

A Classification Method of the Incident By Extraction From Text

Kazuhiro Morita, Kenzaburo Yashiro, Masao Fuketa

Abstract—In this research, in order to use for prediction of the accident risk which prevents serious accident and disaster, the method of detecting and classifying an incident from a text is proposed. A multi-attribute matching machine is used for detection and a classification. The feature expression is extracted from the incident case sentence currently released, and detection of an incident and the classification of an accident kind are carried out by the matching rule created from extraction data. Although classification precision was mostly as good as 0.783 as a result of the evaluation experiment, the room for an improvement for extraction precision was seen. The incident which was able to be managed with flawlessness or a slight injury although it was likely to get injured can warn of a big accident, and can urge evasion of it. Therefore, this research which leads to an early warning by detecting and classifying mechanically is meaningful. A future subject is an improvement of extraction precision.

Keywords—incident detection, incident classification, multi-attribute pattern matching

I. INTRODUCTION

In the daily life people sometimes encounter situations such as “You noticed the ladder which you had used was stepping on a hammer just after you got down.”, or “You almost bumped into someone at a corner.”. Those instances that “You’re almost hurt, but you weren’t” or “You got hurt, but the injury wasn’t severe.” were called Hiyari-Hatto (HH) in Japanese. HH is also called “incident” or “near-miss”.

According to Heinrich's law[1], there are many minor accidents and HHs before a major accident happens, and it is revealed that a major accident follows a minor accident, also a minor accident follows a HH. In other words, we can predict minor and major accidents to reflect on HHs. Many case studies are conducted manually[2][3][4][5][6]. Case study is an attempt to prevent a recurrence interpreting the case carefully, finding the cause, and improving the behavior etc. by analyzing the relevant information individually or multi-party discussions. Case studies have two types of purposes. One is to study HH likely to lead to a specific accident among the major accidents assumed in a certain scene, and promote the avoidance of the accident. The other is to predict what kind of major accident will occur, focusing on frequently occurring HHs.

In this study, medical-related HH is used, as well as daily and industrial accident's HH. The target of the medical-related HH include not only medical staffs, but also all visitors such as patients, patients' carers, sales persons. Falling or crashing occurs in the same way with daily and industrial accident's HHs, but there is a difference in that it also generates HHs related to medical equipment, medicine,

surgery, and inspection in. For example, “Delay in time of use due to incorrect settings or parts of the medical device.”, “Someone was about to prescribe the wrong drug.” “Almost perform an operation on the wrong patient.” or “Forgot to show what to do before the test.”. Each case may put patients at risk of danger. If they did not pay attention, a medical accident might have occurred. In the medical-related HH case study, an understanding of HHs for nursing students is investigated[2]. In this research, changes of an understanding to hazarder cognition and a medical accident before and after learning of the HH scene which was performed by the simulated patient were investigated. In this way, to prevent an accident in daily life, medical sites and education sites, many case studies of HHs are performed manually. While there are few experiments to process HH mechanically. Although manual case studies have a demerit like it cost too much to find a target case from a large amount of past cases.

In order to use for prediction of the accident risk which prevents serious accident and disaster, this research detects the information on a HH and classifies it according to the kind of accident by using the extracted HH expressions from the HH example currently exhibited. A multi-attribute matching machine[7] is used for detection and a classification. Detection of a HH sentence and the classification by an accident kind are carried out by the matching rule created from extraction data. Moreover, the tweet of Twitter is used for extraction of the HH expression because necessity to report speedily is high and the HH lurking in everyday life is able to acquire. The sentence include expression which leads to a HH is detected inside the tweet which contained a hash tag “#hiyari-hatto” and the “hiyari-hatto case” by keyword searches. By being classified simultaneously with detection, HH information is accumulated efficiently.

II. METHODOLOGY & COLLECTION AND EXTRACTION OF HH EXPRESSION

Collection of the experience case sentence include HH expression is explained. Although related research was aimed at the HH and accident which have occurred in the hospital to which oneself belongs, they are collected from the hiyari-hatto case currently released in order to collect HHs broadly in this research. The kind of accident is beforehand given to the collected HH cases. However, by the collecting source, since the accident kind was scattering, it unified to 14 kinds. The sentence of a HH case is shown in Table 1, and the number of the collected HH cases is shown in Table 2 according to accident kinds.

Revised Version Manuscript Received on September 14, 2019.

Kazuhiro Morita, Tokushima University, Japan
Kenzaburo Yashiro, Tokushima University, Japan
Masao Fuketa, Tokushima University, Japan

A CLASSIFICATION METHOD OF THE INCIDENT BY EXTRACTION FROM TEXT

Next, the HH expression for matching rule creation is extracted. Extraction of HH expression decomposes a sentence into morphemes, applying a case sentence to a morphological analysis first. From parsed morphemes, nouns, verbs, auxiliary verbs, and so on which lead to a HH are extracted divided into a total of 14 kinds of accident kinds. An example of extraction of HH expression is given.

Table 1. Example of HH cases

Accident kind	Example
Tumble	When one box of B4 size copy paper (five bundle of 500 sheets is entered) was raised among consultative-sales business at the second floor counter around 3 p.m., I lost the balance of the body, fell and nearly crashed into the angle of a desk.
Electric shock / Fire	When the repair work of the welding part was done in the ship block, the sparks of the grinder jumped into the can containing the paint for anticorrosives placed near.
Pinch	While cleaning near the chair, having held the frame of the door of a passenger car and supporting the body on the enclosure of a place of business around 9 a.m., the door was suddenly closed by the wind and the hand nearly pinched.
Cut / Rub	When the power was turned off and the hand was brought close to a ground meat exit, in order for a ground meat machine to wash a machine after the end of processing operation of meat, the finger almost touched a cutter under inertia rotation.

Table 2. Number of collected HH cases

Accident kind	Number of HH cases
Tumble	72
Electric shock / Fire	18
Pinch	43

Cut / Rub	20
Traffic accident	22
Crash	77
Fall	63
Touch with high / low temperature thing	9
Drop (Something comes down from above)	56
Collapse	15
Medicine	266
Medical treatment care	118
Operation	9
Others	18

“When it tries to send out the car which oil supply finished around 8:30 a.m. and I come out to a driveway, it seemed to be hit by a running car.”

In the case of this example, noting the portion of "be hit by a car", it turns out that it is HH expression which leads to a traffic accident from "be hit" with a "car." For this reason, the noun "car" is extracted as a vehicle, and it parses "be hit" into morphemes and extracts as HH expression which leads to a traffic accident.

Next, the matching rule creation using extracted HH expression is explained. Since the matching rule needs to detect many expression patterns with the small number of rules, a fundamental matching rule uses a matching element <separation>. This element can detect an object, even when arbitrary elements exist between matching elements. The example of a matching rule is shown in Fig. 1. Two or more expressions can be registered into one element (enclosed by <>), and combination of expressions included in elements is detected. At this time, by describing a <separation> element, even if the <vehicle> and the <traffic accident> are separated, it is detectable. Moreover, if there is expression included in each element like Example 1 and Example 2, it is detectable with the same rule. The matching rule of 14 kinds of accident kinds was created by making this rule into a basic form.

<vehicle> + <separation> + <traffic accident>

car	hit
bicycle	run over
truck	bump into
bus	
motorcycle	

Example 1: While driving the car, I nearly ran over the child who jumped out.

Example 2: When it was running by the bicycle, it was just going to bump into the pedestrian.

Figure 1. Example of matching rule



III. DETECTION AND CLASSIFICATION OF A TWEET INCLUDE HH EXPRESSION

By using the matching rule and multi-attribute matching machine which were created from the experience of a scary incident case sentence, the tweet including HH expression is detected from Twitter, and it classifies into 14 kinds of

accident kinds. About "others" of an accident kind, although classification information is insufficient, it is aimed at the sentence which includes HH expression of "it was chilled", "having been startled", etc., and has been detected as a HH. Fig. 2 shows the flow of detection and a classification.

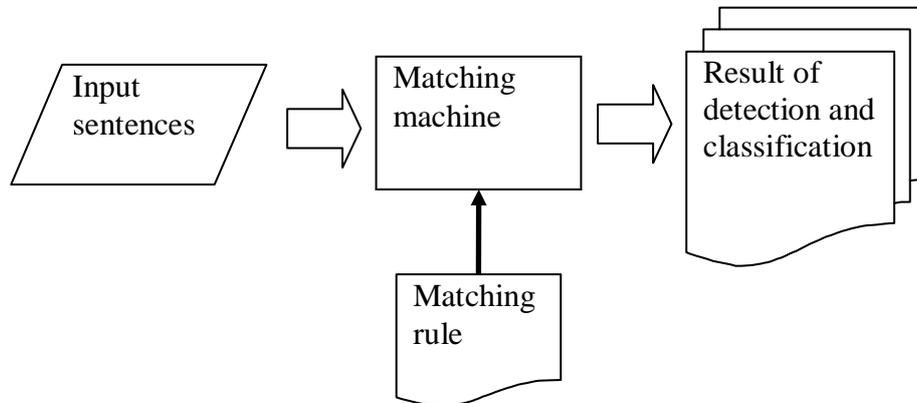


Figure 2. Flow of detection and classification

IV. EVALUATION & RESULTS

An experiment setup which evaluates detection / classification accuracy of the tweet including HH expression is explained. Tweets were collected on condition of the following as experimental data.

- The tweet including a hash tag "#scary minor incident" or "#Hiyari-Hatto"
- The tweet containing a keyword "today's scary minor incident" and "today's scary minor incident"

Moreover, the tweet applicable to the following rules is beforehand excepted from the candidate for collection.

- Reply (users' conversation)
- Retweet (quotation of an utterance of the others)
- The tweet containing URL
- The tweet including "periodical", "automatic", and "diffusion hope"
- The tweet of Bot account
- The tweet which overlaps with the already collected tweets

200 tweets including the HH which does damage physically among the tweets collected on the above conditions were taken up by hand, and were used as experimental data. The example of the collected tweets is shown below.

“When coming out from an alley to a national highway, it almost ran over the pedestrian. I was fearful although not fortunately run over. Possibly consciousness was lost just for a moment.”

“It went into the running lane in a form which is thrust into the right-hand side of my car, and turned the wheel to the left involuntarily. If not avoided here, an accident was caused rightly.”

“The son's finger was pinched with the slide door of the car. Since it cooled immediately and a turgidity was not carried out, although there was bleeding, I looked at the

situation at the house. Although it became a scar, it was normal to a bone. It was good. Since it was an accident from bite-size inattention, I reflected on slack of mind intensely.”

Moreover, in order to evaluate a detection accuracy, the following experimental data was prepared.

- 200 tweets which do not include HH expression
- 480 HH examples used for matching rule creation

4.1. Evaluation of a detection accuracy

A detection rate expresses the rate which detected the sentence which included HH expression to input data. The evaluation result of a detection rate to experimental data is shown in Table 3.

Table 3. Evaluation results for detection rates

	Input Sentences	Detected Sentences	Detection rate
HH examples used for matching rule creation	480	450	0.937
Tweets including HH	200	97	0.485
Tweets not including HH	200	0	0.000

Since it was not detected from the tweet which does not include HH expression, it is thought to be robust for incorrect detection. However, even the HH example used for matching rule creation besides the tweet including HH expression brought a result with the omission in detection.

The cause can consider that the sentence which does not include a direct HH expression (a "mistake", "he forgets", etc.) is not detectable.



Since the matching rule is detected / classified on the basis of the "object of a scary minor incident" + "scary minor incident expression", it is difficult to detect the following sentences which predict a scary minor incident in consideration of the context.

“Indication were received and processed after the end of operating of long daywork. Since the patient whom the catheter blockaded as a result of bloody urine corresponded to immediately before at that time, the medicine which received indication was put into the patient's cart.”

Moreover, the scantiness of a rule to proper nouns is also considered to be the cause. An example of the tweet which was un-detecting is shown below.

“It comes to a T junction, since I would like to bend, speed is reduced slowly, and it is to the right. The Hiace passes at about 60 km in a crossing. I who was not able to check by a mirror turned the wheel vice versa suddenly, in order to avoid the Hiace who appeared suddenly.”

Although "Hiace" is a subject in this example, common nouns, such as a "car", a "motorbike", and a "track", are used for the matching rule. For this reason, proper nouns, such as a vehicle type and a brand name, are un-detecting, without corresponding. In order to prevent such un-detecting, it is possible to create the matching rule for which it does not depend on a proper noun. If it is the above-mentioned example, the detection rule combined with HH expressions, such as "suddenly" and "avoid", will be created by making a subject into a "T junction", a "crossing", etc.

4.2. Evaluation of a classification accuracy

Precision shows the rate which the detected sentence has classified correctly. The evaluation result of precision to experimental data is shown in Table 4.

Table 4. Evaluation results for precision

	Detected sentences	Correctly classified sentences	Precision
HH examples used for matching rule creation	450	392	0.871
Tweets including HH	97	76	0.783

Although it was a mostly good result, the case where it was incorrect-classified was seen. It is thought that classification information was lacking as a cause. The sentence which there is no subject even though it includes HH expression, or contained the verb which is not in a matching rule is classified into "others." An example is shown below.

“It is windy today. It was buffeted along the sea, and I was also chilled.”

Because a subject is not found although "chilled" was detected as HH expression, this example was classified to "other." It turns out that the right classification becomes a "fall" from "windy" and "buffeted." In order to prevent such an incorrect classification, the rule classified from "expression which leads to a hazarder", and "expression showing a situation" is required.

Moreover, the case where it was incorrect-classified into both "medical treatment care" and a "medicine" was seen.

Because the matching rule has given classification information for every rule, if it matches two or more rules, it will be classified into the all. The solution determined as a specific classification can be considered by setting a priority to each rule.

V. CONCLUSION

In this paper, detection and the classification technique of the HH sentence using multi-attribute matching were proposed. As a result of the evaluation experiment which performs detection and a classification with the matching rule created using the collected experience of a scary incident case sentences, although classification accuracy was mostly good, the room for an improvement for extraction accuracy was seen.

Future works are creation of the matching rule not dependent on a proper noun, extension of a classification rule, and a setup of a priority.

VI. ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number JP16K00372.

REFERENCES

1. H. W. Heinrich and E. R. Grannis, Industrial accident prevention: a scientific approach; 4th ed., McGraw-Hill, New York, NY (1959)
2. T. Yoneda, K. Itami, O. Yasuhara, K. Seki, Y. Kawabata, T. Maesako and L. Zhe, Changes in Subjective Understanding of an Accident and Risk Awareness in First-Year Nursing Students Following Medical Accident Simulation-Based Experimental Learning, Proceedings of 2017 International Conference of Educational Innovation through Technology (EITT), Osaka, (2017), pp. 159-164.
3. K. Seta, T. Okazawa, M. Umamo and M. Ikeda, Ontology Based Organizational Risk Knowledge Creation Support Based on Incident Reports, Proceedings of the 11th WSEAS International Conference on SYSTEMS, Agios Nikolaos, Crete Island, Greece, July 23-25, (2007), pp.337-342.
4. Abe, F. Naya, H. O. Itoh, K. Sagara, N. Kuwahara and K. Kogure, Risk Management by Focusing on Critical Words in Nurses' Conversations, Knowledge-Based Intelligent Information and Engineering Systems, pp.1167-1173, Springer Berlin Heidelberg, Berlin, Heidelberg, (2005).
5. Abe, H. O. Itoh, N. Kuwahara and K. Kogure, Scenario Violation in Nursing Activities: Nursing Risk Management from the Viewpoint of Chance Discovery, Soft Computing, 11, 8, pp.799-809, (2006).
6. H. Igarashi, M. Fukushi, N. Shinoda, A. Miyamoto, M. Hirata, M. Ishidate, M. Kuraishi and K. Doi, Investigation of errors by radiological technologists and evaluation of preventive measures: general and mobile X-ray examinations, Radiological Physics and Technology, 3, 2, pp.136-143,(2010).
7. K. Ando, M. Okada, M. Shishibori and Jun-Ichi Aoe, Efficient multi-attribute pattern matching using the extended Aho-Corasick method, 1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation, Orlando, FL, USA, (1997), pp. 3936-3941 vol.4.



AUTHORS PROFILE



Kazuhiro Morita received B.Sc., M.Sc. and Ph.D. degrees in information science and intelligent systems from Tokushima University, Japan, in 1995, 1997 and 2000, respectively. He had been a research assistant from 2000 to 2006 in information science and intelligent systems, Tokushima University, Japan. He is currently an

associate professor in the Department of Computer Sciences, Graduate School of Technology, Industrial and Social Sciences, Tokushima University, Japan. His research interests are sentence retrieval from huge text databases, double-array structures and binary search tree.



Kenzaburo Yashiro received B.Sc. degree in information science and intelligent systems from Tokushima University, Japan, in 2018. He is currently a master course student at Tokushima University.



Masao Fuketa received B.Sc., M.Sc. and Ph.D. degrees in information science and intelligent systems from Tokushima University, Japan, in 1993, 1995 and 1998, respectively. He had been a research assistant and an associate professor from 1998 to 2000 and from 2000 to 2015 in information science and intelligent systems, Tokushima University, Japan, respectively. He is currently a

professor in the Department of Computer Sciences, Graduate School of Technology, Industrial and Social Sciences, Tokushima University, Japan. He is a member of the Information Processing Society in Japan and the Association for Natural Language Processing of Japan. His research interests are information retrieval and natural language processing.