Desempenho de técnicas de Aprendizado de Máquina em Reconhecimento de Voz

RESUMO EXPANDIDO - Disciplina de TCC290009

Mário André Lehmkuhl de Abreu

Estudante do Curso de Engenharia de Telecomunicações

Ramon Mayor Martins

Professor orientador

Semestre 2019-2

Resumo- *Com os avanços tecnológico ocorrido nos últimos anos o mundo vem passando por transformações, sendo notável cada vez mais a interação dos seres humanos com aplicações computacionais. Assistentes pessoais ativados por voz, carros autônomos, reconhecimento facial, marcação automática em fotos de redes sociais entre outros já são uma realidade. Para a interação com essa aplicações, uma tecnologia que vem ganhando destaque na ultima década vem sendo usada, o Aprendizado de Maquina (Machine Learning). O Aprendizado de Máquina (Machine Learning) é uma área da inteligencia artificial (IA) que investiga como as máquinas podem aprender através da extração de padrões a partir de um conjunto de dados. A partir disso varias aplicações podem ser desenvolvidas em vários cenários diferentes. Uma que vem ganhando bastante presença no cotidiano da sociedade, é o reconhecimento de voz. Seu uso reflete em mais eficiência, na realização de ações, não sendo necessário utilizar as mãos. Além disso, podem dar maior acessibilidade às pessoas com limitações motoras. Desse modo esse trabalho tem como objetivo analisar e comparar o desempenho de várias técnicas de aprendizado de máquina úteis no reconhecimento de voz. Para servir como uma contribuição de informação, e verificar qual técnica apresenta o melhor resultado de uso, no desenvolvimento de sistema com reconhecimento de voz. Todas as análises serão realizadas em uma base de dados de vozes selecionadas. Os resultados obtidos serão avaliados em termos de eficiência, precisão (acurácia), sensibilidade e especificidade.*

Palavras-chave: Reconhecimento de voz, Inteligencia artificial, Técnicas de aprendizado de máquina.

1 Introdução

Com os avanços tecnológico ocorrido nos últimos anos o mundo vem passando por transformações, sendo notável cada vez mais a interação dos seres humanos com aplicações

computacionais. Assistentes pessoais ativados por voz, carros autônomos, reconhecimento facial, marcação automática em fotos de redes sociais entre outros já são uma realidade. Para a interação com essa aplicações, uma tecnologia que vem ganhando destaque na ultima década vem sendo usada, o Aprendizado de Maquina (Machine Learning)(DATA SCIENCE ACADEMY, 2019).

O Aprendizado de Máquina (Machine Learning) é uma área da inteligencia artificial (IA) que investiga como as máquinas podem aprender através da extração de padrões a partir de um conjunto de dados.(ARRIGONI, 2018) Em uma definição fundamental, representa o uso de algoritmos para extrair informações de dados brutos e representá-los por meio de algum tipo de modelo matemático. Com esse modelo conclusões são feitas a partir de outros conjuntos de dados. Para esse processo existem vários algoritmos disponíveis (DATA SCIENCE ACADEMY, 2019).

No entanto aprender é uma ação muito elaborada e complexa. No Aprendizado de Máquina (Machine Learning), fazer isso requer a extração de um padrão para que o mesmo seja primeiramente reconhecido. Isso visa detectar automaticamente regularidades dentro de um conjunto de dados, possibilitando que a máquina possa tomar decisões, como a seleção de dados em diferentes categorias ou classes. Essa classificação se baseia em conhecimentos preliminares ou dedutivos, e em informações estatísticas conseguidas a partir de referências. Para avaliar o resultados obtidos, termos de eficiência, precisão (acurácia), sensibilidade e especificidade são usados.(ARRIGONI, 2018).

A partir disso varias aplicações podem ser desenvolvidas em vários cenários diferentes, como na medicina e na saúde, transportes, agronegócio, comercio em geral, serviços públicos, cidades inteligentes e na indutaria, com o conceito de Indústria 4.0, um termo recente que explica a aplicação das novas tecnologias nos principais processos industriais(BARBOSA, 2017). Segundo o Gartner, até 2020 todos os softwares corporativos terão alguma funcionalidade ligada ao Aprendizado de Máquina (Machine Learning)(DATA SCIENCE ACADEMY, 2019).

Diante desses cenários, um que vem ganhando bastante presença no cotidiano da sociedade, é o reconhecimento de voz. Seu uso reflete em mais eficiência, na realização de ações, não sendo necessário utilizar as mãos (TRINDADE, 2018). Além disso, podem dar maior acessibilidade às pessoas com limitações motoras (ANDRADE et al., 2016). Computadores, smartphones e sistemas multimídias de carros, são alguns exemplos em que o reconhecimento de voz esta inserido. No caso dos computadores e smartphones, as aplicações do momento são os assistentes virtuais como, Siri (Apple), assistente Google, Cortana (Microsoft) e Amazon Echo com Alexa, tendo como objetivo atender seus usuários (TRINDADE, 2018).

Porém trabalhar com a voz não é uma tarefa fácil. O sinal de voz produzido pelo sistema vocal humano é muito complexo, resultando em uma difícil tarefa de caracterizá-lo. Inúmeros modelos matemáticos sofisticados existem que tentam simular a produção da voz humana, entretanto, sua capacidade de modelamento ainda é muito limitada.

Pelo fato de o sinal de voz carregar muita informação, não é possível sintetizar e analisa-lo diretamente. Desse modo para desenvolver sistemas de reconhecimento de voz, é usado a técnica de sistemas modernos de Reconhecimento Automático de Voz (RAVs).

Eles são responsáveis por produzir uma representação textual do sinal de voz. Para isso utilizam a combinação de algoritmos de extração e reconhecimento de características. Com seu uso muitas aplicações podem ser desenvolvidas como, comandos de voz, diálogos interativos, transcrição de discursos gravado, ditado e pesquisa em documentos de áudio.

Os algoritmos de extração de características são usados para remover do sinal qualquer informação que não seja útil para o reconhecimento, ressaltando aspectos que contribuem para a identificação de diferenças. Dessa extração, vetores de características são gerados. Em seu uso tem-se disponível algumas técnicas para o modelamento acústico como, MFCC (Mel-Frequency Cepstral Coefficients), LPC (Linear Predictive Coding) e PLP (Perceptual Linear Prediction Coefficients).

No que diz respeito ao reconhecimento de características, etapa que classifica e reconhece a informação do sinal analisado, tem-se disponivel técnicas como HMM (Hidden Markov Models - Modelo oculto de Markov), modelamento linguístico, e técnicas de aprendizado de máquina como RNA (ou ANN, Artificial Neural Networks), SVM (Support Vector Machines - Máquina de vetores de suporte), Random F, Naive Bayes, arvore de decisão, KNN e mais recentemente, DNN (Deep Neural Networks)(THIAGO, 2017).

Este trabalho tem como objetivo analisar o desempenho de técnicas de aprendizado de máquina em reconhecimento de voz, para servir como uma contribuição de informação, e verificar qual técnica apresenta o melhor resultado de uso, no desenvolvimento de sistema com reconhecimento de voz. No algoritmo de extração será utilizado o MFCC (Mel-Frequency Cepstral Coefficients). No uso do sinal de voz, uma base de dados com MFCC, composta por palavras isoladas (dígitos de zero a nove, pronunciados em inglês) será usado. No reconhecimento de características será usado as técnicas de aprendizado de máquina RNA (ou ANN, Artificial Neural Networks), SVM (Support Vector Machines), Random F, Naive Bayes, arvore de decisão e KNN. Para a analise de desempenho serão avaliados os termos de eficiência, precisão (acurácia), sensibilidade e especificidade.

2 Metodologia

O projeto será dividido em cinco etapas que se inicia em: Análise e especificação dos requisitos, Estudo das tecnologias utilizadas, Desenvolvimento, Testes de utilização e Verificação do desempenho.

2.1 Análise e especificação dos requisitos

Nesta etapa será feito o levantamento dos requisitos que o projeto possuirá.

2.2 Estudo das tecnologias utilizadas

Com a analise e especificação concluída , o próximo passo é fazer os estudos das tecnologias que serão utilizadas para o desenvolvimento do projeto.

2.3 Desenvolvimento

Após o estudo e entendimento das tecnologias utilizadas, o desenvolvimento do algoritmos de aprendizado de máquina será iniciado.

2.4 Testes de utilização

Concluído o desenvolvimento, testes de uso serão feitos para verificar se o funcionamento está de acordo com os requisitos especificados. Caso seja detectado algum problema, correções serão feitas.

2.5 Verificação do desempenho

Com o o funcionamento dos algoritmos validados, o teste de desempenho será feito para verificar qual técnica de aprendizado de máquina é mais eficiente no reconhecimento de voz..

3 Considerações Parciais/Finais

Após investigação, analises e desenvolvimento das técnicas de aprendizado de máquina RNA (ou ANN, Artificial Neural Networks), SVM (Support Vector Machines), Random F, Naive Bayes, arvore de decisão e KNN, definidos para o reconhecimento de voz, espera-se verificar qual técnica possui o melhor desempenho no reconhecimento. E com isso contribuir na determinação de qual técnica é melhor de se usar no desenvolvimento de sistemas de reconhecimento de voz.

Referências

ANDRADE, C. Araujo de et al. Avaliação de um módulo de reconhecimento de fala na plataforma arduino. *Revista Principia - Divulgação Científica e Tecnológica do IFPB*, v. 1, p. 19, 06 2016.

ARRIGONI, T. R. Reconhecimento de silhueta de automóveis para carros autônomos utilizando aprendizado de máquina. 2018.

BARBOSA, J. P. *Os três pilares da inovação – Machine Learning, Big Data e IoT*. 2017. Disponível em: <http://igti.com.br/blog/os-tres-pilares-da-inovacao-machine-learning-big-data-e-iot>. Acesso em: 19 Outubro. 2019.

DATA SCIENCE ACADEMY. *Deep Learning e a Tempestade Perfeita*. 2019. Disponível em: <http://deeplearningbook.com.br/deep-learning-a-tempestade-perfeita>. Acesso em: 17 Outubro. 2019.

THIAGO, E. R. de S. Reconhecimento de voz utilizando extração de coeficientes mel-cepstrais e redes neurais artificiais. 2017.

TRINDADE, R. *Siri, Alexa e Google: Veja como usar as assistentes de voz no celular*. 2018. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2018/05/31/siri-alexa-e-google-veja-como-usar-as-assistentes-de-voz-no-celular.htm>. Acesso em: 19 Outubro. 2019.