A Majority Voting Approach for Sentiment Analysis in Short Texts using Topic Models

Rodrigo Rodrigues do Carmo
Centro Federal de Educação
Técnológica de Minas Gerais
Avenida Amazonas, 7675
Belo Horizonte, Minas Gerais
30510-000
rodrodriguescarmo@gmail.com

Anísio Mendes Lacerda
Centro Federal de Educação
Técnológica de Minas Gerais
Avenida Amazonas, 7675
Belo Horizonte, Minas Gerais
30510-000
anisio@decom.cefetmg.br

Daniel Hasan Dalip
Centro Federal de Educação
Técnológica de Minas Gerais
Avenida Amazonas, 7675
Belo Horizonte, Minas Gerais
30510-000
hasan@decom.cefetmg.br

ABSTRACT

Nowadays people can provide feedback on products and services on the web. Site owners can use this kind of information in order to understand more their public preferences. Sentiment Analysis can help in this task, providing methods to infer the polarity of the reviews. In these methods, the classifier can use hints about the polarity of the words and the subject being discussed in order to infer the polarity of the text. However, many of these texts are short and, because of that, the classifier can have difficulties to infer these hints. We here propose a new sentiment analysis method that uses topic models to infer the polarity of short texts. The intuition of this approach is that, by using topics, the classifier is able to better understand the context and improve the performance in this task. In this approach, we first use methods to infer topics such as LDA, BTM and MedLDA in order to represent the review and, then, we apply a classifier (e.g. Linear SVM, Random Forest or Logistic Regression). In this method, we combine the results of classifiers and text representations in two ways: (1) by using single topic representation and multiple classifiers; (2) and using multiple topic representations and a single classifier. We also analyzed the impact of expanding these texts since the topic model methods can have difficulties to deal with the data sparsity present in these reviews. The proposed approach could achieve gains of up to 8.5% compared to our baseline. Moreover, we were able to determine the best classifier (Random Forest) and the best topic detection method (MedLDA).

CCS CONCEPTS

• **Computing methodologies → Machine learning**; *Classification and regression trees*; Natural language processing;

KEYWORDS

Sentiment Analysis; Topic Models; Text Expansion

WebMedia '17, October 17–20, 2017, Gramado, Brazil

ACM ISBN 978-1-4503-5096-9/17/10…$15.00
https://doi.org/10.1145/3126858.3126861

ACM Reference format:
Rodrigo Rodrigues do Carmo, Anísio Mendes Lacerda, and Daniel Hasan Dalip. 2017. A Majority Voting Approach for Sentiment Analysis in Short Texts using Topic Models. In *Proceedings of WebMedia '17, Gramado, Brazil, October 17–20, 2017,* 7 pages.
https://doi.org/10.1145/3126858.3126861

1 INTRODUCTION

Nowadays, on-line reviewing has becoming an important tool for users to provide feedback on products and services on the web. On the other hand, the popularization of this kind of interaction helps site owners to understand their public and interact with them in a more effective way.

This interaction has several challenges, one of them is to determine the user assessment of a particular item by analyzing a given interaction. The technique, known as Sentiment Analysis, detects, extracts and classify opinions, sentiment and attitudes [12]. For instance, a text can be classified as positive, negative or neutral depending on its content or the author intention [3].

A variety of studies have investigated different strategies to overcome the sentiment analysis problem, which consists in identifying the sentiment expressed by users referring to a product or service [3], [13], [14], [2]. These methods commonly use the text as the input in order to predict the class of a given review, which can be classified as one from a set of categories (e.g., positive, negative or neutral). Hence, the problem is often modeled as a classification problem.

In these methods, the classifier can use hints about the polarity of the words and the subject being discussed in order to infer the polarity of the text. However, many of these texts are short and, because of that, the classifier can have difficulties to infer these hints.

In this paper, we propose a Sentiment Analysis Method in which we represent the text of reviews as topics in a lower dimensional space. We believe that this new representation is effective in avoiding the high dimensionality problem existing in the bag of words model. Moreover, as the text in this domain is small, we propose to use a text expansion technique in order to predict better topics to the text. The use of topic and text expansion has never been used to predict the polarity of a text before. Our hypothesis is that, by doing this, we can give more information about the context to the classifier and, then, increase the prediction performance.

To accomplish this, we compare our method with baselines using a sample of Yelp reviews. Through experiments, we have shown that our method achieved 81.7% of Macro-F1 score. Then, our new contributions are (1) the proposal a technique to infer the polarity taking the topics into account with gains of up to 8.5% against the baselines; (2) an study of which text representation is better to use in this task.

This paper is organized as it follows. Section 2 describes our related work. We describe the proposed method in Section 3. Our experiments and results are presented in Section 4. Finally, we conclude and presents the future work in Section 5.

2 RELATED WORK

In this section we present studies about Sentiment Analysis. Most of the methods use two techniques: (1) lexical methods and (2) performing machine learning techniques. The lexical methods analyze the polarity of words in order to predict the sentiment of the text.

Proposed by [4], SentiWordNet is a lexicon based method in which the words are grouped in synonym sets (namely synsets). By using the WordNet lexical database, this method associates each synset to a score representing the polarity of the text (positive, negative or neutral). Other study proposed the method Happiness index which uses a sentiment scale based on the Affective Norms for English Words (ANEW) containing 1,034 words with the affective dimensions of valence, arousal, and dominance.

SentiStrength, proposed by [16], uses an expanded version of LIWC, a dictionary which associates words with sentiments and other categories such as personal concerns (e.g. work, home, leisure), cognitive processes (e.g. seeing, hearing, thinking), body (e.g. ache, heart, cough). Defined by [6], EmoLex, also known as NRC Emotion Lexicon, is a lexical method which uses up 10,000 pairs of word and 8 basic emotions: joy, sadness, anger, fear, trust, disgust, anticipation and surprise.

Some other studies use machine learning techniques in order to predict the polarity of texts. For example, Sentiment Analyser (SASA) uses the Naïve Bayes classifier on unigram features to perform this task [17]. Stanford Recursive Neural Tensor Network uses a deep learning technique in order to predict the sentiment of the text, taking also the text structure into account [15].

Note that, these methods are commonly used in short texts in which texts can have few information in order to help the classifier to predict the polarity of the text. Because of that, a method which expands the text can help to identify the context and more information about the text. Then, our study presents a method to expand the text and show how this can be used in order to increase the performance in this task. As far as we know, this has never been applied before in sentiment analysis.

In a different domain, [1, 9] expand the text in order to predict the topic in short texts. According to the authors, the importance of expanding the text in this domain is that the state-of-art techniques rely on word co-occurrence which is rare in short texts. In this study, they expand the text using word vectors. By doing this, they have shown that their method could predict topics in texts better than the state-of-the-art methods.

3 TEXT POLARITY PREDICTIONS USING MAJORITY VOTING AND TOPICS REPRESENTATIONS

In this work we propose a method to predict the sentiment of a short text, which can be a review of a product or a movie, by using the text topics and a combination of classifiers.

Given a set of short texts that we want to evaluate, the first step is to make a preprocessing of these texts by lowercasing the words, removing stop-words and words with two characters or less. After that, we apply our approach, as presented in Figure 1. Since it is harder to detect topics in short text, we first expand the text in order to evaluate by using the DREx method [1]. After that, we can represent the texts using topics. We can have multiple topics representations depending on the topic identification method we use on. Then, in order to obtain the predictions, we apply a classifier for each representation. As we are going to have more then one polarity for the same text, we use majority voting to determine the sentiment of a given text.

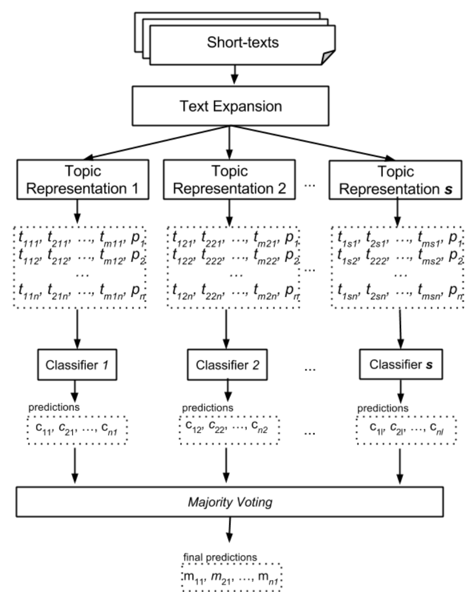

Figure 1: Method

More formally, let $S = \{s_1, s_2, ..., s_n\}$ be a set of short texts we want to evaluate. Each text s_i is represented by a set of m topics $T = \{T_{1k}, T_{2k}, T_{3k}, ..., T_{mk}\}$ when used the topic representation k. Then, we represent the short text as the vector $s_i = (t_{1ki}, t_{2ki}, ..., t_{mki})$ where t_{kji} is the value representing how close the text s_i is to the topic T_{jk} by using the topic representation k. The topic representations are discussed in Section 3.2.

After representing each short text as topics, we can classify each one according to its polarity. To accomplish this, we assume that we have access to some training data $\{(s_1, p_1), (s_2, p_2), ..., (s_o, p_o)\}$ where each pair (s_i, p_i) represents the text s_i and its corresponding sentiment polarity p_i. This polarity can assume 1 for positive (i.e. whether the text transmit something positive) or -1 for negative. Then, for each topic representation k, we create a model using the

the training data and the classifier k in order to evaluate the short texts and, then, obtain the predictions $\{c_{1k}, c_{2k}, ..., c_{nk}\}$.

By doing this we have, for each short text s_i, k polarity values, one for each topic representation (i.e. $\{c_{i1}, c_{i2}, ..., c_{ik}\}$). In order to obtain the final polarity prediction f_i for a given short text i, we apply majority voting. In other words, by using the predictions per topic representation, we count how many times it were assigned positives and negatives polarity for the text and apply the Equation 1.

$$f_i = \begin{cases} -1, \text{ if } n_{in} > n_{ip} \\ +1, \text{ if} n_{in} < n_{ip} \\ pol_{max} \text{ if} n_{in} = n_{ip} \end{cases} \tag{1}$$

where n_{ip} is the the number of times which was assigned a positive polarity to the text i and, n_{in} is the number of times which was assigned the negative polarity, and pol_{max} is the polarity which appeared more in the training set which we use to break ties.

Then, our solution to this problem consists in: (1) determine methods to create the topics representation k $\{T_{k1}, T_{k2}, T_{k3}, ..., T_{km}\}$; (2) choose a classification method to predict the polarity p_i of a any text s_i; and (3) as we can have many combinations of classification and topics representation, we need to choose a combination of classifiers and topic representations in order to use in our method.

We describe our topic representation in Section 3.2 and classification in Section 3.3. Finally, we define our method setup in Section 3.4.

3.1 Text Expansion

In order to identify topics in short texts, it is important to perform a text expansion. We used the DREx method in order to make the expansion because its performance expanding the texts had gains from 1.5% to 15% against the state-of-art methods [1].

Then, let a short text s we want to expand and a vocabulary V we use to make the expansion, in this method, we need to determine which are the documents we use to create V. For the vocabulary V we use either (1) the training set texts, defined as Train; or (2) a external document source from the GloVe repository [10], defined as GloVe. Then, we can use our method considering one of the two proposed sources for V.

3.2 Topic Representation

After expanding the text, we can represent the text in topic by using a topic identification method. We evaluated here three different topic methods. The first method is LDA, the most known technique to discover topics, in which we the implementation used was GibbsLDA++ [11] with the number of iterations set to 50 in order to speedup the experiments. We also used two other methods, BTM [18] and Gibbs MedLDA [19].

These two algorithms are known to be helpful in short texts since the first one discover the topics by using the word co-occurrence and the second, which is a supervised algorithm, by mixing the mechanism of the max-margin prediction models with the mechanism behind the hierarchical Bayesian topic models. These methods were set to run with the default parameters.

3.3 Classifiers

In our approach we used three classifiers from the Scikit-Learn library [8]: Support Vector Machine (SVM) with linear kernel and the C parameter set to 100 which yielded good results for short texts in [5], Logistic Regression and Random Forest.

We tuned these algorithms using Scikit-Learn's `GridSearchCV` method in which the Random Forest was set to run with 5, 20, 50 and 100 trees and Logistic Regression was set to use 1, 10 and 100 as the C parameter. Moreover, we used `balanced` as the `class_weight` parameter of these classifiers, which automatically adjust weights inversely proportional to class frequencies in the input data so we could address the cases of imbalanced data.

3.4 Method Setup

As we can have multiple classifiers and multiple topic representations, we can make different setups for our method.

The first one is when we use a single topic representation and multiple classifiers (i.e. Classifier 1=Linear SVM, Classifier 2=Logistic Regression, and Classifier 3=Random Forest, in Figure 1). In this approach, we use a single topic representation. That is, in Figure 1, Topic Representation 1, 2 and 3 is going to be the same (using one of the three topic identification method Section 3.2). By doing this, we can evaluate the impact of a single classifier in our method.

In the second approach we use multiple topic representations and a single classifier. In this case, we use the three topic representations each one using a different topic identification method (i.e. LDA, MedLDA or BTM). Then, for the classifier, we use just one for inferring the polarity. In other words, in Figure 1, the Classifier 1, 2, 3 and 4 is going to be the same (one of the three classifier presented in Section 3.3). By doing this, our intention is to evaluate the impact of a single topic representation in our method.

We also used our method with multiple topic representations and multiple classifier. To accomplish this, we use all the topic representations (LDA, MedLDA and BTM) and, for each one, we applied all the classifiers presented in Section 3.3.

4 EXPERIMENTS

4.1 Dataset

For our experiments, we used a random sample of Phoenix's business reviews extracted from the Yelp Dataset Challenge.

The dataset was not previously labeled with positive or negative sentiments. It used a different grading system which can be converted into these polarities. More specifically, when a user review an item, they give a score, ranging from 1 to 5 represented by giving a certain number of stars to a review. Closer to 5 the more satisfied the user is to a product or service.

Then, we need to transform these scores into polarities (positive or negative). To accomplish this, we made a similar transformation as [7]. They have shown that users can have different criteria to give a certain score, then they take into account the average score μ_u of the user u. Thus, if a given score is higher than its average, it is considered a positive review, otherwise, it is negative. Specifically, the average μ_u is as shown in Equation 2.

$$\mu_u = \frac{1}{N} \sum_{s_{ui} \in S_u} s_{ui} \tag{2}$$

where S_u is the set of scores s_{ui} of the user u for their review i and N is the total of reviews made by the user u. Then the Equation 3 computes the polarity p_{ui} for the user u and review i.

$$p_{ui} = \begin{cases} -1 & s_{iu} < \mu \\ +1 & s_{iu} \geq \mu \end{cases} \tag{3}$$

In our experiments, we use a sample for the Yelp dataset with 30,886 reviews. We also processed then using SuperSense Tagger (SST), in which we only consider the nouns in order to have a better meaning of the review. All the statistics of these datasets are shown in Table 1.

Table 1: Datasets Statistics

Datasets	# reviews	Word Count	Average words per review
Yelp	30,886	3,305,591	107.03
Yelp SST	32,812	1,031,958	31.45

4.2 Evaluation Methodology

To evaluate our experiments, we utilized an n-fold cross validation procedure. Each dataset was randomly split into n parts, such that, in each run, one part was used as a test set, one part was used as the validation set for parameter tuning and the remaining parts were used as the training set. The split on training and test sets was the same in all experiments. For all datasets we used 10-fold cross validation.

Since the topic modeling algorithms require to inform the number of topics, we decided that, in order to evaluate how each method are affected by the number of topics, we chose from 10 topics to 100 with a step size of 10.

Our work was compared to the Bag of Words method, where we calculated the TF-IDF score for each word in our corpus generating a sparse matrix where each column represent a word and each line a document. We then use these values as features for our classifiers.

In order to analyze the performance of our method we computed the F1 measure as:

$$\text{Macro-F1} = \frac{1}{q} \sum_{\lambda=1}^{q} \frac{P_\lambda R_\lambda}{P_\lambda + R_\lambda} \tag{4}$$

In which precision, P, is the proportion of instances correctly predicted for the class λ and recall, R, is the proportion of instances of class λ correctly classified. To calculate this metric we used the Sklearn `f1_score` method with the default parameters. The parameter chosen for this metric was `macro` which means it returns the macro average.

4.3 Research Questions

4.3.1 Could only expanding the text lead to better results? Many of the topic modeling algorithms have problems to deal with short texts because of the data sparsity problem, where we have few words to model a mixture of topics. To address that, we have expanded our reviews and compared the Bag of Words representation for the expanded review with the normal review so we could analyze if there is some gain while inferring the sentiment.

In Table 2 we have the Macro-F1 results for our dataset, where the highlighted values are the best obtained. We can note that the expansion did not lead to better results since the best value obtained was for the dataset with non-expanded reviews. Moreover, the experiments also showed that using an external pre-trained set of word vectors gives better performances.

Table 2: Bag of Words Comparison

Yelp Dataset			
Expansion	Linear SVM	Random Forest	Logistic Regression
No expansion	0.656	**0.753**	0.659
Train	0.594	0.481	0.664
GloVe	0.622	0.589	0.706

4.3.2 What is the best topic model? It is known that many of the topic models deal better with long texts. Moreover we can compare if expanding the text can lead to a better topic inferring and therefore a better sentiment analysis. Table 3 shows the results applied on both expanded and non-expanded texts and separated by the expansion approach in which the highlighted results are the best ones.

We can see that the results of the expansion were not enough to overcome the ones obtained with the normal text. We can note that in Table 3, we could achieve 82.7% average Macro-F1 per topic with LDA, while for the expanded review, our best approach, LDA, yielded 81% of Macro-F1, a small difference of 2.1%. Therefore we can conclude that LDA was the best topic model since it was the one that yielded the highest Macro-F1 values.

Since we were able to notice that our approach of using the training set for expansion did not yielded satisfactory results compared to the external GloVe dataset, from now on we are going to show only the results for the GloVe expansion approach.

Table 3: Topic Modeling Comparison for Yelp Dataset

Topic Model	Expansion Approach	Linear SVM	Random Forest	Logistic Regression
LDA	No expansion	0.577	**0.814**	**0.827**
	GloVe	0.749	0.810	0.746
	Train	0.591	0.795	0.618
MedLDA	No expansion	0.761	0.759	0.759
	GloVe	**0.793**	0.797	**0.797**
	Train	0.775	0.780	0.729
BTM	No expansion	**0.560**	**0.800**	**0.617**
	GloVe	0.522	0.793	0.586
	Train	0.552	0.793	0.584

4.3.3 Which is the best classifier when using our approach with multiple topic representations and a single classifier? How does it compare to the best Bag of Words approach? Since the expansion of the text did not give an expressive result across all datasets, we modeled the problem by combining the predictions using the Majority Voting. Table 4 shows the results obtained after we applied the Majority Voting using the multiple topic representation and a single classifier. The highlighted results are better than the Bag of Words approach.

Table 4: Results when using multiple topic representation and a single classifier for the Yelp dataset. Best baseline result: 0.753

Topics	Linear SVM		Random Forest		Logistic Regression	
	Expanded	Not Expanded	Expanded	Not Expanded	Expanded	Not Expanded
10	**0.778**	0.388	**0.803**	**0.803**	**0.785**	**0.817**
20	**0.769**	0.502	**0.803**	**0.803**	**0.780**	**0.814**
30	**0.774**	0.483	**0.803**	**0.807**	**0.779**	**0.816**
40	**0.768**	**0.813**	**0.803**	**0.804**	**0.785**	**0.812**
50	**0.789**	**0.817**	**0.804**	**0.803**	**0.794**	**0.817**
60	**0.784**	**0.817**	**0.804**	**0.805**	**0.789**	**0.817**
70	**0.786**	**0.764**	**0.804**	**0.808**	**0.794**	**0.764**
80	**0.762**	**0.816**	**0.808**	**0.803**	**0.774**	**0.816**
90	**0.786**	0.702	**0.807**	**0.811**	**0.789**	0.718
100	**0.774**	**0.815**	**0.808**	**0.803**	**0.782**	**0.817**

Concerning the Yelp Dataset, in Table 4, we can conclude that our approach of combining the predictions of different methods for the same classifier without using expansion achieved the best results obtaining a gain of 8.5% compared to the baseline, in which both Linear SVM and Logistic Regression achieved the best score of 81.7% of Macro-F1 against 75.3% of the baseline. As we can see, the expanded dataset also performed better than the baseline, but not yielding better values of Macro-F1 than the non-expanded version of our method. Furthermore, it is important to notice that Random Forest was always better than the Bag of Words since it achieved a minimum Macro-F1 of 80.3% while the best baseline was 75.3%. In addition, we could observe that with the expansion we had more results overcoming the baseline.

4.3.4 What is the best classifier when using multiple topic representations and a single classifier approach? How does it compare to our best baseline?

We have used three different classifiers to learn and predict the sentiment from the topic representation and therefore we can point out which one was the best in our experiments. Figure 2 shows the performance of each classifier in our single topic multiple classifiers approach. As we can see, Random Forest performed best for the both unexpanded and expanded Yelp datasets, always being better than the baseline in all the topics tested.

4.3.5 Which is the best topic representation when using our approach of a single topic representation and multiple classifiers? How does it compare to the best bag of words approach?

We can also infer which topic model was the best in order to infer the sentiment. To accomplish this, we present in Figure 3 all the methods evaluated in this study. We can see that MedLDA yielded better results, although it had a some variable performance in the experiments. On other hand, LDA produced good results with a more stable behavior. Moreover, we can also note that the methods in general achieved better results with the expansion, in which we observed that more results overcame the baseline. Furthermore, we concluded that the best topic model is MedLDA since its results were more consistent across the experiments.

4.3.6 Which one is the best approach? If we compare all the topic representations, can we improve the result?

In Figure 4 we have evaluated the combination of all topic models and all the classifiers. By doing this, we expected to improve the performance when some topic representation or classifiers did not

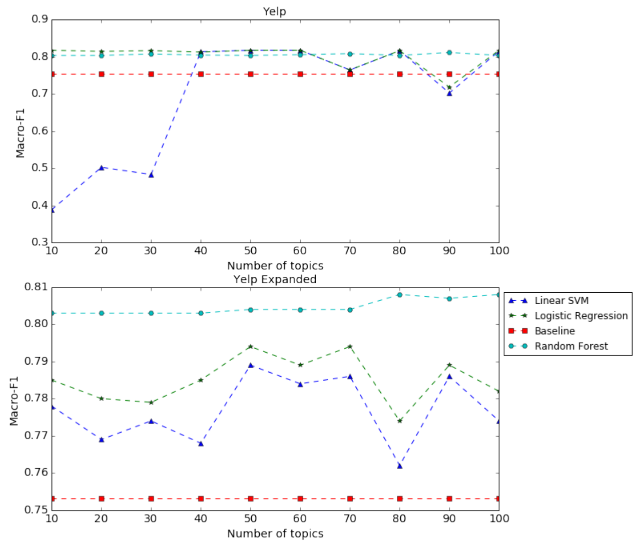

Figure 2: Classifiers Comparison

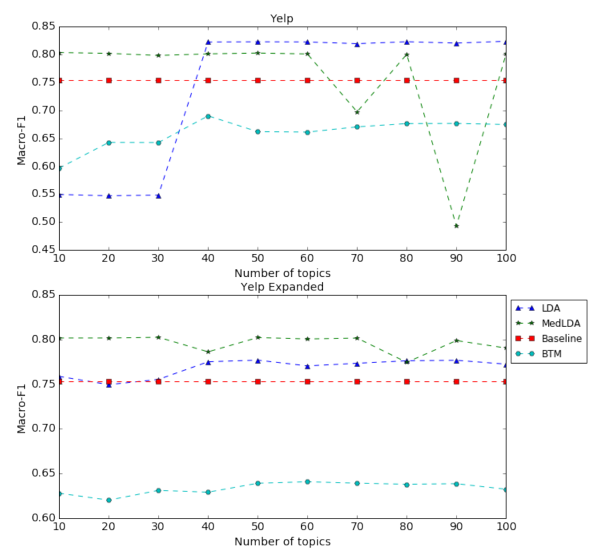

Figure 3: Topic modeling comparison

performed well. We can observe that for the unexpanded dataset we had better results using our single classifier multiple topic representation approach since Random Forest appeared with the highest values. For the expanded dataset we saw a different behavior as our approach of combining multiple classifiers and multiple topic representations, represented by Total, performed better than the other techniques used. It is interesting to notice that the performance of the multiple classifiers and multiple topic representations is similar to the single classifier multiple topic representation, which means that combining the predictions of the same classifier and using all predictions lead to similar results.

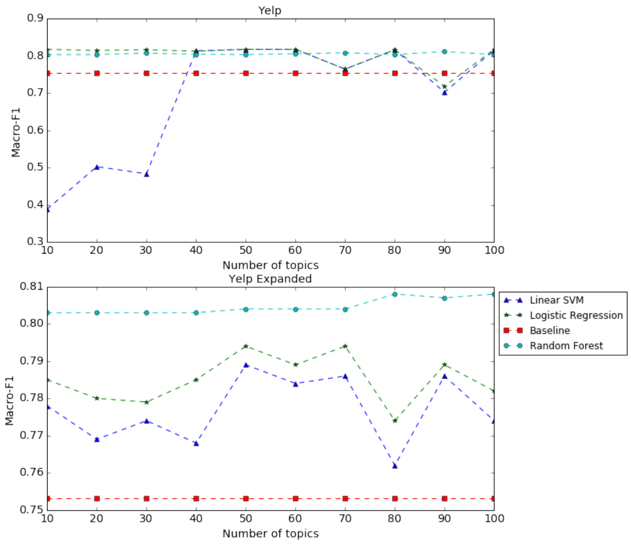

Figure 4: Approaches comparison

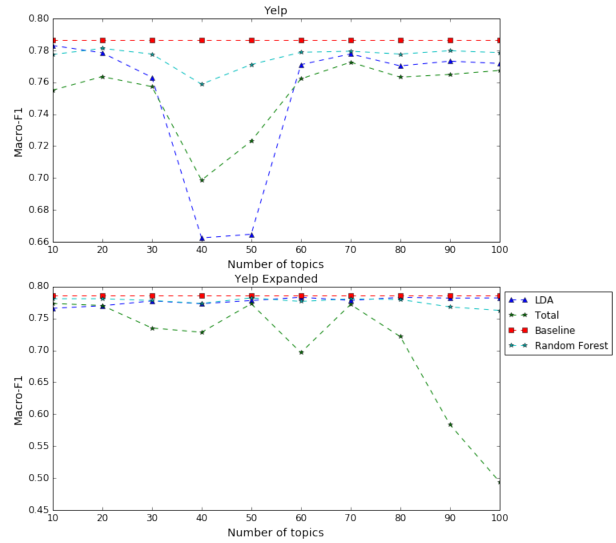

Figure 5: Approaches comparison for SST processed datasets

4.3.7 *What is the impact of using only nouns in our approaches?*

Figure 5 shows the results for both datasets but preprocessed with the SuperSense Tagger (SST), in which we aimed to analyze if only using nouns would result in a better performance. We can note that for the unexpanded and expanded datasets, we could not achieve results that clearly overcame the respective Bag of Words. However we can see that the expansion really made a difference regarding the number of results that were superior to the baseline.

This approach of using only nouns did not yield better results compared to the datasets with all the words because we could not see expressive results. For instance, we had results that were very close to the baseline but not enough to yield a great performance. Therefore this approach proved to not be very helpful for our study.

5 CONCLUSIONS AND FUTURE WORK

We proposed a Sentiment Analysis Method in which we represented the text of reviews as topics in a lower dimensional space for both non-expanded and expanded texts. Our results showed that only expanding the text was not enough to obtain some gains while inferring the sentiment using the Bag of Words approach. Moreover, regarding our expansion technique, we noted using an external dataset of pre-trained word vectors is better than using a set made from the training data.

Our single classifiers, multiple topic representations approach achieved good results expanding or not the text with both Linear SVM and Logistic Regression overcoming by 8.5% the baseline. Therefore we can validate that this approach in fact can be helpful to identify more accurate sentiment in a short text. Moreover the expansion helped in this process, as we have more results overcoming the baseline and also a better performance in the topic models.

Our multiple classifiers, single topic representation approach also allowed us to infer which classifier performed the best and which method yielded the best topic representation. For the best classifier we concluded that Random Forest was the best for since it had the highest average Macro-F1 per topic. For the topic representation we concluded that MedLDA was the best since it produced a greater average Macro-F1 across the topics in the experiments conducted.

Moreover we compared our approaches including the one with multiple classifiers and multiple topic representations and we concluded that for the non-expanded dataset, our single classifier multiple topic representations approach performed best since Random Forest yielded the best results. On the other hand, for the expanded reviews we noted that our approach of combining multiple classifiers and multiple topic representations produced more consistent results,

Furthermore, using only nouns to discover sentiments seemed to not be a good technique for our study since it did not yield expressive results, for both expanded and non-expanded texts, as we could not see a major improvement compared to the approach using all the words.

During this study, we observed some opportunities for future work. For instance, one of them is an approach that identifies, in the training process, the best number of topics, which can improve the results since in this work we deal only with a scale from 10 to 100.

We can also propose to use other text indicators in order to verify if we can achieve better results. Another proposal is to analyze the impact of our approaches in more datasets like Twitter. Finally, we also intend to explore other approaches for combining the classifiers, like the ensemble.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support of CEFETMG; CNPq under Proc. 431458/2016-2; and FAPEMIG under Procs. Universal FAPEMIG, process number APQ-03445-16, and FAPEMIG-PRONEX-MASWeb, Models, Algorithms and Systems for the Web, process number APQ-01400-14.

REFERENCES

[1] Paulo Bicalho, Marcelo Pita, Gabriel Pedrosa, Anisio Lacerda, and Gisele L Pappa. 2017. A general framework to expand short text for topic modeling. *Information Sciences* 393 (2017), 66–81.

[2] Guanhao Chen, Yan Wan, and Xiaoxin Xu. 2016. An analysis of the sales and consumer preferences of e-cigarettes based on text mining of online reviews. In *2016 3rd International Conference on Systems and Informatics (ICSAI)*. IEEE. https://doi.org/10.1109/icsai.2016.7811105

[3] Christine Day. 2015. The Importance of Sentiment Analysis in Social Media Analysis. Internet website. (2015). https://www.linkedin.com/pulse/importance-sentiment-analysis-social-media-christine-day?published=u

[4] Andrea Esuli and Fabrizio Sebastiani. 2007. SENTIWORDNET: A high-coverage lexical resource for opinion mining. *Evaluation* (2007), 1–26.

[5] Svetlana Kiritchenko, Xiaodan Zhu, and Saif M Mohammad. 2014. Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research* 50 (2014), 723–762.

[6] Saif M Mohammad and Peter D Turney. 2013. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence* 29, 3 (2013), 436–465.

[7] Gustavo Arroyo Figueroa Obdulia Pichardo Lagunas, Oscar Herrera Alcántara. 2015. Experimental Results. In *Advances in Artificial Intelligence and Its Applications: 14th Mexican International Conference on Artificial Intelligence, MICAI 2015, Cuernavaca, Morelos, Mexico, October 25-31, 2015, Proceedings, Parte 2*. Springer, 99–100.

[8] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[9] Gabriel Pedrosa, Marcelo Pita, Paulo Bicalho, Anisio Lacerda, and Gisele L Pappa. 2016. Topic modeling for short texts with co-occurrence frequency-based expansion. In *Intelligent Systems (BRACIS), 2016 5th Brazilian Conference on*. IEEE, 277–282.

[10] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. In *Empirical Methods in Natural Language Processing (EMNLP)*. 1532–1543. http://www.aclweb.org/anthology/D14-1162

[11] Xuan-Hieu Phan and Cam-Tu Nguyen. 2007. GibbsLDA++: AC/C++ implementation of latent Dirichlet allocation (LDA). *Tech. rep.* (2007).

[12] Kumar Ravi and Vadlamani Ravi. 2015. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems* 89 (2015), 14–46.

[13] Renata Lopes Rosa, Demostenes Zegarra Rodriguez, and Graca Bressan. 2013. SentiMeter-Br: A new social web analysis metric to discover consumers' sentiment. In *2013 IEEE International Symposium on Consumer Electronics (ISCE)*. IEEE. https://doi.org/10.1109/isce.2013.6570158

[14] Renata L. Rosa, Demsteneso Z. Rodriguez, and Graca Bressan. 2015. Music recommendation system based on user's sentiments extracted from social networks. *IEEE Transactions on Consumer Electronics* 61, 3 (aug 2015), 359–367. https://doi.org/10.1109/tce.2015.7298296

[15] Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, Christopher Potts, et al. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*, Vol. 1631. Citeseer, 1642.

[16] Mike Thelwall. 2013. Heart and soul: Sentiment strength detection in the social web with sentistrength, 2013. (2013).

[17] Hao Wang, Dogan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan. 2012. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations*. Association for Computational Linguistics, 115–120.

[18] Xiaohui Yan, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng. 2013. A biterm topic model for short texts. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 1445–1456.

[19] Jun Zhu, Ning Chen, Hugh Perkins, and Bo Zhang. 2013. Gibbs Max-Margin Topic Models with Fast Sampling Algorithms.. In *ICML (1)*. 124–132.