Classification based approach for Summarizing Opinions in Blog Posts

by

Kiran Sarvabhotla, B. Kranthi Reddy, Vasudeva Varma

in

4th Indian International Conference on Artificial Intelligence

Tumkur, near Bangalore, India

Report No: IIIT/TR/2009/253

Centre for Search and Information Extraction Lab
International Institute of Information Technology
Hyderabad - 500 032, INDIA
December 2009
Classification based approach for Summarizing Opinions in Blog Posts

Kiran Sarvabhotla, Kranthi Reddy, Vasudeva Varma

International Institute of Information Technology, Hyderabad, India

Abstract. With the growth of web, people are using it as a medium for expressing their opinions, thoughts through blog posts, reviews (in the form of ratings), and forums. Blogosphere is a place where people read, write their views and make comments on others views or thoughts there by exchanging information. It will be very difficult for any business, organization or individual to go through and understand thoughts expressed by others on a product or topic which they are interested in. Hence a summarization system which extracts, analyze and summarize opinions will be useful. Our Summarization system exactly does the same for blog posts. The entire process of summary generation is done in three stages, extract sentences which are sources for opinion (Opinion Mining), then analyze the extracted opinions to determine polarity (Opinion Analysis) and finally ranking the opinion sentences (Opinion Summarization). We present a classification based approach for extracting and analyzing opinions and how we use this approach to rank sentences for generating summary.

Keywords: Classification, Word Association, Polarity Estimation.

1 Introduction

Summarizing opinions [3] expressed by users in blogs will be useful for many businesses and organizations where they analyze the sentiments [1] of the people on a product, for individual(s) who are curious to know opinions of other people. It will be very difficult for any business, organization or individual to go through and understand thoughts expressed by others on a product or topic which they are interested in. Hence a summarization system which extracts, analyze and summarize the opinions will necessitate their requirements. Opinion summarization differs from normal text summarization [5, 6] from the fact that we have to extract opinions and analyze them [4]. Our system was developed keeping in view of the above requirements.

We define the term Polarity Estimation which will be used in our summary generation as a feature for ranking opinions in blog posts. In addition to Polarity Estimation feature we use Query Independent (QI) and Query Dependent
(QD) features for ranking [10, 11]. In this paper we present a novel approach for extracting and analyzing opinions [4] at sentence level.

To predict a sentence as an opinion or fact and to determine the polarity (positive or negative) [9] of an opinion sentence is a challenging task. The notion of opinion and polarity varies from context to context. Against traditional dictionary based and language processing approaches where set of rules and dictionaries are used, we modeled a statistical approach using a two class classifier which classifies each sentence of a blog post as an opinion or non-opinion and then determine polarity associated with each opinion sentence (positive/negative) [4, 9]. We will explain the features used in classification and the procedure in Our Approach section in detail.

2 Paper organization

The whole paper is divided into six sections. In section 1, we established the need for an Opinion Summarization [4] system and we have introduced the exact problem and our approach. In section 2, we provide details on work related to Opinion Summarization. In section 3, we describe various corpora used in training and testing phases. In section 4, we describe our approach which includes various features we used in building our system. In section 5, we explain how we tested and evaluated our system with Text Analysis Conference(TAC) 2008 1 data and conclude our work by providing future directions where we can extend our work in section 6.

3 Related Work

Previously research scientists have determined the orientation of a sentence being an opinion by framing a set of rules, preparing seed list of words [4] which will determine the semantic orientation of sentences and dictionaries which are exclusively used for opinion analysis. One of them is Senti Wordnet 2 with pre-computed scores for each word which are the sources of subjectivity in a sentence. Other language processing techniques include tagging based approaches [7] to extract features of a sentences which is an opinion. They are opinion holder [2, 8] (person or organization who holds opinion), object [2, 8] (features on which opinions are expressed) and opinion words [2, 8]. Linguist’s use these tagging approaches for opinion analysis.

But problems with traditional dictionary based approaches are:

- Dictionaries can not disambiguate the semantics of a sentence by themselves as the notion of opinion and polarity differs from context to context.

1 http://www.nist.gov/tac/tracks/2008/summarization/
2 http://sentiwordnet.isti.cnr.it/
– Scalability.
– Computationally expensive.

To make systems which are scalable, computationally inexpensive, researchers started focusing on statistical learning based approaches using SVM, Naïve Bayes or Maximum Entropy [8, 12]. But these methods are computationally heavy and time consuming. We will present a classification approach using word association as a feature which is computationally fast and highly scalable as there is no heavy computation.

4 Corpus Description

Training data for opinion/non-opinion classification: We have downloaded IMDb (Internet Movie Database) movie data which have sentences tagged as opinions and facts/non-opinions.4

<table>
<thead>
<tr>
<th>Class</th>
<th># samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opinions</td>
<td>5000</td>
</tr>
<tr>
<td>Facts</td>
<td>5000</td>
</tr>
</tbody>
</table>

Table 1. Opinion/Fact Data

Training data for Polarity Estimation: We crawled reviews of about 1,30,000 on several products from Amazon5 using nutch6 crawler. We have tagged each review based on the ratings given at the end of each review. All the reviews with ratings 4 or 5 tagged as positive and those with either 1 or 2 or 3 tagged as negative.

<table>
<thead>
<tr>
<th>Class</th>
<th># samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>90000</td>
</tr>
<tr>
<td>Negative</td>
<td>40000</td>
</tr>
</tbody>
</table>

Table 2. Polarity Data

Samples of training data

– Opinion: You really have to salute writer-director hanoke for making a film that isn’t nearly as graphic but much more powerful, brutally shocking and difficult to watch.

3 http://www.imdb.com
4 http://www.cs.cornell.edu/people/pabo/movie-review-data/
5 http://www.amazon.com
6 http://lucene.apache.org/nutch/
– Fact or non-opinion: Soon, the team begins to suspect that Knowles main objective is actually to recover the prototype of a DNA testing machine called the Huxley project, which his company has spent years and millions of dollars developing.

– Positive: A good read. Well worth the time to get a new perspective on our ”leaders” that you will not get from the nightly news or CNN.

– Negative: The book by Dick Morris is pure Fox News, Anti Progressive Garbage. Do not waste your money on this propaganda crap!

Testing data (Blog Posts): We have used TAC 2008 Opinion Pilot Task data. It has 25 topics and each topic has one or two queries followed by a set of blog posts where answers for the queries are likely to be found. Each query is a Squishy List question which expects descriptive answers. A Query classifier which predicts the orientation of query terms has been built for this task only by using a seed list of positive and negative words. The seed list contains very few positive and negative words and it does not cover the entire vocabulary of opinion words. It is collected for query classification exclusively for TAC task.

Samples of testing data

– Topic: Windows Vista

– Query: What features do people like about Vista? Positive aspect of Vista with presence of word like

– Query: What features do people dislike in Vista? Negative aspect of Vista with presence of word dislike

5 Our Approach

To determine each sentence as an opinion or not and to determine its polarity, we have used a classification based approach with bag of words as features.

5.1 Word Association and Rainbow Classifier

Word Association a variant of bi-gram is used as the feature for classification. We tokenize each sentence in opinion/fact and polarity training data into words and associate each token with all other tokens in the sentence. The motivation behind this approach is that the characteristic of opinion or polarity of a sentence is not determined by a single token; rather it is the combination of tokens or bag of words which determines it. The sentence classification is done in two phases. In the first phase it is an opinion or non-opinion classification and in the second phase sentence which is an opinion is classified as positive or negative. We built models for classification using Rainbow\textsuperscript{7} text classifier. It has several in built methods for classification like Naive-Bayes, SVM, KL, TFIDF, k-nearest neighbor and Probabilistic Indexing [15]. We have used Probabilistic Indexing\textsuperscript{15} as the classification method.

\textsuperscript{7} http://www.cs.cmu.edu/~mccallum/bow/rainbow/
5.2 Algorithm:

for each review do
    Tokenize(R) $\rightarrow$ Sentences(S)
    for each sentence $s$ in S do
        Tokenize(S) $\rightarrow$ Words(W)
        for each word $w$ in W do
            associate ($w$, $W'$) where $W'$ is word vector not containing word $w$. And submit it to the classifier for learning.
        end for
    end for
end for

5.3 Polarity Estimation Feature for Ranking

Each sentence which is classified as an opinion or positive/negative will have a score which will be the likelihood of the sentence belonging to that class. We used these scores to estimate the polarity associated with each sentence and used it as a feature to rank opinion sentences.

Polarity Estimate: $PE(S|C) = 0.3 \times P(S|O) + 0.7 \times P(O(S)|C)$

Where $S$ is the input sentence, $O(S)$ is the opinion score of a sentence (sentences which are classified as opinions by the classifier), $O$ is opinion and $C$ is polarity class.

5.4 Testing our System

We tested our system on TAC 2008 data set (blog posts) which were provided for Opinion Summarization pilot task. As explained earlier in section 2, each topic in this task has one or two queries. We predict the polarity class($C$) of the query by determining whether the query focuses on positive aspect of the topic or its negative aspect. The class which is predicted here will act as a filter for the above Polarity Estimate($PE$) feature. A positive class predicted above will estimate the polarity score of each opinion sentence being positive and negative class predicted above will estimate the polarity score of sentence being negative.

In addition to this polarity feature we have used another two features to rank sentences. They are:

- Query Independent Feature($QI$)
- Query Dependent Feature($QD$)

Query Dependent ($QD$) feature boosts the sentences which are most relevant to the query (Only if query is given).

Query Independent ($QI$) feature boosts the most informative sentences in the given content using KL divergence [13].
The final score for each sentence will be a linear combination of the above three features with appropriate smoothing factor or weights associated with each feature.

\[
\text{Score}(S) = \lambda_1 \times QI + \lambda_2 \times QD + \lambda_3 \times PE.
\]

Where QI, QD and PE are Query Independent, Query Dependent and Polarity Estimate features and \( \lambda_1, \lambda_2, \lambda_3 \) are their associated weights. We set the values of \( \lambda_1, \lambda_2 \) and \( \lambda_3 \) to 0.35, 0.25, and 0.4 respectively.

The entire process is presented in the figure below.

![Process Flow Diagram](image_url)

We started with arbitrary weights for opinion and polarity feature intuitively giving more importance to polarity score. Since the notion of polarity is associated more with positive or negative feature of a sentence, the polarity estimate
feature will be dependent more on polarity score. Later we carried out experiments on our classifier by giving test sentences and determined how accurately we are able to estimate the polarity of each sentence with different combinations of weights. The above values 0.3 and 0.7 have finally gave better estimates for polarity of each sentence. Same is the case with weights of QI, QD and PE features in final ranking of sentences. Our Query Independent (QI) feature picks most informative sentences from the content, so we have given high priority to it. These weights may not be optimal but we are working on that aspect.

6 Results

We evaluated our system by using "Nugget Judgments" provided for each topic in TAC 2008. Each judgment has a nugget score or nugget weight associated with it. The most relevant judgment to the topic will have high nugget score. Though there are 25 topics in TAC data set, judgments are available for only 22. We evaluated our approach on the judgments given for 22 topics.

**Nugget Recall (NR):** (sum of weights of all nuggets returned in summary/sum of weights of all nuggets related to the topic).

**Nugget Precision (NP):** (Allowance/length) where allowance = 100*number of nuggets returned in the summary and length is the # of non white space characters in the summary.

Table 3. Average scores with $\lambda_1=0.35, \lambda_2=0.25, \lambda_3=0.40$

<table>
<thead>
<tr>
<th>Nugget Recall (NR)</th>
<th>0.287</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nugget Precision (NP)</td>
<td>0.164</td>
</tr>
<tr>
<td>F-Measure ($\beta=1$)</td>
<td>0.209</td>
</tr>
</tbody>
</table>

The average scores for 22 topics are listed in the above table. Average F-measure with ($\beta=1$, giving equal importance to precision and recall) score we obtained through our system is better than scores obtained by many systems submitted to TAC 2008. Out of the thirty automatic runs submitted to the task only nine reported to have better score than our F-Measure with best being 0.489.

7 Conclusion

We have presented a general overview of building a system which extracts, analyzes and summarize opinions from blog posts. We have clearly explained the need for summarizing blog posts and challenges in summarizing them. We have explained the problems with dictionary based approaches and presented a simple classification based approach for summarizing opinions in blog posts. We built a
system using this approach and evaluated its performance. The system is scalable since it did not use any dictionary in classifying sentences as opinions or estimating polarity. Using **Word Association** as a feature in classification and ranking sentences using Query Independent (QI), Query Dependent (QD) and Polarity Estimate (PE) as features, we were able to get better results than many systems submitted to TAC 2008. This indicates that the proposed approach can be tuned to achieve higher results. There can be many extensions to the current approach. Few of them will be, while classifying we are simply associating words. Here we can associate using distributional clustering method by estimating mutual information between words. The words with highest mutual information can be associated together. It can reduce vocabulary size without affecting the accuracy of classifier. Associating words based on new co-occurrence can also be explored for short texts as opinions are short texts [14]. Each sentence in the post is scored using QI, QD and PE features. In addition to these three features, author information, date on which post was written, number of comments for each post can also be used for ranking. So the sentences from the blog post which is written by very active member in the community and from the post which is commented by the most of the people will be given more importance. We have to explore more on this and see how it affects the ranking of sentences. We also have to work on the optimal values for weights of each feature used in ranking sentences.

8 Acknowledgments

We thank National Institute of Standards and Technology (NIST) for providing blog posts data which helped us in testing our system. We thank Rohit Bharadwaj and P.V.Sai Krishna for their help in building the summarization system.

References

7. Xiaowen Ding and Bing Liu. 2007. The Utility of Linguistic Rules in Opinion Mining, SIGIR’07