

EXTRAÇÃO DE ASPECTOS EM PORTUGUÊS UTILIZANDO BERT

ÉMERSON PHILIPPE LOPES¹; LARISSA ASTROGILDO DE FREITAS²

¹UFPel – eplopes@inf.ufpel.edu.br

²UFPel – larissa@inf.ufpel.edu.br

1. INTRODUÇÃO

Análise de sentimentos é o campo de estudo que analisa as opiniões, sentimentos, atitudes e emoções de pessoas sobre entidades e seus atributos em textos (LIU, 2015). A análise de sentimentos é estudada em 3 níveis de granularidade: nível de documento, nível de sentença e nível de aspecto (ZHANG; LIU, 2014).

Em níveis de documento ou sentença, considera-se que todo o texto contém uma única opinião, e a tarefa é classificar a polaridade (BECKER; TUMITAN, 2013). No nível de aspectos, é necessário descobrir quais aspectos de uma entidade estão sendo comentados no texto, e a partir disso encontrar suas polaridades. Para isso, existem duas principais tarefas: extração de aspectos e classificação da polaridade (LIU, 2015). Este trabalho se concentra apenas na extração dos aspectos.

Extração de aspectos é a primeira das etapas da análise de sentimentos baseada em aspectos (ABSA). É considerada uma tarefa importante porque falhas nela podem gerar erros nas outras fases (PEREIRA, 2020). Essa etapa consiste em identificar os termos que representam os aspectos presentes em um documento. Diferentes técnicas, baseadas em léxico ou em Aprendizado de Máquina (AM), podem ser usadas nesta tarefa. Dentre as técnicas de AM, o BERT (*Bidirectional Encoder Representations from Transformers*), desenvolvido por DEVLIN et al. (2018), conseguiu obter os melhores resultados nos últimos anos em várias tarefas de Processamento de Língua Natural (PLN).

O BERT é um modelo de linguagem pré-treinado, que é treinado originalmente em vários textos não anotados, e usa *transfer learning* para ser treinado mais rapidamente para várias tarefas, precisando de menos textos anotados.

O propósito deste trabalho é avaliar o desempenho do BERT na tarefa de ABSA, especificamente na extração de aspectos em textos em português, usando o modelo pré-treinado para o idioma, disponibilizado por SOUZA et al. (2019).

2. METODOLOGIA

Neste trabalho foi utilizado o *dataset* do TripAdvisor desenvolvido por (Freitas e Vieira 2015), o qual contém 194 *reviews* de 10 hotéis, anotados no nível de aspecto, com suas respectivas polaridades. Para usar o classificador de pares de sentenças do BERT, foi usado o *review* como primeiro texto e o aspecto como segundo texto. O dataset foi dividido em 85% para o treinamento, e 15% para teste.

Os aspectos presentes no dataset para um *review* foram anotados como *related*, e os que não existiam para aquele aspecto foram anotados como *unrelated*. Por causa dessa geração dos aspectos não relacionados ao *review*, o dataset ficou muito desbalanceado. Para compensar isso, foram usados pesos diferentes para cada label, baseado em MADABUSHI et al. (2019)

Os testes foram feitos usando os modelos *base-multilingual-cased* e *base-portuguese-cased*. Os modelos *cased* usam textos que podem conter letras maiúsculas e acentuação. Embora esses modelos geralmente obtenham resultados piores, os modelos em português não tem versões *uncased*, então para que os modelos em Português e Multilingue pudessem ser comparados, ambos foram usados nessa versão. Isso reduz uma etapa do pré-processamento dos dados, que podem ser usados na sua forma original.

Os modelos *base* são as menores versões do BERT que foram inicialmente disponibilizadas, e a menor disponível em português. Esses modelos obtém resultados um pouco piores, porém foram usados devido a menor necessidade de memória.

3. RESULTADOS E DISCUSSÃO

Os valores preditos são divididos entre os verdadeiro positivo (*vp*), nos quais o aspecto é relacionado ao texto, e o modelo acerta, os verdadeiro negativo (*vn*), nos quais o aspecto não é relacionado, e o modelo acerta, os falso positivo (*fp*), nos o modelo classifica como relacionado um aspecto que não é, e os falso negativo (*fn*), nos quais o modelo classifica não relacionado um aspecto que é.

Para avaliar os resultados, foram usadas métricas tradicionais apresentadas na literatura. Essas métricas utilizadas para avaliar os resultados foram a precisão, revocação, medida f e acurácia. A precisão é a proporção de casos que foram classificados como positivos que realmente são. A revocação é a proporção de positivos que foram classificados corretamente. A medida f é a média harmônica entre a precisão e a revocação. A acurácia é a proporção dos acertos do modelo em relação ao total de exemplos.

Na Tabela 1 são comparados os resultados obtidos para os modelos em Português e Multilingue, treinados em 4 e 8 épocas, usando *batches* de tamanho 16, sentenças de até 128 *tokens*, e pesos 20, para *unrelated* e 1 para *related*.

Modelo	Épocas	Precisão	Revocação	Medida f	Acurácia
Português	8	85.61	86.23	85.92	92.49
Português	4	90.19	66.67	76.67	89.21
<i>Multilingual</i>	8	94.68	64.49	76.72	89.59
<i>Multilingual</i>	4	93.83	55.07	69.41	87.09

Tabela 1 - Comparação entre modelos Português e Multilingue

Esses resultados mostram que o modelo em Português consegue obter desempenho significativamente melhor que o modelo Multilingue.

Na Tabela 2 são apresentados os resultados obtidos com tamanho de sentença maior (256 e 512 *tokens*), usando 4 e 8 épocas, e pesos 20 para *unrelated* e 1 para *related*. Devido ao grande consumo de memória para sentenças maiores, foram usados *batches* de tamanho 3, para sentenças de 512 *tokens*, e 6, para sentenças de 256 *tokens*.

Tam. Sentença	Épocas	Precisão	Revocação	Medida f	Acurácia
512	8	91.49	93.48	92.47	95.95
512	4	87.78	88.41	88.09	93.64
256	8	91.49	93.18	92.47	95.95
256	4	88.33	76.81	82.17	91.14

Tabela 2 - Comparaçāo entre diferentes tamanhos de sentença

Esses resultados indicam que tamanhos de sentença maiores, quando necessários para o *dataset*, resultam em melhorias nos resultados obtidos.

4. CONCLUSĀOES

Os resultados deste trabalho indicam que o BERT, principalmente treinado para o Português, consegue obter ótimos resultados para a etapa de extração de aspectos na tarefa de ABSA. Mais testes, com *datasets* maiores e mais variados podem obter resultados diferentes, porém existem poucos *datasets* disponíveis em português anotados no nível de aspecto. Como trabalhos futuros pretendemos avaliar o desempenho dos modelos Português e Multilingue na etapa de classificação de polaridade na tarefa de ABSA.

5. REFERĀNCIAS BIBLIOGRÁFICAS

LIU, B. **Sentiment analysis: mining opinions, sentiments, and emotions**. Cambridge University Press, 2015.

ZHANG, L.; LIU, B. **Aspect and Entity Extraction for Opinion Mining**. Berlim: Springer, 2014.

BECKER, K.; TUMITAN, D. Introdução à Mineração de Opiniões: Conceitos, Aplicações e Desafios. In: FERREIRA, J. E. (Org.). **Lectures of the 28th Brazilian Symposium on Databases**. 1ed. Pernambuco: CIN - UFPE, 2013, v., p.27-52.

PEREIRA, D. A survey of sentiment analysis in the Portuguese language. **Artificial Intelligence Review**. Dordrecht, online, p. 1-29, 2020.

DEVLIN, J; CHANG, M.; LEE, K; TOUTANOVA, K; BERT: Pre-Training of deep bidirectional transformers for language understanding. **arXiv preprint arXiv:1810.04805**, 2018.

SOUZA, F.; NOGUEIRA, R.; LOTUFO, R.; Portuguese Named Entity Recognition using BERT-CRF. **arXiv preprint arXiv:1909.10649**, 2019.

FREITAS, L.A.; VIEIRA, R. Exploring Resources for Sentiment Analysis in Portuguese Language. In: **2015 BRAZILIAN CONFERENCE ON INTELLIGENT**

SYSTEMS (BRACIS), Natal, 2015. 2015 Brazilian Conference on Intelligent Systems (BRACIS), 2015. p. 152-5

MADABUSHI, T.; KOCHKINA, E.; CASTELLE, M.; Cost-Sensitive {BERT} for Generalisable Sentence Classification on Imbalanced Data. In: **SECOND WORKSHOP ON NATURAL LANGUAGE PROCESSING FOR INTERNET FREEDOM: CENSORSHIP, DISINFORMATION, AND PROPAGANDA**, Hong Kong, 2019. Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda. Hong Kong: Association for Computational Linguistics, 2019. p. 125-134