Chapter 6
Rules Extraction using Data Mining in Historical Data

Manish Kumar
IIIT, Allahabad, India

Shashank Srivastava
IIIT, Allahabad, India

ABSTRACT

Rules are the smallest building blocks of data mining that produce the evidence for expected outcomes. Many organizations like weather forecasting, production and sales, satellite communications, banks, etc. have adopted this mode of technological understanding not for the enhanced productivity but to attain stability by analyzing past records and preparing a rule-based strategy for the future. Rules may be extracted in different ways depending on the requirements and the dataset from that has to be extracted. This chapter covers various methodologies for extracting such rules. It presents the impact of rule extraction for the predictive analysis in decision making.

INTRODUCTION

Data mining has become the most prominent and assuring methodology for decision making. It can be used for extracting various rules in any historical data set and present them as an approach for efficient predictions. Rule extraction using data mining approaches is a tool for efficient decision making where preprocessed historical records like weather reports, healthcare data, geospatial data, sales records etc. act as an input for training and generating the rules through analysis. Based upon generating rules, predictions and decision making can be done in the respective areas. The whole process of rule extraction is divided into various phases (Mishra, Addy, Roy, & Dehuri, 2011), which can be customized accordingly. There are different paradigms of rule extraction including association rule extraction such as Apriori algorithm (Agrawal & Srikant, 1994), decision tree (Apté & Weiss, 1997), hypothesis testing, rough set rules and many other algorithms. Apriori algorithm has several application areas like educational data mining (Yang & Hu, 2011) that helps arranging courses, quality education and educational model. Other application areas include medical domain (Yuguang & Chunyan, 2011), Electric Multiple unit fault data analysis (Zhang, Xie, Zhang, Li, & Liu, 2011), Electronic Commerce (Yang, 2012). Rule extraction using classification has become

a very promising technique in various domains for bringing innovative touches including speech recognition (Zhou, Kang, Fan, & Zhang, 2011), real estate development scheme optimization (Wang, 2013) and fraud detection (Zou, Sun, Yu, & Liu, 2012) etc. Especially decision tree has vast applications including economical statistical data processing (Jinguo & Chen, 2011), electric power marketing (Meng & Yang, 2012) etc. It is being used in several other application domains like market segmentation, prediction, fraud detection, weather forecasting, trend analysis, time series analysis and interactive marketing etc. Various rule extraction tools are available in accordance with the requirements, which are categorized according to different approaches (Laender, Ribeiro-Neto, da Silva, & Teixeira, 2002) like natural language processing tools, modeling and ontology based tools etc.

Our world is full of records and facts. There is a bit difference between facts and records. All records are not facts; we have to extract facts from records. These facts only help to reach a decision in different areas. Thus keeping this extraction of fact from records several models have been proposed but most of them lack implementation. The major reason behind their failure was the lack of transparency. Transparency is nothing but a measure of quality of rules extraction. Intelligent and efficient process of discovering and extracting the logical rules out of records is known as *Rules Extraction*. Decision making is predictive in nature. Facts are discovered using the rule extraction and these facts motivate us to reach a prediction of a decision. Therefore there exists a direct relation between rules extraction and prediction, if the rules are extracted efficiently then our predictions will also be efficient.

The rules can be extracted in two ways:

1. Top down extraction.
2. Bottom up extraction.

Rule extraction starts from the top level and then extract rules for sub modules, such mode of top to down rule extraction is known as top-down extraction while extracting rules for individual sub modules first and then aggregate them to form a composite rule set, it is known as bottom-up extraction. Extracting the facts from the bulk of the records, particular data set is used as a reference for extracting certain rules, this is known as *Training Set* or *Trained Data set*. Once the rules are discovered from the training set, another data set is used to apply those discovered rules on it and perform the predictions; this data set is called Test *Set* (Han, Kamber, & Pei, 2011).

PHASES OF RULES EXTRACTION

Figure 1 shows the phases of rule extraction (Mishra et al., 2011):

- **Rule Requirement:** It covers the rule's requirement for the class generation.
- **Rule Generation:** Based on the requirements and domain constraint rules are generated.
- **Rule Coverage:** It ensures that all the required rules are generated and also all the generated rules are required.
- **Rule Selection:** This is the final phase where all the selected rules are listed and applied to the available dataset to test the accuracy.

PARADIGM OF RULE EXTRACTION

Associative Rule Extraction

This paradigm deals with the extraction of the rules on the basis of the observable associative relationship between the attributes which can

Related Content

iTrade: A Mobile Data-Driven Stock Trading System with Concept Drift Adaptation
Yong Hu, Xiangzhou Zhang, Bin Feng, Kang Xieand Mei Liu (2015). *International Journal of Data Warehousing and Mining (pp. 66-83).*
www.irma-international.org/article/itrade/122516

HASTA: A Hierarchical-Grid Clustering Algorithm with Data Field
Shuliang Wangand Yasen Chen (2014). *International Journal of Data Warehousing and Mining (pp. 39-54).*
www.irma-international.org/article/hasta/110385

Dimensionality Reduction with Unsupervised Feature Selection and Applying Non-Euclidean Norms for Classification Accuracy
Amit Saxenaand John Wang (2010). *International Journal of Data Warehousing and Mining (pp. 22-40).*
www.irma-international.org/article/dimensionality-reduction-unsupervised-feature-selection/42150

Issues in Feathers Application in the Seoul Metropolitan Area
Won Do Lee, Chang-Hyeon Joh, Sungjin Choand Bruno Kochan (2014). *Data Science and Simulation in Transportation Research (pp. 71-85).*
www.irma-international.org/chapter/issues-in-feathers-application-in-the-seoul-metropolitan-area/90066

Privacy Preserving Data Mining, Concepts, Techniques, and Evaluation Methodologies
Igor Nai Fovino (2008). *Successes and New Directions in Data Mining (pp. 277-301).*
www.irma-international.org/chapter/privacy-preserving-data-mining-concepts/29963