Formal concept analysis of medical incident reports

Baba, Takahiro
Graduate School of Information Science and Electrical Engineering, Kyushu University

Liu, Lucing
Graduate School of Information Science and Electrical Engineering, Kyushu University

Hirokawa, Sachio
Research Institute for Information Technology, Kyushu University

https://hdl.handle.net/2324/820673

出版情報：Knowledge-Based and Intelligent Information and Engineering Systems (Lecture Notes in Artificial Intelligence). 6278 LNAI (PART 3), pp.207-214, 2010-11-23
バージョン：
権利関係：

Formal Concept Analysis of Medical Incident Reports

Takahiro Baba[1], Lucing Liu[1] and Sachio Hirokawa[2]

[1] Graduate School of Information Science and Electrical Engineering,
[2] Research Institute for Information Technology
Kyushu University, Fukuoka 812-8581, JAPAN
hirokawa@cc.kyushu-u.c.jp http://matu.cc.kyushu-u.ac.jp

Abstract. It is known that a lot of incidents has happened ahead of a serious accident. Such experiences have been collected in medical sites as incident reports. The text mining is expected as a method that discovers the factors of incidents and the improvement of the situation. This paper proposes a method to analyse the co-occurrence relation of the words that appear in the medical incident reports using concept lattice.

1 Introduction

Organizational efforts to collect incident reports are being made in hospitals, factories, traffic controls and network security where a small mistake would cause a terrible damage to the society or to individuals. Incident reports contain detailed situations which are very close to accidents. The incident reports contain not only the situation and the reason of the incident, but also how it was prevented to occur an accident. So, it is worthwhile to analyse incident reports to discover some hits to prevent accidents.

In the field of medicine, even a small mistake may cause a death of an client. Many hospitals are collecting these incident reports systematically. Indeed, the ministry of health and labor of Japanese government published the announcements #0330008 of medical policy, and #0330010 of medicine and food, to strengthen the activity to file the incident reports. However, analysing the reports is time consuming hard work for doctors and nurses who are working the actual situation. They do not have separate time to consider the report deeply. Nowadays, the reports are being kept as digital texts from the beginning. Hence, the number of reports are increasing [6]. There are strong needs in introducing ICT to support analysing reports.

In [8], keyword extraction methods are applied to incident reports to analyse particular reports that are specified by a metadata. In [7], the method of SOM(Self Organizing Map) are applied to discover similar reports. However, the method of [8] is restricted to metadata. The method of [7] does not explain the meaning of documents with characteristic words. Neither of the methods provide interaction with the user. There are many researches, e.g. [2, 3], for clustering documents. However, the most of these approach adapts vector space model to represent document. The results vary according to the definition of similarity of the vectors and to the threshold. There is no general rule to justify how we formulate.

In this paper, we demonstrate that the theory of FCA (formal concept analysis) is applicable to the analysis of incident reports.

2 Incident Reports

An incident report describes an experience of doctors and of nurses, where they faced a dangerous situation that was very close to an accident. The reports contain meta data such as date, time, place, type of an incident, the name of person who reports. They contain free texts that explain the situation. Table 1 is a typical example. Incident reports are similar to accident reports in nature. Both describe the situation of the accident/incident and the considerable factors of the accident/incident. The most important difference between incident reports and accident reports is that the former may contain some reason that prevented an accident. There are many lessons that we can learn from these incidents.

It is known, as the Heinrig law, that there are much large number of incidents behind an accident. So, we can expect in collecting large number of incident reports than that of accident reports, if we do our efforts.

Table 1. An Example of Incident Report

date	yyyy/mm/dd
location	Medical Examination Room
contents	The syringe drivers for vein injection and for epidural injection were in the same tray. I almost gave the wrong injection by mistake.

This paper analyses the 47 incident reports that are described and analysed in the book [5]. The reports contain not only free texts but also metadata that specify the causes of the incident and possible improvement factors. The table 2 shows these metadata.

3 Related Keywords for metadata and their Weight

The analysis is based on the term*document matrix. Firstly, the set of the documents that contain a metadata are searched using the matrix. Then, the top 5 words of highest weight are retrieved using the matrix conversely as related words for the metadata. Table 3 shows the related words for each metadata. For example, the characteristic words for the metadata x:C, which indicates the cause of incidents in the organization, are drip, injection, bottle, direction and

Table 2. Metadata for Cause and Improvement

x:A	Patient and their environment
x:B	Stuff and their experience
x:C	Organization
x:D	Other
y:1	Communication between Stuff
y:2	Design of Commodity
y:3	Design of Equipment and Operation
y:4	Maintenance
y:5	Inspection
y:6	Nursing Procedure
y:7	Paper Work
y:8	Information Sharing
y:9	Physical Environment
y:10	Workplace
y:11	Personnel
y:12	Management of Equipment and Medicine
y:13	Other Management Issue in Hospital
y:14	Educational Activity
y:15	Organizational Culture

patient. We might imagine a situation where a doctor gives a direction to a nurse to make an injection to a patient, or to connect a bottle to "drip". Without these actual words, we cannot think of the real situation from the metadata "x:C". In this sense, these related word is useful. However, we do not know if the 5 words is enough or not. There is no criteria to determine the number of related words. There is no justification to determine the threshold of the weight.

4 Analysis System using Formal Concept Lattice

4.1 Formal Concept Lattice

Given a set of documents and the set of keywords that appear in the documents, the formal concept lattice represents the relationship of documents and keywords and further, represents the hierarchical structure of keywords.

In the theory of concept lattice [1, 4], a tuple (G, M, I) of a finite set G of objects, a finite set M of attributes and a relation $I \subseteq G \times M$ is called a context. When an object g has an attribute m, we denote $(g, m) \in I$ or gIm. Given an object g, the set of all attributes of g is represented as $nbr(g)$, i.e, $nbr(g) = \{m \in M \mid (g, m) \in I\}$. For a set of objects $X \subseteq G$, by $attr(X)$, we denote the set of attributes that are common to all objects g in X, i.e.,$attr(X) = \cap_{g \in X} nbr(g)$. Given an attribute $m \in M$, the set $nbr(m)$ of objects that has the property is defined by $nbr(m) = \{g \in G \mid (g, m) \in I\}$. For a set of attribute $J \subseteq M$, by $obj(J)$, we denote the set of all objects that have all the attributes in J, i.e., $obj(J) = \cap_{m \in J} nbr(m)$. A pair (A, B) of a set $A \subseteq G$ of objects and a set $B \subseteq M$

Table 3. Top 5 Related Words for Metadata

freq	cause/improvement	word(weight)
32	x:C (organization)	drip(8.375) injection(8.171) bottle(8.033) direction(7.761) patient(7.553)
21	x:B (stuff)	injection(7.125) inject(6.876) confirm(6.578) ample(6.418) new stuff(6.217)
5	x:A (patient)	fit(3.834) spill(2.558) slip(2.558) return(2.558) reach(2.558)
12	y:6 (nurse)	nurse(5.101) prepare(4.558) tube(3.767) merge(3.546) mistake(3.522)
10	y:8 (information sharing)	told(6.057) bottle(4.643) new stuff(4.620) drip(4.376) intramuscular injection(4.117)
8	y:14 (education)	new stuff(6.217) ample(4.934) location(4.164) told(4.117) intramuscular injection(4.117)
5	y:12 (eq. management)	location(4.164) same(3.577) inject(3.268) connect(2.477) color sringe(2.477)
3	y:7 (paper work)	sharp(3.784) infection(2.249) without notice(2.249) break(2.249) duty(2.249)
2	y:1 (comm. stuff)	cards(2.477) "ha"(2.477) lubrine(2.477) se(2.477) check(2.477)

of attributes are said to be a concept iff $A = obj(B)$ and $B = attr(A)$. An order relation $\prec$ for two concepts $C_1 = (A_1, B_1)$ and, $C_2 = (A_2, B_2)$ are defined by $C_1 \prec C_2 \iff (A_1 \subseteq A_2) \wedge (B_1 \supseteq B_2)$. When $C_1 \prec C_2$, C_2 is said to be a lower concept of C_1 and C_1 is said to be an upper concept of C_2. When $C_1 \prec C_2$ and there is no concept E except C_1 or C_2 such that $C_1 \prec E \prec C_2$, C_1 is said to be a direct lower concept of C_2 and C_2 is said to be a direct upper concept of C_1. When (A, B) is a concept, the set of objects A and the set of attributes B characterize each other. The direct lower concepts represent a clustering of the objects A according to attributes. The direct upper concepts represent a clustering of the attribute B according to objects.

Table 1 shows a context of reports D_1, D_2, D_3, D_4, and attributes $x : a, x : b, x : c, y : d, y : e$, where "a","b" etc represent words and "x","y" represent the category of the words. Fig 2 is the concept lattice of this matrix. We can see that the set of reports $\{D_3, D_4\}$ is characterized by the attribute $\{y : e\}$ and is classified into the two direct lower concepts D_3 and D_4 according to the attributes $x : a$ and $x : b$.

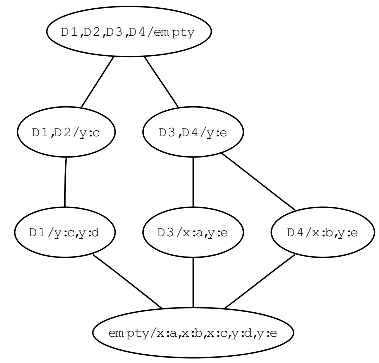
4.2 Clustering with Words and Clustering with Attributes

If we consider the reports as objects and the words that appear in the reports as attributes, we can construct a concept lattice by which we can analyse the relation of reports and keywords and the relation between keywords. However, the analysis should depend on the purpose of the analyser. The viepoint of the analysis is crucial. In the case of the incident report of Table 3, the metadata

Fig. 1. Object*Attribute Matrix

	x:a	x:b	y:c	y:d	y:e
D_1			$\surd$	$\surd$	
D_2			$\surd$		
D_3	$\surd$				$\surd$
D_4		$\surd$			$\surd$

Fig. 2. Concept Lattice for Table1



x:A,x:B,x:C,y:1,y:2,... and y:15 represent the view points. The metadata x:A, x:B and x:C classifies the cause of the incidents. The metadata y:1,..., y:15 represent the improvement points to prevent the incidents.

A concept lattice is constructed from a context matrix that represents the relation of the objects and the attributes. A novelty of our approach is that we do not consider the documents as the objects. The attributes of the incident report are chosen as the objects. Thus, the row of the context matrix consists of attributes and the columns consists of terms. In other words, we use the term*attribute matrix as the context for the concept lattice.

Let X be the term*document matrix where $X[i, j] = 1$ iff the word or the attribute w_i occurs in a document d_j. To construct the term*attribute matrix, we have to determine the set of words that correspond to an attribute $x : P$. There are two way to formulate the set. The standard formulation (disjunctive construction) is to choose all the words in the documents that contains the attribute $x : P$. Another formulation (conjunctive construction) is to choose the words that appear in all documents that contain the attribute $x : P$.

Given a term*document matrix X, we construct attribute*term matrices Y (disjunctive construction) and Z (conjunctive construction) as follows. Here, the document d_k ranges under the condition $x : P_i \in d_k$ and $y : P_j \in d_k$. In this paper, we adapt the disjunctive construction.

$$Y[x : P_i, y : Q_j] = \Pi_{d_k}\{X[x : P_i, d_k] * X[y : Q_j, d_k]\}$$
$$Z[x : P_i, y : Q_j] = \Sigma_{d_k}\{X[x : P_i, d_k] * X[y : Q_j, d_k]\}$$

We constructed a system that accepts words and attributes as input, and that outputs the concept. The user can combine multiple words to form a "AND query" and "OR query". The system searches the documents that satisfy the query q, and then obtain the set $obj(q)$ of attributes that co-occur with the query q. Be Aware that the objects of the concept are metadata x:A,x:B,x:C,y:1,y:2,..., and y:14 and that the attributes are keywords. The system retrieve the set $attr(obj(q))$ keywords that co-occur with all of the metadata in $obj(q)$. The pair $(obj(q), attr(obj(q)))$ is the concept for the query.

A user can use the system iteratively and interactively. Once a user send an input, the system displays related keywords and attributes. The user only has to click one of the words for further analysis. The user easily can reach an upper concept and a lower concept. An adjacent concept is shown with a keyword which is used as ankar text. So, the user can obtain the corresponding concept simply by clicking the keyword.

5 Case Studies

5.1 Concepts for Cause

Table4 shows the concepts for each metadata. All words are listed in a line if they belong to the concept. This is the most crucial distinction compared to Table3 of basic analysis where only top 5 words are chosen according to their weight. For example, the metadata m6 that corresponds to the nurse procedure is completely characterized by the metadata x:B,x:C,y:12,y:14,y:2,y:6,y:8 and other 75 words according to the conjunctive construction. On the other hand, no words appear in the concept except for y:1 in the disjunctive construction. In other words, the disjunctive construction is too restrictive that the metadata for cause and improvement cannot be characterized by words. However, the relation of metadata can be seen in the disjunctive construction. For example, $\{x:B, x:C\}$ and $\{x:A, x:C\}$ form concepts, but $\{x:A, x:B\}$ does not. This implies that no incident is observed with respect to the patient and the stuff. They are observed only when the relation of stuff is concerned.

Each case is worthwhile to analyse. The concept for metadata mB, that represents some cause of incident in stuff, contains mC that represents organizational cause. This can be interpreted that we should pay much attention to the relationship of stuff than the individual problems of stuff to prevent incidents. Organizational improvement is expected to solve the individual problems.

The concept for y:14, that represents improvement by education, consists of x:B, x:C and y:14. This implies that no incidents that could be prevented are observed unless organizational problems are concerned.

5.2 Analysis by Cause and Improvement

Given an attribute of cause or improvement, we can obtain the related words that characterize the attribute. Thus, we can analyse depending on particular target of analysis. However, these detailed analysis do not give a birds-eye view of the whole reports. We cannot grasp the whole picture of the reports. The visualization of the lattice is useful for this purpose. Fig 3 is the concept lattice for cause*improvement. It is worthwhile to notice that the lattice is very small and that there are only 5 concepts. It is not because the number of the document is 47 and is very small. It is because there are only 3 attributes x:A,x:B and x:C as the attributes to describe the cause. The attributes are determined according to the purpose of the analysis. Therefor, the number of the attributes is considered

Table 4. The Concepts for Metadata

freq	metadata	Conjunctive Construction
32	x:C	A B C 1 11 12 14 2 5 6 7 8 one "5 minutes" IVH give after ...(378)
21	x:B	B C 1 11 12 14 2 6 7 8 one "5 minutes" IVH saying said ...(246)
5	x:A	A C 6 7 give after in convulsion spill ...(74)
12	y:6	A B C 12 6 8 one "5 minutes" IVH give always ...(178)
10	y:8	B C 12 14 5 6 8 "5 minutes" saying always say told ...(127)
8	y:14	B C 12 14 8 told understood up atoniun alarm ...(77)
5	y:12	B C 12 14 2 6 8 IVH saying told do connect ...(75)
3	y:7	A B C 7 after said in please moreover ...(69)
2	y:1	B C 1 not after-operation before-operation card Ha ...(29)

freq	metadata	Disjunctive Construction
32	x:C	C:organizational
21	x:B	B:stuff C:organizational
5	x:A	A:patient C:organizational
12	y:6	C:organizational 6:nurse
10	y:8	C:organizational 8:message
8	y:14	B:stuff C:organizational 14:education
5	y:12	B:stuff C:organizational 12:equipment management
3	y:7	C:organizational 7:official procedure
2	y:1	B:stuff C:organizational 1:"stuff communication" direction

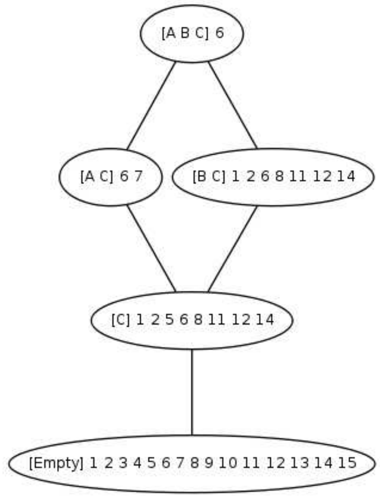
to small enough compared to the number of reports. Since the number of concepts are bound by the exponential of the number of attributes, we can expect that the concept lattice is small enough to draw.

6 Conclusion and Further Work

This paper proposed a method to analyse medical incident reports using concept lattice. The metadata that specify the cause of incidents or the possible improvements are considered as objects and the words are considered as attributes. For each metadata, The set of words that characterizes the metadata are obtained and analysed.

The number of incident reports that analysed in this paper is 47. They may be too small for general evaluation. However, each sample was selected from

Fig. 3. Concept Lattice for Cause*Improvement



expert point of view in the book [5]. They are valuable examples worthwhile to analyse in detail. Nonetheless, we need quantitative evaluation of the proposed method.

References

1. C. Carpineto, G. Romano, Concept Data Analysis Theory and Application, John Wiley and Sons, 2004
2. Y. Cheng, G.M. Church, Biclustering of expression data. Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology, pp.93-103, 2000
3. I. S. Dhillon, S. Mallela, D. S. Modha, Information-Theoretic Co-clustering, Proceedings of The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD), pp 89-98, 2003
4. B. Ganter, R. Wille, C. Franzke, Formal Concept Analysis:Mathematical Foundation, Springer, 1999
5. H. Kawamura, Willing to write incident reports – lessons learned to prevent accidents(in Japanese), Igaku-Shoin, 2000
6. K. Kaneko, Reports on activities in Kyushu University Hospital for Medical Safety Management Database (in Japanese), 12th Web Intelligence Workshop, 2008.
7. H. Kawanaka, Y. Otani, K. Yamamoto, T. Shinogi, S. Tsuruoka, Tendency discovery from incident report map generated by self organizing map and its development, IEEE International Conference on Systems, Man and Cybernetics, pp.2016-2021, 2007
8. T. Okabe, T. Yoshikawa, T. Furuhashi, A Proposal of Analysis System for Medical Incident Reports using Metadata and Co-occurrence Information (in Japanese), Journal of Knowledge, Information and Fuzzy Society, Japan vol.18,No.5,pp.689-700, 2006