

Corporate social responsibility information disclosure objective or not

Companies
based on text
mining

An empirical research of Chinese listed companies based on text mining

519

Zhao Duan

CSR Research Center, Huazhong Normal University, Wuhan, China, and

Yajuan He and Yuan Zhong

School of Information Management, Huazhong Normal University, Wuhan, China

Abstract

Purpose – Based on the text mining tools, this paper aims to propose a new method to evaluate the subjectivity and objectivity of corporate social responsibility information disclosure.

Design/methodology/approach – The authors build up a text subjectivity evaluation model of corporate social responsibility reports through meta-analysis; a text mining is conducted to all sample CSR reports released by Chinese listed companies until March 2016[1]. Furthermore, the authors made an overall and quantitative analysis of the situation which contained changing state, characteristics and abnormal value on the subjectivity and objectivity of information disclosure.

Findings – The results show that the subjectivity scores of social responsibility reports of Chinese listed companies are generally in a normal distribution. The diagram turns out to be a rising trend over the years and increases linearly from 2011 to 2013. Also, the industry heterogeneity and policy control are the main reasons for the formation of the differences, which are significant between different industries and different years.

Originality/value – This paper provides not only an important empirical basis for the research of corporate social responsibility but also a new idea for the non-financial information disclosure as well as objective evaluation of normative text.

Keywords Text mining, Corporate social responsibility report, Objectivity evaluation

Paper type Research paper

Introduction

The corporate social responsibility reports[2], as a major vehicle for the disclosure of non-financial information, provide important information on sustainable development capability of companies to stakeholders through the description of labor practice, environmental protection, customer relations, philanthropy and charity, corporate governance and other topics. With the emergence of socially responsible investment (SRI), securities markets in many countries, including China, have successively launched the corporate social responsibility index. Disclosure of social responsibility information not only plays an important role in the discovery of market prices and the optimization of asset allocation (Cao *et al.*, 2012; Tao and Jin, 2013; Meng *et al.*, 2010; Ling and Sultana, 2015) but also can strengthen the management level within the company, enhance financial stability, improve corporate image, acquire social prestige and reduce operational risks (Li and Peng, 2010; Zhu *et al.*, 2014; Tian *et al.*, 2014; McDonnell and King, 2013).



At present, the large-scale comparison of vertical and horizontal descriptive information of different companies is still a problem to be resolved due to the lack of a consensus on motivation, optimality criteria and economic consequences of social responsibility information disclosure in theoretical research and practice. Moreover, opportunistic disclosers possibly disguise or obscure their performance with the manipulation and impression management by means of readability, rhetoric, topic, typography structure, performance attribution, etc. (Merkel-Davies and Brennan, 2007; Li, 2008) and therefore, the function of corporate social responsibility information as an “indicator” of future financial performance is weakened. In this regard, the situation is even more serious for the securities market of China, which needs further improvement. In recent years, although the awareness of corporate social responsibility in China has been gradually strengthened and the form and content of disclosure have been somewhat improved, the information disclosure is incomplete and reports only the good news, not the bad news, indicates general deficiencies and improvements without going into details, and pays more attention to the form than the essence, so it is difficult to have a substantial impact on stakeholders’ evaluation and decision. (Qi *et al.*, 2016) The quality of Chinese social responsibility information disclosure still have long way to go comparing with the international practice (KPMG China, 2016).

In response to above problems, some researchers propose an effective approach by strengthening the inspection and evaluation of corporate social responsibility reports based on the formation of a set of systematic, quantifiable and comparable methods and standards, so as to promptly identify various disclosure manipulations, effectively improve the quality of information disclosure and enhance the completeness and symmetry of market information (Gallego-álvarez and Quina-Custodio, 2016; Sun, 2012; Wang *et al.*, 2013). In practice, the world trend is also moving toward CSR attestation (Shen, 2007), which imposes a new requirement for the innovation of inspection and evaluation methods. Currently, common approach generally include content analysis, grounded research, reputation index or KLD index, etc. For example, Preston (Preston, 1978) and Campopiano (Campopiano and De Massis, 2015) applied the content analysis for scoring specific items in reports and measuring the corporate social responsibility status; Song Xianzhong (Song and Gong, 2007) and Koladkiewicz (Koladkiewicz, 2009) used the grounded theory to summarize the useful information in corporate reports and come to valuable conclusions; Cochran *et al.* (Cochran and Wood, 1984) used the social reputation of companies to indirectly evaluate their social responsibility; Scholtens *et al.* (Scholtens *et al.*, 2008) made use of the KLD index to measure the corporate social responsibility. Although based on a series of social responsibility standards in their application, these methods may not be consistent regarding the calculation of horizontal and vertical indicators. The case judgment based on the expert evaluation is closely related to the attitudes and positions of evaluators. Generally, there are no consistent standards, but problems such as lack of comparability, strong subjectivity and difficulty in the validity and reliability testing (Huang *et al.*, 2009; Chen and Guan, 2015). Meanwhile, without a holistic analysis with objective and consistent standards, it is difficult to reflect the general trend, characteristics and changes of social responsibility information disclosure, thus affecting the inspection and evaluation functions and the pertinence of the formulation of relevant regulatory policies.

In view of deficiencies in existing approaches, this paper focuses on the subjectivity and objectivity of corporate social responsibility reports. Based on literature review, we build up a text subjectivity evaluation model of corporate social responsibility reports through meta-analysis; a text mining is conducted to all sample reports released by Chinese listed companies (as of March 2016) for a quantitative analysis of the subjectivity of corporate social responsibility reports released by listed companies of all industries and in these years;

and we also give relevant policy suggestions. By a horizontal and vertical comparison, inference of the subjective intentions and the latent sentiment of report preparers, this study can not only provide a relatively objective, systematic, comparable and quantifiable evaluation support system for evaluators in carrying out the expert evaluation but also be conducive to the prompt and effective recognition of descriptive manipulations. What's more, this paper empirically reveals the overall situation, changing trend and industry features of the subjectivity of corporate social responsibility reports of Chinese listed companies, and regulatory effects of relevant policies on the social responsibility information disclosure, thereby providing the important empirical evidence for related theoretical research.

2. CSR report subjectivity recognition model

2.1 Literature review

The subjectivity recognition of report texts mainly aims at preventing opportunistic behaviors of companies and the market by taking advantage of information asymmetry. Related studies pointed out the motivations for corporate social responsibility behaviors usually include altruistic motivation (Zhu *et al.*, 2014; Campbell *et al.*, 1999), strategic motivation (Porter and Kramer, 2002), political motivation (Li *et al.*, 2011; Liang *et al.*, 2010; Zhang *et al.*, 2013) and manager self-interest motivation (Galaskiewicz, 1997). Obviously, according to the latter three motivations, companies may exaggerate good news and avoid adverse information for acquiring resources, enhancing their position or improving their image in the short term (Li *et al.*, 2013); even if some important issues related to social responsibility cannot be concealed, the description may be tactful to increase semantic ambiguity, thereby influencing the decision of stakeholders, that is, enhancing the semantic ambiguity by subjective descriptions and reducing the quality of information disclosure. Consequently, even the accuracy and completeness of disclosed information cannot be directly recognized through the subjectivity and objectivity analysis of texts; they can be indirectly manifested in the degree of subjectivity of texts. Our study also finds that many reports with high subjectivity scores, especially samples within the interval of abnormal value, often have serious accuracy and completeness problems. Moreover, the vertical and horizontal comparisons based on years and industries are useful for determining the quality of the report information disclosure.

Subjective texts generally express the intention, sentiment and attitude, but objective texts are usually factual descriptions based on relations between entities (Yao *et al.*, 2008). These two types of texts are significantly different in content and form. Recently, with increasing research on the opinion mining under the big data environment, the subjective and objective text classification has become a hot topic in the natural language processing (Kim and Hovy, 2004). Some researchers tend to recognize the subjectivity and objectivity of texts by using different proportions of subjective and objective texts after text classification (Ye *et al.*, 2007; Lin and Lin, 2009).

The subjective and objective text classification is generally achieved by machine learning. Wiebe selected the part of speech, punctuation and sentence position in the text as the features, classified subjective and objective texts with the Naive Bayes classifier and obtained good classification results (Wiebe, 1994). Later, Nigam chose word counts as the feature and used the maximum entropy and naive Bayes to improve the text classification method (Nigam, 1999); Yu and Hatzivassiloglou applied three sentence-level statistical techniques for subjective and objective text classification, among which the Naive Bayes classifier has a high recall and precision (80-90 per cent) (H Y, V H, 2003). In the past two years, Jiang applied the LDA machine learning model to recognize subjective sentences.

This method could simultaneously consider the influence of multiple topics on subjective expression recognition task (Jiang, 2014); Cheng Xianyi recognized opinion sentences by the machine learning algorithm by integrating the idea of maximum flow and minimum cut of the graph theory and significantly improved the classification performance and stability of opinion sentence (Cheng *et al.*, 2009); Song Hongwei presented a Chinese subjective and objective sentence classification system based on the fuzzy inference machine and conducted experiments on the NTCIR-6 data, and the results demonstrated the feasibility of the method (Song *et al.*, 2015). On the whole, the classification rules generated by machine learning are easy to understand and can improve text classification through constant training, with a high accuracy. However, machine learning requires a large number of training samples, which must be established manually after reading them one by one.

To solve the problem of an insufficient number of training samples, some researchers attempt to define subjective text features based on the semantic study. For example, Yao Tianfang *et al.* acquired language rules for subjective and objective sentences through manual definition based on syntactic and language analysis (Yao and Lou, 2007); Karamibekr established rules based on subjective verbs and extracted triples with subjective expression from texts for recognizing subjective sentence (Glass, 1976). After Ye Qiang proposed an approach to measure the subjective strength of sentences using patterns of continuous two words combination (2-POS model) (Ye *et al.*, 2007), this model was further applied by many researchers. For example, Lin Shiping extracted ten 2-POS subjective patterns for recognizing subjective information in text and verified the feasibility of this method (Lin and Lin, 2009); Zhang Wenwen classified network texts by six features, such as 2-POS, and extracted subjective sentence from texts, and the final classification results showed the classification based on 2-POS model had good effects (Zhang and Wang, 2013). In general, for the Chinese text analysis, classification methods based on the 2-POS model achieved good results.

With the establishment and implementation of international and domestic corporate social responsibility standards and regulations, corporate social responsibility reports show strong standardization and formatting features in texts, which are quite different from common text analysis objects, such as user reviews, microblog, news and similar types of samples. Meanwhile, prior to use, the existing subjective and objective text classification methods require a large number of training samples to train the models, and manual annotation of numerous texts after reading them one by one for collecting the training sample set. However, until now, China has a very limited number of social responsibility reports of listed companies as our samples. We only collect over 3,000 CSR reports released publicly in 2015 and before, and the report words are subtle without obvious sentiment features. Although we tried to use the above machine learning algorithms to extract subjective and objective classification features of these reports at the early stage, the results of text classification using extracted feature items are not satisfactory. To achieve an automatic classification of subjective and objective report texts, this paper uses the text representation model to process the text information in these reports. After comparison and analysis of several different text representation models, the subjectivity judgment based on 2-POS model is finally selected.

2.2 Subjectivity classification based on 2-POS patterns

For reducing the influence of subjectivity, partiality and text heterogeneity during the model selection, we use the effect size in the meta-analysis for a reanalysis of collected and sorted subjectivity models and their weights. Meta-analysis, also known as secondary analysis, according to Glass, is a statistical analysis that combines the results of multiple scientific

studies of the same type; (Ye *et al.*, 2014) Ye Shuhang argued that the meta-analysis has no unique correct method because the most suitable is the most correct (Cui and Ning, 2010). The meta-analysis can maximize the scientificity, accuracy and objectivity of results, reduce errors possibly occurring in a single study and effectively overcome inconsistency of results across studies (Bruce, 1999).

Based on the literature review, we collected 2-POS subjective patterns and weights from 20 articles of CNIK journals, master and doctoral theses and dissertations and conference papers published from 2007 to 2014, including 9 master and doctoral theses and dissertations and 11 articles from journals, such as *Journal of Management Sciences in China*, *China Journal of Information Systems* and *Application Research of Computers* (for more details, please refer to Annex A). The research corpuses of these articles are collected from the internet, involving comments, social news and economic areas. It is found that the effective subjective patterns are basically consistent. Furthermore, the effective 2-POS subjective patterns are all calculated by the chi-squared formula and are sorted by chi-squared values. The precision of these 2-POS subjective patterns in the corpus is uniformly selected as the subjective weight of the 2-POS model.

To make them more reliable, this paper quantifies and standardizes these research results in all articles containing Chinese 2-POS patterns by using different values of effect size and according to the different fields which the corpus belongs to; the low effect value 0.2 is assigned to the results with big difference in the text natures of corporate social responsibility reports, and the high effect value 0.8 is assigned to the result with a similar nature of report texts; these subject patterns are screened according to the features of corporate social responsibility reports, weighted average by weight, ranked and manually intervened, and finally form a set of effective subject patterns and weights (Table I).

For each 2-POS item in a sentence, if there is a subjective pattern that effectively expresses sentiment (i.e. 11 patterns listed in Table I), the subjective weight is added to the total score of the sentence, and other 2-POS patterns are assigned the value 0. This can be denoted as: $w(U_j) = \sum w(i)$, where $w(U_j)$ is the subjectivity score of the sentence (U_j), and $w(i)$ is the weight of the effective 2-POS subjective pattern of the sentence. A threshold is set for scoring. If the final score of the sentence is larger than the value, it is a subjective sentence; otherwise, it is an objective sentence.

2.3 Validity test of subjective patterns

To ensure that subjective patterns can be applied to CSR reports of different discourse types, we used the stratified sampling method during the validity test. First, we conducted a

No.	Effective 2-POS subjective patterns	Initial word	Last word	Subjective weight
1	d-a	adverb	adjective	0.811
2	r-d	pronoun	adverb	0.712
3	v-r	verb	pronoun	0.676
4	d-d	adverb	adverb	0.605
5	d-v	adverb	verb	0.560
6	a-udel	adjective	udel	0.538
7	v-d	verb	adverb	0.530
8	v-v	verb	verb	0.494
9	r-v	pronoun	verb	0.481
10	v-a	verb	adjective	0.425
11	r-n	pronoun	noun	0.229

Table I.
Effective subjective
patterns and
subjective weights

preliminary sampling, randomly selected 100 reports from all samples and put them into five categories according to sample presentation and distribution features; then, four reports are randomly selected from each category (20 reports in total) and stratified 20 texts by topic and style features; finally, a number of report fragments are randomly selected from each layer to form the testing corpus after combination and adjustment, including a total of 1,633 natural sentences. Next, based on the subjectivity annotation of sentences proposed by Bruce (Waddock and Graves, 1997), we designed an annotation description for this study. In view of the goal of this paper, we simulated the situations by expert attestation. Four people with relevant professional backgrounds were independently invited to annotate the selected testing corpus and further had discussions for achieving consensus on controversial and uncertain sentences. After comparison and judgment, they deleted 193 sentences that cannot reach consensus. At last, they annotated 944 subjective sentences and 496 objective sentences. Meanwhile, the Python software is used for annotation participles and part-of-speech and filtering stop words, and then sentences are separated by the paragraph mark and stored. With the subject pattern and weight table and formula, the software is used to code and automatically calculate the subjectivity score of each sentence of annotated samples. The subjective-objective situation calculated with the subjective patterns is compared with the manual annotation.

After analysis and judgment, when the threshold is set as 0, subjective sentences have the recall and precision higher than 90 per cent, which proves the effectiveness of extracted 2-POS subjective patterns. Therefore, this article uses 11 subjective patterns as the effective patterns to recognize the subjectivity of corporate social responsibility reports and defines sentences with the subjectivity score of 0 as objective sentences, and those with a subjectivity score not equal to 0 as subjective sentences. Considering different lengths of these reports, based on the total subjectivity score S_i , the subjective sentence count $Scount_i$, and the objective sentence count $Ocount_i$, calculated for each report, we divided the total score by the total number of sentences of each report, so as to standardize their subjectivity score. The mean subjectivity score of each report is $A_i = \frac{S_i}{Scount_i + Ocount_i}$. Finally, we recognize the subjectivity-objectivity of a report by the value of this variable A_i .

3. Empirical analysis and results discussion

3.1 Samples and data

After determining the subjectivity recognition model, we collected all samples made up of 3,211 corporate social responsibility reports disclosed by Chinese listed companies between 2009 and 2015 from “cninfo.com.cn”, the website designated by China Securities Regulatory Commission for information disclosure. The reason why we selected samples since 2009 was that Shenzhen Stock Exchange promulgated the Social Responsibility Instructions to Listed Companies in September 2006 and encouraged listed companies to periodically disclose their social responsibility situation and existing problems as required by the Instructions. Then, a small number of listed companies began to voluntarily release corporate social responsibility reports according to relevant policies. Besides, in May 2008, Shanghai Stock Exchange (SSE) issued the Notice on Strengthening Listed Companies’ Assumption of Social Responsibility, which mandated that listed companies must regularly disclose their CSR reports at the SSE website. As a result, the disclosure of CSR reports by listed companies multiplied since 2009. Moreover, only 96 CSR reports were disclosed by listed companies between 2006 and 2008. The sample size was too small, so we chose 2009 as the beginning year of our study and analysis. The year 2015 was selected as the ending year because our collection expired in March 2016 when only a few companies issued their CSR reports. In China, listed companies generally made a summary and prepared a report in the

middle of the following year by considering the corporate social responsibility situation in the previous year.

According to the Guidelines for the Industrial Classification of Listed Companies formulated by China Securities Regulatory Commission in the fourth quarter of 2015, we first classified all collected reports into 18 categories by industries and further eliminated environmental reports with similar content and industries with a very small sample size. After screening, we finally obtained 2,970 effective samples within ten industries in seven years from 2009 to 2015. Then, for all samples in different years and industries, the subjectivity recognition model proposed above was used for annotation participles and part-of-speech and filtering stop words, and sentences of each report were separated by paragraph marks; next, the subjectivity score of each report was automatically calculated with the subjective pattern and weight table and formula; at last, the mean values of subjectivity scores of all reports were stored in an Excel spreadsheet by industries and years.

3.2 Descriptive statistics

According to the available data, descriptive statistics tools of SPSS20.0 are used to analyze mean subjectivity scores of reports in different industries and years. The subjectivity scores of reports of different industries are presented in [Table II](#). It can be found that the most of the disclosed reports are concentrated in the Manufacture, accounting for 58.42 per cent of total samples, followed by the Finance accounting for 6.90 per cent, while the Farming, Forestry, Animal Husbandry and Fishery only accounts for 1.31 per cent, reflecting a significant difference in the number of reports of different industries. The Finance gets the highest mean subjective score 1.7410, which is higher than 1.6087, the mean score of all samples. The Transport, Storage and Postal Services also have a high score, 1.6956. The Construction has the lowest subjectivity score, 1.4710, and the Production and Supply of Electric Power, Heat Power, Gas and Water also gets a relatively low score, 1.4739. Their scores are much lower than the mean of total samples. The above analysis shows that there are also significant differences in the subjectivity scores across different industries.

The subjectivity scores by years are presented in [Table III](#) and [Figure 1](#). [Table III](#) shows that the reports are released mostly in 2015 and least in 2009 and suggest an increasing trend year over year. The mean subjectivity score of reports is the highest in 2013, being 1.6333, which is higher than 1.6087, the mean score of all samples, while the lowest mean score is 1.5618 in 2011. This indicates that there are some differences in the subjectivity scores in different years.

Internally, one possible explanation for the year-over-year increase of CSR reports release by listed companies is that companies expect to gain the trust of the government and external parties by assuming social responsibility and bring benefits to their future management; or that the management has the awareness that undertaking social responsibility can enhance the corporate social image and social status, ([Waddock and Graves, 1997](#)) and reflect the market performance in the period, thereby forming a positive feedback effect; externally, more negative news exposures of corporate social responsibility in recent years, demonstration of excellent companies, guide by the government policy files and active promotion by industry associations may increase the emphasis of more companies on the social responsibility information disclosure. Besides, the number of corporate social responsibility reports released after 2009 shows an explosive growth, which is possibly the result of the guidelines issued by Shanghai Stock Exchange in 2008 with the assistance of GRI, ([Shen et al., 2016](#)) and companies' fear of falling behind their competitors.

Table II.
Mean subjectivity
score of reports of
different industries

	<i>N</i>	Mean	SD	SE	95 % mean confidence interval		Minimum	Maximum
					Lower limit	Upper limit		
Farming, forestry, animal husbandry and fishery	39	1.4807	0.2211	0.0354	1.4091	1.5524	1.0839	1.9736
Mining	147	1.6802	0.2997	0.0247	1.6314	1.7291	1.0271	2.6308
Manufacture	1735	1.6103	0.2998	0.0072	1.5962	1.6244	0.7652	3.1181
Production and supply of electric power, heat power, gas and water	159	1.4739	0.2980	0.0236	1.4272	1.5206	0.8296	2.3172
Construction	63	1.4710	0.2325	0.0293	1.4125	1.5296	0.9904	2.1182
Wholesale and retail trade	157	1.6135	0.3132	0.0250	1.5641	1.6628	1.0094	3.0561
Transport, storage and postal services	145	1.6956	0.3396	0.0282	1.6399	1.7514	0.7923	2.8803
Information transmission, software and information technology services	129	1.6022	0.2416	0.0213	1.5601	1.6443	1.0468	2.1982
Finance	205	1.7410	0.3631	0.0254	1.6910	1.7910	0.5809	3.4835
Real estate	191	1.5154	0.2763	0.0200	1.4759	1.5548	0.9464	2.6070
Total	2970	1.6087	0.3076	0.0056	1.5976	1.6198	0.5809	3.4835

In Figure 1, the overall trend of subjectivity scores of reports varies significantly with years, and the subjectivity scores of reports published in 2011-2013 show a linear increase.

The descriptive statistics and the statistical graph of mean values can only reflect the final mean subjectivity scores of all reports, but it is impossible to judge how many reports follow the change trend in these years. The study further classifies subjectivity scores of 2,970 reports by company names. There are 736 companies in total. The report scores in the continuous years of each company are compared and the results are shown in Figure 2.

Combining Figures 1 and 2, we can find that the trend of the number of companies with increasing subjectivity scores in the years listed in Figure 2 is basically consistent with that in Figure 1. That is, if the mean subjectivity score of a year increases or decreases compared with the previous year, the number of companies with increasing subjectivity scores also increases or decreases. This finding also verifies that the subjectivity scores of reports of companies match the change of the total mean subjectivity score of all reports.

The linear growth in mean subjectivity scores of reports over the period from 2011 to 2013, and the continuous increase in the number of companies, in our opinion, are the results of the lagging effect of some policies and the enhancement by a series of external policies in this period. On January 9 of 2009, Shanghai Stock Exchange issued a notice on the CSR reporting of listed companies, provided the social responsibility report format and the working papers for directors reviewing the reports and encouraged listed companies to actively disclose their social responsibility reports; in November 2010, the International Organization for Standardization issued the standard ISO26000 Guidance on Social

	N	Mean	SD	SE	95% mean confidence interval		Minimum	Maximum
					Lower limit	Upper limit		
2009	163	1.5667	0.3167	0.0248	1.5177	1.6157	0.9451	3.1181
2010	167	1.5720	0.2953	0.0228	1.5269	1.6171	0.7923	2.6024
2011	414	1.5618	0.3044	0.0150	1.5324	1.5912	0.5809	2.9509
2012	498	1.5972	0.3224	0.0144	1.5688	1.6256	0.8409	3.4835
2013	545	1.6333	0.3068	0.0131	1.6075	1.6591	0.7652	2.7342
2014	587	1.6272	0.2947	0.0122	1.6033	1.6511	0.8423	2.8121
2015	596	1.6320	0.3061	0.0125	1.6074	1.6566	0.8296	3.2660
Total	2970	1.6087	0.3076	0.0056	1.5976	1.6198	0.5809	3.4835

Table III.
Mean subjectivity
score of reports in
different years

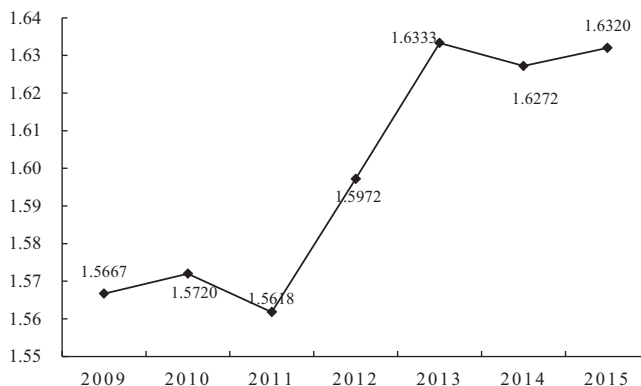


Figure 1.
Mean of CSR report's
subjectivity scores in
different years

Responsibility, and in March 2011, Chinese Academy of Social Sciences released the Guidelines on Preparation of Corporate Social Responsibility Reports by Chinese Companies (CASS-CSR2.0). Because the corporate social responsibility is still under development in China, most companies do not understand the concept and knowledge of corporate social responsibility and sustainable development. They cannot selectively disclose relevant key information from their own perspective and prepare report conforming to their development status and the information requirements of stakeholders, but solely cater for the standards and rigidly follow the existing reporting guidelines and standards. Therefore, with the promulgation of new policies, the subjectivity of reports is different to some extent for catering to the policy. Of course, this means existing policies have exerted a great effect on the implementation.

3.3 Data analysis

Given that the simple comparison of mean values cannot reflect some important results, to more accurately verify the difference in subjectivity scores of CSR reports of different industries and in different years, we conducted a significance test of this variable. Figure 3 presents a normal distribution test histogram of subjectivity scores of reports. The histogram shows that the variable follows a normal distribution and reflects that the subjectivity scores of most reports are at the medium level and only a few mean scores are at the extreme level. Meanwhile, the Levene’s test of homogeneity-of-variance is performed to the data (the table omitted). The results indicate that the homogeneity-of-variance significance Sig is less than 5 per cent and the variance is not homogeneous, but the significance Sig of the mean subjectivity scores of reports in different years is larger than 5 per cent, and the scores are equal variance. The inhomogeneous variance of subjectivity scores of reports in different industries is mainly because of obviously different industry features. The industry belongs to an endogenous variable, and there are big differences in different data groups. The year is an exogenous variable, so there is no essential difference in data and the subjectivity scores of reports in different years are equal variance. This also indicates that the industry is an important variable for the difference in the report subjectivity.

Subjectivity Scores of Reports of Different Industries. Based on results of the normal distribution test and Levene’s test of homogeneity-of-variance above, the Kruskal–Wallis H test on ranks is used for testing the mean subjectivity scores of CSR reports of companies in different industries. The test results are shown in Table IV.

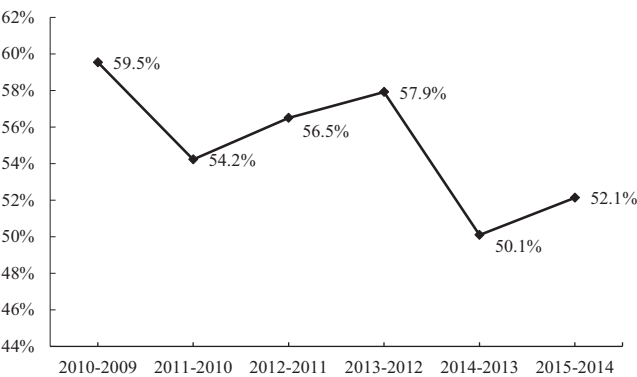


Figure 2.
Proportion trends of
companies with
increased subjectivity
score

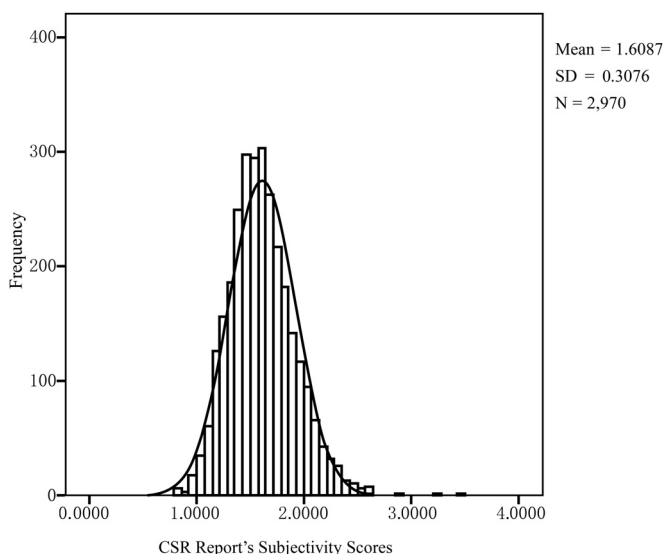


Figure 3.
Histogram of CSR
report's subjectivity
scores

	Mean subjectivity score of report
Chi-Square	938.536
df	9
Asymp.Sig.	0.000

Table IV.
Results of Kruskal–
Wallis H test on
ranks (N = 2970)

As can be seen from [Table IV](#), after the Kruskal–Wallis H test on ranks, the mean subjectivity scores of reports grouped by industries are statistically significant with the Sig value less than 5 per cent. This indicates that there is a significant difference in the mean subjectivity scores of reports of different industries, which is consistent with the results of the descriptive statistics. Through the comparison of samples, this is probably because of the preparers' selection of presentation mode and topic distribution. Under the influence of internal and external factors, companies in the same industry tend to converge and the difference exists in companies of different industries due to the industry features. The presentation mode refers to the way used by preparers to express, such as the use of quantitative mode or qualitative mode; and the topic distribution refers to which topics selected and highlighted by report preparers and how to arrange the text.

Generally, companies prepare reports pursuant to various standards and guidelines. Different companies often use different methods to present the same topic owing to different understandings of the connotation, denotation and topics of corporate social responsibility and the different subjective intentions for disclosure. The companies emphasize relevant topics and increase description according to their own social practice. If a topic is closely related to a company in its business practice, it is easy to find some facts and data and this topic is often presented in a highly objective manner with more description in the text; if a topic is not closely linked to a company or without relevant social practice, the company possibly presents it by semantic transformation or using a fuzzy language, tries to get close to the topic through subjective descriptions or reduces description of this topic. In this case,

the difference in presentation mode and topic distribution of a report is reflected in the subjectivity scoring.

Influenced by external policies, industry features, pressure from peer disclosure and industry imitation, companies in the same industry often share commonalities in presentation mode and topic distribution strategies, but companies in different industries have obvious differences. Since the current guidelines for the preparation of Corporate Social Responsibility Report are applicable to all industries and cover large and complete topics. Some topics are highly relevant to some industries, but others are not. However, according to relevant policies, some topics are mandatorily required to be disclosed in the report, so that some industries are forced to adjust presentation mode and text distribution. This topic pertinence is often related to industry features, thus resulting in the differentiation across industries. For example, the Notice on Strengthening Listed Companies' Assumption of Social Responsibility and on Issuing the Guidelines on Listed Companies' Environmental Information Disclosure issued by Shanghai Stock Exchange in 2008 encourages listed companies to disclose social responsibility reports and stipulates that the report must include "the company's effort in promoting the environmentally and ecologically sustainable development". Because some industries do not have similar practices to the environmental topic and there are no guidelines on preparing social responsibility reports generally applicable across industries, in the face of the disclosure pressure from competitors and to cater for promulgated policies and the pressure of attestation and decrease the search cost incurred to the organization for reducing uncertainty, the discourse strategy of larger, more successful or more prestigious companies in the industry may be imitated, thereby leading to the convergence in presentation in a same industry. Further, after associations, like China Banking Association and China National Textile and Apparel Council, release their social responsibility reports, companies in the same industry prepare their own reports pursuant to or referring to the industry standards, thus further enhancing the difference across industries. An empirical study of Shen Hongtao, Huang Changyin *et al.* also reflected this point to some extent (Huang and Zhang, 2011; Li, 2006).

3.3.2 *Subjectivity Scores of Reports in Different Years.* Mean subjectivity scores of reports in different years are both in a normal distribution and equal variance. ANOVA (analysis of variance) is used in this paper for testing mean subjectivity scores and the results are presented in Table V.

The ANOVA results show that the significance level (Sig.) of *F* value is less than 0.05. This means that significant differences exist in mean subjectivity scores of reports in different years. However, the results of a simple ANOVA cannot clearly indicate in which years there are differences. Thus, after ANOVA, LSD of the post hoc test is further used for a multiple comparison of samples, and the test results are presented in Table VI.

Table VI shows no significant difference in subjectivity scores of reports released in the four years from 2009 to 2012 ($p > 0.05$) and in the four years from 2012 to 2015 ($p > 0.05$), but significant difference in subjectivity scores of reports disclosed in any of the three years from 2009 to 2011 and any year from 2013 to 2015. The data reveal that the year 2012 is an important turning point for the change in subjectivity scores of reports. Through document

Table V.
Analysis results of
mean subjectivity
scozres of reports in
different years

	Sum of square	Df	Mean square	<i>F</i>	Significance
Inter-group	2.342	6	0.390	4.153	0.000
Intra-group	278.489	2963	0.094		
Total	280.831	2969			

(I) Year	(J) Year	Mean difference (I-J)	Standard error	Significance	95% confidence interval	
					Upper limit	Lower limit
2009	2010	-0.0052721	0.0338	0.876	-0.0715	0.0609
	2011	0.0048864	0.0283	0.863	-0.0507	0.0605
	2012	-0.0304618	0.0277	0.271	-0.0847	0.0238
	2013	-0.0666052*	0.0274	0.015	-0.1203	-0.0129
	2014	-0.0604744*	0.0271	0.026	-0.1137	-0.0073
	2015	-0.0652587*	0.0271	0.016	-0.1184	-0.0121
2010	2009	0.0052721	0.0338	0.876	-0.0609	0.0715
	2011	0.0101584	0.0281	0.718	-0.0449	0.0653
	2012	-0.0251897	0.0274	0.358	-0.0789	0.0286
	2013	-0.0613331*	0.0271	0.024	-0.1145	-0.0082
	2014	-0.0552024*	0.0269	0.040	-0.1079	-0.0025
	2015	-0.0599866*	0.0268	0.026	-0.1126	-0.0074
2011	2009	-0.0048864	0.0283	0.863	-0.0605	0.0507
	2010	-0.0101584	0.0281	0.718	-0.0653	0.0449
	2012	-0.0353481	0.0204	0.083	-0.0753	0.0046
	2013	-0.0714916*	0.0200	0.000	-0.1107	-0.0323
	2014	-0.0653608*	0.0197	0.001	-0.1039	-0.0268
	2015	-0.0701450*	0.0196	0.000	-0.1086	-0.0317
2012	2009	0.0304618	0.0277	0.271	-0.0238	0.0847
	2010	0.0251897	0.0274	0.358	-0.0286	0.0789
	2011	0.0353481	0.0204	0.083	-0.0046	0.0753
	2013	-0.0361434	0.0190	0.057	-0.0734	0.0011
	2014	-0.0300127	0.0187	0.108	-0.0666	0.0066
	2015	-0.0347969	0.0186	0.062	-0.0713	0.0017
2013	2009	0.0666052*	0.0274	0.015	0.0129	0.1203
	2010	0.0613331*	0.0271	0.024	0.0082	0.1145
	2011	0.0714916*	0.0200	0.000	0.0323	0.1107
	2012	0.0361434	0.0190	0.057	-0.0011	0.0734
	2014	0.0061308	0.0182	0.737	-0.0296	0.0419
	2015	0.0013466	0.0182	0.941	-0.0343	0.0370
2014	2009	0.0604744*	0.0271	0.026	0.0073	0.1137
	2010	0.0552024*	0.0269	0.040	0.0025	0.1079
	2011	0.0653608*	0.0197	0.001	0.0268	0.1039
	2012	0.0300127	0.0187	0.108	-0.0066	0.0666
	2013	-0.0061308	0.0182	0.737	-0.0419	0.0296
	2015	-0.0047842	0.0178	0.788	-0.0397	0.0302
2015	2009	0.0652587*	0.0271	0.016	0.0121	0.1184
	2010	0.0599866*	0.0268	0.026	0.0074	0.1126
	2011	0.0701450*	0.0196	0.000	0.0317	0.1086
	2012	0.0347969	0.0186	0.062	-0.0017	0.0713
	2013	-0.0013466	0.0182	0.941	-0.0370	0.0343
	2014	0.0047842	0.0178	0.788	-0.0302	0.0397

Note: * The significance level of the mean difference is 0.05

Companies
based on text
mining

531

Table VI.
Multiple comparison
of mean subjectivity
scores of reports in
different years

comparison, the possible reason is that Chinese Academy of Social Sciences released the Guidelines on Preparation of Corporate Social Responsibility Reports by Chinese Companies (CASS-CSR2.0) in 2011, which is very important in China's social responsibility research. This file aimed to provide guidance on the preparation of corporate social responsibility reports. China Corporate Social Responsibility Report (2015) also pointed out that the most of companies referred to these guidelines for preparing their reports. Therefore, it is

reasonable to take 2012 as a dividing line. After that year, the content of most CSR reports of China undergoes major changes and the subjectivity changes accordingly, thus resulting in a significant difference in subjectivity scores of reports released between continuous years from 2009 to 2011 and from 2013 to 2015.

3.3.3 Abnormality in Subjectivity Scoring. Considering the significant difference in subjectivity scores of reports of different industries and released in different years, the boxplot is used for quickly identify abnormal values in subjectivity scores. Abnormal values refer to too large or too small values in the observation data. Figure 4 presents the boxplot of subjectivity scores of reports in different industries. (Note: In the boxplot, abnormal values are denoted by “O” and extreme values denoted by “*”).

The boxplot shows that only Information Transmission, Software and Information Technology Services has no abnormal values, which means that CSR reports in this industry achieve a balance between subjectivity and objectivity. In the industry of Manufacture, CSR reports corresponding to maxima numbered 536 and 538 are those disclosed by First Automobile Workshop in 2009 and 2011. The report released in 2009 has 16 pages and the mean subjectivity score of 3.1181, and the report in 2010 has 31 pages and the mean subjectivity score of 2.9509. The scores are much higher than 1.6103, the mean score of the Manufacture. The human review finds that although they disclose some specific data on corporate social responsibility, subjective texts in the speech of general manager and company overview account for much of the reports, so there is an abnormality in their mean subjectivity score. Likewise, the CSR report numbered 2813 in the Finance industry is a report released by China Pacific Insurance in 2012. It has 81 pages and the mean subjectivity score of 3.4835, which is much higher than 1.7410, the mean score of the industry. After removing a large number of pictures, these reports have a very large proportion of subjective texts, so the subjectivity score is abnormal.

In general, the analysis of large sample data with the boxplot can quickly recognize suspected abnormality in mean subjectivity scores of reports of different industries and in different years. A further analysis of the content of reports with abnormal scores can

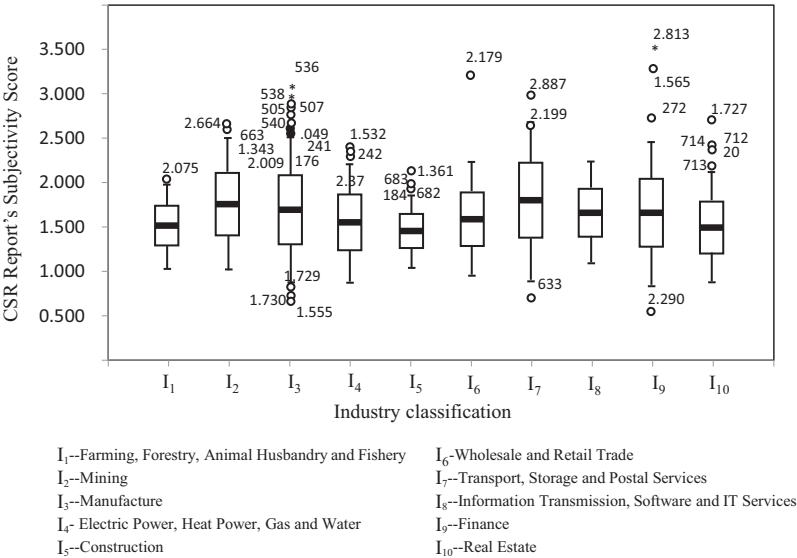


Figure 4.
Boxplot of CSR
report's subjectivity
score in different
industries

provide targeted guidance to companies and a basis for regulation of disclosure manipulating and appropriate intervention.

3.4 Verification of applicability and stability

The expert scoring method is used to verify the applicability and stability of the subjectivity evaluation system proposed in this paper. Expert scoring is measured by the seven-point Likert scale, in which 7 means “very strong subjectivity” and 1 means “very weak subjectivity”. In total, 2,970 reports are divided into seven intervals according to the normal distribution. The specific intervals of mean subjectivity scores, subjectivity strength of report and seven-point Likert scale are shown in [Table VII](#).

We chose seven reports corresponding to seven intervals above for the evaluation experiment of CSR reports, including five reports released after March 2016, namely, China Pacific Insurance Corporate (ZGTB) Social Responsibility Report 2015 (2.2178), Jiangnan Water Corporate (JNSW) Social Responsibility Report 2015 (1.9541), Shanghai AJ Group Corporate (AJJT) Social Responsibility Report 2015 (1.5511), Zhejiang Juhua (JHGF) Corporate Social Responsibility Report 2015 (1.4797) and Huaneng Power International (HNGJ) Corporate Social Responsibility Report 2015 (1.1474). Besides, to verify whether the method in this paper could effectively identify reports within intervals of abnormal values, we selected Xiandai Investment (XDTZ) Corporate Social Responsibility Report 2010 with the subjectivity score of 0.8113, as well as China Merchants Bank (ZSYH) Corporate Social Responsibility Report 2014 for manual intervention, in which the most of objective descriptions are obscured so that the subjectivity score reaches 3.2660.

Then we contacted 12 experts who are experienced in text analysis and divided them into three groups based on their years of service t. Experts in Group 1 have more than eight years’ work experience, experts in Group 2 have three to eight years’ work experience and those in Group 3 have less than three years’ work experience. Experimental group and control group were set up under each group. These experts scored the subjectivity of seven reports according to their own knowledge and experience. Before scoring, experts in experimental groups were informed of mean subjectivity scores of reports calculated in this paper (as shown in [Table VII](#)); in contrast, no information was provided to experts in control groups, who gave scores directly.

After the first round of scoring, the results of the three groups of experts were analyzed and summarized (as shown in [Table VIII](#) below), and the comprehensive results were fed back to them. Then, experts revised their scoring and began the next round. After three rounds of feedback and revisions, the scoring results of the three groups experts were basically consistent. The final scoring results are shown in [Table IX](#). To test the consistency

Interval of mean subjectivity score	Subjective strength of report	Seven-point Likert scale
$A_i < 0.9$	Very weak	1
$0.9 \leq A_i < 1.2$	Weak	2
$1.2 \leq A_i < 1.5$	Relatively weak	3
$1.5 \leq A_i < 1.8$	Medium	4
$1.8 \leq A_i < 2.1$	Relatively strong	5
$2.1 \leq A_i < 2.4$	Strong	6
$A_i \geq 2.4$	Very strong	7

Note: The table above is obtained by dividing normal distribution results of subjectivity scores of 2970 reports into seven intervals

Table VII.
Mean subjectivity
scores

of the final scoring results of experts, we compared and analyzed final scores of different experts, and the test results showed that there was no significant difference between the scores of different experts.

In Table VIII, six experts of the experimental groups are consistent in their subjectivity scoring of reports in different groups, while six experts of the control groups give quite different scores. This indicates that the reference results provided in this paper can influence expert scoring to some extent. Results in Table IX show that the subjectivity score results of experimental groups and control groups converge within the group and that the results of control groups deviate toward the results of experimental groups. This further means that the subjectivity scores calculated in this paper play a considerate guiding role for expert decision. In addition, the comprehensive scores of 12 experts in Table IX calculated by the weighted evaluation calculation of the expert scoring method, and the seven-point Likert scale corresponding to mean subjectivity scores of reports are presented in Table X.

In Table X, the two groups total expert score results are basically consistent with the results by direct judgment of subjectivity scores of reports. After special treatment, the corporate social responsibility report of China Merchants Bank gets the highest subjectivity score. This means the evaluation system can provide effective decision support for expert review and verify the applicability and stability of the evaluation method.

Table VIII.
First-round expert
scoring

CSR report	Group 1 ($t \geq 8$)				Group 2 ($3 \leq t < 8$)				Group 3 ($0 < t < 3$)			
	Experimental group		Control group		Experimental group		Control group		Experimental group		Control group	
	Expert	Expert	Expert	Expert	Expert	Expert	Expert	Expert	Expert	Expert	Expert	Expert
	1	2	3	4	5	6	7	8	9	10	11	12
ZSYH	6	7	6	4	6	6	4	6	7	6	5	3
ZGTB	6	5	5	3	6	6	3	5	5	6	3	6
JNSW	5	4	4	2	4	5	3	6	5	5	5	3
AJJT	4	4	3	5	5	4	5	2	3	4	3	5
JHGF	4	3	2	4	3	3	2	5	3	4	4	2
HNGJ	2	3	4	1	3	2	4	2	2	3	2	4
XDTZ	2	2	2	4	1	2	3	2	2	1	2	3

Table IX.
Final expert scoring

CSR report	Group 1 ($t \geq 8$)				Group 2 ($3 \leq t < 8$)				Group 3 ($0 < t < 3$)			
	Experimental group		Control group		Experimental group		Control group		Experimental group		Control group	
	Expert	Expert	Expert	Expert	Expert	Expert	Expert	Expert	Expert	Expert	Expert	Expert
	1	2	3	4	5	6	7	8	9	10	11	12
ZSYH	7	7	7	6	7	6	7	6	7	7	6	6
ZGTB	6	5	6	6	6	6	5	6	5	6	6	5
JNSW	5	4	4	5	4	5	5	5	5	5	5	4
AJJT	4	4	5	4	4	4	4	5	4	4	4	5
JHGF	4	3	3	4	3	3	3	4	3	4	4	3
HNGJ	2	3	2	2	2	2	3	2	2	2	2	3
XDTZ	2	1	1	2	1	2	1	1	2	1	1	2

4. Conclusions and outlook

4.1 Conclusions and suggestions

With the continuous development of the social and economic environment, the usefulness of corporate social responsibility information of listed companies has been widely recognized by the academic world and the industry. The effective decision-making of the stakeholders depends not only on the availability of high-quality disclosure information, but also increasingly on the comparison of horizontal and vertical descriptive information of different companies within a large range. Meanwhile, the improvement of information disclosure, as the basis for the existence and development of securities market, relies on timely and effective regulations. Therefore, an overall and dynamic grasp of the disclosure of social responsibility information is of great significance. Based on existing methods, considering the in-depth text mining of corporate social responsibility reports, on the one hand, it can make up the deficiencies of limited evaluation objects caused by subjectivity scoring, importance attached only to the quantity of information, subjectivity noises and heavy workload; on the other hand, it can avoid the problems of limitations of non-financial indicators and lack of comparability.

From the perspective of preventing opportunistic behaviors of companies, we focus on the subjectivity and objectivity analysis of the texts of corporate social responsibility reports. Through an analysis of all samples of CSR reports released by Chinese listed companies from 2009 to 2015, we have important findings in the description of changing state by industries and years, discovery of abnormal values, validity test of policies and standards and empirical test of existing theories. Besides, the method is easy to operate and popularize. If the number of sample reports continues to increase in the future, its validity and reliability will be further enhanced. This provides a new idea for the evaluation of non-financial information of companies and objectively helps to enhance the credibility of evaluation.

Generally, subjectivity scores of corporate social responsibility reports of Chinese listed companies are in normal distribution and show a rising trend year over year; the multiple comparison of samples in these years by LSD of the post hoc test indicates that there is a turning point in the overall changes of the subjectivity of corporate social responsibility reports; and by the ANOVA and the rank test, the analysis results of mean subjectivity scores of reports reveal significant differences in different years and across industries. These features reflect the important influence of external factors such as the promulgation of industry policies, guidelines for preparation of reports in different years and new policies promulgated by the state. Moreover, the results of the homogeneity-of-variance analysis also demonstrate that the industry is a critical variable for the difference in the report subjectivity. The empirical results adequately support related studies undertaken by Li Zheng (2006), [Holder-Webb and Cohen \(2009\)](#), [Lori et al. \(2009\)](#) and [Song and Sheng \(2009\)](#). Consequently, three suggestions are made: first, relevant social responsibility information disclosure policies and guidelines need further

CSR report	Comprehensive score of experts	Direct subjectivity scores
China Merchants Bank	6.583	7
China Pacific Insurance	5.667	6
Jiangnan Water	4.667	5
Shanghai AJ Group	4.250	4
Zhejiang Juhua	3.417	3
Huaneng Power International	2.250	2
Xiandai Investment	1.417	1

Table X.
Subjectivity score
comparison table

revision and improvement to purposively prevent increasing subjectivity; second, a recognition and intervention mechanism is established for violations and manipulations in the corporate social responsibility information disclosure; third, competent industry authorities are encouraged to promulgate their respective preparation guidelines, implementation rules and evaluation standards based on existing reporting guidelines and standards and considering industry features and differences.

4.2 Deficiencies and outlook

The evaluation method proposed in this paper mainly aims at the text subjectivity evaluation. This method can reflect the intention of discloser to some extent, but it is impossible to directly recognize the accuracy and completeness of disclosed information. To solve this problem, our future study will carry out a quantifiable systematic analysis of corporate social responsibility reports by using topic extraction, sentiment analysis and data mining. The subjectivity recognition model is obtained by the meta-analysis, thereby subjectivity features of report texts may not be fully measured due to the representativeness and timeliness of collected documents. Although this error is consistent for all reports and belongs to a systematic bias having a marginal influence on the overall trend and conclusions of comparison and analysis, there may be situation different from expert opinion in the large-scale practical application, which has an impact on the convergence of expert opinion. For this problem, with the increase of samples in the future, the subjectivity recognition model can be extracted by machine learning so as to further enhance the objectivity and reliability of the evaluation. Unfortunately, this paper only analyzes Chinese texts from the corporate social responsibility reports of Chinese listed companies. If a further comparison can be made across countries and different company types, this will be more theoretically significant. Breakthroughs are expected in further studies.

Notes

1. The years mentioned in this paper are those when the CSR reports are disclosed.
2. CSR reports mentioned in this paper include Corporate Social Responsibility Report and Sustainability Report disclosed by companies.

References

- Bruce, R.E.F. (1999), "Recognizing subjectivity: a case study of manual annotation", *Natural Language Engineering*, Vol. 1 No. 1, pp. 1-16.
- Campbell, L., Gulas, C.S. and Gruca, T.S. (1999), "Corporate giving behavior and decision-maker social consciousness", *Journal of Business Ethics*, Vol. 19 No. 4, pp. 375-383.
- Campopiano, G. and De Massis, A. (2015), "Corporate social responsibility reporting: a content analysis in family and non-family firms", *Journal of Business Ethics*, Vol. 129 No. 3, pp. 511-534.
- Cao, Y., Wang, J. and Yu, L. (2012), "An empirical study of corporate social responsibility information disclosure and investment efficiency", *Management World*, Vol. 12, pp. 183-185.
- Chen, X. and Guan, C. (2015), "Problem analysis in 'social responsibility report of publishing companies' in recent years", *Modern Publishing*, Vol. 2, pp. 21-23.
- Cheng, X., Yang, T. and Zhu, Q. (2009), "Research of text filtering based on semantic orientation", *Application Research of Computers*, Vol. 12, pp. 4460-4462, *et al.*
- Cochran, L.F. and Wood, R.A. (1984), "Corporate social responsibility and financial performance", *Academy of Management Journal*, Vol. 27 No. 1, pp. 42-56.

- Cui, Z. and Ning, Z. (2010), "Quantitative literature review and meta-analysis", *Statistics and Decision*, Vol. 19, pp. 166-168.
- Galaskiewicz, J. (1997), "An urban grants economy revisited: corporate charitable contributions in the twin cities, 1979-81, 1987-89", *Administrative Science Quarterly*, Vol. 42 No. 3, pp. 445-471.
- Gallego-álvarez, I. and Quina-Custodio, I.A. (2016), "Disclosure of corporate social responsibility information and explanatory factors", *Online Information Review*, Vol. 40 No. 2, pp. 218-238.
- Glass (1976), "Primary G V secondary and meta-analysis of research", *Education Research*, Vol. 5 No. 6.
- H Y, V H (2003), "Towards Answering Opinion Questions: separating facts from opinions and identifying the polarity of opinion sentences", *Proc. of the 2003 Conf. on Empirical Methods in Natural Language Processing*, pp. 129-136.
- Holder-Webb, L. and Cohen, J. (2009), "The supply of corporate social responsibility disclosures among US Firms", *Journal of Business Ethics*, Vol. 84 No. 4, pp. 497-527.
- Huang, Q., Peng, H., Zhong, H. and Zhang, E. (2009), "Evaluating the level of responsibility management and CSR information disclosure of top 100 companies in China", *China Industrial Economics*, Vol. 10, pp. 23-35.
- Huang, C. and Zhang, T. (2011), "Difference among industries of voluntary information disclosure of listed companies—an quantitative perspective based on XBRL taxonomy", *Securities Market Herald*, Vol. 7, pp. 56-61.
- Jiang, W. (2014), "Study on identification of subjective sentences in product reviews based on weekly supervised topic model", *Journal of Software*, Vol. 7 No. 9, pp. 1952-1959.
- Kim, S. and Hovy, E. (2004), "determining the sentiment of opinions", *Proceedings of the COLING Conference*, pp. 1367-1373.
- Koladkiewicz, I. (2009), "Corporate social responsibility in Poland: the responsible business forum 2002-2007 report perspective", *Social Responsibility Journal*, Vol. 5 No. 1, pp. 48-61.
- KPMG China (2016), *KPMG Survey of Corporate Responsibility Reporting 2015 – Currents of Change*,
- Li, F. (2008), "Annual report readability, current earnings, and earnings persistence", *Journal of Accounting and Economics*, Vol. 45 No. 2-3, pp. 221-247.
- Li, Z. (2006), "A study on relation of corporate social responsibility and corporate value: empirical evidence from shanghai securities exchange", *China Industrial Economy*, Vol. 2, pp. 77-83.
- L. P. Analyzing Corporate Social Performance: methods and results (1978), *Journal of Contemporary Business*, Vol. 1 No. 7.
- Liang, J., Chen, S. and Qing'en, G. (2010), "Political participation, governance structure and charity donation", *Management World*, Vol. 7, pp. 109-118.
- Li, Z., Guan, F. and Li, Z. (2013), "An empirical research on drivers of corporate social responsibility report attestation—empirical Evidence from chinese listed companies", *Auditing Research*, Vol. 3, pp. 102-112.
- Li, B., Gu, H. and Gao, W. (2011), "An empirical study on the effect of institutional pressures on corporate social responsibility of tourism enterprises", *Nankai Business Review*, Vol. 14 No. 6, pp. 67-75.
- Lin, H. and Lin, S. (2009), "Research on extracting subjective sentence in Chinese sentiment orientation analysis", *20th National Computer Technology and Application Academic Conference (CACIS-2009) and 1st National Key Technology and Application Academic Conference*, p. 6.
- Ling, T.C. and Sultana, N. (2015), "Corporate social responsibility: what motivates management to disclose?", *Social Responsibility Journal*, Vol. 11 No. 3, pp. 513-534.
- Li, X. and Peng, H. (2010), *An Empirical Study on the Relationship between Corporate Social Responsibility Information Disclosure and Corporate Reputation. Reform of Economic System*, Vol. 3, pp. 74-76.
- Mcdonnell, H.M. and King, B. (2013), "Keeping up appearances: reputation threat and prosocial responses to social movement boycotts", *Administrative Science Quarterly*, Vol. 58 No. 3, pp. 387-419.

- Meng, X., Xiao, Z. and Qu, J. (2010), "Interactive relation between corporate social responsibility disclosure and the cost of capital: an analysis framework based on asymmetric information", *Accounting Research*, Vol. 9, pp. 25-29.
- Merkel-Davies, D.M. and Brennan, N.M. (2007), "Discretionary disclosure strategies in corporate narratives: incremental information or impression management?", *Journal of Accounting Literature*, Vol. 26, pp. 116-196.
- Nigam, K. (1999), "Using maximum entropy for text classification", in *IJCAI-99 Workshop on Machine Learning for Information Filtering*, pp. 61-67.
- Porter, M. and Kramer, M. (2002), "The competitive advantage of corporate philanthropy", *Harvard Business Review*, Vol. 80 No. 12, pp. 56-68.
- Qi, L., Zhang, B. and Li, T. (2016), "A research on the quality evaluation of corporate social responsibility report", *Science Research Management*, Vol. S1, pp. 644-651.
- Scholtens, B., Cerin, P. and Hassel, L. (2008), "Sustainable development and socially responsible finance and investing", *Sustainable Development*, Vol. 16 No. 3, pp. 137-140.
- Shen, H. (2007), "Corporate features and corporate social responsibility information disclosure—empirical evidence from Chinese listed companies", *Accounting Research*, Vol. 3, pp. 9-16.
- Shen, H., Chen, T. and Huang, N. (2016), "Reluctance or willingness: event history analysis of social responsibility report attestation decision", *Accounting Research*, Vol. 3, pp. 79-86.
- Song, X. and Gong, M. (2007), "Evaluation of social responsibility information quality and decision value—content analysis of annual accounting reports of listed companies", *Accounting Research*, Vol. 2, pp. 37-43.
- Song, J. and Sheng, C. (2009), "A research on corporate social responsibility measurement in listed manufacturing companies: a stakeholder theory perspective", *China Soft Science*, Vol. 10, pp. 153-163.
- Song, H., Song, J. and Fu, G. (2015), "Chinese subjective sentence recognition based on fuzzy inference machine", *Journal of Chinese Information Processing*, Vol. 5, pp. 160-166.
- Sun, Y. (2012), "The effect of corporate responsibility report's disclosure clarity and attestation on investment decisions—an experimental research", *Auditing Research*, Vol. 4, pp. 97-104.
- Tao, W. and Jin, Z. (2013), "The relationship between corporate social responsibility disclosure and corporate financial performance under media exposure and its implications: an empirical study based on the CSR report of Chinese listed firms", *Chinese Journal of Management Science*, Vol. 4, pp. 162-170.
- Tian, M., Li, C. and Keng, S. (2014), "The impact of corporate social responsibility activities on brand evaluation", *Nankai Business Review*, Vol. 17 No. 6, pp. 19-29.
- Waddock, A.S. and Graves, S.B. (1997), "The corporate social performance-financial performance link", *Strategic Management Journal*, Vol. 18 No. 4, pp. 303-319.
- Wang, J., Wang, Q. and Jia, W. (2013), "The determinants of corporate social responsibility information disclosure quality: based on Chinese finance industry", *Statistics and Information Forum*, Vol. 3, pp. 43-48.
- Wiebe, J.M. (1994), "Tracking point of view in narrative", *Computational Linguistic*, Vol. 2 No. 20, pp. 233-287.
- Yao, T. and Lou, D. (2007), "Research on semantic orientation analysis for topics in chinese sentences", *Journal of Chinese Information Processing*, Vol. 5 No. 21, pp. 73-79.
- Yao, T., Cheng, X., Xu, F., Uszkoreit, H. and Wang, R. (2008), "A survey of opinion mining for texts", *Journal of Chinese Information Processing*, Vol. 3, pp. 71-80.
- Ye, S., Guo, D. and Ge, H. (2014), "The influencing factors of external knowledge transfer of transformation business—based on meta-analysis technology", *Studies in Science of Science*, Vol. 6, pp. 909-918.

-
- Ye, Q., Zhang, Z. and Law, R. (2007), "Automatically measuring subjectivity of Chinese sentences for sentiment analysis to reviews on the internet", *China Journal of Information Systems*, Vol. 1, pp. 79-91.
- Zhang, M., Mi, L. and Zhang, W. (2013), "Government-enterprise bonding effect of corporate charity donation—empirical evidence based on chinese listed companies", *Management World*, Vol. 7, pp. 163-171.
- Zhang, W. and Wang, T. (2013), "Unsupervised subjective sentence extraction for non-standard texts", *Computer and Digital Engineering*, Vol. 1, pp. 64-68.
- Zhu, H., Tu, R. and Jin, F. (2014), "Are good deeds to be sustained? A study on the change of consumers' attitudes towards firms after they quit social responsibility", *Nankai Business Review*, Vol. 17 No. 6, pp. 4-12.

Further reading

- Cao, L., Qu, X. and Zhang, D. (2011), "Corporate social responsibility reporting and accounting information quality—an empirical study on listed companies in Shenzhen", *Collected Essays on Finance and Economics*, Vol. 3, pp. 99-105.
- Gao, Y., He, X. and Li, L. (2011), "Business owners' social status, economic achievement and corporate philanthropy", *Economic Research Journal*, Vol. 12, pp. 111-123.
- Jia, X. and Liu, Y. (2014), "External environment, internal resource, and corporate social responsibility", *Nankai Business Review*, Vol. 17 No. 6, pp. 13-18.
- Karamibekr, M. and Ghorbani, A.A. (2013), *Sentence Subjectivity Analysis in Social Domains. Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*.
- Li, S. and Xie, X. (2014), "Corporate social responsibility, political relationship and debt financing of private enterprises: evidence from Chinese capital market", *Nankai Business Review*, Vol. 17 No. 6, pp. 30-40.

Corresponding author

Zhao Duan can be contacted at: duanzhao@mail.ccnu.edu.cn

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.