

Evaluation of Total Quality Management Using CSR Company Reports

Shu Ochikubo^(✉), Fumiaki Saitoh, and Syohei Ishizu

Aoyama Gakuin University, Shibuya, Japan
occc.shhh@outlook.jp,
{saitoh, ishizu}@ise.aoyama.ac.jp

Abstract. In recent years, serious quality accidents and quality troubles such as recalls have occurred frequently, and Total Quality Management (TQM) is important as effective management to prevent quality troubles beforehand. There is also a Quality Management Level Research [6] (TQM research) that is jointly implemented by the Union of Japanese Scientists and Engineers and the Nikkei. In the TQM survey, the rankings on the six criteria of quality management and comprehensive rankings have been announced. However, concrete TQM activities have not been announced. Meanwhile, Corporate Social Responsibility (CSR) report has published abundant descriptions about the role of customers, employees, society and management who are stakeholders of companies. In this research, we aim to evaluate and extract the characteristics of the company's quality management activities according to the six criteria of the TQM survey using corporate CSR reports.

Keywords: TQM · PLSA · MDS · Word colud

1 Introduction

Serious quality problems have occurred in recent years: complaints about products and services and product recalls may in fact pose a critical risk to the sustainability of affected companies. Total Quality Management (TQM) is important for preventing this problem and refers to “a process, an organization, a person, and a system being combined organically, and company management being conducted from the viewpoint of customers.” Because quality problems are ultimately caused by problems with the company systems, TQM is important for their prevention and is also relevant in the context of corporate social responsibility. A research of the extent to which TQM is employed by Japanese companies was jointly administered by the Union of Japanese Scientists and Engineers and the Nikkei. The evaluation of the research was based on responses related to the six criteria shown in Table 1.

Using the TQM research, a ranking of companies according to these six criteria was published, revealing which companies are excellent in each criterion. The research responses were not published, however, and we had difficulty understanding the beneficial practices of the highly ranked companies based on the results. Furthermore, many studies on TQM concern the introduction of TQM to a company, such as Kamio [1], while there are few studies of concrete TQM plans already in use at companies.

Table 1. Six criteria evaluated in the research

Customer orientation	Establishment and compliance of the processes
Top management commitment	Personnel training for TQM
Management of relief, safety and trust	Utilization and deployment power of the systems

On the other hand, most companies publish Corporate Social Responsibility (CSR) reports in order to fulfill their accountability to stakeholders. These CSR reports usually present significant efforts on the part of companies to the stakeholders. The aims of CSR and TQM are different, but their scopes are closely related. Good practices of companies within the scope of TQM may therefore be discovered by making use of this relationship. In this study, we thus use text data from CSR reports that contain information about the quality management of companies and analyze the data according to the TQM criteria in order to identify best TQM practices for each criterion. We use dimension reduction techniques, such as Probabilistic Latent Semantic Analysis (PLSA), which is effective for such hyperspace data.

Focusing on references that companies make to stakeholders, Kitora [5] investigates how each company influences CSR efforts. Many prior studies attempt to quantify corporate characteristics that affect corporate information disclosure. By directly performing text mining, corporate orientation to stakeholders can be grasped directly, and thus qualitative corporate characteristics can also be discerned. Clarifying the influence on information disclosure using regression is advantageous. However, although it is possible to extract the word of the factor influencing the problem, it is impossible to extract the document can be mentioned.

In the research of Saitoh [3], a large sparse matrix can be translated to a small dense matrix using PLSA of a customer review. After this conversion, the arrangement of review sentences is visualized by making a self-organizing map. Because a large sparse matrix can be converted to a small dense matrix using PLSA as in Saitoh's research, it is considered to be very effective for analysis of text in high dimensional data. For nonlinear data, such as a frequency matrix, when visualization is performed by principal component analysis or correspondence analysis and when the cumulative contribution rate to the number of dimensions is low, a lot of information is lost, and only a small amount remains. Although it is necessary to explain the data with information, PLSA, which can approximate the high dimensional semantic space of the original data with a few dimensions, can be effective for the visualization of text data.

2 Analysis Procedure

The overall analysis procedure will first be explained to provide an outline. The Steps 1 to 4 correspond to the interested companies selection, and Steps 5 to 7 correspond to the TQM information extraction.

Step 1. Data collection

From the company ranking produced by the TQM research evaluation, we selected the top ten companies (first to tenth place) and the ten companies in

the 89th to 98th places. Also the CSR reports contain varying amounts of documents, they are divided into two groups according to the quantity of documents. The results of this are shown in Table 1.

Step 2. PLSA

PLSA is effective as a technology for extracting useful knowledge from big data and is often used as a method for high dimensional text data. In this study, this method is used for document evaluation and document extraction.

Step 3. Calculation of company similarity by Cosine similarity and Visualization of similarity using Multi-dimensional Scaling (MDS)

We calculate the cosine similarity to visually grasp similarity in the PLSA results of Step 2. We also visualize similarity with MDS using the result of the cosine similarity.

Step 4. Company evaluation and selecting companies of interest

We evaluate companies from the results of Steps 1 to 3 and select companies of interest.

Step 5. Data collection

We subdivide CSR reports of the companies of interest for specific TQM activity extraction. In order to facilitate the TQM activity extraction, we prepare a frequency table of nouns by adding the six criteria from the research to the data.

Step 6. PLSA

As in Step 2, PLSA is performed on the frequency table of nouns created in Step 5.

Step 7. TQM activity extraction

We extract TQM activity of companies from the results of Step 6.

3 Analysis Method

3.1 Probabilistic Latent Semantic Analysis (PLSA)

PLSA is an effective method for extracting useful knowledge from big data and is also used as a clustering method. It is an effective method for high dimensional text data. In thinking of PLSA, it is assumed that there is a latent class z that becomes a common topic between the document d and the word w appearing therein. This latent class is generated probabilistically. The concept of PLSA can be illustrated with the graphical model in Fig. 1. It can be represented by three kinds of random variables, and as shown in Eq. 1, the number of words w included in document d is $n(d, w)$. We maximize i Eq. 1 by taking the log-likelihood. A document d belongs to the document set D , and a word w belongs to the word set W .

$$L = \sum_{d \in D} \sum_{w \in W} n(d, w) \log P(d, w) \quad (1)$$

We use the Expectation Maximization (EM) algorithm to maximize Eq. 1. The EM algorithm is a learning algorithm for incomplete data that can be used for parameter

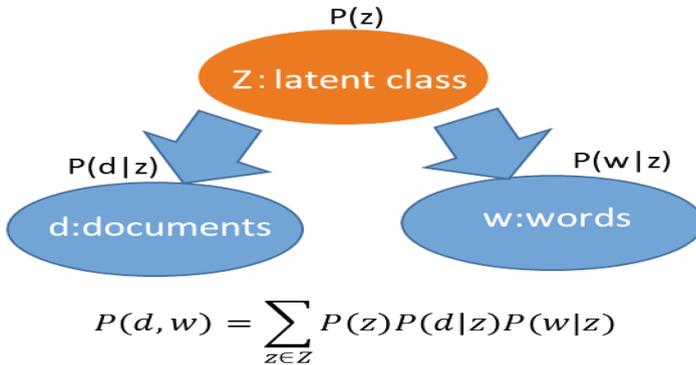


Fig. 1. PLSA graphical model

estimation of a mixed distribution model. It is a method that approaches the optimal solution by successively improving the solution, and it converges quickly in the initial stage. In the E step, the expected value of the log-likelihood function is obtained according to the current parameter value. In the M step, the parameter is updated so as to maximize the expected value of the log-likelihood obtained in the E step. The E step is shown in Eq. 2, and the M step is shown in Eqs. 3 to 5.

E-step

$$P(z|d, w) = \frac{P(z)P(d|z)P(w|z)}{\sum_{z \in Z} P(z)P(d|z)P(w|z)} \tag{2}$$

M-step

$$P(w|z) = \frac{\sum_d n(d, w)P(z|d, w)}{\sum_{d, w'} n(d, w')P(z|d, w')} \tag{3}$$

$$P(d|z) = \frac{\sum_w n(d, w)P(z|d, w)}{\sum_{d'', w} n(d'', w)P(z|d'', w)} \tag{4}$$

$$P(z) = \frac{\sum_{d, w} n(d, w)P(z|d, w)}{\sum_{d, w} n(d, w)} \tag{5}$$

3.2 Word Cloud

We extracted meaning by visualizing topics generated by PLSA with a word cloud. It is possible to select a plurality of frequently occurring words in the document and impress the contents of the method also sentence using size to correspond to the frequency. In this study, the probability that a word belongs to a topic, rather than the word's frequency, is expressed by the size of the letters.

3.3 Cosine Similarity

Cosine similarity is a similarity calculation method that is used to compare documents in a vector space model. It expresses similarity by angles between vectors. A value close to 1 indicates similarity, while a value close to 0 indicates dissimilarity. The calculation formula for cosine similarity can be expressed by Eq. 6

$$\cos(a, b) = \frac{\sum a_i b_i}{\|a\| \cdot \|b\|} \tag{6}$$

where a_i and b_i are the i^{th} elements of a and b , respectively, the numerator on the right side is the inner product of a and b , and the denominator is the product of the lengths of the two vectors. The closer the cosine is to 1, the more similar to vectors are considered to be. In this research, because the similarity calculated in Eq. 6 is treated as a distance, we convert similarity to distance using Eq. 7.

$$\cos(a, b) = d = 1 - \cos(a, b) \tag{7}$$

3.4 Multi-dimensional Scaling (MDS)

MDS is a method of multivariate analysis and is a way of placing novel data between individuals in two- or three-dimensional space in the vicinity and other things in the distant place. A matrix $D_{m \times m}$ is as shown in Fig. 2 below. The distance can be freely defined, but it must satisfy the distance axioms of Eqs. 8 to 10. In this research, the content of the data is the distance found using the cosine similarity as shown in Eq. 7. The distance d in Eq. 7 is data satisfying the axioms of distance.

$$d_{ij} \geq 0 \tag{8}$$

$$d_{ij} = d_{ji} \tag{9}$$

$$d_{ij} + d_{jk} \geq d_{ik} \tag{10}$$

$$D_{m \times m} = \begin{matrix} & \begin{matrix} \text{data 1} & \text{data 2} & \dots & \text{data } j & \dots & \text{data } m \end{matrix} \\ \begin{bmatrix} 0 & d_{12} & \dots & d_{1j} & \dots & d_{1m} \\ d_{21} & 0 & \dots & d_{2j} & \dots & d_{2m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ d_{i1} & d_{i2} & \dots & d_{ij} & \dots & d_{im} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ d_{m1} & d_{m2} & \dots & d_{mj} & \dots & 0 \end{bmatrix} & \begin{matrix} \text{data 1} \\ \text{data 2} \\ \dots \\ \text{data } i \\ \dots \\ \text{data } m \end{matrix} \end{matrix}$$

Fig. 2. Matrices handled by MDS

By converting the distance matrix $D_{m \times m}$ obtained in Fig. 2 as shown in Eq. 11, the eigenvector of the matrix $Z_{m \times m}$ is taken as the coordinate value of the point.

$$z_{ij} = -\frac{1}{2} \left(d_{ij}^2 - \sum_{i=1}^m \frac{d_{ij}^2}{m} - \sum_{j=1}^m \frac{d_{ij}^2}{m} + \sum_{i=1}^m \sum_{j=1}^m \frac{d_{ij}^2}{m^2} \right) \tag{11}$$

4 Result of Analysis

We first present the analysis results of the interested company selection STEP.

4.1 Data Collection

We obtained CSR reports from the companies shown in Table 2 and created a noun frequency table for each group.

Table 2. Grouping by document quantity

Companies with large document volume	Word count	Companies with low document volume	Word count
Sony	177,000	GC	47,000
Casio	169,000	AMADA	44,000
Toshiba	165,000	SAPPORO	43,000
FUJIFILM	121,000	Konica Minolta	43,000
Fujitsu	104,000	DNP	40,000
SHARP	100,000	Canon	37,000
Kubota	76,500	Asahi	35,000
SHOWA DENKO	70,000	TOKAI RIKA	35,000
LIXIL	65,000	SHIMADZU	35,000
		NISSHIN OilliO	26,000
		Nomura Research Institute	12,000

4.2 Introduction of PLSA

In the noun frequency tables, PLSA was used to generate ten topics and a visualization of the top 40 words with a high probability of belonging to each topic in the word cloud to illustrate the meaning of each topic. We determine the meaning of each topic using the word cloud and then decide which company to focus on. Figure 3 shows a word cloud for Topic 4 of Companies with low document volume. Figure 4 presents a diagram showing the affiliation probability of a company to Topic 4.

In Topic 4, it seems that information about stakeholders is described because it contains words such as “customer,” “product,” “employee,” and “production.” Konica Minolta is a company with a higher probability of being affiliated with subjects in



Fig. 3. A word cloud of Topic 4 for companies with low document volume

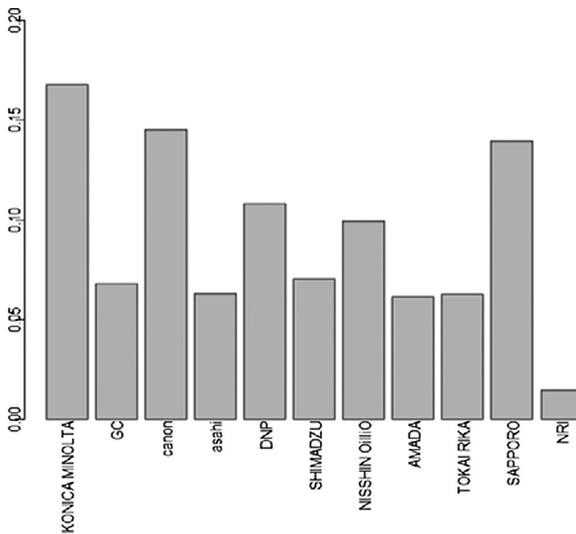


Fig. 4. Diagram showing the affiliation probability of a company to Topic 4

Topic 4. In this way, a meaning for each topic is posited. Table 3 summarizes topics of companies with a high document volume, and Table 4 summarizes topics of companies with a low document volume.

4.3 Calculation of Company Similarity with Cosine Similarity and Visualization of Similarity with MDS

By using each company’s probability of affiliation to each topic, we calculate the similarity of the company using cosine similarity as shown in Eq. 6 in Sect. 4, and

Table 3. Topic summary of companies with high document volume

Topic1	Thought to describe projects related to the environment.
Topic2	Thought to describe the emission of environmental substances
Topic3	Thought to concern customer-oriented information, human resource development, etc.
Topic4	Thought to describe environmental activities and services.
Topic5	Thought to describe chemical emissions and employees.
Topic6	Thought to contain information about safety management at the time of trading and product development.
Topic7	Thought to describe environmental activities and conservation
Topic8	Thought to describe safety management of products and transactions in the supply chain.
Topic9	Thought to describe corporate social activities
Topic10	Thought to describe each stakeholder.

Table 4. Topic summary of companies with low document volume

Topic1	Thought to describe the company's tasks and activities.
Topic2	Thought to contain information about the environment and social contributions.
Topic3	Thought to contain information about safety management, improvement and so on for production in the supply chain.
Topic4	Thought to describe each stakeholder.
Topic5	Thought to describe the company's environmental activities.
Topic6	Thought to describe efforts to achieve customer satisfaction.
Topic7	Thought to describe environmental and social activities.
Topic8	Thought to describe efforts related the global environment, such as energy conservation.
Topic9	Thought to mainly describe the company's products.
Topic10	Thought to describe employee health management.

following the distance axiom in Eq. 7, we recalculate the distance. Each company's affiliation probability to a topic was converted to similarity, but it was converted for 10 topics (10 dimensions), making visualization in a single figure difficult. Visualization is therefore performed with MDS, which can be used to arrange multidimensional information in a two-dimensional or three-dimensional space. Figures 5 and 6 contain visualizations of the relationship between companies with a high document volume and companies with a low document volume. (Red letters denote the top 10 companies; blue letters denote the lower 10 companies.)

In Fig. 5, the grouping takes a relatively near form, with the top ten companies tending to shift toward the positive side of the y-axis, while the lower ten companies tend to shift towards the negative side. In contrast, the lower ten companies showed relatively greater variability in Fig. 6. The reason for this is that in Fig. 5, the document volume is large and the contents of the document are comparatively similar, whereas

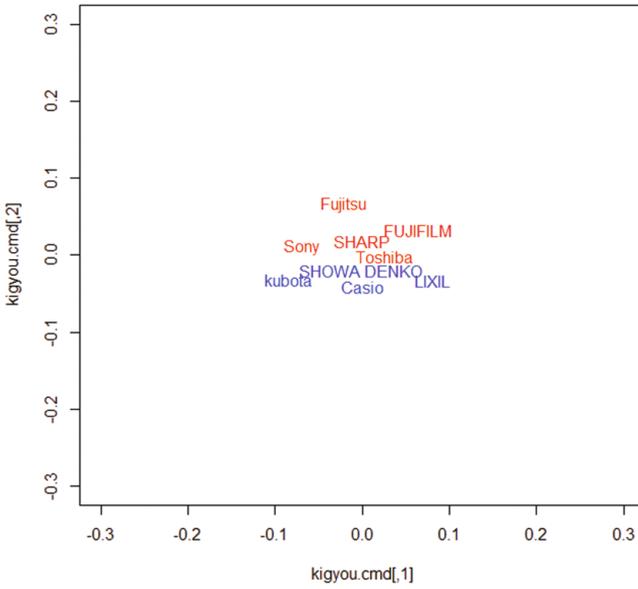


Fig. 5. Visualization of the relationship between companies with high document volume (Color figure online)

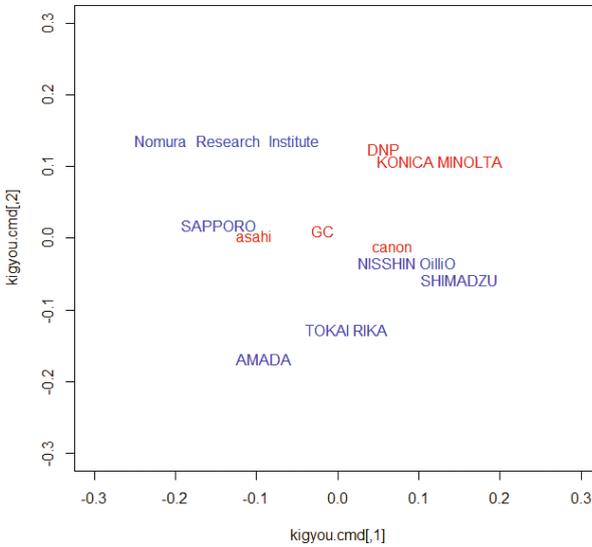


Fig. 6. Visualization of the relationship between companies with low document volume (Color figure online)

Table 5. Companies of interest

Highly ranked companies that have a high probability of affiliation to TQM topics	Konica Minolta GC Toshiba
Low ranked companies with a low probability of affiliation to TQM topics	Kubota
Low ranked companies with higher affiliation probability to environmental topics	Sapporo

the document volume is small in Fig. 6 meaning that the content may depend on the company.

Based on these analyses, companies of interest were selected. The selected companies and the reasons for their selection are summarized in Table 5.

5 TQM Information Extraction Step

Because the TQM information extraction step is similar to the step for selecting companies of interest mentioned in Sect. 5, we only describe differences between the extraction procedure and the evaluation procedure in this section. To make this easier to understand, we focus in particular on Konica Minolta in the explanation.

5.1 Data Collection

The target of the data is five companies, “Toshiba,” “GC,” “Konica Minolta,” “Sapporo,” and “Kubota,” listed in Table 5. For the extraction of quality management information in this section, we use Evaluation and Difference to evaluate concrete TQM, finely divide each company’s CSR reports, and obtain frequency matrices of nouns. The TQM is evaluated according to the six criteria of the TQM research described in Table 1. When considering the meaning of the topic, the quality management information contains many duplicate words and similar words. We therefore add items related to each criteria of the TQM research as data. Table 6 shows the question items of the TQM research.

5.2 Introduction of PLSA

We first focus on Konica Minolta as an example and to derive topic meanings (Fig. 7).

In word cloud for Topic 3 in Fig. 8, words related to quality management such as “people,” “employees,” “quality management,” “president” can be seen, so we connect this to “human resource development to achieve quality management,” “individual commitment” and the like. In the diagram of probability of affiliation to “manager’s commitment” and “personnel training for quality management” is also high. We posit the meaning of topics for the other four companies in this same way.

Table 6. Question items of the TQM research

Customer intention	Do you conduct a customer satisfaction research?
Manager commitment	How do you position and disseminate the concept of quality emphasis, such as being “customer-oriented” and “putting quality first,” in the corporate philosophy and management philosophy?
Personnel training for TQM	Do you create a mid- to long-term plan for quality management education throughout the company?
Management of relief, security, and trust	Do you conduct employee satisfaction researchs?
Establishment, observance of the process	Do you clarify the roles and processes of each organization in manufacturing so as to ensure and improve the quality of products to be offered to customers?
Utilization, development power of the systems	Do you investigate what kind of products customers want and reflect this in new products in a timely way?

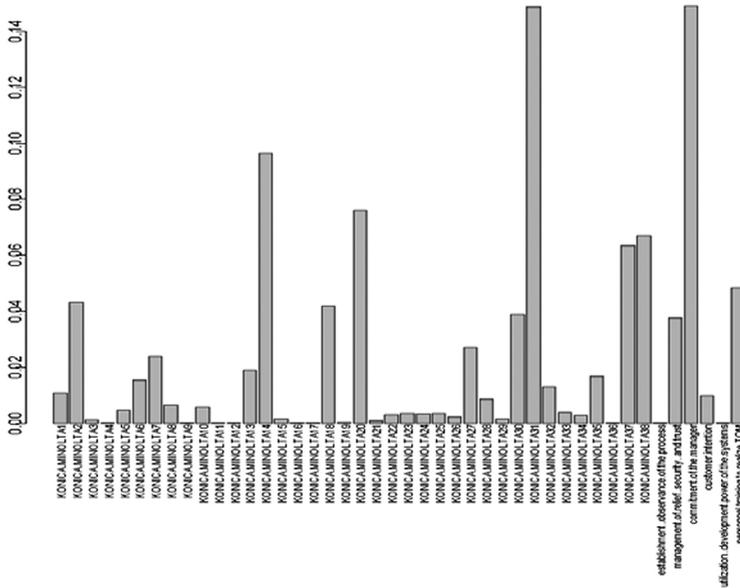


Table 9. TQM information extraction result

Konica Minorta	TQM information includes descriptions of executive personnel education, top management topics for human resource development of employees, and training of business leaders
GC	TQM information includes a description of the management philosophy and quality management, a concrete statement that the president has confirmed to what extent company policy has been implemented by going to the site
Toushiba	TQM information includes descriptions of management’s commitment, such as measures for risk compliance management
Sapporo	TQM information includes information about human resource development, establishment of a group quality assurance system, compliance, and information on the six criteria is mixed
Kubota	TQM information includes descriptions of specific measures such as securing of quality and training of human resources to expand production bases and improve customer satisfaction

Table 10. CSR report contents for each company based on analysis of main contents

Konica Minorta	Contained information about “environment,” “effort for customers,” and “human resource development”
GC	Contained information about “environment,” “management commitment,” and “human resource development”
Toushiba	Contained information about “environment,” “effort for customers,” and “management structure”
Sapporo	Contained information about “environment,” “social contribution,” and “management structure”
Kubota	Contained information about “environment” and “social contribution”

“personnel training for TQM.” The criterion for placement of the circles in these tables is shown in Eq. 12.

$$\left\{ \frac{100}{\text{Number of documents} + 6} \right\} \times 2 \tag{12}$$

The number of documents in Expression 12 refers to the number of CSR reports of each company as a subdivision, and the 6 in the denominator represents the six criteria. The value obtained in Eq. 12 is rounded off to the third decimal place, and the probability of exceeding Eq. 12 is judged to be high for topics. Table 9 shows the TQM information extracted for each company.

6 Discussion

In this study, we evaluated TQM and extracted TQM information using corporate CSR reports. We selected companies of interest using the analysis result of PLSA. Companies with a high probability in the topic on quality management were “Konica

Minolta,” “GC,” and “Toshiba.” Conversely, “Kubota” was a company with a low probability of topics related to TQM, while “Sapporo” had a high probability of topics on environmental information. Based on these results, we selected these five companies to focus on. In the MDS results (Figs. 5 and 6), companies with high document volume were located relatively close together. The reason for this is the large amount of documents and their contents are not biased. Because many of the top companies are electrical manufacturers, their CSR reports are considered to be similar. We subsequently extracted each company’s TQM information. The results of extraction indicate that most of the TQM information included in the CSR report is related to “manager commitment,” “customer intention,” and “personnel training for TQM.” Table 10 summarizes the main contents described in each company’s CSR report.

7 Conclusions

In this study, using the text data of the corporate CSR reports, we used PLSA to probabilistically classify and evaluate the company’s CSR reports. The extent to which each company disclosed TQM information to customers was extracted. In the corporate evaluation, we analyzed contents using PLSA and posited topic meanings with word clouds so that we could focus our attention on companies with a high probability of affiliation to topics related to quality management. By classifying the CSR report of the companies of interest and adding the six criteria from the TQM level research to the data, classification of documents related to quality management was facilitated. Finally, we were able to obtain information on TQM activities of companies from Table 9.

References

1. Kamio, M.: A study of management in the aspect of human-being. *Prod. Manage.* **3**(1), 30–33 (1996). (In Japanese)
2. Shibayama, N., Nakagawa, H.: Introducing off-diagonal elements to singular value matrix in probabilistic latent semantic indexing. *Trans. Japan. Soc. Artif. Intell.* **26**(1), 262–272 (2011). (In Japanese)
3. Saitoh, F.: Visualization of online customer reviews and evaluations based on Self-organizing Map. In: *Proceedings of 2014 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 182–187
4. Hofmann, T.: Probabilistic latent semantic analysis. In: *Proceedings of Uncertainty in Artificial Intelligence*, pp. 289–296 (1999)
5. Kitora, Y.: Relationship between the corporate stance toward stakeholders and disclosures on corporate websites. *Environ. Conserv. Eng.* **39**(2), 103–111 (2010). (In Japanese)
6. Union of Japanese Scientists and Engineers: The result of the 9th JUSE Quality Management Level Research announced. JUSE 10, March 2016. (Online)