A Combined Web Mining Model and Its Application in Crisis Management

Xingsen Li^{1,*}, Lingling Zhang^{1,2}, Maoliang Ding⁴, Yong Shi^{2,3}, and Jun Li¹

¹ School of Management, Graduate University of the Chinese Academy of Sciences, Beijing, 100080

lixingsen@126.com

² The CAS Research Center on Data Technology & Knowledge Economy ; Beijing, 100080

yshi@gucas.ac.cn, zll933@163.com

³ Graduate University of the Chinese Academy of Sciences, Beijing, 100080

⁴ Information center, Shandong Jinjing Group, Zibo, 255200

Jamesdingcn@msn.com

Abstract. Crisis management is an increasingly important activity of management. Organizations' critical crises can induce severe consequences or even disasters if wrong decisions are made. The World Wide Web provides an unprecedented freedom for the public to discuss one's opinions which impact much on the progress of the crisis. Based on the study of web mining and its relative techniques, a combined web mining model was presented. It consists of web mining, text mining, data mining and human experience knowledge. All knowledge from each part was saved in database through a knowledge management platform. Its application shows that utilizing the combined model for crisis management can improve the quality of decisions.

Keywords: Web Mining Model, Crisis Management, Data Mining, Knowledge Management.

1 Introduction

The World Wide Web has become a huge data source. Billions of pages are publicly available and it is still growing. The web provides an unprecedented freedom to discuss public issues, crisis events and politics etc. The general publics enjoy the right to speak on BBS, message board, self-owned sites and blogs. One can also take part in interactive actions online and even voting. Web Mining, that can discover knowledge from huge amounts of web pages, has become an urgent research area in computer science [1]. Crisis management is a relatively new field of management. Its activities include forecasting potential crises and planning how to deal with them, such as recover any damage to public image. Crisis management has many directions. Most research has relied heavily on case study methods yet often with rhetorical or even inconsistent suggestions [2]. The interactive features of the Web make crisis

* Corresponding author.

management become more complex, but also provide a great convenience to guide or steer public opinions for the management of crisis. Our goal in this paper is to present a novel application model combined with web mining, data mining and human experience knowledge sharing, and introduce its process in crisis management.

2 Framework of the Combined Web Mining Model

Web mining is powerful, but it also needs to combine with other technologies and management rules to enhance the business. Figure 1 shows a framework of combined web mining model:

Fig. 1. Framework of the combined web mining model

Web mining extracts knowledge from web data, i.e. web content, web structure, and web usage data [1]. Web content mining is the process of extracting useful information from the contents of web documents. Web mining involves two main steps: Extracting the pages relevant to a query and ranking them according to their quality. Text mining deals with text documents in general, such as emails, letters, reports, and articles, that exist in both intranet and internet environment. Text mining is concerned with the analysis of very large document collections and the extraction of hidden knowledge include topic discovery and tracking, extracting association patterns or abstract, clustering of web documents and classification from text-based data.

Data mining has been used in many fields [3, 4]. During the process of web mining and text mining, statistic data such as visit-times, last-visit-date, browser numbers, weights, number of total words can be collected from mining software or by special software tools, these data sets are acted as training sets for data mining. We define two classes for this dataset using a label variable: The pages should be pay attention to (label = 1) and not (label = 0). After the classification model has been trained, the effect of new pages can be predicted by scoring in earlier time.

Finally, knowledge management is necessary for collecting experts' experience and compared with knowledge from data mining. Its working process is similar with the process presented in paper [5].

3 Working Process for the Aftermath of Crisis

In handling the aftermath of the crisis, the most important thing to do is to change the passive situation. Most organizations make decisions rely on experience of skilled persons [2]. How to guide managers to invoke corresponding reactions? We get the answers in a proper methodology based on web mining. The process has 7 main steps:

Fig. 2. The work process of the combined model

Step 1. Define event ontology. Similar with the Personal Ontology concept [6] and the construction of ontology by mining the Semantic Web data [7], we put forward user defined event ontology in the dynamic environment.

Step 2. Searching the web. Search information on the web manually or by web crawler tools. Useful pages and its web address are collected to software called integrator. Web crawler counts the information by statistic and transforms it into structured data set then stores into data base.

Step 3. Files integration. Transfer web pages been searched into text or XML files and integrate some local files together by integrator. Noise on web pages should be cleaned before applying mining tasks to these files.

Step 4. Mining. By web mining and text mining, we can discover interesting topics, extracting association patterns or abstract, clustering of web documents and classification. By data mining, we can get scoring knowledge and know which pages should pay more attention to and decide when to take actions combined with experience.

Step 5. Monitoring. Identifying communities of users and information sources [1] on the web. The communities play a more important role and can even identifying malicious competitions. Monitor all useful pages from mining and the new comments. OLAP or other visualization tools can be used here to monitor efficiently, and redo step 2 to step 4 if necessary.

Step 6. Take actions. Since people are easily affected by the group consciousness. Proper actions can change the passive situation similarly, such as publish the true facts and guide browsers to see the positive side on the key websites, say the right words to invoke good corresponding reactions, advertisement on TV on the right time, communicating with customers or press conference, etc.

Step 7. Feeding back. After the actions been taken, the effect can be found by re-mining. At the same time, the experts' opinions can also be saved as text files or store in the knowledge base. This can make the future crisis management wiser and wiser.

This model had been implemented in a famous bedding company (Shumian) in Beijing, China. Its application shows that the model has a good effect.

4 Conclusions

We proposed a combined model based on web mining and described its working process. This approach gives a new tool to manage crisis problem and can also be applied to other fields, such as strategy, society security, web fraud, and advertisement etc. There are several issues need to be researched deeply, such as how to collect related pages efficiently by web crawler, how to mark the label of data set more precisely, how to mining other un-structured data. We also notice that the speed of the whole process should be increased and the Event Ontology needs to be updated by web mining results in the future study.

Acknowledgments. This research has been partially supported by a grant from National Natural Science Foundation of China (#70621001, #70531040, #70501030, #70472074), National Natural Science Foundation of Beijing #9073020, 973 Project #2004CB720103, National Technology Support Program #2006BAF01A02, Ministry of Science and Technology, China, and BHP Billion Co., Australia.

References

1. Srivastava, J., Desikan, P., Kumar, V.: Web Mining- Concepts, Applications and Research Directions, In Data Mining: Next Generation Challenges and Future Directions. AAAI/MIT Press, Boston, MA (2003)
2. Lin, Z.: Organizational Performance Under Critical Situations—Exploring the Role of Computer Modeling in Crisis Case Analyses, Computational & Mathematical Organization Theory, Springer Netherlands, Vol. 6, No. 3, (2000) 277–310
3. Kou, G., Y. Peng, etc.: Discovering Credit Cardholders' Behavior by Multiple Criteria Linear Programming, Annals of Operations Research Vol.135, (2005) 261–274

4. Kou, G., Peng, Y., Shi, Y., etc.: Network Intrusion Detection by Using Multiple-Criteria Linear Programming, Proceedings of 2004 International Conference on Service Systems and Service Management, Beijing, China (2004)
5. Li, X.S., Shi, Y., etc.: A Knowledge Management Platform for Optimization-based Data Mining, Sixth IEEE International Conference on Data Mining, Hong Kong, China, Dec. 18-22, (2006)
6. Nakayama, K., Hara, T., Nishio, S.: A Web Mining Method Based on Personal Ontology for Semi-structured RDF, Web information systems engineering. International workshop, New York NY, Vol. 3807, (2005) 227–234
7. Maedche, A., Staab, S.: Ontology Learning for the Semantic Web, IEEE Intelligent Systems, Vol. 16, No. 2, (2001) 72–79