Sentimental Analysis of Web Financial Reviews
Opportunities and Challenges

Changxuan Wan, Tengjiao Jiang, Dexi Liu and Guoqiong Liao
School of Information and Technology, Jiangxi University of Finance and Economics, Nanchang, China
Jiangxi Key Laboratory of Data and Knowledge Engineering, Jiangxi University of Finance and Economics, Nanchang, China

Keywords: Web Financial Reviews, Web Financial Indexes, Opinion Targets, Sentiment Words, Sentiment Analysis.

Abstract: Web financial reviews are real-time, comprehensive and authentic. The construction and quantification of Web financial indexes based on Web financial reviews is of great significance for the financial early warning for enterprises. Comparing with product reviews and news commentaries, in Web financial reviews, the opinion targets have more diverse compositions, the frequencies of opinion targets' occurrence vary greatly, and the sentiment words' have more diverse parts of speech. These characteristics make the extraction of opinion targets, the construction of Web financial indexes, and opinion targets-based sentimental analysis all more complicated, posing new challenges to natural language processing.

1 INTRODUCTION

With the aggravation of the global financial market's instability, more attention is paid to the financial crisis prediction for enterprises. Currently, in most studies, financial crisis prediction is done by a prediction model established based on the data in financial statements.

However, there are several drawbacks in using financial statements. First, financial statements are easily manipulated; second, data in financial statements are static, ignoring the characteristic of time series of enterprise financial ratios; thirdly, financial statements are released yearly, the data in them are not real-time; and fourthly, the influence of historic accumulation of financial ratios on present situation is not considered. Therefore, determining risks merely based on financial indicators in financial statements is doomed to cause deviation of the prediction.

According to the theory of modern corporation performance evaluation, in the era of knowledge economy, comprehensive performance evaluation equation must be introduced into the performance evaluation of corporations; various non-financial indicators should be added on the basis of conventional financial indicators.

Comparing with financial indicators, the biggest advantage of non-financial indicators is that they are future-oriented, while financial indicators are past-oriented. In The Choice of Performance Measures in Annual Bonus Contracts published by Wharton in 1995, it is pointed out that non-financial indicators are better indicators reflecting the management's performance and the company's development prospects. Unlike financial indicators, non-financial indicators can not be obtained by calculating financial data, so the non-financial index also has some shortcomings, such as data collection and quantify are difficult. Then, how should we overcome the aforementioned drawbacks of financial indicators and non-financial indicators and introduce non-financial indicators on the basis of financial indicators to establish a uniform indicator evaluation system?

With the arrival of the era of big data, large amount of financial reviews occur on the Internet every minute. Those reviews cover all aspects of enterprises' past and current operation and prospects of enterprises, and contain thorough analyses of enterprises' financial and non-financial indicators. Therefore, not only are Web financial reviews real-time and readily available, they are also comprehensive and have wide coverage. In addition, Web financial reviews contain expert interpretation, practical experience of investors, as well as customer experience, thus are in-depth and authentic. These characteristics of Web financial reviews, on the one hand, overcome the drawbacks of indicators from financial statements,

Wan C., Jiang T., Liu D. and Liao G..
Sentimental Analysis of Web Financial Reviews - Opportunities and Challenges.
DOI: 10.5220/0005137403660373
In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval* (KDIR-2014), pages 366-373
ISBN: 978-989-758-048-2

and on the other hand, reveal the influences of financial and non-financial indicators on enterprises' future development. By classifying the opinion targets extracted from Web-financial reviews, the financial and non-financial indicators in the reviews can be obtained, and we call them the **Web financial indexes**.

Being real-time, comprehensive, and in-depth, Web financial reviews make possible comprehensiveness and systematicness of financial and non-financial indicators in enterprise financial early warning models. In addition, the sentiment inclination of the reviews makes possible quantification of Web financial indexes.

The extraction and sentimental quantification of Web financial indexes based on Web financial reviews is a meaningful, yet extraordinarily challenging project.

1) Extraction of opinion targets in Web financial reviews and construction of Web financial indexes

The opinion targets refer to the objects modified by the evaluative words. For example, the words component, function and service of a product in product reviews, and the words people, event and subject of conversation in news commentaries are all opinion targets. Groups of opinion targets constitute topics, for example, the opinion targets in product reviews can be classified into product group. In Web financial reviews, opinion targets can be a national policy, a sub-item in financial statements, or a subject of conversation. The grouping of opinion targets results in Web financial indexes. Now available studies mostly focus on the extraction and grouping of opinion targets in product reviews, and by comparison, the extraction and classification of opinion targets in Web financial reviews are much more complicated because those reviews involve wide range of areas.

(1) The opinion targets in product reviews are generally nouns or noun phrases, such as 'Apple', 'screen' and 'keyboard layout' in a cellphone review. In financial reviews, in addition to being nouns or noun phrases such as 'raw material' and 'stock price', opinion targets can also be subordinate clauses. For example, in the sentence 'Share price rise quickly is good.', the opinion target of the sentiment word 'good' is a verb phase ' share price rise quickly '. Therefore, the extraction of opinion targets from Web financial reviews is more complicated.

(2) In product reviews, opinion targets are more evenly distributed. For instance, in a cellphone review, the user would generally comment on the appearance, audio, image display, etc. Financial reviews may contain interpretation of financial statements, deciphering of macro policies, and analysis of personnel movement. The different numbers of comments in different categories lead to very different frequencies of opinion targets' occurrence. Consequently, the construction of Web financial indexes based on opinion targets grouping is much more complicated.

2) Quantification of Web financial indexes

In opinion target-based sentimental analysis, the sentiment value of each opinion target is first calculated based on the sentiment phrase, and then the opinion target is classified into corresponding topic/indicator based on the grouping of opinion targets.

In product reviews, sentiment words are usually adjectives, and available studies mostly perform sentimental analysis based on adjective sentiment words. Different from product reviews, Web financial reviews contain sentiment words that have more diverse parts of speech. Besides being adjective, those sentiment words may be verb, adverb or noun, especially verb. For example, in the previous example 'Share price rise quickly is a good thing', the word 'rise' is a verb sentiment word, and the phrase 'good thing' is a noun sentiment word. The diversity of sentiment words' parts of speech in financial reviews makes the identification of sentiment words and the calculation of those words' polarity and intensity more difficult. In addition, this diversity results in more flexible components the sentiment words serve as in sentences, thus the sentiment word-based extraction of opinion targets is also more difficult.

The diversity of opinion targets' composition, differences in opinion targets' frequencies, and richness of sentiment words' parts of speech in Web financial reviews make the extraction of opinion targets, the construction of Web financial indexes, as well as the opinion target-based sentimental analysis all more complicated, and bring new challenges to natural language processing.

2 RELEVANT STUDIES

2.1 Relation between Web Financial Reviews and Enterprises' Financial Statuses

The first research on Web financial reviews was done by Wysocki (1999). After investigating the 50 listed companies that had the greatest amounts of information during Jan 1998 to Aug 1998, Wysocki noticed that with information from the notice board, the trading volume and abnormal stock returns of the next day can be predicted. Das and Chen (2007), after exploring information from Yahoo, Amazon and other forums, found that the information contain contents significantly correlated with the return on assets. Tetlock *et al.* (2008) conducted research on the relationship between the negative words in news reports about S&P500 companies from 1980 to 2004 and the companies' profits and stock returns, and discovered that the negative words in news reports about listed companies can be used for the prediction of those companies' future lower profits. Si *et al.* (2013) reported that the topic-based public sentiments in Web financial reviews could help improve the accuracy of stock price prediction. The study of Bian *et al.* (2013) revealed that the result of sentimental analysis of Web financial information can be used as important indicator for the financial early warning for listed companies.

The literature discussed above reveals the influence of the amount, popularity and content of Web financial reviews on investors as well as their reflection in stock trading. However, available studies, when performing mining of Web financial reviews, only pay attention to sentimental polarity at document level, or roughly count the numbers of positive and negative sentiment words in a document to determine the document's polarity. In fact, Web financial reviews contain far more information. In general, every document contains multi-facet (financial indicators and non-financial indicators) interpretation of an enterprise. After determining the sentimental polarity of each financial or non-financial indicator, the sentiment value of the indicator can be applied to financial early warning models in order to perform more detailed, accurate, and refined analysis and prediction. The construction and quantification of Web financial indexes are detailed in the next few sub-sections in three aspects, the extraction of targets, extraction of target-sentiment word pairs, and target-based sentimental analysis of texts.

2.2 Extraction of Targets

The opinion targets in product reviews are also called aspects or attributes. For the extraction of opinion targets, there are mainly three types of methods, which are methods based on frequent nouns and rules-based methods, machine learning-based methods, and topic model-based methods.

1) Frequent nouns and rules-based methods

The frequent noun and rule-based methods generally extract opinion targets based on the following heuristic rule: the opinion targets in product reviews are generally noun or noun phrase, extract these words first, and then use the opinion target-sentiment word relation to extract new opinion targets and sentiment words. Hu *et al.* (2004) proposed to, based on massive corpuses of a certain field, first identify nouns based on part of speech labeling, and then use Apriori algorithm to find frequent nouns or noun phrases as the opinion targets. On the basis of this study of Hu *et al.*, Popescu *et al.* (2005) improved the accuracy of the algorithm by further filtering the nouns or noun phrases. They proposed to determine whether a noun or noun phrase is an opinion target by calculating the pointwise mutual information (PMI) between the noun or noun phrase and the classification of the opinion target to be extracted. This method has improved opinion target extraction accuracy, yet somewhat decreased recall rate. Moghaddam *et al.* (2010) tried to determine the occurrence pattern of product features based on standard product features defined in fine-grained product reviews, and filter high-frequency words based on the hit rates of high-frequency nouns or noun phrases in the pattern.

2) Machine learning-based methods

Some other researchers applied machine learning to the identification of opinion targets. Jakob *et al.* (2010) used conditional random field (CRF) model to extract opinion targets. Jin *et al.* (2009) treated the extraction of characteristic words and sentiment words as a sequence labeling task, where each word in a review corresponds to a category label, and proposed to adopt lexicalized hidden Markov model (HMM) to search for the most possible label sequence. Su *et al.* (2008) proposed a novel mutual strengthening criterion for the mining of hidden association between opinion targets and sentiment words, and to identify hidden opinion targets based on clustering.

3) Topic model-based methods

In recent years, with topic model gradually becomes popular, scholars are applying it to the field of sentimental analysis. It is used for identification and classification of opinion targets.

Titov *et al.* (2008) found that standard latent Dirichet allocation (LDA) model is not suitable for extraction of fine-grained opinion targets. They proposed a multi-grain latent Dirichlet allocation (MG-LDA) model and a multi-aspect sentiment (MAS) model, which are able to discover not only general opinion targets, but also fine-grained opinion targets. Andrzejewski *et al.* (2009) first proposed DF-LDA model with constraints. They introduced two types of constraints, must-link and cannot link, as priori knowledge. However, as the number of documents increases, the computational complexity of this model increases exponentially. Zhai *et al.* (2011) proposed a constrained-LDA model for the extraction and classification of opinion targets from product reviews. They also set two types of constraints, must-link and cannot-link; the former classifies opinion targets with the same composition to the same topic, while the latter classifies opinion targets in the same sentence to different topics. Moghaddam *et al.* (2013) presented a factorized LDA (FLDA) model for cold start items, which models opinion targets and reviewers at the same time, and performs opinion target extraction and rating on the basis of opinion target classification.

The methods based on frequent nouns and rules may cause omitting of some opinion targets that occur at lower frequencies. In addition, not all frequent nouns are opinion targets. With the machine learning-based methods, training sets need to be labeled manually, and data sets from different fields have poor transferability. The topic model-based methods cluster opinion targets with similar semantics to the same topic, and are able to explain the membership degrees of individual words to a topic, thus well serve the purpose of opinion target extraction and classification.

2.3 Extraction of Opinion Target- Sentiment Word Pairs

There are two types of methods for the extraction of opinion target-sentiment word pairs, machine learning based methods and syntactic rule-based methods.

1) Machine Learning-based Methods

Jin *et al.* (2009) considered the extraction of product features and sentiment words as a sequence labeling task, and proposed to label the most possible label sequence based on hidden Markov model, and then use the sequence to identify opinion targets and sentiment words. Lakkaraju *et al.* (2011) proposed to use hidden Markov model to describe the syntactic dependencies between opinion targets and sentiment words, and extract product features and corresponding sentiment words based on contextual consistency.

2) Syntactic Rules-based Methods

Kobayashi *et al.* (2004) adopted templates to express the modification relationship between opinion targets and sentiment words; they designed eight templates for this purpose. Bloom *et al.* (2007) adopted Stanford Parser and manually constructed 31 syntactic rules to obtain appraisal expressions. Bloom *et al.* (2009) used confidence rating method for the automatic learning of rules. First, all possible appraisal expressions in a sentence are identified, then rules are extracted based on these appraisal expressions, and the extracted rules are eventually used to match the opinion targets and sentiment words in the sentence in order to find rules with higher confidence levels. Qiu *et al.* (2011) reported a method called Double Propagation, which performs identification and extraction of sentiment words and opinion targets at the same time. Kamal *et al.* (2012) designed rules for the extraction of appraisal expressions on the basis of linguistic and syntactic analyses of reviews.

The machine learning-based methods generally treat the extraction of opinion targets and sentiment words as a sequence labeling task; they often require manual labeling of training sets. In contrast, syntactic analysis-based methods extract the syntactic relationships between opinion targets and sentiment words at syntactic level instead of field level, thus demonstrating better adaptability to different fields.

2.4 Opinion Target-based Sentimental Analysis of Texts

Methods for opinion Target-based sentimental analysis could generally be divided into machine learning-based methods and syntactic analysis-based methods.

1) Machine Learning-based Methods

Wilson *et al.* (2009) proposed to identify sentiment words with supervised modification methods. Their experiment showed that supervised learning classifier that fuses multiple features can greatly improve the extraction of opinion targets and sentiment words. Fang *et al.* (2012) used latent structural model to realize the opinion targets in product reviews, and their method is capable of identifying the opinion targets' level in the meantime. Liu *et al.* (2012) realized fine-grained opinion mining based on word-based translation model. Lu *et al.* (2011) proposed a segmented topic model (STM), which performs joint modeling of topic distribution at document and sentence levels with two-parameter Poisson-Dirichlet process, labels sentences with weak supervision to strengthen the direct correlation between topic and aspect/opinion targets, and obtains better multi-aspect evaluative polarity by combining overall rating and sentence labeling. Kontopoulos *et al.* (2013) proposed an ontology-based dispatching method that effectively improves Twitter sentimental analysis. This method allocates sentiment scores to relevant attributes/opinion targets of each topic, and is thus able to conduct more detailed sentimental analysis of each specific topic.

These methods generally omit the syntactic information of the sentences that the sentiment words are in. This issue has been noticed by some researchers, who have tried to determine the fine-grained polarity of texts through syntactic analysis.

2) Syntactic Analysis-based Methods

Feng *et al.* (2012), based on dependency parsing, obtained ADV dependency pair in adverbial-verb structure with sentiment word as the center, and on this basis, obtained the sentiment values of sentiment sentences in micro-blogs. Wan *et al.* (2013) proposed to determine the sentimental polarity of Web financial reviews through sentimental analysis of dependency pairs.

Machine learning-based methods treat the association between opinion targets and sentiment words as a sequence label, without taking the syntactic association between opinion targets and sentiment words into consideration. At present, the opinion target-based polarity analysis of Web financial reviews is not adequately thorough.

3 CHALLENGES IN SENTIMENTAL ANALYSIS OF WEB FINANCIAL REVIEWS AND CORRESPONDING STRATEGIES

3.1 Extraction of Opinion Targets

The available studies of opinion target extraction mostly focus on the extraction of opinion targets in product reviews. They generally limit opinion targets as nouns or noun phrases and then perform further identification. As the analysis in section 2.2 indicates, the frequent nouns and rules-based methods essentially count the occurrence of nouns or noun phrases. With supervised learning, better results can be obtained when there is adequate training data. However, with the rapidly growing information on the Internet, newly occurred information may not be labeled and become training corpuses before they are outdated. Although the various emerging semi-supervised learning methods are trying to remedy this defect, iterative learning started from a seed set will exhibit deviations after large amount of training, and the consequent manual deviation rectification and adjustment are massive. In recent years, statistic topic model is becoming a popular method for the topic discovery in massive documents. The advantage of topic model is that besides opinion targets discovery, it can also perform opinion targets clustering.

The above analysis shows that in Web financial reviews, the composition of opinion targets is more diverse. Instead of being noun or noun phrase, opinion targets can be verb phrase, verb-object phrase, or even sentences. Therefore, we define opinion targets in Web financial reviews as opinion target expressions. Opinion target expressions can be nouns or noun phrases, or composed of noun (phrase) and verb. On the one hand, the nouns in opinion target expressions are conducive to further classification of opinion targets, and on the other hand, the verbs in opinion target expressions help determine the oddity of opinion targets. An odd opinion target reverses the polarity of the sentiment words that modify it. Meanwhile, since in Web financial reviews, the number of comments on each Web financial indicator varies, which result in very different frequencies of the opinion targets. Therefore, direct application of topic model to the extraction and classification of opinion targets in Web financial reviews does not work well.

An opinion target refers to the object of modification of a sentiment word. Dependencies usually exist between sentiment words and opinion targets. Therefore, it is possible to extract opinion targets based on sentiment words and those dependencies.

(1) The diversity of sentiment words' parts of speech makes them able to be different structural components of sentences. In Web financial reviews, sentiment words may be adjective, verb, adverb or noun, while opinion targets may be noun, noun phrase, subject-verb phrase, verb-object phrase, or sentence. Therefore, the rules of syntactic paths between sentiment words and opinion targets in Web financial reviews are far more complex.

(2) When using syntactic paths to extract opinion targets, we noticed that even when sentiment words serve as the same component of sentences, the components their opinion targets serve as may be different. For example, in the sentence '我看中这家公司的发展前景 (I prefer this company's development prospects).' and the phrase '股价上涨 (Stock price rises).', the verb sentiment words '看中 (prefer)' and '上涨(rise)' both serve as predicates, yet their opinion targets' positions are different. The opinion target of 'prefer' is the object of the sentence, while the opinion target of 'rise' is the subject of the sentence. Therefore, on the basis of syntactic rules, the understanding of semantics should be added. For example, psychological verbs modify the objects of the sentences they are in, while non-psychological verbs modify the objects. A natural question is then whether it is possible to identify the components that the adjective sentiment words' opinion targets serve

as based on syntax.

Therefore, it should be considered to extract opinion targets based on both syntactic paths and semantics. Such methods are detailed as follows.

1) Machine Learning and Semantic Analysis Combined Methods

The first step is to label a sentiment word and its candidate opinion targets. Since the opinion targets of sentiment words may be subject or object of a sentence, it is required to label the subject or object the sentiment word modifies.

The second step is to extract all syntactic paths (phrase syntax or dependency syntax) between the sentiment word and all candidate opinion targets.

The third step is to generalize the syntactic paths and form a syntactic path library by descending order of the paths.

The last step is to match the sentiment word with the semantics (synonym or co-occurrence) of subject sentiment words (or object sentiment words). When the threshold similarity reaches a certain value, the subject (or object) is chosen as the sentiment word's opinion target. When the threshold similarity is lower than a certain value, the strategy of 'syntactic path library' plus 'sentiment word' is used to identify the opinion target of the sentiment word.

2) Syntactic Rules and Semantic Analysis Combined Methods

By combining syntactic rules and semantic analysis, it can be directly determined whether a sentiment word's opinion target is the subject or the object.

A sentiment word can be verb, adjective, or adverb. These different conditions are analyzed as follows.

(1) When a sentiment word is a verb, it may be psychological verb or non-psychological verb. It has been proved in studies that when a sentiment word is a psychological verb, it modifies the object. For example, the opinion targets of psychological words '喜欢(like)' and '看中(prefer)' are objects. When the sentiment word is a non-psychological verb, its opinion target is the subject. For example, in the sentence '我看中这家公司的发展前景(I prefer this company's development prospects)', the opinion target of sentiment word ' prefer' is the object of the sentence 'prospects', and it can be further extended to ' this company's development prospects.'; And in the sentence '股价上涨(Stock price rises)', the opinion target of sentiment word 'rise' is the subject of the sentence, i.e. 'stock price'.

(2) When a sentiment word is an adjective or a verb, can a pattern about its opinion targets be summarized? A sentiment word may serve as the predicate of a sentence, the modifiers of the predicate, the object, and the object's modifiers. The following assumptions are then made:

a) When the sentiment word is the predicate, its opinion target is the subject;

b) When the sentiment word is a modifier of the predicate, its opinion target is the subject-predicate structure;

c) When the sentiment word is the object, its opinion target is the subject-predicate structure;

d) When the sentiment word is a modifier of the object, its opinion target is the subject-predicate-object structure;

e) When the sentiment word is the attribute, its opinion target is the word it modifies.

3.2 The Construction of Non-financial Index System

In Web financial reviews, the number of comments on each Web financial indicator varies, which results in very different frequencies of opinion targets' occurrence. For the construction and quantification of Web financial indicators, low-frequency opinion targets are often non-financial indicators, yet they are of great importance in financial early warning for enterprises. Therefore, direct use of topic model for the extraction and grouping of the opinion targets in Web financial reviews would lead to incompleteness of the constructed Web financial indexes. The idea is detailed as follows.

(1) Extract the opinion targets with the syntactic path and semantic analysis combined method introduced in section 3.1, use these opinion targets as the aspects of topic model, and then group the opinion targets with topic model in order to construct the Web financial indexes.

(2) Classify opinion targets according to different topics the Web financial reviews describe, for example the deciphering of macro policies, interpretation of financial statements, and analysis of corporate culture, and then construct a hierarchical LDA model for the hybrid classification of opinion targets of different topics and frequencies.

3.3 Opinion Target-based Sentimental Analysis of Web Financial Reviews

In opinion target-based sentimental analysis of Web financial reviews, since one sentence may contain multiple sentiment words and opinion targets, it is needed to calculate the sentiment value of each opinion target based on the sentiment phrase. Sentiment phrase is a three-component group in the form of < *opinion target expression*, *sentiment word*, *contextual modifiers of sentiment word*>.

1) Influence of Opinion Target Expression on Polarity of Sentiment Phrase

Opinion targets that are noun or noun phrase may exhibit oddity, and an odd opinion target will change the polarity of the sentiment word that modifies it. For example, the word '减少(decrease)' generally exhibits negative polarity, and the phrase '营业收入减少(the operating income decreases)' exhibits negative polarity, while the phrase '损失减少(the loss decreases)' exhibits positive polarity. This is because the word 'loss' is an odd target.

For a opinion target composed of noun and verb, sometimes this verb may also be a sentiment word. For example, in sentences '股价上涨得很快(the stock price rises rapidly)' and '股价下跌得很快(the stock price drops rapidly)', the sentiment word 'rapidly' modifies 'stock price rises' and 'stock price drops', respectively. At this time, the sentimental polarity and intensity of the entire opinion target expression need to be determined first.

2) Influence of Sentiment Word's Contextual Modifiers on Polarity of Sentiment Phrase

The contextual modifiers of sentiment words are mainly negative adverbs and adverbs of degree, and their influences on the polarity of sentiment phrases include:

(1) Influence of negative adverbs on polarity of sentiment words;

(2) Influence of adverbs of degree on polarity of sentiment words;

(3) The distance between negative adverbs or adverbs of degree and sentiment words is called edit distance. When a negative adverb and an adverb of degree modify the same sentiment word simultaneously, different combinations of their edit distances from the sentiment word result in different influences on the polarity and intensity of the sentiment word.

4 CONCLUSIONS

The beginning of the era of big data brings us both opportunities and challenges. Applying data mining to Web financial reviews, which contain abundant information, could help with investors' investment decision-making, enterprise operators' management decision-making, as well as credit rating in the finance and insurance industry.

However, the mining of Web financial reviews faces many challenges, for example the diversity of sentiment words' parts of speech, the diversity of the opinion targets expressions, and the complexity of the construction of Web financial indexes, as well as the sentimental quantification of Web financial indexes caused by these three features. In the meantime, this challenging task is very meaningful.

REFERENCES

Andrzejewski, D., Zhu, X. , and Craven M.(2009). Incorporating domain knowledge into topic modeling via Dirichlet forest priors. *In ICML'09: Proceedings of the International Conference on Machine Learning*, Montreal, Quebec, Canade, 14-18 June 2009. ACM, 25-32.

Bian, H., Wan, C., Liu, D., Jiang, T. (2013). A study of financial crisis prediction model for listed companies taking into account Web financial information. *Computer Science,* 40(11), 295-298, 315. (in Chinese)

Bloom, K., Argamon, S. (2009). Automated learning of appraisal extraction patterns. *Language and Computers*, 71(1), 249-260.

Bloom, K., Garg, N., Argamon, S. (2007). Extracting appraisal expressions. *In EMNLP'07: Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing*, Prague, Czech Republic, 28-30 June 2007. ACL, 308-315.

Das, S., Chen, M. (2007). Yahoo! For Amazon: Sentiment Extraction from Small Talk on the Web. *Management Science*, 53(9),1375-1388.

Fang, L., Huang, M. (2012). Fine granular aspect analysis using latent structural models. *In ACL 2012: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics,* Jeju, Korea,8-14 July 2012. *ACL*, 333-337.

Feng, S., Fu, Y., Yang, F., et al., 2012. Blog sentiment orientation analysis based on dependency parsing. *Journal of Computer Research and Development*, 49(11), 2395-2406. (in Chinese)

Hu, M., Liu, B. (2004). Mining and summarizing customer reviews. *In KDD'04: Proceedings of the 10th ACM International Conference on Knowledge Discovery and Data Mining,* Seattle, Washington, USA, 22-25 August 2004. New York: ACM, 168-177.

Jakob, N., Gurevych, I. (2010). Extracting opinion targets in a single and cross-domain setting with conditional random fields. *In EMNLP'10: Proceedings of Conference on Empirical Methods in Natural Language Processing*. Cambridge, MA, 12-14 October 2010. ACL,1035-1045.

Jin, W., Ho, H., Srihari, R. (2009). OpinionMiner: a novel machine learning system for Web opinion mining and extraction. *In KDD'09: Proceedings of ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, Paris, 19-21 September 2009. New York: ACM, 1195-1204.

Kamal, A., Abulaish, M., Anwar, T. (2012). Mining feature-opinion pairs and their reliability scores from

Web opinion sources. *In WIMS'12: Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics*，Craiova, Romania, 13-15 June 2012. ACM,15.
Kobayashi, N., Inui, K., Matsumoto, Y., Tateishi K., and Fukushima T. (2004). Collecting evaluative expressions for opinion extraction. *In IJCNLP'04: Proceedings of the International Joint Conference on Natural Language Processing*, Hainan Island, China, 22-24 March 2004. Vol.3248, Springer, 596-605.
Kontopoulos, E., Berberidis, C., Dergiades, T., Bassiliades, N. (2013). Ontology-based sentiment analysis of Twitter posts. *Expert Systems with Applications*, 40(10), 4065-4074.
Lakkaraju, H., Bhattacharyya, C., Bhattacharya, I., Merugu, S.(2011). Exploiting coherence for the simultaneous discovery of latent facets and associated sentiments. *In SDM'11: Proceedings of 2011 SIAM International Conference on Data Mining*, Mesa, Arizona USA, 28-30 April 2011. SIAM, 498-509.
Liu, K., Xu, L., Zhao, J. (2012). Opinion target extraction using word-based translation model. *In EMNLP'12: Proceedings of the Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Jeju, Korea,8-14 July 2012. *ACL*, 1346-1356.
Lu, B., Ott, M., Cardie, C., Tsou, B. (2011). Multi-aspect sentiment analysis with topic models. *In ICDM'11: Proceedings of the 11th IEEE International Conference on Data Mining Workshops*, Vancouver, Canada, 11-14 December 2011. IEEE,81-88.
Moghaddam, S., Ester, M.(2013). The FLDA model for aspect-based opinion mining: addressing the cold start problem. *In WWW'13: Proceedings of International Conference on World Wide Web*, Rio, 13-17 May 2013. ACM, 909-918.
Moghaddam, S., Ester, M. (2010). Opinion digger: an unsupervised opinion miner from unstructured product reviews. *In CIKM'10: Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, Toronto, Canada, 26-30 October 2010 ACM, 1825-1828.
Popescu, A., Etzioni, O. (2005). Extracting product features and opinions from reviews. *In HLT/EMNLP'05: Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing,* Vancouver, Canada, 6-8 October 2005. ACL, 339-346.
Qiu, G., Liu, B., Bu, J., Chen, C. (2011). Opinion word expansion and target extraction through double propagation. *Computational Linguistics*, 37(1), 9-27.
Si J., Mukherjee A., Liu, B., Li Q., Li H., Deng X. (2013). Exploiting topic based Twitter sentiment for stock prediction. *In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics,* Sofia, Bulgaria, 4-9 August 2013. ACL, 24-29.
Su, Q., Xu, X., Guo, H., Guo Z., Wu X., Zhang X., Swen B., and Su Z. (2008). Hidden sentiment association in Chinese Web opinion mining. *In WWW'08: Proceedings of the 17th International Conference on World Wide Web*, Beijing, 21-25 April 2008. ACM, 959-968.
Tetloek, P., Saar-Tsechansky, M., Macskassy, S. (2008). More Than Words: Quantifying Language to Measure Firms' Fundamentals. *The Journal of Finance*, 63(3),1437-1467.
Titov, I., McDonald, R. (2008). Modeling online reviews with multi-grain topic models. *In WWW'08: Proceedings of the 17th International Conference on World Wide Web*, Beijing, 21-25 April 2008. ACM, p111-120.
Wan, C., Jiang, T., Zhong, M., Bian, H., 2013. Sentiment computing of Web financial information based on the part-of-speech tagging and dependency parsing. *Journal of Computer Research and Development*, 50(12), 2554-2569. (in Chinese)
Wilson, T., Wiebe, J., Hoffmann, P. (2009). Recognizing contextual polarity: an exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, 35(1), 399-433.
Wysocki, P. (1999). Cheap talk on the Web: the determinants of postings on stock message boards. Working Paper, University of Michigan Business School. No.98025. November 1999.
Zhai, Z., Liu, B., Xu, H., Jia, P.(2011). Constrained LDA for grouping product features in opinion mining. *In PAKDD'11: Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Shenzhen, China, 24-27 May 2011. Springer Berlin Heidelberg, 448-459.