Chapter 4.30
Fuzzy Miner:
Extracting Fuzzy Rules from Numerical Patterns[*]

Nikos Pelekis
University of Piraeus, Greece
UMIST Manchester, UK

Babis Theodoulidis
UMIST Manchester, UK

Ioannis Kopanakis
UMIST Manchester, UK

Yannis Theodoridis
University of Piraeus, Greece

ABSTRACT

We study the problem of classification as this is presented in the context of data mining. Among the various approaches that are investigated, we focus on the use of Fuzzy Logic for pattern classification, due to its close relation to human thinking. More specifically, this paper presents a heuristic fuzzy method for the classification of numerical data, followed by the design and the implementation of its corresponding tool (Fuzzy Miner). The initial idea comes from the fact that fuzzy systems are universal approximators of any real continuous function. Such an approximation method coming from the domain of fuzzy control is appropriately adjusted into pattern classification and an "adaptive" procedure is proposed for deriving highly accurate linguistic if-then rules. Extensive simulation tests are performed to demonstrate the performance of Fuzzy Miner, while a comparison with a neuro-fuzzy classifier of the area is taking place in order to contradict the methodologies and the corresponding outcomes. Finally, new research directions in the context of Fuzzy Miner are identified, and ideas for its improvement are formulated.

INTRODUCTION

Recently, our capabilities of both generating and collecting data have increased rapidly. Consequently, data mining has become a research area with increasing importance. Data mining, also referred to as knowledge discovery in databases (Chen et al., 1996), is the search of relationships and global patterns that exist "hidden" among vast amounts of data. There are various problems that someone has to deal with when extracting knowledge from data, including characterization, comparison, association, classification, prediction, and clustering (Han & Kamber, 2001). This paper elaborates on the problem of *classification*. Broadly speaking, pattern classification (or recognition) is the science that is concerned with the description or classification of measurements. More technically, pattern classification is the process that finds the common properties among a set of objects in a database and classifies them into different classes according to a classification model.

Classical models usually try to avoid *vague, imprecise*, or *uncertain* information, because it is considered as having a negative influence in the inference process. This paper accepts the challenge of dealing with such kind of information by introducing a fuzzy system, which deliberately makes use of it. The main idea of fuzzy systems is to extend the classical two-valued modeling of concepts and attributes like *tall, fast*, or *old* in a sense of gradual truth. This means that a person is not just viewed as *tall* or *not tall,* but as tall to a certain degree between 0 and 1. This usually leads to simpler models, which are handled more easily and are more familiar to the human way of thinking.

After providing a brief comparative overview of pattern classification approaches (Section 2) and a short specification of the pattern classification domain in fuzzy systems (Section 3), the paper follows the above paradigm and describes an effective fuzzy system for the classification of numerical data (Section 4). The initial idea comes from the fact that fuzzy systems are universal approximators (Kosko, 1992; Wang, 1992) of any real continuous function. Such an approximation method (Nozzaki et al., 1997) coming from the domain of fuzzy control systems is appropriately adjusted, extended, and implemented in order to produce a powerful working solution in the domain of pattern classification. An "adaptive" process is also introduced, developed, and incorporated into the previous mechanism for automatically deriving highly accurate linguistic if-then rules. The description of the methodology is combined with the illustration of the design issues of the tool *Fuzzy Miner.* The current work is evaluated (Section 5) by extensive simulation tests and by providing a comparison framework with another tool of the domain that employs a neuro-fuzzy approach, NEFCLASS (Nauck & Kruse, 1995). Finally, the paper concludes (Section 6) by identifying promising directions for future work pointed to by this research effort.

COMPARATIVE OVERVIEW OF PATTERN CLASSIFICATION APPROACHES

Already, when the field was still in its very infancy, it was realized that the statistics and probability theory (Berger, 1985) had much to offer to pattern classification (Schalkoff, 1992). The question of whether or not a given pattern "belongs" to some pattern class may naturally be treated as a special case of the statistical decision theory problem. Effective, though, as it is, the statistical approach has built-in limitations. For instance, the theory of testing statistical hypotheses entails that a clear-cut yes or no answer should always decide upon the membership of a pattern in a given class. Clearly, not all of the real life patterns admit of such coarse decisions. Sometimes information in a pattern is not simply in the presence or the absence of a set of features, but rather the interconnection of fea-

21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/fuzzy-miner-extracting-fuzzy-rules/7753

Related Content

Materialized Hypertext View Maintenance
Giuseppe Sindoni (2005). *Encyclopedia of Data Warehousing and Mining (pp. 710-713).*
www.irma-international.org/chapter/materialized-hypertext-view-maintenance/10689

Mining for Web-Enabled E-Business Applications
Richi Nayak (2005). *Encyclopedia of Data Warehousing and Mining (pp. 785-789).*
www.irma-international.org/chapter/mining-web-enabled-business-applications/10703

Reference Scheme Modeling
Terry Halpin (2019). *New Perspectives on Information Systems Modeling and Design (pp. 227-254).*
www.irma-international.org/chapter/reference-scheme-modeling/216340

Duplicate Record Detection for Data Integration
 (2014). *Innovative Techniques and Applications of Entity Resolution (pp. 339-358).*
www.irma-international.org/chapter/duplicate-record-detection-for-data-integration/103256

Homeland Security Data Mining and Link Analysis
Bhavani Thuraisingham (2005). *Encyclopedia of Data Warehousing and Mining (pp. 566-569).*
www.irma-international.org/chapter/homeland-security-data-mining-link/10661