

Compreensão De Linguagem Natural: Uma solução em português brasileiro*

Augusto Vicente Neto¹, Guilherme Feulo do Espirito Santo², Alfredo Goldman²

¹Faculdade de Filosofia, Letras e Ciências Humanas - Universidade de São Paulo

²Instituto de Matemática e Estatística - Universidade de São Paulo

augusto.vicente@usp.br, {feulo, gold}@ime.usp.br

Abstract. *The A.D.A (Advanced Distributed Assistant) Project aims to develop an open source personal assistant able to interact with users from an ecosystem of devices through voice commands in brazilian portuguese. In this framework, we intend to interpret commands in natural language and translate them to a compilable formal language. To this work, we did a preliminary study in order to investigate among currently available solutions, based on artificial neural networks capable of capturing different kinds of semantic and syntatic relations, if and how, they could contribute to the development of the solution desired by the project.*

Resumo. *O Projeto A.D.A (Assistente Distribuída Avançada)¹ tem como objetivo desenvolver uma assistente pessoal de código aberto capaz de interagir com o usuário a partir de um ecossistema de dispositivos por meio de comandos de voz em português. Dentro desse escopo, visamos interpretar os comandos em linguagem natural e traduzi-los para uma linguagem formal compilável. Para o presente trabalho, fizemos um estudo preliminar a fim de investigar entre as soluções atualmente disponíveis, baseadas em redes neurais artificiais capazes de capturar diversas relações semânticas e sintáticas, se, e como, estas podem contribuir para o desenvolvimento de uma solução nos moldes requeridos pelo projeto.*

1. Introdução

O cenário de desenvolvimento de assistentes virtuais comandadas por voz traz diversos desafios, dentre estes, o de fazer computadores compreenderem comandos em linguagem natural. Estas por sua vez, devido ao seu aspecto simbólico, nos impõem barreiras como ambiguidade, elipse, polissemia, entre outros fatores, que dificultam o acesso imediato ao significado de comandos produzidos na interação humano-computador, sendo necessária a sua tradução para uma linguagem formal.

Em paralelo, a área de Processamento de Linguagem Natural (PLN) já vem, há tempos, buscando formas de representar o significado de palavras, e o resultado desses esforços foram arquiteturas de *embedding* profundas, como o ELMo (Embedding for Language Models) [Peters et al. 2018], e arquiteturas para modelamento de linguagem

*Este projeto foi financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e pelo grupo de extensão CodeLab uclab.xyz/site

¹www.uclab.xyz/ada

genérico a partir de grandes bases de dados, como o OpenAi Transformer (Open Artificial Intelligence Transformer) [Vaswani et al. 2017] e o ULMFiT (Universal Language Model Fine-Tuning for Text Classification) [Howard and Ruder 2018],

Todos esses modelos obtidos através de aprendizado profundo com redes neurais têm em comum o fato de possuir componentes complexos (*encoders, decoders, LSTMs*), associados entre si, geralmente, lidando com uma grande quantidade de dados para desempenhar diversas tarefas, tais como: inferência de linguagem, análise de sentimento, tradução de máquina etc.

Atualmente, existem poucos trabalhos que realizem essa tarefa de tradução de comandos de linguagem natural para uma linguagem formal a partir do Português Brasileiro. Um deles é o [Roman 2001], que realiza essa tarefa baseado nas estruturas do discurso propostas por [Grosz and Sidner 1986] junto da teoria de multiagentes. E também [Luz 2019], centrado na arquitetura de redes *LSTM encoder-decoder*, no entanto, a linguagem formal destino é uma linguagem de consulta, o SPARQL.

Neste cenário, este trabalho tem a intenção de investigar a utilização das tecnologias de modelamento de linguagem para saber se, e de que maneira, contribuem para o desenvolvimento de um módulo de tradução de linguagem natural para uma linguagem formal com o propósito de integrar o Projeto A.D.A.

2. BERT e seus precursores

Como visto, existem algumas alternativas de modelos de linguagem para realizar pré-treinamento, sendo o BERT [Devlin et al. 2018], atualmente, o estado da arte em 11 tarefas de PLN, não surpreendentemente, já que reúne muito do que seus precursores já apresentavam.

O ELMo demonstrou que condicionar ambos os lados, mesmo que treinados independentemente, no processo de *embedding* traz bons resultados. Já o ULMFiT foi um dos primeiros passos na tentativa de solucionar o problema da escassez de *corpora* não etiquetado para tarefas específicas e, finalmente, o OpenAI Transformer permite que isso seja feito com alterações mínimas na arquitetura do modelo e populariza o uso de *transformers* para pré-treinamento.

O BERT junta tudo isso: *transformers (self-attention)*, *transfer learning* e *embedding* bidirecional, com o treinamento em uma tarefa de previsão de palavras a partir da sua posição mascarada, inspirado na tarefa de Cloze [Taylor 1953], bem como uma tarefa de Predição de Próxima Sentença (PPS).

Na primeira etapa, o modelo passa por um treinamento semi-supervisionado em uma grande quantidade de textos (livros, wikipédia). Esse treinamento é feito através de uma tarefa que escolhe algumas posições dos *tokens*, aleatoriamente, com o objetivo de prever a palavra nessa posição. Na segunda etapa, o modelo está pronto para ser treinado em uma tarefa específica como análise de sentimento, equivalência semântica, inferência etc.

A Figura 1 mostra o processo mencionado acima, ilustrando como exemplo de refinamento, uma tarefa de detecção de *spam* em e-mails. O conjunto de dados, ou seja, as sentenças, são etiquetadas como "spam" ou "não spam", mas a incorporação do significado das palavras, que funciona como parâmetro para o reconhecimento de padrões, foi

feito no treinamento robusto da primeira etapa.

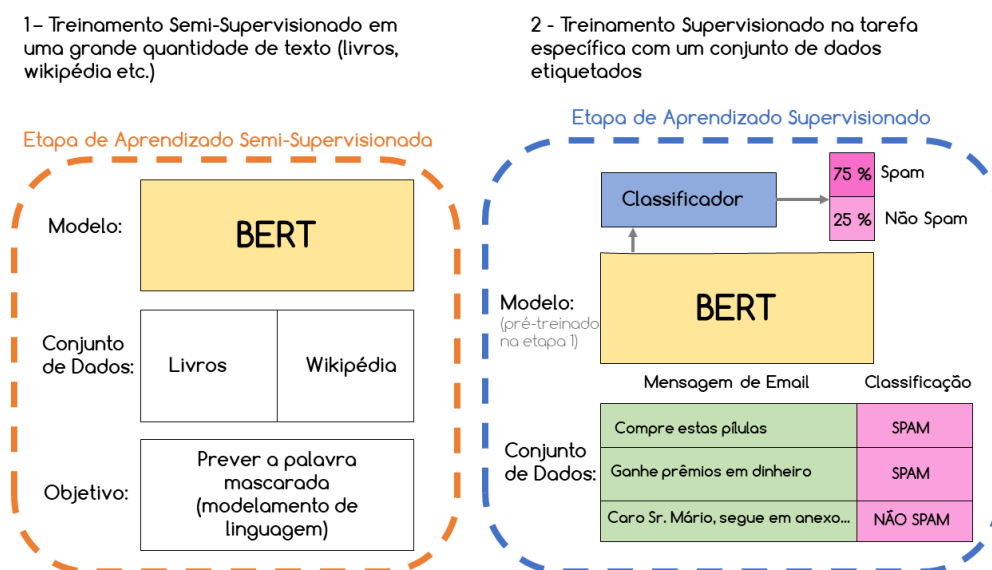


Figura 1. Etapas de treinamento do BERT.

3. Uma solução de Compreensão de Linguagem Natural para o Projeto A.D.A

O Projeto A.D.A - Assistente Distribuída Avançada, propõe a criação de uma assistente virtual voltada para IoT - Internet das Coisas, cuja interação com usuários através de comandos em linguagem natural requer uma solução de Compreensão de Linguagem Natural em Português Brasileiro. A partir dos estudos do modelo de linguagem, pretendemos escolher a opção mais adequada para as necessidades do projeto. Como visto, o BERT tem se mostrado o mais apropriado dado seu processo complexo para captura de relações semânticas e sintáticas capazes de se relacionar com o contexto e, não obstante, oferecendo saídas no nível do *token* e da sentença.

Este, já conta com uma versão em Português Brasileiro disponível [Souza et al. 2019], portanto, só seria necessário realizar a segunda etapa com o treinamento da tarefa específica. Contudo, mesmo que as tarefas específicas atualmente disponíveis interessem aos problemas oriundos das linguagem naturais, como inferência, análise de sentimento, similaridade textual, classificação de sentenças, nenhuma delas é capaz de gerar como saída um comando em uma linguagem formal.

Logo, o BERT se apresenta como a solução ideal para captura dos diferentes usos de uma palavra e até mesmo dependências de longa distância, porém, a tarefa do presente trabalho necessita que, ao receber uma sentença seja possível identificar um único sentido para ela, e em seguida, mapeá-la para um comando interpretável para um computador, restringindo a utilidade do modelo na sua primeira etapa.

4. Próximos Passos

Dessa maneira, os próximos passos seriam investigar como as tarefas disponíveis contribuem para realização do objetivo principal e, em caso negativo, o desenvolvimento de uma arquitetura específica para a realização desse.

Uma hipótese seria a utilização da tarefa de reconhecimento de entidades, que consiste no reconhecimento de entidades nomeáveis (nomes pessoais, organizações, locais, tempo e quantidade). Principalmente com um conjunto de dados formatado no formato CoNLL-2003² que conta com classificações morfológicas, sintáticas, para sintagmas, e por fim, reconhecimento de entidades.

Todas contribuem significativamente para colocação em termos formais: a primeira pode identificar verbos, por exemplo, definindo-o como a ação principal do comando; a segunda, a subordinação de termos relacionados, como identificar que "azul" trata-se de um termo subordinado a "casa" no sintagma "casa azul"; e finalmente, identificar entidades de maneira a distinguir locais que podem ser ambientes das ações do comando, assim como, tempo e quantidade.

Referências

- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Grosz, B. J. and Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational linguistics*, 12(3):175–204.
- Howard, J. and Ruder, S. (2018). Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146*.
- Luz, F. F. (2019). *Deep neural semantic parsing: translating from natural language into SPARQL*. PhD thesis, Universidade de São Paulo.
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. (2018). Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*.
- Roman, N. T. (2001). Estudo de dialogos orientados a tarefa usando a teoria de multia-gentes. Master's thesis, Universidade Estadual de Campinas, São Paulo, Brazil.
- Souza, F., Nogueira, R., and Lotufo, R. (2019). Portuguese named entity recognition using bert-crf. *arXiv preprint arXiv:1909.10649*.
- Taylor, W. L. (1953). Cloze procedure: a new tool for measuring readability. *Journalism Bulletin*, pages 415–433.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.

²<https://www.clips.uantwerpen.be/conll2003/ner/>