Postprint: Sentiment Analysis of Financial Forum Texts Integrating Syntactic Information

Authors: Lan Qiujun, Liu Wenxing, LI Weikang, Hu Xingye

Date: 2017-10-11T00:00:00+00:00

Abstract

[Purpose] To accurately identify the sentiment orientation of financial forum texts, a sentiment analysis method based on dependency parsing is proposed. [Method] Based on the analysis results of dependency parsing, sentiment backbone extraction is performed on sentences; then, according to different types of dependency relations and different part-of-speech collocations, sentiment computation rules are defined to calculate the sentiment orientation of sentences. [Results] Experimental results show that the overall accuracy of the method is 84.46%; the average precision and recall for the bullish category are 82.84% and 87.14% respectively, with an F-value of 84.94%; the average precision and recall for the bearish category are 86.28% and 81.74% respectively, with an F-value of 83.95%. [Limitations] The correlation between clauses was not fully considered in sentiment computation. [Conclusion] The use of dependency parsing can effectively improve the accuracy of sentiment computation for financial forum texts.

Full Text

Sentiment Analysis of Financial Forum Text Incorporating Syntactic Information

Lan Qiujun, Liu Wenxing, Li Weikang, Hu Xingye
(School of Business Administration, Hunan University, Changsha 410082, China)

Abstract

[Objective] To accurately identify sentiment orientation in financial forum text, this paper proposes a sentiment analysis method based on dependency parsing. [**Methods**] Building upon dependency parsing results, we extract sentiment stems from sentences and define sentiment computation rules according to different dependency relation types and part-of-speech combinations to calculate sentence-level sentiment orientation. [**Results**] Experimental results demonstrate that the proposed method achieves an overall accuracy of 84.46%. For the bullish class, the average precision and recall rates are 82.84% and 87.14%, respectively, with an F-measure of 84.94%. For the bearish class, the average precision and recall rates are 86.28% and 81.74%, respectively, with an F-measure of 83.95%. [**Limitations**] The method does not fully consider relationships between clauses in sentiment computation. [**Conclusions**] Dependency parsing can effectively improve the accuracy of sentiment analysis for financial forum text.

Keywords: Sentiment analysis; Dependency parsing; Financial forum text; Text mining

1. Introduction

With the development and popularization of the Internet, people are no longer satisfied with passively receiving online information. An increasing number of users actively express their opinions and emotions online. Against this backdrop, text sentiment analysis technology has emerged. Also known as opinion mining, sentiment analysis belongs to the field of natural language processing and aims to automatically identify evaluations, attitudes, and emotions toward products, services, organizations, events, and other entities in text [**?**]. This technology undoubtedly holds tremendous potential for understanding public sentiment, monitoring opinion trends, improving product quality, and enhancing service levels. In the financial domain, behavioral finance theory has demonstrated that investor sentiment is a crucial variable in financial markets. The idea of mining investor sentiment from online forums, news, microblogs, and other sources to serve as a basis for investment decisions has attracted considerable attention from financial analysts [**?**]. However, financial forum corpora exhibit characteristics of short text, including sparse features and high noise, which pose significant challenges to traditional sentiment analysis methods [**?**]. To accurately identify sentiment orientation in financial forum text, this study employs dependency parsing technology to uncover semantic modification relationships between words, thereby improving sentiment analysis performance for financial forum corpora.

Current sentiment analysis techniques for financial forum text fall into two main categories: dictionary-based methods and machine learning methods. Dictionary-based methods are the simplest, relying primarily on the bag-of-words model that treats text as an unordered collection of words. These methods identify sentiment words in text based on a sentiment dictionary and obtain the overall text sentiment orientation by accumulating sentiment scores of individual words. For instance, Duan et al. [**?**] categorized emotions into five levels and determined the sentiment of entire posts based on matching results between vocabulary in post content and predefined keyword libraries for each level. References [?] explored the impact of investor sentiment on stock markets using the ROST system developed by Wuhan University, which identifies sentiment words and degree words in text based on sentiment and degree dictionaries to determine single-sentence sentiment orientation. Machine learning methods currently dominate the sentiment analysis field and are particularly common in processing financial forum corpora. For example, references [?] employed support vector machine classification to perform sentiment classification on online public opinion data from Eastmoney, Sina, and Hexun stock forums, constructing sentiment indices for stock price prediction. References [?] used the widely applied open-source software Weka, comparing multiple algorithms and ultimately selecting the best-performing KNN algorithm for sentiment classification while constructing sentiment indices to study their impact on stock markets.

Among these two approaches, dictionary-based methods offer the advantage of simple concepts and straightforward implementation, relying on a well-defined sentiment dictionary with simple accumulation of individual word scores. Machine learning methods, while not requiring a sentiment dictionary, can automatically extract information from large corpora to build sentiment computation models and demonstrate good performance in practice. However, machine learning methods require a sufficiently large, annotated corpus as training data. It should be noted that both current approaches are based on word statistics without analyzing or utilizing deep syntactic structures and semantic relationships in text.

In fact, Chinese is a highly complex language where the same words can produce different semantic relationships under different syntactic structures, resulting in vastly different sentiment orientations. Consequently, an increasing number of scholars have begun using syntactic analysis to improve sentiment analysis accuracy. For example, Xia et al. [?] utilized syntactic analysis and conditional random fields (CRFs) to extract candidate evaluation objects for sentiment classification of microblog posts using SVM. Zhang et al. [?] constructed ternary dependency features consisting of head words, dependent words, and dependency relations through dependency parsing, employing support vector machines and deep belief networks for sentiment classification of hotel review corpora. Nakagawa et al. [?] used English and Japanese dependency trees as feature inputs for CRFs models for text sentiment classification. Xiao et al. [?] employed syntactic analysis to identify different roles played by words in sentences (subject, predicate, object, attribute, adverbial, complement) and assigned different weights to these roles to calculate sentence sentiment indices. While these methods have improved sentiment analysis performance by applying syntactic information from various perspectives, they rarely consider modification relationships between words, particularly the impact of modification relationships and part-of-speech combinations on sentence sentiment.

This paper leverages dependency parsing technology to obtain syntactic structures and modification relationships between words for sentiment transfer. Innovatively, we treat subject-predicate-object relationships and sentence cores as sentiment stems. Based on statistical analysis of large-scale corpora, we propose several sentiment computation rules and ultimately construct a sentiment computation model. Experimental results demonstrate that this model significantly improves accuracy and recall compared to traditional machine learning methods.

3. Methodology

3.1 Dictionary Construction For sentiment analysis purposes, we constructed three dictionaries: a sentiment dictionary, a negation dictionary, and a degree dictionary. Due to the informal nature of forum corpora and the particular requirements of financial domain sentiment analysis, existing Chinese sentiment dictionaries such as HowNet and NTUSD are inadequate. Therefore, we employed the SO-PMI method [?] to build a domain-specific sentiment dictionary. This method uses a set of positive sentiment words (pLists) and negative sentiment words (nLists) as benchmark words. The sentiment orientation of a target word is determined based on the difference in pointwise mutual information between the word and pLists versus nLists. Let N represent the total number of documents in the corpus, df(x&y) denote the number of documents where words x and y co-occur, and df(z) denote the number of documents containing word z. The calculation formula is as follows:

SO-PMI(word) p pLists N df (word & p) df (word) df (p) n nLists N df (word & n) df (word) df (n)

Words with SO-PMI values greater than 0 are classified as positive sentiment words, while those with values less than 0 are classified as negative sentiment words. Through manual screening and adjustment of the sentiment dictionary generated by the SO-PMI algorithm, we obtained 1,404 positive sentiment words and 926 negative sentiment words. The degree dictionary was downloaded from Datatang , containing 61 words. The negation dictionary was manually constructed with 21 words.

3.2 Dependency Parsing Dependency parsing is a crucial technology in natural language processing that automatically analyzes input text to obtain its syntactic structure [?]. Through dependency parsing, we can understand modification relationships between words in a sentence, which can be conveniently applied to sentence sentiment orientation analysis. Current dependency parsing tools primarily include the LTP Language Technology Platform from Harbin Institute of Technology [?], Fudan University NLP Dependency Parser, and the Stanford Parser. Among these, the LTP platform is a relatively mature Chinese natural language processing platform that provides an efficient, accurate, and open set of text processing modules, achieving second place in the SANCL 2012 web data dependency parsing evaluation. Considering both openness and accuracy, this paper selects the LTP platform for dependency parsing implementation. LTP defines 14 types of dependency relations, as shown in .

[Figure 1: see original paper] provides an example of LTP dependency parsing results. Each dependency relation consists of a head word and a modifier word. In LTP analysis results, the head word points to the modifier word through a dependency arc, with the specific dependency relation type annotated on the arc [?].

3.3 Sentiment Stem Extraction and Sentiment Transfer From a Chinese grammatical perspective, subject-predicate-object relationships serve as the main structure of a sentence and essentially express what the narrator intends to convey. Consider the example: "Regardless of how things develop in the end, this is obviously bad news for commodity-exporting countries." Without considering the narrator' s target, focusing solely on their view and attitude toward the event, the subject-predicate-object structure "this is bad news" basically expresses the narrator' s opinion. However, due to the short-text nature and informal language characteristics of forums, some sentences lack subject-predicate-object structures. In such cases, according to LTP analysis results, the HED (head) relation serves as an excellent alternative. Based on LTP output, regardless of how brief or irregular a sentence may be, it always contains an HED relation (i.e., sentence core). The HED relation describes the core of the entire sentence, summarizing its central idea and representing the primary component for understanding the narrator' s attitude. Based on these considerations, this paper proposes the following sentiment stem extraction strategy: use the sentence' s subject-predicate-object structure as the main stem; if such structure does not exist, use the sentence core as the sentiment stem.

After obtaining sentiment stems, we further extract modifier words for each stem word. Sentiment transfer propagates the sentiment values of modifier words to the stem words. For example, in "the door to rise is about to open," the subject-predicate relation is "door open," which itself carries no sentiment orientation. However, there exists a dependency arc between "rise" and the subject "door," indicating that "rise" modifies "door." We can first transfer the sentiment value of "rise" to "door," calculating the combined sentiment value of "rise + door." At this point, "door" possesses a sentiment value, and when examining "door open," it already carries sentiment. It should be noted that not all dependency relations can transfer sentiment. Following Wan et al. [?], we only consider sentiment transfer across six types of dependency relations, as shown in .

3.4 Sentiment Computation Rules In reality, under different dependency relations and part-of-speech combinations, the sentiment impact of modifier words on head words (i.e., stem words) varies, resulting in differences in sentiment transfer. Therefore, specific sentiment computation rules must be established based on part-of-speech combinations in dependency relations. Existing research primarily analyzes part-of-speech combinations in dependency relations from a purely linguistic perspective. However, forum corpora feature casual narration and heavy colloquialism, making it difficult to cover all part-of-speech combinations from a purely linguistic standpoint. Therefore, through analysis of large-scale financial forum corpora, we statistically examined part-of-speech combinations appearing in the six dependency relations that may affect text sentiment orientation. lists the top six most frequent part-of-speech combinations for each dependency relation, formatted as "modifier + head." The results show that except for the ATT relation, the cumulative frequency of the top six part-of-speech combinations for all other dependency relations exceeds 80%, while ATT approaches 60%. For simplicity, we only establish sentiment computation rules for the top six frequent part-of-speech combinations of each dependency relation; for all other cases, we simply use the sum of sentiment scores of head and modifier words as the combined sentiment score.

Based on observations and analysis of extensive financial forum corpora, and drawing upon relevant research findings from reference [?], we established eight categories of sentiment computation rules primarily according to six types of inter-word dependency relations and main structural relations. For convenience, let S(·) denote the sentiment score of a word or clause, D(·) denote the degree value of degree adverbs, P(·) denote the sentiment polarity of words, and Score represent the calculated score according to rules. Additionally, let mw denote modifier words, cw denote head words, dd denote degree adverbs, nd denote negation adverbs, and other part-of-speech symbols are shown in .

(1) ADV Rules

ADV represents the adverbial-head relation where modifier words function as adverbials modifying head words. When the modifier is an adverb, it changes the sentiment intensity or reverses the polarity of the modified word. For example, in "the financing purchase amount is too large, decided not to enter tomorrow," the degree adverb "too" intensifies the sentiment of "large," so the combined sentiment value can be set as the product of the adverb' s degree value and the verb' s sentiment value. The negation adverb "not" reverses the polarity of "enter," so the combined sentiment value can be set as the negative of the verb' s sentiment value. For "adjective + verb" combinations such as "steadily taking over" or "successfully breaking through," since the adjective modifies the verb but the focus remains on the verb, we set the combined sentiment value as a weighted sum with the adjective weight lower than the verb, specifically half of the verb's weight. For "verb + verb"combinations, based on corpus observations, in most cases both words share the same sentiment polarity and it is difficult to distinguish which is more important, such as "enter and grab chips,"so we use the sum of both words as the combined sentiment value. When modifiers are time nouns, prepositions, measure words, etc., these generally carry no sentiment orientation, so the combined sentiment value equals the head word' s sentiment value. Therefore, this rule category can be expressed as:

if (mw, cw) is ((d, v) or (d, a)) and mw is dd then Score=D(mw)× S(cw);
if (mw, cw) is ((d, v) or (d, a)) and mw is nd then Score= -S(cw);
if (mw, cw) is (a, v) then Score 0.5 S(mw) S(cw)

if (mw, cw) is (v, v) then Score S(mw) S(cw)

Note that for “verb + verb”combinations where modifier and head share the same part-of-speech, this situation also exists in other dependency relation categories below, and in most cases the importance of sentiment cannot be distinguished. Therefore, subsequent rules handle these similarly. Additionally, when modifiers are time nouns, prepositions, measure words, etc., similar situations exist in other rule categories and are handled identically, without further elaboration.

(2) ATT Rules

ATT represents the attribute-head relation between an attributive and its head word. When modifiers are verbs or adjectives and the head is a noun, the verb or adjective functions as an attributive modifying the noun. For example, in “this is a big conspiracy,” although “big” modifies “conspiracy,” the sentiment orientation of the entire dependency relation depends on the polarity of the noun “conspiracy.” Therefore, this rule category can be expressed as:

if (mw, cw) is ((r, n) or (m, n) or (q, n)) then Score= S(cw);
if (mw, cw) is (n, n) then Score S(mw) S(cw)
if (mw, cw) is ((v, n) or (a, n)) then Score S(mw) P(cw)

(3) COO Rules

COO represents a coordinate relation between two words, such as “suppress and beat down” where “suppress” and “beat down” have equal status. Therefore, we use the sum of modifier and head word scores as the combined sentiment value.

Score S(mw) S(cw)

(4) SBV Rules

SBV represents the subject-verb relation, which forms the main structure of a sentence. For “noun + verb” and “noun + adjective” combinations, the sentiment orientation of the entire dependency relation largely depends on the noun’ s sentiment. For example, in “bulls will not easily waver,” the positive sentiment of “bulls” carries significant weight. Therefore, this rule category can be expressed as:

if (mw, cw) is ((r, v) or (nh, v) or (ns, v)) then Score = S(cw);
if (mw, cw) is (v, v) then Score S(mw) S(cw)
if (mw, cw) is ((n, v) or (n, a)) then Score = S(mw) + 0.5 $\times$ S(cw);

(5) VOB Rules

VOB represents the verb-object relation. For “noun + verb” and “adjective + verb” combinations, the noun or adjective serves as the object of the verb action, so sentiment is primarily expressed through the verb. For example, in “stock price breaks through resistance,” the sentiment is mainly reflected in “breaks through.” Therefore, this rule category can be expressed as:

if (mw, cw) is ((r, v) or (m, v) or (q, v)) then Score= S(cw);
if (mw, cw) is (v, v) then Score S(mw) S(cw)
if (mw, cw) is ((n, v) or (a, v)) then Score 0.5 S(mw) S(cw)

(6) CMP Rules
CMP represents the complement relation, which supplements the action expressed by a verb. Statistical analysis shows this relation rarely appears in financial forum corpora, so we simply use the sum of modifier and head word scores.

Score S(mw) S(cw)

(7) IS-DO Rules
Based on different predicates, subject-predicate-object relations can be divided into two categories: "be" type and "do" type. The "be" type (with predicates like "is," "is exactly," "for," etc.) explains what the subject is, with focus on the object, such as "bad news is an excellent buying opportunity." The "do" type explains what the subject does or what happens under the predicate verb' s action, such as "major players are pushing up stock prices." For "be" type relations, since the focus is on the object, when the object is a sentiment word, we use its sentiment score as the entire relation' s sentiment value; otherwise, we return the subject' s sentiment score. For "do" type relations, we calculate sentiment scores for subject-predicate and predicate-object pairs separately, using their sum as the final sentiment value.

(8) Inter-clause Rules
We observe that financial forum corpora rarely contain complex clausal relationships such as transitions. To simplify computation, we define the sentiment value of an entire sentence as the sum of its individual clause sentiment values. The computation rule is therefore:

Score S(s) S(s) S(s) is for each clause.

3.5 Sentiment Computation Model We use the NLPIR Chinese word segmentation system from the Institute of Computing Technology, Chinese Academy of Sciences [?] to segment each text for analysis. The segmented text is submitted to LTP for processing in XML format via POST requests. Based on LTP' s returned results, we extract sentence sentiment stems and perform inter-word sentiment transfer and computation on these stems according to the constructed dictionaries and sentiment computation rules, ultimately obtaining the overall sentence sentiment category. The sentiment computation model is illustrated in [Figure 2: see original paper].

4. Experiments

4.1 Data Collection and Preprocessing Experimental corpora were collected from the Eastmoney stock forum using a web crawler. We selected five stocks from the biomedical sector, using each post' s title as the collection object. After noise removal, we obtained 31,815 data entries, as shown in .

The corpora were annotated by financial professionals using five sentiment levels: "strongly bullish," "weakly bullish," "neutral," "weakly bearish," and "strongly bearish.” Considering that different annotators showed some disagreement on the three middle levels but relatively high consistency on “strongly bullish” and “strongly bearish,” we only selected data annotated as “strongly bullish” and “strongly bearish” from each stock’ s data, categorizing them into “bullish” (p) and “bearish” (n) classes. This yielded a total of 5,430 data entries, as shown in .

4.2 Experimental Results and Analysis Reference [**?**] tested four common algorithms—KNN, Naive Bayes, Decision Tree, and Support Vector Machine—for sentiment mining in stock forums, with KNN achieving the highest accuracy. To verify the effectiveness of our method, we use KNN as a baseline for comparison. Additionally, N-Gram establishes conditional probability models based on statistical dependencies between words and is a common text classification method. Cui et al. [**?**] suggest that N>3 yields good results, so we also include it as a baseline.

During experiments, two-thirds of the dataset were used for training and one-third for testing. For each method, we performed 10 experiments using random sampling, recording accuracy, precision, recall, and F-measure on the test set for each experiment. We compare the average results of our method across 10 experiments with the best single-run results of KNN and N-Gram.

The experimental results are shown in . Our method achieves an overall accuracy of 84.46%, representing a clear improvement over KNN and N-Gram text classification methods. Examining bullish and bearish classes separately, our method’s bullish recall rate of 87.14% shows substantial improvement over KNN’ s 83.18%. The improvement in bearish recall is even more pronounced. The F-measure comprehensively considers precision and recall. Our method’ s bullish F-measure of 84.94% represents improvements of 6.82% and 14.15% over KNN (78.12%) and N-Gram (70.79%), respectively. The bearish F-measure of 83.95% shows improvements of 9.16% and 12.65% over KNN (74.79%) and N-Gram (71.30%), respectively. These results fully demonstrate that syntax structure-based sentiment computation methods outperform pure word frequency-based machine learning methods.

5. Conclusion

This paper proposes a sentiment analysis method for financial forum corpora based on dependency parsing. Compared with machine learning methods, our approach shows significant improvements in accuracy, recall, and F-measure, fully demonstrating the value of syntactic structure and semantic information for text sentiment analysis.

Due to the complex structure and rich, varied expressions of the Chinese language, our proposed method does not fully exploit syntactic structure and semantic relationship information. For example, just as articles and paragraphs have topic sections and sentences, individual clauses contribute differently to overall sentence sentiment orientation, yet our study does not differentiate these contributions. We also do not consider changes in part-of-speech for words in subject-predicate-object and sentence core structures after sentiment transfer. Additionally, our method relies on LTP analysis results. Although LTP demonstrates outstanding performance among existing systems, its accuracy still has room for improvement. We believe that as this technology further develops, better results can be achieved.

Future research will focus on the sentiment weights of different clauses in financial forum text and the establishment of deeper-level sentiment computation rules. Additionally, the application of sentiment analysis technology remains our research interest.

References

[1] Liu B. Sentiment Analysis and Opinion Mining [M]. California: Morgan and Claypool Publishers, 2012.
[2] Smailović J, Grčar M, Lavrač N, et al. Stream-based Active Learning for Sentiment Analysis in the Financial Domain [J]. Information Sciences, 2014, 285: 181-203.
[3] Van de Kauter M, Breesch D, Hoste V. Fine-grained Analysis of Explicit and Implicit Sentiment in Financial News Articles [J]. Expert Systems with Applications, 2015, 42(11): 4999-5010.
[4] Hagenau M, Liebmann M, Neumann D. Automated News Reading: Stock Price Prediction Based on Financial News Using Context-capturing Features[J]. Decision Support Systems, 2013, 55(3): 685-697.
[5] Kiritchenko S, Zhu X, Mohammad S M. Sentiment Analysis of Short Informal Texts [J]. Journal of Artificial Intelligence Research, 2014, 50: 723-762.
[6] Hu Yongjun, Jiang Jiaxin, Chang Huiyou. A New Method of Keywords Extraction for Chinese Short-text Classification [J]. New Technology of Library and Information Service, 2013(6): 42-48.
[7] Duan Jiangjiao, Liu Hongzhong, Zeng Jianping. A Study of the Influence Between Investor Sentiment Index, Analyst Recommendation Index and Stock Index Return —–Based on Eastmoney.com and Sina Analyst Shares Rating Data[J]. Shanghai Finance, 2014 (11): 60-64.
[8] Lin Bingcan. Statistical Studies of Network Public Opinion' s Investor Sentiment' s Impact on the Stock Price[D]. Chengdu: Southwestern University of Finance and Economics, 2013.
[9] Chen Jiangpeng. A Study of China' s Stock Market Effectiveness Testing Based on the Network Public Opinion [D]. Chengdu: Southwestern University of Finance and Economics, 2013.
[10] Liu Dingping. Empirical Research on the Volatility of Stock Price Brought by Investor' s Sentiment in the Emergency Environment[D]. Chengdu: Southwestern University of Finance and Economics, 2014.
[11] Zhang Shijun, Cheng Guosheng, Cai Jihua, et al. Stock Price Prediction Base on Network Public Opinion and Support Vector Machine [J]. Mathematics

in Practice and Theory, 2013, 43(24): 33-40.
[12] Song Minjing. Stock Prediction Model Based on Sentiment Analysis Research[D]. Harbin: Harbin Institute of Technology, 2013.
[13] Jin Xuejun, Zhu Yu, Yang Xiaolan. Effects of Online Media on Stock Market: An Empirical Study on Eastmoney.com[J]. Journalism & Communication, 2013(12): 36-51.
[14] Shen Hanbin. The Effect of Investor Home Attention on Stock Return -A Study Based on Stock Message Boards with Text Mining Techniques [D]. Hangzhou: Zhejiang University, 2014.
[15] Xia Mengnan, Du Yongping, Zuo Benxin. Micro-blog Opinion Analysis Based on Syntactic Dependency and Feature Combination [J]. Journal of Shandong University: Natural Science, 2014, 49(11): 22-30.
[16] Zhang Qingqing, Liu Xilin. Sentiment Analysis Based on Dependency Syntactic Relation [J]. Computer Engineering and Applications, 2015, 51(22): 28-32.
[17] Nakagawa T, Inui K, Kurohashi S. Dependency Tree-based Sentiment Classification Using CRFs with Hidden Variables [C]. In: Proceedings of the 11th Annual Conference of the North American Chapter the Association for Computational Linguistics. Stroudsburg, USA: ACL, 2010: 786-794.
[18] Xiao Hong, Xu Shaohua. Analysis on Web Public Opinion Orientation Based on Syntactic Parsing and Emotional Dictionary [J]. Journal of Chinese Computer Systems, 2014, 35(4): 811-813.
[19] Turney P D, Littman M L. Measuring Praise and Criticism: Inference of Semantic Orientation from Association [J]. ACM Transactions on Information Systems, 2003, 21(4): 315-346.
[20] Yao Tianfang, Lou Decheng. Research on Semantic Orientation Analysis for Topics in Chinese Sentences [J]. Journal of Chinese Information Processing, 2007, 21(5): 73-79.
[21] Che W X, Li Z H, Liu T. LTP: A Chinese Language Technology Platform [C]. In: Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010), Beijing, China.2010: 13-16.
[22] Liu Ting, Ma Jinshan. Theories and Methods of Chinese Automatic Syntactic Parsing: A Critical Survey[J]. Contemporary Linguistics, 2009, 11(2): 100-112.
[23] Wan Changxuan, Jiang Tengjiao, Zhong Minjuan, et al. Sentiment Computing of Web Financial Information Based on the Part-of-Speech Tagging and Dependency Parsing[J]. Journal of Computer Research and Development, 2013, 50(12): 2554-2569.
[24] NLPIR/ICTCLAS Chinese Segmentation System [EB/OL]. [2013-07-02]. http://ictclas.nlpir.org/.
[25] Cui H, Mittal V, Datar M. Comparative Experiments on Sentiment Classification for Online Product Reviews[C]. In: Proceedings of the 21st National Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2006: 1265-1270.

Author Contributions

Lan Qiujun: Conceived the research idea, designed the study, and revised the final manuscript.
Liu Wenxing, Li Weikang: Conducted literature review.
Liu Wenxing, Li Weikang, Hu Xingye: Collected, cleaned, and analyzed data.
Liu Wenxing: Designed programs and drafted the manuscript.

Conflict of Interest Statement

All authors declare no conflict of interest.

Supporting Data

Supporting data is available in the online version of the journal at http://www.infotech.ac.cn.
[1] Lan Qiujun, Liu Wenxing, Li Weikang. dic.txt. Sentiment dictionary generated using SO-PMI and manual processing.
[2] Lan Qiujun, Liu Wenxing, Li Weikang. TestData.rar. Manually annotated test data.
[3] Lan Qiujun, Liu Wenxing, Li Weikang. Dependency relation part-of-speech combination corpora.rar. Corpus for analyzing part-of-speech combination frequencies.
[4] Lan Qiujun, Liu Wenxing, Li Weikang. parsingCount.xls. Frequencies of different part-of-speech combinations in each dependency relation.

Received: 2015-10-14
Revised: 2016-01-10

Note: Figure translations are in progress. See original paper for figures.