IJCAT - International Journal of Computing and Technology, Volume 4, Issue 4, April 2017
ISSN : 2348 - 6090
www.IJCAT.org

Opinion Analysis Using Polarity Shift Model

[1]Amit Jangde; [2]Pravin Malviya

[1,2] MTech. CSE, Department, SBITM COE, Betul, India

Abstract - The user-generated text volume on the Web in the form of reviews, blogs, and social networks has grown dramatically in recent years. This was mirrored by an increasing interest, from both the academic and the business world, in the field of opinion analysis. The polarity shift problem is a major factor that affects classification performance of machine-learning-based sentiment analysis systems .Here we proposing a three-stage cascade model to address the polarity shift problem in the context of document-level sentiment classification. Firstly splitting of each document into a set of sub sentences and build a hybrid model that employs rules and statistical methods to detect explicit and implicit polarity shifts, respectively is done. Secondly, proposing a polarity shift elimination method, to remove polarity shift in negations. Afterwards, training the base classifiers on training subsets divided by different types of polarity shifts, and uses a weighted combination of the component classifiers for sentiment classification. On this basis, also proposing a dual training algorithm to make use of original and reversed training reviews in duality for learning a sentiment classifier, and a dual prediction algorithm to classify the test results by keeping in mind both of two phases of one result. An extended framework from polarity (positive-negative) classification to 3-class (positive-negative-neutral) classification has to be done, by taking the neutral reviews into consideration. Finally, a corpus-based approach is constructed for pseudo-antonym dictionary, which elimination of Dual Sentential approach's dependency on an external antonym dictionary for review reversion.

Keywords - sentiment analysis, polarity

1. Introduction

In late years, with the increasing of online reviews gettable on the internet , sentiment analysis and opinion mining, as a special text mining task for studying or interpret the subjective attitude (i.e., sentiment) assert by the text, is becoming a difficulty in the field of data mining and natural language processing . Sentiment classification is a primary task in sentiment analysis, with its focus to classify the sentiment (e.g., positive or negative) of a given text. The extensive practice in sentiment classification chases the techniques which are widely used in topic-based text classification, where the bag-of-words (BOW) model as a rule used for text representation. In the BOW model, a review text is represented by a vector of autonomous words The statistical machine learning algorithms (such as naive Bayes, maximum entropy classifier, and support vector machines) are then applied to train a sentiment classifier.
Although the BOW model is very elementary and quite efficient in topic-based text classification, it is actually not very suitable. However, the BOW model disrupts word order, breaks the syntactic structures and discards some semantic information of the text. Therefore, it brings about some fundamental deficiencies including the polarity shift problem. Polarity shift refers to a linguistic phenomenon in which the polarity of sentiment can be reversed [1] (i.e., positive to negative or vice versa) by some special linguistic structures called polarity shifters, e.g., negation ("*I don't like this movie*") and contrast ("*Fairly good, but not my style*"). Obviously, in the BOW model, it is hard to capture the sentiment reversion caused by polarity shifters, because two sentiment-opposite texts (e.g., "*I don't like this movie*" and "*I like this movie*") are regarded to be very similar in the BOW representation [1].

However, most of them focused on either modeling polarity shift in phrase/sub sentence-level sentiment classification, or encoding polarity shift in rule-based term-counting methods. Even there were few of them dealing with polarity shift by using machine learning methods for document-level sentiment classification, their performances were not satisfactory, e.g., the improvements were less than 2% after considering polarity shift[3]. In this work, proposing a three-stage model, namely Polarity Shift Detection, Elimination and Ensemble (PSDEE), to address polarity shift for document-level sentiment classification. Firstly, proposing a hybrid polarity shift detection approach, which employs a rule-based method to detect some polarity shifts such as explicit negations and contrasts, and a statistical method to detect some implicit polarity shifts such as sentiment inconsistencies. Secondly, a novel polarity shift elimination algorithm to eliminate polarity shifts in negations is implemented. For example, the review "*this movie is not interesting*" is reversed to "*this movie is boring*". It can make the BOW representation more feasible due to the elimination of negations.

Finally, by separating the training and test data into four component subsets, i.e., negation subset, contrast subset, sentiment-inconsistency set as well as polarity-unshifted subset, and train the base classifiers based on each of the component subset. A weighted ensemble of four component predictions are finally used in testing, with the motivation to distinguish texts with different types of polarity shifts such that the polarity-unshifted part will have a higher weight, while the polarity-shifted part will have a lower weight in sentiment prediction. A systematical evaluation our PSDEE model by conducting experiments on four sentiment datasets, three kinds of classification algorithms and two types of features.

2. Literature Review

According to the levels of granularity, tasks in sentiment analysis can be divied into four categorizations: document level, sentence-level, phrase-level, and aspect-level sentiment analysis. Focusing on the phrase/subsentence- and aspect-level sentiment analysis, Wilsonetal. [22]discussed effects of complex polarity shift. They began with a lexicon of words with established prior polarities, and identify the "contextual polarity" of phrases, based on some refined annotations. Choi and Cardie [4] further combined different kinds of negators with lexical polarity items though various compositional semantic models,both heuristic andmachine learned,to improved subsentential sentiment analysis. Nakagawa et al. [29] developed a semi-supervised model for subsentential sentiment analysis that predicts polarity based on the interactions between nodes independency graphs,which potentially can induce the scope of negation. In aspect-level sentiment analysis, the polarity shift problem was considered in both corpus- and lexicon based methods[8],[9],[10],[13]. For document- and sentence-level sentiment classification, there are two main types of methods in the literature : term-counting and machine learning methods. In term counting methods, the overall orientation of a text is obtained by summing up the orientation scores of content words in the text, based on manually-collected or external lexical resources [8], [13]. In machine learning methods, sentiment classification is regarded as a statistical classification problem, where a text is represented by a bag-of words; then, the supervised machine learning algorithms are applied as classifier [35]. Accordingly, the way to handle polarity shift also differs in the two types of methods. The term-counting methods can be easily modified to include polarity shift. One common way is to directly reverse the sentiment of polarity-shifted words, and then sum up the sentiment score word by word [4], [16], [17], [34]. Compared with term counting methods, the machine learning methods are more widely discussed in the sentiment classification literatures. However, it is relatively hard to integrate the polarity shift information into the BOW model in such methods. For example, Das and Chen [6] proposed a method by simply attaching "NOT" to words in the scope of negation, so that in the text "I don't like book", the word "like" becomes a new word "like NOT". Yet Pang et al. [35] reported that this method only has slightly negligible effects on improving the sentiment classification accuracy. There were also some attempts to model polarity shift by using more linguistic features or lexicalresources. For example, Na et al. [28] proposed to model negation by looking for specific part-of-speech tag patterns. Kennedy and Ink pen [17] proposed to use syntactic parsing to capture three types of valence shifters (negative, intensifiers, and diminishers).

Their results showed that handling polarity shift improves the performance of term-counting systems significantly, but the improvements upon the baselines of machine learning systems are very slight (less than 1 percent). Ikeda et al. [14] proposed a machine learning method based on a lexical dictionary extracted from General Inquirer1 to model polarity-shifters for both word-wise and sentence-wise sentiment classification. There were still some approaches that addressed polarity shift without complex linguistic analysis and extra annotations. For example, Li and Huang [19] proposed a method first to classify each sentence in a text into a polarity-unshifted part and a polarity-shifted part according to certain rules, then to represent them as two bags-of-words for sentiment classification. Li et al. [21] further proposed a method to separate the shifted and unshifted text based on training a binary detector. Classification models are then trained based on each of the two parts. An ensemble of two component classifiers is used to provide the final polarity of the whole text. Orimaye et al. [34] proposed a sentence polarity shift algorithm to identify consistent sentiment polarity patterns and use only the sentiment-consistent sentences for sentiment classification. A preliminary version of this paper was published in [34]. In this paper, we extend our previous work in three major aspects. First, strengthening of the DSA algorithm by adding a selective data expansion procedure. Second, extending the DSA framework from sentiment polarity classification to positive-negative-neutral sentiment classification. Third, a corpus-based method to construct a pseudo-antonym dictionary that could remove DSA's dependency on an external antonym dictionary is implemented

3. Problem Definition

This paper mainly works on text classification which applies a rule-based method to recognize some polarity shifts such as explicit negations and contrasts, and a statistical method to recognize some implicit polarity shifts such as sentiment difference. A novel polarity shift elimination algorithm to eliminate polarity shifts in negations. A dual training algorithm to make use of original and reversed training reviews in duality for learning a sentiment classifier, and a dual prediction algorithm to classify the test results by keeping in mind both of two phases of one result.

4. Proposed Methodology

The architecture consists of two major components as Textual preprocessing, Sentiment analysis, opinions. Each component performs functions namely, preprocessing the raw dataset(text, document, topic relevant Dataset) from the user or from the datasets and Classifying the datasets by using the opining related resource generator of textual data into lexicons, patterns, training Set. FigureIV-1 presents the architecture of the proposed system.

4.1 System Architecture

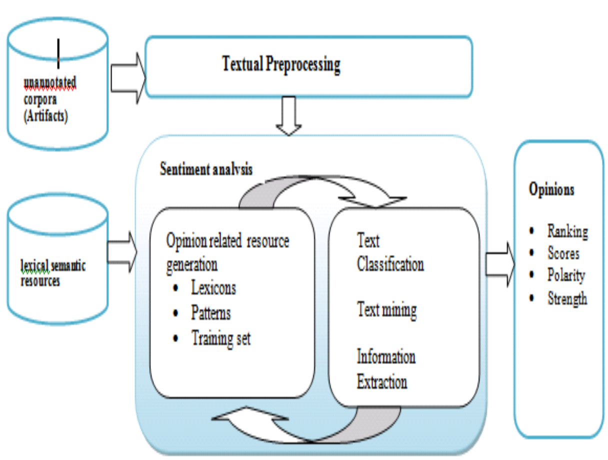
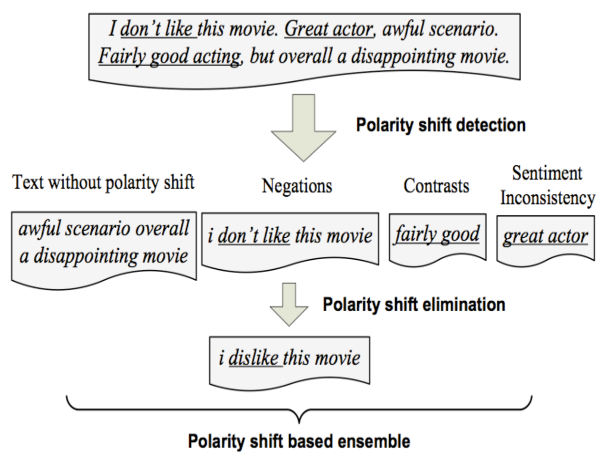


Figure IV-1: System Architecture

The architecture shows the internal working of the sentiment analyzer which is data driven approach for extraction of opinion expressions, their holders and targets require reliable annotated data at the expression level. It collects the un annotated data into the text preprocessing module which refine the text into annotated form. Which is taken by the sentiment analyzer for classification by performing text mining ad data Extraction, it filters individual sentences regarding their topic relevancy and the existence of an opinion or factual information. Then it identifies opinion (Sentiments) expression including the respective opinion Scores, Ranking, polarity (Neagtive,Positive, Neutral) and strength(Weak, average, Strong).

4.2 Proposed Approach

This work presents a three-stage cascade model for document-level sentiment classification. The three stages are
1) hybrid polarity shift detection,
2) polarity shift elimination in negations,
3) polarity shift based ensemble model.



5. Conclusion

Considering the feature extraction and tracking of object. The set of datasets were categorized to analyzing and extracting of feature in less time with high performance and more type of sentiments are analyzed. In existing system there is more occlusion, less feature extraction but in proposed system there will be occlusion reduction, and more feature extraction in less time. Random forest classifier is an ensemble of decision trees. Each decision tree can be constructed independently making this an embarrassingly parallelizable problem. Consequently, it can very easily incorporate in the system where each process was responsible for creating a decision tree. When classifying a piece of text, the random forest passes the item to each decision tree and the output is the majority vote.

References

[1] Rui Xia1, Feng Xu2, Jianfei Yu1, Yong Qi1,2 and Erik Cambria*3, Polarity Shift Detection, Elimination and Ensemble: A Three-stage Model for Document-level Sentiment Analysis

[2] Rui Xia, Feng Xu, Chengqing Zong, Qianmu Li, Yong Qi, and Tao Li, Dual Sentiment Analysis: Considering Two Sides of One Review, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 8, AUGUST 2015

[3] C. Lin, Y. He, R. Everson, and S. Ruger, "Weakly supervised joint sentiment-topic detection from text," IEEE Trans. Knowl. Data Eng., vol. 24, no. 6, pp. 1134–1145, Jun. 2012 [4] Dialogic Corporation, "Introduction to Cloud Computing'" white paper, 2010

[5] Z. Hai, K. Chang, J. Kim, and C. C. Yang, "Identifying features in opinion mining via intrinsic and extrinsic domain relevance," IEEE Trans. Knowl. Data Eng., vol. 26, no. 3, pp. 447–462, Mar. 2014. [6] Ericsson Corporation, "Cloud computing in Telecommunications," White Paper, 16 December 2012.

[7] S. Li and C. Huang, "Sentiment classification considering negation and contrast transition," in Proc. Pacific Asia Conf. Lang., Inf. Comput., 2009, pp. 307–316.

[8] A. Kennedy and D. Inkpen, "Sentiment classification of movie reviews using contextual valence shifters," Comput. Intell., vol. 22, pp. 110–125, 2006.

[9] S. Li, S. Lee, Y. Chen, C. Huang and G. Zhou, "Sentiment classification and polarity shifting," in Proc. Int. Conf. Comput. Linguistics, 2010, pp. 635–643.

[10] Ana C.E.S.Lima,Leandro De Castro "Automatic sentiment analysis Of Tweeter messages," IEEE Trans. Knowl. DataEng., vol. 24, no. 6, 978-1-4673-5/12, Jun. 2012.

[11] Theresa Wilson,Paul hoffmanan "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysi," in Synthesis Lectures on Human Language Technologies, vol. 5, no. 1. San Rafael, CA,USA: Morgan & Claypool, 2012, pp. 1–165.

.[12] B. Liu, "Sentiment analysis and opinion mining," in Synthesis Lectures on Human Language Technologies, vol. 5, no. 1. San Rafael, CA, USA: Morgan & Claypool, 2012, pp. 1–165 [13] Mei,W., Chan,T. Tse,A , "Tale of clouds : paradigm comparis on sandsome thoughtson research issues," ProceedingsoftheAsia-PacificServices Computing Conference, APSCC' IEEE, 2008, pp - 464–469.

[14] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment classification using machine learning techniques," in Proc. Conf. Empirical Methods Natural Language Process., 2002, pp. 79–86.

[15] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Found. Trends Inf. Retrieval, vol. 2, no. 1/2, pp. 1–135, 2008. [37] L. Polanyi and A. Zaenen, "Contextual lexical valence shifters," in Proc.AAAISpringSymp.ExploringAttitudeAffectText,2004,pp.1–10.

[16] P Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," Proc. Annu. Meeting Assoc. Comput. Linguistics, 2002, pp. 417–424.

[17] Y. Choi and C. Cardie, "Learning with compositional semantics as structural inference for subsentential sentiment analysis," in Proc. Conf.EmpiricalMethodsNaturalLanguageProcess.,2008,pp.793–801.

[18] Amit Pimpalkar, 2Tejashree Wandhe, 3 M. Swati Rao, 4 Minal Kene, Review of Online Product using Rule Based and Fuzzy Logic with Smiley's, IJCAT - International Journal of Computing and Technology Volume 1, Issue 1, February 2014

[19] Pushpak Bhattacharyya, Akhel Agrawal, "Sentiment Analysis: A New Approach for Effective Use of Linguistic Knowledge and Exploiting Similarities in a Set of Documents to be Classified", International Conference on Natural Language Processing, IIT Kanpur, India, December, 2005.

[20] Mike Thelwall and Rudy Prabowo, "Sentiment Analysis: A Combined Approach", Journal of Informatics, School of Computing and Information Technology, University of Wolverhampton, UK, Volume 3, Issue 2, pp. 143-157, April 2009.

[21] Pooja Manna,Sonali Bodhke, International Journal of Advance Computation Engineering and Networkinng,IssN:2320-2106, Volume-3,Issue-12,Dec-2015

[22] X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion mining," in Proc. Int. Conf. Web Search Data Mining, 2008, pp. 231–240.

[23] J. Li, G. Zhou, H. Wang, and Q. Zhu, "Learning the scope of negation via shallow semantic parsing," in Proc. Int. Conf. Comput. Linguistics, 2010, pp. 671–679.

[24] S. Li and C. Huang, "Sentiment classification considering negation and contrast transition," in Proc. Pacific Asia Conf. Lang., Inf. Comput., 2009, pp. 307–316.

[25] X.Yu, Y. Liu, X. Huang, and A. An, "Mining online reviews for predicting sales performance: A case study in the movie domain," IEEE Trans. Knowl. Data Eng., vol. 24, no. 4, pp. 720–734, Apr. 2012.

[26] T. Nakagawa, K. Inui, and S. Kurohashi. "Dependency tree-based sentiment classification using CRFs with

hidden variables," in Proc. Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics, 2010, pp. 786–794.

[27] A. McCallum and K. Nigam, "A comparison of event models for naive Bayes text classification," in Proc. AAAI Workshop Learn. Text Categorization, 1998, pp. 307–314.

[28] Finding Opinion Strength Using Fuzzy Logic on Web Reviews Animesh Kar, Deba Prasad Mandal International Journal of Engineering and Industries, volume 2, Number 1, March, 2011

[29] Aurangzeb khan, Baharum Baharudin, Sentiment Classification by Sentence Level Semantic Orientation using SentiWordNet from Online Reviews and Blogs, Int. J Comp Sci. Emerging Tech, Vol-2 No 4 August, 2011.

[30] B. Ohana, "Opinion mining with the SentWordNet lexical resource," Dublin Institute of Technology, 2009

[31] J. Kamps, M.J. Marx, R.J. Mokken, and M. De Rijke, Using wordnet to measuresemantic orientations of adjectives, European Language Resources Association (ELRA), 2004.

[32] A. Khan, B. Baharudin, and K. Khan, "Sentiment Classification Using Sentencelevel Lexical Based Semantic Orientation of Online Reviews," vol. 6, 2011, pp. 11411157.

[33] B. Liu, "Sentiment analysis and subjectivity," Handbook of Natural Language Processing,, 2010, pp. 978-1420.

[34] P.D. Turney, "Mining the web for lexical knowledge to improve keyphrase extraction: Learning from labeled and unlabeled data," Journal of Artificial Intelligence Research, Arxiv preprint cs/0212011, 2002, pp. 1-34.