

Lexicon-based Sentiment Analysis for Reviews of Products in Brazilian Portuguese

Lucas V. Avanço, Maria G. V. Nunes

Interinstitutional Center for Computational Linguistics (NILC)

Institute of Mathematical and Computer Sciences, University of São Paulo

avanco@usp.br, gracan@icmc.usp.br

Abstract—This paper presents some results on lexicon-based classification of sentiment polarity in web reviews of products written in Brazilian Portuguese. They represent a first step towards a robust opinion miner from reviews of technology products. The evaluation shows the performance of 3 different sentiment lexicons combined with simple strategies. It is also discussed the risk of considering the rating provided by the writers for the purpose of evaluating the algorithms. The results show that the better combination is the version of the algorithm that deals also with negation and intensification and uses the sentiment lexicon Sentilex. The average F-measure achieved 0.73.

I. INTRODUCTION

Since the creation of the Web, and mainly after the spread of Web 2.0 technologies, we have been drowned by data produced in social networks, micro blogging and forums. All these data hold a lot of rich information such as opinions about products or services and texts regarding political and social issues. This constitutes a great source of information to perform business and government intelligence. Considering the vital importance of understanding the sentiment present in texts, a new avenue for research has been flourished, called Sentiment Analysis and Opinion Mining [7], [3].

Sentiment Analysis can be seen as a natural language processing (NLP) task that aims to analyze opinions, sentiments, and emotions expressed in unstructured data [5]. A common task in this research area is polarity classification, which consists in classifying the overall sentiment present in a document or sentence. Usually this task is simplified by classifying a text or a sentence in 3 classes: positive, negative or neutral. In order to build sentiment classifiers, two main approaches have been investigated: lexicon-based methods [14], [16] and machine learning algorithms [8].

In this paper, we present and evaluate a classifier of reviews of products written in Brazilian Portuguese and published in specialized web forums. This represents a first step towards a robust opinion miner for Brazilian Portuguese (BP).

The classification of a text or a sentence according to its semantic orientation or polarity (positive, negative or neutral) can be performed by machine learning or lexicon based methods or even hybrid methods. Most of machine learning approaches use algorithms such as Support Vector Machine, Naive Bayes and Maximum Entropy, which are trained on a particular dataset for one specific domain [8]. The usual features used include unigrams (bag-of-words), bigrams and part-of-speech tags. Despite of the high accuracy reached by these approaches,

when the classifier is used for another domain, its performance decreases significantly [1].

Lexicon-Based methods, on the other hand, rely only on linguistic knowledge, and they are more robust across domains and texts [14]. Nevertheless, high accuracy is harder to achieve. Basically they use a sentiment lexicon consisting in a set of pairs of word and its polarity. Words belonging to a sentiment lexicon are called sentiment words. It is important to notice that not every word has a polarity value (and hence belong to the lexicon); usually adjectives, adverbs and some substantives and verbs have polarity values. Moreover, some rules for handling negation and intensity are used to sophisticate the simpler methods. Whereas there is no known robust opinion miner for BP, some basic linguistic resources have been built. In this paper, we evaluate 3 different sentiment lexicons for BP (described in Section II) in the context of polarity classification.

Hybrid methods [10] combine lexicon-based and supervised learning, and even manually written linguistic rules. Different unsupervised learning methods can also be used in a cascade way such that whether one classifier fails, the next one tries to classify, and so on, until the text is classified or there is no more classifier to use.

No matter which method is used, some important challenges must be faced when aiming to classify texts from informal web forums. They all refer to a necessary preprocessing of the input text for optimal functioning of the methods. In the corpus considered in this work (described in Section IV), we face the following major noises which can jeopardize the final results: case folding, punctuation, spelling and the use of internet slang. The effects of these noises and some procedures to minimize them are discussed in [4].

Another important issue is related to the rating given by the writers about the product under evaluation. This rating is usually taken into account for evaluating the classifier accuracy. The classifier output is compared with the writer's own rating. In this paper we show that one additional rating of the reviews which was made by different readers reveals a significant discrepancy. We show how the classifier accuracy changes when one shifts from one rating to other. This problem had already been pointed by [6], [9].

Hence, this paper relates the performance of 3 different versions of a lexicon-based classifier depending on some improvements based on linguistic knowledge, as well as its evaluation based on the reviews rated by the original reviewers and also on the ones rated by independent readers.

Section II relates some close works to this proposal. Section III describes the method and the construction of the classifier. Section IV presents the results and the comparative evaluation of the different versions of the classifier. Finally, Section V concludes the paper and points to some future work.

II. RELATED WORK

In [14] it is presented a method to calculate semantic orientation (Semantic Orientation CALCulator - SO-CAL) for lexicon-based classification. Words have prior and context-independent polarity, and it is possible to compute a numerical value to express semantic orientation (polarity). Hence, the semantic orientation of a whole text can be obtained by combining the values of words polarities. The SO-CAL is also based on valence shifters to handle negation and intensification contexts. Negation handling is performed by shifting the base value of the polarity word when it is in a negation context. For intensification it is used a percentage value for each booster or downtoner word, to increase or decrease the strength of a polarity word. The authors present results for versions applied in different domains (books, cars, computers, cookware, hotels, movies, music and phones). The mean values are: using just polarity words, 0.6604 of accuracy; polarity words and negation handling, 0.6835 of accuracy; polarity words with negation and intensification handling, 0.7135 of accuracy. The most complex version of SO-CAL, which considers negation and intensification contexts, modals, negative and repetition weighting, achieves a higher accuracy, 0.7874.

The basic and most important resource for sentiment classifiers is the sentiment lexicon. It can be created manually or automatically. Frequently, the automatic methods use seed words and verify the distance between one specific adjective (in general) and one or more seed words. It is common to obtain the association with a seed word by applying the Pointwise Mutual Information as proposed by [16]. For BP there are some sentiment lexicons created by different methods.

The OpinionLexicon [13] uses three approaches to build the lexicon: corpus-based, thesaurus-based and translation. The corpus-based approach consists in verifying the mutual information of words regarding to a set of seed words. Thesaurus-based approach looks for the minimal paths from a word and a list of synonyms and antonyms. Translation is used to make use of existing sentiment lexicons for other languages. The resulting lexicon has 31,150 words: 8,675 positive, 14,469 negative and 8,005 neutral.

SentiLex [11] is a sentiment lexicon for the Politics domain, using both linguistic knowledge and machine learning. It has 82,347 inflected forms: 20,663 are positive, 53,656 are negative, 7,704 are neutral.

The BP translation of a subset of LIWC [2] is composed just by the words whose tag is “posemo” (positive emotion) and “negemo” (negative emotion). It has a total of 27,993 polarity words: 12,878 positive words and 15,115 negative words.

An evaluation of agreement between these three lexicons is presented by [2]. First a normalization was performed because there is no uniform format shared by the lexicons. For example, LIWC does not have neutral words, OpinionLexicon

and SentiLex do; LIWC is composed just by unigrams while SentiLex and OpinionLexicon have multi-words. Analyzing 1,871 entries, LIWC has 80% in common with OpinionLexicon, and 75% in common with SentiLex comparing 7,310 entries. OpinionLexicon and SentiLex have 97% in common, considering 13,880 entries.

Few works deal with sentiment classification for BP. In [12] the authors propose a lexicon-based classifier for tweets in BP. Both lexicons OpinionLexicon and SentiLex are used. Their best results are 0.55 of F1-Score for positive class and 0.45 for negative. In [2], for a domain of books’ reviews, the authors use the same three lexicons we use in this paper. Their system achieved 0.7037 of F1-Score for positive, and 0.6025 for negative classification. Another work for BP, in a different scenario, is [15], which performs a case study over opinions expressed about politicians as a reaction to news. The system applies lexicon-based classification using a modified lexicon, adding domain-dependent words that express sentiments. The best results were: 0.5214 of accuracy, 0.3723 of F1-Score for positive, and 0.6552 of F1-Score for negative.

III. LEXICON-BASED CLASSIFIER

The method for building the lexicon-based classifier (LBC) proposed by this paper is basically a variation of the method developed by [14]. It considers the prior polarity of words according to a sentiment lexicon and uses some linguistic knowledge about contextual valence shifting (negation and intensification) to compute the polarity value of each sentence and text. In this work we consider the classes positive and negative; the class neutral is not being considered yet.

The prior polarities are defined by the sentiment lexicon. We have used separately 3 sentiment lexicons for BP: OpinionLexicon, SentiLex and a subset of LIWC.

The polarity value of a text is the sum of the prior polarities of its sentiment words, eventually modified by the contextual valence. If the sum is positive (strictly greater than zero) the opinion is classified as positive, otherwise it is classified as negative. To deal with contextual shifting, a set of negation words and booster-reducer words is used, as presented in Table I. The algorithm LBC is presented in Frame 1.

TABLE I. THE SETS OF NEGATION, AMPLIFIER AND DOWNTONER WORDS, IN PORTUGUESE

Negation	Amplifier	Downtoner
<i>jamaís, nada, nem, nenhum, ninguém, nunca, não, tampouco</i>	<i>mais, muito, demais, completamente, absolutamente, totalmente, definitivamente, extremamente, frequentemente, bastante</i>	<i>pouco, quase, menos, apenas</i>

Three possible scenarios demand change of polarity value: negation context, intensification context, and both of them together. The context is defined by a window whose size was empirically chosen as of 4 words (choosing 3 or 5 words produced worse results) to the left of the sentiment word. In the first case (only negation), if there is some sentiment word and a negation word in the same context, the polarity value is flipped (line 17 of algorithm in Frame 1). If a sentiment word occurs in an intensification context, its polarity value is

tripled if a booster word was found (line 8 of algorithm), or divided by three if a reducer word was found (line 14). When negation and intensification words are in the same context, the amplifier turns a downtoner, or the opposite (lines 6 and 12).

Frame 1. Algorithm to calculate the overall sentiment in a text

```

1: overall_sentiment  $\leftarrow$  0
2: while there is sentiment_word in text do
3:   polarity  $\leftarrow$  read lexicon(sentiment_word)
4:   if booster word in context then
5:     if negation word in context then
6:       polarity  $\leftarrow$  polarity/3
7:     else
8:       polarity  $\leftarrow$  polarity * 3
9:     end if
10:  else if reducer word in context then
11:    if negation word in context then
12:      polarity  $\leftarrow$  polarity * 3
13:    else
14:      polarity  $\leftarrow$  polarity/3
15:    end if
16:  else if negation word in context then
17:    polarity  $\leftarrow$  -1 * polarity
18:  end if
19:  overall_sentiment = overall_sentiment + polarity
20: end while

```

See the sentences below to better understand the computation of the semantic orientation performed by the algorithm.

“O celular é bom, apesar da bateria não ser muito boa”

bom (+1) + não ser muito boa (+1/3) = +1.33

“O celular não é muito ruim, mas é um pouco lento e as vezes apaga de repente”

não é muito ruim (-1/3) + pouco lento (-1/3) + apaga (-1) = -1.66

IV. RESULTS

To evaluate our classifier we used a dataset composed by reviews of products crawled from the database of one of the most traditional online services in Brazil, called Buscapé, where customers post their comments about several products. The comments are written in a free format within a template with three sections: Pros, Cons, and Opinion. The reviews selected are specific about mobiles and smartphones. We selected a sample of 2000 reviews, such that 1000 were set as positive, and 1000 were set as negative. This classification is based on the reviewer’s final recommendation (or not) about the product. However the analysis of a small sample has revealed how inconsistent can be the classification (final

recommendation) given by the writer when one takes into account the corresponding text.

Some examples of inconsistent polarity attribution for the texts in our dataset (original reviews, some with misspelling and grammar errors):

- Positive for Writer, but Negative for Reader:

“O custo poderia ser mais em conta.”

“Após 10 meses a bateria já era - viciadíssima.”

“Sua fragilidade atrapalha, por ser fácil de arranhar.”

- Negative for Writer, but Positive for Reader:

“para quem gosta de modernidade e agilidade”

“ESSE CELULAR SÓ É BOM POR UM ANO UMA COISA QUE ELE É RUIM É A CAMERA MAIS VALE APENA”

Therefore, a second version of the same dataset was used: the one whose reviews were manually classified. The same 2000 reviews were reclassified independently by 2 readers. The initial readers agreed to classify approximately 80.6% of reviews. The result of this revision was: 748 reviews rated as negative, 1085 as positive, 71 as neutral and 96 have both positive and negative aspects together. The latest ones were discarded for evaluation purposes.

The classifier LBC described in the previous section was broken into 3 different versions in order to evaluate the impact of each improvement on the use of word polarity only. The first version, LBC-p, considers only the prior polarity of sentiment words. The second one, LBC-pn, also considers negation contexts. The third and complete version, LBC-pni, adds the treatment of intensification contexts too.

We have used the most common measures, F1-Score and Accuracy, to evaluate our classifier. F1-Score is a harmonic mean of precision and recall, expressed by $F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$. Accuracy is the ratio between the total number of opinions correctly classified and the total number of opinions submitted to the classifier.

Table II shows the evaluation results for the 3 versions (col. 1) of the classifier, varying the sentiment lexicon (col. 2), for the original dataset - reviews rated by the original writers (2000 reviews: 1000 positive and 1000 negative). Table III presents the correspondent results for the revised data - reviews classified by independent readers (1833 reviews: 1085 positive and 748 negative). The best results are in bold.

The performance analysis for the different datasets shows an important increase of Average F1-Score and Accuracy in favor of the revised dataset in every combination of version and lexicon (col. 4 and 5). In average, the difference is of 0.063 in Average F1-Score, and of 0.075 in Accuracy. This stresses the fragility of the spontaneous rating by the opinion writers. This means that evaluation methods for classifiers must take this into account. This is still more important when machine learning methods are used.

TABLE II. EVALUATION OF LBC VERSIONS ON THE ORIGINAL DATASET - RATED BY THE WRITERS

LBC	Sentiment Lexicon	F1-Score		Average F1-Score	Accuracy
		Positive	Negative		
LBC-p	LIWC	0.6547	0.3026	0.4786	0.5382
	OpinionLexicon	0.6868	0.5924	0.6396	0.6459
	SentiLex	0.6225	0.5433	0.5829	0.5868
LBC-pn	LIWC	0.6860	0.4585	0.5722	0.6026
	OpinionLexicon	0.6474	0.4944	0.5709	0.5846
	SentiLex	0.6733	0.6557	0.6645	0.6649
LBC-pni	LIWC	0.6875	0.5114	0.5994	0.6189
	OpinionLexicon	0.6512	0.4781	0.5646	0.5820
	SentiLex	0.6879	0.6524	0.6701	0.6712

TABLE III. EVALUATION OF LBC VERSIONS ON THE DATASET RATED BY THE INDEPENDENT READERS

LBC	Sentiment Lexicon	F1-Score		Average F1-Score	Accuracy
		Positive	Negative		
LBC-p	LIWC	0.7481	0.3560	0.5520	0.6380
	OpinionLexicon	0.7529	0.6013	0.6771	0.6950
	SentiLex	0.7251	0.5895	0.6573	0.6708
LBC-pn	LIWC	0.7787	0.5291	0.6539	0.6990
	OpinionLexicon	0.7182	0.5026	0.6104	0.6403
	SentiLex	0.7632	0.6938	0.7285	0.7330
LBC-pni	LIWC	0.7799	0.5768	0.6783	0.7105
	OpinionLexicon	0.7272	0.4983	0.6127	0.6467
	SentiLex	0.7751	0.6945	0.7348	0.7410

It is also interesting to notice that the difference between positive and negative F1-Scores is lower when the lexicon SentiLex is used. Indeed, the lexicon SentiLex presented the better results, followed by LIWC. We can also see the increasing of F1-Score and Accuracy measures from the first to the latest version, for LIWC and SentiLex, but surprisingly not for OpinionLexicon - what must be further investigated.

One more experiment was conducted to evaluate the performance of LBC, now using as the sentiment lexicon the union of the three above lexicons. Actually this new sentiment lexicon is an extension of the SentiLex, so the entries present in LIWC or OpinionLexicon, which are not in SentiLex, were added to SentiLex. The resulting lexicon is called UnionLexicon.

TABLE IV. EVALUATION OF LBC-PNI WITH UNIONLEXICON FOR EACH DATASET

Dataset	F1-Score		Average F1-Score	Accuracy
	Positive	Negative		
Original (writers)	0.6978	0.6288	0.6633	0.6670
Revised (readers)	0.7726	0.6607	0.7166	0.7278

Comparing these latest results to the best ones of Tables I and II, one can see that the performance of LBC-pni with UnionLexicon is worse, in general. One possible reason for that is the presence of sentiment words which are more dependent on the context. The polarity of words like “grande (big)” and “pequeno (small)” are totally dependent on context. For example, some aspects of a product are expected to be big, while others, small. The UnionLexicon has the following entries (grande,+1) and (pequeno,-1), in contrast with the SentiLex, which has no entries for these words.

V. CONCLUSIONS AND FUTURE WORK

This paper proposed a lexicon-based classifier for reviews of products in BP which has reached very good evaluation measures when compared to analogous ones for BP. In addition, our comparison between original and revised rated

datasets confirms the results obtained by [6], [9]: the ratings of reviewers are not reliable, so if they are supposed to be used as reference in an evaluation, they must be revised.

Future works include some improvements towards a more realistic classification when positive and negative impressions are expressed in the same review. Many lexicon-based methods propose to neutralize sentiment polarities when opposite values occur in the same sentence. This does not seem to produce good results. For example, “*this mobile is great but very expensive*” is not considered nor positive nor negative, therefore it is considered neutral by most lexicon-based methods. For many purposes, however, it is far from a neutral opinion. The source of this problem may be in the text region one looks for polarity values: word, sentence, text. Smaller text chunks should be considered, and probably more sophisticated NLP techniques will be necessary for handling them. Machine Learning techniques will also be investigated for classifying opinions in BP, and the results will be contrasted to the ones presented here.

VI. ACKNOWLEDGMENTS

Part of the results presented in this paper were obtained through research activity in the project titled “Semantic Processing of texts in Brazilian Portuguese”, sponsored by Samsung Eletrônica da Amazônia Ltda. under the terms of Brazilian federal law number 8.248/91.

REFERENCES

- [1] Aue, A., Gamon, M.: Customizing sentiment classifiers to new domains: A case study. In: Proceedings of Recent Advances in Natural Language Processing (RANLP). 2005. p. 2, (2005)
- [2] Balage Filho, P.P., Pardo, T.A.S., Aluísio, S.M.: An Evaluation of the Brazilian Portuguese LIWC Dictionary for Sentiment Analysis. In the Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology, pp 215-219, (2013)
- [3] Becker, K., Tumitan, D.: “Introdução à Mineração de Opiniões: Conceitos, Aplicações e Desafios.” Simpósio Brasileiro de Banco de Dados (2013).
- [4] Duran, S. M., Avanço, L.V., Aluísio, S.M., Pardo, T.S., Nunes, M.G.V.: Some Issues on the Normalization of a Corpus of Products Reviews in Portuguese, Proceedings of the 9th Web as Corpus Workshop (WaC-9), Gothenburg, Sweden, Association for Computational Linguistics, pp. 22-28, (2014)
- [5] Liu, B.: Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, v. 5, n. 1, pp. 1-167, (2012)
- [6] Maks, I., Vossen, P.: Sentiment Analysis of Reviews: Should we analyze writer intentions or reader perceptions?, in: Proceedings of Recent Advances in Natural Language Processing (RANLP-2013), Hissar, Bulgaria, September 7-14, pp. 415-419, (2013)
- [7] Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and trends in information retrieval, v. 2, n. 1-2, pp. 1-135, (2008)
- [8] Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002. pp. 79-86, (2002)
- [9] Pang, B., Lee, L.: Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2005. pp. 115-124, (2005)
- [10] Prabowo, R., Thelwall, M.: Sentiment analysis: A combined approach. Journal of Informetrics, v. 3, n. 2, pp. 143-157, (2009)

- [11] Silva, M.J., Carvalho, P., Sarmiento, L.: Building a Sentiment Lexicon for Social Judgement Mining, In Lecture Notes in Computer Science (LNCS) / Lecture Notes in Artificial Intelligence (LNAI), International Conference on Computational Processing of Portuguese (PROPOR), 17-20 April, 2012, Coimbra, (2012)
- [12] Souza, M., Vieira, R.: Sentiment analysis on twitter data for portuguese language. In: Computational Processing of the Portuguese Language. Springer Berlin Heidelberg, 2012. pp. 241-247, (2012)
- [13] Souza, M., Vieira, R., Busseti, D., Chishman, R., Alves, I.M.: Construction of a portuguese opinion lexicon from multiple resources. In: The 8th Brazilian Symposium in Information and Human Language Technology (STIL 2012), Cuiabá, Brazil, (2011)
- [14] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. Computational linguistics, v. 37, n. 2, pp. 267-307, (2011)
- [15] Tumitan, D., Becker, K.: Tracking Sentiment Evolution on User-Generated Content: A Case Study on the Brazilian Political Scene. In: Brazilian Symposium on Databases (SBBD). 2013. p. 6.
- [16] Turney, P.D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, pp. 417-424, (2002)