



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
Centro Tecnológico
Colegiado do Curso de Engenharia de Computação

COMUNICADO DE DEFESA DO PROJETO DE GRADUAÇÃO **(Anexo I – Resolução 02/2013-CCEC)**

Nome do Estudante: Gabriel Soares Xavier

Título: Classificador de Oradores em uma Comunicação via Rádio

Data da apresentação: 31/01/2023

Horário: 10:30

Local: <https://meet.google.com/xnf-guih-gwe>

Banca Examinadora:

Prof.: Dr. Jorge Leonid Aching Samatelo (Depto. de Engenharia Elétrica - UFES) (Orientador)

Prof.: Eng. Lucas Grigoletto Scart (Depto. de Engenharia Elétrica - UFES) (Co-orientador)

Prof.: Raquel Frizera Vassallo (Depto. de Engenharia Elétrica - UFES)

Prof.: Flavio Miguel Varejão (Depto. de Informática - UFES)

Resumo do trabalho:

O problema de Diarização de Locutor consiste em responder duas perguntas, "Alguém está falando?" e "quem está falando?". Portanto, o objetivo é identificar em um áudio os diferentes segmentos que ocorrem momentos de falas e os oradores responsáveis pelos segmentos. A tarefa de Diarização demonstrou ter ganhos

significativos para análises de áudios em tarefas do dia a dia, onde a indexação de diálogos é extremamente importante, como, por exemplo, consultas médicas, procedimentos jurídicos, reuniões e convenções, entre outros tipos de comunicação. Além disso, na literatura já se mostrou que a inclusão da diarização em outras tarefas adiciona ganhos em desempenho, por exemplo, em tarefas de Reconhecimento de Fala Automatizada. Este trabalho foi realizado em parceria da empresa Vale S.A, que apresentou uma necessidade de indexação e histórico das comunicações via rádio de seus funcionários. Dito isso, o projeto desenvolve uma versão de diarização específica para o problema da Vale S.A, onde algumas alterações são realizadas para simplificar e atender melhor as dores da empresa. A proposta deste projeto é modificar a diarização para ser uma tarefa de classificação dos segmentos de fala, onde cada classe representa um grupo específico de responsabilidade na empresa. Ao longo deste texto, serão apresentados os fundamentos que cobrem a teoria da tarefa de diarização, seguindo pela metodologia de desenvolvimento da solução, onde é baseada no pipeline desenvolvido pelo Pyannote. Os modelos utilizados foram treinados no conjunto de dados públicos VoxConverse e ajustados aos dados privados da Vale S.A. Neste projeto, foram realizados alguns testes, como a utilização do clustering para a separação das classes, método clássico na diarização, e a solução proposta que utiliza um classificador para a identificação da classe do segmento. Para validar o classificador neural, foram utilizados modelos clássicos como Random Forest e KNN para observar o desempenho dos 3 classificadores quando aplicado aos dados de teste. Durante os experimentos, o melhor pipeline quando se olha isoladamente para o Diarization Error Rate, é o que utiliza o classificador neural desenvolvido e descarta a etapa de Speaker Change Detection, tendo um desempenho de 3.80%. Já quando se observa o Jaccard Error Rate, o melhor pipeline é o que adota o KNN como classificador e descarta a etapa de Speaker Change Detection, tendo um Jaccard Error Rate de 8.45%. Durante os experimentos, é discutido o porquê da remoção da camada de Speaker Change Detection. Por fim, apresentação de possíveis melhorias do modelo proposto são apresentados na conclusão deste texto.

Palavras-chaves: Diarização de Locutor; Redes Neurais; Redes Neurais Recorrentes; Redes Convolucionais; Processamento de Fala; Classificação de Fala.

