Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 35 (2014) 701 – 709

Procedia Computer Science

18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems - KES2014

Rough-Fuzzy classifier modeling using data repository sets

Jiri Krupka*, Pavel Jirava

Institute of System Engineering and Informatics, Faculty of Economics and Administration, University of Pardubice, Studentska 84, 532 10 Pardubice, The Czech Republic

Abstract

This paper reflects the trends of the past years based on the diffusion of various traditional approaches and methods when tackling new problems. Two components of the computational intelligence (CI) are applied, rough and fuzzy sets theory. These components permit one to operate with uncertainty data. The current knowledge in the investigated field is summarized and briefly explained. It also deals with uncertainty in an information system and the two approaches, the fuzzy sets (FSs) and rough sets theory (RST), for operating it. The proposal and implementation of a rough-fuzzy classifier (RFC) is modified. RFC uses the rules generated by RSTbox. The databases IRIS and WINE were chosen for verification. The classification results were compared with the results of other classification methods are applied on these databases. Finally, we summarized the presented problems. Based on the above stated facts it can be claimed that the proposed modified algorithm, RSTbox and RFC model are functional. The model is relatively successful (compared to other approaches), and by using it two classification databases can be carried out. This model is proposed in MATLAB.

Peer-review under responsibility of KES International.

Keywords: Classification; fuzzy inference system; rough-fuzzy approach; rules generation; UCI data repository

1. Introduction

In this section we will review the basic concepts and definitions of FSs [1-3], rough sets [4-6], and fuzzy-rough approach [7-10] in relation to the process classification of systems. Systems can be usually described and defined

* Corresponding author. Tel.: +420 466 036 144; fax: +420 466 036 010.
E-mail address: Jiri.Krupka@upce.cz.

Peer-review under responsibility of KES International.
doi:10.1016/j.procs.2014.08.152

by humans. This section analyses the past years trend in the diffusion of various traditional methods and approaches when tackling new problems.

CI methods are used in creating of system models. Areas of CI (FSs, neural networks, genetic algorithms, rough sets etc.) belong to a fast developing field in the applied research. It is composed of several theories and approaches which despite being different from one another, have two common denominators. They are the non-symbolic representation of pieces of knowledge [11] and 'bottom-up' architecture, in which the structures and paradigms appear from an unordered beginning [4], [11].

These two denominators have been successfully used in various uncertainty information processing systems. The RST [12-14], attributed by prof. Pawlak, is based on the research in the logical properties of information systems, and the uncertainty in information systems which are expressed by a boundary region. Every investigated object is related to a specific piece of information. The objects which are characterized by the same pieces of information are mutually undistinguishable from the point of view of the accessible pieces of information. This is expressed in RST by the indiscernibity relations. The theory of FSs, attributed by prof. Zadeh, is a known approach to uncertainty. In this theory an element belongs to a set according to the membership function values (membership degree) [3], [15-16], i.e. to a closed interval. Theory of FSs is an expansion of the traditional sets theory in which an element either is or is not a set member. If we attempt to describe and model a particular reality problem we encounter a certain discrepancy. On one hand there is the accuracy of mathematical methods by which a specific problem is described, and on the other hand there is a very complicated reality demanding a range of simplifications and the consequent inaccuracy, infidelity of the model arising from them.

RST and FSs are applied in classifier modeling. The goal of the paper is a synthesis and analysis of the original RFC model with using of the RST tool (henceforth called RSTbox).

2. Classification based on rough and fuzzy sets

The role of a classification is to classify objects, events, and real-life situations into classes. Each of the reviewed objects is unique, original, and its classification means a certain degree of generalization. Let's define a system for the particular objects, i.e. input and output variables, elements (objects), and their mutual relations. Defining and collecting the data of input/output variables cannot be generalized, even though this stage influences the classification result. An application of classification methods based on CI represents an effective tool for the realization of a classification model.

On the basis of achieved classification results, it seems to be effective and up-to-date to tackle the classification problem using a hybrid approach combining rough sets and FSs. Both of them belong to the field of the CI research.

The application of the classification methods based on CI represents an effective tool for the classification model implementation [6], [17]. For example, we can speak about probabilistic rough classifiers [18], fuzzy classifiers [19] etc. The probabilistic rough classifier combines all positive aspects of rule induction systems with the flexibility of statistical techniques for classification. Two natural approaches [19] to classifier design are: to ask experts how they solve the problem and try to encapsulate the knowledge in a fuzzy-base classifier; to collect input-output data (i.e. a labeled data set) and extract the classifier parameters from the data. The first model represents transparent approach (is interpretable in the domain context) and the second based on data, may or may not be interpretable. Fuzzy classifier models are deemed to be able to integrate both approaches: human and data sources.

On the basis of specialized literature we can define some known interesting approaches to rough fuzzy hybridization [20-26].

They are divided into two groups - supervised and unsupervised learning [20]. Referring to the former, we can speak about supervised learning and information retrieval. In [27] a fuzzy-rough ownership function that involves