Proceedings of the 2011 International Conference on Machine Learning and Cybernetics, Guilin, 10-13 July, 2011

A SENTIMENT ANALYSIS MODEL FOR HOTEL REVIEWS BASED ON SUPERVISED LEARNING

HAN-XIAO SHI, XIAO-JUN LI

School of Computer Science & Information Engineering, Zhejiang Gongshang University, Hangzhou China, 310018
E-MAIL: hxshi@mail.zjgsu.edu.cn

Abstract:

As the widespread use of computers and the high-speed development of the Internet, E-Commerce has already penetrated as a part of our daily life. For a popular product, there are a large number of reviews. This makes it difficult for a potential customer to make an informed decision on purchasing the product, as well as for the manufacturer of the product to keep track and to manage customer opinions. In this paper, we pay attention to online hotel reviews, and propose a supervised machine learning approach using unigram feature with two types of information (frequency and TF-IDF) to realize polarity classification of documents. As shown in our experimental results, the information of TF-IDF is more effective than frequency.

Keywords:

Sentiment classification; supervised machine learning; Online reviews

1. Introduction

With the high-speed development of the Internet, E-Commerce has already penetrated as a part of our daily life. Currently most information on the Internet exists in unstructured or semi-structured forms, and presented the explosive growth. Besides, along with the development of E-Commerce, the more appeared reviews not only help potential consumers to make decisions of the products in a certain extent, but also provide some good feedback for merchant. For instance, when a consumer plans to select comfortable hotel for his trip, he will surf on the BBS or review sites to read the opinions from experienced consumers.

However, for a popular product, the number of reviews can be in hundreds or even thousands. This makes it difficult for a potential customer to read them to make an informed decision on whether to purchase the product. It also makes it difficult for the merchant of the product to keep track and to manage customer opinions. Moreover, there is also some noise such as misleading articles often appearing in the first few pages, which would affect the comprehensiveness of browsers' information acquisition and correctness of their judgment. Nowadays, some websites have made quantized expressions for the sentiment orientation (SO) of their local review information, such as *Amazon.com*, which has coarse-grained rating (5-star scale) for each review on its website, and the 5-star is the best, while the 1-star is the worst, then giving the total rating.

In the past few years, many researchers transfer their interests from text classification [1, 2] to sentiment analysis [3,4,5,6,7,8,9]. Current researches mainly focused on proposing novel analyzing and processing technologies based on different domains, according to the large scale review data acquired from Internet, such as 1) Using Part-Of-Speech (POS) to tag the sentences, several researchers summarized some rules to focus the object of opinion and sentiment items[6, 10], and utilized the distribution rules of POS to extract the corresponding template for sentiment analysis [11]; 2) some scholars also started to make researches in sentiment orientation of different sentence structures [12, 13]; 3) Kim and Hovy used the technology of semantic role labeling (SRL), which was mainly utilized in news and public's opinion analysis, to help identify two main components of opinions: opinion holder and topic [14]. Of course there were also some scholars aiming at the reviews in a special website, and proposing special processing methods, so as to obtain the sentiment orientation of reviews [15].

This article mainly has two contributions: 1) to propose a supervised machine learning approach to realize sentiment classification of online hotel reviews; 2) to utilize TF-IDF information to set up unigram feature, this information is more effective than frequency evaluated by our experiment results.

The rest of this paper is organized as follows. Section 2 presents related work. Section 3 introduces the theory of machine learning of support vector machine. Section 4 describes the experiment set and shows the experimental results. Finally, we conclude and prospect our work in Section 5.

978-1-4577-0308-9/11/$26.00 ©2011 IEEE

2. Related work

There is presently a huge amount of online hotel reviews which are beyond the visual capacity of any human beings. Hence, there is an urgent need for innovative techniques that can automatically analyze the attitudes of customers in their reviews. As such, sentiment classification (sentiment analysis or opinion mining) can perform the tasks of automatically understanding the online reviews [3,4]. Mining opinions from reviews on web pages, however, is a complex process, which requires more than just text mining techniques. The complexity is related to a couple of issues. First, data of reviews are to be crawled from websites, in which web spiders or search engines can play an important role. Moreover, it is necessary to separate the data of reviews from non-reviews. The sentiment classification process can then be conducted. Pang et al. found text mining algorithms on sentiment classification do not perform as well as that on traditional topic-based categorization [4]. Topics can be identified by keywords but sentiment would be expressed in a more subtle manner. As such, sentiment classification requires more understanding than the usual topic-based classification [4].

Sentiment classification aims to extract the text of written reviews of customers for certain products or services by classifying the reviews into positive or negative opinions according to the polarity of the review [5,16]. The method has been attempted in different domains such as movie reviews, product reviews, customer feedback reviews, and legal blogs [4,17]. Other potential applications include extracting opinions or reviews from discussion forums such as blogs, and integrating automatic review mining with search engines to automatically provide useful statistical data of search results or to build sentiment analysis systems for specific products or services. Tourist destinations, naturally, would be one of the good application areas. In relation to opinion mining applications, the extant literature indicates two types of techniques have been utilized, including machine learning and semantic orientation [3].

The first one is based on the simple statistic. Turney, Nasukawa and Yi mainly made simple statistics for orientation values to obtain the whole tendency of texts [3, 18]. This method is generally applied to the document-level sentiment analysis, such as Tsou makes statistics for the sentiment orientation of news articles and measures the opinions of celebrities from the public through calculating the sentiment orientation of the words and comprehensively considers the spread, density and semantic intensity of the polarity elements [19]. Although the sentiment analysis based on the simple statistic belongs to coarse-grained orientation classification, because of its simple realization and not bad accuracy, it occupies a particular weight in the beginning of the orientation study.

The second one is based on machine learning, generating orientation classification model through the training of numerous labeled corpuses, and then classifying the test texts using generated model. Pang adopted the technology of standard bag-of-words and three machine learning methods (naive bayes, maximum entropy classifications and support vector machine (SVM)) to make text orientation classification for the film reviews, and respectively compared them with the outcome of manual classification [4,20]. The result of experiment shows that the method of SVM has the best effectiveness among several classifications. Chaovalit and Zhou also used the methods of machine learning and sentiment orientation to deeply mine the film reviews [21]. Mullen and Collier proposed an approach to sentiment analysis which uses SVM to bring together diverse sources of potentially pertinent information, including several favorability measures for phrases and adjectives and, where available, knowledge of the topic of the text [22]. It indicates that the accuracy improves by joining the category feature of semantic orientation. Whitelaw presented a method for sentiment classification based on extracting and analyzing appraisal groups which is represented as a set of attributes values in several task-independent semantic taxonomies [23]. They used semi-automated methods to build a lexicon of appraisal adjectives and their modifiers and classified movie reviews using features based on these taxonomies combined with standard bag-of-words features, and reported the accuracy of 90.2%.

Previous experiments show the method of SVM has the best effectiveness [4]. This study therefore makes an initial attempt to apply supervised machine learning algorithms of SVM to the realistic online reviews of some hotels.

3. Support vector machines

A classification task usually involves with training and testing data which consist of some data instances. Each instance in the training set contains one target value or class labels and several attributes/features. The goal of support vector machine (SVM) is to produce a model which predicts target value of data instances in the testing set which are given only the attributes.

SVM has been shown to be highly effective at traditional text categorization, generally outperforming Naive Bayes (Joachims, 1998)[24]. They are large-margin, rather than probabilistic, classifiers, in contrast to Naive Bayes and MaxEnt. In the two-category case, the basic idea behind the training procedure is to find a hyperplane, represented by vector $\vec{w}$, that not only separates the document vectors in one class from those in the other, but for which the separation, or margin, is as large as possible (Fig. 1). This search corresponds to a constrained optimization problem; letting $y_j \in \{1,-1\}$ (corresponding to positive and negative) be the correct class of document d_j, the solution can be written as

$$\vec{w} = \sum_j \alpha_j y_j \vec{d_j}, \quad \alpha_j \geq 0, \quad (1)$$

where the α_j are obtained by solving a dual optimization problem. Eq.(1) shows that the resulting weight vector of the hyperplane is constructed as a linear combination of $\vec{d_j}$. Only those examples that contribute to which the coefficient α_j is greater than zero. Those vectors are called support vectors, since they are the only document vectors contributing to $\vec{w}$. We used Joachim's SVMlight package[1] for training and testing, with all parameters set to their default values, after first length-normalizing the document vectors, as is standard (neglecting to normalize generally hurt performance slightly). As well as Radial basis function is selected as kernel function in our experiment because it has better generality.

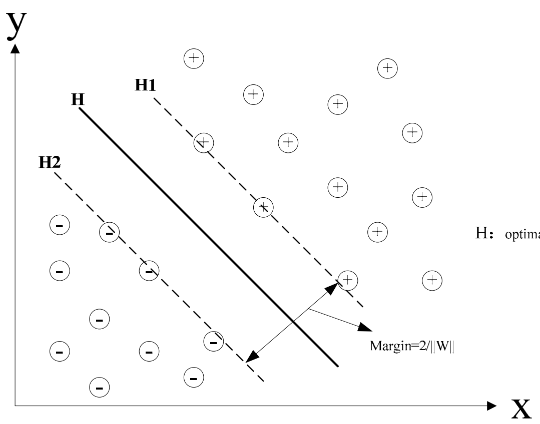

Figure 1. The theory model of SVM

4. Experiment

The basic mechanism of sentiment classification by supervised machine learning algorithms is depicted in Fig. 1. In this research, we applied SVM-based supervised machine learning models for sentiment classification of online hotel reviews.

[1] http://svmlight.joachims.org

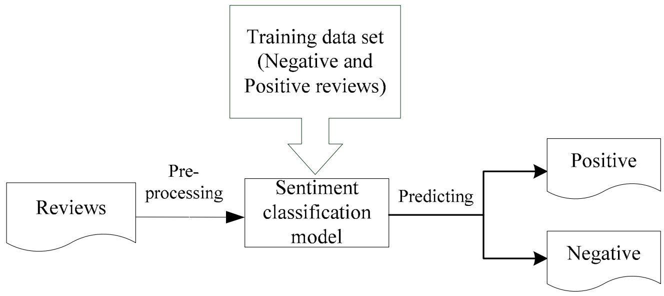

Figure 2. The flow chart of supervised machine learning algorithms

4.1. Experimental Set-up

We used documents from the hotel-review corpus presented by Tan S.H. (http://www.searchforum.org.cn/tansongbo/corpus-senti.htm), which have reviews of 4000 (half positive and the other half negative, respectively). We then divided this data into four equal-sized folds, maintaining balanced class distributions in each fold. All results reported below are the average four-fold cross-validation results on this data (of course, the baseline algorithms had no parameters to tune). Each document is first segmented using ICTCLAS developed by Chinese Academy of Sciences (http://www.nlp.org.cn/project/project.php?proj_id=6).

For this study, we focused on features based on unigrams (segmented Chinese word). There are 12745 unigrams appearing at least one time in our 4000-document corpus.

4.2. Performance evaluations

To evaluate the performance of sentiment classification, we adopted three indexes that are generally used in text categorization: Recall, Precision, and F-score. The indexes can be calculated according to the figures in Table 1 and the following formulas, respectively,

Recall(pos)=A/(A+C), Precision(pos)=A/(A+B),
Recall(neg)=D/(B+D), Precision(neg)=D/(C+D)
*F = 2*Recall*Precision/(Recall + Precision)*

TABLE.1 PERFORMANCE EVALUATIONS

	Actual positive reviews	**Actual negative reviews**
Predict positive	A	B
Predict negative	C	D

Here, Recall(pos) and Precision(pos) are the recall ratio and precision ratio for actual positive reviews. Recall(neg) and Precision(neg) are the recall ratio and precision ratio for actual negative reviews. F-score is the overall evaluation of certain sentiment classification models.

4.3. Experiment result

Experiment result is shown in table 2. Unigrams are the segmentation sequence based on hotel reviews.

TABLE. 2 EXPERIMENT RESULT

	Features	Frequency or TF-IDF?	Recall	Precision	F-score
(1)	unigram	Frequency	88.4%	84.5%	86.4%
(2)	unigram	TF-IDF	89.2%	85.2%	87.2%

In one experiment, we select TF-IDF as one feature.

The TF-IDF feature (term frequency–inverse document frequency) is a weight often used in information retrieval and text mining. This weight is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. The importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the corpus. This result shows the information of TF-IDF is more effective than frequency.

5. Conclusion

In this paper, we first analyze previous research on sentiment classification, and then propose a supervised machine learning approach using unigram feature with two types of information (frequency and TF-IDF) to realize polarity classification of documents. As shown in our experimental results, the information of TF-IDF is more effective than frequency.

In our future work, we will explore semi-supervised machine learning to increase training data set, thus improve the effectiveness of experiment. As well as we will introduce some natural language processing (NLP) techniques to further improve the performance of sentiment analysis.

Acknowledgements

This paper is supported by Project No.Y1090688 and Z1110551 under Zhejiang Provincial Natural Science Foundation of China.

References

[1] Subramanian, S.K. and N. Ramaraj, 2007. Automated classification of customer emails via association rule mining. Inform. Technol. J., 6: 567-572.

[2] Umer, M.F. and M.S.H. Khiyal, 2007. Classification of textual documents using learning vector quantization. Inform. Technol. J., 6: 154-159.

[3] Turney, P.D., 2002. Thumbs up or thumbs down? Sentiment orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, July 11-12, Philadelphia, USA, pp: 417-424.

[4] Pang, B., L. Lee and S. Vaithyanathan, 2002. Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the Conference on Empirical Methods on Natural Language Processing, , July 6-7, Association for Computational Linguistics Press, Philadelphia, USA pp: 79-86.

[5] Dave, K., S. Lawrence and D.M. Pennock, 2003. Mining the peanut gallery:opinion extraction and semantic classification of product reviews. In: Proceedings of the WWW, May 20-24, Budapest, Hungary, ACM Press, pp: 519-528.

[6] Hu, M. and B. Liu, 2004a. Mining and summarizing customer reviews. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data. Aug. 22-25, ACM Press, Washington, USA pp: 168-177.

[7] Hu, M. and B. Liu, 2004b. Mining opinion features in customer reviews. In: Proceedings of the 19th National Conference on American Association for Artificial Intelligence, July 25-29, AAAI Press, San Jose, California, pp: 755-760.

[8] Carenini, G., R. Ng and A. Pauls, 2006. Multi-document summarization of evaluative text. In: Proceedings of the11th EACL 2006, Sydney, Australia, July 22-23, Association for Computational Linguistics Press, pp: 305-312.

[9] Liu, J.J., Y.B. Cao, C.Y. Lin, Y.L. Huang and M. Zhou, 2007. Low-quality product review detection in opinion summarization. In: Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, June 28-30, Association for Computational Linguistics Press, Prague, Czech Republic pp: 343-342.

[10] Baccianella, S., A. Esuli and F. Sebastiani, 2009. Multi-facet Rating of Product Reviews. In: ECIR'2009.

Springer Press, Toulouse, France, April 6-9, Springer-Verlag Press, pp: 461-472.
[11] Liu, J.J. and S. Seneff, 2009. Review Sentiment Scoring via a Parse-and-Paraphrase Paradigm. In: EMNLP'2009, Singapore, August 6-7, Association for Computational Linguistics Press, pp: 161-169.
[12] Narayanan, R., B. Liu, and A. Choudhary, 2009. Sentiment Analysis of Conditional Sentences. In: EMNLP'2009, Singapore, August 6-7, Association for Computational Linguistics Press, pp: 180-189.
[13] Ku, L. W., T.H. Huang and H.H. Chen, 2009. Using Morphological and Syntactic Structures for Chinese Opinion Analysis. In EMNLP'2009, Singapore, August 6-7, Association for Computational Linguistics Press, pp: 1260-1269.
[14] Kim, S.M. and E. Hovy, 2006. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text, In Proceedings of the Workshop on Sentiment and Subjectivity in Text at the joint COLING-ACL'06, Sydney, Australia, Association for Computational Linguistics Press, pp: 1-8.
[15] Danescu-Niculescu-Mizil, C., G. Kossinets, J. Kleinberg and L. Lee, 2009. How Opinions are Received by Online Communities: A Case Study on Amazon.com Helpfulness Votes. In: WWW'2009, Madrid, Spain, April 20-24, ACM Press, pp: 141-150.
[16] Okanohara, D., and Tsujii, J, 2005. Assigning polarity scores to reviews using machine learning techniques. Lecture Notes in Computer Science (LNAI), 3651, pp: 314-325.
[17] Beineke, P., Hastie, T., and Vaithyanathan, S., 2004. The sentimental factor: Improving review classification via human-provided information. ACL'2004, pp: 263-270.
[18] Nasukawa, T. and J. Yi, 2003. Sentiment analysis: Capturing favorability using natural language processing. In: Proceedings of the 2nd International Conference on Knowledge Capture, Whistler, BC, Oct. 28-31, ACM Press, pp: 70 - 77.
[19] Tsou, B.K.Y, R.W.M Yuen, O.Y. Kwong, T.B.Y Lai and W.L. Wong, 2005. Polarity Classification of Celebrity Coverage in the Chinese Press. In: Proceeding of the 2005 International Conference on Intelligence Analysis, McLean VA, May 2-6.
[20] Pang, B. and L. Lee, 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In: ACL'2005, Ann Arbor, Michigan, June 25-30, Association for Computational Linguistics Press, pp: 115 -124.
[21] Chaovalit, P. and L. Zhou, 2005. Movie review mining: A comparison between supervised and unsupervised classification approaches. In: Proceedings of the 38th Hawaii International Conference on System Sciences. Hawaii.
[22] Mullen, T. and N. Collier, 2004. Sentiment analysis using support vector machines with diverse information sources. In EMNLP'2004, Barcelona, Spain, July 25-26, Association for Computational Linguistics Press, PP: 412-418.
[23] Whitelaw, C., N. Garg and S. Argamon, 2005. Using appraisal groups for sentiment analysis. In: CIKM'2005, Bremen, Germany, Oct. 31 – Nov. 5, ACM Press, pp: 625-631.
[24] Joachims, T., 1998. Text categorization with support vector machines: Learning with many relevant features. Proceedings of 10th European Conference on Machine Learning (ECML-98), Chemnitz, Germany, pp: 137-142. Springer.