

Case Study of Text Analytics Applied to Accident Reports of a University

Rumiko HAYASHI^{1*}, Tsubasa YAMADA², Kouhei SHINKAWA², Kengo TOMITA¹, Tadashi NISHIKIMI¹, Shizuaki MURATA³ and Hidekazu KURIMOTO^{2,4}

¹Environment, Health & Safety Office, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan

²Graduate School of Environmental Studies, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan

³Professor Emeritus, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan

⁴Institute of Liberal Arts and Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan

Abstract Many accidents have occurred in universities and the accident reports are accumulated in most universities. The information described in the accident reports must be used effectively to prevent a recurrence of the accidents. In this study, we applied text analytics to the description written in 373 accident reports in a university as a case study. Information mining method was adopted for the contents analysis, and 9 factors based on m-SHEL and human error, that is “software”, “hardware”, “environment”, “liveware2”, “management” “slip”, “lapse”, “mistake”, and “violation” were used for morphological analysis for description in report. The factors in each category of accident situation were extracted, and it is suggested that text analytics is one of the most effective methods to analyse the accident reports in universities.

1 Introduction

Many accidents have occurred in universities, such as fire, burn, cut by glassware, burst of chemicals, chemical injury, explosion, falling, traffic accident, and so on, not only in laboratories but also out of laboratories. Accident cases have been collected in most universities and countermeasures have been taken to prevent the recurrence of the accidents. Some of the major accidents were analysed deeply, and their lessons were shared with university members in various ways such as posting on web pages, posters on bulletin board, e-mail newsletters, safety training courses, and so on. However, information on minor accidents including near misses is often not analysed and remains underutilised due to the large number of cases. Even if all the information is fed back to university members, they will consider that most accidents irrelevant to their own activities. Therefore, there is a need for a method to extract useful lessons from minor accidents and to extract more relevant accidents by individual activities and status from the accident data base.

There are few reports of analysis of university accidents. Some universities have published accident statistics and serious accident information on the web. Nishikimi *et al.* (2019) estimated injury risk for the students involved in experimental works by using accident data reported in Nagoya university. Tachikake *et al.* (2016) classified the chemical-related accidents by academic major and status at university and suggested that cases using chemicals as tools were more frequent

than cases using them as the actual subjects of the research. These statistical analysis and classification considering university-specific situations are very important and informative, however, it depends on the point of view of the analyst and the analysis method has not been well established.

We focus on text analytics as a method to classify accident cases and to extract useful information from accident data base. Text analytics was adopted for accidents and near misses analysis in various fields, such as oil industry (Minowa *et al.*, 2012; Minowa and Munesawa, 2014), drug safety (Kimura *et al.*, 2005), construction accidents (Fan and Li, 2013), traffic accidents (Sakai *et al.*, 2006), and so on. Adopting this method for accident descriptions, accidents factors and similarities among accidents can be extracted from a vast database of accident cases. In this study, text analytics was adopted for a university accident database to examine the applicability of this method to extract the characteristics and factors of the accidents as a case study.

2 Accident Report Data Used in This Study

We adopted text analytics for accidents and disaster data reported in a large integrated university. Accident reports are mandatory in the university including injury, fires and explosion with no injury, chemical leaks to environment, chemical exposures without any symptoms. In this study, text analysis was conducted on 373 cases

* Corresponding author: hayashi@esmc.nagoya-u.ac.jp

reported during two fiscal years (2013 FY and 2014 FY), including a lot of minor cases with no need of any medical cares and no days off from their work.

The report form consists of three main parts, basic accident information, injured or reported person information, and actions after accident and current situation. The basic accident information part includes the date and time of the accident, location, and free description about the accident. Person's information includes age, sex, status, and injured parts. The actions after accident part is also described by free text. Some of the entries in the form is not text form, therefore, we call the analysis method adopted in this study "information mining". All the descriptions excluding personal name and accident location were used for information mining. All the entries in the report form are written in Japanese.

Each accident is classified into 6 accident categories ("experiment and field work", "non-experiment work", "commuting", "club activities", "traffic accident in campus", and others) according to the situation of the accident. Category of each accident was decided referring to the basic information part of the report. The normalized number of cases in each category was shown in Figure 1. The others include accidents during work at medical sites, at business trips, injury of non-university member, and so on.

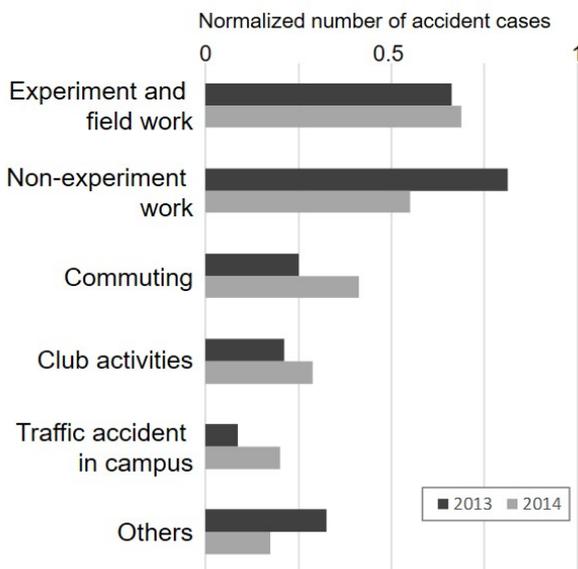


Figure 1. Normalized number of accident cases reported in 2013 FY and 2014 FY classified in each category.

3. Procedures of Information Mining

3.1 Morphological analysis

First, text data written in accident report is pre-processed to convert sentences of natural language to computer-readable form. In this step, a sentence is divided into each morpheme, which is smallest grammatical unit which has meaning. The software Text Mining Studio 5.2 (NTT DATA Mathematical Systems Inc.) was used for morphological analysis.

3.2 Specific word analysis

Next, pre-processed data converted in section 3.1. are used for identification of the factors of each accident. In this study, 9 points of accident factors based on m-SHEL model (Kawano, 1999) and 4 human factors, that is "software", "hardware", "environment", "liveware2", "management", "slip", "lapse", "mistake", and "violation" are adopted. The summary of the concept of 9 factors of accidents used in this study is shown in Figure 2. "Liveware1" means the person concerned of the accident, such as injured person, and "liveware2" means the person not concerned directly. "Slip", "lapse", "mistake", and "violation" are the terms for human error. To identify the factors of each accident, we created a table of "specific word" list, which is a kind of dictionary. By reading the free text description, we determined the specific word representing each factor of accidents. The words are classified in each factor which the word is related to. To ensure the objectivity of specific words, two different people independently extracted specific words from accident reports, and confirmed that there were no significant differences between the extracted words. We also have confirmed that there was no major inconsistency in the factor extraction results for each accident case. Specific word table used in this study was created by comparing two tables carefully and extracting the parts in common, to make it more universal. The appearance of specific words in each factor was counted for each accident category focused on the analysis. Even if the specific words appear more than two times for each factor, the count is one for the factor for one accident. In this study, totally 324 words are extracted as specific words for analysis.

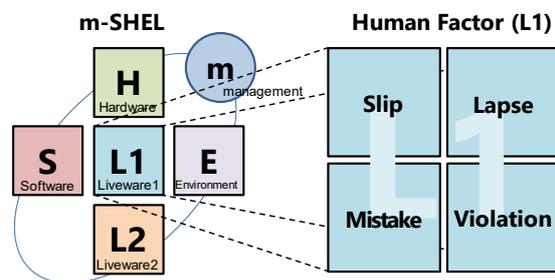


Figure 2. 9 factors of accidents considered in this study by combination of m-SHEL and 4 human factors.

3.3. Evaluation and visualization

At the last step, the computer-readable form results should be converted to human-understandable expression of results. In this study, "Complementary similarity measure" (Yamamoto and Umemura, 2002) was used to evaluate the number of the specific word for each factor in each category focused on each analysis, which was resulted in the section 3.2. The values of complementary similarity measures were used for the visualization of results.

4. Results and Discussion

Accident factors were identified in 144 out of 188 cases for 2013 FY, and 124 out of 185 cases for 2014 FY. For each category of accidents, the identified factors were visualized by complementary similarity measure. The tendencies of factors were compared between 2013 FY and 2014 FY. Some examples of spider charts of the identified factors for each accident category are shown in Figure 3 in terms of the normalized complementary similarity measure. Note that this measure indicates the strength of the relevance of each factor to each accident category, and absolute values of complementary similarity measure should not be compared between different years.

From the spider charts, the similar tendency between 2013 FY and 2014 FY for some accident categories were visualized, and different for the others.

Regardless of the year, “liveware2” and “environment” were extracted as a factor for commuting accident category, and only “liveware2” was extracted strongly for club activities. It is suggested that these factors can be the main factor related to the accident category. “Liveware2” was extracted as a factor for the commuting accidents such as rushing-out of bicycles, rear-end collisions of following cars, and “Environment” was extracted for the accidents on a frozen road surface or snow. For the club activities category, it is considered that liveware2 was extracted because of the many accidents that a player collided with other players during practice. Furthermore, short description in accident reports may be one of the reasons that only one factor liveware2 was extracted.

For experiment and field work category, “lapse”, “mistake”, and “violation” were commonly extracted as factors in both years although there are some differences. These 3 factors are about human errors, which are rarely extracted factors in other categories except for non-experiment work. It is suggested that descriptions related

to human errors are much more written in the accident reports for categories for working than those in the other non-working situation categories. Generally, most accidents occur due to multiple factors including non-human factors. However, in some cases, it may be difficult for reporting persons to notice the other factors such as “management” and “environment” soon after the accident. It may be helpful to devise a format of accident report in order to have a description related to multiple factors.

For non-experiment work, common factors were not extracted in two years. General accidents such as falls, bumps, rubs in general office work are classified into this category, and the accident in various situations are included. The category itself should be changed or divided into detailed categories to extract more useful information for accident prevention. In addition, the descriptions of reports are short in many cases in this category. This also suggests that the necessity of reviewing the reporting format to write more information easily. The method proposed in this study is also expected to be applied as a tool for automatically judging what kind of description is missing in the accident report to clarify the accident factors. It will facilitate self-checking of the factors behind the accident by people involved in the accident, and will assist in their introspection. It can be an educational tool to prevent the recurrence of accidents.

5. Conclusions

Descriptions of the 373 accident reports in a university were analysed by information mining as a case study. Dominant factors related to each accident situation category were extracted with the specific words determined for 9 factors in this study. It is suggested that text analytics is an effective method to extract the accident factors from the information of accident reports in universities. The tendencies of factors for 2013 FY

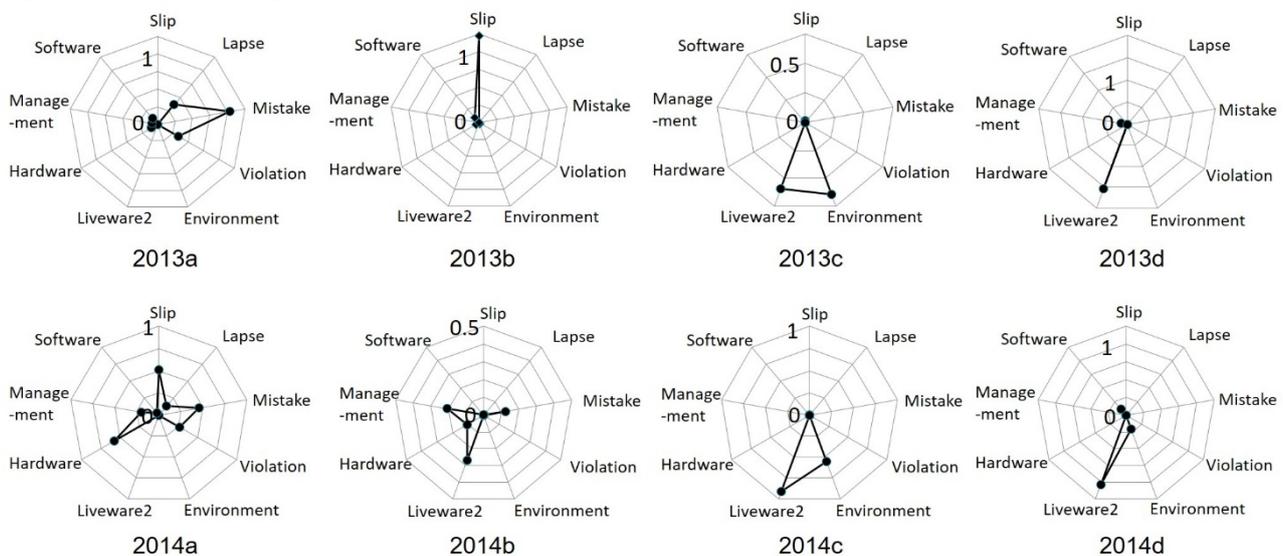


Figure 3. Spider charts of normalized complementary similarity measure in 2013 and 2014 fiscal year for “experiment and field work” (2013a, 2014a), “non-experiment work” (2013b, 2014b), “commuting” (2013c, 2014c), and “club activities” (2013d, 2014d).

accident cases were the similar with those for 2014 FY cases in some accident categories, and the tendencies were not similar between two years in the others. This indicates that setting appropriate categories is important for proper analysis, which is also important for effectively utilizing the results for accident prevention. In order to improve this method, it is necessary to increase the number of accidents applied and to improve the table of specific word list. It is also important to improve the accident reporting format for universities, which makes it easier to extract the accident factors with this method. At the same time, it is expected that information mining adopted in this study can contribute to the improvement of accident report format tailored to the university.

Acknowledgements

This research was partially supported by Grants-in-Aid for Scientific Research, Japan Society for the Promotion of Science (JSPS), No. 17K18624.

References

- Fan, H. and H. Li; "Retrieving Similar Cases for Alternative Dispute Resolution in Construction Accidents Using Text Mining Techniques," *Automation in Construction*, **34**, 85–91 (2013)
- Kawano, R.; "An Idea Generation Procedure for Human Error Reduction Techniques : Principles of Error Proof Technique," *The Journal of the Society for Industrial Plant Human Factors of Japan*, **4**(2), 121–130 (1999) (in Japanese)
- Kimura, M., H. Furukawa, H. Tsukamoto, H. Tasaki, M. Kuga, M. Ohkura, and F. Tsuchiya; "The Analysis of Questionnaires About Safety of Drug Use, the Application of Text Mining to Free Description Questionnaires," *The Japanese Journal of Ergonomics*, **41**(5), 297–305, (2005) (in Japanese)
- Minowa, H., Y. Munesawa, and K. Suzuki; "Method to Extract Remarkable Words and Grouped Accidents According to Co-occurrence and Frequency of Words," *Journal of Japan Society for Safety Engineering*, **51** (5), 319–326 (2012) (in Japanese)
- Minowa, H. and Y. Munesawa; "Integrated Analyzing Method for the Progress Event based on Subjects and Predicates in Events," *Journal of Japan Society for Safety Engineering*, **53** (5), 317–324 (2014) (in Japanese)
- Nishikimi, T., K. Tomita, R. Hayashi, and S. Murata; "Injury Risk Estimation for Students Involved in Experimental Work in Nagoya University," *Journal of Environment and Safety*, **10**(1), 1–7 (2019) (in Japanese)
- Sakai, H., S. Umemura, and S. Masuyama; "Extraction of Expressions Concerning Accident Cause Contained in Articles on Traffic Accidents," *Journal of Natural Language Processing*, **13**(2), 99–123 (2006) (in Japanese)
- Tachikake, T., H. Momose, K. Tomita, I. Shibata, H. Yamamoto, and S. Usui; "Decline in Risk Perception When Using Chemicals as Tools - Suggestions for Laboratory Safety," *Journal of Environment and Safety*, **7**(2), 133–139 (2016)
- Yamamoto, E., and K. Umemura; "A Similarity Measure for Estimation of One-to-Many Relationship in Corpus," *Journal of Natural Language Processing*, **9** (2), 45–75 (2002) (in Japanese)