

ANÁLISE DE SENTIMENTO EM NÍVEL DE ASPECTO COM BASE EM ONTOLOGIAS DE DOMÍNIO E ANÁLISE DE DEPENDÊNCIA SINTÁTICA

FRANCISCO DIAS FRANCO¹; LARISSA ASTROGILDO DE FREITAS²;
ULISSES BRISOLARA CORRÊA³

¹ Universidade Federal de Pelotas – fdfranco@inf.ufpel.edu.br

² Universidade Federal de Pelotas – larissa.freitas@ufpel.edu.br

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

1. TRATAMENTO DOS LÉXICOS DE SENTIMENTO

Ao total foram testados doze Léxicos de Sentimento: AffectPT-br¹ (com e sem Wildcard Expansion), EmoLex², LeIA³, LIWC⁴, OpLexicon (SOUZA et al., 2011; SOUZA; VIEIRA, 2012), OntoPT⁵, Reli-Lex⁶, SentiLex⁷, SentiWordNet-PT-BR⁸, UNILEX⁹ e WordNetAffectBR¹⁰.

Em tese, o tratamento dos léxicos consiste em uma padronização de cada léxico, de forma a facilitar o acesso à informação e a implementação do sistema proposto. Foi gerado um dicionário referente à cada léxico utilizado, as chaves são compostas pelos termos (palavras, *stemming* de palavras, gírias, expressões ou elementos de linguagem (*hashtags*, *emoticon*)) utilizados pelo recurso léxico e o valor das chaves são correspondentes ao valor de polaridade do termo utilizado na chave.

No entanto, como nem todos os léxicos possuem a coluna de polaridade, ou não utilizavam "1" para positivo ou "-1" para negativo, foi necessário fazer uma transformação dos dados. Por exemplo:

- Para recursos como o WordNetAffectBR, que possuem símbolos para representar cada polaridade, foi criado uma nova coluna chamada de "polaridade"

¹O recurso AffectPT-br (CARVALHO; SANTOS; GUEDES, 2018) pode ser baixado através do link: <<https://github.com/LaCAfe/AffectPT-br>>.

²O recurso EmoLex (MOHAMMAD; TURNEY, 2010; MOHAMMAD; TURNEY, 2013) pode ser baixado através do link: <<https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>>.

³O recurso LeIA (HUTTO; GILBERT, 2014; ALMEIDA, 2018) pode ser baixado através do link: <<https://github.com/rafjaa/LeIA>>.

⁴O recurso LeIA (FILHO; PARDO; ALUÍSIO, 2013) pode ser baixado através do link: <https://sites.icmc.usp.br/sandra/LIWC/LIWC2007_Portugues_win.dic>.

⁵O recurso OntoPT (Gonçalo Oliveira; GOMES, 2014) pode ser baixado através do link: <http://ontopt.dei.uc.pt/index.php?sec=download_ontopt>.

⁶O recurso Reli-Lex (FREITAS et al., 2012) pode ser baixado através do link: <<https://www.linguateca.pt/Repositorio/ReLi/>>.

⁷O recurso SentiLex (SILVA; CARVALHO; SARMENTO, 2012) pode ser baixado através do link: <<https://b2share.eudat.eu/records/93ab120efdaa4662baec6adee8e7585f>>.

⁸O recurso SentiWordNet-PT-BR (ESULI; SEBASTIANI, 2006; BACCIANELLA; ESULI; SEBASTIANI, 2010) pode ser baixado através do link: <<https://github.com/Pedro-Thales/SentiWordNet-PT-BR>>.

⁹O recurso UNILEX (SOUZA; PEREIRA; DALIP, 2017) pode ser baixado através do link: <<https://dicionariounilex.wixsite.com/unilex>>.

¹⁰O recurso WordNetAffectBR (PASQUALOTTI; VIEIRA, 2008) pode ser baixado através do link: <<https://www.inf.pucrs.br/linatural/wordpress/recursos-e-ferramentas/wordnet affectbr/>>.

e atribuído a polaridade correspondente ao símbolo utilizado. Para o símbolo "+" foi atribuído "1" e para símbolo "-" foi atribuído "-1".

- Para recursos como o SentiWordNet-PT-BR, que possuem uma porcentagem de polaridade positiva ou negativa, foi criado uma nova coluna chamada de "polaridade" e atribuído "1" ou "-1" dependendo do maior valor de polaridade. Se o valor de polaridade positiva for maior é atribuído o valor "1", caso contrário, é atribuído o valor "-1".
- Para recursos como o Reli-Lex, que possuem listas para cada polaridade, foi criado uma nova coluna chamada de "polaridade" e atribuído a polaridade correspondente ao *dataset* utilizado. Para *datasets* com sentimento positivo foi atribuído "1" e para sentimento negativo foi atribuído "-1".

Além disso, em alguns léxicos também foi necessário corrigir erros de codificação e em outros foi necessário arrumar erros com o separador, pois, utilizavam um tipo de separador diferente ou mais de um separador.

Os léxicos Reli-Lex, LeIA, WordNetAffectBR utilizavam a codificação "ISO-8859-1", os outros utilizavam a codificação "UTF-8".

O léxico WordnetAffectBR utilizava o separador ";", os léxicos OntoPT, OpLexicon, LIWC utilizavam o separador "," e os léxicos AffectPT-br, Emolex, LeIA, SentiWordNet utilizavam o separador "\t". Os léxicos UNILEX e SentiLex utilizavam dois separadores, eles usavam, respectivamente, ";" e "\t", "." e ",".

5. REFERÊNCIAS

ALMEIDA, R. J. A. **LeIA - Léxico para Inferência Adaptada**. [S.l.]: GitHub, 2018. <<https://github.com/rafjaa/LeIA>>.

BACCIANELLA, S.; ESULI, A.; SEBASTIANI, F. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: **Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'10)**. Valletta, Malta: European Language Resources Association (ELRA), 2010. Disponível em: <http://www.lrec-conf.org/proceedings/lrec2010/pdf/769_Paper.pdf>.

CARVALHO, F.; SANTOS, G.; GUEDES, G. P. Affectpt-br: an affective lexicon based on liwc 2015. In: **2018 37th International Conference of the Chilean Computer Science Society (SCCC)**. [S.l.: s.n.], 2018. p. 1–5.

ESULI, A.; SEBASTIANI, F. SENTIWORDNET: A publicly available lexical resource for opinion mining. In: **Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)**. Genoa, Italy: European Language Resources Association (ELRA), 2006. Disponível em: <http://www.lrec-conf.org/proceedings/lrec2006/pdf/384_pdf.pdf>.

FILHO, P. B.; PARDO, T. A. S.; ALUÍSIO, S. An evaluation of the brazilian portuguese liwc dictionary for sentiment analysis. In: **Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology**. [S.l.: s.n.], 2013.

FREITAS, C.; MOTTA, E.; MILIDIÚ, R.; CÉSAR, J. Vampiro que brilha... rá! desafios na anotação de opinião em um corpus de resenhas de livros. **Encontro de Linguística de Corpus**, v. 11, p. 22, 2012.

Gonçalo Oliveira, H.; GOMES, P. ECO and Onto.PT: A flexible approach for creating a Portuguese wordnet automatically. **Language Resources and Evaluation Journal**, Springer, v. 48, n. 2, p. 373–393, 2014. Disponível em: <<http://dx.doi.org/10.1007/s10579-013-9249-9>>.

HUTTO, C.; GILBERT, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. **Proceedings of the International AAAI Conference on Web and Social Media**, v. 8, n. 1, p. 216–225, May 2014. Disponível em: <<https://ojs.aaai.org/index.php/ICWSM/article/view/14550>>.

MOHAMMAD, S.; TURNEY, P. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In: **Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text**. [S.l.: s.n.], 2010. p. 26–34.

MOHAMMAD, S. M.; TURNEY, P. D. Crowdsourcing a word–emotion association lexicon. **Computational intelligence**, Wiley Online Library, v. 29, n. 3, p. 436–465, 2013.

PASQUALOTTI, P. R.; VIEIRA, R. Wordnet affect br: uma base lexical de palavras de emoções para a língua portuguesa. **RENOTE**, v. 6, n. 1, 2008.

SILVA, M. J.; CARVALHO, P.; SARMENTO, L. Building a sentiment lexicon for social judgement mining. In: SPRINGER. **Computational Processing of the Portuguese Language: 10th International Conference, PROPOR 2012, Coimbra, Portugal, April 17-20, 2012. Proceedings 10**. [S.l.], 2012. p. 218–228.

SOUZA, K. F. de; PEREIRA, M. H. R.; DALIP, D. H. Unilex: Método léxico para análise de sentimentos textuais sobre conteúdo de tweets em português brasileiro. **Abakós**, v. 5, n. 2, p. 79–96, 2017.

SOUZA, M.; VIEIRA, R. Sentiment analysis on twitter data for portuguese language. In: **Proceedings of the 10th International Conference on Computational Processing of the Portuguese Language**. Berlin, Heidelberg: Springer-Verlag, 2012. (PROPOR'12), p. 241–247. ISBN 978-3-642-28884-5. Disponível em: <http://dx.doi.org/10.1007/978-3-642-28885-2_28>.

SOUZA, M.; VIEIRA, R.; Buseti, D.; CHISHMAN, R.; ALVES, I. M.; UNISINOS, F. D. L. Construction of a portuguese opinion lexicon from multiple resources. In: **8th Brazilian Symposium in Information and Human Language Technology - STIL, Mato Grosso**. [S.l.: s.n.], 2011.