THE ROLE OF DATA MINING TECHNIQUES IN EMERGENCY MANAGEMENT

Ning Chen
GECAD, Instituto Superior de Engenharia do Porto, Instituto Politecnico do Porto, Portugal

An Chen
Institute of Policy and Management, Chinese Academy of Sciences, Beijing, P. R. China

Keywords: Emergency management, Data mining, Intelligent decision support system, Assessment, Planning.

Abstract: Emergency management is becoming more and more attractive in both theory and practice due to the frequently occurring incidents in the world. The objective of emergency management is to make optimal decisions to decrease or diminish harm caused by incidents. Nowadays the overwhelming amount of information leads to a great need of effective data analysis for the purpose of well informed decision. The potential of data mining has been demonstrated through the success of decision-making module in present-day emergency management systems. In this paper, we review advanced data mining techniques applied in emergency management and indicate some promising future research directions.

1 INTRODUCTION

Emergency management (EM) is described as a process integrating various social resources to control and manage emergency based on the analysis of causation, evolution and consequences (L. Ji, 2006). The objective of EM is to make optimal decisions to decrease or diminish harm caused by incidents. Generally, incidents can be categorized into natural disaster, accident, public sanitary incident, and social security incident. There are some concepts closely related to emergency management in literature, such as risk management, disruption management, alert management, crisis management, early warning etc. From the generalized perspective they can be regarded as parts of EM though the information and strategies are slightly different. Risk management is a very general concept. There are many kinds of risks in our life, including economics risk, social risk, natural risk, etc. In our opinion, risk has different forms regarding the intensity. The lightest is disruption situation which can be recovered to the original status easily. The worst is emergent situation which must have caused loss of the value, and only a part of the whole value can be recovered. The intensity of crisis lies between disruption and emergency. Crisis management is a turning point in the sense that a successful intervention can transfer the bad situation to a better one, while a failing intervention often induces a worse status. Early warning and alert management mainly focus on early provision before incidents. Early warning is to find the symptom of an incident or disaster. It is always based on the prediction of the incident, and able to give an alert for the coming event. Alert Management is to manage the signal and early warning information.

Generally, emergency management can be described as a four-step process, namely, assessment, planning, action and monitoring (M. de la Asuncion, 2005). As outlined in Figure 1, assessment is to evaluate the occurrence or evolution of incidents with respect to the scenario of incidents and situation. Planning is to construct the response (plan), i.e., a proposal of control activities, based on both current situation and foreseeable evolvement, and then dispatch it to the responsible person or organization (emergency manager). The successive step is to implement the actions in the plan and evaluate the impact to decrease the damages caused by incident or overcome the problematic situation. Finally, the monitoring module detects the execution of plan in dynamic environment, concerning incomplete information and uncertainty factors during the episode. If specified conditions satisfy, the original plan is revised and a new process is

Chen N. and Chen A. (2009).
THE ROLE OF DATA MINING TECHNIQUES IN EMERGENCY MANAGEMENT.
In *Proceedings of the 11th International Conference on Enterprise Information Systems - Artificial Intelligence and Decision Support Systems*, pages 118-123
DOI: 10.5220/0001961601180123

triggered. The emergency manager is responsible for the execution of the whole process by means of validating the plan and modifying it at necessary.

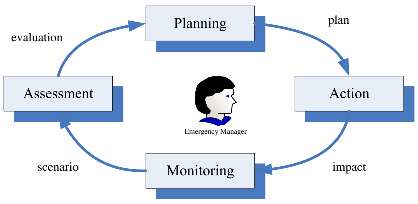

Figure 1: Emergency management process.

Traditional emergency management mostly relies on the skill and experience of emergency manager or domain expert to evaluate the situation, define the goal of intervention and choose the strategies to achieve the goal. Nowadays, due to the rapid development of information techniques, the increasing amount of data makes the decision-making task more difficult and complicated. The data involved includes numerical data captured by sensors or communication devices reflecting current situation, qualitative data from domain experts expressing personal preference, and historical data from data base. The data needs to be converted into valuable knowledge which helps to generate well informed plans or strategies. Due to the complexity of emergency management problems and explosive growth of data processed, there is a great need to enhance and supplement the capabilities of emergency management under the support of specialized problem-solving tools. Data mining (DM) is a good candidate for this purpose, in particular in assessment and planning phases.

Since effective management of incidents is based on reasonable assessment and decisions, decision making is the kernel of emergency management. The success of current emergency management systems is largely dependent on the decision-making module which enables managers to take advantage of the capabilities of DM techniques. In literature, a variety of data mining techniques were employed in the area of emergency management to provide the functionality for the purpose of improving the intelligence of emergency management systems. In such systems, data mining techniques analyze the information, track the evolution of episode of emergency and share the decisions with emergency manager. The potential of data mining techniques have been demonstrated for valuable plan construction in both military domains, e.g., air camping design, noncombatant evacuation, and civil domains, e.g., oil spill, flood, forest fire. As a related work, (U. Cortes, 2000) shows the usage of artificial intelligence techniques in environmental decision support system. However to our knowledge, there is little attempt on reviewing DM applications in the field of emergence management.

The remainder of the paper is organized as follows. Section 2 addresses state-of-the-art DM techniques, including association rule mining, classification and sorting, clustering, and case-based reasoning as well as their contributions in emergency management. Some emergency management systems integrated with DM techniques are introduced in section 3. Section 4 concludes the paper and presents some interesting future research directions.

2 DATA MINING TECHNIQUES IN EMERGENCY MANAGEMENT

Formally, data mining is an analytical technique to discover hidden, implicate, and nontrivial knowledge from database. Data mining, especially predictive data mining acts as a model-driven process for risk prediction in pre-incident phase and evolvement forecast in post-incident phase. A broad range of DM techniques have been employed in the management of emergencies and play important roles to provide emergency manager with high-level information in order to take proper decisions at right time. The contributions of DM can be summarized in four aspects: (1) In assessment phase, it is used for knowledge acquisition to identify potential incident and predict the possible consequence from relevant factors and their relationship; (2) In response planning phase, it is used to suggest the desirable plan to responsible manager; (3) In plan action phase, it is used to evaluate the performance of plan concerning the accuracy, reliability, and usefulness. (4) In monitoring phase, it is used to acquire high quality information from raw data. In this section, we introduce advanced data mining techniques and their representative applications in emergency management.

2.1 Association Rule Mining

Association rule (AR) mining is one of the most widely applied technique to discover interesting relations between variables in large databases. Mining association rules is introduced for the first time to discover regularities between products from basket data recorded by point-of-sale (POS) systems in supermarkets (R. Agrawal, 1993). In a market basket database, each transaction contains the set of products (items) purchased by a customer. Association rule mining attempts to discover important association among items such that the presence of some items in a transaction will imply that of other items in the same transaction. So far, association rule has been extended to different types and applied in a vast variety of areas.

Some experiences show that effective performance of DM is the key to the success of risk management, which relies on the accurate prediction of emergency to some extent. For this purpose, association rule mining is employed to predict the possibility of occurrence of particular incidents. A geospatial decision support system (S.K. Harms, 2001) is developed to improve the quality and accessibility of draught risk management. The relationship between user defined target episodes and other climate events is exploited from geospatial and temporal data and consequently used to predict the occurrence of target episodes. It was reported that the association rule between weather station data, crop yields and sea surface thermal properties provides reasonable strategies like adjusting the planting date and population to farmers in US.

In network event management, abnormal event is seen as a kind of emergency essential to network management. Three types of events, namely, burst patterns, periodic patterns and mutually dependent patterns, that indicate underlying problems are mined from historical event data (J. L. Hellerstein, 2002). These patterns do not occur frequently enough in well-maintained production environments, but they are of particular interest to event management. Mining burst patterns is achieved by first finding periods in which event rates are higher than a specified threshold, then mining patterns common to the periods identified. Periodic patterns are explored by finding period lengths for each event type followed by temporal associations. Mining mutually dependent patterns is similar to association rule mining except that it focuses on infrequent patterns related to problem situations.

2.2 Classification and Sorting

Classification is a typical supervised learning method to derive models between features (independent variables) and class (target variable) in the sense that the class information is considered during learning. The process of classification is to first develop models through training on the preclassified data and then assign new data to a class on the basis of the model developed. Sorting also involves the assignment of an alternative to a predefined group expect that the groups are ordered in an ordinal way. In emergency management, classification and sorting are particularly useful in the assessment of incidents, which analyze whether the incident occurs or to what extent the damage is caused.

Decision tree is one of the fastest and easily interpreted algorithms to solve the task of classifying instances. The tree structure is constructed by a “divide and conquer” strategy, with leaves representing classification and branches representing conjunctions of feature tests. Decision tree algorithms have many variations differing in alternate methods of choosing the split attribute and value, pruning method, and data access methods used for large training dataset. In (A. Gerbesioti, 2001), the classification model derived by C4.5 decision tree algorithm (Quinlan, 1993) relating to some problem-specific instances is used by an expert system for ranking infected buildings.

Artificial neural network (ANN) is a computational modelisation inspired from biological neural network to explore nonlinear patterns from complicated and large data. It consists of a set of interconnected artificial neurons for data computation through a connectionist approach. ANN has been widely used in various categories of risk assessment problems. E.g, an early warning system is composed of knowledge acquisition, knowledge representation and knowledge-based reasoning using ANN (B. Yang, 2001). The applicability of ANN is demonstrated on predicting storm surge and surge deviation in real-world data from Suao Harbor station of Taiwan (Lee, 2008).

2.3 Clustering

Clustering is an unsupervised process to partition a set of data into homogeneous clusters in the sense that there is no preclassified data as it were in classification. The objects are grouped together on the basis of intrinsic similarity to maximize the intraclass similarity and minimize the interclass similarity. Clustering is able to generate compressed representatives of raw data, especially for large, complex data set with many variables and a lot of internal structures. Applications of clustering in emergency management mainly focus on high quality information acquisition from a mass of raw data with redundancy and noise.

Since real data usually contains reduplicate and false alarms, the alarms should be filtered and aggregated to achieve a better understanding of situation. Through a density-based clustering algorithm which finds dense regions separated by low-density regions and clusters together the objects in the same dense region, the alarms form into clusters based on the associated attributes and then false alarms are identified from the uniformly behaving neighborhoods. As a result, the aggregated alerts are generated from a large number of incoming alarms (Vandana, 2005).

Kohonen's self-organizing map (SOM) (Kohonen, 1997) is called topology mapping in the sense that the similarity of input data is preserved in the output layer through a competitive learning. As a visualized approach, SOM makes possible to detect internal data structure from the map grid. In (Y. Atoji, 2000), SOM is used as an information filtering method to select useful information from a large amount of data related to emergency. Utilizing the clustering ability of SOM, the incoming data that contains similar content is mapped to a particular category and then a number of representative keywords are extracted for each pattern. Also, SOM is used for security assessment of power system (K.S. Swarup, 2006). The line flows under different component cases are input to a two-dimensional grid network to obtain the cluster of components based on their loading limits. From the output patterns, the violation of constraints can be identified and consequently the secure and insecure status are classified accordingly. The utilization of SOM eliminates the repetitive computation required in traditional methods.

2.4 Case-based Reasoning

Case-based reasoning (CBR) originates from traditional behavior of human problem solving to solve new problems based on past cases. For example, a doctor diagnoses the disease of a patient by recalling another patient who has similar symptoms. Similarly, a lawyer advocates a case in a trial based on a judge in the past. Formally, the process of CBR consists of four steps (A. Aamodt, 1994): (1) Retrieve: extract relevant cases from case base based on a distance measurement; (2) Reuse: map the similar cases to the target case by adapting to current situation; (3) Revise: apply the solution in the real world and test the influence with respect to the feedback; (4) Retain: store the new case and solution if accepted. Compared with rule-based reasoning, CBR does not need the complicated phase of rule extraction, however, it is critically depended on the distance metric for similar case retrieval and sensitive to the computational cost when a large number of cases are available for comparison. Moreover, the adaption is important to solution reuse taking into account the difference between matched cases and current situation, which is probably more difficult than generating a new solution.

In emergency management, CBR methodology is regarded as a powerful method to create the emergency plan based on the solutions of similar past incidents. The popularity of CBR benefits from the presence of cases available in the databases. For example, CBR is applied to plan the initial attack to forest fires with respect to past interventions stored in the case base (P. Avesani, 1999). The adaption of retrieved plan to domain requirements and constraints is implemented by a constraint propagation algorithm. CBR is also used to generate incident response plans automatically in the domain of computer and network security (G. Capuzzi, 2006). The past cases including attacks (incidents) and corresponding responses (plans) are stored as a sequence of concrete events. When a new attack occurs, it searches the case memory for similar attacks on the basis of a distance metric, then reuses the past response by substituting the concrete action and parameters with current action type and parameters.

3 EMERGENCY MANAGEMENT SYSTEMS AND APPLICATIONS

Emergency management system (EMS) is an intelligent decision support system (IDSS) capable of enhancing the ability of human to manage the emergent situations. It plays an increasing role to reduce the risks by appropriate and proactive preparation or minimize the negative impact caused by incidents by immediate and efficient response. Although a number of emergency management systems have been developed, the capabilities of emergency management systems should be addressed:

- Acquire, filter, select, transform and visualize data in diverse types and from different resources to provide well understanding of relevant factors and identify the potential problems;
- Explore, represent, structure and manage background knowledge, expert subjective knowledge and predictable models by means of statistical analysis, data mining or other approaches;
- Provide functions of problem formulization, diagnosis, planning and optimization using rule-based reasoning or case-based reasoning;
- Assist emergency manager to evaluate the impact and cost of strategies, choose appreciate plans for a given problem;
- Monitor the performance of plans and adapt to dynamic environment taken into consideration uncertainty factors;
- Provide user-friend interface which allows users to offer subjective criteria, justify decision impact, and intervene the process interactively.

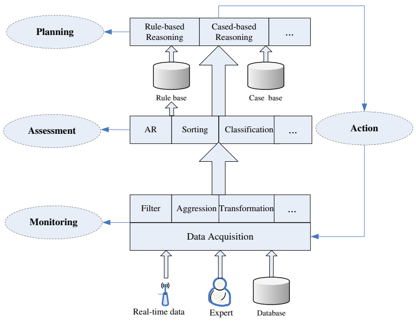

Figure 2: Emergency management flow.

As shown in Figure 2, the flow of emergency management starts from data acquisition followed by data preprocessing of filter, aggression and transformation. Then using data mining based analysis tools, rules are extracted and reserved in rule base for the usage of assessment and planning. Finally, the decisions are made through rule-based reasoning, case-based reasoning or hybrid approaches. So far a variety of EMS at regional, national, or international level have been designed in centralized or distributed architecture for various domains. Some representative emergency management systems are listed in the following.

- ARTEMIS (J.Z. Hernandez, 2001): a general architecture of knowledge-based environmental emergency management system. The evolution of incident is represented by a collection of causal models, classification models and temporal models structured in a hierarchy. The effect of primary causes is propagated top-down through a rule inference process and the expected impact of plans is estimated by simulators.
- PHOENIX (P.R. Cohen, 1989): a multi-agent system capable of providing a distributed, adaptive and real-time planner to control simulated forest fires.
- CHARADE (P. Avesani, 1999): an interactive decision support system for situation assessment and activities planning in forest fire emergencies application domain. This system is also used for knowledge acquisition from the cases, which enable users to directly extract various forms of knowledge, e.g., dependencies between features, clusters of cases, feature statistical description.
- SIADEX (M. de la Asuncion, 2005): an interactive fire fighting planner, integrating several artificial intelligence techniques for designing and redesigning plans against forest fires. The system provides an easy-to-use tool capable of not only assisting the manager during the whole process of decision making in an online fashion, but also reproducing the episodes for manager training in an offline fashion.
- RODOS (W. Raskob, 2005): a real-time decision support system for nuclear emergency management, providing countermeasure and remediation strategies from pre-release phase, release phase to long-time phase.
- AGrIP (J. Luo, 2007): a flood decision support system integrating a multi-strategy data mining tool MSMiner in the algorithm library.
- SIPE (Bienkowski, 1995): a crisis management system for plan generation to oil threats in the sea requiring users to provide knowledge or interaction to the system.
- CubeView (C.T. Lu, 2005): a road transportation system applying advanced data mining and visualization techniques to road traffic data for emergency situation control and management planning. Visualization of loop-detector traffic data helps to recognize potentially important patterns embedded in the data. Outlier detection discovers traffic patterns applicable in the detection of, prevention of and recovery from road crisis.
- IMASH (E. Iakovou, 2001): an intelligent integrated dynamic information management tool to provide comprehensive data pertaining to emergency planning and response for hurricane disasters. It can offer support for a wide range of hurricane disaster related activities including hurricane preparedness, hurricane response and restoration.

4 CONCLUDING REMARKS AND FUTURE PERSPECTIVES

With the increasing demand of data analysis involved in emergency management, there is an urgent need for automatic and effective tools which allow emergency managers to make quick and desirable intervention to incidents and disasters. The intelligent decision support systems based on data mining techniques become more and more popular in diverse domains related to emergency management. This paper reviews the up to date data mining techniques and their applications in the field of emergency management.

A future research direction is the investigation of fuzzy data mining for emergency management. The data related to the incident and situation are usually represented in linguistic assessments instead of deterministic numerical values, thus, the fuzzy strategies are needed to deal with the imprecise and uncertain information. Since decision-making is usually carried out in a distributed environment to achieve a satisfactory solution, organizational learning and collaborative learning is another direction of need in future study. Additionally, special attentions should be given to the implementation of emergency management system. As pointed out by several authors, an interactive system is more realistic than a completely automated system, in which human is responsible of personnel knowledge specification, knowledge interpretation and decision selection. The selection can be performed through a multi-criteria evaluation from several perspectives such as cost, effort, feasibility, public acceptance, psychological and political implication, preference of decision makers (W. Raskob, 2005). Since EM managers are not experts on DM, the comprehension is quite important for easy access, e.g, derived rules are easily understandable and applicable in decision reasoning. The scalability of DM is worth noting for decision-making due to the information flood occurring at the inception of emergency, when real-time response becomes difficult.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial grant of GECAD/ISEP-Knowledge Based, Cognitive and Learning Systems (C2007-FCT/442/2006).

REFERENCES

A. Aamodt, E. P. (1994). Case-based reasoning: Foundational issues, methodological variations and system approaches. *Artificial Intelligence Communications*, 1:39–52.

A. Gerbesioti, V. Delis, Y. T. S. A. (2001). Developing decision support tools for confronting seismic hazards. In *Proc. of the 8th Panhellenic Conference on Informatics*, pages 247–253.

B. Yang, L.X. Lib, H. J. J. X. (2001). An early warning system for loan risk assessment using artificial neural networks. *Knowledge-Based Systems*, 14(5-6):303–306.

Bienkowski, M. (1995). Demonstrating the operational feasibility of new technologies: the arpi ifds. *IEEE Expert*, 10(1):27–33.

C. T. Lu, L.N. Sripada, S. S. R. L. (2005). Transportation data visualization and mining for emergency management. *International Journal of Critical Infrastructures*, 1(2/3):170–194.

E. Iakovou, C. D. (2001). An information management system for the emergency management of hurricane disasters. *International Journal of Risk Assessment and Management*, 2(3-4):243–262.

G. Capuzzi, E. Cardinale, I. P. L. S. (2006). An incident response support system. *International Journal of Computer Science and Network Security*, 6(10):72–78.

J. L. Hellerstein, S. Ma, C.-S. P. (2002). Discovering actionable patterns in event data. *IBM Systems Journal*, 41(3):475–493.

J. Luo, L. Xu, J.-P. J. L. Z. Z. S. (2007). Flood decision support system on agent grid: Method and implementation. *Enterprise Information Systems*, 1(1):49–68.

J. Z. Hernandez, J. S. (2001). Knowledge-based models for emergency management systems. *Expert Systems with Applications*, 20:173–186.

Kohonen, T. (1997). *Self-organizing Maps*. Springer Verlag, Berlin.

K. S. Swarup, P. C. (2006). Power system static security assessment using self-organizing neural network. *J. Indian Inst. Sci.*, 86:327–342.

L. Ji, H. Chi, e. (2006). *Emergency Management*. High Education Press.

Lee, T.-L. (2008). Prediction of storm surge and surge deviation using a neural network. *Journal of Coastal Research*, 24(sp3):76–82.

M. de la Asuncion, L. Castillo, F. J. G. A. P. F. (2005). Siadex: an interactive knowledge-based planner for decision support in forest fire fighting. *AI Communications*, 18(4):257–268.

P. Avesani, A. P. (1999). Cases on fire: Applying cbr to emergency management. *New Review of Applied Expert Systems*, 6:175–190.

P. R. Cohen, M.L. Greenberg, D. H. A. H. (1989). Trial by fire: Understanding the design requirements for agents in complex environments. *AI Magazine*, 10(3):32–48.

Quinlan, J. (1993). *C4.5: Programs for Machine Learning*. Morgan Kauffman, Los Altos, CA.

R. Agrawal, T. Imielinski, A. S. (1993). Mining association rules between sets of items in large databases. In *Proc. of ACM SIGMOD*, pages 207–216.

S. K. Harms, S. Goddard, S. R. W. W. T. T. (2001). Data mining in a geospatial decision support system for drought risk management. In *Proc. of the National Conference on Digital Government Research*, pages 9–16.

U. Cortes, M.S. Marre, L. C. I. R. (2000). Artificial intelligence and environmental decision support systems. *Applied Intelligence*, 13:77–91.

Vandana, P. (2005). Dm-ams: Employing data mining techniques for alert management. In *Proceedings of National Conference on Digital Government Research*, volume 89 of *ACM International Conference Proceeding Series*, pages 103–111.

W. Raskob, V. Bertsch, J. G. S. B. F. G. (2005). Demands to and experience with the decision support system rodos for off-site emergency management in the decision making process in germany. In *Proc. of the 2nd International ISCRAM Conference*.

Y. Atoji, T. Koiso, S. N. (2000). An information filtering method for emergency management based on self-organizing map. *Human Interface*, 2(3):1–6.