

Processamento de textos com Few-shot, One-shot, Zero-shot Learning, and Transfer Learning

Universidade do Minho, Braga

Resumo Neste artigo é realizada uma revisão da literatura sobre novas tendências e aplicações relacionadas com processamento de textos utilizando técnicas como *Few-shot Learning*, *One-shot Learning*, *Zero-shot Learning* e *Transfer Learning*. É apresentado o GPT-3, um modelo de linguagem natural pré-treinado que utiliza a técnica de Transformer para gerar texto de alta qualidade em diversas tarefas de processamento de linguagem natural. Além disso, é discutido o BERT, um modelo de linguagem natural pré-treinado desenvolvido pelo Google que se destaca por sua capacidade de compreender o contexto e as nuances do idioma. Ambos os modelos são amplamente utilizados em diversas indústrias para a criação de chatbots, produção de conteúdo automatizado e outras aplicações relacionadas ao processamento de texto.

Keywords: Processamento de texto · *Few-shot Learning* · *One-shot Learning* · *Zero-shot Learning* · *Transfer Learning* · GPT-3 · BERT.

1 Introdução

O artigo aborda as novas tendências e aplicações relacionadas com o processamento de textos utilizando técnicas de aprendizagem profunda, incluindo Few-shot Learning, One-shot Learning, Zero-shot Learning e Transfer Learning. Através da revisão da literatura, o artigo descreve as principais características de cada uma dessas abordagens e suas aplicações práticas no campo do processamento de linguagem natural.

Tem como objetivo fornecer uma visão geral do estado da arte do processamento de textos com técnicas de aprendizagem profunda e apresentar as possibilidades e desafios que essas abordagens oferecem para a criação de soluções de processamento de linguagem natural cada vez mais avançadas.

2 Domínio de Investigação

2.1 *Few-shot Learning*

Few-shot learning é uma abordagem de aprendizagem máquina que visa treinar modelos capazes de generalizar a partir de um conjunto muito pequeno de exemplos de treinamento. Esta abordagem é útil em casos em que os dados de

treinamento são limitados, permitindo treinar modelos com poucos exemplos de cada categoria. Uma das técnicas mais populares é o modelo de redes neurais convolucionais com aprendizagem por siameses, que compara pares de exemplos e identifica semelhanças e diferenças entre eles.

2.2 *One-shot Learning*

One-shot learning é uma abordagem de aprendizagem máquina que visa treinar modelos capazes de realizar uma tarefa de classificação correta após terem sido apresentados a apenas um único exemplo de cada categoria durante a fase de treinamento. Esta abordagem é útil em casos em que o número de exemplos de treinamento é extremamente limitado. É possível implementar o *one-shot learning* utilizando técnicas como redes neurais siamesas, que comparam o exemplo de entrada com cada um dos exemplos de treinamento e calculam a distância entre eles para definir a classe do exemplo de entrada.

2.3 *Zero-shot Learning*

Zero-shot learning é uma abordagem de aprendizagem máquina que visa treinar modelos capazes de realizar tarefas de classificação para categorias que não foram vistas durante o treinamento. Essa abordagem é útil em casos em que é difícil recolher exemplos de treinamento para todas as possíveis categorias de um problema. Para implementar o *zero-shot learning*, são utilizadas técnicas de *transfer learning* e *embedding* de palavras. O modelo é treinado com exemplos de algumas categorias e aprende a representar cada categoria em um espaço semântico, podendo generalizar para novas categorias com base em uma descrição textual ou semântica.

2.4 *Transfer Learning*

Transfer learning é uma abordagem de aprendizagem máquina em que um modelo pré-treinado em uma tarefa é reutilizado como ponto de partida para uma nova tarefa relacionada. Isso é útil em casos em que há uma quantidade limitada de dados de treinamento disponíveis para a nova tarefa ou quando o treinamento de um modelo do zero seria muito demorado ou caro. Existem diferentes técnicas de *transfer learning*, como *fine-tuning* e *feature extraction*, amplamente utilizadas em áreas como visão computacional, processamento de linguagem natural e reconhecimento de fala, entre outras. O conhecimento adquirido pelo modelo pré-treinado em uma tarefa pode ser transferido para outras tarefas relacionadas, acelerando e melhorando o processo de treinamento.

2.5 Técnicas

***Name Entity Recognition*(NER)** *Named Entity Recognition* (NER) é uma técnica de processamento de linguagem natural (PLN) que tem como objetivo

identificar e extrair entidades nomeadas de um texto, tais como nomes de pessoas, organizações, locais, datas, entre outros. O objetivo do NER é identificar os nomes e classificá-los em diferentes categorias, de forma a facilitar a análise do texto e a identificação de informações relevantes. O NER é uma técnica amplamente utilizada em diversas áreas, incluindo processamento de documentos, análise de redes sociais, análise de sentimentos, tradução automática e recuperação de informações.

Sentiment Analysis *Sentiment Analysis*, é uma técnica de processamento de linguagem natural que consiste em identificar a opinião, emoção ou atitude expressa em um texto, geralmente um texto escrito, como uma crítica, um comentário em uma rede social, um review de um produto, entre outros. Esta pode ser feita através da classificação do texto em diferentes categorias, como positivo, negativo ou neutro, ou em outras categorias mais específicas, dependendo do contexto da análise. Essa técnica é muito utilizada em áreas como marketing e atendimento ao cliente para entender a percepção dos consumidores em relação a produtos e serviços, bem como na análise de dados de redes sociais e na avaliação da reputação de empresas e marcas.

Text Classification *Text classification*, é uma técnica de processamento de linguagem natural que consiste em classificar automaticamente um texto em uma ou mais categorias pré-definidas. Isso pode ser feito por meio de algoritmos de aprendizagem máquina, que analisam as características do texto e aprendem a identificar padrões que indicam a categoria à qual ele pertence. Essa técnica é utilizada em diversas aplicações, como análise de sentimentos, detecção de spam, classificação de documentos, entre outras.

Text Generation *Text generation*, é uma tarefa de processamento de linguagem natural que envolve a criação de texto em linguagem natural por meio de um modelo de aprendizagem máquina. Em outras palavras, a ideia é que um computador seja capaz de gerar texto que pareça ter sido escrito por um ser humano. Para fazer isso, o modelo de aprendizagem máquina é treinado em um *dataset* que contém exemplos de texto em linguagem natural e, em seguida, usa esse conhecimento para gerar novas frases e parágrafos de texto. A geração de texto pode ser usada para uma variedade de aplicações, incluindo chatbots, assistentes virtuais, resumos automáticos e até mesmo para criação de conteúdo em redes sociais e marketing digital.

3 GPT-3 (*Generative Pre-trained Transformer 3*)

O GPT-3 é um modelo de linguagem natural pré-treinado criado pela OpenAI que utiliza a técnica de Transformer para gerar texto de alta qualidade em uma variedade de tarefas de processamento de linguagem natural, como tradução, resumo, resposta a perguntas, geração de texto, entre outras.

O modelo GPT-3 é treinado em um grande *dataset* de texto não rotulado usando a técnica de previsão de palavras seguintes, onde o modelo é alimentado com uma sequência de palavras e é treinado para prever a próxima palavra na sequência. Durante o treinamento, o modelo é capaz de aprender padrões nos dados de texto e gerar representações de palavras contextualizadas.

Uma das principais vantagens do GPT-3 é sua capacidade de gerar texto de alta qualidade e coerente, tornando-se uma ferramenta poderosa para a geração de texto em tarefas como a criação de resumos de texto, tradução automática e geração de texto automatizada. Além disso, o GPT-3 é capaz de realizar tarefas de processamento de linguagem natural complexas, como responder perguntas e até mesmo criar histórias e artigos completos.

O GPT-3 é um dos modelos mais avançados e complexos de processamento de texto disponíveis atualmente e tem sido amplamente utilizado em uma variedade de indústrias, desde a criação de chatbots até a produção de conteúdo automatizado para sites e redes sociais.

Investigações

1. O artigo [1] relata sobre o experimento dos autores utilizando o GPT-3 para responder perguntas em dados tabulares. Os pesquisadores descobriram que o GPT-3 foi capaz de aprender a estrutura da tabela a partir de uma representação JSON serializada de matrizes e responder perguntas de busca e perguntas de comparação simples em linguagem natural sem nenhum ajuste fino. Eles também descobriram que a engenharia de prompts, incluindo exemplos estáticos de perguntas e respostas de *few-shot learning*, melhorou significativamente a precisão e a mistura de texto de passagem melhorou ainda mais a precisão em dados heterogêneos. Eles aplicaram sua abordagem a um novo *dataset* de tabelas simples em infográficos de jornais com resultados promissores.

TID	Question	Expected	GPT-3 (Zero-Shot)	GPT-3 (Few-Shot)
1	What is the lowest value in the column "Price per kg as of Monday"?	110	The lowest value in the "Price per kg as of Monday" column is 90.	110
1	Which variety of Mango saw the least hike in price and what is the hike?	Mallika,Imam Pasand, 20	The variety of mango that saw the least price hike is Badami, with a price hike of 40%	Mallika, 20
2	What is the highest strike rate of any player?	91.64	78.04	91.64
2	Sort the players in increasing order of their average scores.	Mithali Raj, Smriti Mandhana, Harmanpreet Kaur	[[Mithali Raj 7, 7, 182, 68, 26.0, 62.97, 0/2], [Harmanpreet Kaur , 7, 318, 109, 53.0, 91.64, 1/2], [Smriti Mandhana , 7, 327, 123, 46.71, 78.04, 1/2]]	Mithali Raj, Smriti Mandhana, Harmanpreet Kaur
3	Which crop has the highest return over cost percentage?	Bajra	Sesame has the highest return over cost percentage at 50%	Bajra
3	What was the MSP of Bajra in "2021-22"?	2250	The MSP of Bajra in 2021-22 was 2350	2,250
4	How many Indian players are above Virat Kohli in terms of strike rate?	1	3	1
4	Which country's player has the highest strike rate?	RSA	The player with the highest strike rate is Aaron Finch of Australia	RSA
5	Which city has the least difference between minimum and maximum temperatures?	Bengaluru	Mysuru	Bengaluru
5	Which city has the least maximum temperature?	Bengaluru	Mysuru	Bengaluru
6	What is the coverage of LPG in Maharashtra?	98%	79.7%	98%

Figura 1. Casos de insucesso para perguntas e respostas de *zero-shot learning*, que são abordados utilizando *few-shot learning*.

Model	Accuracy
TAPAS	18.4
GPT-3 (Zero-Shot)	73.6
GPT-3 (Few-Shot)	86.8

Figura 2. Comparação entre o TAPAS e o GPT-3, com *zero-shot learning* e *few-shot learning* no novo *dataset*.

- O artigo [2] discute o problema dos modelos de processamento de linguagem natural gerarem linguagem prejudicial, que pode direcionar-se a grupos marginalizados com base em sua raça, sexo, orientação sexual e outras características. Estes modelos são frequentemente treinados em *datasets* que contêm declarações falsas, preconceitos humanos e linguagem abusiva. O artigo se concentra em saber se os grandes modelos de linguagem podem identificar ou aprender a identificar linguagem prejudicial. O estudo considera 120 extratos diferentes categorizados como "*racist*", "*sexist*" ou "*neither*" em configurações de categoria única (zero-shot, one-shot e few-shot) e 243 extra-

tos diferentes em configurações few-shot mistas de várias categorias. O estudo pede ao GPT-3 para classificá-los com base no *zero*, *one* e *few-shot learning*, com e sem instrução. Os resultados sugerem que os modelos de linguagem podem ser usados para identificar linguagem abusiva, potencialmente contrariar a produção de linguagem abusiva por humanos e fazer a autovigilância. No artigo utiliza um *dataset* ETHOS, que contém comentários do YouTube e Reddit, e concentra-se em discurso de ódio baseado em raça e gênero. Seleciona 60 comentários sexistas e 60 comentários racistas para criar um *dataset* equilibrado e comparável a 120 comentários classificados como não discurso de ódio. Os resultados mostram que o GPT-3 desempenha bem em identificar discurso de ódio em todas as configurações, mas com a configuração de *few-shot learning* produz a maior precisão.

4 BERT - *Bidirectional Encoder Representations from Transformers*

BERT é um modelo de linguagem pré-treinado desenvolvido pelo Google para o processamento de texto e tarefas de aprendizado de máquina relacionadas à linguagem natural.

Ao contrário de modelos de linguagem pré-treinados anteriores que usavam uma abordagem unidirecional de processamento de texto, o BERT usa um modelo de codificador transformador bidirecional que permite que o modelo veja todo o contexto em uma sequência de texto ao mesmo tempo.

Este modelo é pré-treinado em grandes quantidades de dados não rotulados, o que permite que ele desenvolva uma compreensão profunda da estrutura e do significado da linguagem natural. Este pré-treinamento é então usado para ajustar o modelo em tarefas específicas de aprendizado supervisionado, como classificação de texto, identificação de entidades nomeadas e previsão de palavras ausentes.

O BERT tornou-se numa referência no processamento de texto e alcançou resultados de ponta em vários benchmarks de aprendizagem máquina relacionados à linguagem natural. É amplamente usado em várias aplicações, incluindo chatbots, assistentes virtuais, análise de sentimento, tradução automática e muito mais.

Investigações

1. O artigo [3] concentra-se na tarefa de detecção de sarcasmo, uma tarefa desafiadora no processamento de linguagem natural devido à presença densa de linguagem figurativa em comunidades de mídia social. Os autores estabelecem linhas de base sólidas com base em modelos de linguagem pré-treinados BERT para esta tarefa e propõem melhorar os modelos BERT, ajustando-os em tarefas intermediárias relacionadas antes de ajustá-los na tarefa-alvo de detecção de sarcasmo. Estudam o desempenho de seus modelos BERT em três *datasets* de tamanhos e características diferentes recolhidos de diferentes

plataformas sociais e fornecem uma análise da eficácia do BERT e o *transfer learning* de tarefa intermediária com o BERT na tarefa de detecção de sarcasmo.

É realizado um *fine-tune* do modelo BERT em duas tarefas intermediárias, classificação de sentimento e detecção de emoção, antes de um *fine-tuning* na tarefa alvo de detecção de sarcasmo. Também comparam o desempenho do modelo BERT com modelos neurais padrão, como BiLSTM e CNN, em diferentes *datasets* com diferentes características, recuperados de diferentes sites de redes sociais e com tamanhos diferentes. Os *datasets* usados nos experimentos incluem o Corpus de Sarcasmo V2, SARC e Twitter.

O Corpus de Sarcasmo V2 alcançou resultados de ponta usando um modelo BERT com *transfer learning* como tarefa intermediária, superando modelos anteriores que usavam engenharia de recursos com SVMs. Foi constatado que o *dataset* pré-treinado do IMDB foi benéfico para o *transfer learning* do Corpus de Sarcasmo V2. Houve também evidências de uma correlação entre sarcasmo e sentimento. No conjunto de dados SARC, os melhores modelos superaram modelos de ponta existentes, incluindo aqueles que utilizam CNNs para extrair recursos de conteúdo, sentimento, emoção e personalidade.

2. artigo [4] os autores propõem a realização de várias tarefas de PLN em cenários de *zero-shot* e *few-shot* utilizando uma tarefa original de pré-treinamento do BERT, o *Next Sentence Prediction* (NSP). Ao contrário das técnicas de nível de token, o método de prompt NSP-BERT de nível de frase não precisa fixar o comprimento do prompt ou a posição a ser prevista, permitindo lidar com tarefas como a vinculação de entidades com facilidade. O artigo avaliou o desempenho do NSP-BERT em várias tarefas de compreensão de linguagem natural com *datasets* em inglês e chinês. Foram utilizados baselines como *finetuning*, PET e EFL7 para comparação, e o NSP-BERT teve um desempenho superior no cenário de *zero-shot*, especialmente em tarefas de classificação de vários tópicos. No cenário de *few-shot*, o desempenho do NSP-BERT foi comparável ao do PET na maioria dos conjuntos de dados e muito melhor do que o EFL sem pré-treinamento no conjunto de dados NLI. Em conclusão, o NSP-BERT mostrou um desempenho melhorado e a velocidade de convergência mais rápida em comparação com os baselines em tarefas de compreensão de linguagem natural.

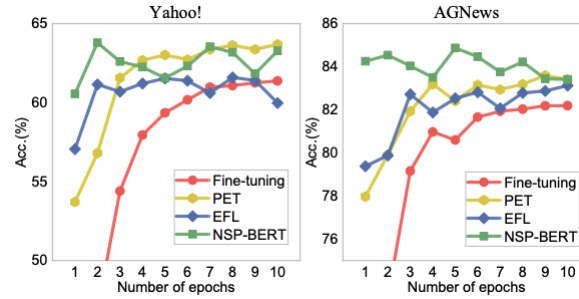


Figura 3. Precisão dos 4 métodos para cada *epoch* durante *few-shot learning* no Yahoo! e AGNews.

5 Conclusão

Os quatro artigos mencionados tratam do uso de aprendizagem máquina para processamento de linguagem natural em diferentes cenários, e todos utilizam abordagens de *few, one, zero-shot learning* ou *transfer learning*. Em geral, essas abordagens são úteis para situações em que a quantidade de dados disponíveis é limitada ou quando o modelo pré-treinado em uma tarefa pode ser adaptado para uma tarefa relacionada.

No artigo [1], o GPT-3 foi utilizado para responder perguntas em dados tabulares com sucesso. Descobriram que incluindo exemplos estáticos de perguntas e respostas de *few-shot learning*, melhorou significativamente a precisão. A utilização do GPT-3 em tarefas relacionadas a tabelas é promissora, uma vez que muitas informações valiosas estão contidas em tabelas e base de dados.

O artigo [2] destaca o problema do viés humano nos modelos de processamento de linguagem natural, que muitas vezes levam a uma linguagem prejudicial. O estudo descobriu que os modelos de linguagem podem ser usados para identificar linguagem abusiva, potencialmente contrariando a produção de linguagem abusiva por humanos e fazer a autovigilância.

O artigo [3] concentra-se na detecção de sarcasmo em redes sociais, que é uma tarefa desafiadora devido à presença densa de linguagem figurativa. Os autores estabeleceram linhas de base sólidas com base em modelos de linguagem pré-treinados BERT para esta tarefa e propuseram melhorar os modelos BERT, ajustando-os em tarefas intermediárias relacionadas antes de ajustá-los na tarefa-alvo de detecção de sarcasmo.

O artigo [4] utiliza o BERT para lidar com o problema de tradução de sentenças de uma língua para outra. Ele explora o uso do *zero-shot* e do *few-shot learning* para a tarefa de tradução, utilizando modelos pré-treinados de diferentes línguas. O estudo descobriu que o *zero-shot* e o *few-shot learning* podem ser usados para a tradução, mesmo entre línguas sem relação direta entre si, o que é um resultado interessante.

Referências

1. Srivastava P., Ganu T., Guha S, Towards Zero-Shot and Few-Shot Table Question Answering using GPT-3(2022)
2. Chiu K., Collins A., Alexander R., Detecting Hate Speech with GPT-3(2022)
3. Savini E., Caragea C., Intermediate-Task Transfer Learning with BERT for Sarcasm Detection(2022)
4. Sun Y., Zheng Y., Hao C., Qiu H., NSP-BERT: A Prompt-based Few-Shot Learner Through an Original Pre-training Task-Next Sentence Prediction(2021)