Chapter III
Introduction to Fuzzy Data Mining Methods

Balazs Feil
University of Pannonia, Hungary

Janos Abonyi
University of Pannonia, Hungary

ABSTRACT

This chapter aims to give a comprehensive view about the links between fuzzy logic and data mining. It will be shown that knowledge extracted from simple data sets or huge databases can be represented by fuzzy rule-based expert systems. It is highlighted that both model performance and interpretability of the mined fuzzy models are of major importance, and effort is required to keep the resulting rule bases small and comprehensible. Therefore, in the previous years, soft computing based data mining algorithms have been developed for feature selection, feature extraction, model optimization, and model reduction (rule based simplification). Application of these techniques is illustrated using the wine data classification problem. The results illustrate that fuzzy tools can be applied in a synergistic manner through the nine steps of knowledge discovery.

INTRODUCTION

In our society, the amount of data doubles almost every year. Hence, there is an urgent need for a new generation of computational techniques and tools to assist humans in extracting useful information (knowledge) from the rapidly growing volumes of data. When we attempt to solve real-world problems, like extracting knowledge from large amounts of data, we realize that they are typically ill-defined systems, difficult to model, and with largescale solution spaces. In these cases, precise models are impractical, too expensive, or nonexistent. Furthermore, the relevant available information is usually in the form of empirical prior knowledge and input-output data representing instances of the system’s behavior. Therefore, we need an approximate reasoning system capable of handling such imperfect information. Computational intelligence (CI) and soft computing (SC) are recently coined terms describing the use of many emerging computing disciplines. According to Zadeh (1994),
“in contrast to traditional, hard computing, soft computing is tolerant of imprecision, uncertainty, and partial truth.” In this context, fuzzy logic (FL), probabilistic reasoning (PR), neural networks (NNs), and genetic algorithms (GAs) are considered main components of CI. Each of these technologies provides us with complementary reasoning and searching methods to solve complex, real-world problems. What is important to note is that soft computing is not a melange. Rather, it is a partnership in which each of the partners contributes a distinct methodology for addressing problems in its domain. In this perspective, the principal constituent methodologies in CI are complementary rather than competitive.

As the title of this chapter shows, its aim is to give an overview about fuzzy data mining methods. It should be emphasized that this title is equivocal in some sense because it has two meanings. Fuzzy data mining methods can mean data mining methods that are fuzzy methods as well; on the other hand, it can also mean approaches to analyze fuzzy data. In some sense, the later ones are fuzzy methods as well but the conceptions are different. Fuzzy data mean imprecise, vague, uncertain, ambiguous, inconsistent, and/or incomplete data. Therefore, the source of uncertainty is the data themselves. It is very important to develop methods that are able to handle this kind of data because data from several information sources might be fuzzy (e.g., from human expert who describes their knowledge in natural language). In Giordani and Kiers (2004), a modified principal component analysis is presented for symmetric fuzzy data. Krol, Kukla, Lasota, and Trawinski (2006) deal with models on the basis of fuzzy data. There are algorithms to cluster, classify, or visualize fuzzy data (e.g., Bandemer, 2006; Esogbue, 1986; Gershon, 1992; Pang, Wittenbrink, & Lodha, 1997). This chapter deals with data mining methods based on fuzzy techniques. The data are crisp and can be given in absolute (\(N \times n\) matrix) or relative form (\(N \times N\) matrix, where \(N\) and \(n\) are the numbers of samples and attributes, respectively. By absolute data, the values of the attributes are given. “Relative data” means that the data’s values are not known, but their pairwise distance is known. The approaches the data are analyzed with handle the uncertainty on the basis of fuzzy logic; for example, by clustering a problem, a sample can be the member of each cluster simultaneously with different degrees between 0 and 1. In the following, several fuzzy methods will be presented for several kinds of problems.

The remainder of this chapter is organized as follows. The aim of the Steps of Knowledge Discovery section is to show how the elements of CI can be used in data mining and how fuzzy information processing can be situated within this general and comprehensive process. In the remaining sections, basic definitions, widely applied methods and tools for clustering (Classical Fuzzy Cluster Analysis section), visualization (Visualization of High Dimensional Data section), classification (Fuzzy Classifier Systems for Effective Model Representation section), and association rule mining (Fuzzy Association Rule Mining section) are discussed, including the related knowledge representation, identification, and reduction methods. A detailed bibliographical view of the methods and tools is given. The mentioned approaches are evaluated using illustrative examples. Finally, conclusions are given.

**STEPS OF KNOWLEDGE DISCOVERY**

According to Fayyad, Piatetsk-Shapiro, and Smyth (1996), “historically the notion of finding useful patterns in data has been given a variety of names including data mining, knowledge extraction, information discovery, and data pattern processing. The term data mining has been mostly used by statisticians, data analysts, and the management information systems (MIS) communities.” The term knowledge discovery in databases (KDD) refers to the overall process of discovering knowledge from data, while data mining refers to a particular step of this process. Data mining is
39 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:
www.igi-global.com/chapter/introduction-fuzzy-data-mining-methods/20349?camid=4v1

This title is available in InfoSci-Books, InfoSci-Database Technologies, Business-Technology-Solution, Library Science, Information Studies, and Education, InfoSci-Library and Information Science. Recommend this product to your librarian:
www.igi-global.com/e-resources/library-recommendation/?id=1

Related Content

**Dynamic Semantics of UML State Machines: A Metamodeling Perspective**
www.igi-global.com/article/dynamic-semantics-uml-state-machines/3285?camid=4v1a

**Adoption, Improvement, and Disruption: Predicting the Impact of Open Source Applications in Enterprise Software Markets**
Michael Brydon and Aidan R. Vining (2008). *Journal of Database Management* (pp. 73-94).
www.igi-global.com/article/adoption-improvement-disruption/3386?camid=4v1a

**3-D Virtual Worlds in Education: Applications, Benefits, Issues, and Opportunities**
www.igi-global.com/article/virtual-worlds-education/3397?camid=4v1a

**Preparing Clinical Text for Use in Biomedical Research**
www.igi-global.com/article/preparing-clinical-text-use-biomedical/3350?camid=4v1a