

APLICAÇÃO DE REDES NEURAIS EM GRAFOS COM FOCO PARA ANÁLISE DE SENTIMENTO EM NÍVEL DE ASPECTO

GABRIEL A. GOMES¹, LARISSA A. FREITAS², ULISSES B. CORRÊA³

¹Universidade Federal de Pelotas - gagomes@inf.ufpel.edu.br

²Universidade Federal de Pelotas - larissa@inf.ufpel.edu.br

³Universidade Federal de Pelotas - ulisses@inf.ufpel.edu.br

1. INTRODUÇÃO

Com tantos dados em formato textual disponíveis, se torna desejável, e até mesmo necessário, desenvolver técnicas para automatizar a extração de informações importantes. Dados como orientação do sentimento do autor, extração de tópicos, criação de resumos de textos muito grandes, entre diversas outras características que podemos retirar deste formato.

Processamento de Língua Natural (PLN) é uma vertente da área de Inteligência Artificial (IA) que busca extrair essas e outras informações de dados em texto, e assim agregar mais valor para esses dados.

Dentro da área de PLN, encontra-se a Análise de Sentimento (AS). Esse campo de PLN foca em identificar opiniões e sentimentos do autor em relação à uma entidade (FELDMAN, 2013). A importância dessa área se dá para a automação na tarefa de processar computacionalmente comentários, *reviews* de produtos e serviços. Podendo melhorar a forma em que empresas, prestadores de serviços, vendedores e até consumidores têm de visualizar *feedbacks* sobre um produto, criando formas de visualização destes comentários mais dinâmicas, tornando possível a criação de algo como um "resumo" das opiniões referidas ao item de interesse.

É possível fazer a AS em diferentes níveis de avaliações, como em nível de documento, de sentença e de aspecto (BEHDENNA; BARIGOU; BELALEM, 2018). A análise em nível de documento tem como objetivo estimar a orientação do sentimento para o documento inteiro, precisando assim assumir que cada documento expressa uma única opinião para uma única entidade. Para a análise em nível de sentença, é necessário primeiro verificar se a sentença expressa uma opinião, e então estimar a polaridade dessa opinião expressada.

Porém, um dos problemas da AS em nível de sentença e documento é que ao atribuir uma única orientação de sentimento podemos suprimir informações, em uma sentença, por exemplo, podemos nos referir a mais de uma entidade e com diferentes opiniões. Uma forma de contornar este problema é fazer a AS em nível de aspectos. Aspectos são todas as entidades que são possíveis de se atribuir uma opinião ou sentimento (LIU, 2012).

Na AS em nível de aspectos, é realizado primeiramente uma identificação de todos os possíveis aspectos em um documento, após isso é realizada a estimativa da polaridade da opinião, uma frase para exemplificar isso em um contexto de *reviews* de um hotel é "O quarto é bom, porém a cama é dura", nesta frase o autor se refere a dois aspectos, sendo eles "quarto" e "cama", assim como diferentes opiniões para cada aspecto, no caso da "cama" a polaridade da opinião é negativa e do "quarto" a polaridade da opinião é positiva.

Agradecemos à CAPES, à FAPERGS e à NVIDIA Corporation pelo financiamento parcial deste trabalho.

Indo além, também é possível agregar mais conhecimento aos dados por meio de outras representações do texto, como textos são formados seguindo uma sintaxe, podemos utilizar as relações linguísticas entre as palavras como mais uma informação de entrada, e uma boa forma de fazer essa representação é por meio de grafos.

GNNs (do Inglês - Graph Neural Networks) são redes neurais que utilizam grafos como formato de entrada, podendo assim serem utilizadas para diversos campos de problemas, envolvendo imagens, textos ou até mesmo ambos. Textos aparentam poder fazer um bom uso desse tipo de abordagem, pois em uma transformação de texto para grafos podemos representar as relações linguísticas através das arestas dos nodos, que por sua vez vão ser as representações das palavras, criando assim novas informações sobre os dados para a classificação.

Ainda existem poucos trabalhos na literatura sobre a utilização de GNN para a tarefa de ABSA, principalmente voltadas para a língua Portuguesa. Portanto, o trabalho propo analisar e avaliar a utilização de uma adaptação de uma GNN, com resultados estado da arte para a tarefa de ABSA na língua Inglesa, com a tarefa de ABSA na língua Portuguesa.

2. METODOLOGIA

Nossa abordagem consiste na transformação dos dados textuais para grafos e, após esta etapa, a utilização destes grafos junto com os aspectos presentes em cada *review* para o treinamento da GNN.

O *dataset* escolhido para a aplicação da metodologia é criado a partir dos trabalhos de Corrêa (2021) e Freitas (2015). Existem poucos conjunto de dados para o português brasileiro com aspectos já anotados e classificados quanto à orientação de sentimento, por isso, selecionamos esses dois conjuntos, que se encaixam nesta categoria.

O *dataset* é formado por *reviews* de hotéis de Nova Iorque, Las Vegas e Paris. Todos os dados foram extraídos do site TripAdvisor, estão escritos em Português e possuem pelo menos 300 caracteres. Cada anotação possui a polaridade da orientação de sentimento ligada ao aspecto em questão. A polaridade pode assumir os valores de 1 se a orientação de sentimento expresso em relação ao aspecto for positiva, 0 se for neutra e -1 se for negativa. O Conjunto de dados possui 3.797 anotações, de 1.031 *reviews* com 77 aspectos.

É importante informar que o *dataset* é desbalanceado em relação às polaridades dos aspectos, com mais anotações positivas do que negativas e neutras somadas. Por isto, é preciso pensar em técnicas para lidar com o desbalanceamento e utilizar métricas específicas para esse tipo de distribuição.

Uma possível forma de lidar com o desbalanceamento dos dados é com o uso de *Oversampling*. Essa técnica consiste em replicar valores aleatórios presentes nas classes minoritárias até balancear com a classe majoritária. Vale salientar que esse processo só é realizado com os dados destinados a treinamento, que no nosso caso, representa 80% do total de exemplos.

Após a expansão de dados, é realizada a etapa de representação em grafos. A forma que adotamos para esta representação é dada por:

1. Cada nodo do grafo representa uma palavra.
2. Cada aresta representa a relação linguística entre palavras.

Os grafos foram gerados utilizando a biblioteca spaCy (HONNIBAL; MONTANI, 2017), que possui diversas ferramentas para NLP, inclusive a de geração de árvores de dependência sintática, utilizada para a criação dos grafos. Além disso, utilizamos um modelo GloVe treinado com dados em português (HARTMANN et al., 2017) para a criação dos *embeddings*.

A arquitetura de GNN selecionada para a tarefa é a DualGCN (LI et al., 2021), que foi escolhida por ter demonstrado ótimos resultados, em comparação ao seus modelos *baselines*, para tarefas de ABSA utilizando *datasets* em inglês. Esta arquitetura recebeu o nome de *DualGCN* por seus autores.

Esta arquitetura é composta por três módulos diferentes, o primeiro módulo, chamado *SynGCN*, é responsável pela análise sintática dos dados, fazendo uso de um analisador de dependência estado da arte. O segundo módulo, conhecido como *SemGCN*, observa as correlações semântica entre os nodos, utilizando um mecanismo de atenção. E o último módulo é encarregado de trocar informações entre os dois primeiros módulos, por meio de uma função *BiAffine*. A função de perda utilizada pela arquitetura foi criada pelos próprios autores, e consta com dois regularizadores, o ortogonal e diferencial. Utilizamos o método *hold-out validation* para a avaliação do modelo final.

3. RESULTADOS E DISCUSSÃO

A competição *Aspect-Based Sentiment Analysis in Portuguese* (ABSAPT) (SILVA et al., 2022), realizada em 2022, sob organização do Hub de Inovação em Inteligência Artificial, da Universidade Federal de Pelotas, teve como foco as subtarefas de ABSA, Extração de Aspectos e Determinação da Orientação do Sentimento. A competição foi realizada utilizando dados de *reviews* do setor de acomodações extraídos do site TripAdvisor.

Neste trabalho, utilizamos o mesmo conjunto de dados disponibilizados pela organização do desafio, tanto os dados de treino como os dados de teste. Na Tabela 1 apresentamos uma comparação dos resultados obtidos com as quatro abordagens com melhor desempenho na competição, considerando a tarefa de Determinação da Orientação de Sentimento.

Tabela 1: Resultados Obtidos no Conjunto de Testes.

Nome	Bacc	F1	Precision	Recall
<i>Team Deep Learning Brasil</i>	0,82	0,81	0,81	0,82
<i>Team PiLN</i>	0,78	0,77	0,76	0,78
<i>DualGCN_PT_over</i>	0,75	0,75	0,76	0,75
<i>DualGCN_PT</i>	0,74	0,74	0,76	0,74
<i>Team UFSCAR</i>	0,62	0,61	0,65	0,62
<i>Team UFPR</i>	0,62	0,61	0,65	0,62

Os modelos gerados pela nossa metodologia superaram o terceiro e quarto colocados da competição. Porém, não conseguimos superar o primeiro e segundo colocados. Podemos notar uma melhora com a utilização da técnica *Oversampling* para o balanceamento dos dados, um comportamento que nos era esperado, dado o desbalanceamento dos dados.

4. CONCLUSÃO

Este trabalho teve como objetivo analisar a adaptação de uma arquitetura de GNN aplicada para a tarefa de ABSA na língua Inglesa, conhecida como *DualGCN*, para a língua Portuguesa. Foram explorados alguns tipos de expansão de dados, visto que, os dados eram desbalanceados e isso acarreta em uma maior representação de uma classe e menor representação das demais.

Nossa abordagem gerou modelos com métricas superiores as do terceiro colocado na competição ABSAPT, porém inferiores as do segundo.

5. REFERÊNCIAS

BEHDENNA, S.; BARIGOU, F.; BELALEM, G. Document level sentiment analysis: A survey. **EAI Endorsed Transactions on Context-aware Systems and Applications**, v. 4, p. 154339, 03 2018.

CORRÊA, U. B. **Análise de sentimento baseada em aspectos usando aprendizado profundo: uma proposta aplicada à língua portuguesa**. 123 p. Tese (Doutorado em Computação) — Universidade Federal de Pelotas, Pelotas, 2021.

FELDMAN, R. Techniques and applications for sentiment analysis. **Commun. ACM**, Association for Computing Machinery, New York, NY, USA, v. 56, n. 4, p. 82–89, apr 2013. ISSN 0001-0782. Disponível em: <<https://doi.org/10.1145/2436256.2436274>>.

FREITAS, L. A. d. **Feature-level sentiment analysis applied to brazilian portuguese reviews**. 94 p. Tese (Doutorado em Ciência da Computação) — Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2015.

HARTMANN, N.; FONSECA, E.; SHULBY, C.; TREVISO, M.; RODRIGUES, J.; ALUISIO, S. **Portuguese Word Embeddings: Evaluating on Word Analogies and Natural Language Tasks**. arXiv, 2017. Disponível em: <<https://arxiv.org/abs/1708.06025>>.

HONNIBAL, M.; MONTANI, I. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear. 2017.

LI, R.; CHEN, H.; FENG, F.; MA, Z.; WANG, X.; HOVY, E. Dual graph convolutional networks for aspect-based sentiment analysis. In: **Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)**. Online: Association for Computational Linguistics, 2021. Disponível em: <<https://aclanthology.org/2021.acl-long.494>>.

LIU, B. **Sentiment Analysis and Opinion Mining**. [S.l.]: Morgan & Claypool Publishers, 2012.

SILVA, F. L. V. da; XAVIER, G. da S.; MENSENBURG, H. M.; RODRIGUES, R. F.; SANTOS, L. P. dos; ARAÚJO, R. M.; CORRÊA, U. B.; FREITAS, L. A. de. Absapt 2022 at iberlef: Overview of the task on aspect-based sentiment analysis in portuguese. **Procesamiento del Lenguaje Natural**, v. 69, n. 0, p. 199–205, 2022. ISSN 1989-7553. Disponível em: <<http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6440>>.