Time Series Analysis on the Determinants of Environmental Costs Expenditure Using Text Mining Technique

Toshiyuki MAEDA*, Naoya KAWAKAMI*, Yoshimi CHUJO*, and Eunjee PARK†

* Faculty of Management Information, Hannan University, Japan

† Faculty of Economics, Kagawa University, Japan

***Abstract*—This study aims to read from environmental report which corporate management policy and strategy would promote the motivation of corporate behavior, especially the environment-related activities by listed companies. To do that, we analyze the relation between qualitative data and quantitative description using text mining technique. Our study aims to derive the result of more detailed analysis to confirm the change of positiveness factor by analyzing with text mining technique such as corresponding analysis, for environmental report in 2000-2011 in Japan. Time series analysis reveals that as CSR concept is widely disseminated, the focus of massage shifts from pollution aids to more comprehensive activities, which implicitly indicates that the environmental report is currently recognized as a useful tool to effectively communicate a firm's social activities.**

Keywords: Environmental report; Corporate Social Responsibility; Text mining; Corresponding analysis

I. Introduction

This study aims to read from environmental report which corporate management policy and strategy would promote the motivation of corporate behavior, especially the environment-related activities by listed companies. To do that, we analyze the relation between qualitative data and quantitative description using text mining technique.

These days, disclosure of social-related information by company has been expanding by dissemination of Corporate Social Responsibility (CSR). In other hand, the method of text mining in various researches has spread as main tool of information gathering. However, this text mining has hardly been applied in management study fields except marketing.

Shirata et al. analyzed the annual securities report qualitatively and descriptively, and tried to read the type of corporate behavior from text data. In general, many studies which have analyzed the annual securities report are interested in quantitative data, the financial information included. Her study used the text mining with several kinds of indices to extract keywords to characterize the bankrupt company and going concern [1]. The results from her analysis provide a new clue for management studies in Japan that they are able to analyze the dynamism of corporate behavior in a different viewpoint.

To retain and improve various society and environment-related activities is becoming a factor to gain an advantage of the corporate competitiveness. This attribute will be reflected in the information of company's disclosure [2]. For example, Kitora [3] analyzed the descriptive answer of CSR policy in "the CSR Corporate Survey 2006" of Toyo Keizai Shinpo Sya with the text mining.

In contrast, Murai et al. even analyzed the factor to define the positiveness. They regarded the amount that companies pay for the environment conservation cost as a proxy of the positiveness for environmental activities in Japanese companies. They extracted directly keywords which support the positiveness from the description of environmental report, and then analyzed with the environmental information [4]. To develop further, our study aims to derive the result of more detailed analysis to confirm the change of positiveness factor by analyzing with text mining technique for environmental report in 2000-2011 in Japan.

II. Environmental report

A. Definition of the environmental report

Ministry of the Environment in Japan defines as follows in the Environmental Reporting Guidelines (2012 Version) [5];

Environmental reporting allows enterprises to fulfill their accountability to society as businesses that use natural resources, to provide stakeholders with useful information that may affect their judgment, and to promote environmental communication with them.

These days, environmental report is becoming a basic tool in environmental communication between company and stakeholders including consumer. Many companies disclose "the sustainability report," or "the social and environmental report," or "the CSR report" except "the environmental report" which describes initiatives intended to fulfill CSR. We study to regard them as environmental report because any report used for environmental reporting, regardless of name, is regarded as "an environmental report" in the guidelines [5].

The main contents of the environmental report are always "environment issues" , but they also add economic and social aspects related the environment in the 2012 version. Our study

omits explanations for particular items because the subject is not all of them. The environmental accounting information of our study's subject is in "the Economic Contexts of environmentally focused management" in the new guideline.

B. Environmental accounting

Ministry of the Environment in Japan defines as follows in the Environmental Accounting Guidelines (2005 Version) [6];

Environmental accounting aims at achieving sustainable development, maintaining a favorable relationship with the community, and pursuing effective and efficient environmental conservation activities. These accounting procedures allow a company to identify the cost of environmental conservation during the normal course of business, identify benefit gained from such activities, and provide the best possible means of quantitative measurement (in monetary value or physical units) and support the communication of its results.

In other words, environmental accounting is the system to measure how much capital investment and what kind of the environmental conservation activities, and provide them to inter- and intra-company stakeholders. Even if the inter-company stakeholders can read the environmental accounting and will understand the situation of environmental conservation activities of the company. These days, many companies in Japan adapt this environmental accounting system.

The reasons why companies introduce the environmental accounting and disclose the information are as follows. First, companies need to grasp and evaluate cost vs benefit in aspects of environmental activities for themselves. These can be by using in monetary value or physical units for environmental conservation activities, environmental conservation benefit and economic benefit. Second, companies have an accountability widely for society in aspects of CSR, so companies disclose the information by environmental accounting as one of methods to accomplish the accountability. Third, the society with comprised various stakeholders asks the environment-related information. Consumers who are interested in the environmental issue would positively buy the environmentally focused products and services. Investors who are interested in the environmental issues would positively have one of information when they choice the invested company. Like these, companies and stakeholders would need to use the environmental accounting [7].

The contents of environmental accounting information in environmental report are comprised of "the Environmental Conservation Cost" , "the Environmental Conservation Benefit" , and "the Environmental Benefit Associated with Environmental Conservation Activities" . Our study target that companies pay "the Environmental Conservation Cost" of them. The environmental conservation cost is divided into the investment and costs, measured in monetary value, allocated for the prevention, reduction, and/or avoidance of environmental impact, removal of such impact, restoration following the occurrence of a disaster, and other activities [6].

III. Research methodology

A. Research target

In this study, we use environmental reports published in 2010, which are based on the period from April in 2009 to March in 2010. We especially focus on "top messages" in environmental reports as the targets of text mining analysis.

"Top messages" present swearness of CEO to consumers and stake-holders, and concentrated Those imply that contents of environmental reports

Among the companies which present environmental reports for 12 years on Web, we sorted companies which make groups more than five companies per one group.

We thus have ten groups such as home electric industry, precise industry, chemical industry, transportation industry, communication industry, control industry, other electric industry, construction industry, food industry, metal industry.

We collect the environmental report data from various Web sites such as corporation Web pages linked from METI (Ministry of Economy, Trade and Industry), Internet Archives, and so on.

B. Research method

We extract the top messages of the environmental reports mentioned above, and make groups from those reports with three high ranks and low ranks at the environmental cost ratio with the sales amounts.

We furthermore divide into three periods of each four-year time so that it is easy to capture the trend of changes, and then analyses with respect to the change of the top messages for 12 years.

The reason we analyze the ratio of environmental cost in the sales, is that analysis should be independent with company size, and should find the characteristics of the environmental cost ratio for each industry.

In addition, the reason for the split 12 years into three periods by four years for the analysis of time series is that the number of documents for each period is easily comparable because of the same amount of documents, and the number of groups is not so much for the analysis of the time series variation.

Extracted data from the top message are grouped and normalized as follows.

1) Cost ratio is it be classified in descending order and small order in three companies group of every industry. Each group is named as "upper group" and "lower group" respectively.
2) 12 years are divided into the period of 2000 - 2003, 2004 - 2007, and 2008 - 2011 by the year of report publication with the above groups. Each period is named as "former period", "middle period", and "later period" respectively.
3) Morphemes extracted from the top message are limited only to nouns (except proper noun, person's name, and place name) and adjectives.
4) The word usage frequency for each report is normalized with 200 words, and the top 50 words are extracted by

usage frequency, though some keywords, which do not make sense for analysis purpose such as "year", "billion" and so on, are omitted.

5) Above extracted Keywords are checked, and the keywords that are used in more than one group are picked up.

After above arrangement, the corresponding analysis are performed to confirm visually the relationship between the keywords and the frequency of respective group.

In addition, the analysis for multi-sector is performed as described above, and then groups are re-summarized based on the per-cost ratio. After that the correspondence analysis are performed in the same manner as above.

To enable to compare numerically with the environmental conservation costs of the sample companies, we calculate the "cost rate (%) = [environmental conservation costs] / [sales] * 100".

This allows us to compare how much the environmental conservation cost is spent by companies comparing to the percentage expenditure to sales.

C. Analysis method

First of all, sentences must be split into words to analyze with the text mining techniques. If words are separated by white space as English, it is quite easy. However, we do not write Japanese sentences leaving some space between words or grammatical units so that it is very difficult to separate the words correctly.

Therefore, we use a technique called morphological analysis in order to extract words accurately from the sentence in Japanese. The morpheme is the “the smallest semantic unit” in terminology used in linguistics [8]. Accuracy to split from the sentence to the morphemes (keywords) is achieved over 90 percent in the current state of the art.

D. Analysis environment

We introduce “RMecab”, “R” with “Mecab” module. “R” is a programming language and its developing environment born in 1992, and that has graphics making functions and statistical techniques. In this study, "MeCab" is used for morphological analysis, because word or phrase sets are said to be better than letters to extract semantic information in Japanese.

IV. Results and Discussion

A. Comparison of EC ratio

Table I exhibits the average proportion of environmental costs expenditure to sales revenue (hereafter EC ratio) for each industry. The mean EC ratio across industries is 0.90%, varying from 0.34% (precision equipment) to 1.87% (construction). It is argued that over half of the sample firms expend over time for the environmental conservation by as much as 1% of the amount of sales.

B. Frequency of key words

Typically, the result of text mining is accompanied with the frequency of key words gauged by the morphological analysis. Focusing on those key words which appear many times in the text usually provides a practical measure with which to know the general characteristics describing the overall text. Hence, this study initially classifies the sample firms of each industry into 3 groups, which consist of top (HIGH), medium (MID) and bottom (LOW) 3 firms in terms of EC ratio, to investigate whether there is any relationship between environmental costs expenditure and the word choices. As is explained in the previous section, the sample period (2000-2011) is divided threefold: first (2000-2003), second (2004-2007), and third (2008-2011) period. Given the 3 firm groups and the 3 time periods, this study normalizes the morphemes as well as the key words extracted from the management message section included in the environmental report. Then, it compares the common characteristics of the words or morphemes across these groups.

C. Correspondence analysis of industry specific words

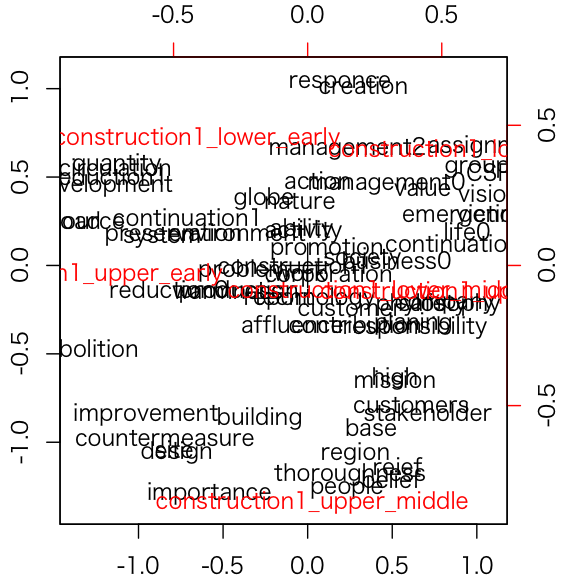

Figure 1: Correspondence analysis of Construction.

Figure 1 shows the correspondence analysis of construction industry. Construction is located in the left-hand side for the first period and in the right-hand side for the second and third period, indicating that the emphasis of the top message has significantly changed between the periods. The HIGH group appears for the first time near an array of words such as ’reduction of environmental load’ or ’reduction of wastes’, while the Great East Japan Earthquake and subsequent economic stagnation likely make this group more involved in the CSR activities. The LOW group also recognized issues regarding the pollution from early stages, and started actions

Type	Construction	Chemistry	Transport	Iron & Metal	Home Electric	Food
Proportion(percent)	1.87	1.56	1.49	0.81	0.70	0.64

Type	Control Machinery	Communication	Precision Machinery	Others	All
Proportion	0.53	0.49	0.34	0.52	0.90

Table I: Average proportion of environmental costs expenditure.

to resolve the issues by adopting more efficient environmental techniques, but due to the occurrence of the earthquake such activities come to be largely constrained. For transportation device, subsequent to the first time period, there arises a significant difference between the EC ratio groups. Especially, the HIGH group consistently stresses the importance of the 'product concepts' such as 'diesel engine' or 'truck', which results in the nontrivial distance from the LOW group.

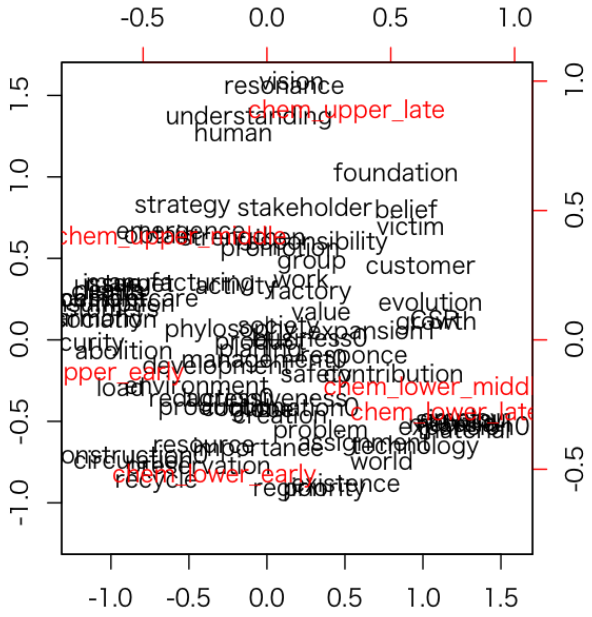

Figure 2: Correspondence analysis of chemical industry.

Figure 2 shows the correspondence analysis of chemical industry. That indicates a disperse deployment of characteristic words between time periods. From the second period, the HIGH group begins to depart from the remaining two groups. It is suggested from the words that this group tries to shift its focus toward a new direction following the occurrence of the great earthquake. Likewise, in the iron non-ferrous industry, the HIGH group initially promotes activities such as 'energy-saving' or 'recycling', whereas switches the target into 'CSR activities' from the second period. On the other hand, in the latter periods, the LOW group is surrounded by those words, the contexts of which are somewhat inconsistent. Consumer electronics is initially centered with how to resolve the environmental problems, but from the second period, there arises a significant difference between the HIGH and LOW group: the former heads toward 'innovation' and the latter toward 'CSR activities'.

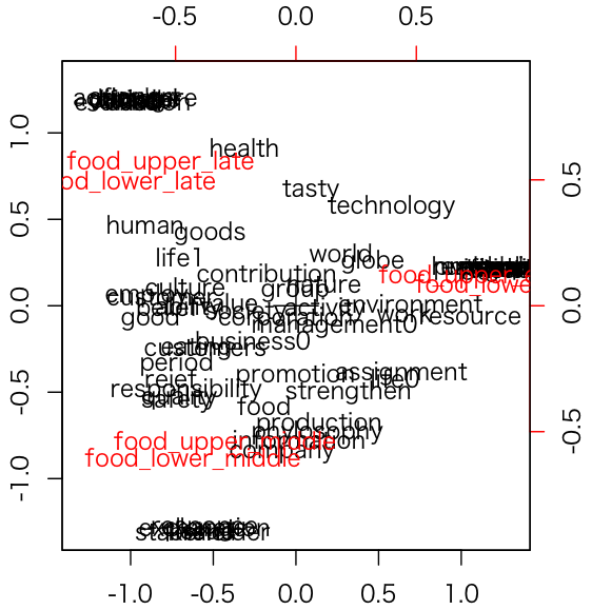

Figure 3: Correspondence analysis of food industry.

Figure 3 shows the correspondence analysis of food industry. Food would be a peculiar case in that there is no identifiable difference except for the inter-temporal one. Alternatively, whether to spend for the environment does not matter for food firms, and common targets such as 'responsibility' (second period) and 'consumers' health' (third period) are shared. Anyhow, the words seem to commonly represent the strategic characteristics of this industry.

Figure 4 shows the correspondence analysis of control equipment industry. As for controlled equipment, the words appearing in the third period considerably departs from those characterizing the beginning two periods. Currently the HIGH group mainly advocates innovation and the LOW group is stuck to customer services.

Figure 5 shows the correspondence analysis of communication industry. As for communication, similar to the construction, the distinction of words manifests itself between the first and subsequent time periods but there is little difference between the EC ratio groups. In the latter periods, the disaster recovery actions might become a common objectives to undertake for both HIGH and LOW groups.

Figure 6 shows the correspondence analysis of precision machine industry. As for precision mechanical equipment, the same trend can be observed in that there is a significant

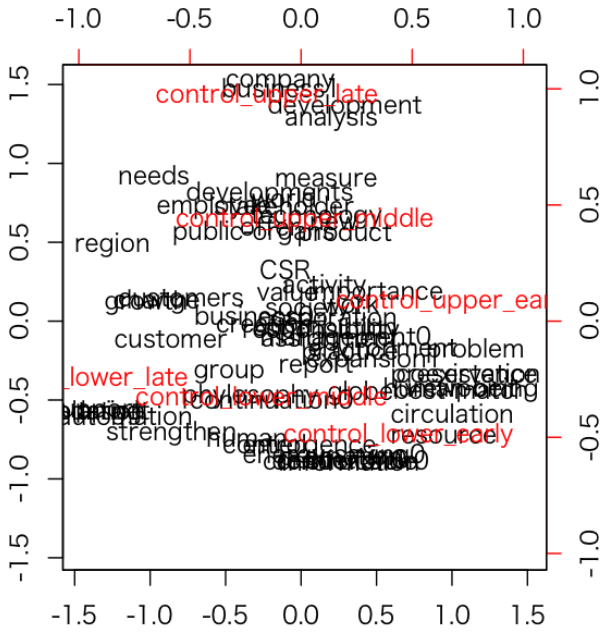

Figure 4: Correspondence analysis of control equipment industry.

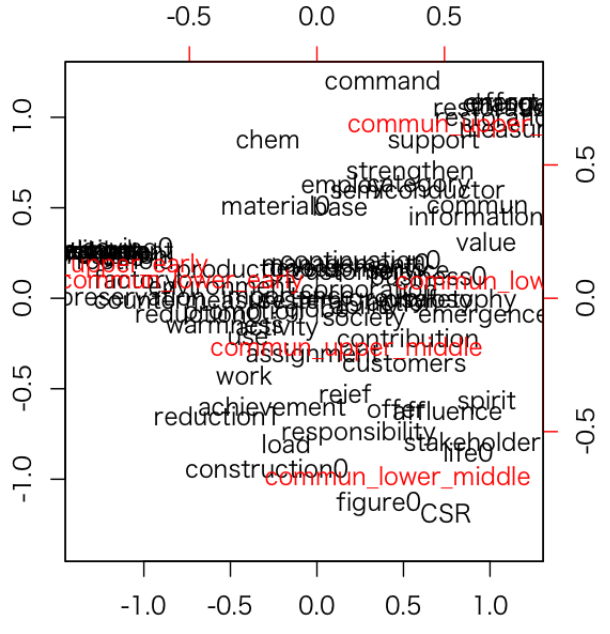

Figure 5: Correspondence analysis of communication industry.

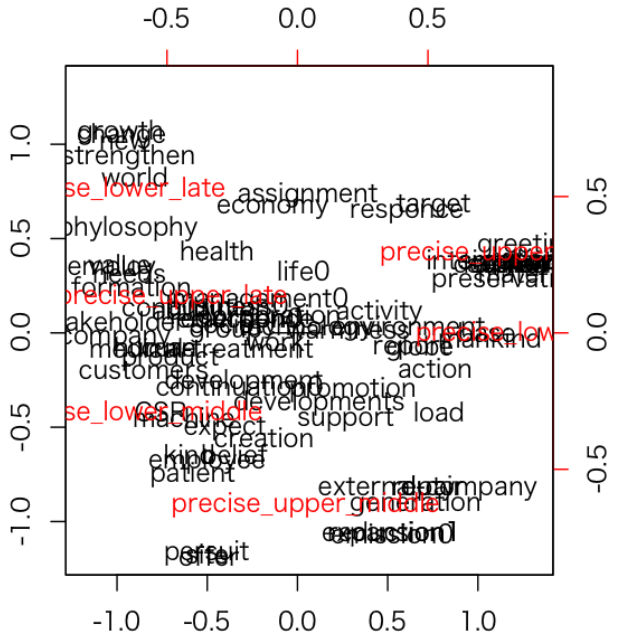

Figure 6: Correspondence analysis of precision machine industry.

divergence between the first and subsequent periods. Simultaneously, even in the latter period, while the HIGH group is located near many words addressing various kind of visions this group encompasses, the LOW group is surrounded by a few words, which indicates that this group is encountered with some difficulties in actively conducting environmental activities since the occurrence of the earthquake.

V. Concluding Remarks

This study applies the text mining to exploring what kind of key words exclusively characterize the top management message attached to the environmental report which Japanese firms have disclosed. Not only addressing the key words, it also visualize the relationship between the words and the firm group based on how much expenditure is devoted for environmental activities.

As Murai et al. suggest, the more a firm expends for environment, the more involved it would be in developing new products and commercializing such products as well as more interested in the CSR activities [4]. Oppositely, the less expenditure for environment means that the firm does not envisage any clear objectives to conduct the environmental conservation and CSR activities.

Time series analysis reveals that as CSR concept is widely disseminated, the focus of massage shifts from pollution aids to more comprehensive activities, which implicitly indicates that the environmental report is currently recognized as a useful tool to effectively communicate a firm's social activities. In addition, despite the variability of targets in developing new products across industries, Japanese firms become more

sensitive regarding the extent to which they should understand the needs of their surrounding society as well as they can receive the approval from the community. In this sense, this study gives some realities to CSR literature by linking the amount of environmental cost expenditure to the text with which management speaks to not only shareholders but also communities.

The limitations of the study should also be noted. The sample included in the analysis consists of only ten industries which are deemed to be substantially subject to environmental regulations, so the results derived above do not suffice to address a general conclusion about the motivations for Japanese firms to invest more in environmental activities. Further, this study deals with the environmental report available on line: those reports undisclosed via website are hardly acquired. A potential sample selection bias might work toward favoring those key words reported in the previous section. The future study would be directed to incorporate the missing samples and elaborate on the results presented in this study.

ACKNOWLEDGMENT

Part of this research was supported by JSPS KAKENHI Grant Number 15K03802. This research was also partially conducted under a sponsorship of Grant of Institute of Industrial and Economic Research, Hannan University. The authors greatly appreciates the supports.

REFERENCES

[1] Y. Shirata, H. Takeuchi, S. Ogino, and H. Watanabe, "Financial analysis using text mining technique : Empirical analysis of bankrupt companies (in japanese)," *Business Analysis Association Annual Report*, vol. 25, pp. 40–47, 2009.
[2] T. Toyosumi, *Strategic Environment Management*. Chuo Keizaisha, 2007.
[3] Y. Kitora, "Classification of companies based on the basic corporate social responsibility (csr) policy by using the text mining approach (in japanese)," *Journal of Socio-Informatics*, vol. 13, no. 1, pp. 17–29, 2009.
[4] T. Murai, Y. Chujo, E. Park, and T. Maeda, "An analysis of environmental reports using text mining methods [in japanese]," in *Abstracts of Annual Conference of Japan Society for Management Information 2011f*, 2011, pp. F3–2.
[5] G. o. J. Ministry of Environment, "Environmental reporting guidelines (fiscal year 2012 version)," http://www.env.go.jp/en/policy/economy/erg2012.pdf, 2012.
[6] ——, "Environmental accounting guidelines 2005," https://www.env.go.jp/en/policy/ssee/eag05.pdf, 2005.
[7] H. Ishibashi, *Environment and Consumers (in Japanese)*. Keiou Gijuku University Publishing, 2010.
[8] M. Ishida, *Introduction of Text Mining Using R (in Japanese)*. Morikita Shuppan, 2009.