ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

Review on Efficient Sentiment Analysis with Fine-grained Opinion Mining

Ankita Kale, Prof. Jayshree Aher

Student, Department of Computer Engineering, MITCOE, Savitribai Phule Pune University, Pune, India

Professor, Department of Computer Engineering, MITCOE, Savitribai Phule Pune University, Pune, India

ABSTRACT: Sentiment Analysis or Opinion Mining is an imperative research area of Natural Language Processing. It is the process of identifying and extracting subjective information in source materials. Basic task in sentiment analysis is classifying the polarity by categorizing a piece of text into positive, negative or neutral. The present survey gives an idea of the work done by several researchers on Sentiment analysis.

KEYWORDS: Opinion mining, Sentiment analysis, Product aspect.

I. INTRODUCTION

One of the emerging aspect of web is e-commerce has been on the improvement since the beginning of 21st century. Nowadays more items are sold on the web and that's why the customer database is expanding. The online shippers are requesting to their customers to provide their important feedback at the time of delivery of some item, in order to upgrade their future services and company policies. As the huge amount of people makes online purchasing, the survey of each item is growing and thereby it leads to processing of huge amount of data. The Service providers may face difficulty because of the amount of inputs coming through from the users. So this feedback system is an important part of any Service Providing Company. Specifically online reviews on websites greatly influence shopping decisions of customer. Sometimes customers may provide a negative feedback just by going through one to two previous feedbacks.

In recent days sentiment analysis becomes an innovative, challenging at the same time vast research area. In this area, main task is not only to identify when sentiment is expressed but also identifying attributes of the sentiment. Sentiments attribute includes who is expressing the sentiment, about what or whom the sentiment is being expressed along with the polarity and intensity or rating of the sentiment.

Most of the time sentiments plays vital part in the decision-making process. Many of us ask our friends to recommend a digital camera or to share experience of travel on a tourist destination, request reference letters regarding job applicants from colleagues, or consult experts' reviews in magazines to decide which computer to buy. In this way, sentiments are derived from experiences of a number of people. At the same time more people share their own opinions with world through internet. As more people provides the sentiments and experiences which are easily available online, analyzing them has growing importance in our daily life.

The deep understanding of the meaning of text is the most important and critical element of the automated processing. The modern technologies and methods of big data, with artificial intelligence, has already been helping researchers to automate the process of content analysis. Machine translation interprets the data in same way as human translation. These innovations give the opportunity to conduct large-scale research and to monitor social media in real-time.

In Section II, we give the Literature review of previous work on different sentiment analysis approaches with methodologies and datasets used. And Section III consists of conclusion part.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

II. RELATED WORK

Z. J. Zha et. al. have proposed a product aspect ranking framework to identify the important aspects of products from multiple number of consumer reviews. The approach consist of three main components, i.e., identification of product aspect, classification of aspect sentiment, and ranking of aspect. They also have conducted extensive experiments to systematically evaluate the given framework. The experimental corpus which contains 94,560 consumer reviews of 21 popular products in eight domains [1].

K. Liu et al. developed a method for co-extracting opinion targets and opinion words by using a word alignment model. Main contribution in this paper is focused on detecting opinion relations between opinion targets and opinion words. Proposed method captures opinion relations more precisely and therefore is more effective for opinion target and opinion word extraction. After that they construct an Opinion Relation Graph to model all candidates and the detected opinion relations among candidates, with a graph co-ranking algorithm to estimate the confidence of each candidate [2].

R. Xia et. al. proposed a data expansion approach, known dual sentiment analysis (DSA), to address the polarity shift problem in sentiment classification. The basic idea of DSA is to create reversed reviews that are sentiment-opposite to the original reviews, and uses the original reviews and reversed reviews in couple to train a sentiment classifier and make predictions. DSA is focused by the technique of one to one correspondence data expansion and the manner of using a pair of samples in training or dual training and prediction or dual prediction. A wide range of experiments demonstrate that the DSA model is very useful for classification of polarity and it significantly outperforms several alternative methods of considering polarity shift [3].

Evert, Stefan, et al stated that TKE is a versatile tool for the identification of topics in medium to large sets of small text documents and the general sentiments expressed towards each topic. From a user point of view, the strength available in its quick and cost efficient way of analysis. The system copes easily with amounts of material 725 that render manual processing virtually impossible. With the increasing popularity of online surveys (and other applications such as trend mining in the Web or social media networks) data sets of this size are no longer uncommon. Authors also stated that TKE provides a potent instrument for the analysis of textual data in market research and depicts promising potential for a range of same applications [4].

Khan, Farhan Hassan, Usman Qamar, and Saba Bashir makes use of SentiWordNet and treats it as the labeled corpus for training process. A dictionary of sentiments, SentiMI, builds upon the mutual information calculated from these terms. Also stated that a complete framework is developed by using feature selection and extracting mutual information for the selected features. Training, evaluation and testing of the proposed framework are conducted on a big dataset which contain 50,000 reviews of movie. A good performance improvement of 7% in accuracy, 14% in specificity, and 8% F-measure is earn by the proposed framework as compared to the baseline SentiWordNet classifier [5].

Gerani, Shima et. al. propose and investigate three alternative content selection and structuring models for the automatic construction of an AHT in our summarization framework: 1) Rhetorical model, which captures the aspects' importance and relationship by looking at the way people discuss and relate the aspects when expressing opinion in their reviews. 2) Conceptual model, which exploits a common-sense knowledge base (e.g. ConceptNet) to find the conceptual association between aspects. 3) Hybrid model, which exploits both the rhetorical and conceptual information. Designed abstractive summarization framework has the potential to implement one of the proposed models depending on the application or apply all three models and let a user choose the output, depending on his/her desire to use the conceptual, rhetorical or both sources of information [6].

Khan, Farhan Hassan, Usman Qamar, and Saba Bashir concentrated on improve SentiWordNet performance and designed a complete sentiment analysis and classification approach is based on SentiWordNet vocabulary. Seven publically available datasets are used for performance evaluation and for comparative analysis. A performance

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

improvement of 13.4% in accuracy, 14.2% in precision, 6.9% in recall and 11.1% f-measure is noted by evaluating the proposed eSAP against the baseline SWN classifier [7].

Li, Fangtao, et al. proposed a two-stage framework for co-extraction of sentiment and topic lexicons across domains where we have no labeled data in the target domain but have plenty of labeled data in another domain. In the first stage, they have proposed a simple strategy to generate a few high-quality sentiment and topic seeds for the target domain. In the second stage, they propose a novel Relational Adaptive bootstraPping (RAP) method to expand the seeds, which can exploit the relationships between topic and opinion words, and make use of part of useful source domain labeled data for help [8].

Liu, Kang, Liheng Xu, and Jun Zhao have praposed approach to extract opinion targets based on wordbased translation model (WTM). Authors apply WTM in a monolingual scenario to mine the associations between opinion targets and opinion words. Then, a graphbased algorithm is exploited to extract opinion targets, where candidate opinion relevance estimated from the mined associations, is incorporated with candidate importance to generate a global measure. By using WTM, developed method can capture opinion relations more precisely, especially for long-span relations [9].

Li, Fangtao, et al. formulated the review mining task as a joint structure tagging problem. Also a new framework based on Conditional Random Fields is proposed. The framework can employ rich features to simultaneously extract object features, positive opinions and negative opinions. With the proposed framework, authors investigate the chain structure, conjunction structure and syntactic tree structure for review mining. A new unified model, called skip tree CRFs, is proposed for review mining [10].

From above discussion it is observed that sentiment analysis takes different learning approach such as Supervised, Unsupervised or Semi-Supervised. And can work at different levels as Word-level, Sentence-level, Aspect-level or Document-level. Task of object aspect identification is performed using frequency based models or word based model at corpus level. On the contrary polarity shift or summarization tasks are performed at sentence level through clustering. As shown in table 1, literature review of various papers has been listed, giving possibility of research gap.

Table 1: Survey Table

Sr no.	Title	Publication	Techniques	Advantages	Research gap	Future scope
1.	Product Aspect Ranking and Its Applications	IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, MAY 2014. Zheng - Jun Zha,Jianxing Yu,Meng Wang Member, Tat-Seng Chua	1.product aspect ranking framework 2.Probabilistic Aspect Ranking Algorithm	Identifying product aspects considering aspect frequency and influence of consumer opinion determined by sentiment classifier. Facilitates real word applications i.e. Document level sentiment classification	Averaging aspect weights of reviews which can Lead to errors than the overall aspect weights are identified Directly from consumer reviews	Use both aspect ratings and overall rating of each review To learn the overall aspect weights
2.	Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model	IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, MAR 2015. Kang Liu, Liheng Xu, and Jun Zhao	1.Partially Supervised Word Alignment Model 2.Graph Based Co-ranking Algorithm	Mining opinion targets and opinion words to detect opinion relation using partially-supervised alignment model. A graph based co ranking algorithm is exploited to estimate the confidence of each candidate.	Opinion relation between words can also be used to identify topic relations through Opinion Relation Graph. Also performance of the system can be improved.	Understand topic relations between words from Opinion Relation Graph
3.	Dual Sentiment Analysis: Considering Two Sides of One	IEEE TRANSACTIONS ON KNOWLEDGE AND DATA	1.Dual Sentiment Analysis Model 2.Dual Training Algorithm	Model handling polarity shift problem in sentiment classification by dual learning and	DSA algorithm is specific to the application. Doesn't consider complex	Build more generalized DAS algorithm for other tasks.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

	Review	ENGINEERING, AUG 2015 Zheng - Jun Zha, Jianxing Yu, Meng Wang Member, Tat-Seng Chua	3.Dual Prediction Algorithm	prediction algorithm considering both side of one review.	polarity shift pattern in creating reversed reviews.	Reversed reviews generation can be made more accurate.
4.	A distributional approach to open questions in market research	ELSEVIER Computers in Industry, OCT 2015. Stefan Evert, Paul Greiner, Joao Filipe Baigger, Bastian Lang	1. The Klugator Engine (TKE) semi-automatic System	Transforms textual input into structured corpus by automatic detection of polarity scores. Label each topic cluster with set of silent keywords, evaluating sentiment associated with the topic.	Response time of system is very high i.e.20 min	Response time can be improved with significant improvements in the system still lesser than 20min
5.	Sentimi: Introducing point-wise mutual information with Sentiwordnet to improve sentiment polarity detection	ELSEVIER Applied Soft Computing, NOV 2015. Farhan Hassan Khan,Usman Qamar,Saba Bashir	1. A sentiment dictionary, sentimi based on sentiwordnet (SWN)	Synset along with sentiment scores & part of speech information gives mutual information.Stanford POS tagger is used.Increases accuracy of document classification	Proposed algorithm shows improvement of 7% in accuracy, 14% in specificity, and 8% in F-measure only	All these factors can be significantly improved.
6.	Modeling content and structure for abstractive review Summarization	ELSEVIER Computer, Speech and Language, JUN 2016. Shima Gerani, Giuseppe Carenini,Raymond T. Ng	1. Aspect Hierarchy Tree 2. Rhetorical model 3. Conceptual model 4. Hybrid model	Aspect base framework generates abstract from multiple reviews without training data. Three alternative methods for expressing opinions in the reviews are described for Aspect Hierarchy Tree.	Unable to differentiate relation types.	Improvement In content selection, summary generation component and also integrating extractive and abstractive approaches.
7.	Esap: A Decision Support Framework for Enhanced Sentiment Analysis and Polarity Classification	ELSEVIER Information Sciences, JUL 2016 Farhan Hassan Khan, Usman Qamar, Saba Bashir	1. Enhanced Sentiment Analysis and Polarity Classification Framework 2.SWN-V vocabulary	A revised Sentiment score vocabulary based on sentiwordnet Uses SVM for model learning and classification .Improved accuracy for precision, recall and F-measure.	Performance is still less and can be improved.	Improving the performance through other approaches like cosine similarity and information gain can be implemented.

III. CONCLUSION

This paper analyses various techniques used for Sentiment Analysis. This is a vast research area with multiple dimensions to work on providing numerous research opportunities. Sentiment Analysis systems modeling product aspects along with their features can be significantly improved by considering individual as well as overall rating for the aspect. Association between product and its feature is easy task next challenge is topic relations. Handling complex polarity shift in the reviews is half solved problem and need more improvements. All in all Improvement at response time and accuracy of Existing Sentiment Analysis system is the demand of today's fast growing world.

REFERENCES

1. Z. J. Zha, J. Yu, J. Tang, M. Wang and T. S. Chua, "Product Aspect Ranking and Its Applications," in IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 5, pp. 1211-1224, May 2014.
2. K. Liu, L. Xu and J. 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, pp. 636-650, March 1 2015.
3. R. Xia, F. Xu, C. Zong, Q. Li, Y. Qi and T. Li, "Dual Sentiment Analysis: Considering Two Sides of One Review," in IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 8, pp. 2120-2133, Aug. 1 2015.
4. Evert, Stefan, et al. "A distributional approach to open questions in market research." Computers in Industry 78 (2016): 16-28.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

5. Khan, Farhan Hassan, Usman Qamar, and Saba Bashir. "SentiMI: Introducing point-wise mutual information with SentiWordNet to improve sentiment polarity detection." Applied Soft Computing 39 (2016): 140-153.
6. Gerani, Shima, Giuseppe Carenini, and Raymond T. Ng. "Modeling content and structure for abstractive review summarization." Computer Speech & Language (2016).
7. Khan, Farhan Hassan, Usman Qamar, and Saba Bashir. "eSAP: A decision support framework for enhanced sentiment analysis and polarity classification." Information Sciences 367 (2016): 862-873.
8. Li, Fangtao, et al. "Cross-domain co-extraction of sentiment and topic lexicons." Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1. Association for Computational Linguistics, 2012.
9. Liu, Kang, Liheng Xu, and Jun Zhao. "Opinion target extraction using word-based translation model." Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Association for Computational Linguistics, 2012.
10. Li, Fangtao, et al. "Structure-aware review mining and summarization."Proceedings of the 23rd international conference on computational linguistics. Association for Computational Linguistics, 2010