

Chapter 51

Multi-Objective Genetic and Fuzzy Approaches in Rule Mining Problem of Knowledge Discovery in Databases

Harihar Kalia

Seemanta Engineering College, India

Satchidananda Dehuri

Ajou University, Korea

Ashish Ghosh

Indian Statistical Institute, India

ABSTRACT

Knowledge Discovery in Databases (KDD) is the process of automatically searching patterns from large volumes of data by using specific data mining techniques. Classification, association, and associative classification (integration of classification and association) rule mining are popularly used rule mining techniques in KDD for harvesting knowledge in the form of rule. The classical rule mining techniques based on crisp sets have bad experience of “sharp boundary problems” while mining rule from numerical data. Fuzzy rule mining approaches eliminate these problems and generate more human understandable rules. Several quality measures are used in order to quantify the quality of these discovered rules. However, most of these objectives/criteria are conflicting to each other. Thus, fuzzy rule mining problems are modeled as multi-objective optimization problems rather than single objective. Due to the ability of finding diverse trade-off solutions for several objectives in a single run, multi-objective genetic algorithms are popularly employed in rule mining. In this chapter, the authors discuss the multi-objective genetic-fuzzy approaches used in rule mining along with their advantages and disadvantages. In addition, some of the popular applications of these approaches are discussed.

DOI: 10.4018/978-1-4666-6042-7.ch051

INTRODUCTION

Knowledge Discovery in Databases (KDD) (Fayyad, Piatetsky-Shapiro, & Smyth, 1996) refers to the non-trivial, iterative, and interactive process of discovering high-level/abstract knowledge from low level/raw datasets in the form of patterns, models, rules, etc., which are convincing, narrative, potentially useful, and comprehensible.

Data mining (Cios, Pedrycz, Swiniarski, & Kurgan, 2012) is used as a central step of KDD process, performed either through semi-automatic, or automatic methods. In particular, it involves inferring algorithms for tasks like classification, association rule mining, associative classification, clustering, data visualization, and sequential pattern analysis (Han & Kamber, 2006) for exploring the data and discovering previously unknown patterns.

Fuzzy logic/fuzzy sets (Zadeh, 1988) used in knowledge discovery process, make them more flexible in comparison to classical hard computing techniques and make the system tolerant to vagueness, uncertainty, biased, and approximation. Furthermore, it makes use of this tolerance in achieving tractability, robustness and low cost solution to real-world problems. Use of linguistic variables and linguistic terms used in knowledge mining/rule mining process enhance the knowledge representation and facilitates the interpretation of knowledge/rules in linguistic terms as well as successfully evades unnatural boundaries arise in the partitioning of the attribute domains.

Rule mining (Fayyad, Piatetsky-Shapiro, & Smyth, 1996) is an emerging approach in data mining employed to handle and analyze the huge amount of data in order to extract the wealth of knowledge. It is a process of discovering knowledge in the form of IF-THEN rules from large datasets. These knowledge representation schemes in IF-THEN rule format are popular as the discovered knowledge is represented at a high level of abstraction using few logical conditions, which make knowledge representation more comprehensible.

The IF-THEN rule is an expression of the form:

IF (*A given set of conditions are satisfied*) THEN
Conclusion(s) (1)

The “IF” part of the rule is called antecedent or precondition, which consists of one or more attribute tests that are logically connected and “THEN” part is called consequent.

Basing on the types of set utilized in attribute representation the IF-THEN are classified as: crisp rules which uses crisp set and fuzzy rules which uses fuzzy sets.

Crisp rules are best suited for categorical/ Boolean data. In case of quantitative/continuous data, these rules use a sharp cutoff in selecting attributes, which is a major drawback.

As an example, let us consider a rule used for a job selection;

IF (*Age* ≥ 18) AND (*Height* ≥ 165) AND (*Weight* ≥ 55) THEN *Selected* (2)

In this rule, the candidate either with one day less than age 18 or having height 164.9cm or having weight 54.95kg is not selected, which is very unsympathetic.

On the contrary fuzzy rules uses linguistic variables and linguistic values that are defined by context-dependent fuzzy sets whose meanings are spell out by membership functions. In particular very-young, young, middle-aged, old and very-old for age attribute, short, medium and tall for height attribute and low-weight, medium-weight and high-weight for weight attribute and determine the membership degree in $[0,1]$ of the attribute using fuzzy set. Hence the height of 164 is regarded as tall with membership degree less than 165 as tall.

Hence using fuzzy set the above crisp rule towards job selection may look like;

IF (*Age is Young*) AND (*Height is Tall*) (*Weight is Medium*) THEN *Selected* (3)

30 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/multi-objective-genetic-and-fuzzy-approaches-in-rule-mining-problem-of-knowledge-discovery-in-databases/108765

Related Content

Musical Information Dynamics as Models of Auditory Anticipation

Shlomo Dubnov (2011). *Machine Audition: Principles, Algorithms and Systems* (pp. 371-397).

www.irma-international.org/chapter/musical-information-dynamics-models-auditory/45494

Tools for the Automatic Generation of Ontology Documentation: A Task-Based Evaluation

Silvio Peroni, David Shotton and Fabio Vitali (2014). *Computational Linguistics: Concepts, Methodologies, Tools, and Applications* (pp. 839-865).

www.irma-international.org/chapter/tools-for-the-automatic-generation-of-ontology-documentation/108754

Digital Speech Technology: An Overview

H. S. Venkatagiri (2010). *Computer Synthesized Speech Technologies: Tools for Aiding Impairment* (pp. 28-49).

www.irma-international.org/chapter/digital-speech-technology/40857

Can Computers Create Comics and Animations?

Miki Ueno, Kiyohito Fukuda and Naoki Mori (2020). *Natural Language Processing: Concepts, Methodologies, Tools, and Applications* (pp. 1058-1084).

www.irma-international.org/chapter/can-computers-create-comics-and-animations/239979

Portable Text Summarization

Martin Hassel and Hercules Dalianis (2012). *Applied Natural Language Processing: Identification, Investigation and Resolution* (pp. 17-32).

www.irma-international.org/chapter/portable-text-summarization/61040