SAR:
An Algorithm for Selecting a Partition Attribute in Categorical-Valued Information System Using Soft Set Theory

Rabiei Mamat, Universiti Malaysia Terengganu, Malaysia
Tutut Herawan, Universiti Malaysia Pahang and Data Engineering Research Centre, Malaysia
Mustafa Mat Deris, Universiti Tun Hussein Onn Malaysia, Malaysia

ABSTRACT

Soft-set theory proposed by Molodstov is a general mathematic tool for dealing with uncertainty. Recently, several algorithms have been proposed for decision making using soft-set theory. However, these algorithms still concern on Boolean-valued information system. In this paper, Support Attribute Representative (SAR), a soft-set based technique for decision making in categorical-valued information system is proposed. The proposed technique has been tested on three datasets to select the best partitioning attribute. Furthermore, two UCI benchmark datasets are used to elaborate the performance of the proposed technique in term of executing time. On these two datasets, it is shown that SAR outperforms three rough set-based techniques TR, MMR, and MDA up to 95% and 50%, respectively. The results of this research will provide useful information for decision makers to handle categorical datasets.

Keywords: Data Mining, Decision Making, Partition Attribute Selection, Soft-Set Theory, Support Attribute Representative (SAR)

INTRODUCTION

In today’s fast moving world, decision making is a very critical issue in any organization. It is known that good decision making is supported by good information. Unfortunately, some information is uncertain. Therefore, handling uncertain data is very important since there are lots of real life problems in reality still involving uncertain data, for example in field of engineering, medical, social sciences, etc. (Maji, Roy, & Biswas, 2002). Theoretically, probabilities theories, theory of fuzzy set as well as theory of rough set can be considered as the mathematical tools for dealing with uncertainties.

In 1999, Molodstov proposed a theory of soft-set as a new way for managing uncertain data. Molodstov pointed out that one of the main advantages of using soft-set theory because it is free from the inadequacy of the parameterization tools, unlike in the theories mentioned above.
The soft-set theory uses parameterization sets, as its main solution for problem solving, which makes it very convenient and easy to apply in practice. Therefore, many applications based on soft-set theory have already been demonstrated by Molodstov, such as the smoothness of functions, game theory, decision research, Rieman integration, Perron integration, probability theory, and measurement theory (Molodtsov, 1999).

Presently, great progress of study on soft-set theory has been achieved. Maji et al. (2003) firstly introduced some definitions of the related operation on soft-set. Ali et al. (2009) took into account some errors of formers and put forward some new operations on soft-set. The great progress has been achieved when adopting soft-set theory in applications. Maji et al. (2002) employed soft-set theory to solve the decision-making problem. Maji et al. (2001) presented a novel method of object recognition from an imprecise multi-observer data to deal with decision making based on fuzzy soft-set, which was revised by Kong et al. (2009). Feng et al. (2010) showed an adjustable approach to fuzzy soft-set based decision making by means of level soft-set. It is worthwhile to mention that some effort has been done to such issues concerning reduction of soft-set. Chen et al. (2005) pointed out that the conclusion of soft-set reduction offered in Maji et al. (2002) was incorrect, and then present a new notion of parameterization reduction in soft-set in comparison with the definition to the related concept of attributes reduction in rough set theory. The concept of normal parameter reduction is introduced in Kong et al. (2008), which overcome the problem of suboptimal choice and added parameter set of soft-set. An algorithm for normal parameter reduction is also presented in Kong et al. (2008) to overcome the problem of sub-optimal decision making in Chen et al. (2005). However, these soft-set-based algorithms still concern on a Boolean-valued information system.

In this paper, Support Attribute Representative (SAR), a soft-set based technique for selecting the best partitioning attribute in categorical-valued information system is proposed. The proposed technique has been tested on three datasets. Furthermore two UCI benchmark datasets are used to elaborate the performance of the proposed technique in term of executing time. It is shown that SAR outperforms three rough set-based techniques TR, MMR, and MDA. The results of this research will provide useful information for decision makers to handle categorical datasets.

The rest of the paper is organized as follows. First, we described the soft-set theory. Then, the proposed technique is described. An experiment and analysis is described afterwards. Finally, the conclusion of this work is described.

**SOFT-SET THEORY**

**Information System**

An information system is defined as a 4-tuple (quadruple) $S = (U, A, V, f)$, where $U = \{u_1, u_2, u_3, \ldots, u_n\}$ is a non-empty finite set of objects, $A = \{a_1, a_2, a_3, \ldots, a_m\}$ is a non-empty finite set of attributes, $V = \bigcup_{a \in A} V_a$, $V_a$ is the domain (value set) of attribute $a$, $f : U \times A \rightarrow V$ is an information function such that $f(u, a) \in V_a$, for every $(u, a) \in U \times A$, called information (knowledge) function. An information system can be intuitively expressed in terms of an information table (Table 1).

In many applications, there is an outcome of classification that is known. This posteriori knowledge is expressed by one (or more) distinguished attribute called decision attribute; the process is known as supervised learning. An information system of this kind is called a decision system. A **decision system** is an information system of the form $D = (U, A \bigcup \{d\}, V, f)$, where $d \notin A$ is the decision attribute.

Let $S = (U, A, V, f)$ be an information system. If $V_{\{0,1\}}$, for every $a \in A$, then
A Study on Models and Methods of Information Retrieval System
www.igi-global.com/chapter/a-study-on-models-and-methods-of-information-retrieval-system/198553?camid=4v1a