

Data Mining for Risk Management in Hospital Information Systems

Shusaku Tsumoto and Shoji Hirano Department of Medical Informatics,
Shimane University, School of Medicine,
89-1 Enya-cho, Izumo 693-8501 Japan
Email: tsumoto@computer.org, hirano@ieee.org

Abstract

Organizations in our modern society grow larger and more complex to provide advanced services due to the varieties of social demands. Such organizations are highly efficient for routine work processes but known to be not robust to unexpected situations. According to this observation, the importance of the organizational risk management has been noticed in recent years. On the other hand, a large amount of data on the work processes has been automatically stored since information technology was introduced to the organizations. Thus, it has been expected that reuse of collected data should contribute to risk management for large-scale organizations. This paper proposes risk mining, where data mining techniques were applied to detection and analysis of risks potentially existing in the organizations and to usage of risk information for better organizational management. We applied this technique to the following three medical domains: risk aversion of nurse incidents, infection control and hospital management. The results show that data mining methods were effective to detection of risk factors.

1. Introduction

It has passed about twenty years since clinical information are stored electronically as a hospital information system since 1980's. Stored data include from accounting information to laboratory data and even patient records are now started to be accumulated: in other words, a hospital cannot function without the information system, where almost all the pieces of medical information are stored as multimedia databases. Especially, if the implementation of electronic patient records is progressed into the improvement on the efficiency of information retrieval, it may not be a dream for each patient to benefit from the personal database with all the healthcare information, "from cradle to tomb". However, although the studies on electronic patient record has been progressed rapidly, reuse of the stored data has not yet been discussed in details, except for laboratory

data and accounting information to which OLAP methodologies are applied. Even in these databases, more intelligent techniques for reuse of the data, such as data mining and classical statistical methods has just started to be applied from 1990's[2, 3].

Human data analysis is characterized by a deep and short-range investigation based on their experienced "cases", whereas one of the most distinguished features of computer-based data analysis is to enable us to understand from the different viewpoints by using "cross-sectional" search. It is expected that the intelligent reuse of data in the hospital information system provides us to grasp the all the characteristics of university hospital and to acquire objective knowledge about how the hospital management should be and what kind of medical care should be served in the university hospital.

This paper focuses on application of data mining to medical risk management. To err is human. However, medical practice should avoid as many errors as possible to achieve safe medicine. Thus, it is a very critical issue in clinical environment how we can avoid the near misses and achieve the medical safety. Errors can be classified into the following three type of errors. First one is systematic errors, which occur due to problems of system and workflow. Second one is personal errors, which occur due to lack of expertise of medical staff. Finally, the third one is random error. The important point is to detect systematic errors and personal errors, which may be prevented by suitable actions, and data mining is expected as a tool for analysis of those errors.

For this purpose, this paper proposes *risk mining* where data including risk information is analyzed by using data mining methods and mining results are used for risk prevention. We assume that risk mining consists of three major processes: risk detection, risk clarification and risk utilization, as shown in Section 2.

As an illustrative example, we applied risk mining process to analysis of nurses' incident data. First, data collected in 6 months were analyzed by rule induction methods, which detects several important factors for incidents (risk detection). Since data do not include precise infor-

mation about these factors, we recollect incident data for 6 months to collect precise information about incidents. Then, rule induction is applied to new data. Domain experts discussed all the results obtained and found several important systematic errors in workflow (risk clarification). Finally, nurses changed workflow to prevent incidents and data were recollected for 6 months. Surprisingly, the frequency of medication errors has been reduced to one-tenth (risk utilization).

This paper is organized as follows. Section 2 shows background of our studies. Section 3 proposes three major processes of risk mining. Section 4 gives an illustrative application of risk mining. Finally, Section 5 concludes this paper.

2. Background

A hospital is a very complicated organization where medical staff, including doctors and nurses give a very efficient and specialized service for patients. However, such a complicated organization is not robust to rapid changes. Due to rapid advances in medical technology, such as introduction of complicated chemotherapy, medical workflow has to be changed in a rapid and systematic way. Such rapid changes lead to malpractice of medical staff, sometimes a large-scale accident may occur by chain reaction of small-scale accidents.

Medical accidents include not only careless mistakes of doctors or nurses, but also prescription errors, intrahospital infections or drug side-effects. The cause for such accidents may not be well investigated and it is unknown whether such accidents can be classified into systematic errors or random errors. Since the occurrence of severe accidents is very low, case studies are used for their analysis. However, in such investigations, personal errors tend to be the cause of the accidents. Thus, it is very important to discover knowledge about how such accidents occur in a complicated organization and knowledge about the nature of systematic errors or random errors.

On the other hand, clinical information have been stored electronically as a hospital information system(HIS). The database stores all the data related with medical actions, including accounting information, laboratory examination, treatment and patient records described by medical staffs. Incident or accident reports are not exception: they are also stored in HIS as clinical data. Thus, it is now expected that mining such combined data will give a new insight to medical accidents.

3. Risk Mining

In order to utilize information about risk extracted from information systems, we propose risk mining which inte-

grates the following three important process: risk detection, risk clarification and risk utilization.

3.1. Risk Detection

Patterns or information unexpected to domain experts may be important to detect the possibility of large scale accidents. So, first, mining patterns or other types of information which are unexpected to domain experts is one of the important processes in risk mining. We call this process *risk detection*, where acquired knowledge is referred to as *detected risk information*.

3.2. Risk Clarification

Focusing on detected risk information, domain experts and data miners can focus on clarification of modelling the hidden mechanism of risk. If domain experts need more information with finer granularity, we should collect more data with detailed information, and apply data mining to newly collected data. We call this process *risk clarification*, where acquired knowledge is referred to as *clarified risk information*.

3.3. Risk Utilization

We have to evaluate clarified risk information in a real world environment to prevent risk events. If risk information is not enough to prevention, then more analysis is required. Thus, additional data collection is evoked for a new cycle of risk mining process. We call this process *risk utilization*. where acquired knowledge is referred to as *clarified risk information*.

Figure 1 shows the overview of risk mining process.

3.4. Elemental Techniques for Risk Mining

Mining unbalanced data. A large scale accident rarely occur: usually such it can viewed as a large deviation of small scale accidents, called incidents. Since even the occurrence of incidents is very low, the probability of large accidents is nearly equal to 0. On the other hand, most of the data mining methods depend on "frequency" and mining such unbalanced data with small probabilities is one of the difficult problems in data mining research. Thus, for risk mining, techniques for mining unbalanced data are very important to detect risk information.

Interestingness. In conventional data mining, indices for mining patterns are based on frequency. However, to extract unexpected or interesting knowledge, we can introduce measures for unexpectedness or interestingness to extract patterns from data, and such studies have been reported in data mining literature.

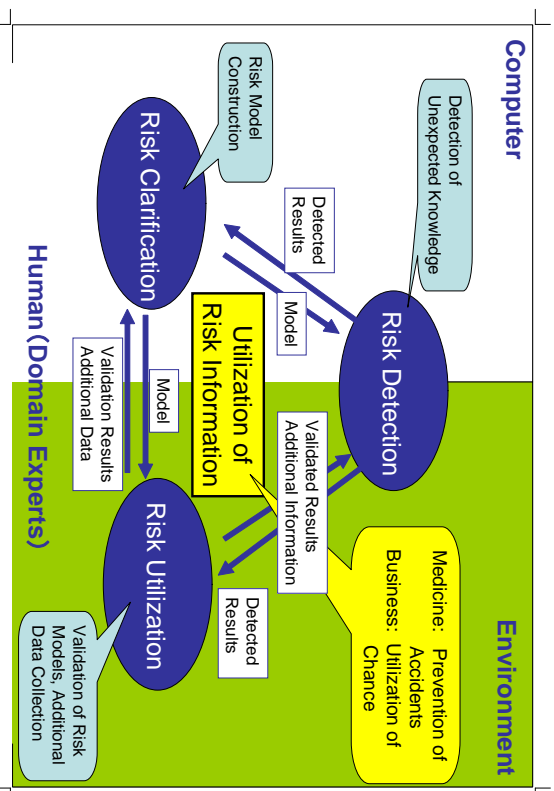


Figure 1. Risk Mining Process: Overview

Uncertainty and Granularity: Granular Computing. Since incident reports include information about human actions, these data are described by subjective information with uncertainty, where we need to deal with coarseness and fineness of information (information granularity). Granular computing, including fuzzy sets and rough sets, are closely related with this point.

Visualization. Visualizing cocurrence events or items may enable domain experts to detect risk information, to clarify the mechanism of risk, or to utilize risk information.

Structuration: Graph Mining. Risk may be detected or clarified only by relations between several items in a large network structure. Thus, extracting partial structure from network hidden in data is a very important technique, focusing on risk information based on relations between items.

Clustering. Similarity may find relations between similar objects which seems not to be similar. Or events which seems to occur independently can be grouped into several “similar” events, which enables us to find dependencies between events. For this purpose, clustering is a very important technique.

Evaluation of Risk Probability. Since probability is formally defined as a Lebeque measure on a fixed sample space, its performance is very unstable when the definition of sample space is unstable. Especially, when we collect data dynamically, such as instability frequently occurs. Thus,

deep reflection on evaluation of risk probability is very important.

Human Computer Interaction. This process is very important for risk mining process because of the following reasons. First, risk information may be obtained by deep discussions on mining results among domain experts because mining results may show only small part of the total risk information. Since domain experts have knowledge, which is not described in a datasets, they can compensate for insufficient knowledge to obtain a hypothesis or explanation of mining results. Second, mining results may lead to domain experts’ deep understanding of workflow, as shown in Section 4. Interpretation of mining results in risk detection may lead to new data collection for risk clarification. Finally, human computer interaction gives a new aspect for risk utilization. Domain experts can not only performance of risk clarification results, but also look for other possibilities from the rules which seems to be not so important, compared with rules for risk clarification and also evaluate the possibility to design a new data collection.

4. Case Study: Prevention of Medical Errors

4.1. Risk Detection

Dataset. Nurses’ incident data were collected by using the conventional sheet of incident reports during 6 months from April, 2001 to September, 2001 at the emergency room in Osaka Prefectural General Hospital.

The dataset includes the types of the near misses, the patients' factors, the medical staff's factors and the shift (early-night, late-night, and daytime) and the number of items of incidents collected was 245.

We applied C4.5[1], decision tree induction and rule induction to this dataset.

Rule Induction. We obtained the following interesting rules.

(medication error):
If late-night and lack of checking,
then medication errors occur:
probability (53.3%, 8/15).

(injection error):
If daytime and lack of checking,
then injection incidents occur:
probability (53.6%, 15/28).

(injection error):
If early-night, lack of checking,
and error of injection rate,
then injection incidents occur:
probability (50%, 2/4)

Those rules show that the time shift of nurse and lack of checking were the principal factors for medication and injection errors. Interestingly, lack of expertise (personal errors) was not selected. Thus, time shift and lack of checking could be viewed as risk factor for these errors. Since the conventional format of incident reports did not include future information about workflow, we had decided to ask nurses' to fill out new report form for each incident. This is the next step in risk clarification.

4.2. Risk Clarification

Dataset. Just after the first 6 months, we had found that the mental concentration of nurses may be important factors for medical errors. During the next 6 months from October 2001 to March 2002, the detailed interference factors were included in the additional incident report form as the items of "environmental factors".

Figure 2 shows a sheet for additional information. The additional items included the duration of experience at the present ward, the number of nurse, the degree of business, the number of serious patients whether the nursing service was interrupted or not and so on.

We applied C4.5[1], decision tree induction and rule induction to this dataset.

Rule Induction. The following rules were obtained:

(medication error):
If the number of disturbing patients
is one or more,
then medication errors occur:
probability (90%, 18/20).

(medication error):
If nurses' work interrupted,
then medication errors occur:
probability (80%, 4/5).

By addition of "the environmental factors", these high probability rules of medication errors were extracted.

Rule Interpretation. With these results, the nurses discussed their medication check system.

At the emergency room, the nurses in charge of the shift prepared the medication (identification, quantity of medicines, etc.). The time of preparation before the beginning of the shift was occasionally less than 30 minutes when the liaison conference between shifts took time. In such cases, the sorting of medicines could not be made in advance and must be done during the shift.

If nurses' concentration was disturbed by the restless patients in such situations, double check of the preparation for medicine could not be made, which leads to medication errors.

4.3 Risk Utilization

Therefore, it was decided that two nurses who had finished their shifts would prepare medicines for the next shift, and one nurse in charge of the medication would check the dose and identification of medicines alone (triple check by a total of 3 nurses). (However, heated discussions among domain experts (nurses) needed for this decision, as shown in Section 5.) Improvement was applied to the check system as a result of their discussion. During the last 6 months (April 2002 to October 2002), incident reports were collected.

After introducing the triple check system, the total number of the medication errors during the last 6 months decreased to 24 cases. It was considered that the nurses' medication work was improved by the triple check system during the last 6 months.

5. Discussion for Case Study

5.1. Risk Utilization as Information Sharing

For discussion among domain experts, mining results were presented to medical staffs as objective evidence. Discussion on mining results give a very interactive discussion

Circumstances of incident

Person who noticed the incident
Whether the incident was anticipated or not
Degree of busyness, etc.

Environment of incident

Number of patients in A ward
Number of patients in B ward
Number of patients isolated due to infections
Number of restless patients
Whether there were new arrivals of patients or not
Whether treatment was made or not, etc.

Additional check sheet

この日 発症した患者の年齢・性別・病名を記入し、
発症の時間・経過を記入して分析のために追加記入欄に記入し、
* どの患者の発症が原因かを分析するために記入してください。

2. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

3. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

4. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

5. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

6. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

7. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

8. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

9. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

10. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

11. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

12. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

13. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

14. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

15. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

16. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

17. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

18. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

19. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

20. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

21. 予兆の有無を確認してください。 あり あり/不明 なし なし/不明

この日 発症した患者の年齢・性別・病名を記入し、
発症の時間・経過を記入して分析のために追加記入欄に記入し、
* どの患者の発症が原因かを分析するために記入してください。

Figure 2. Sheet for Additional Information

among the staff of the department of emergency and finally achieve common understanding of the problem on its workflow. Then, it is found that changes in workflow is required for solving the problem: If the staff assigned to the shift cannot prepare medicines, other members who are free should cooperate. However, this idea met a fierce objection in the department at first because of disagreement among nurses about the responsibility of those who prepare medicines. After repeated discussions, it was decided that nurses in charge of medication were responsible for mistakes rather than those who made preparations and nurses in the preceding shift should prepare medicines for the next shift.

During the last 6 months, medication errors were reduced markedly by creating the common perception that liaison (overlapping of shift margins, or paste margins) is important among nurses, and the initial opposition completely subsided. Following this nursing example, we could extend this policy of “paste margins”, i.e. mutual support by free staff members, to the entire department. This process also shows that information granularity is a very important issue for risk clarification. Items in a conventional report form, such as “lack of checking, lack of attention, etc.” are too coarse for risk clarification. Rather, detailed description of environmental factors are much more important to evoke domain experts’ discussion and their risk utilization.

6. Conclusion

Since all the clinical information have been stored electronically as a hospital information system(HIS), it is now expected that mining such combined data will give a new insight to medical accidents.

In order to utilize information about risk extracted from information systems, we propose risk mining which integrates the following three important process: risk detection, risk clarification and risk utilization.

As an illustrative example, we applied risk mining process to analysis of nurses’ incident data. First, data collected in 6 months were analyzed by rule induction methods, which detects several important factors for incidents (risk detection). Since data do not include precise information about these factors, we recollect incident data for 6 months to collect precise information about incidents. Then, rule induction is applied to new data. Domain experts discussed all the results obtained and found several important systematic errors in workflow (risk clarification). Finally, nurses changed workflow to prevent incidents and data were recollected for 6 months. Surprisingly, the frequency of medication errors has been reduced to one-tenth (risk utilization).

References

[1] J. Quinlan. *C4.5 - Programs for Machine Learning*. Morgan Kaufmann, Palo Alto, 1993.

[2] S. Tsumoto. Knowledge discovery in clinical databases and evaluation of discovered knowledge in outpatient clinic. *Information Sciences*, (124):125–137, 2000.

[3] S. Tsumoto. G5: Data mining in medicine. In W. Kloesgen and J. Zytkow, editors, *Handbook of Data Mining and Knowledge Discovery*, pages 798–807. Oxford University Press, Oxford, 2001.