

Evaluating an Aspect Extraction Method for Opinion Mining in the Portuguese Language

Breno Gomes Cardoso, Denilson Alves Pereira

Departamento de Ciência da Computação
Universidade Federal de Lavras
breno.cardoso@estudante.ufla.br
denilsonpereira@ufla.br

Abstract. The opinion issued by consumers of products and services has become increasingly valued, both by other consumers and by companies. The automatic interpretation of review texts to generate information is of paramount importance. With opinion mining at the aspect level, it is possible to extract and summarize opinions about different components of a product or service. This paper evaluates the behavior of a method for extracting aspects using natural language processing tools for the Portuguese language. The aim is to investigate the maturity of the tools for Portuguese compared to the already consolidated tools for the English language. The evaluation was carried out in three datasets from two different domains with original texts in Portuguese and their translations into English, and vice versa, and the results indicate that there is no difference between languages.

CCS Concepts: • **Information systems** → **Sentiment analysis**; • **Computing methodologies** → **Information extraction**.

Keywords: aspect extraction, opinion mining, Portuguese language, sentiment analysis

1. INTRODUÇÃO

Com o crescimento do número de pessoas com acesso à Internet, tem se tornado muito comum a publicação de opiniões pessoais sobre diversos assuntos, tais como política, eventos, filmes, livros, hotéis e produtos diversos. Essas opiniões podem, por exemplo, influenciar eleitores na escolha de um candidato em uma disputa política, auxiliar consumidores na compra de um produto ou na contratação de um serviço. Opiniões também são de grande interesse das empresas, como forma de melhorar seus produtos, manter um bom relacionamento com seus clientes ou definir estratégias de marketing. Entretanto, coletar um grande número de opiniões e analisá-las de forma automática, com objetivo de extrair conhecimento, não é uma tarefa trivial.

A análise de sentimentos é uma área de estudos que visa usar recursos computacionais para analisar opiniões, sentimentos e emoções, os quais podem ser capturados a partir de texto, expressão facial, tom de fala, entre outros [Poria et al. 2016]. A análise de sentimentos pode ser dividida em duas tarefas: mineração de emoção, cujo objetivo é identificar as emoções emitidas (ex: alegria, tristeza, raiva), e mineração de opinião, a qual objetiva analisar as opiniões, identificando sua polaridade como um grau de positividade ou negatividade [Yadollahi et al. 2017].

A mineração de opinião em um texto pode ser analisada em três níveis de granularidade: documento, sentença e aspecto [Liu 2012]. A nível de documento, é analisada a opinião do texto como um todo. A nível de sentença, é analisada separadamente a opinião emitida em cada sentença do documento. E a nível de aspecto, o objetivo é identificar a opinião expressa sobre cada aspecto de uma entidade,

Esse trabalho foi parcialmente financiado pela FAPEMIG e pela UFLA.

Copyright©2020 Permission to copy without fee all or part of the material printed in KDMiLe is granted provided that the copies are not made or distributed for commercial advantage, and that notice is given that copying is by permission of the Sociedade Brasileira de Computação.

por exemplo, a limpeza de um hotel.

A nível de aspecto, uma opinião pode ser definida pela quintupla (e, a, s, h, t) , onde e é a entidade, a é um dos aspectos da entidade e , s é o sentimento em relação ao aspecto a , h é o autor da opinião, e t é tempo quando a opinião foi emitida [Liu 2012]. Por exemplo, na frase “O som dessa televisão é estourado”, o autor expressa o sentimento “estourado” em relação ao aspecto “som” da entidade “televisão”. Nesse exemplo, o autor e o tempo não estão explícitos, mas poderiam ser identificados na etapa de coleta de dados.

Identificar e extrair aspectos é basicamente uma tarefa de extração de informação. Existem diferentes abordagens para essa tarefa, e elas podem ser agrupadas em três categorias, de acordo com a ferramenta utilizada: regras de linguagem, modelo sequencial e modelo de tópico [Zhang and Liu 2014]. A abordagem de regras de linguagem utiliza os padrões de escrita e as relações entre as palavras, o modelo sequencial utiliza a posição das palavras na frase, e o modelo de tópicos considera que um documento é conjunto de tópicos diferentes.

Os métodos de análise de sentimentos normalmente utilizam-se de ferramentas de Processamento de Linguagem Natural (PLN) para processar textos e identificar atributos que auxiliam na análise. Muitas dessas ferramentas e recursos linguísticos são dependentes de idioma, e como consequência, algumas línguas possuem ferramentas mais maduras que outras [Lo et al. 2017]. A maioria das pesquisas em análise de sentimentos está voltada para a língua Inglesa [Chaturvedi et al. 2018] [Cambria et al. 2018] [Zhang et al. 2018]. Para a língua portuguesa, apesar dos avanços nos últimos anos, ainda existe muito espaço para melhorias [Pereira 2020].

Para a tarefa de extração de aspectos, alguns métodos foram desenvolvidos para a língua portuguesa [Catharin and Feltrim 2018] [Machado et al. 2017] [Saiaes et al. 2018]. [Farias et al. 2016] avaliaram o comportamento de um método de identificação de aspectos em uma base de dados de produtos com textos em português e sua respectiva tradução automática para o inglês. O experimento mostrou que o uso do método com uma ferramenta de PLN para inglês aplicado à base de dados traduzida obteve melhor resultado do que o mesmo método aplicado à base de dados original em português, o que pode indicar que as ferramentas para a língua portuguesa ainda não estão tão maduras quanto aquelas para a língua Inglesa.

Este trabalho avalia o comportamento de um método supervisionado para extração de aspectos usando ferramentas de PLN para português e inglês. O objetivo é avaliar a maturidade das ferramentas em português para extração de aspectos. O método implementado foi proposto por [Jakob and Gurevych 2010], e utiliza um modelo sequencial baseado em *Conditional Random Fields* (CRF) [Lafferty et al. 2001]. Em [Farias et al. 2016], as bases de dados de teste foram traduzidas por meio de um tradutor automático. Entretanto, foi usada apenas uma pequena base de dados de produtos com texto em português traduzido para o inglês. Neste trabalho, nós avaliamos nossa implementação em três bases de dados, sendo duas em português e uma em inglês. As bases de dados em português foram traduzidas para o inglês, e nós também avaliamos o processo inverso, traduzindo a base de dados em inglês para o português. Os resultados indicam um empate estatístico em duas bases de dados e uma pequena melhora em outra com textos em português, sugerindo que não vale a pena traduzir os textos para extração de aspectos na língua portuguesa.

Uma contribuição adicional deste trabalho é a criação de uma das bases de dados em português e os ajustes na tradução. Esses ajustes se referem aos nomes dos aspectos, pois na tradução, os nomes que estão nos gabaritos podem ser traduzidos de forma diferente dos mesmos nomes que estão no texto, requerendo uma correção manual para facilitar a verificação automática. Assim, deixamos disponíveis para a comunidade científica as bases de dados em português e em inglês, as quais podem ser úteis em outros experimentos com extração de aspectos.

2. TRABALHOS RELACIONADOS

A pesquisa em análise de sentimentos para a língua portuguesa tem avançado bastante nos últimos anos. [Pereira 2020] apresenta uma ampla cobertura dos trabalhos desenvolvidos e descreve os principais recursos disponíveis para a língua portuguesa, tais como ferramentas de PLN, léxicos, ontologias e bases de dados.

Neste trabalho, avaliamos ferramentas em português para extração de aspectos. A maturidade de ferramentas de análise de sentimentos para outros idiomas diferentes do inglês já foi objeto de pesquisa de alguns trabalhos [Araújo et al. 2020] [Farias et al. 2016]. Para análise a nível de sentença, [Araújo et al. 2020] avaliaram métodos em inglês e em outras línguas. Bases de dados de quatorze idiomas, incluindo o português, foram traduzidas, de forma automática, para o inglês. Os resultados sugerem que traduzir as bases de dados e aplicar métodos desenvolvidos para o inglês é mais eficaz do que utilizar os métodos desenvolvidos para as línguas nativas. Na análise a nível de aspectos, [Farias et al. 2016] também identificaram resultados que indicam vantagem para a tradução dos textos para o inglês. Eles apresentaram uma ferramenta para extração de aspectos e classificação da polaridade da opinião para quatro idiomas, incluindo o português. O método implementado para extração de aspectos identifica os substantivos mais frequentes e extrai aqueles com uma frequência mínima especificada.

Alguns trabalhos implementaram abordagens para extração de aspectos na língua portuguesa. [Machado et al. 2017] analisaram abordagens não supervisionadas para extrair aspectos em revisões de livros. As abordagens se baseiam na identificação da frequência dos substantivos e estratégias de podas. Os melhores resultados foram obtidos com o uso de Word2Vec para processar termos infrequentes. [Freitas and Vieira 2019] trataram a análise de sentimentos a nível de aspectos usando uma abordagem baseada em ontologia para identificar aspectos explícitos e implícitos. O trabalho usou diferentes ferramentas de PLN para classificação gramatical de palavras, léxicos e regras linguísticas. [Catharin and Feltrim 2018] e [Saias et al. 2018] apresentaram propostas de extração de aspectos baseadas em regras de linguagens. [Aires et al. 2018] usaram uma abordagem baseada em *deep learning* para identificar e classificar aspectos. O trabalho usa uma lista fixa de nomes de aspectos e, assim, não é capaz de descobrir novos aspectos.

O trabalho de [Jakob and Gurevych 2010] teve como objetivo avaliar uma abordagem baseada em CRF para extração de aspectos, comparando o desempenho quando o método é treinado e testado no mesmo domínio e quando ele é treinado em um domínio e testado em outro diferente. Neste trabalho, nós reimplementamos o método deles e o usamos para comparar o seu desempenho em textos em português e em inglês.

3. MÉTODO PARA EXTRAÇÃO DE ASPECTOS

O método para extração de aspectos implementado neste trabalho foi proposto por [Jakob and Gurevych 2010]. Ele é um método de aprendizagem de máquina supervisionado que utiliza um modelo sequencial baseado em CRF [Lafferty et al. 2001]. Descrevemos a seguir os seus detalhes, bem como os atributos extraídos pelas ferramentas de PLN utilizados pelo método.

A extração de aspectos é feita por sentença. O método implementa um classificador, que recebe como entrada a sequência de *tokens* que compõe cada sentença, juntamente com seus atributos extraídos, e retorna a classe predita para cada *token*. Um *token* representa uma palavra ou o lema extraído da palavra por meio de um processo de lematização obtido por uma ferramenta de PLN. A classificação é feita por um esquema conhecido como *IOB*, com três classes, sendo que um *token* só pode pertencer a uma classe. Um rótulo *B* identifica o primeiro *token* de um aspecto, um rótulo *I* identifica a continuação de um aspecto, caso ele seja composto por mais de um *token*, e um rótulo *O* identifica os demais *tokens* (não aspectos). O CRF utiliza um modelo de cadeia linear, onde um *token* está diretamente relacionado com seus vizinhos.

O classificador utiliza um conjunto de atributos extraído dos textos por meio de ferramentas de PLN. Elas são usadas para obter o lema, a classificação gramatical das palavras e a árvore de dependência das sentenças. A seguir, são detalhados cada um dos atributos.

Token: Representa a palavra ou sua forma canônica (lema). Por exemplo, as palavras *devemos* e *deveria* possuem o mesmo lema, *dever*. Em certos domínios de textos, como comentários de produtos eletrônicos, a quantidade de aspectos tende a ser pequena. Assim, é provável que o conjunto de treino do classificador contenha a maioria dos *tokens* de aspectos, otimizando a aprendizagem. Esse atributo é denominado **wd** (palavra) ou **lm** (lema) nas tabelas de resultado da Seção 4.4.

PoS: Representa a classe gramatical de cada *token* em uma sentença (ex: substantivo, verbo, artigo). É útil para desambiguar *tokens* com significados diferentes, por exemplo, *envio* pode se referir tanto ao ato de enviar (verbo) quanto à ação de enviar (substantivo). Também, pode ser útil na identificação de aspectos formados por mais de um *token*. Esse atributo é denominado **pos**.

Dependência direta: Indica os *tokens* que têm uma ligação direta com uma expressão de sentimento na árvore de dependência. Por exemplo, na sentença “A tela é escura”, existe uma relação de predicado nominal (*nsubj*) entre *escura* e *tela*. Esse atributo é denominado **dln**.

Distância entre palavras: Indica o(s) *token(s)* que fazem parte do substantivo mais próximo de cada expressão de sentimento. Por exemplo, na sentença “Essa televisão possui uma imagem incrível”, o substantivo *imagem* está mais próximo da palavra de sentimento *incrível*, e portanto será marcado, já o substantivo *televisão* não será marcado. Esse atributo é importante, pois nem sempre a expressão de sentimento e o aspecto estão relacionados diretamente na árvore de dependência, como mostrado em [Zhuang et al. 2006], e assim, o atributo *dln* anterior não seria útil. Esse atributo é chamado de **wds**.

Expressão de sentimento: Indica todos os *tokens* que aparecem em sentenças contendo alguma expressão de sentimento. É importante para o classificador distinguir quando um determinado *token* aparece em sentenças com ou sem opinião. Esse atributo é chamado de **sen**.

4. AVALIAÇÃO EXPERIMENTAL

Esta seção descreve os experimentos e os resultados obtidos na avaliação do método descrito para a extração de aspectos a partir de textos na língua portuguesa e inglesa. O objetivo é verificar a maturidade de ferramentas disponíveis para a língua portuguesa comparada com aquelas para a língua inglesa.

4.1 Bases de dados

Nos experimentos, foram utilizadas três bases de dados com comentários de usuários em dois diferentes domínios. Todas elas possuem a marcação dos aspectos de cada sentença. A primeira base de dados, denominada *ReLi* [Freitas et al. 2014], contém comentários sobre 13 livros de 7 autores diferentes, na língua portuguesa. A segunda, também na língua portuguesa, é uma base de dados de revisões de usuários sobre um modelo de televisão, coletada de sites de comércio eletrônico do grupo B2W Digital pelo grupo de pesquisa dos autores deste artigo. As anotações foram feitas por dois estudantes de cursos de Ciência da Computação, um de graduação e outro de pós-graduação. A base de dados contém a marcação dos aspectos e das expressões de sentimento. Estão anotados tanto os aspectos explícitos quanto os implícitos, mas neste trabalho foram usados somente os aspectos explícitos. A base de dados resultante, denominada *TV*, contém somente as marcações em que houve concordância entre os anotadores. A terceira base de dados, denominada *HuLiu*, está em inglês, e foi criada por [Hu and Liu 2004]. Ela contém comentários sobre 5 diferentes produtos eletrônicos. Uma descrição quantitativa dessas bases de dados é apresentada na tabela I.

	ReLi	TV	HuLiu
Revisões	1435	1091	311
Sentenças	10507	2329	3947
Aspectos marcados	1983	2388	1831
Aspectos distintos	493	350	251
Sentenças com aspectos	16.2%	61.7%	38.2%
Sentenças com sentimento	67.4%	68.7%	61.7%

Table I. Detalhamento das bases de dados

As bases de dados foram traduzidas utilizando a ferramenta de tradução automática Azure Translator¹. Os textos em português foram traduzidos para o inglês, e os textos em inglês foram traduzidos para o português. Assim, tem-se as três bases de dados em português e em inglês. Os gabaritos com as expressões distintas de aspectos foram traduzidos separadamente dos textos com os comentários, e isso gerou algumas traduções diferentes para a mesma expressão. Assim, foi necessário um ajuste manual para garantir a consistência. As bases de dados estão publicamente disponíveis².

4.2 Métricas de Avaliação

A aplicação do método gera como resultado uma sequência de rótulos das classes correspondentes à classificação dos *tokens* de cada sentença. Para medir a qualidade da extração, foram utilizados os seguintes critérios: (i) se todos os *tokens* de um aspecto forem rotulados corretamente, é contado um verdadeiro positivo (TP), (ii) se os *tokens* forem rotulados parcialmente, é contado como um falso positivo (FP) e (iii) se um *token* que não é aspecto for rotulado como aspecto, também é contado como falso positivo. Por exemplo, considerando o aspecto anotado “brilho da tela”, se todos os três *tokens* forem rotulados como aspecto, é contado um verdadeiro positivo. Se for rotulado como aspecto apenas o *token* “tela” ou apenas “brilho”, então será considerado um falso positivo.

Foram utilizadas as métricas Precisão, calculada como $P = \frac{TP}{TP+FP}$, Revocação, calculada como $R = \frac{TP}{T_A}$, sendo T_A a quantidade de aspectos anotados, e Medida F1, calculada como $F1 = 2 * \frac{P * R}{P + R}$.

4.3 Configuração Experimental

A implementação do método de [Jakob and Gurevych 2010] foi feita neste trabalho com algumas modificações em relação à implementação original. Como ferramenta para o CRF, foi utilizada a CRFSuite³, em vez de Mallet⁴. Para as tarefas de PLN, foi usada a ferramenta CoreNLP [Manning et al. 2014] para os textos em inglês, como no original, e a ferramenta spaCy⁵ para os textos em português. Para identificar sentenças com expressões de sentimento, foram utilizados dois léxicos, o SentiWordNet [Baccianella et al. 2010] para os textos em inglês e o OpLexicon [Souza et al. 2011] para os textos em português. O trabalho original não usou léxicos, a avaliação foi feita com as expressões anotadas nas bases de dados. Além disso, neste trabalho, foi utilizado o lema da palavra como *token*, em vez da própria palavra. Em uma avaliação experimental, não houve diferença estatística entre o uso do lema e da palavra.

Os léxicos utilizados classificam os sentimentos de formas diferentes. No SentiWordNet, cada expressão recebe uma pontuação para negatividade e outra para positividade, representadas por valores entre 0 e 1, e o nível de subjetividade é calculado como a soma desses valores. Já o OpLexicon utiliza uma classificação mais simples, -1 para sentimento negativo, 0 para neutro e 1 para positivo. O critério para classificar uma sentença como contendo uma expressão de sentimento leva em conta as

¹<https://azure.microsoft.com/en-us/services/cognitive-services/translator/>

²<https://www.kaggle.com/brenexdev/aspect-extraction-portuguese>

³<http://www.chokkan.org/software/crfsuite/>

⁴<http://mallet.cs.umass.edu/>

⁵<https://spacy.io/>

particularidades de cada léxico. Para sentenças em inglês, usando o SentiWordNet, uma expressão foi considerada como de sentimento se o seu valor de subjetividade for maior ou igual a 0,5, e para sentenças em português, usando o OpLexicon, foi a presença no léxico com um valor diferente de 0.

Os textos das bases de dados passaram por uma etapa de pré-processamento, onde as revisões foram separadas em sentenças, convertidas em minúsculas, os *tokens* separados e lematizados. Não foram efetuadas remoções de *stopwords* e nem de acentuações.

Para avaliar qual a melhor combinação de atributos para a classificação dos aspectos, foram realizados experimentos com 9 diferentes configurações, como mostram as tabelas de resultados da Seção 4.4. Foi utilizada a abordagem de validação cruzada com 10 partições, onde em cada execução 9 partes da base de dados são usadas para treinamento e uma parte distinta é usada para teste. O resultado final de cada experimento é calculado como a média aritmética dos resultados de cada partição para cada uma das métricas utilizadas. Além disso, para o treinamento, foi feito um balanceamento dos dados, de forma que cada conjunto de treino contenha as sentenças com aspectos e o mesmo número de sentenças sem aspectos, escolhidas aleatoriamente.

4.4 Resultados e Discussão

As Tabelas II, III e IV apresentam os resultados para cada uma das bases de dados e para cada conjunto de atributos. As colunas da esquerda mostram os resultados para a língua portuguesa e as direita, para a língua inglesa. Os valores em negrito são os maiores de cada coluna.

Atributos	Português			Inglês		
	Precisão	Revocação	F1	Precisão	Revocação	F1
lm	0.671	0.338	0.450	0.673	0.348	0.459
lm, pos	0.653	0.363	0.467	0.676	0.380	0.486
lm, pos, sen	0.644	0.380	0.478	0.651	0.395	0.491
lm, pos, dln	0.665	0.387	0.490	0.657	0.403	0.500
lm, pos, wds	0.659	0.384	0.485	0.663	0.383	0.486
lm, pos, sen, dln	0.660	0.397	0.496	0.664	0.398	0.498
lm, pos, sen, wds	0.644	0.386	0.483	0.655	0.397	0.494
lm, pos, dln, wds	0.661	0.387	0.488	0.661	0.403	0.500
lm, pos, sen, dln, wds	0.644	0.392	0.487	0.660	0.395	0.494

Table II. Resultados obtidos com a base de dados ReLi

Atributos	Português			Inglês		
	Precisão	Revocação	F1	Precisão	Revocação	F1
lm	0.818	0.545	0.654	0.814	0.546	0.654
lm, pos	0.817	0.587	0.683	0.816	0.586	0.682
lm, pos, sen	0.807	0.593	0.683	0.814	0.589	0.683
lm, pos, dln	0.819	0.592	0.687	0.824	0.595	0.691
lm, pos, wds	0.817	0.587	0.683	0.813	0.593	0.686
lm, pos, sen, dln	0.822	0.584	0.683	0.825	0.590	0.688
lm, pos, sen, wds	0.818	0.588	0.684	0.813	0.591	0.685
lm, pos, dln, wds	0.824	0.589	0.687	0.825	0.596	0.692
lm, pos, sen, dln, wds	0.811	0.590	0.683	0.820	0.593	0.688

Table III. Resultados obtidos com a base de dados TV

Ao comparar a contribuição dos conjuntos de atributos, pode-se observar que adicionar outros atributos além do lema causa pouca mudança para o resultado, com base na medida F1, indicando que o lema é fundamental para o aprendizado nessa abordagem. Somente na base de dados ReLi, tanto em português quanto em inglês, o melhor conjunto de atributos é significativamente melhor do

Atributos	Português			Inglês		
	Precisão	Revocação	F1	Precisão	Revocação	F1
lm	0.716	0.289	0.411	0.669	0.275	0.390
lm, pos	0.699	0.326	0.445	0.665	0.315	0.427
lm, pos, sen	0.704	0.327	0.447	0.668	0.310	0.423
lm, pos, dln	0.703	0.334	0.452	0.673	0.320	0.433
lm, pos, wds	0.701	0.328	0.447	0.668	0.317	0.430
lm, pos, sen, dln	0.699	0.333	0.451	0.683	0.320	0.436
lm, pos, sen, wds	0.696	0.332	0.450	0.666	0.319	0.431
lm, pos, dln, wds	0.697	0.332	0.450	0.673	0.321	0.435
lm, pos, sen, dln, wds	0.687	0.331	0.447	0.672	0.326	0.439

Table IV. Resultados obtidos com a base de dados HuLiu

que o lema sozinho. O teste estatístico usado foi o *one-way repeated-measure ANOVA*, com 95% de confiança.

Esse resultado indica que a configuração mais simples, usando somente o lema, pode ser usada sem grandes perdas de qualidade. Entretanto, isso ocorreu porque as bases de dados usadas não têm um grande número de aspectos distintos (ver Tabela I), e pode não se repetir se for usado para domínios distintos, como foi observado por [Jakob and Gurevych 2010].

Comparando os resultados obtidos com as bases de dados em português e em inglês pela medida F1, não houve diferença estatística significativa para as bases de dados ReLi e TV. Já para a HuLiu, os resultados são significativamente melhores para o português, com ANOVA $F(1,9) = 6.28$, $p < .05$, para o conjunto de atributos $\{lm, pos, dln\}$. Isso se repete para quase todos os conjuntos de atributos. Esse resultado indica que as ferramentas para português estão, pelo menos, tão maduras quanto aquelas para o inglês.

5. CONCLUSÃO

Este trabalho apresentou uma avaliação sobre a maturidade de ferramentas para extração de aspectos para a língua portuguesa. Um método baseado em CRF foi implementado, com os valores dos atributos do classificador obtidos por meio de ferramentas de PLN para português e comparados com aqueles obtidos por ferramentas para o inglês. Os resultados sugerem que não vale a pena fazer a tradução dos textos para o inglês para se obter maior eficácia. Em geral, a tradução automática introduz perdas na qualidade dos textos, pois as máquinas de tradução são capazes de capturar principalmente os padrões genéricos compartilhados entre os idiomas, perdendo muito do conhecimento cultural específico. Entretanto, devido à predominância do atributo lema, é necessário efetuar mais testes com outras abordagens para ser conclusivo.

Este trabalho também contribuiu com a disponibilização de bases de dados em português e em inglês para experimentos com extração de aspectos. Em trabalhos futuros, pretendemos usá-las na avaliação de outros métodos para a língua portuguesa.

ACKNOWLEDGMENT

Nossos agradecimentos a Gustavo Dimas Franco Freitas pela sua contribuição na coleta e classificação da base de dados de TV, e a Erick Galani Maziero pelas sugestões de como anotá-la.

REFERENCES

- AIRES, J. P., PADILHA, C., QUEVEDO, C., AND MENEGUZZI, F. A deep learning approach to classify aspect-level sentiment using small datasets. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*. Rio de Janeiro, Brazil, pp. 1–8, 2018.

- ARAÚJO, M., PEREIRA, A., AND BENEVENUTO, F. A comparative study of machine translation for multilingual sentence-level sentiment analysis. *Information Sciences* vol. 512, pp. 1078 – 1102, 2020.
- BACCIANELLA, S., ESULI, A., AND SEBASTIANI, F. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Valletta, Malta, pp. 2200–2204, 2010.
- CAMBRIA, E., PORIA, S., HAZARIKA, D., AND KWOK, K. Senticnet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. In *Proceedings of the The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI)*. AAAI Press, New Orleans, USA, pp. 1795–1802, 2018.
- CATHARIN, L. G. AND FELTRIM, V. D. Finding opinion targets in news comments and book reviews. In *Proceedings of the 13th International Conference on Computational Processing of the Portuguese Language (PROPOR)*. Springer International Publishing, Canela, Brazil, pp. 375–384, 2018.
- CHATURVEDI, I., CAMBRIA, E., WELSCH, R. E., AND HERRERA, F. Distinguishing between facts and opinions for sentiment analysis: Survey and challenges. *Information Fusion* vol. 44, pp. 65–77, 2018.
- FARIAS, D. S., MATSUNO, I. P., MARCACINI, R. M., AND REZENDE, S. O. Opinion-meter: A framework for aspect-based sentiment analysis. In *Proceedings of the 22nd Brazilian Symposium on Multimedia and the Web (Webmedia)*. ACM, Teresina, Brazil, pp. 351–354, 2016.
- FREITAS, C., MOTTA, E., MILIDIÚ, R., AND CÉSAR, J. *Sparkling Vampire... lol! Annotating Opinions in a Book Review Corpus*. In Sandra Aluísio and Stella E. O. Tagnin (eds.), *New Language Technologies and Linguistic Research: A Two-Way Road*. Cambridge Scholars Publishing, 2014.
- FREITAS, L. A. D. AND VIEIRA, R. Ontology-based feature-level sentiment analysis in portuguese reviews. *International Journal of Business Information Systems* 32 (1): 30–55, 2019.
- HU, M. AND LIU, B. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA, pp. 168–177, 2004.
- JAKOB, N. AND GUREVYCH, I. Extracting opinion targets in a single- and cross-domain setting with conditional random fields. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. USA, pp. 1035–1045, 2010.
- LAFFERTY, J., MCCALLUM, A., AND PEREIRA, F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th International Conference on Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, USA, pp. 282–289, 2001.
- LIU, B. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies* 5 (1): 1–167, 2012.
- LO, S. L., CAMBRIA, E., CHIONG, R., AND CORNFORTH, D. Multilingual sentiment analysis: from formal to informal and scarce resource languages. *Artificial Intelligence Review* 48 (4): 499–527, Dec, 2017.
- MACHADO, M. T., PARDO, T. A. S., AND RUIZ, E. E. S. Analysis of unsupervised aspect term identification methods for portuguese reviews. In *Anais do XIV Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*. SBC, Uberlândia, Brazil, pp. 239–249, 2017.
- MANNING, C. D., SURDEANU, M., BAUER, J., FINKEL, J., BETHARD, S. J., AND MCCLOSKEY, D. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Baltimore, Maryland, pp. 55–60, 2014.
- PEREIRA, D. A. A survey of sentiment analysis in the portuguese language. *Artificial Intelligence Review*, July, 2020.
- PORIA, S., CAMBRIA, E., HOWARD, N., HUANG, G.-B., AND HUSSAIN, A. Fusing audio, visual and textual clues for sentiment analysis from multimodal content. *Neurocomputing* vol. 174, pp. 50–59, 2016.
- SAIAS, J., AO, M. M., AND OLIVEIRA, E. Detailing sentiment analysis to consider entity aspects: An approach for portuguese short texts. *Transactions on Machine Learning and Artificial Intelligence* 6 (2): 26–35, 2018.
- SOUZA, M., VIEIRA, R., BUSETTI, D., CHISHMAN, R., AND ALVES, I. M. Construction of a Portuguese opinion lexicon from multiple resources. In *Proceedings of the 8th Brazilian Symposium in Information and Human Language Technology*. Mato Grosso, Brazil, 2011.
- YADOLLAHI, A., SHAHRAKI, A. G., AND ZAIANE, O. R. Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys* 50 (2): 25:1–25:33, May, 2017.
- ZHANG, L. AND LIU, B. *Aspect and entity extraction for opinion mining*. Springer, 2014. Book Chapter, Volume 1 of the series Studies in Big Data.
- ZHANG, L., WANG, S., AND LIU, B. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8 (4): e1253, 2018.
- ZHUANG, L., JING, F., AND ZHU, X.-Y. Movie review mining and summarization. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management*. New York, NY, USA, pp. 43–50, 2006.