An Evaluation of the Brazilian Portuguese LIWC Dictionary for Sentiment Analysis

Pedro P. Balage Filho, Thiago A. S. Pardo, Sandra M. Aluísio

Interinstitutional Center for Computational Linguistics (NILC)
Institute of Mathematical and Computer Sciences, University of São Paulo
Sao Carlos - SP, Brazil

{balage,taspardo,sandra}@icmc.usp.br

Abstract. *This work presents an evaluation of the Brazilian Portuguese LIWC dictionary for Sentiment Analysis. This evaluation is conducted by comparison against two other sentiment resources for Portuguese language: Opinion Lexicon and SentiLex. We conducted an intrinsic and an extrinsic evaluations and show how LIWC dictionary could be used in sentiment analysis projects.*

1. Introduction

Linguistic Inquiry and Word Count (LIWC) is a text analysis software that calculates the degree of use for different categories of words across a wide array of texts [Pennebaker et al. 2001]. The core of this program is a lexicon resource, best known as LIWC dictionary, which recently has been made available for Portuguese Language [1].

Sentiment analysis, or opinion mining, is a relatively new topic of research in natural language processing that has gained lots of attention due to the growth of the social web. A common task in sentiment analysis is text classification. In this task, a text, sentence or piece of opinion may be classified as positive, negative or neutral.

Sentiment classification is commonly categorized in two basic approaches: machine learning and lexicon-based [Taboada et al. 2011]. Machine learning approach uses a set of features, usually the vocabulary, which are learned from annotated corpora or labelled examples. The lexicon-based approach uses a lexicon to provide the polarity, or semantic orientation, for each word or phrase in the text. This last approach does not require an annotated corpora, and it is known for its domain independence, while the machine learning approach tends to adapt to the domain the classifier was trained [Aue and Gamon 2005]. The main component for the lexicon-based sentiment classifier is the lexicon resource, which needs to be precise and have a good vocabulary coverage.

In this work, we evaluated the use of Brazilian Portuguese LIWC dictionary for sentiment classification in Brazilian Portuguese texts. In this evaluation we compared this resource against two other sentiment resources available for Portuguese: the Opinion Lexicon and the SentiLex. Two evaluations are presented: an intrinsic and an extrinsic evaluation. The aim of this work is to provide new insights about how these lexicons would be useful in sentiment analysis and their main characteristics.

[1]This resource was kindly provided by the researchers: Profa. Rove Chishman, Profa. Sandra Maria Aluísio and Rosângela Lopes Toledo Checchia. In the moment, we are not aware of any publication introducing the resource.

Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology, pages 215–219,
Fortaleza, CE, Brazil, October 21–23, 2013.

2. Sentiment Lexicons

2.1. LIWC Dictionary

The LIWC dictionary is a recent lexicon which was built from the original English LIWC dictionary [Pennebaker et al. 2001]. The purpose of this lexicon is to group words into categories that can further be used to analyze psycho-linguistic features in texts.

The LIWC dictionary has 127,149 entries, where each entry can be assigned to one or more categories. The two categories used in this evaluation are *posemo* (12,878 entries), which stands for positive emotion, and *negemo* (15,115 entries), which stands for negative emotion. Other categories would also be useful (e.g., affect, anger, sad, etc), however, we decided it would be a fair comparison against the other lexicons if we used only these two.

The English version for LIWC dictionary has been used for a number of relevant works in sentiment analysis. For example, SentiStrength [Thelwall et al. 2010] uses LIWC dictionary for building its internal word list, which is the core of this sentiment classifier; Ott et al. [Ott et al. 2011] uses it to identify fictitious opinions that have been deliberately written to sound authentic; Kim et al. [Kim et al. 2012] uses it to classify anonymous texts.

2.2. OpinionLexicon

The OpinionLexicon [Souza et al. 2011] is a dictionary built for sentiment analysis task. To construct this resource, the authors applied three methods from the literature: a corpus-based, a thesaurus-based and an automatic translation system. The lexicon in the version 2.1 is composed of 30,678 entries (30,236 words and 442 phrases).

Opinion Lexicon was used by Souza and Vieira [Souza and Vieira 2012] for twitter sentiment analysis; and by Ribeiro Junior et al. [Ribeiro Junior et al. 2012] to assess vehicle features in blogs.

2.3. SentiLex

SentiLex [Silva et al. 2012] is a lexicon constructed for social judgments domain. The lexicon approach was built starting from publicly available language resources and enlarged by a combination of a linguistic-based and machine learning strategies.

The SentiLex (version 2) is made up of 82,347 inflected forms, organized in adjectives (16,863), nouns (1,280), verbs (29,504) and idiomatic expressions (34,700). SentiLex was used by Morgado [Morgado 2012] in sentiment classification for Portuguese On-line news and by Santos et al. [Santos et al. 2012] to evaluate positive and negative polarity propagation of words.

3. Agreement Evaluation

First, in order to conduct a fair evaluation, we normalized all lexicons into a comparable format. In this normalization, we only took the unigrams and their polarity. For entries containing more than one polarity, we choose the polarity for the first entry. This assumption was made in order to have a comparable format among the dictionaries (because LWIC does not have PoS or semantic roles). As LIWC dictionary has not a neutral category, we conducted the comparison against polar words only.

In our experiment, we analyzed the pairwise agreement, i.e., number of lexicon entries with equal polarity, among the lexicons. Table 1 shows this agreement.

Table 1. Lexical Agreement for polar entries

Agreement	LIWC	Opinion Lexicon	SentiLex
LIWC	x	80.17% (of 1,871 entries)	74.83% (of 7,310 entries)
Opinion Lexicon	x	x	97.04% (of 13,880 entries)
SentiLex	x	x	x

As we can see from Table 1, LIWC stands for 80% of agreement with Opinion Lexicon and 75% with SentiLex. The correlation between Opinion Lexicon and SentiLex is 97%. The next section shows an extrinsic evaluation based on a sentiment analysis task.

4. Sentiment Classification

The purpose of an extrinsic evaluation is to measure the performance of each lexicon in the sentiment classification task. For this reason, we choose to perform a simple lexicon-based sentiment classification where the lexicon performs almost all the work.

The algorithm adopted for this task is similar to the SO-CAL described in [Taboada et al. 2011]. This algorithm computes the individual polarity for each word in the lexicon and then sums up all these polarities to form the text polarity. If the sum is zero, the text is classified as neutral; if it is greater than zero, the text is classified as positive; otherwise, negative. The SO-CAL also accounts for intensification, *irrealis* and negation in the text.

There are few corpora available for sentiment classification in Portuguese language. For conducting this evaluation, we choose the ReLi [Freitas et al. 2012], a corpus from a Brazilian social network of book reviews. We choose this corpus because it is relatively new and we are not aware of any work that uses it yet, so we took the chance to understand better this resource. In addition, we also believed this corpus would not bias any of the lexicons in the evaluation.

The corpus is composed by 2,056 reviews from 13 different books (approximately 200 reviews each). The corpus has 300,000 words and 15,000 sentences that were annotated with PoS and chunks. The sentiment annotation is present in the opinion and sentence levels. The corpus has 4,210 positive opinion spans and 1,024 negative opinion spans. In the level of sentence, the corpus has 2,883 positive sentences and 596 negative ones.

For the evaluation, we choose to use both opinion and sentence levels of sentiment annotation. Therefore, we conduct two main experiments: opinion classification and sentence classification. The opinions present in the corpus are either positive or negative. Despite this factor, we did not change our lexicon classifier, so the classifier output could be positive, negative or neutral. Table 2 shows the results for the opinion classification in terms of precision, recall, F-measure and accuracy, which are common metrics for text classification evaluation. We omit the results for the neutral class.

As we can see from Table 2, all systems had a similar accuracy for opinion classification. When we compare the individual classes, the LIWC Lexicon had a

Table 2. Results for Opinion Classification

Lexicon	Class	Precision	Recall	F-measure	Accuracy
LIWC	Positive	88.93%	58.22%	70.37%	52.02%
	Negative	65.80%	34.51%	45.28%	
Opinion Lexicon	Positive	86.87%	55.42%	67.66%	50.53%
	Negative	58.18%	36.72%	45.02%	
SentiLex	Positive	95.74%	53.85%	68.93%	53.35%
	Negative	71.73%	51.95%	60.25%	

better performance in positive texts (F-measure of 70.37%) while SentiLex had a better performance in negative texts (F-measure of 60.25%).

Our second experiment was to perform sentence classification. Table 3 shows the results.

Table 3. Results for Sentence Classification

Lexicon	Class	Precision	Recall	F-measure	Accuracy
LIWC	Positive	86.42%	65.43%	74.48%	57.33%
	Negative	40.06%	22.66%	28.95%	
Opinion Lexicon	Positive	87.85%	50.95%	64.49%	47.42%
	Negative	35.96%	32.35%	34.06%	
SentiLex	Positive	91.67%	43.22%	58.74%	44.17%
	Negative	46.34%	48.26%	47.28%	

As we can see in Table 3, LIWC continues to have difficult to label negative examples (F-measure of 28.95%), but it have a high score for positive class (F-measure of 74.48%). By these results, we may assume that LIWC dictionary performs better indicating positivity than negativity.

5. Final Conclusions

In this paper, we presented an evaluation for the Brazilian Portuguese LIWC dictionary. This evaluation aims to guide future works in lexicon-based sentiment analysis. We conducted two evaluations: an intrinsic evaluation, by measuring the agreement compared with two other lexicons; and an extrinsic evaluation, by measuring the lexicon impact in a sentiment classification task.

Acknowledgements

We would like to thank the effort of those who produced the lexicons and the corpus used in this evaluation. We also would like to thank FAPESP and CNPq for financial support.

References

Aue, A. and Gamon, M. (2005). Customizing sentiment classifiers to new domains: A case study. In *Proceedings of recent advances in natural language processing (RANLP)*, volume 1, pages 2–1.

Freitas, C., Motta, E., Milidiú, R., and Cesar, J. (2012). Vampiro que brilha... rá! Desafios na anotação de opinião em um corpus de resenhas de livros. In *Proceedings do XI Encontro de Linguística de Corpus (XI ELC)*, São Carlos - SP.

Kim, I. J., Yoon, J. Y., Lim, J. Y., Kim, S. K., and Kim, U. M. (2012). Integration of opinion mining service flow with anonymous information. In *Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication*, ICUIMC '12, pages 104:1–104:4, New York, NY, USA. ACM.

Morgado, I. C. (2012). Classification of sentiment polarity of portuguese on-line news. In *Proceedings of the 7th Doctoral Symposium in Informatics Engineering*, pages 139–150.

Ott, M., Choi, Y., Cardie, C., and Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 309–319, Stroudsburg, PA, USA. Association for Computational Linguistics.

Pennebaker, J. W., Francis, M. E., and Booth, R. J. (2001). Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*.

Ribeiro Junior, S. S., Junior, Z., Meira Junior, W., and Pappa, G. L. (2012). Positive or negative ? Using blogs to assess vehicles features. In *Proceedings of the IX Encontro de Inteligência Artificial (ENIA)*, Curitiba-PR, Brazil.

Santos, A., Oliveira, H., Ramos, C., and Marques, N. (2012). The role of language registers in polarity propagation. In Caseli, H., Villavicencio, A., Teixeira, A., and Perdigão, F., editors, *Computational Processing of the Portuguese Language*, volume 7243 of *Lecture Notes in Computer Science*, pages 235–240. Springer Berlin Heidelberg.

Silva, M. J., Carvalho, P., and Sarmento, L. (2012). Building a Sentiment Lexicon for Social Judgement Mining. *Computational Processing of the Portuguese Language*, pages 218–228.

Souza, M. and Vieira, R. (2012). Sentiment Analysis on Twitter Data for Portuguese Language. *Computational Processing of the Portuguese Language*, 7243:241–247.

Souza, M., Vieira, R., Chishman, R., and Alves, I. M. (2011). Construction of a Portuguese Opinion Lexicon from multiple resources. In *8th Brazilian Symposium in Information and Human Language Technology - STIL*, Mato Grosso, Brazil.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, 37(2):267–307.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558.