Phrase based Sentiment Analysis using topical n-grams

Abstract

Sentiment Analysis is mainly the classification of text into two classes viz. positive and negative. Joint sentiment and topic models have been used to tackle this classification problem. Despite having a hierarchical structure, these generative models are essentially extensions of bag of words model. Due to this fact, they tend to misclassify texts having sentiment in the form of phrases. LDA and it's extensions don't work properly with phrases. To tackle this situation, we propose an unsupervised approach to sentiment analysis using topical n-grams which have been shown to be effective with phrases. We train the topical n-grams model using two topics i.e; positive and negative, list of positive and negative words, and rules to detect positive and negative phrases. New documents are then classified using this trained model. The system gives better results than the existing Joint Sentiment Topic model. We also propose an approach to generate list of positive and negative words using LDA.

1 Introduction

LDA (Latent Dirichlet Allocation) as shown by Blei et al. (2003) is generative model used to discover topics in a document collection. It gives two types of distributions as output, document-topic and word-topic distributions. The document-topic distribution gives the proportion of topics in each document and the word-topic distribution gives the probability of a word being in each topic. LDA works on the principle of co-occurrence. It assumes that words tending to appear together belong to the same topic. JST (Joint Sentiment Topic) as explained by Lin and He (2009) is a probabilistic generative model which extends LDA and discovers both sentiment and topic simultaneously in a document collection. JST has shown promising results on binary sentiment classification.

There are many extensions of the basic LDA model which try to combine both sentiment and topic to solve the problem of sentiment analysis. All these models including LDA have one underlying assumption which makes them unsuitable for text classification purposes. They assume that each word is generated separately and independent of other words. This is essentially the bag of words assumption. However, text being a sequence of words, the correct meaning of the text cannot be understood by merely capturing co-occurrences. In addition to this, we also need to consider collocation of words. A phrase is a collocation of words which usually has more meaning than the individual words making up that phrase. There is a subtle difference between a phrase and collocation of words. Not all collocations of words can be considered as a phrase. We need a model which takes into account phrases to completely understand the meaning of the text.

Topical n-grams as described by Wang et al. (2007) is one such generative model which takes into account not only co-occurrences but also collocations of words. It also decides whether a particular collocation of words should be considered as a phrase or not. We train this model using 2 topics viz. positive and negative, using prior list of positive and negative words, and some rules to identify the subjective nature of phrases. The phrases in our experiments are restricted to bigrams. We then use this trained model to infer the topic distribution for new documents. The topic having higher proportion is considered to be the class of the document.

Rest of the paper is organized as follows. Section 2 discusses related work in this area. Section 3 explains our process in detail. The experimental setup is explained in Section 4. Results of the experiments are presented in Section 5. Section 6 discusses these results followed by explanation of the use of topic models for resource generation in Section 7. Section 8 concludes and hints to some future work are given in Section 9.

2 Related Word

There are many supervised, semi-supervised and unsupervised approaches to solve the sentiment analysis problem. One supervised method based on n-gram analysis is explained by Bespalov et al. (2011). In this they map the n-grams to a low-dimensional latent semantic space where a classification function can be defined. Our approach being unsupervised, we will go through some of the related work in that direction. One more motivation to consider only semi-supervised and unsupervised approaches is the fact that they don't need any corpus and don't have any domain specific limitations. Rule based systems can also be considered as unsupervised systems but usually due to exceptions to these rules, their performance is affected. Due to this, we are not considering them in further discussion.

Turney and Littman (2002) used an unsupervised learning algorithm based on mutual information between document phrases and a small set of positive/negative paradigm words called seed words to classify the semantic orientation at the word/phrase level. Another unsupervised approach to classify the text at document level was proposed by Turney (2002). Eguchi and Lavrenko (2006) created a generative model that jointly models sentiment words, topic words and sentiment polarity in a sentence as a triple. Mei et al. (2007) proposed another generative model,called TSM (Topic Sentiment Mixture) model which can be used to discover topics in blogs as well as their associated sentiments. A novel generation model that unifies topic-relevance and opinion generation by a quadratic combination was proposed by Zhang and Ye (2008). A probabilistic generative model based on LDA called JST (Joint Sentiment Topic) model was shown to perform well for sentiment analysis of reviews by Lin and He (2009). It is a fully unsupervised method and shows good result when priors are used for training. Another extension of LDA which tries to unify aspect and sentiment was proposed by Jo and Oh (2011). Most of the unsupervised methods discussed here operate at the word level. Due to this they lose out on the information provided by phrases which may lead to incorrect classification.

3 Phrase based Sentiment Analysis

We will explain the use of basic LDA and Topical n-grams for Sentiment Classification in this section. We won't go into the mathematical details of these models for the constraint of space. Let us first list the basic steps to use any topic model for discovering topics.

3.1 Using Topic models

- Set number of topics.
- Remove stop-words as they do not belong to any topic.
- Estimate probabilities using some inference method.
- Use the trained model for inference of topics in new documents.

3.2 Using LDA for Sentiment Classification

To use basic LDA as a sentiment classifier, we add one more step to remove objective words. Also, during Gibbs sampling the first step assigns topics randomly to words. We introduce a prior information about the positivity and negativity of words to aid in the topic detection. The steps are as follows.

- Set number of topics, 2 in this case viz. positive and negative.
- Remove stop-words.
- Remove objective words as they won't affect sentiment.
- Gibbs Sampling with prior using lists of positive and negative words.
- Use the trained model to classify a new document as positive or negative.

3.3 Using Topical n-grams Model for Sentiment Classification

To make use of topical n-grams model for sentiment classification, we use a similar approach.

- Set number of topics equal to 2.
- Remove stop-words.

- Remove objective words as they won't affect sentiment. The objective words in this case do not include the negation words like *don't, doesn't, won't, no*, etc. This is to ensure that we can catch negation of polarity when they are used with subjective words.
- Gibbs Sampling with prior. The prior used in this case is more sophisticated and can handle both words and phrases. In case of words, it simply uses a list of positive and negative words. There are some rules to detect and assign topics to phrases which are explained next.
- Use the trained model to classify a new document as positive or negative.

Rules for Topic assignment of phrases

At present, our rules are restricted to bigrams. We plan to extend them as explained in Section 9. In the following rules, we mean topic when we say polarity. The use of polarity makes it easy to understand the rules as they are concerned with subjectivity.

1. If the first word in the bigram is a negation word and the second word is subjective then the polarity of the bigram is opposite to the polarity of the second word.
   Examples: *won't like, won't regret, etc.*. Here, *won't like* is assigned negative polarity and *won't regret* is assigned positive polarity.
2. If both the words in the bigram are subjective then are two cases. If both words are of the same polarity then resultant polarity is the same. But if their polarities are different, then the polarity of the first word is assigned to the bigram.
   Examples: *beautifully amazing* is positive as both words as positive. *lack respect* is assigned negative as per the rules.

4 Experimental Setup

Analysis was performed for binary sentiment classification task. The language used in this case was English. We conducted experiments on 4 models, BOW using SVM, LDA (Blei et al., 2003), JST (Lin and He, 2009), and Topical n-gram model (Wang et al., 2007). We used two settings for the topic models, with and without prior. The word lists used are those specified in (Bing Liu, 2010). The implementations of SVM, LDA and Topical n-gram in Mallet [1] have been used for evaluation. For JST, we have used the implementation provided by the authors [2]. The default settings for the hyper-parameters were used in all these implementations.

Dataset

To create the dataset we made use of the amazon reviews dataset provided by SNAP [3]. These reviews are not tagged with sentiment but they have ratings from 1 to 5. We used this information to create a sentiment tagged corpus. The reviews with ratings less than 3 were tagged as negative and others were tagged as positive. We conducted the experiments on 6,00,000 reviews containing equal number of positive and negative reviews.

5 Results

The results presented here are for 10-fold cross validation. For the topic models, due to the randomness involved during sampling, the best result obtained for each fold has been used for calculating the average. The Bag of words system performs better than all the models when no prior information is provided. The performance of topic models significantly increases when prior information is provided. Our approach to use Topical n-gram outperforms all the systems when a prior is used.

System	**Avg. accuracy (%)**
BOW-SVM	82.45
LDA	65.34
LDA with prior	80.19
JST	68.64
JST with prior	84.43
Topical n-gram	63.57
Topical n-gram with prior	**87.32**

Table 1: Evaluation of systems

6 Discussion

The performance of our system is better due to the capacity to handle phrases. All the other systems, consider each word separately. The performance improvement over JST is 3% which is statistically significant. Though the system performs

[1] http://mallet.cs.umass.edu/

[2] https://github.com/linron84/JST

[3] http://snap.stanford.edu/

better for this dataset, it still has some obvious limitations. The system highly depends on the rules used for initial assignment of topics which is evident from the results. These rules don't apply to all the bigrams. Let us consider, the bigram *insanely good*. In this case, the first word is negative and second word is positive. The rules will assign a negative polarity to this bigram. But in this phrase *insanely* is used to increase the intensity of *good*. The rules apply only for bigrams, we need to add more rules to handle n-grams. A phrase like *I don't think it's good* won't be handled by the system. During error analysis, we also found that some words like *engrossing, blockbuster, bravo, etc.*, are not present in the word lists. These words convey a positive sentiment but our system fails to correctly classify reviews containing such words. Also, words like *awsome* which is spelling mistake of *awesome* are often found in reviews. We thought that the accuracy might be increased if we could detect the correct subjective nature of these words and also the bigrams using them. For this, we propose an approach for resource generation using LDA.

7 Resource generation using LDA

LDA can be used for resource generation of positive and negative words. The steps to do so are explained below.

- Set number of topics equal to 3 i.e., positive, negative and objective. We do not remove the objective words in this case as we want to find out which of them are positive or negative.
- Remove stop-words.
- Gibbs Sampling with prior using lists of positive and negative words. In the initial step, words present in the list are assigned that specific topic bu
- Get the top words in the positive and negative topics.

The word lists we generated using this approach were used to test the systems using priors. The results for the same are shown in Table 2. We can see a marginal increase in the accuracy in this setting.

System	Avg. accuracy (%)
LDA with prior	80.21
JST with prior	86.37
Topical n-gram with prior	**89.83**

Table 2: Evaluation of systems

8 Conclusion

This paper made use of topical n-gram model for the binary sentiment classification problem. The motivation behind this approach was to use not only words but also phrases to classify documents. The model was trained with documents containing only subjective and negation words and bigrams formed by the combination of these. It also, made use of rules to assign topics to words and bigrams in the intialization stage of estimation. Results using a prior show statistically significant improvement over the JST model. The paper also shows the use of LDA for resource generation, prompted by observations made during error analysis. The system shows marginal improvement in accuracy after using the resources generated by this technique.

9 Future Work

The system is focused on bigrams at the moment. It can be extended to handle n-grams by adding rules to detect and assign topics. Instead of adding rules, we can make use of machine learning to detect the subjective nature of a phrase.

References

Bespalov, Dmitriy and Bai, Bing and Qi, Yanjun and Shokoufandeh, Ali 2011. *Sentiment classification based on supervised latent n-gram analysis. Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 375-382. ACM.

Jo, Yohan and Oh, Alice H 2011. *Aspect and sentiment unification model for online review analysis. Proceedings of the fourth ACM international conference on Web search and data mining*, pages 815-824. ACM.

Zhang, Min and Ye, Xingyao 2008. *A generation model to unify topic relevance and lexicon-based sentiment for opinion retrieval. Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 411-418. ACM.

Mei, Qiaozhu and Ling, Xu and Wondra, Matthew and Su, Hang and Zhai, ChengXiang 2007. *Topic sentiment mixture: modeling facets and opinions in weblogs. Proceedings of the 16th international conference on World Wide Web*, pages 171-180. ACM.

Eguchi, Koji and Lavrenko, Victor 2006. *Sentiment retrieval using generative models*. *Proceedings of the 2006 conference on empirical methods in natural language processing*, pages 345-354. ACM.

Turney, Peter D 2002. *Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews*. *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417-424. Association for Computational Linguistics.

Turney, Peter and Littman, Michael L 2002. *Unsupervised learning of semantic orientation from a hundred-billion-word corpus*. *Handbook of natural language processing*, volume-2, pages 627-666. CoRR, cs.LG/0212012.

Liu, Bing 2010. *Sentiment analysis and subjectivity*. *Handbook of natural language processing*, volume-2, pages 627-666. Chapman & Hall.

Wang, Xuerui and McCallum, Andrew and Wei, Xing. 2007. *Topical n-grams: Phrase and topic discovery, with an application to information retrieval*. *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*, pages 697-702. IEEE.

Lin, Chenghua and He, Yulan. 2009. *Joint sentiment-topic model for sentiment analysis*. *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 375-384. ACM.

Blei, David M and Ng, Andrew Y and Jordan, Michael I. 2003. *Latent dirichlet allocation The Journal of machine learning research*, volume-3, pages 993-1022. JMLR.