Chapter 1

Overview of Knowledge Discovery in Databases Process and Data Mining for Surveillance Technologies and EWS

Inci Batmaz
Middle East Technical University, Turkey

Gülser Köksal
Middle East Technical University, Turkey

ABSTRACT

Development of more effective early warning systems (EWSs) for various applications have been possible during the past decade due to advancements in information, detection, data mining (DM) and surveillance technologies. These application areas include economy, banking, finance, health care, bioinformatics, production and service delivery, hazard and crime prevention and minimization of other social risks involving the environment, administrations, politics and human rights. This chapter aims to define knowledge discovery in databases (KDD) process in five steps: Data preparation, data preprocessing, DM, evaluation and interpretation, and implementation. DM is further explained in descriptive and predictive mining categories with their functions and methods used or likely to be used in EWSs. In addition to well-known structured data types, mining of advanced data types such as spatial, temporal, sequence, images, multimedia and hypertexts is also introduced. Moreover, it presents a brief survey of overview and application papers and software in the EWS literature.

INTRODUCTION

Early warning systems (EWSs) are designed to warn of a potential or an impending problem such as earthquakes, tsunamis, wildfires, hurricanes, pandemics, famine, human rights violations, conflicts, financial crisis, fraud and market shifts. They typically utilize a network of sensing or surveillance devices for the problem or risk detec-

tion. Organisation for Economic Co-operation and Development [OECD] (2003) provides an analysis of many of the systemic risks that surround us.

Knowledge discovery in databases (KDD) is a general process of discovering hidden patterns in data for better decision making. A major part of KDD involves data mining (DM). DM has become a widely accepted approach for risk management. Tsumoto and Washio (2007) provide a perspective about application of DM to risk management, in general. Similarly, Baesens et al. (2009) outline a series of upcoming trends and challenges for DM, covering applications in credit risk, fraud detection and so on.

Typical applications of DM are observed in management of risks in economy, banking and financial operations, such as economic crisis prediction and fraud detection based on historical data about the economy or banking/financial operations and customers.

Environmental, public health and disaster monitoring systems use surveillance technologies such as remote sensing (satellite imaging) and the data collected are analyzed by DM approaches for early warnings. Several EWSs for health and infectious disease control exist (such the Global Outbreak Alert and Response Network (GOARN) of World Health Organization (WHO)) that make more and more use of DM methods.

Similarly, EWS applications and attempts are observed in the literature and in practice for network security, intelligence, protection against natural or man-made hazards, risks from poor product quality or competitors' threats, and so on. Gurr and Davies (1998) present original essays that examine complex political and humanitarian crisis early warning factors and develop recommendations for effective early warning and response; Carragata (1999) examines early warning signs and designing EWSs for business survival and growth; Goldstein et al. (2000) study banking and currency crises in emerging markets to identify most reliable early warning signals; Gilad (2003) discusses early warning development

for competitiveness in the market; Meier (2006) discusses an EWS for preventing environmental conflicts; Gasparini et al. (2007) analyze major earthquake EWSs; Brown (2008) focuses on remote sensing measurements and how they are used to in famine EWSs, specifically in the US Agency for International development (USAID)'s Famine EWS Network; Glantz (2009) reviews EWSs for water- and weather-related hazards; OECD (2009) presents a review and analysis of early warning and response mechanisms and instruments for preventing violence, war, and state collapse, violent conflict. There is a need to develop many others and improve the existing ones for detecting important economic and social risks including hungers, epidemics, crimes, child abuses and unemployment.

KDD and DM, in particular, can boost these developments if utilized to their full extends. The main propose of this chapter is to define the KDD process and DM with their functions and methods for EWS developers and users.

In the following, first, a brief overview of DM applications in EWSs is provided based on a review of the literature. Then, the KDD process is defined in five major steps: Data preparation, data preprocessing, DM, evaluation and interpretation, and implementation. DM is further explained in descriptive and predictive mining categories with their functions and methods. Mining of advanced data types such as images, multimedia and hypertexts, on the other hand, is presented in a separate section. Afterwards, a review of DM software used in EWSs is provided. At the end of the chapter, future research directions and concluding remarks are given.

SOME DATA MINING APPLICATIONS FOR EWS DEVELOPMENT IN THE LITERATURE

Several attempts exist in the literature for EWS development in various industries/areas by us-

Related Content

Analysing Architecture and Transaction Model in Securing Mobile Commerce

Poonam Ahuja Narangand Basanti Pal Nandi (2016). *Securing Transactions and Payment Systems for M-Commerce (pp. 193-216).*

www.irma-international.org/chapter/analysing-architecture-and-transaction-model-in-securing-mobile-commerce/150076

Exploring the Roles of Police Leaders in Countries in Transition

Gerald Dapaah Gyamfi (2020). *International Journal of Risk and Contingency Management (pp. 1-17).*

www.irma-international.org/article/exploring-the-roles-of-police-leaders-in-countries-in-transition/261205

Anomaly Detection in IoT Frameworks Using Machine Learning

Phidahunlang Chyne, Parag Chatterjee, Sugata Sanyaland Debdatta Kandar (2020). *Applied Approach to Privacy and Security for the Internet of Things (pp. 88-112).*

www.irma-international.org/chapter/anomaly-detection-in-iot-frameworks-using-machine-learning/257905

Assessing the Risk of Leveraging Technology in Small Businesses Entering the European Union

Fation Losha, Kenneth David Strangand Narasimha Rao Vajjhala (2017). *International Journal of Risk and Contingency Management (pp. 56-69).*

www.irma-international.org/article/assessing-the-risk-of-leveraging-technology-in-small-businesses-entering-the-european-union/188682

Security Issues for Cloud Computing

Kevin Hamlen, Murat Kantarcioglu, Latifur Khanand Bhavani Thuraisingham (2010). *International Journal of Information Security and Privacy (pp. 36-48).*

www.irma-international.org/article/security-issues-cloud-computing/46102