

APLICAÇÃO WEB PARA PROCESSAMENTO DE LINGUAGEM NATURAL EM PORTUGUÊS UTILIZANDO BERT

RESUMO EXPANDIDO - Disciplina de TCC29009

João Leonardo Martins

Estudante do Curso de Engenharia de Telecomunicações

Mário de Noronha Neto, Dr.

Professor orientador

Semestre 2022.1

Resumo- *O emprego de técnicas de Inteligência Artificial (IA) tem permitido avanços muito relevantes em nosso cotidiano, nos mais diversos setores: desde assistentes virtuais, passando por anúncios e feeds personalizados em redes sociais, até chegar nos atendentes virtuais. Muito disso passa pela compreensão profunda das intenções de buscas e interações dos usuários, analisado atentamente pelos algoritmos utilizados por empresas, graças ao uso de técnicas de Processamento de Linguagem Natural (PLN). Neste trabalho busca-se entender os conceitos estudados nos principais ramos de IA relacionados ao PLN, com enfoque no algoritmo BERT (Bidirecional Encoder Representations from Transformers), da empresa Google, que teve seu código fonte aberto e é amplamente utilizado no buscador da mesma. De forma a observarmos o uso do BERT na prática, por fim, será desenvolvida uma API para que o modelo possa ser treinado (como é chamada a parte de refinamento do algoritmo) e assim seja capaz de classificar o sentimento por trás de frases em português, as identificando como negativas, neutras ou positivas.*

Palavras-chave: Processamento de Linguagem Natural. BERT. API. Aplicação Web.

1 Introdução

Recentemente pudemos acompanhar avanços importantes em IA, cada vez mais presente no nosso cotidiano, seja em algoritmos de busca, detecção de doenças ou em atendimentos de empresas, através de *chatbots* (ALBAYRAK; ÖZDEMİR; ZEYDAN, 2018). Com a situação econômica de diversas empresas prejudicada pela pandemia, surge

a necessidade de investimentos pontuais que impliquem num maior retorno com a menor contrapartida possível, é quando entra a IA, podendo ser aplicada em publicidades mais nichadas, de forma a impactar um público mais específico e com grande probabilidade de retorno (JAREK; MAZUREK, 2019).

Sempre quando há um avanço tecnológico que possibilite que uma tarefa passe a ser executada por uma máquina, há um receio geral de que pessoas perderão seus empregos. Porém, o que é possível observar é que, num primeiro momento, os frutos da IA são utilizados como auxiliares, como por exemplo, em um pré-atendimento, seja por *web chat* ou através da telefonia convencional. Isso traduz em produtividade que, por sua vez, possibilita a implantação de uma estratégia em voga: o conceito de *Omnichannel*, que é, basicamente, a convergência de todos os canais utilizados por uma empresa (SCHROTENBOER, 2019).

Quando utilizamos os buscadores modernos, fatalmente nos deparamos com IA. E com o que há de mais moderno, pode-se dizer. Por exemplo, quando utilizamos os mecanismos de buscas do Google, estamos em contato com um algoritmo chamado BERT (*Bidirectional Encoder Representations from Transformers*), que é o que há de mais atual em PLN (Processamento de Linguagem Natural) e que permite uma busca muito mais refinada, com resultados muito mais próximos ao que o usuário realmente se refere (daí o conceito de PLN). O autopreenchimento em buscas é mais um recurso possível graças ao emprego do BERT (DEVLIN et al., 2018).

Outro conceito interessante que observamos nos buscadores, porém não apenas nos mesmos, é o de Visão Computacional, que é o ramo de IA que estuda o processamento de imagens pelos computadores. Com isso, é possível realizarmos uma busca a partir de uma figura, ou então garantir a entrada de pessoas (reconhecimento facial), reconhecimento de padrões de modo a evitar acidentes (MARTINEZ; AL-HUSSEIN; AHMAD, 2019), leitura de placas de forma automatizada, entre outras aplicações.

Sobre PLN, como mencionado, houve um avanço bem interessante desde o início da pandemia de Covid-19, em que muitas empresas buscaram formas de automatizar o seu atendimento, fomentando o setor (ZAGO, 2021). Além do emprego em atendimento ao cliente e buscas, como já citado, o PLN pode ser aplicado em tradução de idiomas, filtro de SPAM, reconhecimento de fala, aplicativos assistentes pessoais, entre outros. Tudo isso teve início na década de 1940, quando surge uma maior necessidade, com os aprendizados da Segunda Guerra Mundial, de máquinas que pudessem traduzir de uma língua para outra rapidamente.

Ainda que o conceito de PLN esteja se popularizando e cada vez mais presente em nossas vidas, há muitas possibilidades de aperfeiçoamento e implementação dessa tecnologia, que cada vez mais vem atuando como uma interface entre máquinas e humanos. Assim sendo, gradativamente iremos observar avanços nessa área, seja no aperfeiçoamento no entendimento das solicitações de usuários em buscadores, por exemplo, como também numa maior integração com outros segmentos de IA, além de outras áreas, como IoT (*Internet of Things*).

2 Metodologia

Os tópicos abaixo detalham de que forma chegaremos no resultado final esperado para este trabalho.

2.1 Estudo sobre IA e PLN

Inicialmente, iremos aprofundar os conhecimentos sobre IA em geral, com o objetivo de estabelecer uma visão ampla a respeito dessa área para, em seguida, aprofundarmos o domínio do segmento que dá sentido a este trabalho: Processamento de Linguagem Natural. Faremos um estudo, já dentro de PLN, de *Transformers*, que é uma arquitetura que utiliza modelos de outro ramo da IA, chamado *Deep Learning*, para, basicamente, transformar sequência de elementos de dada sentença em outros, estabelecendo relações entre as palavras.

2.2 Coleta de Dados

Para que o modelo BERT seja refinado, há um etapa importante, que será tratada na seção 2.3, chamada treinamento. Porém, para a obtenção do resultado esperado, deve-se alimentar o algoritmo com uma base de dados que relacione uma avaliação em nota, por exemplo, com um comentário, como é possível observar no próprio *Google, site* em que as empresas podem ser avaliadas pelos usuários, bem como aplicativos em suas respectivas lojas (*Play Store* ou *App Store*).

2.3 Treinamento

Nesta etapa, conforme mencionado anteriormente, é quando há a parte de afinação do modelo. Deve-se relacionar uma avaliação textual a uma métrica já definida. Conforme (DEVLIN et al., 2018), o BERT representa o estado da arte em PLN, isso porque utiliza o modelo *Transformers*, que revoluciona como cientistas trabalham com dados textuais. Isso tudo graças à facilidade de relacionar palavras dentro de uma sentença.

De modo a facilitar a implementação, utiliza-se uma biblioteca *transformers* pré-treinada da comunidade *Hugging Face*, preparada para o idioma português. Também será trabalhado com o modelo BERT já no nosso idioma, facilitando, e muito, a implementação do projeto. Em resumo, a etapa de pré-treinamento já estará pronta, bastando apenas a criação de um *dataset* para completar a etapa de *fine-tuning* ou refinamento.

2.4 API

Por último, será desenvolvida uma API em que será possível submeter o *dataset*, bem como possível obter a análise de sentimento de uma frase, classificando-a em negativa, neutra ou positiva. Tudo isso em cima de uma interface web para facilitar a interação com o usuário.

3 Considerações Parciais

Com o serviço proposto em produção, espera-se facilitar o entendimento da opinião da população a respeito de algum serviço, bem como, por exemplo, o sentimento dos usuários a respeito de algum assunto em redes sociais, exemplificativamente, o *Twitter*.

Referências

ALBAYRAK, N.; ÖZDEMİR, A.; ZEYDAN, E. An overview of artificial intelligence based chatbots and an example chatbot application. In: IEEE. *2018 26th signal processing and communications applications conference (SIU)*. [S.l.], 2018. p. 1–4.

DEVLIN, J. et al. BERT: Pre-training of deep bidirectional transformers for language understanding. out. 2018.

JAREK, K.; MAZUREK, G. Marketing and artificial intelligence. *Central European Business Review*, v. 8, n. 2, 2019.

MARTINEZ, P.; AL-HUSSEIN, M.; AHMAD, R. A scientometric analysis and critical review of computer vision applications for construction. *Automation in Construction*, Elsevier, v. 107, p. 102947, 2019.

SCHROTENBOER, D. The impact of artificial intelligence along the customer journey: a systematic literature review. University of Twente, 2019.

ZAGO, B. *Panorama de Atendimento online pré e pós-pandemia E tendências para 2021*. 2021. Disponível em: <<https://www.ecommercebrasil.com.br/artigos/panorama-de-atendimento-online-pre-e-pos-pandemia-e-tendencias-para-2021/>>. Acesso em: 24 de jul. de 2022.