



Projeto de Formatura – 2020 – Press Release

PCs - Departamento de Engenharia de Computação e Sistemas Digitais

Engenharia Elétrica – Ênfase Computação

Tema:

Solução do Desafio de Esquemas de Winograd em Português Brasileiro

Estudantes da Escola Politécnica da Universidade de São Paulo avançam o estado da arte em desafio de inteligência artificial

Nos dias 15 e 16 de dezembro de 2020, os alunos Guilherme Busato Vecchi e Thábata Pontes Lazarou Risso apresentarão o seu trabalho de formatura no curso de Engenharia Elétrica com Ênfase em Computação, desenvolvido sob a orientação do Professor Doutor Glauber de Bona.

Desafios de inteligência artificial são muito importantes para incentivar o desenvolvimento de algoritmos em tarefas específicas, de modo a generalizar os benefícios obtidos para uso em outras aplicações. O Desafio de Esquemas de Winograd (WSC) é um desafio focado na área de processamento de linguagem natural (PLN), composto por um conjunto de frases ambíguas que, para terem a ambiguidade resolvida corretamente, necessitam de conhecimento do senso-comum, intrínseco à natureza humana. O desenvolvimento de soluções a este desafio insere este conhecimento nos algoritmos de PLN, o que é importante para que estes algoritmos possam compreender e emular, de forma mais adequada, a comunicação humana. Este desafio apresenta semelhanças com o Teste de Turing, visando entender a capacidade de pensamento das máquinas.

Implementaram-se duas abordagens de soluções diferentes ao desafio, no intuito de obter uma acurácia superior ao estado da arte nos 273 esquemas do desafio em português brasileiro (45,5%) e de testar a possibilidade de transferência de conhecimento entre idiomas, dado que o estado da arte em inglês é bastante avançado (90,1%), e, com isso, aproximar o desempenho em português do desempenho humano (96,5%). Na primeira abordagem, os esquemas em português são traduzidos para inglês e um modelo de linguagem pré-treinado e com ajuste fino em inglês é executado. Já na segunda abordagem, é utilizado um modelo de linguagem pré-treinado em português mas, para seu ajuste fino, utiliza-se o dataset WINOGRANDE, composto por esquemas semelhantes ao WSC, que passa por um processo de tradução para português.

Em ambas as abordagens o modelo utilizado é o BERT (Bidirectional Encoder Representations from Transformers) que se mostrou muito eficaz por ser bidirecional e apropriado à tarefa de Modelagem de Linguagem mascarada utilizada. Os resultados foram satisfatórios nas duas abordagens, superando o estado da arte em ambas. Na primeira, conseguiu-se chegar a uma acurácia de 72,6% e na segunda, 65,9%.

Integrantes: Guilherme Busato Vecchi
Thábata Pontes Lazarou Risso

Professor(a) Orientador(a): Professor Doutor Glauber de Bona