MULTI-ASPECT SENTIMENT SCRUTINY SYSTEM BY MEANS OF SUPERVISED LEARNING PRACTICES

R.MEENA GOMATHI
PG Scholar, Computer Science and Engineering Department,
K.L.N. College of Engineering, Sivagangai- 630612 INDIA

---***---

ABSTRACT-*Sentiment scrutiny is an application of text analytic methods for the identification of subjective opinions in text data. It usually comprises the classification of text into sorts such as positive, negative. There are even circumstances where diverse forms of a single word will be related with unalike sentiments. We propose a system to analyze sentiments across different domains by creating sentiment sensitive thesaurus using datum from multiple source domains to find the association between words that express similar sentiments in different domains. The created thesaurus is then used to expand feature vectors to train a binary classifier. Unlike previous cross-domain sentiment classification methods, our method can competently learn from numerous source domains. Our method significantly outperforms numerous baselines and returns results that are better than previous cross-domain sentiment classification methods when equated on a benchmark dataset containing user reviews for different types of products.*

Keywords*:* Cross-Domain Sentiment Classification, Binary Classifier, Sentiment Sensitive Thesaurus

1. INTRODUCTION

Sentiment analysis, also called opinion mining, [1] is the field of study that analyses people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. When an organization wants to find opinions of the general public about its products and services, it conducts surveys. However, due to the tremendous growth of the social media content in the past few years, the people post their reviews of products at merchant sites and express their views on almost anything in discussion forums and blogs, and at social networks.

Sentiment Analysis refers to computational methods for investigating the opinions that are extracted from various sources like the blog posts, observations on forums, reviews about products, policies or any topic on social networking sites or tweets. It aims at determining the attitude of a user about some topic. The Web is a huge repository of structured and unstructured data. The analysis of this data to extract underlying user's opinion and sentiment is a challenging task. An opinion can be described as a quadruple consisting of a Topic, Holder, Claim and Sentiment. The holder trusts a claim about the topic and expresses it through an associated sentiment.

In opinion mining, [2] one fundamental problem is to extract opinion targets and opinion words, which are defined as the objects on which users have conveyed their opinions as nouns or noun phrases. This task is very important because customers are not satisfied with the overall sentiment polarity of a product, but they expect to find the fine-grained opinions about an aspect or a product feature mentioned in reviews. Mining opinions from online reviews become more useful for manufacturers to obtain feedbacks from customers. Hence, mining opinion relations in sentences and estimating associations between opinion targets and opinion words are keys for opinion target extraction.

There is a large expansion of e-commerce, more and more products are sold on the Web, and more and more people are also buying products online. In order to enhance customer satisfaction and shopping experience, it has become a common practice for online merchants to enable their customers to review or to express opinions on the products that they have purchased. With more and more common users becoming more comfortable with the Web, an increasing number of people are writing reviews.

With the growth of online social networking sites, for example, forums, review sites, blogs, and micro blogs, the enthusiasm towards opinion mining has expanded essentially. Today online opinions have transformed into a sort of virtual profit for business organizations looking to market their items, recognize new trends and deal with their position. Many organizations are currently utilizing opinion mining systems to track customer inputs in online shopping sites and review sites. Opinion mining is additionally helpful for organizations to analyze customer opinions on their products and features. If one wants to buy a product, he/she is no longer limited to ask one's friends and families because there are many user reviews on the web. For a company, it may no longer need to conduct surveys or focus groups in order to gather consumer opinions about its products and those of its competitors because there is a plenty of such information publicly.

Let us use the following review segment on iPhone as an instance to introduce the general problem (a number is associated with each sentence for easy reference) "(i) I bought an iPhone 2 days ago. (ii) It was such a nice phone. (iii) The touch screen was really cool. (iv) The voice quality was clear too. (v) However, my mother was mad with me as I did not tell her before I bought it. (vi) She also thought the phone was too expensive, and wanted me to return it to the shop". The question is, what we want to extract from this review? The first thing that we notice is that, there are several opinions in this review. Sentences (ii), (iii) and (iv) express three positive opinions, while sentences (v) and (vi) express negative opinions. We also notice that all the opinions have some targets on which they are expressed. The opinion in sentence (ii) is on iPhone as a whole, and the opinions in sentences (iii) and (iv) are on the "touch screen" and "voice quality" features of iPhone respectively. The opinion in sentence (vi) is on the price of iPhone, but the opinion/emotion in sentence (v) is on "me", not iPhone. This is a vital outlook. In an application, the user may be interested in opinions on certain targets, but not on all. The source of the opinions in sentences (ii), (iii) and (iv) is the author of the review ("I"), but in sentences (v) and (vi) it is "my mother". With this example in mind, we can define sentiment analysis or opinion mining. We begin with the opinion target.

2. PROBLEM DEFINITION

Applying sentimental classifier for a particular domain gives better performance, but when it is being applied for multiple domains there is lack of performance and also it leads to feature mismatch problem

3. EXISTING SYSTEM

A cross domain sentiment classification system must overcome two main challenges [8]. First, it is necessary to identify the source domain features are related to which target domain features. Second, the information should have the relatedness of both the source and target domain features. In the existing system there is a feature mismatch problem. To overcome that problem in cross-domain sentiment classification, we use labeled data from multiple source domains and unlabeled data from source and target domains to compute the relatedness of features and construct a sentiment sensitive thesaurus. This competently offers higher level of accuracy.

4. PROPOSED SYSTEM

We create a sentiment sensitive thesaurus using both labeled and unlabeled data from multiple source domains to find the association between words that express similar sentiments in different domains. The created thesaurus is then used to expand feature vectors to train a binary classifier. The sentiment sensitive thesaurus is constructed in order to overcome the shortcomings of classification using a single domain. The sentiment sensitive thesaurus captures relatedness of words as used in different domains. In addition to the existing positive, negative and neutral opinions, we go for classifying the mixed opinion.

5. ARCHITECTURAL COMPONENTS

- A sentiment sensitive thesaurus is created in order that it aligns different words expressing the same sentiment in different domains.
- Labeled data from multiple source domains and unlabeled data from source and target domains are utilized to represent the distribution of features. [14]
- The lexical elements (unigrams and bigrams of word lemma) and Sentiment elements (rating information) are used to represent a user review.
- Next, for each lexical element its relatedness is measured to other lexical elements and group related lexical elements to create a sentiment sensitive thesaurus.[9]
- The thesaurus captures the relatedness among lexical elements that appear in source and target domains based on the contexts in which the lexical elements appear (its distributional context).
- A distinctive aspect of our approach is that, in addition to the usual co-occurrence features typically used in exemplifying a word's distributional context, we make use, where possible, of the sentiment label of a document i.e., sentiment labels form part of our context features. This is what makes the distributional thesaurus "sentiment-sensitive".
- Unlabeled data is cheaper to collect compared to labeled data and is often available in large quantities.
- The use of unlabeled data enables us to accurately estimate the distribution of words in source and target domains.
- The proposed method can learn from a large amount of unlabeled data to leverage a robust cross domain sentiment classifier.[15]

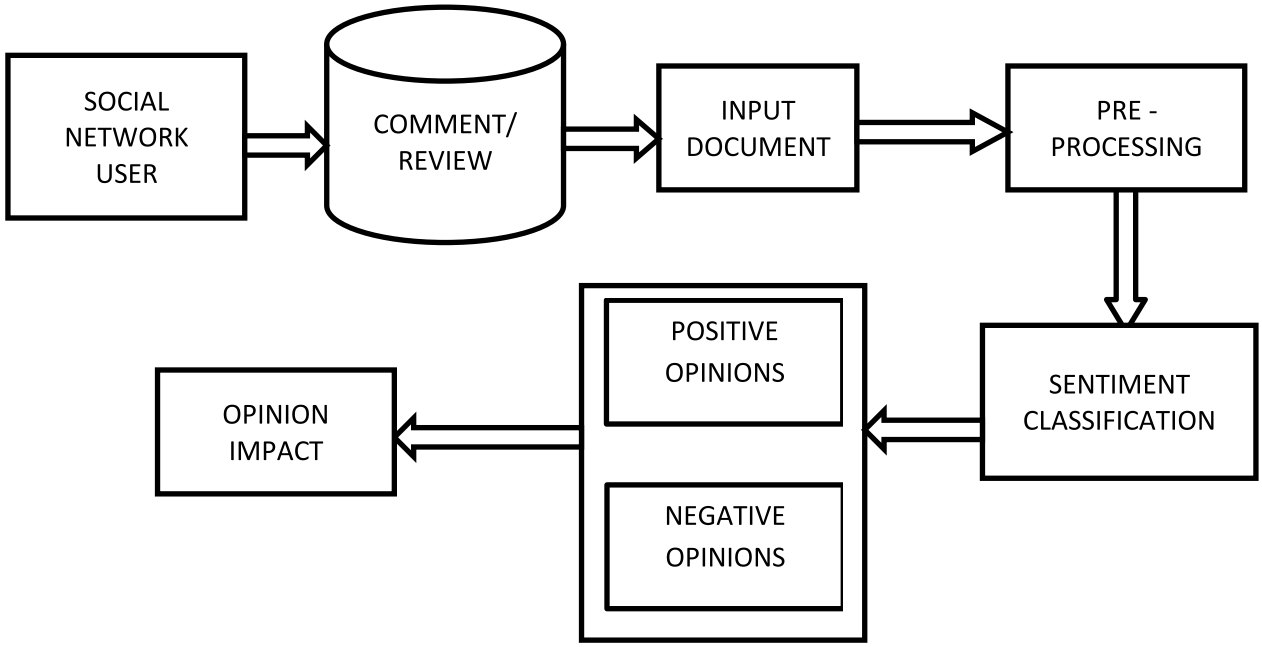

Figure 1. Flow diagram of sentiment analysis

6. LIST OF MODULES

i. Pre-processing
ii. Building sentimental sensitive thesaurus
iii. Classifications
iv. Identifications

7. Module Description:

7.1. Pre-processing

Data in the real world is incomplete, noisy, and inconsistent. Noisy data (incorrect values) may come from faulty data collection instruments, human or computer error at data entry and errors in data transmission. So here we pre-process datasets before classifying document sentiment.

7.2. Building Sentiment Sensitive Thesaurus:

We built a sentiment sensitive thesaurus to find the prior information. This thesaurus contains number of sentimental words related to opinion mining. We extracted the words with strong positive and negative orientation and performed stemming in the pre-processing. The words whose polarity changed after stemming were removed automatically. The thesaurus used here is fully domain-independent and bear supervised information.

7.3. Classification and Identification

This module is used to classify the polarity of a given text in the document, whether the expressed opinion in a document, is positive, negative, neutral or mixed. [12] On Identification, we define that a document 'd' is classified as a positive-sentiment document if its probability of positive sentiment label is greater than its probability of negative sentiment label in the given document. The document sentiment is classified based on P(l|d), the probability of a sentiment label given document. First classify positive and negative opinions.
The prior information incorporated merely contributes to the positive and negative words, and consequently there will be much more influence on the probability distribution of positive and negative labels for a given document. Therefore, we define that a document d is classified as a positive-sentiment document if the probability of a positive sentiment label $P(l_{pos}|d)$ is greater than its probability of negative sentiment label $P(l_{neg}|d)$, and vice versa.

8. CONCLUSION AND FUTURE WORK

This paper proposes a novel scheme for co-extracting opinion targets and opinion words. The foremost influence is motivated on perceiving opinion relations between opinion targets and opinion words.

While equated to preceding means grounded on nearest neighbour rules and syntactic patterns, a sentiment-sensitive thesaurus bridges the gap amid source and target domains in cross-domain sentiment classification via manifold source domains. Experimental outcomes using a dataset for cross–domain sentiment classification illustrates that our proposed method can improve classification accuracy and therefore is more effective for opinion target and opinion word extraction in a sentiment classifier. In future, the proposed method is intended to apply to other domain adaptation tasks.

9. REFERENCES

[1].B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs Up? Sentiment Classification Using Machine Learning Techniques,” Proc. ACL-02 Conf. Empirical Methods in Natural Language Processing (EMNLP ’02) pp. 79-86, 2002.

[2]. P.D. Turney, “Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews,” Proc. 40th Ann. Meeting on Assoc. for Computational Linguistics (ACL’02),pp. 417-424, 2002.

[3]. Y. Lu, C. Zhai, and N. Sundaresan, “Rated Aspect Summarization of Short Comments,” Proc. 18th Int’l Conf. World Wide Web (WWW ’09), pp. 131-140, 2009.

[4].T.-K. Fan and C.-H. Chang, “Sentiment-Oriented Contextual Advertising,” Knowledge and Information Systems, vol. 23, no. 3, pp. 321-344, 2010.

[5].M. Hu and B. Liu, “Mining and Summarizing Customer Reviews,” Proc. 10th ACM SIGKDD International Conference Knowledge Discovery and Data Mining (KDD ’04), pp. 168-177, 2004.

[6]. J. Blitzer, M. Dredze, and F. Pereira, “Biographies, Bollywood, Boom-Boxes and Blenders: Domain Adaptation for Sentiment Classification,” Proc. 45th Ann. Meeting of the Assoc. Computational Linguistics (ACL ’07), pp. 440-447, 2007.

[7]. S.J. Pan, X. Ni, J.-T. Sun, Q. Yang, and Z. Chen, “Cross-Domain Sentiment Classification via Spectral Feature Alignment,” Proc. 19th Int’l Conf. World Wide Web (WWW ’10), 2010.

[8]. H. Fang, “A Re-Examination of Query Expansion using Lexical Resources,” Proc. Ann. Meeting of the Assoc. Computational Linguistics (ACL ’08), pp. 139-147, 2008.

[9].G. Salton and C. Buckley, “Introduction to Modern Information Retrieval,” McGraw-Hill Book Company.

[10] Kang Liu, Liheng Xu, and Jun Zhao, “Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model”, in IEEE Transactions On Knowledge And Data Engineering, vol. 27, NO. 3, March 2015, P.636-650.

[11] Zheng-Jun Zha, Jianxing Yu, Jinhui Tang, Meng Wang, Tat-Seng Chua, “Product Aspect Ranking and its Applications” in IEEE transactions on knowledge and data engineering, vol.26,No.5, May 2014,pp.1211-1224.

[12] Lisette Garcia-Moya, Henry Anaya-Sanchez, Rafael Berlanga-Llavori, "Retrieving Product Features and Opinions from Customer Reviews", IEEE Intelligent Systems, vol.28, no. 3, pp. 19-27, May-June 2013.

[13] MagdaliniEirinaki, S. Pisal, and J. Singh. "Feature-based Opinion Mining and Ranking" Journal of Computer and System Sciences 2012, pp.1175-1184.

[14] XU Xueke, CHENG Xueqi, TAN Songbo, LIU Yue, SHEN Huawei, ”Aspect-Level Opinion Mining of Online Customer Reviews” in Journal IEEE, China Communications -March 2013,pp.25-41.

[15] Zhen Hai, Kuiyu Chang, Jung-Jae Kim, and Christopher C. Yang, “Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance” in IEEE transactions on knowledge and data engineering, volume. 26, No. 3, May 2014, pp.623-634.