Prediction of Accident by Using Decision Tree and Display Accident Information

Takahiro Sasaki, Katsutoshi Kanamori, and Hayato Ohwada
Tokyo University of Science, Japan
Email: 7413607@ed.tus.ac.jp

***Abstract*—Recently, product accidents have been occurring frequently. We developed a system that predicts injury that is likely to occur and displays accident information in an effort to reduce accidents. To develop this system, we use patient data in the National Electronic Injury Surveillance System (NEISS). NEISS data are collected from hospitals and include such information as how the patient was injured and a description of the situation. We used this information to create decision trees and to display past accident information. The decision tree is used for prediction.**

***Index Terms*—decision tree, Surveillance system, Product accident**

I. Introduction

Recently, product accidents have been occurring frequently [1], [2], with 400,000 product accident reports registered in the National Electronic Injury Surveillance System (NEISS) [1], [3], [4] in one year.

Studies for reducing such product accidents focus on companies, such as reviewing the product design [1], [5]. However, no research has focused on users. Many product accidents occur due to carelessness or incorrect usage [5]. In order to reduce product accidents, before using a product the customer must be aware of the likelihood of a product accident. Therefore, we developed a system that predicts injury that has a high possibility of occurring and displays accident information for reducing accidents. In order to develop this system, we use a decision tree created with information on NEISS. These decision trees predict the injury that is likely to occur from consumers' features and the product that will be used. Then to use decision trees' predictions and consumers' feature and the product in order to display accident information.

II. System

This system consists of a preparation part and a system part. The processing is showed as Fig. 1. In the preparation part, a decision tree is created based on NEISS data. We used as an example data on one incident in NEISS (Fig. 2). NEISS includes 400,000 such incidents per year, and we can easily obtain a CSV of these data [4]. In the system part, an injury is predicted based on a decision tree, age, gender, and product. Then past accident information nearing the user of a system are displayed.

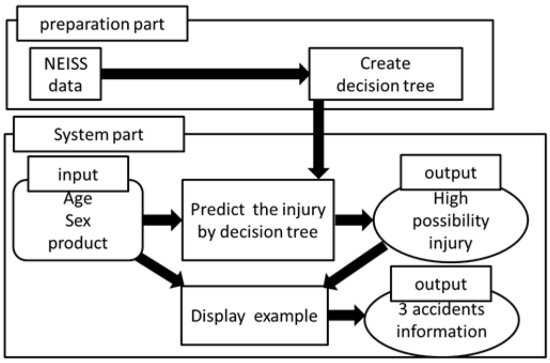

Figure. 1. System summary

In this system, the decision tree was created using data on 792,498 product accidents collected by NEISS in 2009 and 2010. The data that will be displayed as a past accident information are choose from NEISS 2009 and 2010.

A. Preparation

In this section, we describe creation of the decision tree used for prediction. The decision trees are created using J48 algorithm by weka [6]. The decision trees' attributes include past accident information: patients' age, patients' gender, and products that caused the accidents.

The decision trees' class is diagnosis of injury. Use of this decision tree allows prediction of injury, considering the possibility that it will happen based on the user's gender and age, and the product used.

Before creation of the decision tree, NEISS data of 2009 and 2010 are pretreated. First, we extract patients' age and gender, diagnosis of injury, and the product leading to the accident Age is divided into six groups (Table I).

TABLEI: Age Groups

Group	1	2	3	4	5	6
Age	0~	6~	14~	20~	40~	60~

1) Decision tree creation

The decision trees' attributes are patients' age and gender, and the product leading to an accident. Class is patients' injury. Numbers of attributes and class are indicated in Table II.

Manuscript received May 15, 2014; revised August 4, 2014.

doi: 10.12720/lnit.2.2.186-190

TABLE II. NUMBER OF ATTRIBUTES AND CLASS

Attributes			Class
Age	Gender	Product	Injury
6	3	798	30

Here, we create three decision trees: tree A, tree B, and tree C (in order of creation). Tree B is created in order to predict a highly likely injury that tree A did not predict. Tree C is created in order to predict a highly likely injury that trees A and B did not predict. These three decision trees were created as depicted in Fig. 3.

CPSC Case #: 130108519 Treatment Date: 01/01/2013 PSU: 58 Weight: 14.8537 Stratum: V
Age: 26 - 26 YEARS Sex: 2 - FEMALE Race: 0 - N.S. Race Other:
Diagnosis: 41 - INGESTED OBJECT Diag Other:
Body Part: 0 - INTERNAL
Disposition: 1 - TREATED & RELEASED, OR EXAMINED & RELEASED WITHOUT TRTMNT
Location: 1 - HOME Fire Involvement: 0 - NO FIRE OR NO FLAME/SMOKE SPREAD
Products: 1616 - JEWELRY
Narrative: A 26 YR OLD FE ACCIDENTALLY SWALLOWED HER LIP RING DX FOREIGN BODY GI TRACT

Figure. 2. NEISS data example.

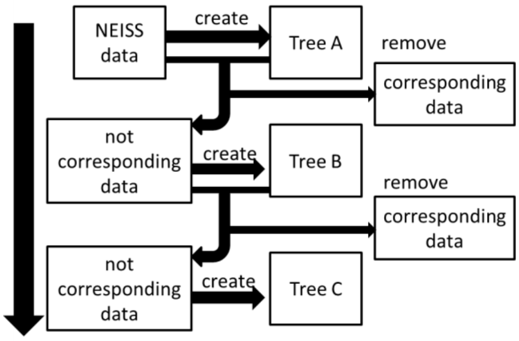

Figure. 3. Flow of decision tree creation

First, decision tree A is created from NEISS data. Next, using tree A and the data used for creation of tree A (age, gender, and product) output injuries. The data used to predict injury and the corresponding diagnosis are removed. Tree B is created from the remaining data. Tree C is created similarly. We then use trees A, B, and C as the system part.

B. System Part

Output values from user information (age, gender, and product) and decision trees A, B, and C are treated as prediction of injury that is likely to occur [7]. We explain the system which display previous accident data of NEISS (2009, 2010).

1) Prediction of injury

Figs. 4, 5, and 6 depict parts of decision trees A, B, and C. For example, for product 102, the age group is 1; the gender is male; tree A output is 59; tree B output is 54; and there is no tree C output. For product 106, the age group is 1, the gender is male, tree A output is 59, tree B output is 67, and tree C output is 65. These outputs indicate the likelihood that an injury will occur.

In some cases, there were fewer than three outputs, or different decision trees output the same injury prediction. In this system, if trees' outputs are fewer than three, we do not add other predictions; we use only trees' outputs. In addition, if different decision trees output the same injury prediction, they are treated as only one output.

For example, if tree A output is 59, tree B output is 59 and there is no tree C output, we use only 59 for injury prediction.

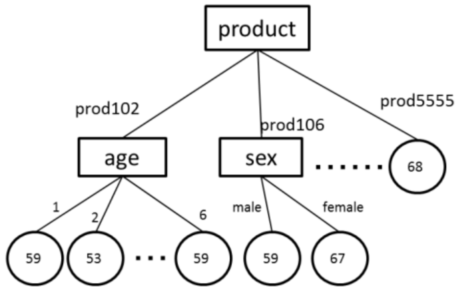

Figure. 4. Part of tree A

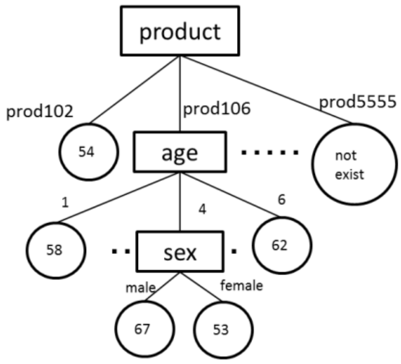

Figure. 5. Part of tree B

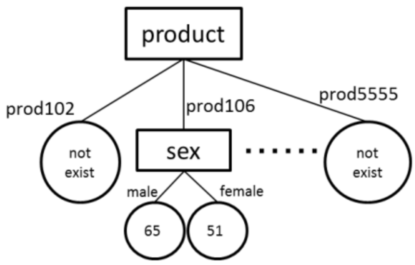

Figure. 6. Part of tree C

2) Display past accident information

We display past accident data using injury predictions output by decision trees and user information (age, gender, and product). The accident information is displayed one by one prediction.

The accident information is data in NEISS 2009 and 2010. The displayed information includes diagnosis of injury, injured body part, date when accident occurred, product that caused the accident, patient age, patient gender, and narrative.

Displayed accident information are selected information that have the patient closer to the user. First, in data that meets the following: ((1) patient used the product that was chosen in system (2) patient suffered the same injury that was predicted by decision tree), system count and find the data have most frequently injured body part (3). In data satisfying (1) (2) (3), this system display 1 data in priority order as listed below:

- Same age group and gender
- Same age group but different gender
- Different age group but same gender
- Different age group and gender.

If some data have the same priority, then the system displays recent data. System displays one by one each prediction by the method described above.

III. Experiment Result

A. Decision Tree

Next we write decision tree information. Trees were developed using NEISS 2009 and 2010 data (Table III).

Decision tree A was developed using 792,498 data (all NEISS 2009 and 2010 data), the number of leaves is 3683, and the size of the tree is 4588. Decision tree B was developed using 475674 data, the number of leaves is 4116, and the size of the tree is 5203. Decision tree C was developed using 319,329 data, the number of leaves is 4090 and the size of the tree is 5182.

TABLE III. TREE INFORMATION

	Tree A	Tree B	Tree C
Input	792,498	3683	4588
Leaves	475,674	4116	5203
Size	319,329	4090	5182

B. Prediction Accuracy

In order to examine prediction accuracy, we compare prediction and actual diagnosis of injury using 10,000 data chosen randomly from NEISS 2011 data. For comparison, we conducted experiments with prediction from injury incidence for each product. It is that treated as predict the top three injury incidence for each product in the accident information in 2009, in 2010 without considering patients' age and gender.

First, we consider the accuracy of the decision trees. For tree A, the number of predictions is 9999, and the number of agreements between prediction and actual diagnosis of injury is 3813; thus, accuracy is 38.1%. For tree B, the number of predictions is 9975, and the number of agreements between prediction and actual diagnosis of injury is 1887; thus, accuracy is 18.9%. For tree C, the number of predictions is 9926, and the number of agreements between prediction and actual diagnosis of injury is 1395; thus, accuracy is 14.1% (Table IV).

TABLE IV: ACCURACY OF DECISION TREES

	tree A	tree B	tree C	Total
Prediction	9999	9975	9926	29900
Agreement	3813	1887	1395	7095
Precision	38.1%	18.9%	14.1%	

Next, we compare predictions from incidence and prediction by decision tree (Table V).

TABLE V. COMPARISON OF TREE AND INCIDENCE

	Decision tree	Incidence
Input	10000	10000
Output	9999	9999
Prediction	29,900	29,997
Agreement	7095	6778
Accuracy	71.0%	67.8%

To calculate prediction accuracy, we divided the total number of agreements between prediction and actual diagnosis of injury by the number of inputs.

The number of outputs is 9999 for both decision tree and incidence. The number of outputs (9999) is less than the number of inputs (10,000), because NEISS 2011 included new product data. Next, in prediction by decision tree, the number of predictions is 29,900, and the number of agreements is 7095; thus, accuracy is 71.0%. In prediction from incidence, the number of predictions is 29,997, and the number of agreements is 6778; thus, accuracy is 67.8%. The number of predictions by decision tree is less than that by incidence, and the number of agreements of prediction by decision tree is more than that by incidence.

These results indicate that use of the decision tree to predict injury is valid and decision trees developed with removing data that match are also effective.

C. Extracting Accident Information

We evaluate displayed accident information using data on 100 injury incidents in NEISS 2011. For each incident, we enter the age, gender, and the product that caused the accident, and then compare information on the actual incident and the prediction. If the accident happened in similar circumstances, we define it is the good answer.

Among data whose prediction of injury match the actual diagnosis, look at the items in the details of the accident. Decisions were based on actions and circumstances that caused accidents.

TABLE VI. BREAKDOWN OF DATA

	Total	Good	Bad
Match	74	50	24
Not match	26		

First, 74 predictions are consistent with diagnosis of actual injuries. There are 50 good displays and 24 bad displays (Table VI).

In addition, 27 predictions are consistent with actual diagnoses and injured body parts, and there are 24 good displays (Table VII). Therefore, it can be said that part of the body that was injured and diagnosis of injury if the match, similar information to almost be able to display. In addition, result of the experiment, less prone displayed accident similar enough high severity was observed. We suspect this is because the number of occurrences little big accident than a small accident.

TABLE VII. BREAKDOWN OF DATA (BODY IS MATCH OR NOT)

	Total	Good	Bad
Body part match	27	24	3
Not match	47	26	21

TABLE VIII: EXAMPLE OF INPUT

age	Gender	Prod
22	1	381
	1 male, 2 females	AIR CONDITIONERS

TABLE VIIII: EXAMPLE OF OUTPUT

Diag	STRAIN, SPRAIN	Narr
Diag other		38 YOM LIFTED A HEAVY AIR CONDITIONER AND PULLED MUSCLE IN LOW BACK DX. LOW BACK SPRAIN
Body part	LOWER TRUNK	
Date	2010/9/11	
Prod	AIR CONDITIONERS	
Age	38	
Gender	Male	
Diag	LACERATION	Narr
Diag other		39 YOM DX RT INDEX FINGER LACERATION - S/P PT C/O PARTIAL NAIL AVULSION DUE TO AN AIR CONDITIONER ACCIDENTALLY CUT FINGER.
Body part	FINGER	
Date	2010/7/11	
Prod	AIR CONDITIONERS	
Age	39	
Gender	Male	
Diag	NOT STATED	Narr
Diag other	LOW BACK PAIN	LOW BACK PAIN/39YOM C/O LOW BACK PAIN 3 DAYS AFTER LIFTING AIR CONDITIONER.
Body part	LOWER TRUNK	
Date	2010/10/14	
Prod	AIR CONDITIONERS	
Age	39	
Gender	Male	

IV. RESULTS AND DISCUSSION

Table VIII presents examples of input, and Table VIIII presents examples of accident information of output prior to the input of Table VIII. In this case, for the input, the age is 22, the gender is male, and the product number is 381 (AIR CONDITIONERS). For the output, the victims' ages are 38, 39 and 39; and all victims are male. Because group 4 includes ages 20 through 39 years, both input and output are the same age group. Thus, cases of victims who are similar to the input user can be displayed.

In the narratives of the first and third incidents, similar contents are displayed. We assume that caused the "other" of diagnosis. It includes those mild symptoms of other diagnostics. In this case, the symptom is sprain of the lower trunk in the first incident, but low back pain in the third incident. In addition, the injury in the second incident occurred during use, whereas those in the first and third incidents did not occur during use. Since we made this system by assuming the use, it can't be said that the appropriate output. We are considering that if we use when using or preparation to predict, accuracy goes up more.

V. CONCLUSION

In this study, we developed a system to predict injury and display accident information for preventing accidents. Results indicated that with new type products, the decision tree cannot output prediction. This problem could be solved by using simple features of the product. In addition, similar situations from one input are output often. In future work, we would like to adapt to these problems.

REFERENCES

[1] K. Zhang, K. Nakahira, and Y. Mikam, "Making of a risk-matrix based on product accident data–its application to child toy accident," *Social Technology Research Papers*, vol. 7, pp. 66-75, Mar. 2010

[2] N. Furuichi, K. Suzuki, Y. Ihara, T. Kirikara, H. Onoda, and K. Nagata, "Construction and analysis of the product safety database," *Transactions of the Japan Society of Mechanical Engineers, Symposium on Environmental Engineering*, vol. 20, pp. 163-166, 16 June, 2010.

[3] K. Zhang, J. X. Wang, K. Nakahira, and Y. Mikam, "A study on the construction of an injury system," *Social Technology Research Papers*, vol. 8, pp. 111-123, Apr. 2011

[4] National Electronic Injury Surveillance System. [Online]. Available: http://www.cpsc.gov/library/neiss.html

[5] K. Suzuki and K. Aoki, "Nine principles for error prevention to avert safety problems at customer usage stage," *Journal of the Japanese Society for Quality Control*, vol. 39, no. 4, pp. 79-91, 2009.

[6] Weka 3: Data Mining Software in Java. [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka/

[7] S. Akama, Understanding Data Mining, Kougakusha.

Takahiro Sasaki holds a Master of Industrial Administration, Tokyo University of Science. He received a degree from Tokyo University of Science in 2012.

Katsutoshi Kanamori is an Assistant Professor of Industrial Administration, Tokyo University of Science. He received a Doctor of Science from Tokyo University of Science in 2009. He has been working on Artificial Intelligence and Formulation of Creativity.

Hayato Ohwada is a Professor of Industrial Administration and the Director of the Division of Next Generation Data Mining Technology, Tokyo University of Science, E-mail: ohwada@rs.tus.ac.jp. He received a Doctor of Engineering from Tokyo University of Science 1988. He has been working on Machine Learning and Inductive Logic Programming.