

Classificação de Sotaques Brasileiros usando Redes Neurais Profundas^{*}

Wagner A. Tostes^{*,**} Francisco A. Boldt^{*} Karin S. Komati^{*}
Filipe Mutz^{*}

^{*} Programa de Pós-Graduação em Computação Aplicada (PPComp)
Instituto Federal do Espírito Santo (IFES), Campus Serra, ES
(e-mail: {franciscoa, kkomati, filipe.mutz}@ifes.edu.br)
^{**} VixTeam Consultoria e Sistemas SA, ES
(e-mail: wagner.arca.tostes@gmail.com)

Abstract: The automatic classification of accents has several potential applications, for instance, the identification and authentication of users, forensic investigation tools and the selection of specialized models in text-to-speech and speech-to-text systems. In this work, we propose and evaluate several architectures of artificial neural networks for accent classification. The performance of these architectures in the Braccent dataset was compared with the methods GMM-UBM, GMM-SVM and iVector. Experimental results show that 4 out of 6 architectures achieve better values of accuracy, precision and recall than the previous methods. The best architecture reached 90% of accuracy, with precision, recall and F1-score of 0.92, 0.84 and 0.87, respectively.

Resumo: A classificação automática de sotaques possui diversas aplicações potenciais como a identificação e autenticação de usuários, ferramentas de investigação forense e a seleção de modelos especializados para *text-to-speech* e *speech-to-text*. Neste trabalho, propomos e avaliamos diversas arquiteturas de redes neurais artificiais para classificação de sotaques. A performance das arquiteturas na base de dados Braccent foi comparada com os métodos GMM-UBM, GMM-SVM e *iVector*. Resultados experimentais mostram que 4 das 6 arquiteturas alcançam valores melhores de acurácia, precisão e revocação que os métodos anteriores. A melhor arquitetura alcançou 90% de acerto, com precisão, revocação e F1-score de 0.92, 0.84 e 0.87, respectivamente.

Keywords: Accent Recognition; Convolutional Neural Networks; Recurrent Neural Networks;
Palavras-chaves: Reconhecimento de Sotaques; Redes Neurais Convolucionais; Redes Neurais Recorrentes;

1. INTRODUÇÃO

O processo de construção da fala (Brescancini, 2017) é influenciado por diversas características pessoais (e.g., timbre e velocidade da fala), de saúde do aparelho fonador (e.g., rouquidão e cansaço), estado emocional, traços demográficos (e.g., gênero e faixa etária), bases socioeducacionais, além do fator regional, o sotaque. O sotaque se refere à maneira distinta da fala de uma pessoa em uma língua. Diferentes sotaques podem ser identificados por variações no tom, ênfase e extensão da pronúncia de sílabas de uma palavra.

O sotaque é um dos principais fatores variáveis na fala humana, que representa um grande desafio para a robustez dos sistemas de reconhecimento automático de fala (Shi

et al., 2021). Assim, é comum que seja feito uma classificação de sotaques, para ser usado em fase anterior ou em conjunto com o modelo de reconhecimento de fala. Além disso, esta classificação pode ser usada para reconhecimento automático de falantes em sistemas de identificação e autenticação, ou ainda em aplicações de investigação forense (Rose, 2002).

Devido às características únicas dos sotaques em diferentes línguas, os trabalhos na área em geral têm foco em línguas específicas. Por ser uma das línguas mais faladas no mundo (Shi et al., 2021), vários trabalhos estudam os sotaques na língua inglesa (Ahmed et al., 2019; Wang et al., 2020; Zhang et al., 2021). Encontram-se artigos sobre sotaque árabe (Biadisy et al., 2009), sotaque francês (Lazaridis et al., 2014), sotaque em mandarim (Weninger et al., 2019), dialetos da Nigéria Salau et al. (2020), dentre outros. Em português, destacam-se o trabalho realizado por Ynoguti (1999) e por Batista et al. (2019).

Uma grande contribuição do trabalho de Batista et al. (2019) foi a elaboração da base de dados Braccent que contém 1.757 áudios de sete sotaques diferentes: nortista, baiano, fluminense, mineiro, carioca, nordestino e sulista.

^{*} Agradecemos à FAPES e a CAPES pelo apoio financeiro dado por meio do PDGP (Parcerias Estratégicas nos Estados da CAPES) (PROCESSO: 2021-2S6CD, TO/nº FAPES: 132/2021). Também agradecemos ao Propós (Programa Institucional de Apoio à Pós-graduação Stricto Sensu) do IFES pela apoio financeiro. Filipe Mutz agradece ao Instituto Federal do Espírito Santo (IFES) por incentivar sua pesquisa via o Programa Pesquisador de Produtividade (PPP) - portaria n. 1072 de 21 de maio de 2020.