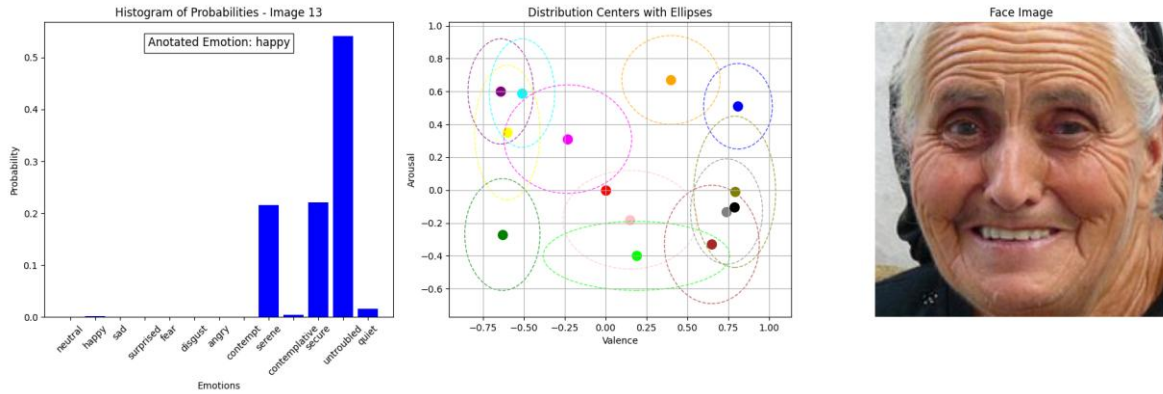
Devido essa característica ambígua da representação emocional existe grande dificuldade em mensurar quais são as demonstrações e como elas podem estar ligadas com a interação atual. Dentro desse contexto foram propostas as seis emoções universais (felicidade, surpresa, tristeza, raiva, medo e nojo) mais a não exibição emocional alguma (neutro) [6]. Apesar de ser a base para a maioria dos rótulos utilizados nos *datasets* para reconhecimento de expressão facial (*Facial Expression Recognition – FER*) essa categorização é amplamente discutida, uma vez que aspectos culturais e sociais podem afetar como a representação emocional é percebida pelos atores envolvidos em uma interação. Sendo assim a teoria das seis emoções universais é amplamente discutida e não deve ser vista mais como uma forma adequada de representação emocional [7] [8] [9] [10] [11].

Dentro desse contexto uma possível forma mais justa de avaliar as emoções é vincular sua representação com medidas dimensionais, sendo uma que diz a intensidade da sua ativação, denominada *arousal*, e outra dizendo se a emoção é positiva ou negativa, denominada *valence*. Utilizando essas duas medidas é possível descrever uma representação emocional facial de forma que não esteja essencialmente ligada com emoções, mas sim o quanto e a qualidade da expressão facial.

Apesar do reconhecimento de expressão facial dimensional ser mais flexível e não estar vinculado com os vícios do reconhecimento emocional ele torna a interpretação mais complexa, uma vez que dois pontos dimensionais em um espaço possuem lacunas na sua facilidade de interpretação. Dentro desse contexto os autores em [12] propuseram 151 termos que descrevem diversos estados emocionais e, a partir de dois estudos com 500 pessoas no total, descrevem a média e o desvio padrão de três dimensões para a representação emocional. No caso foram descritos *pleasure*, *arousal* e *dominance*. Sendo assim se torna claro que é possível utilizar as médias e os desvios padrões das medidas de *pleasure*, que é equivalente a *valence*, e *arousal* para criar um modelo de mistura gaussiana (*Gaussian Mixture Model – GMM*) que, dado duas medidas dimensionais, é possível estimar a probabilidade do pertencimento a cada um dos componentes das misturas. A Fig. 1 exibe um resultado dessa abordagem onde, com base nos

valores de *valence* e *arousal* de uma face é calculado a probabilidade de ela pertencer a um dos componentes da GMM, no caso cada estado emocional.

Fig. 1 – Estimativa da probabilidade de pertencimento em diferentes componentes da mistura de gaussianas.



Fonte: (Autores, 2025)

Sendo assim o presente trabalho se limita a analisar o conjunto de termos, formas de mapear termos emocionais mais complexos e que possibilitem alternar entre sua especificidade de descrição.

2. Materiais e métodos

2.1. Materiais

O presente trabalho foi todo desenvolvimento com linguagem Python, a biblioteca sklearn para implementação dos GMMs e a biblioteca Pytorch para a implementação de redes neurais. Para os experimentos com reconhecimento de expressão faciais foi utilizado o *dataset* AffectNet [13]. O *dataset* possui mais de um milhão de faces coletadas da internet, sendo 440.000 imagens anotadas manualmente a presença de até oito rótulos emocionais (neutro, feliz, triste, surpreso, medo, enojado, raiva e desprezo) e as medidas dimensionais de *valence* e *arousal*. O trabalho atual utilizou as imagens que possuem anotação para ser possível o cálculo das medidas de desempenho para o modelo proposto.

2.2. Metodologia

O principal objetivo do projeto atual é a mensuração das discrepâncias das anotações das categoriais emocionais e das suas medidas dimensionais, isso será medido pela distância Euclidiana entre o valor de *valence* e *arousal* da face e o centro da distribuição emocional (1).

$$D(x, y) = \frac{1}{N} \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad (1)$$

Para cada termo dos rótulos propostos pelo *dataset* AffectNet foi encontrado, de forma manual, o termo mais próximo no trabalho de [12] e foi construída uma distribuição multivariada utilizando as médias dos valores das variáveis *pleasure* como *valence* e *arousal* como o vetor de médias da distribuição. Sendo uma distribuição normal multivariada é

Anais da VIII Mostra de Docentes em RJ

necessário à sua matriz de covariância, uma aproximação possível é utilizar o desvio padrão de cada variável como o valor da diagonal principal da matriz, deixando o resto com o valor de 0.

Uma vez construídas as distribuições foram verificadas que não existia representação emocional que possuía valores de *valence* positivos e *arousal* negativos. Para tornar o espaço emocional mais representativo foram utilizados todos os termos que atendem essas duas características e foi calculado o nível de intersecção entre as distribuições, quando o nível passava de 80% ambas as distribuições eram unificadas, utilizando a média de suas médias e dos desvios padrão. Ao fim desse processo foram adicionados mais cinco estados emocionais, sendo eles: sereno, contemplativo, seguro, despreocupado e quieto.

No próximo passo um GMM foi construído utilizando os parâmetros já conhecidos das distribuições e, com base nesse passo, foi calculado o nível de discordância entre os rótulos emocionais e qual a emoção com maior probabilidade com base nos valores dimensionais. O próximo passo foi gerar novos rótulos para os dados do *dataset*, sendo que esses rótulos são a probabilidade de cada imagem facial pertencer a uma das distribuições.

Uma vez com os novos rótulos uma resnet18 foi treinada utilizando os novos rótulos como anotações, assim ensinando a rede neural a extrair qual a probabilidade de cada face pertencer a um estado emocional. Para o treinamento foi utilizado a função de LOSS *Binary Cross Entropy*, o otimizador ADAM e foram utilizadas 20 épocas no conjunto de treinamento do *dataset* AffectNet.

Ao fim, foi utilizado o conjunto de validação para medir um conjunto de métricas da saída da rede treinada e entender a performance do modelo. Para tal foram utilizadas *Brier Score* (2), a divergência de Kullback-Leibler (3) e *Ranked Score Probability* (4).

$$Brier\ Score = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C (p_{ij} - y_{ij})^2 \quad (2)$$

$$KL(Y||P) = \sum_{j=1}^C y_j \log \left(\frac{y_j}{p_j} \right) \quad (3)$$

$$RPS = \frac{1}{C-1} \sum_{i=1}^N \left(\sum_{j=1}^C (Y_{ij} - P_{ij}) \right)^2 \quad (4)$$

3. Resultados e Discussão

Nesta seção serão discutidos os resultados obtidos pelos experimentos elencados na seção de metodologia. Inicialmente foi calculado a discrepância entre os rótulos emocionais e os valores de *valence* e *arousal*. A Tab. 1 traz a avaliação desse passo. É possível visualizar a existência de uma discrepância grande entre as anotações dimensionais e dos rótulos emocionais, provavelmente isso é devido ao viés dos anotadores, nesse sentido existe um grande nível de incerteza dentre as anotações. Nesse contexto se torna possível uma forma mais próxima de anotação com os valores dimensionais, que possibilite a mudança de granularidade das representações emocionais.

Tab. 1 - Discrepância entre anotações emocionais e dimensional.

	Concordância	Discrepância	Total
Absoluto	2257	1742	3999
Percentualmente	56.44%	43.56%	100%

Fonte: (Autores, 2025)

O próximo passo, com base na geração dos novos rótulos, é o treino de um modelo definido na literatura, no caso resnet18, por 20 épocas. Abaixo segue a Tab. 2, que lista os resultados das fórmulas 1, 2 e 3.

Tab. 2 - Resultados das métricas.

Métrica	Resultado
Divergência de Kullback-Leibler	1.795
Brier Score	0.380
Ranked Score Probability	0.075

Fonte: (Autores, 2025)

É possível ver que a divergência de Kullback-Leibler possui um resultado afastado de 0, isso quer dizer que existe uma divergência grande entre os valores das probabilidades reais e estimada. Por um outro lado o Brier Score e o Ranked Score Probability estão próximos de 0, isso pode ser um indicativo que, apesar dos valores das probabilidades possuir uma divergência grande os itens mais prováveis se mantem próximos. Isso é uma evidência que a ordem das representações emocionais está próxima do correto, mas as distribuições em si necessitam uma aproximação mais precisa.

4. Considerações finais

É possível ver que a metodologia de proposta de novos rótulos emocionais aumenta a flexibilidade da representação. Ainda assim o modelo não generaliza de forma satisfatória e possivelmente seja necessário da alteração da função de Loss ou mudança dos hiper parâmetros durante o treinamento. Também é possível pensar em modelos mais complexos para a melhor generalização da aprendizagem das distribuições latentes de cada estado emocional. Além disso também é pensado adicionar mecanismos de atenção e um modelo de linguagem para gerar rótulos emocionais mais diversos.

Referências

- [1] M. T. MOTLEY and C. T. CAMDEN, Facial expression of emotion: A comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting, *Western Journal of Communication*, pp. 1-22, 1988.
- [2] K. WOLF, Measuring facial expression of emotion, *Dialogues in Clinical Neuroscience*, pp. 457-462, 2022.
- [3] S. DU e A. M. MARTINEZ, Compound facial expressions of emotion: from basic research to clinical applications, *Dialogues in Clinical Neuroscience*, pp. 443-455, 2022.
- [4] C. P. ZEKI, The importance of non-verbal communication in classroom management, *Procedia - Social and Behavioral Sciences*, pp. 1443-1449, 2009.

Anais da VIII Mostra de Docentes em RJ

- [5] M. KHAN e S. ZEB, Academic Conversation: A Case to Exemplify the Influence of Non-Verbal Communication on Human Interaction, *NUST JOURNAL OF SOCIAL SCIENCES AND HUMANITIES*, pp. 79-98, 2021.
- [6] P. EKMAN e W. V. FIESEN, Constants across cultures in the face and emotion, *Journal of Personality and Social Psychology*, pp. 124-129, 1971.
- [7] J. A. RUSSELL, Is There Universal Recognition of Emotion From Facial Expressions? A Review of Cross-Cultural Studies, *Psychological Bulletin*, pp. 102-141, 1994.
- [8] L. F. BARRETT, R. ADOLPHS, S. MARSELLA, A. M. MARTINEZ e S. D. POLLAK, Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements, *Psychological science in the public interest*, pp. 1-68, 2019.
- [9] T. L. MAU, K. HOEMANN, S. H. LYONS, J. M. B. FUGATE, E. N. BROWN, M. GENDRON e L. F. BARRETT, Professional actors demonstrate variability, not stereotypical expressions, when portraying emotional states in photographs, *Nature Communications*, 2021.
- [10] J. I. DURÁN e J.-M. FERNÁNDEZ-DOLS, Do emotions result in their predicted facial expressions? A meta-analysis of studies on the co-occurrence of expression and emotion, *Emotion*, pp. 1550-1569, 2021.
- [11] K. A. LINDQUIST, L. F. BARRETT, E. BLISS-MOREAU e J. A. RUSSELL, Language and the perception of emotion, *Emotion*, pp. 125-138, 2006.
- [12] J. A. RUSSELL e A. MEHRABIAN, Evidence for a three-factor theory of emotions, *Journal of Research in Personality*, pp. 273-294, 1977.
- [13] A. MOLLAHOSSEINI, B. HASANI e M. H. MAHOOR, AffectNet: A New Database for Facial Expression, Valence, and Arousal Computation in the Wild, *IEEE Transactions on Affective Computing*, 2017.