

Creating a Portuguese context sensitive lexicon for sentiment analysis

Mateus T. Machado^{1,3,4} Thiago A.S. Pardo^{2,4} Evandro E.S. Ruiz^{3,4}

- ¹ Programa de Pós-Graduação em Computação Aplicada
- ² Departamento de Ciências de Computação, ICMC
- ³ Departamento de Computação e Matemática, FFCLRP
- ⁴ Universidade de São Paulo (USP)

PROPOR 2018, October 25th.

- 1 Introduction
- 2 Data & Methods
 - Lexicons
 - Aspect-based polarity detection
- 3 Results
- 4 Conclusion

Computing and Math Dept., USP Ribeirão Preto

Applied Computing, M.Sc.



1 Introduction

2 Data & Methods

- Lexicons
- Aspect-based polarity detection

3 Results

4 Conclusion

Sentiment Analysis (SA)

SA: A definition

Sentiment Analysis uses NLP techniques to **extract and classify** opinions, emotions, evaluations, and attitudes related to products, services, organizations, people, events and subjects expressed in free text.¹

¹[Liu, 2012]

²[Hussein, 2016]

Sentiment Analysis (SA)

SA: A definition

Sentiment Analysis uses NLP techniques to **extract and classify** opinions, emotions, evaluations, and attitudes related to products, services, organizations, people, events and subjects expressed in free text.¹

Obstacles²

- Analyzing the meaning of sentiments
- Detecting a suitable sentiment polarity
- *We add ABSA*

¹[Liu, 2012]

²[Hussein, 2016]

Aspect-Based Sentiment Analysis (ABSA)

Definition³

Aspect-Based Sentiment Analysis (ABSA) is a **fine grained form of SA** aiming to identify the aspects of given entities and the related sentiments.

Example

“O **telefone** realmente é muito bom, *fácil de usar*, sua **bateria** é pequena e *dura muito*, só a **tela** que achei pequena demais.”

Is ‘pequena’ a positive or a negative word?

³[Thet et al., 2010]

Aspect-Based Sentiment Analysis (ABSA)

Subtasks⁴

- Aspect Term Extraction
- Aspect Term Sentiment Estimation
- Aspect Aggregation

Aspect Term Sentiment approaches⁵

- Lexicon-based
- Machine Learning

⁴[Pavlopoulos and Androutsopoulos, 2014]

⁵[Taboada et al., 2011]

To achieve

Objective

The objective was to create a lexicon composed only by adjectives, where their **polarities** are identified through the **context** they belong to.

1 Introduction

2 Data & Methods

- Lexicons
- Aspect-based polarity detection

3 Results

4 Conclusion

Dataset

LexReLi

Lexicon specialized in identifying aspects' polarity in opinion texts about books.

Corpus ReLi⁶

- 1,600 reviews of 14 books
- Manually annotated for opinion, aspects and polarities
- $\approx 2,700$ aspects + polarities
- $\approx 2,100$ positives
- ≈ 600 negatives

Extended ReLi

- 6,698 reviews
- 51,148 sentences
- 980,640 words

⁶[Freitas et al., 2012]

Construction

- 1 Pre-processing: pos-tagger (nlpnet)
- 2 Aspect identification: sentences with nouns and adjectives
- 3 Polarity detection: sentence polarity detection with Adjectives Preference method⁷
- 4 Computation of frequency of adjectives polarities
- 5 Adjectives selection and polarity assignment

[Taboada et al., 2011]

LexReLi

Construction

- 1 Pre-processing: pos-tagger (nlpnet)
- 2 Aspect identification: sentences with nouns and adjectives
- 3 Polarity detection: sentence polarity detection with Adjectives Preference method⁷
- 4 Computation of frequency of adjectives polarities
- 5 Adjectives selection and polarity assignment

[Taboada et al., 2011]

Result: LexReLi composition

- 1.543 entries
- 70% positive
- 30% negative

Next...

We still have to check:

- If different lexicon combinations enhance SA; and
- The performance of distinctive polarity detection approaches under these various lexicons.

- 1 Introduction
- 2 Data & Methods
 - Lexicons
 - Aspect-based polarity detection
- 3 Results
- 4 Conclusion

Lexicon arrangements

Used lexicons

- OpinionLexicon⁸
- SentiLex⁹
- Brazilian Portuguese LIWC 2007¹⁰

[Souza et al., 2011, Souza and Vieira, 2012]

[Silva et al., 2010]

[Balage Filho et al., 2013]

Lexicon arrangements

Used lexicons

- OpinionLexicon⁸
- SentiLex⁹
- Brazilian Portuguese LIWC 2007¹⁰

[Souza et al., 2011, Souza and Vieira, 2012]

[Silva et al., 2010]

[Balage Filho et al., 2013]

Combined Lexicons

- 6 lexicons
- \approx 111,000 entries
- 30% positive
- 60% negative
- 10% neutral

Lexicon arrangements

Used lexicons

- OpinionLexicon⁸
- SentiLex⁹
- Brazilian Portuguese LIWC 2007¹⁰

[Souza et al., 2011, Souza and Vieira, 2012]

[Silva et al., 2010]

[Balage Filho et al., 2013]

Combined Lexicons

- 6 lexicons
- \approx 111,000 entries
- 30% positive
- 60% negative
- 10% neutral

Conciliated Lexicon

- 1 lexicon
- \approx 106,000 entries
- 30% positive
- 60% negative
- 10% neutral

1 Introduction

2 Data & Methods

- Lexicons
- Aspect-based polarity detection

3 Results

4 Conclusion

Aspect Based Polarity Detection

Steps

- Locate aspects marked in the ReLi corpus
- Search for adjective near aspect and verifies its polarity in the lexicon
- This polarity is attributed to the aspect
- Negation word between aspect and the adjective: polarity reversed
- In the absence of an adjective or neutral polarity, one of the following methods is applied.

Methods of sentence polarity detection

Steps

- Word 'lookup' in a sentiment lexicon
- Sum the polarity of each word in a sentence
- Negation word reverse the polarity of nearest word
- Value > 0 : positive polarity
- Value < 0 : negative polarity
- Value $=0$: neutral polarity

Methods of sentence polarity detection

Steps

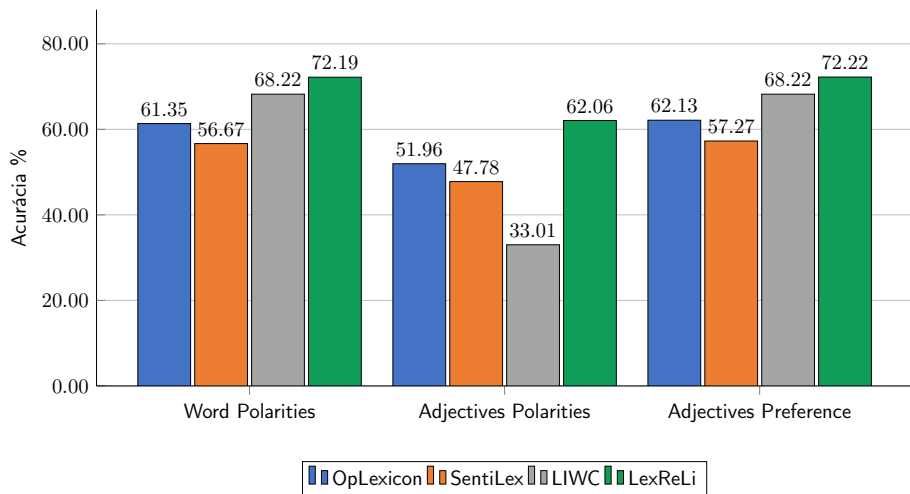
- Word 'lookup' in a sentiment lexicon
- Sum the polarity of each word in a sentence
- Negation word reverse the polarity of nearest word
- Value > 0 : positive polarity
- Value < 0 : negative polarity
- Value $=0$: neutral polarity

Methods

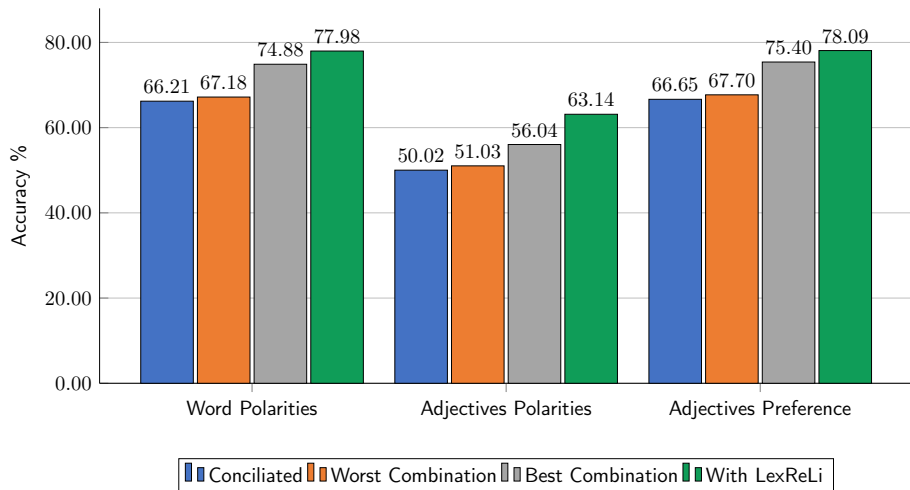
- Words Polarities
- Adjectives Polarities
- Adjectives Preference

- 1 Introduction
- 2 Data & Methods
 - Lexicons
 - Aspect-based polarity detection
- 3 Results**
- 4 Conclusion

Individual Lexicons



Combined Lexicons



- 1 Introduction
- 2 Data & Methods
 - Lexicons
 - Aspect-based polarity detection
- 3 Results
- 4 Conclusion

Conclusion

- Approaches used to create and combine lexicons were effective
- Method used to create LexReLi may be easily used for the creation of lexicons for different contexts
- Conciliation approach proved to be less effective
- Although slightly superior, the results of the preference to the adjectives method are promising

Referências I

- [Balage Filho et al., 2013] Balage Filho, P. P. P., Aluísio, S., and Pardo, T. T. (2013).
An Evaluation of the Brazilian Portuguese LIWC Dictionary for Sentiment Analysis.
9th Brazilian Symposium in Information and Human Language Technology – STIL, pages 215–219.
- [Freitas et al., 2012] Freitas, C., Motta, E., Milidiú, R. L., and César, J. (2012).
Vampiro Que Brilha... Rá! Desafios Na Anotação De Opinião Em Um Corpus De Resenhas De Livros.
In XI Encontro de Linguística de Corpus (ELC 2012), São Paulo.
- [Hussein, 2016] Hussein, D. M. E.-D. M. (2016).
A survey on sentiment analysis challenges.
Journal of King Saud University - Engineering Sciences.
- [Liu, 2012] Liu, B. (2012).
Sentiment analysis and opinion mining.
Synthesis Lectures on Human Language Technologies, 5(1):1–167.
- [Pavlopoulos and Androutsopoulos, 2014] Pavlopoulos, J. and Androutsopoulos, I. (2014).
Aspect Term Extraction for Sentiment Analysis: New Datasets, New Evaluation Measures and an Improved Unsupervised Method.
In Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM)@ EACL, pages 44–52.
- [Silva et al., 2010] Silva, M. J., Carvalho, P., Costa, C., and Sarmiento, L. (2010).
Automatic Expansion of a Social Judgment Lexicon for Sentiment Analysis.
Technical Report TR 1008 University of Lisbon Faculty of Sciences LASIGE, 6694(December).
- [Souza and Vieira, 2012] Souza, M. and Vieira, R. (2012).
Sentiment Analysis on Twitter Data for Portuguese Language.
Proceedings of the 10th international conference on Computational Processing of the Portuguese Language, pages 241–247.

Referências II

- [Souza et al., 2011] Souza, M., Vieira, R., Chishman, R., and Alves, I. M. (2011).
Construction of a Portuguese Opinion Lexicon from multiple resources.
Proceedings of the 8th Brazilian Symposium in Information and Human Language Technology, pages 59–66.
- [Taboada et al., 2011] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. (2011).
Lexicon-Based Methods for Sentiment Analysis.
Computational Linguistics, 37(2):267–307.
- [Thet et al., 2010] Thet, T. T., Na, J.-C., and Khoo, C. S. (2010).
Aspect-based sentiment analysis of movie reviews on discussion boards.
Journal of Information Science, 36(6):823–848.