

Detalhes do Projeto de Pesquisa

Marcos Vinícius Treviso
Universidade de São Paulo

21 de junho de 2016

1 Tema

Detecção de pontuação final e disfluências em narrativas transcritas de pacientes com: Doença de Alzheimer (DA); Comprometimento Cognitivo Leve (CCL); e Controle (i.e. pessoas saudáveis).

Este projeto está contido num projeto de maior escopo cujo objetivo é, justamente, realizar o diagnóstico precoce de pessoas com CCL e DA através de suas narrativas. A objetivo da classificação é decidir se uma pessoa possui DA, CCL, ou nenhuma das duas. A Figura 1 mostra, em vermelho, a fase com o qual este projeto se encaixa no panorama geral.

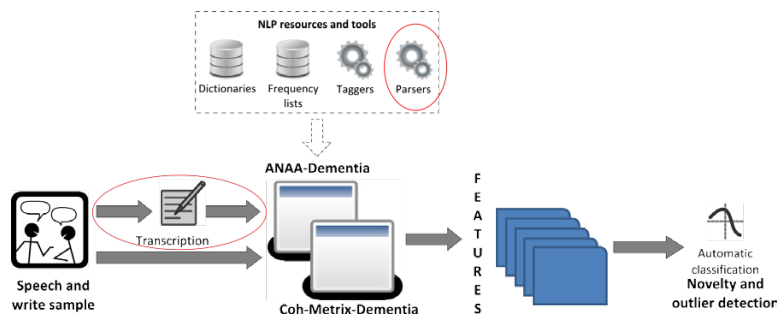


Figura 1: Estrutura dos sistemas, infraestrutura computacional e uso de técnicas de aprendizado de máquina para diagnóstico precoce de DA/CCL. Destacado em vermelho a parte que será realizada neste projeto.

2 Título

Detecção de Disfluências e Limites de Sentenças em Transcrições de Narrativas da Tarefa de Reconto visando a Extração Automática da Densidade de Ideias.

3 Lacuna

Detectar sinais de pontuação (incluindo vírgulas, pontos de exclamação e interrogação) já é uma tarefa bem definida e que vem sendo realizada nos últimos 15

anos, devido ao surgimento da Web 2.0 e de *softwares* que fazem o reconhecimento automático de voz (RAV) e fazem uma transcrição automática. Por lidar com fala, trabalhos que lidam com RAV, a tarefa de detecção de disfluências também vem sendo realizadas nos últimos anos.

No entanto, o nosso cenário é mais fechado, pois trabalhamos com pessoas com deficiências cognitivas, e que são possivelmente idosas. Essas características influenciam na qualidade do texto e da fala.

4 Hipóteses

Acreditamos que para realizar as detecções mencionadas, iremos precisar separar dois modelos computacionais: um que trata a narrativa transcrita (isto é, trata o texto); e outro que trata o áudio (isto é, lida com informações prosódicas).

De acordo com trabalhos relacionados, acreditamos que o desempenho das detecções deverá ser maior para pacientes de Controle, do que para CCL e DA. Visto que a deficiência cognitiva é o grande dificultador.

Além disso, o resultado vindo das detecções será passado para um próximo passo no *pipeline*, que será responsável de extrair a métrica de Densidade de Ideias (DI). Acreditamos que o resultado das nossas detecções vai influenciar diretamente no desempenho dessa métrica, que é calculada através de uma *parser* baseado em regras, e portanto sua entrada deve estar com uma linguagem textual bem formada.

5 Objetivos

Nosso maior objetivo é remover o ruído inserido por problemas de fala e da narrativa do paciente, de modo que a Densidade de Ideias consiga ser extraída sem problemas. Uma vez que esses ruídos forem detectados, eles podem ser facilmente eliminados.

Com isso, podemos definir nossos objetivos em:

1. Desenvolvimento de uma ferramenta que elimine as disfluências de textos de sujeitos com CCL, DA e normais, que são extraídos de maneira automática por um reconhecedor automático de voz.
2. Desenvolvimento de uma ferramenta que segmente automaticamente os textos afásicos extraídos de um RAV em sentenças.
3. Avaliação das ferramentas através da métrica DI.

6 Justificativa

Devido ao fato dos pacientes terem deficiências cognitivas e serem possivelmente idosos, é necessário criar novas técnicas de detecção de fim de sentença e de disfluências que sejam específicas para esse problema. Desse modo, próximos passos de Processamento de Linguagem Natural podem ser aplicados, visto que sentenças são unidades básicas de um texto e várias outras ferramentas se beneficiam disso.

7 Método Científico

Para conseguir realizar as detecções, contamos com um corpus manualmente anotado que contém marcações de disfluências e de segmentação para as narrativas de cada tipo de paciente: Controle, CCL e DA. Além disso, contamos também com o áudio da narrativa, onde o paciente conta em voz alta a narrativa. Esse áudio carrega informações prosódicas que podem ser importantes para as nossas detecções. Por fim, o áudio é alinhado com o texto, de modo que temos a informação de cada fonema para as palavras da narrativa.

Iremos trabalhar com técnicas de *Deep Learning* (DL) para classificar os fins de sentenças e as disfluências. A princípio, contamos com dois modelos: um onde a entrada é o texto representado através de vetores densamente valorados em um espaço n-dimensional; e outro onde a entrada é o áudio representado através de informações prosódicas como *pitch*, energia e duração dos fonemas.

Para ambos os modelos, criamos uma arquitetura que extrai *features* automaticamente através da entrada usando *Convolutional Neural Network* - CNN. Essas *features* são passadas para uma próxima camada que é responsável por lidar com informações sequenciais do texto, isto é, leva em considerações o contexto de uma palavra, nesse caso usamos um *Recurrent Neural Network* especial denominada LSTM, que é capaz de lembrar de previsões passadas para realizar a previsão atual. Essa arquitetura pode ser vista na Figura 2.

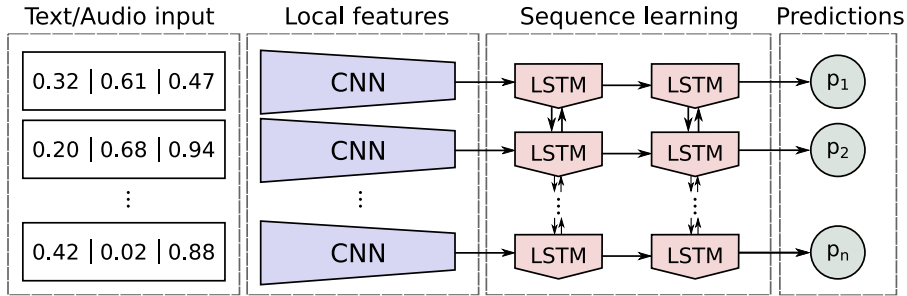


Figura 2: Arquitetura dos modelos de texto e de áudio.

8 Avaliação

Como mencionado anteriormente, pretendemos avaliar nossos modelos num próximo passo do *pipeline* que extrair a métrica de Densidade de Ideias.

Além disso, iremos também utilizar a métrica de F1 para a classe positiva na detecção de fim de sentença, isto é, apenas focar nos pontos finais e não nas outras palavras. E para a detecção de disfluências, pretendemos utilizar a taxa de acertos do modelo.

9 Limitações e Contribuições esperadas

Este trabalho está limitado a trabalhar com textos provindos de narrativas de transcrições de pessoas com a fala comprometida, então provavelmente seu

desempenho será menor em textos bem formados e sem informações prosódicas. Além disso, limitamos nosso escopo para trabalhar com o português.

As contribuições esperadas é que com essa ferramenta é possível automatizar as avaliações de pacientes através do conto de narrativas, e permitindo sua aplicação em larga escala e em uma avaliação longitudinal.