Projeto de Pesquisa de Iniciação Científica

Reconhecimento Inteligente de Locutores na Modalidade Independente do Discurso

[nome do(a) candidato(a) removido(a) em função da análise duplo-cega] [1], [nome do(a) orientador(a) removido(a) em função da análise duplo-cega] [2], [filiação removida em função da análise duplo-cega]

Resumo: A intenção deste trabalho de pesquisa é a de revisar conceitos e, em seguida, projetar e implementar em nível de *software*, um algoritmo para identificação de locutores na modalidade independente do discurso. Valendo-se de uma base específica de sinais de voz, provida por empresa conveniada com a unidade universitária na qual desenvolver-se-á o trabalho, o propósito particular é o de contabilizar, em meio a um discurso de dezenas de horas, quanto tempo determinados locutores fizeram uso da palavra. Após uma etapa de revisão bibliográfica e estudo da base de dados, serão extraídas e comparadas características temporais e espectrais associadas a diversas configurações de classificadores, registrando-se os resultados em matrizes de confusão e adotando-se um procedimento de validação cruzada para treinamento e testes da estratégia desenvolvida.

Abstract: The objective of this research work is to study, design, and implement, in a programming language, a text-independent algorithm for speaker identification. Using a specific speech database, provided by a company with a partnership with the university branch in which the project will be conducted, the particular purpose is that of accounting, in a long speech recording with dozens of hours, how much time certain speakers took part in the discourse. After a bibliographic review and a study on the database, time-frequency features associated with a number of classifiers will be compared, registering the results in confusion matrices and adopting a cross-validation procedure for training and testing the proposed strategy.

[1] Candidato.
[2] Orientador

1. Introdução

Notavelmente, as técnicas de identificação de indivíduos por voz têm se tornado cada vez mais frequentes, principalmente em aplicações de controle de acesso, substituindo tradicionais sistemas baseados em senhas [1] [2]. O potencial exponencialmente crescente dos computadores constitui um forte fator motivador para que os sistemas de reconhecimento de padrões, inclusive aqueles baseados em voz, estejam presentes em aplicações diversas [3]. Nitidamente, o referido tema tem merecido considerável atenção da comunidade científica, conforme é possível notar mediante uma busca realizada na base científica do *Web of Science* [4].

Claramente, estratégias específicas das áreas de Inteligência Artificial e Processamento de Sinais têm oferecido um grande número de possibilidades para a implementação de sistemas de reconhecimento de padrões em sinais de voz. Técnicas do tipo *deep learning* [3], em associação com extratores de características que permitam a obtenção de informações no domínio conjunto tempo-frequência [5], têm possibilitado resultados promissores.

Diante do exposto, este trabalho concentra-se no estudo de conceitos, no projeto e na implementação de uma estratégia computacional para identificação de locutores, na modalidade *text-independent*, conforme especificado adiante. Estima-se que este projeto trará significante contribuição para a formação do candidato que, supervisionado pelo seu orientador, será colocado em contato com um tema de potencial importância na área de reconhecimento de padrões.

2. Objetivos

O objetivo deste projeto de pesquisa, em nível de iniciação científica, é o de permitir que o candidato adquira experiência nas áreas de processamento digital de sinais e sistemas inteligentes, por meio do estudo de conceitos relevantes, do projeto e, também, da implementação em linguagem de programação de alto nível, de um sistema de identificação de locutores. Valendo-se de uma base específica de sinais de voz, provida por empresa conveniada com a unidade universitária na qual desenvolver-se-á o trabalho, o objetivo específico é o de contabilizar, em meio a um discurso de muitas horas, quanto tempo determinados locutores fizeram uso da palavra. A intenção é a de executar os respectivos algoritmos no modo *off-line*, de forma a possibilitar a comparação de características temporais e espectrais e de classificadoresdiversos, analisando os resultados por meio de validações cruzadas e matrizes de confusão. Pretende-se, ainda, apresentar os frutos da pesquisa no Congresso Anual de Iniciação Científica (CIC) da Unesp.

3. Metodologia

A pesquisa será conduzida pelo candidato, com a supervisão do orientador, de forma que ambos terão reuniões periódicas para explicações e avaliações. Inicialmente, será realizada uma revisão dos conceitos das áreas de processamento de sinais e de inteligência artificial, pertinentes ao projeto. Em seguida, a base de sinais de voz, com dezenas de horas de gravações, será estudada em detalhes, permitindo a coleta dos sinais-modelo de cada locutor-alvo.

Na sequência, serão extraídas características, isto é, valores numéricos representativos dos sinais-modelo de voz, visando caracterizar os locutores-alvo independentemente do discurso. Para tal, após uma etapa de pré-processamento dos sinais de voz, que inclui normalizações e operações de pré-ênfase [2], serão delimitadas regiões que estejam na fronteira entre segmentos vozeados e não-vozeados [2], dos quais serão computadas a frequência fundamental (F_0) e as frequências formantes (F_1, F_2, F_3 e F_4) [2], além de características prosódicas de mais alto nível [7] [8] [9] [10] . Uma vez que as melhores características sejam escolhidas, por meio da lógica paraconsistente [11], a qual tem se mostrado relevante para tal fim [12], serão definidas e testadas configurações distintas de classificadores *knowledge-based* do tipo *deep neural network* (DNN) [13] e outros, os qual contemplam o estado-da-arte em termos de reconhecimento de padrões. Nesse quesito, variar-se-ão o número de camadas ocultas e a quantidade de neurônios artificiais ativos não-lineares em cada estrutura de DNN testada, optando-se por aquela que tenha condições de solucionar o problema com a menor complexidade e custo computacionais. Tratamentos que possibilitem, ainda, modelos mais interpretáveis, o que constitui foco de considerável atenção atualmente [14], serão preferíveis.

Particularmente, serão testadas as diversas combinações de características com as diversas configurações de classificadores, com o intuito de analisar a acurácia de cada uma, documentando-se os resultados em matrizes de confusão e permitindo, assim, a consolidação de conclusões que relacionem caracteríticas, classificadores e locutores. Os algoritmos serão implementados em linguagem de programação específica, tal como *Python* e C/C++. Todo o trabalho será documentado em um relatório final e os resultados serão apresentados no CIC da Unesp.

4. Cronograma e Plano de Atividades

A tabela 1 contém o cronograma proposto para a execução do projeto, sendo que as fases, de T_1 até T_5, correspondem às seguintes atividades:

- T_1: revisão bibliográfica dos conceitos pertinentes e trabalhos relacionados com a área do projeto, assim como detalhamento da base de vozes;
- T_2: projeto e o desenvolvimento da técnica proposta, particularmente trabalhando na etapa de extração e seleção de características;

- T_3: projeto e o desenvolvimento da técnica proposta, particularmente trabalhando na etapa dos classificadores *knowledge-based*;
- T_4: testes e refinamentos da técnica proposta e preparação para apresentação no CIC da Unesp;
- T_5: apresentação do projeto, redação e entrega de relatório final.

Tabela 1: Cronograma do projeto

Mês / Fase	T_1	T_2	T_3	T_4	T_5
Setembro de 2022		X			
Outubro de 2022		X			
Novembro de 2022		X			
Dezembro de 2022		X			
Janeiro de 2023			X		
Fevereiro de 2023			X		
Março de 2023			X		
Abril de 2023			X		
Maio de 2023				X	
Junho de 2023				X	
Julho de 2023					X
Agosto de 2023					X

Referências

[1] Beigi, H. **Fundamentals of Speaker Recognition**. Yorktown: Springer, 2011.

[2] Mak, M.W.; Chien, J.T.**Machine Learning for Speaker Recognition**, Cambridge University Press, 2000.

[3] Duda, R. et. al. **Pattern Classification**. 2 ed. New York: Wiley-Interscience, 2000.

[4] http://isiknowledge.com . Acesso em Junho de 2021.

[5] Lyons, R.D. **Understanding Digital Signal Processing**. 2 ed. New Jersey: Prentice Hall, 2004.

[6] Bossi, M.; GOLDBERG, R. **Introducing Digital Audio Coding and Sandards**. Massachusetts: Springer, 2003.

[7] Guido, R.C. A tutorial on signal energy and its applications. *Neurocomputing*, v. 179, pp. 264-282, (2016).

[8] Guido, R.C. ZCR-aided neurocomputing: a study with applications. *Knowledge-based Systems*, v. 105, pp.248-269, (2016).

[9] Guido, R.C. A Tutorial-review on Entropy-based Handcrafted Feature Extraction for Information Fusion. *Information Fusion*, n.41, pp.161-175, (2018).

[10] Behrman, A. **Speech and Voice Science**. 4.ed. Plural Publishing, Inc., 2021.

[11] Avron, A.; Arieli, O.; Zamansky, A. **Theory of Effective Propositional Paraconsistent Logics**. College Publications, 2018.

[12] Guido, R.C. Paraconsistent Feature Engineering. *IEEE Signal Processing Magazine*, v.36, n.1, pp. 154-158, (2019).

[13] Goodfellow, I.; Bengio, Y.; Courville, A. **Deep Learning**. The MIT Press, 2016.

[14] Monte-Serrat, D.M. Interpretability in neural networks towards universal consistency. *International Journal of Cognitive Computing in Engineering*, v.2, pp. 30-39, (2021).