Res., Soc. Dev. 2019; 8(11):e278111470
ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v8i11.1470

Reconhecimento de Gestos em Imagens usando Redes Neurais Artificiais

Gesture Recognition in Images Using Neural Networks

Reconocimiento de gestos en imágenes usando redes neuronales Artificiales

Recebido: 02/08/2019 | Revisado: 10/08/2019 | Aceito: 11/08/2019 | Publicado: 24/08/2019

André Ricardo Nascimento das Neves
ORCID: http://orcid.org/0000-0002-2911-5376
Escola Superior Batista de Manaus, Brasil
E-mail: aricardo.neves@gmail.com
Hugo Kenji Rodrigues Okada
ORCID: http://orcid.org/0000-0002-7364-7986
Escola Superior Batista de Manaus, Brasil
E-mail: hugookadastm@gmail.com
Ricardo Shitsuka
ORCID: http://orcid.org/0000-0003-2630-1541
Universidade Federal de Itajubá, Brasil
E-mail: ricardoshitsuka@unifei.edu.br

Resumo

A Inteligência Artificial é uma área de pesquisa da Computação que é voltada para desenvolver mecanismos e dispositivos que permitam simular o raciocínio humano Dentro desta, uma subarea importante é a do reconhecimento de imagens. O presente artigo tem o objetivo de descrever a parte inicial de uma pesquisa que visa analisar e identificar sentimentos registrados de expressões corporais em vídeos de avaliações de produtos. Foram planejados testes experimentais deverão afim de identificar a melhor técnica para solução do problema. Foram analisadas e testadas algumas formas de identificação de gestos por meio do emprego de redes neurais.

Palavras-chave: Inteligência artificial; Aprendizado de máquina; Identificação de expressões corporais; Reconhecimento de imagens; Analise de sentimentos.

Abstract

Artificial Intelligence is an area of computer research that is focused on developing mechanisms and devices to simulate human reasoning. Within this, an important subarea is

the recognition of images. This article aims to describe the initial part of a research that aims to analyze and identify registered feelings of body expressions in videos of product reviews. Experimental tests have been planned to identify the best technique to solve the problem. Some forms of gesture identification through the use of neural networks were analyzed and tested.

Keywords: Artificial intelligence; Machine learning; Identification of body expressions; Image Recognition; Sentiment analysis.

Resumen

La inteligencia artificial es un área de investigación informática que se centra en el desarrollo de mecanismos y dispositivos para simular el razonamiento humano. Dentro de esto, una subárea importante es el reconocimiento de imágenes. Este artículo tiene como objetivo describir la parte inicial de una investigación que tiene como objetivo analizar e identificar los sentimientos registrados de las expresiones corporales en videos de reseñas de productos. Se han planificado pruebas experimentales para identificar la mejor técnica para resolver el problema. Se analizaron y probaron algunas formas de identificación de gestos mediante el uso de redes neuronales.

Palabras clave: inteligencia artificial; Aprendizaje automático; Identificación de expresiones corporales; Reconocimiento de imagen; Análisis de sentimientos.

1. Introdução

Identificar a opinião da sociedade sobre um determinado produto é algo muito utilizado como *feedback* para a definição de futuros objetivos relacionados a propostas de crescimento e planejamento de mercado para empresas. Há muito tempo esse tipo de estudo é feito através de perguntas e respostas, resultando em um trabalho difícil e de longo prazo. No entanto, existem tecnologias atuais que possibilitam diversas outras formas de se obter estes resultados, de maneira mais precisa e menos complexa.

Uma das estratégias atuais de identificação de opinião é baseada no uso de sistemas inteligentes. Segundo Jaques & Vicari (2005), para entender e utilizar trabalhos que possuem foco em emoções para sistemas inteligentes, é obrigatório entender o que são emoções. Emoção é a expressão de um elemento existente no conjunto de estados afetivos onde se tem relativamente uma resposta. Este sentimento é incentivado por um acontecimento interno ou externo àquela pessoa e à situação encontrada.

Um desses estudos está ligado à computação afetiva, que possui como parâmetro a interpretação dos estados afetivos, para que um algoritmo computacional possa identificar as emoções. Conforme definição de Picard (1995) “Computadores estão começando a adquirir capacidade de expressar e reconhecer afetos, e em breve poderá ser dada a capacidade de ter emoções.”

A área de pesquisa conhecida como análise de sentimentos tem a tarefa de identificar e entender uma opinião humana. Nesse contexto, opiniões expressas na forma de textos, áudio, imagens e vídeo podem ser classificadas como positivas, negativas ou neutras, podendo estimar a extensão e a aceitação de um produto através de determinadas estratégias.

De modo geral, a análise de sentimentos é uma forma de auxílio para diversas áreas, provando que existem várias formas de utilizar a opinião estudada, solucionando muitos problemas de mercado, que poderiam levar muito tempo e não serem solucionados (Bar, 2013, Maynard et al, 2013, Prabowo & Thelwall, 2009, Santos, 2010).

De acordo com Duan et al, (2012), a análise de sentimentos é um fator muito importante para opiniões on-line, blogs, produtos, notícias, dentre outros. Os métodos de coleta de opiniões existentes formam uma base sólida para pesquisa de sentimentos mas, essa base na maioria das vezes possui ideias dispersas para atender às principais necessidades dos clientes, pois o tempo que é levado para sua formação é demorado, podendo haver uma mudança de opinião durante a coleta.

Um exemplo de aplicação de análise de sentimentos é a identificação de opiniões em *vídeo* aferidas por meio de expressões faciais. Esses sentimentos, além da divisão em duas categorias, que são positivo e negativo, também são classificados em uma escala qualitativa: muito bom, bom, satisfatório, ruim e muito ruim, proporcionando assim a interpretação e a classificação do sentimento representado. Através deste tipo de estudo, muitas empresas podem avaliar a aceitação de seu produto ou serviço e determinar estratégias para sua melhoria (Prabowo & Thelwall, 2009).

Atualmente, existem no mercado muitos estudos de análises de opiniões, muitos deles voltados para o uso de dados de texto, buscas em redes sociais, sistemas de identificação de sentimentos faciais em imagens, entre outros.

O trabalho de Prabowo & Thelwall (2009) relata um estudo de análise de sentimentos, utilizando *reviews* de filmes e comentários existentes no MySpace, com uma proposta de um sistema semi-automático para complementar a opinião. Essa proposta necessita de outras aplicações para ter eficácia, sendo esse seu ponto negativo.

Siersdorfer et al (2010) utilizaram imagens existentes no Flickr para estudar o

sentimento, distribuindo-os em positivo e negativo, mostrando ser um trabalho de nível de dificuldade bem elevado, o qual necessita de muitos outros estudos para poder ser aplicado em um sistema real.

Wollmer et al (2013), uniram vários sistemas de medição de expressão facial para poder julgar o áudio e o vídeo, tendo como resultado sentimentos relacionados a filmes. Esse estudo teve grande eficácia pois utilizou para coleta de emoções o áudio e o vídeo. Com o uso do áudio, tornou-se mais fácil identificar palavras chaves que representam o sentimento em questão. Porém, não foi apresentado um estudo aprofundado na identificação de sentimentos através apenas de vídeos.

Sikandar (2014) fez um trabalho onde foram discutidas as possibilidades de coleta de dados a serem examinados por diferentes algoritmos, para entender como uma máquina reconhece o comportamento humano, julgando-os através de seu desempenho e robustez. O autor concluiu que, para obter um bom resultado, provavelmente será necessário unir diferentes técnicas, tornando a aplicação trabalhosa e de difícil uso.

Bittencourt & Osório (2002) realizaram um estudo onde foram usadas redes neurais artificiais na detecção de pele em imagens digitais. Com isso, foi criada uma aplicação para o reconhecimento de gestos, aumentando a interação homem-máquina.

Acharya & Mitra (2007) fizeram uma pesquisa sobre reconhecimento de gestos com especial ênfase nos gestos corporais e expressões faciais, relatando várias aplicações e discutindo-as em detalhes, onde foi criada uma base de dados de exemplos ao treinamento.

Portanto, pode-se considerar que identificar expressões corporais para registrar e interpretar comentários ou opiniões de um determinado serviço ou produto feitos em vídeo, isto é, análise de sentimentos, ainda é um problema não resolvido na literatura. Por outro lado, soluções computacionais que possibilitem identificar se uma opinião manifestada em um vídeo é negativa, positiva ou neutra, por meio unicamente da identificação de gestos, poderão ajudar empresas a criarem planos futuros de crescimento no mercado.

2. Metodologia

Uma pesquisa é a busca por novos saberes. Segundo Pereira, Shitsuka, Parreira & Shitsuka (2018) uma pesquisa quantitativa é aquela na qual há a preocupação com valores numéricos, porcentagens e/ou estatísticas. O presente estudo tem um viés quantitativo nele busca-se resultados de treino, validação e teste para se comparar as três técnicas.

O problema investigado neste trabalho é a identificação automática de gestos do

corpo, principalmente das mãos, em imagens. Tal problemática nos leva ao objetivo geral de identificar gestos por meio de redes neurais artificiais e aos específicos de ajustar parâmetros da rede neural e testar estratégias para a extração de atributos. Para se ir ao encontro deste problema, propôs-se o emprego de três diferentes estratégias de extração que serão comparadas: a da Imagem bruta, a *Canny* e a *Surf*.

3. Técnicas de Extração de Atributos

Para poder efetuar os testes, foram escolhidas três diferentes estratégias a serem comparadas: imagem "bruta", Canny e Surf.

A técnica de imagem "bruta" é o uso da imagem original sem que não tenha ocorrido nenhum tipo de tratamento para remoção de determinada característica.

O Canny é uma técnica para detectar bordas de uma imagem, foi desenvolvido em

1986 por John F. Canny, este algoritmo é muito importante uma vez que extrai informações estruturais dos objetos, assim reduzindo a quantidade de dados a serem analisados, esta solução pode ser usada em diversos tipos de situações.

De acordo com Canny (1986) o processo da detecção acontece em cinco passos:

1. Utiliza o filtro de Gauss para remoção de ruído;
2. Encontra os gradientes da imagem;
3. Procura uma resposta exata para a borda;
4. Utiliza duplo limiar, para detectar bordas importantes;
5. Finaliza detectando as bordas fortes e fracas.

O *Surf* é uma técnica para extrair pontos de interesse de uma imagem, foi proposto por Bay (2006), a literatura o reconhece como um algoritmo rápido e robusto. Esta técnica é considerada como sendo uma evolução do *Sift*, que é outro algoritmo para detecção de pontos de interesse, uma vez que encontra pontos de interesse considerados mais representativos e de forma mais rápida.

Para descobrir esses pontos a imagem é transformada em coordenada e realiza uma cópia da imagem original para uso da Pirâmide Gaussiana ou Laplaciana, garantindo os principais pontos de interesse.

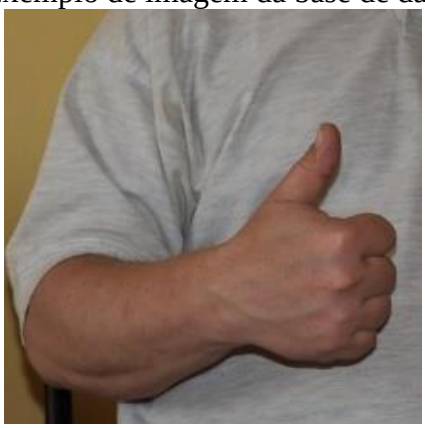
4. Protocolo Experimental

Foi utilizada como base de dados uma parte da base “Database for hand gesture recognition” disponível em http://sun.aei.polsl.pl/~mkawulok/gestures/, esta seleção possui 648 imagens com 12 pessoas, com resolução de 128 por 128 pixels e representa 6 tipos de gestos.

A base de dados foi dividida em 70% da base para treino, 15% para validação e 15% para teste. Todos os testes realizados por meio do emprego do *software* Matlab.

As imagens da base de dados estão sem nenhum tratamento a priori, segue abaixo um exemplo:

Figura 1. Exemplo de imagem da base de dados



Fonte: Arquivo dos autores.

A figura ilustra o exemplo de uma pessoa fazendo um gesto de aprovação que é o sinal manual do “positivo”.

Para identificação dos gestos faz-se o emprego do método de redes neurais do tipo Perceptron multicamadas. Os experimentos são realizados com a versão de redes neurais disponível no *software* Mat-Lab.

Nos experimentos, objetiva-se: definir o melhor conjunto de parâmetros para a rede neural, identificar o melhor filtro a ser usado, fazer o emprego dos atributos de forma, pontos de ângulos, dentre outros e, também se visa treinar, testar e comparar os resultados.

5. Resultados e discussão

Os experimentos foram realizados com alguns parâmetros da rede neural definidos como padrão. No treinamento, foi utilizada a função *Scaled Conjugate Gradient*, que tem

como característica formar a rede e atualizar os valores dos pesos de acordo com a necessidade da rede. Como algoritmo de desempenho, foi usado o *cross entropy*, sendo este o responsável por calcular o desempenho e parâmetros para penalizar as saídas imprecisas da rede.

Os testes estão divididos em 3 experimentos, sendo cada um composto por algum uso de tratamento e filtro, diferentes quantidades de neurônios e em todos os testes houve o tratamento para escala de cinza.

No primeiro teste, não ocorreu nenhum tratamento a nas imagens, apenas o tratamento *a priori*, gerando 16384 características para cada imagem.

Para o segundo teste foi usado o filtro *edge* com o algoritmo *canny*, este filtro retira as bordas das imagens e o algoritmo é um dos mais robusto, pois detecta bordas fortes e fracas, gerando 16384 características para cada imagem.

O terceiro teste teve o uso da técnica surf, este método detecta os pontos de interesse existentes na imagem, assim extraindo suas características principais, gerando 16384 características para cada imagem.

A Tabela 1, a seguir, apresenta dados dos testes realizados.

Tabela 1. Resultados de experimentos realizados.

Neurônios	Teste 1			Teste 2			Teste 3		
	Treino	Valid	Teste	Treino	Valid	Teste	Treino	Valid	Teste
10	46,5%	25,8%	36,1%	100%	44,3%	26,8%	100%	80.4%	79.4%
20	94,5%	41,2%	37,1%	100%	35,1%	39,2%	100%	76,3%	76.3%
30	98,9%	53,6%	36,1%	97,6%	36,1%	29,9%	100%	84,5%	76.3%
40	84,1%	39,2%	41,2%	100%	46,4%	39,2%	100%	86,6%	80,4%
50	85,7%	41,2%	37,1%	36,6%	23,7%	19,6%	100%	77,3%	85.6%
100	97,4%	45,4%	40,2%	87,9%	32,0%	33,0%	100%	80,4%	85.6%
150	77,3%	42,3%	38,1%	96,5%	34,0%	19,6%	100%	87,6%	87,6%
200	91,0%	46,4%	36,1%	94,7%	24,7%	25,8%	100%	82,5%	79,4%
2x 10	91,0%	42,3%	42,3%	78,6%	33,0%	23,7%	100%	71,1%	77.3%
2x 20	91,0%	42,3%	42,3%	92,5%	29,9%	29,9%	100%	80,4%	83,5%
2x 30	65,9%	36,1%	39,2%	86,1%	26,8%	26,8%	100%	83,5%	84,5%
2x 40	73,6%	39,2%	30,9%	33,0%	23,7%	19,6%	100%	88,7%	80,4%
2x 50	86,6%	50,5%	50,5%	40,7%	26,8%	30,9%	100%	82,5%	79,4%
2x 100	94,3%	43,3%	39,2%	45,4%	20,6%	22,7%	100%	78,4%	78.4%
2x 150	75,1%	43,3%	35,1%	48,0%	25,8%	17,5%	100%	85,6%	82,5%
2x 200	98,5%	43,3%	44,3%	81,5%	27,8%	18,6%	100%	67,0%	68,0%
5x 10	45,6%	28,9%	22,7%	31,7%	19,6%	21,6%	100%	79,4%	76,3%
5x 20	80,6%	38,1%	33,0%	66,5%	21,6%	26,8%	99,8%	75,3%	64,9%
5x 30	93,0%	48,5%	40,2%	55,9%	26,8%	26,8%	100%	84,5%	76,3%
5x 40	73,1%	36,1%	30,9%	87,2%	33,0%	21,6%	100%	84,5%	76,3%
5x 50	87,4%	38,1%	37,1%	72,5%	22,7%	22,7%	100%	74,2%	78,4%
5x 100	74,9%	42,3%	46,4%	30,2%	22,7%	21,6%	100%	83,5%	85,6%
5x 200	74,0%	29,9%	37,1%	80,2%	33,0%	34,0%	100%	68,0%	61,9%

Fonte: os autores.

A tabela apresenta na primeira coluna os neurônios e, nas colunas seguintes, os dados coletados para três testes. Observa-se que cada espaço na coluna Neurônios representa uma camada, exemplo “2x 10” possui duas camadas com 10 neurônios em cada, e em cada experimento existem três porcentagens de acerto: treino, validação e teste.

Os melhores resultados obtidos no teste 1 foram: 86,6% no treino, 50,5% na validação e 50,5% no teste, sendo composta por duas camadas com 50 neurônios em cada, isto mostra

que a imagem “bruta”, sem nenhum filtro ou extração, não é um bom parâmetro para reconhecimento de padrões para a rede neural, pois seu resultado mostrou uma porcentagem considerada como incerteza.

No segundo teste os melhores resultados encontrados foram: 100% no treino, 46,4% na validação e 39,2% no teste, sendo composta por apenas uma camada com 40 neurônios, isso representa que a detecção de borda, apesar de ser uma técnica muito usando para reconhecimento de formas e objetos, não mostrou um resultado satisfatório para a solução do problema proposto.

O terceiro teste obteve os melhores resultados dentre todos, estes foram: 100% no treino, 87,6% na validação e 87,6% no teste, sendo composta por uma camada com 150 neurônios, desta forma pode-se verificar que a extração de características, onde são registrados os pontos de interesse, mostrou ser uma ótima solução para o reconhecimento de gestos.

6. Considerações finais

O presente estudo contribui com o saber sobre a identificação de imagens por meio de inteligência artificial de modo a buscar a melhoria da performance deste tipo de ação e é útil para todos interessados em conhecer um pouco mais deste assunto que está em evolução na sociedade atual.

Como parte inicial deste trabalho, os experimentos realizados mostraram que a melhor forma de uma rede neural identificar gestos é através da extração de característica surf, sendo a arquitetura da rede composta por uma camada com 150 neurônios. Os melhores resultados obtidos em termos de acurácia foram: 87,6% na validação e 87,6% no teste. O uso de filtro de detecção de borda não mostrou um bom rendimento. Seu melhor nível de acerto foi com 40 neurônios, com 100% de treino, 46,4% na validação e 39,2% no teste.

Para planos futuros e melhor precisão da rede neural, pretende-se criar uma base de gestos, que se enquadre na necessidade do problema, essa base pode ter a representação do corpo como um todo ou dividir partes do corpo e realizar teste em muitos outros tipos de redes neurais. Esses experimentos são, portanto, iniciais para a pesquisa, a partir de outros experimentos deverão ser realizados para que o objetivo final seja alcançado.

Referências

Acharya, T., Mitra, S. (2007). Gesture Recognition: a survey. *IEEE Transaction onSystems, Man, And cybernetics – Part C: Applications and reviews*, 37(1): 3. Acess on: August, 01, 2019.

Bar, K. (2013). *Sentiment Analysis of Movie Reviews and Twitter Statuses*. Machine Learning–Final Project. Pp. 1-12. Available from: <http://www.cs.tau.ac.il/~kfirbar/mlproject/project-ml.pdf>. Acess on: August, 2nd, 2019.

Bay, H., Tuytelaars, T., Van Gool, L. J. (2006). *SURF: Speeded up robust features*. In: Anal of The 9th European Conference on Computer Vision (ECCV 2006). Graz, Austria, pp. 404-417.

Bittencourt, J. R. & Osório, F. S. (2002). *O uso de redes neurais artificiais na detecção de pele em imagens digitais visando o reconhecimento de gestos*. In: XI SEMINCO – Seminário de Computação 2002 da UNISINOS. Disponível em: <http://www.inf.furb.br/seminco/2002/artigos/Bittencourt-seminco2002-29.pdf>. Acesso em: 02 ago 2019.

Braga, A. P., Carvalho, A. C. P. L. F. & Ludemir, T. B. (2000). *Redes neurais artificiais*: teoria e aplicações. Ed. LTC, Rio de Janeiro/RJ.

Canny, J. (1986). A Computational Approach to Edge Detection. *IEEE Transactions on PatternAnalysis and Machine Intelligence*, 8(6): 679-698.

Duan, D., Qian, W., Pan, S., Shi, L. & Lin, C. (2012). VISA: *A Visual Sentiment AnalysisSystem.* In: VINCI '12 Proceedings of the 5th International Symposium on Visual Information Communication and Interaction. pp. 22-28. ACM, New York. Available from: <http://dl.acm.org/citation.cfm?id=2397700>. Acess on: Aug., 2nd, 2019.

Haykin, S. (2001). *Redes neurais: princípios e prática*. Ed. Bookman, Porto Alegre/RS.

Jaques, P.A., Vicari, R. M. (2005). Estado da Arte em Ambientes Inteligentes de Aprendizagem que Consideram a Afetividade do Aluno. *Revista Informática na Educação: Teoria e Prática*, 8(1).

Maynard, D., Dupplaw, D., Hare, J. (2013). *Multimodal Sentiment Analysis of SocialMedia*. University of Sheffield, Sheffield. Available from: <https://gate.ac.uk/sale/bcs-sgai-2013/arcomem.pdf>. Acess on: 1st. Aug. 2019.

Pereira, A.S, Shitsuka, D.M., Parreira, F.J. & Shitsuka, R. (2018). *Metodologia da pesquisa científica*. [*e-book*]. Ed. UAB/NTE/UFSM, Santa Maria/RS. Disponível em: https://repositorio.ufsm.br/bitstream/handle/1/15824/Lic_Computacao_Metodologia-Pesquisa-Cientifica.pdf?sequence=1. Acesso em: 02 ago. 2019.

Picard, R. W. (1997). *Affective Computing*. M.I.T Media Laboratory Perceptual Computing Section Technical Report. Disponível em: <http://affect.media.mit.edu/pdfs/95.picard.pdf>. Acesso em: 01 ago. 2019.

Prabowo, R., Thelwall, M. (2014). *Sentiment Analysis*: A Combined Approach. Jan.2009. Disponível em: <https://s3.amazonaws.com/academia.edu.documents/34362252/rudy-sentiment-preprint.pdf?response-content-disposition=inline%3B%20filename%3DSentiment_analysis_A_combined_approach.pdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIWOWYYGZ2Y53UL3A%2F20190802%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20190802T181547Z&X-Amz-Expires=3600&X-Amz-SignedHeaders=host&X-Amz-Signature=23e16fddaa9ae9b0a45aa8aa157a8e48ef4863e79d670cbd7058d4c19d92a7da>. Acesso em: August, 2nd. 2019.

Santos, H. C. (2010). *Investigação e implementação de técnicas em Análise de Sentimentos*. 35 f. Monografia apresentada como requisito parcial para obtenção do Grau em Engenharia da Computação, Universidade Federal de Pernambuco, Recife.

Siersodorfer, S., Minack, E., Deng, F. & Hare, J. (2010). *Analyzing and Predicting Sentiment of Images on the Social Web*. Article published in Siersdorfer Sources. Available from: <http://www.l3s.de/~siersdorfer/sources/2010/mm10-siersdorfer.pdf>. Acess on: August, 2nd, 2019.

Sikandar, M. (2014). A Survey for Multimodal Sentiment Analysis Methods. *Int. J. Computer Technology & Applications*, 5(1): 1470-1476, Jul. 2014. Disponível em: <http://www.ijcta.com/documents/volumes/vol5issue4/ijcta2014050421.pdf>. Acess on: August, 1st, 2019.

Wollmer, M., Weninger, F., Knaup, T., Schuller, B., Sun, C., Sagae, K. & Morency, L. (2013). Youtube Movie Reviews: Sentiment Analysis in na Audio-Visual Context. Intelligent Systems, *IEEE*, 28(3), Marc. 2013. Available from: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6487473>. Acess on: August, 1st, 2019.

Porcentagem de contribuição de cada autor no manuscrito

André Ricardo Nascimento das Neves – 35%

Hugo Kenji Rodrigues Okada – 35%

Ricardo Shitsuka – 30%