Chapter III

Generalization Data Mining in Fuzzy Object-Oriented Databases

Rafal Angryk
Tulane University, USA

Roy Ladner
Naval Research Laboratory, USA

Frederick E. Petry
Tulane University & Naval Research Laboratory, USA

Abstract

In this chapter, we consider the application of generalization-based data mining to fuzzy similarity-based object-oriented databases (OODBs). Attribute generalization algorithms have been most commonly applied to relational databases, and we extend these approaches. A key aspect of generalization data mining is the use of a concept hierarchy. The objects of the database are generalized by replacing specific attribute values by the next higher-level term in the hierarchy. This will then eventually result in generalizations that represent a summarization of the information in the database. We focus on the generalization of similarity-based simple fuzzy attributes for an OODB using approaches to the fuzzy concept hierarchy

developed from the given similarity relation of the database. Then consideration is given to applying this approach to complex structure-valued data in the fuzzy OODB.

Introduction

Data mining and knowledge discovery have increasing importance as the amount of data from various sources has rapidly increased. Awash in such volumes of data, data mining techniques attempt to make sense of this data by formulating information of value for decision making. This can vary from deciding on commercial sales promotions to environmental planning to national security decisions. Much of the current work is in the context of conventional relational databases. In this chapter, we will discuss how to apply one valuable data mining approach — attribute-oriented generalization — to a similarity-based fuzzy OODB.

Background

In this section, we survey the general area of data mining, discuss some of the relevant work in fuzzy data mining, and then describe the specific technique of attribute-oriented induction for generalization, which is the focus of this chapter. Additionally, we describe the fuzzy object-oriented model based on similarity relationships that is the context in which we investigate data generalization.

Data Mining

Data mining or knowledge discovery generally refers to a variety of techniques that have developed in the fields of databases, machine learning, and pattern recognition. The intent is to uncover useful patterns and associations from large databases.

Although we are primarily interested here in specific algorithms for knowledge discovery, we will first review the overall process of data mining (Feelders, Daniels, & Holsheimer, 2000). The initial steps of data mining are concerned with preparation of data, including data cleaning intended to resolve errors and missing data and integration of data from multiple heterogeneous sources. Next are the steps needed to prepare for actual data mining. These include selection

Related Content

Optimal Information Acquisition and Sharing Decisions: Joint Reviews on Crowdsourcing Product Design

Jizi Li, Xiaodie Wang, Justin Z. Zhangand Longyu Li (2024). *Journal of Database Management (pp. 1-34).*

www.irma-international.org/article/optimal-information-acquisition-and-sharing-decisions/337971

A Quick Presentation of Evolutionary Computation

Pierre Collet (2010). *Soft Computing Applications for Database Technologies: Techniques and Issues (pp. 22-38).*

www.irma-international.org/chapter/quick-presentation-evolutionary-computation/44380

Business Data Warehouse: The Case of Wal-Mart

Indranil Bose, Lam A.K. Chun, Leung V.W. Yue, Li H.W. Inesand Wong O.L. Helen (2006). *Database Modeling for Industrial Data Management: Emerging Technologies and Applications (pp. 244-257).*

www.irma-international.org/chapter/business-data-warehouse/7894

The Expert's Opinion

Mohammad Dadashzadeh (1991). *Journal of Database Administration (pp. 38-40).*

www.irma-international.org/article/expert-opinion/51097

Service Composition and Interaction in a SOC Middleware Supporting Separation of Concerns with Flows and Views

Dickson K.W. Chiu, Qing Li, Patrick C. K. Hung, Zhe Shan, S. C. Cheung, Yu Yangand Matthias Farwick (2011). *Journal of Database Management (pp. 32-63).*

www.irma-international.org/article/service-composition-interaction-soc-middleware/52992