Chapter 6
Data Mining Problems Classification and Techniques

Nayem Rahman
Portland State University, Portland, USA

ABSTRACT

Data mining techniques are widely used to uncover hidden knowledge that cannot be extracted using conventional information retrieval and data analytics tools or using any manual techniques. Different data mining techniques have evolved over the last two decades and solve a wide variety of business problems. Different techniques have been proposed. Practitioners and researchers in both industry and academia continuously develop and experiment with variety of data mining techniques. This article provides a consolidated list of problems being solved by different data mining techniques. The author presents up to three techniques that can be used to address a particular type of problem. The objective is to assist practitioners and researchers to have a holistic view of data mining techniques, and the problems being solved by them. This article also provides an overview of data mining problems solved in the healthcare industry. The article also highlights as to how big data technologies are leveraged in handling and processing huge amounts of complex data from data mining perspectives.

1. INTRODUCTION

Data mining techniques have been applied in healthcare, retail, marketing, customer relationship management (CRM), finance and banking, insurance, and scientific discoveries to name a few (Dominguez-Morales et al., 2018; Rahman, 2018; Kourou et al., 2017; Abdallah et al., 2016; Liao et al., 2012). Data mining techniques are used to address different business scenarios such as customer recommendations, anomaly detection, development of customer profiles, mining of unstructured data, discovery of new insights, providing accurate predictions, exploring complex patterns of data, providing predictive analytics capabilities, developing interesting patterns in data, developing customer behavior patterns, medical diagnoses, and making scientific discoveries.

In banking, predictive analytics are essential in fraud detection, assessing customer creditworthiness, and bankruptcy predictions (Jabeur, 2017; du Jardin, 2017; Tarvin, 2017). In sales and marketing customers classification and segmentation are important to serve those customers better. In CRM data mining techniques are used to understand customer behavior, attraction, retention, development, make correct classifications, and predicting customers' probability to purchase certain products and services (Ngai et al., 2009). In healthcare data mining models are used in predicting patients' hospital re-admittance, anomaly detection, clinical decision making, tracking chronic disease states in high-risk patients, prediction of Intensive Care Unit readmission, and mortality rate (Herland et al., 2014; Jothi et al., 2015; Koh and Tan, 2005; Presbitero et al., 2017). Insurance companies would be interested in having a prediction model based on classification of customers they would consider high risk, those who have previously made false claims (Rahman and Iverson, 2015) or in predicting accident rates in a certain period of time. In marketing data mining techniques are used for drawing insights into customer characteristics, identifying their purchasing patterns and conducting real-time interactive marketing (Shaw et al., 2001).

Researchers have attempted to conduct reviews of individual data mining techniques to report progress in terms of research and problem solving. They also made attempts to compare between data mining techniques to understand problem solving capability and performance. The following paragraphs provide an account of those.

In his recent publication titled, "A Taxonomy of Data Mining Problems," Rahman (2018) presented a variety of data mining problems and data mining techniques used to solve them. This article provides updates based on the latest research on data mining, and the author proposes up to three techniques in solving distinct types of data mining problems. The author also highlights role of data mining in healthcare. The author states data mining can play a significant role in big data space.

This article identifies various data mining problems and discusses top data mining techniques, their applications, and problem-solving capability. It will first review prominent data mining problems, then provides up to three techniques in each data mining problem area. The author searched relevant articles in EBSCO databases which pulled thousands of articles related to each data mining technique and three years' worth of publications (2016-2018). The author reviewed the articles by title to short-list relevant articles. The research reveals that there are certain techniques that solves some specific data mining problems better than others.

In 2018 it is readily apparent that several data mining techniques are widely used. They include Bayesian networks, neural networks (NN), decision trees, association rules, clustering techniques, support vector machine (SVM), logistic regression, and K-nearest neighbors (Rahman, 2018). Based on an extensive review of extant literature it was found that a handful of real-world data mining problems are solved by the data mining techniques mentioned above. The important problems solved by data mining techniques include classification, forecasting, risk analysis, optimization, prediction, partitioning, relationship between subsets of data pattern recognition, ranking, sequence discovery, anomaly detection, text mining, geographic knowledge discovery, and visualization (Rahman, 2018). More than one data mining techniques could be applied to solve a particular data mining problem, but some are frequently used than others, depending on accuracy of results and performance.

In section two, we discuss related work based on extant literature. We provide a detail account of research and experiments done on individual data mining techniques. We present comparative analysis undertaken using different data mining techniques. In section three, we give a brief description of data mining problems solved by each data mining technique. We also propose up to three data mining techniques that can be used in each of the data mining problems identified based on academic and industry

Related Content

Rethinking Bloom's Taxonomy: Implicit Cognitive Vulnerability as an Impetus towards Higher Order Thinking Skills
Caroline M. Crawfordand Marion S. Smith (2015). *Exploring Implicit Cognition: Learning, Memory, and Social Cognitive Processes (pp. 86-103).*
www.irma-international.org/chapter/rethinking-blooms-taxonomy/120854

Political Sentiment Mining: A New Age Intelligence Tool for Business Strategy Formulation
Nishikant Bele, Prabin Kumar Panigrahiand Shashi Kant Srivastava (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications (pp. 1406-1422).*
www.irma-international.org/chapter/political-sentiment-mining/252089

A Comparative Study of Machine Learning Techniques for Gesture Recognition Using Kinect
Rodrigo Ibañez, Alvaro Soria, Alfredo Raul Teyseyre, Luis Berdunand Marcelo Ricardo Campo (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications (pp. 1096-1117).*
www.irma-international.org/chapter/a-comparative-study-of-machine-learning-techniques-for-gesture-recognition-using-kinect/252073

Advancing Malware Classification With an Evolving Clustering Method
Chia-Mei Chenand Shi-Hao Wang (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications (pp. 1882-1894).*
www.irma-international.org/chapter/advancing-malware-classification-with-an-evolving-clustering-method/252116

Relativity and Cognitive Ethics
Richard Sieb (2019). *Media Models to Foster Collective Human Coherence in the PSYCHecology (pp. 119-139).*
www.irma-international.org/chapter/relativity-and-cognitive-ethics/229333