Research Paper (Short Note)

Text Mining of the Midterm Goals of National Universities for the Development of Their Individual Characteristics

Kahori Ogashiwa*, Masao Mori** and Sachio Hirokawa***

(Received 31 August 2019 and accepted in revised form 8 January 2020)

Abstract This paper attempts to discover the distinguishing characteristics of Japanese national universities, and to determine the strengths of each university by analyzing their midterm goals. We propose a method that extracts words from documents, analyzes them, and visualizes the features of documents on the basis of these extracted words. We developed two systems, a cross-tabulation search engine and a keyword map search engine, for the midterm goal document. In these systems, the user both decides the viewpoint to be compared and analyzed and inputs the search word, so that the relation of the words related to the search word is visualized, and a novel comparison analysis method among university organizations is realized. In this paper, we report the outline of these systems and the results of the verification of the two systems, using “regional contribution” as an example.

Keywords: text mining, keyword map, midterm reports, university evaluation, institutional research

1. Introduction

1.1 Background

As a result of the establishment of the National University Corporation Law on April 1, 2004, national universities became incorporated and obliged to undergo national university corporation evaluations (hereinafter referred to as “corporation evaluation”). The purpose of corporation evaluation is to continuously improve the quality of the universities, their accountability to society, and to calculate their subsidies for operating expenses. The calculation of subsidies for operating expenses, the acquisition of subsidies for competitive funds, and the strengthening of functions are closely related; thus, corporation evaluation tends to be taken very seriously by universities. However, the original purpose of the evaluation is to contribute to the development of the individuality of each university. It is important for universities to know their strengths and to develop and extend them.

To identify the strengths of each university (including its brand or image and the accomplishments of its management), correspondence to the ex-post evaluation (preparation of the report and the presentation of the basis data) is conducted to understand each university’s characteristics. In addition, it is essential to conduct heuristic institutional research (a new idea or discovery based on data with support from information technology).

As mentioned earlier, corporation evaluation aims at developing the strengths of universities. This research focuses on this purpose of corporation evaluation and aims to develop the strengths of its own university. Further, to bring out its own strengths, comparison with another party is an important factor. Therefore, universities need to find other universities that will serve as a benchmark for them. If universities can find similar universities with new perspectives that are not bounded by the existing framework, they can project their own strengths.

1.2 Related Work

Various studies are being conducted to address these issues. Takata researched the progress management system of universities’ midterm goals and midterm plans(1). Fujii studied the types and characteristics of indicators included in the midterm plans(2). Sun et al. conducted research on university personality and indicators using fundamental numerical data from university portraits(3). Funamori conducted a comparative analysis of universities using data from the School Basic
TEXT MINING OF THE MIDTERM GOALS OF NATIONAL UNIVERSITIES FOR THE DEVELOPMENT OF THEIR INDIVIDUAL CHARACTERISTICS

Survey\(^4\). Mori conducted research on a system capable of benchmarking comparative analysis between universities in real time using data from the School Basic Survey analyzed by Funamori\(^5\). Recently, research on the analysis of universities using Twitter text data has become a trend\(^6\), \(^7\). In Ref. (8), we reported a preliminary study that attempted to use text mining techniques to discover the strengths of individual universities.

However, these studies provide analyses from a limited viewpoint, and do not analyze the strengths of the universities as organizations. As far as the authors know, at present, no method of analyzing the strengths of universities as organizations has been established. Therefore, in this paper, we propose a method to visualize the comparative analysis results from different universities to identify their strengths using the text mining method. In addition, the study reports some examples of comparative analysis among universities by employing “cross tabulation” and “keyword map” search engines. These examples can help each university discover an analytical viewpoint for its benchmarked target.

2. Explore the Strengths of the University Using Two Search Engines

These search engines use attributed words. An attributed word is a word with a symbol that identifies where the word appears\(^9\). In this paper, readers should understand that the word “region” in the manuscript of Utsunomiya University, for example, involves Utsunomiya University. Words with attributes include university names and focused words. See 2.1 (2) (Figure 1, Figure 2).

We selected the following 22 words as “focused words.” We have chosen these words as both the axis and the viewpoints of each university with respect to region, international relation, and industry. We distinguished those words from other words:

```
international, global, region, world, innovation, 
earth, standard, overseas, domestic and foreign,
English, wide area, our country, foreign countries, 
citizen, 21st century, local government, national 
universities, next generation, marine, creative na-

tion, people, whole country
```

In both the cross tabulation and the keyword map systems, a specified number of keywords are selected and their co-occurrence is visualized. These 22 “focused words” were chosen as the axis of analysis. At this time, the realized system can select M words from the 22 focused words and N words from the general words separately. The scale of word selection is based on the three scales: the number of documents that contain each

---

Figure 1. Setting Screen. In this Screen, the Word “region” is Entered in the Search Box. The user must also check that the keyword is displayed on the horizontal axis and that the university lies on the vertical axis. See 2.1 for details.

Figure 2. Setting Screen. In this Screen, the Word “region” is Entered in the Search Box. See 2.2 for details.
word, the number of occurrences, and the degree of relevance to the search term. If there is no distinction between the focused words and general words, the result may not include any focused words that define the viewpoint of analysis. By selecting M focused words separately for general words and focused words, we can interpret the visualization.

2.1 How to Use “Cross Tabulation” Search Engine

The “cross tabulation” search engine was developed by diverting an existing system for research search.16

(1) Collect the documents you want to analyze.
(2) Consider what kind of viewpoints through which you want to analyze the documents. Tag each document with the university name. For each document, register the keywords you want to analyze (focused word).
(3) Enter a search word in the search box (Figure 1).
(4) Select two viewpoints to analyze from university name, word (feature word), and keyword (focused word) in order to create a cross tabulation. See (5). Feature words mean words extracted by morphological analysis.
(5) A cross tabulation is displayed.
(6) For example, if a university name is set on the vertical axis and a keyword (focused word) is set on the horizontal axis, the top five positions are displayed among the focused words, and the focused word appears for each university. You can find whether or not the focused words appear.

2.2 How to Use “Keyword Map” Search Engine

The “keyword map” search engine was developed by diverting an existing system for research search.9

(1) to (3) are the same procedure as above (Figure 2).
(4) For example, in the university name, word (feature word), and keyword (focused word), the top 50, 90, and 30 are set to be displayed, while similar groups are displayed in a two-dimensional map.

2.3 Advantages of Using Two Search Engines

(1) The “cross tabulation” search engine can display university names and focused words as a cross-table with the horizontal and vertical axes, so it can be seen at a glance whether focused words appear. However, the relationship between words is not shown.
(2) The “keyword map” search engine can display the university name, feature words, focused words, and all the relationships on a map. In addition, associations between words are displayed.

3. Case Study: “Regional Contribution”

In this case study, we looked for universities that conduct activities related to regional contribution among national university corporations, and discovered the strengths of each university by comparing the differences among these universities.

3.1 Data

We targeted third-term midterm documents (text data)11. The midterm goals were roughly divided into five major items, as follows: (I) Goals concerning quality improvement of university educational research; (II) Goals concerning the improvement and efficiency of business operations; (III) Targets for improvement of financial contents; (IV) Self-inspection / evaluation of targets related to the provision of information pertaining to the situation; and (V) Other important objectives related to the operation of the business. Since this study aims to identify the strengths related to educational activities, we targeted a part of the description of “(1) Objectives related to education” in “(I) Goals concerning quality improvement of university educational research.” Regarding educational goals, nine universities with descriptions of universities and long sentences with different descriptions of item names were excluded. For that reason, we used data from only 77 of 86 national universities.

3.2 Methods

We investigated the strengths of universitys’ educa-
tional activities in terms of “regional contribution.” We analyzed the data based on the following research items: (a) a list of focused words; and (b) university names. We used the following procedures:

1. We set the focused words. In this study, words such as “region,” “international,” and “global” were set as focused words in order to analyze the educational activities on regional contribution, which are important elements with social relevance.

2. We used the “cross tabulation” search engine and extracted the universities that contained the focused word “region” in their documents.

3. Using the “keyword map” search engine, we extracted the universities that contained “region.” For the words appearing in the search result document, the top N words of the focused words were selected, and for co-occurrences, the branch between the two words was displayed.

3.3 Results and Discussion

3.3.1 Result Using “Cross Tabulation” Search Engine

Using the “cross tabulation” search engine, we displayed a cross table of “university name” and “top 5 focused words” and studied universities containing only the focused word “region.” Thirty of 77 universities included “region” (Figure 3). First, “region,” “international,” “global,” “world,” and “citizen” were listed as the five most frequently occurring words in the data from the 30 universities. In this context, we extracted 7 universities with patterns flagged only with the focused word “region” (region: yes, international: none, global: none, world: none, citizen: none). The results obtained from the “cross tabulation” search engine displayed that “international” and “global” were the two most frequently occurring focused words among the 30 universities that contained the word “region.” Thus, several universities that emphasize regional contributions also highlight international contributions and global viewpoints. From the analysis of the keyword “region,” we were able to extract a group of universities that are considered to have strengths in regional contribution regarding educational activities.

![Figure 3. Keyword Appearance Using the “cross tabulation” Search Engine.](image)

A list of university names is displayed on the left side and a list of the five most frequently occurring focused words (“region,” “international,” “global,” “world,” and “citizen”) is displayed on the right side. If a focused word was found for the university, a flag was set and the cell was highlighted in pink. This makes it possible to understand the distribution of “university × keyword” at a glance.

3.3.2 Result Using “Keyword Map” Search Engine

The relationships between the top 90 feature words were mapped using the “keyword map” search engine to 30 universities that included “region” (Figure 4). Further, Utsunomiya University is used as an example in this paper because it was a previous affiliation of one of the authors of the paper. As a premise, Utsunomiya University has established “regional contribution” as one of its midterm goals.

**Research question:** Which universities can appropriately be compared with Utsunomiya University?

**Method:** Universities with characteristics similar to
Utsunomiya University were analyzed using “keyword map” search engine.

(1) Universities that contained the keyword “region” were extracted.

(2) The keyword map displaying the co-occurrence relations between words were noted.

(3) Universities using the same path as Utsunomiya University were discovered.

Figure 4 shows the results of drawing a map with the “search word” as “k: region (focused word).” Further, the Utsunomiya University features that could be extracted from among the 30 universities in the search results were analyzed to exemplify the detailed evaluation. In this map, we asked for a university similar in characteristics to Utsunomiya University by using “n: Utsunomiya University (university name),” somewhere that follows the same path to Utsunomiya University. In this case, four universities were found with the upper left path as the common path. The midterm goal documents of universities with the same path were analyzed, and the strengths of each university were comprehended in detail. This analysis also aided in determining the target university for benchmarking. Using the four extracted universities, Utsunomiya University, Saga University, Shimane University, and Miyagi University of Education, the development of human resources who can play an active role in the region was written into midterm goal documents.

3.3.3 Discussion

The “cross tabulation” search engine offers the advantage of understanding the relationship between the
university name and the keyword at a glance; however, it does not know the relation of the words. In the “keyword map” search engine, the relations between words are also displayed on the map, and a detailed analysis can be performed. It is possible to conduct varied analyses by combining these systems. The relevance of the feature word and the focused word can also be analyzed through the keyword map; universities can thus be classified on the basis of the relevance of these words. The function of visualizing these results as a two-dimensional map is practical and has multiple potential applications. It is possible to classify universities by applying the “keyword map” search engine to documents representing university activities (i.e., a field of university evaluation) by inserting a search word from the user’s viewpoint. Therefore, a powerful piece of evidence can be accessed to extend the distinctiveness of individual universities, which is one of the purposes of corporation evaluation.

4. Conclusion

This paper proposes a method to visualize the results of a comparative analysis of universities to identify their strengths using the text mining method. Utilizing two search engines, we were able to find a group of universities with strengths in the area of regional contribution. In fact, as a result of investigating the composition of graduates and the graduates by employment area, it was confirmed by some universities that the employment rate in their prefecture is the highest(12), and thus, we were able to verify the usefulness of this system. Since the data used in this study were obtained using only one of the educational goals in the midterm goals documents for analysis, in future research we will incorporate more sources of data, such as the midterm plan documents.

Acknowledgments

We would like to thank the reviewers for their valuable comments. Data were provided by the Ministry of Education, Culture, Sports, Science and Technology (MEXT). We sincerely thank Dr. Miho Funamori, Associate Professor of the National Institute of Informatics, for providing data. This work is supported by MEXT, Grant-in-Aid for Scientific Research (C) (JSPS KAKENHI Grant Number JP17K04677).

References

(1) Takata, E.: “The Subject about the Data Collection in the Progress Management of a Mid-Term Plan”, Proc. of IIAI International Conference on Advanced Applied Informatics (IIAIAAI), pp. 527–528 (2014).
(2) Fujii, T.: “University Evaluation and Institutional Research”, Vol. 7, pp. 3–10 (2016), http://iir.ibaraki.ac.jp/jcache/lib/docu/007_h2811/007-h2811-01_fujii.pdf, (accessed 2019.8.31). (in Japanese)
(3) Sun, Y., Todo, N. and Inoue, S.: “Exploration of Higher Education Indicators and Universities’ Characteristics in Japan”, Proc. of IIAI International Conference on Advanced Applied Informatics (IIAIAAI), pp. 411–416 (2014).
(4) Funamori, M.: “Thorough Analysis of School Basic Survey: Enrollment and Retention Management”, http://researchmap.jp/?action=cv_download_main&upload_id=94797, (accessed 2019.8.31). (in Japanese)
(5) Mori, M.: “Benchmark System of School Basic Survey”, http://www.masaomori.net/ir-analysis/BSJE_bench/, (accessed 2019.8.31). (in Japanese)
(6) Pu, X., Chat, MA. and Schroeder, U.: “Wiki-LDA: A Mixed-method Approach for Effective Interest Mining on Twitter Data”, Proc. of 8th International Conference on Computer Supported Education (CSEDU), pp. 426–433 (2016).
(7) Funamori, M. and Mori, M.: “Conceptual Design toward a Visualization System of University’s Web Presence: Simple Analysis and System Development Using Twitter”, Proc. of IIAI International Conference on Advanced Applied Informatics (IIAIAAI), pp. 456–461 (2016).
(8) Ogashiwa, K., Mori M. and Hirokawa S.: “Text Mining of University Midterm Reports for Comparison of Educational Strength”, Proc. of New Perspectives in Science Education, 6th Edition, pp. 322–326 (2017).
(9) Hirokawa, S., Ito, E. and Baba, K.: “Visualization of Keywords for Exploratory Search of Research”, J. of Information Processing and Management, Vol. 58, No. 6, pp. 447–454 (2015). (in Japanese)
(10) Yin, C., Tabata, Y., Wu, X. et al. “Building a Search Engine for Scientific Projects Survey”, Proc. of IEEE International Conferences on Internet of Things and Cyber, Physical and Social Computing, pp. 558–563 (2011).
(11) Midterm Goals and Midterm Plans of National Universities, Ministry of Education, Culture, Sports, Science and Technology, http://www.mext.go.jp/a_menu/koutou/houjin/1368750.htm, (accessed 2019.8.31). (in Japanese)
(12) Japanese College and University Portraits, http://portraits.niad.ac.jp (accessed 2019.8.31). (in Japanese)