Knowledge-Based Systems 70 (2014) 281–292

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Attribute analysis of information systems based on elementary soft implications

Feng Feng [a,*], Muhammad Akram [b], Bijan Davvaz [c], Violeta Leoreanu-Fotea [d]

[a] Department of Applied Mathematics, School of Science, Xi'an University of Posts and Telecommunications, Xi'an 710121, China
[b] Department of Mathematics, University of the Punjab, New Campus, Lahore, Pakistan
[c] Department of Mathematics, Yazd University, Yazd, Iran
[d] Faculty of Mathematics, "Al.I.Cuza" University, 700506 Iaşi, Romania

ARTICLE INFO

Article history:
Received 2 October 2013
Received in revised form 15 July 2014
Accepted 15 July 2014
Available online 24 July 2014

Keywords:
Information system
Soft set
Soft truth degree
Attribute analysis
Rough set
Decision making
Association rule

ABSTRACT

Soft set theory provides a parameterized treatment of uncertainty, which is closely related to soft computing models like fuzzy sets and rough sets. Based on soft sets and logical formulas over them, this study aims to present a new approach for revealing the causal relationship between values of attributes in an information system. The main procedure of our new method is as follows: First, we choose the attributes to be analyzed and construct some partition soft sets from a given information system. Then we compute the extended union of the obtained partition soft sets, which results in a covering soft set. Further, we transform the obtained covering soft set into a decision soft set and consider logical formulas over it. Next, we calculate various types of soft truth degrees of elementary soft implications. Finally, we can rank attribute values and plot some illustrative graphs, which helps us extract useful knowledge from the given information system. We use several examples, including a classical example given by Pawlak and a practical application concerning IT applying features analysis, to illustrate the newly proposed method and related concepts. In addition, we compare soft attribute analysis with rough attribute analysis and also relate it to soft association rules mining.

1. Introduction

Mathematical modelling, analysis and computing of problems with uncertainty is one of the hottest areas in interdisciplinary research involving applied mathematics, computational intelligence and decision sciences. It is worth noting that uncertainty arises from various domains has very different nature and cannot be captured within a single mathematical framework. In addition to probability theory and statistics, we currently have some advanced soft computing methods such as fuzzy sets [46], rough sets [39] and soft sets [23] for dealing with different kinds of uncertainty. Generally speaking, probability theory and statistics emphasize randomness inherent in uncertain phenomenon and rely heavily on distributions of random variables. Fuzzy set theory is based on the method of gradualness and describes fuzzy concepts using membership functions. It provides an effective way for modelling vagueness and ambiguity in human reasoning and intelligent decision making process. The indiscernibility relation generated from the collected data is the mathematical basis of Pawlak's rough sets, which treat uncertainty using the method of granulation expressed by rough lower and upper approximations.

Molodtsov's soft sets [23] provide us a new way of coping with uncertainty from the viewpoint of parameterization. It has been demonstrated that soft sets have potential applications in various fields such as the smoothness of functions, game theory, operations research, Riemann integration, Perron integration, probability theory, and measurement theory [23,24]. Since then research on soft sets has been very active and received much attention from researchers worldwide. Many researchers contributed to extending soft sets using fuzzy set theory [21,22]. Feng et al. [16,13] combined soft sets with rough sets and fuzzy sets, obtaining three types of hybrid models: rough soft sets, soft rough sets, and soft-rough fuzzy sets. Ali [5] discussed the fuzzy sets and fuzzy soft sets induced by soft sets. To extend the expressive power of soft sets, Jiang et al. [26] presented ontology-based soft sets, which extended soft sets with description logics. Ali et al. [6] proposed several new operations in soft set theory. Based on these new operations, Qin and Hong [34] introduced some congruence relations

* Corresponding author. Tel.: +86 29 88166086.
E-mail addresses: fengnix@hotmail.com (F. Feng), m.akram@pucit.edu.pk (M. Akram), davvaz@yazduni.ac.ir (B. Davvaz), foteavioleta@gmail.com (V. Leoreanu-Fotea).

http://dx.doi.org/10.1016/j.knosys.2014.07.010
0950-7051/

on soft sets and discussed certain lattice structures. Xiao et al. [44] proposed exclusive disjunctive soft sets and some related operations. Gong et al. [17] initiated the concept of bijective soft sets. Babitha and Sunil [7] extend the concepts of relations and functions in the context of soft set theory. Moreover, many works have been devoted to application of soft sets in various algebraic structures [1–4,28–36,36,40–46,47]. With regard to some practical applications, soft set theory has been used in classification [25], data analysis [48], forecasting [45] and simulation [27]. Moreover, decision making based on soft sets (also known as soft decision making) has received much attention in recent years [8–10,12,14,35].

In most studies of soft set theory and its applications, so far as we know, much attention has been given to objects described by several related parameters. For instance, the application of soft sets in decision making problems was initiated by Maji et al. [19]. Then most published works devoted to soft decision making have followed the research line there, and the central issue is to select the best alternative (i.e., object) according to “choice value” which represents for the number of parameters satisfied by each alternative. However, it can be found that in some cases we should pay more attention to parameters themselves rather than the objects. In fact, in the scenario of data analysis, we usually hope to discover the causal relationship between the values of various attributes (i.e., parameters) from the collected data. For example, we suppose that a mobile phone producer hope to design a new mobile phone product. In order to obtain a successful design, the producer is more concerned about what kinds of features (i.e., parameters) play the most crucial role in producing the best mobile phones rather than which mobile phones (i.e., objects) are most popular since everyone knows the biggest sellers in the mobile phone market are most popular. Moreover, it is also known that one important thing in decision making is to acquire useful knowledge represented in the form of decision rules from some collected data. From a logical point of view, decision rules are logical implications which describes the dependency relationship between values of attributes (i.e., parameters).

This paper is a continuation of ideas presented in our previous paper on soft set based approximate reasoning [15]. We intend to give a new approach to attribute analysis of information systems based on soft sets and logical formulas over them. In contrast to the existing research on soft sets, we emphasize the parameters themselves rather than the objects consisting of the universe of discourse. The main issue under our consideration is how to estimate the causal relationship between the values of two attributes in an information system. To solve this problem, we propose a formal language for reasoning with soft sets, in which parameters are atomic formulas. Decision soft sets are used to represent decision systems and decision rules are defined to be implicative type of formulas over decision soft sets. We refer to simple decision rules as elementary soft implications, whose predecessor and successor are atomic formulas. Several measures for capturing the soft truth degree of formulas are given to evaluate decision rules (especially elementary soft implications) from various aspects such as sufficiency, necessity, strength of evidence, or certain combinations of them. An algorithm for solving our main problem is presented and illustrated by a classical example considered by Pawlak [38]. We also apply our new method to a practical issue regarding IT applying features analysis and compare it with the classical approach based on rough sets.

The rest of this study is organized as follows. To facilitate our discussion, Section 2 recalls some basic notions in soft set theory. Section 3 gives a brief introduction to information systems and their connections with soft sets. Section 4 introduces a logic framework for reasoning with soft sets. Section 5 gives several measures for capturing soft truth quantitatively. Section 6 presents a new method of attribute analysis based on soft sets and logical formulas over them, illustrated by some examples. Section 7 compares soft attribute analysis with rough attribute analysis and show its connections with soft association rules mining. Finally, conclusions and some possible directions for future research are given in the last section.

2. Soft set theory

Soft set theory was proposed by Molodtsov [23] in 1999. This theory provides a parameterized point of view for uncertainty modelling and soft computing. Let U be the universe of discourse and E be the universe of all possible parameters related to the objects in U. In most cases parameters are considered to be attributes, characteristics or properties of objects in U. The pair (U,E) is also known as a *soft universe*. The power set of U is denoted by $\mathcal{P}(U)$.

Definition 2.1 [23]. A pair $\mathfrak{S}=(F,A)$ is called a *soft set* over U, where $A \subseteq E$ and $F: A \to \mathcal{P}(U)$ is a set-valued mapping, called the *approximate function* of the soft set $\mathfrak{S}$.

By means of parametrization, a soft set produces a series of approximate descriptions of a complicated object being perceived from various points of view. It is apparent that a soft set $\mathfrak{S}=(F,A)$ over a universe U can be viewed as a parameterized family of subsets of U. For any parameter $\epsilon \in A$, the subset $F(\epsilon) \subseteq U$ may be interpreted as the set of *ϵ-approximate elements* [23]. Note that $F(\epsilon)$ may be arbitrary: some of them may be empty, and some may have nonempty intersections [23]. Soft set theory enables us to interpret a complicate uncertain concept using abundant parameters. The absence of any restrictions on the approximate description in soft set theory makes it easily applicable in practice [23]. We can use any suitable parametrization–with the help of words and sentences, real numbers, functions, mappings, etc.

In what follows, the collection of all soft sets over U with parameter sets contained in E is denoted by $\mathcal{S}^E(U)$. Moreover, we denote by $\mathcal{S}_A(U)$ the collection of all soft sets over U with a fixed parameter set $A \subseteq E$. Maji et al. [20] introduced the concept of soft M-subsets and soft M-equal relations in the following manner:

Definition 2.2 [20]. Let (F,A) and (G,B) be two soft sets over U. Then (F,A) is called a *soft M-subset* of (G,B), denoted $(F,A) \tilde{\subseteq}_M (G,B)$, if $A \subseteq B$ and $F(a)=G(a)$ (i.e., $F(a)$ and $G(a)$ are identical approximations) for all $a \in A$. Two soft sets (F,A) and (G,B) are said to be *soft M-equal*, denoted $(F,A)=_M(G,B)$, if $(F,A) \tilde{\subseteq}_M (G,B)$ and $(G,B) \tilde{\subseteq}_M (F,A)$.

Another different type of soft subsets and soft equal relations can be defined as follows.

Definition 2.3 [16]. Let (F,A) and (G,B) be two soft sets over U. Then (F,A) is called a *soft F-subset* of (G,B), denoted $(F,A) \tilde{\subseteq}_F (G,B)$, if $A \subseteq B$ and $F(a) \subseteq G(a)$ for all $a \in A$. Two soft sets (F,A) and (G,B) are said to be *soft F-equal*, denoted $(F,A)=_F(G,B)$, if $(F,A) \tilde{\subseteq}_F (G,B)$ and $(G,B) \tilde{\subseteq}_F (F,A)$.

It is easy to see that for two soft sets $\mathfrak{S}=(F,A)$ and $\mathfrak{T}=(G,B)$, if $\mathfrak{S}$ is a soft M-subset of $\mathfrak{T}$ then $\mathfrak{S}$ is also a soft F-subset of $\mathfrak{T}$. However, the converse may not be true [11]. As shown in [11], the soft equal relations $=_M$ and $=_F$ coincide with each other. Two soft sets over U satisfy such soft equal relations are identical since they have the same parameter sets and the same approximate functions. Hence in what follows, we just call them soft equal relations and simply write $=$ instead of $=_M$ or $=_F$ unless stated otherwise.

Definition 2.4 [6]. Let $\mathfrak{S}=(F,A)$ be a soft set over U. Then