Fuzzy–Rough Data Mining

Richard Jensen
Aberystwyth University, UK

INTRODUCTION

It is estimated that every 20 months or so the amount of information in the world doubles. In the same way, tools that mine knowledge from data must develop to combat this growth. Such techniques must be able to extract useful, meaningful patterns efficiently in the presence of large amounts of noisy, redundant and sometimes misleading information. To be able to achieve this, data mining is often reliant upon suitable data cleaning and reduction processes to allow a better quality of knowledge to be discovered. Indeed, the computational complexity of some techniques prohibits their application to large volumes of data, and the reduction step is a necessity. Fuzzy-rough set theory provides a framework for developing such applications in a way that combines the best properties of fuzzy sets and rough sets, in order to handle uncertainty. As will be discussed later, fuzzy sets and rough sets both model different aspects of uncertainty encountered in the real world. Their combination can therefore result in powerful methods for data mining and reduction.

This chapter presents a general overview of recent fuzzy-rough data mining developments, including feature selection and classifier learning.

BACKGROUND

Lately there has been great interest in developing methodologies which are capable of dealing with imprecision and uncertainty, and the resounding amount of research currently being done in the areas related to fuzzy (Zadeh, 1965) and rough sets (Pawlak, 1991) is representative of this. The success of rough set theory is due in part to three aspects of the theory. Firstly, only the facts hidden in data are analyzed. Secondly, no additional information about the data is required for data analysis such as thresholds or expert knowledge on a particular domain. Thirdly, it finds a minimal knowledge representation for data. As rough set theory handles only one type of imperfection found in data, it is complementary to other concepts for the purpose, such as fuzzy set theory. The two fields may be considered analogous in the sense that both can tolerate inconsistency and uncertainty – the difference being the type of uncertainty and their approach to it.

Fuzzy sets are concerned with modeling the vagueness that is present in real world data through the extension of classical set theory (with binary set membership) to the case of gradual memberships, i.e. elements belong to (fuzzy) sets to a certain degree. This allows greater flexibility in modeling concepts that are inherently vague. For example, the concept “Tall,” where elements are people, could be defined differently by different people based on their subjective views of what constitutes tallness. Indeed, it is a valid observation that there are degrees of tallness: some people are definitely tall, however other people might be slightly tall or moderately tall. This vagueness is difficult to model with classical set theory, but straightforward to do so with fuzzy set theory. With this basis, tools and techniques can be developed for representation and reasoning.

Rough set theory has been used as a tool to discover data dependencies, induce rules, and to reduce the number of features contained in a dataset using the data alone, requiring no additional information (Pawlak, 1991; Polkowski, 2002; Skowron et al., 2002). The rough set itself is the approximation of a vague concept (set) by a pair of precise

DOI: 10.4018/978-1-4666-5202-6.ch092

Copyright © 2014, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

concepts, called lower and upper approximations, which are a classification of the domain of interest into disjoint categories. The lower approximation is a description of the domain objects which are known with certainty to belong to the subset of interest, whereas the upper approximation is a description of the objects which possibly belong to the subset. The approximations are constructed with regard to a particular subset of features, and works by making use of the granularity structure of the data only. This is a major difference when compared with Dempster-Shafer theory and fuzzy set theory which require probability assignments and membership values respectively. However, this does not mean that no model assumptions are made. In fact by using only the given information, the theory assumes that the data is a true and accurate reflection of the real world (which may not be the case).

Many relationships have been established between fuzzy sets and rough sets (Dubois & Prade, 1992; Radzikowska & Kerre, 2002) and more so, most of the recent studies have concluded at this complementary nature of the two methodologies, especially in the context of granular computing (Hu et al., 2007; Skowron et al., 2012). Therefore, it is desirable to extend and hybridize the underlying concepts to deal with additional aspects of data imperfection. Such developments offer a high degree of flexibility and provide robust solutions and advanced tools for data analysis.

MAIN FOCUS

Research on the hybridization of fuzzy sets and rough sets emerged in the early 1990s (Dubois & Prade, 1992) and has flourished recently. It has focused predominantly on fuzzifying the formulas for the lower and upper approximations, so that data objects can belong to given concept to varying degrees. The approximate equality of objects is used, rather than crisp indiscernibility. As a result, objects are categorized into classes, or granules, with soft boundaries based on their similarity to one another. As such, abrupt transitions between classes are replaced by gradual ones, allowing that an element can belong (to varying degrees) to more than one class.

A fuzzy-rough set is therefore defined by two fuzzy sets, fuzzy lower and upper approximations, obtained by extending the corresponding crisp rough set notions. In the crisp case, elements that belong to the lower approximation (i.e. have a membership of 1) are said to belong to the approximated set with absolute certainty. In the fuzzy-rough case, elements may have a membership in the range [0,1], allowing greater flexibility in handling uncertainty. With this basis, powerful techniques can be developed for data mining.

Fuzzy-Rough Feature Selection

Feature selection addresses the problem of selecting those input features that are most predictive of a given outcome; a problem encountered in many areas of computational intelligence. Unlike other dimensionality reduction methods, feature selectors preserve the original meaning of the features after reduction. This has found application in tasks that involve datasets containing huge numbers of features (in the order of tens of thousands) which, for some learning algorithms, might be impossible to process further. Recent examples include text processing and Web content classification (Jensen & Shen, 2008).

There are often many features involved, and combinatorially large numbers of feature combinations, to select from. Note that the number of feature subset combinations with m features from a collection of N total features is N!/[m!(N - m)!]. It might be expected that the inclusion of an increasing number of features would increase the likelihood of including enough information to distinguish between classes. Unfortunately, this is not necessarily true if the size of the training dataset does not also increase rapidly with each additional feature included. A high-dimensional dataset increases the chances that a learning algorithm will find spurious patterns that are not

6 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-rough-data-mining/107300

Related Content

Web Content Analysis of Online Grocery Shopping Web Sites in India
Tanushri Banerjeeand Arindam Banerjee (2018). *International Journal of Business Analytics (pp. 61-73).*
www.irma-international.org/article/web-content-analysis-of-online-grocery-shopping-web-sites-in-india/212635

Challenges of the Implementation of Research, Development, and Innovation Standards: A Case Study From a Glass Bottle Manufacturer
António Carrizo Moreiraand Alexandra Goreti Figueira Evangelista (2018). *Handbook of Research on Strategic Innovation Management for Improved Competitive Advantage (pp. 511-538).*
www.irma-international.org/chapter/challenges-of-the-implementation-of-research-development-and-innovation-standards/204239

A Prescriptive Stock Market Investment Strategy for the Restaurant Industry using an Artificial Neural Network Methodology
Gary R. Weckman, Ronald W. Dravenstott, William A. Young II, Ehsan Ardjmand, David F. Millieand Andy P. Snow (2016). *International Journal of Business Analytics (pp. 1-21).*
www.irma-international.org/article/a-prescriptive-stock-market-investment-strategy-for-the-restaurant-industry-using-an-artificial-neural-network-methodology/142778

Facilitating the Use of Intellectual Capital in a Matrix Multinational Organization
Alan M. Thompson (2010). *Strategic Intellectual Capital Management in Multinational Organizations: Sustainability and Successful Implications (pp. 1-16).*
www.irma-international.org/chapter/facilitating-use-intellectual-capital-matrix/36453

Internationalization Technology and IPO Firm Growth
Fiona Xiaoying Jiand Steve Bin Zhou (2014). *Encyclopedia of Business Analytics and Optimization (pp. 1310-1322).*
www.irma-international.org/chapter/internationalization-technology-and-ipo-firm-growth/107328