

Análise e Proposição de Modelos de Reconhecimento de Fala para Aplicação em Sistemas Robóticos

Bruno Rafael Barbosa Lima, Universidade de Pernambuco (brbl@ecomp.poli.br)

Alexandre Magno Andrade Maciel, Universidade de Pernambuco (amam@ecomp.poli.br)

Os sistemas de reconhecimento automático de fala (RAF) têm sido cada vez mais utilizados em diversas aplicações dos sistemas digitais, são cada vez mais os dispositivos que possuem esta tecnologia tais como smartphones, TVs, computadores pessoais, entre outros. Neste campo de pesquisa, grandes empresas que produzem soluções e serviços computacionais como o Google, o Baidú e a Microsoft, vêm trabalhando para obter avanços que demonstram as capacidades básicas e limitações da tecnologia atual aprendizagem profunda (DENG et al., 2013). As Redes Neurais Profundas (do inglês: DNNs) recebem essa denominação pois possuem várias camadas escondidas, elas são modelos artificiais inspiradas no cérebro humano e demonstram ter um grande potencial para o reconhecimento de voz. Dado este cenário, no Brasil não é igualmente desenvolvido como para outros idiomas, as bases são mais restritas e possuem menos exemplos para serem utilizadas. Com isso, foi desenvolvido o Framework Integrated Voice Environment (FIVE) (MACIEL; CARVALHO, 2010), com o objetivo de auxiliar na construção de motores de fala e a instanciação destes em diferentes ambientes tecnológicos. Este trabalho objetiva realizar uma análise pormenorizada das principais abordagens para reconhecimento de voz, em uma base de dados com áudios no idioma do português do Brasil. Para isto, será realizada uma investigação das principais abordagens e técnicas apresentadas na literatura, que atualmente indicam as redes neurais como solução. Serão confeccionados scripts específicos com a construção das redes neurais, mais adequadas ao ambiente de aplicação, com aprendizagem profunda. Experimentos serão feitos e, em seguida, testes estatísticos serão aplicados aos resultados para validação dos mesmos. Será inicializado um processo de integração ao FIVE para comparações de resultados com outras técnicas. Outro foco é tornar o Framework compatível com a arquitetura robótica, para que se torne possível a geração de motores de reconhecimentos para robôs e estes passem a interagir e receber comandos através da fala humana. O que se espera alcançar é que o FIVE seja compatível com as arquiteturas e restrições dos ambientes robóticos, e que estes possam ter como opção de entrada comandos de voz, aumentando as alternativas de aplicação do software em desenvolvimento. Nos testes realizados alcançamos 25% de erro nos labels, no melhor modelo, com uma base de 13 horas áudio. Assim podemos concluir que temos um projeto viável e com um grande cenário de aplicação, visto que o FIVE irá ampliar o seu portfólio de compatibilidade em vários sistemas digitais.

Palavras-chave: *Reconhecimento de Fala; Robôs; Redes Neurais;*

Referências

DENG, L. et al. Recent advances in deep learning for speech research at Microsoft. In: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, 2013. p. 8604-8608.

MACIEL, A.; CARVALHO, E. Five-framework for an integrated voice environment. In: Proceedings of International Conference on Systems, Signal and Image Processing, Rio de Janeiro. [S.l.: s.n.], 2010.