Available online at www.sciencedirect.com

ScienceDirect

Applied Mathematics Letters 21 (2008) 400–403

Applied Mathematics Letters

www.elsevier.com/locate/aml

Integrating rough set theory and medical applications

Puntip Pattaraintakorn[a,*], Nick Cercone[b]

[a] *Department of Mathematics and Computer Science, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, Thailand*

[b] *Faculty of Science and Engineering, York University, Ontario, Canada M3J 1P3*

Received 5 March 2007; accepted 30 May 2007

Abstract

Medical science is not an exact science in which processes can be easily analyzed and modeled. Rough set theory has proven well suited for accommodating such inexactness of the medical profession. As rough set theory matures and its theoretical perspective is extended, the theory has been also followed by development of innovative rough sets systems as a result of this maturation. Unique concerns in medical sciences as well as the need of integrated rough sets systems are discussed. We present a short survey of ongoing research and a case study on integrating rough set theory and medical application. Issues in the current state of rough sets in advancing medical technology and some of its challenges are also highlighted.

Keywords: Rough set theory; Medical applications; Survival analysis

1. Introduction

Pawlak [1] introduced mathematical rough set theory in the early 1980's. The theory was based on the discernibility of objects. Rough set theory provides systems designers with the ability to handle uncertainty. If a concept is 'not definable' in a given knowledge base, rough sets can 'approximate' with respect to that knowledge. From a medical point of view, the attribute-value boundaries are usually vague. In actual situations, physicians diagnose a patient and decide what is the best way to cure them. To apply rough sets to medical data and imitate this ability, many issues in rough set theory are raised [2]. For example, discretization is necessary, whether uncertainty is subjective or objective, and medical attribute values lead to difficult situations for rough set-based medical applications. These issues are also discussed by [3]. They pointed out that rough sets offer algorithms with polynomial time complexity and space complexity with respect to the number of attributes and examples. They also note that the advantages of the rough sets methodology consist of: (i) the basic tools are lower and upper approximations of the concept (which are well-defined sets) and (ii) rough sets methodology is computed directly from input data.

* Corresponding author. Tel.: +011 662 326 4341; fax: +011 662 326 4354.
E-mail address: kppuntip@kmitl.ac.th (P. Pattaraintakorn).

0893-9659/$ - see front matter
doi:10.1016/j.aml.2007.05.010

2. Theoretical aspects of rough sets

We describe the fundamental theory of rough sets from [1,4]. Given a finite set $U \neq \emptyset$ (*universe*) of objects, any subset $X \subseteq U$ of the universe is called a *concept* in U and any family of concepts in U is referred to as *knowledge*. A family of classifications over U is called a *knowledge base* over U. This formal foundation of rough set theory reveals that we consider the "universe" to be a finite set. Keeping this stability set in mind, all rough set theory in medical database or data warehousing applications is concerned with the meaningfulness of updating sets, for example, the insert, delete and join operations in database systems. Rough set methodology endeavors to discover the variety of data sources while requiring integration of other approaches to handle extensibility of data sets. Let $R \subseteq X \times X$ be an equivalence relation over U. Then R is reflexive (xRx), symmetric (if xRy then yRx) and transitive (if xRy and yRz then xRz). Define U/R as the family of equivalence classes of R and let $[x]_R$ denote a category in R containing an element $x \in U$. Given a knowledge base $K = (U, \mathbf{R})$ if $\mathbf{P} \subseteq \mathbf{R}$ and $\mathbf{P} \neq \emptyset$, then there is an equivalence relation IND($\mathbf{P}$) called the *indiscernibility relation* over $\mathbf{P}$.

The current trend in rough set theory explores the complementary mathematical properties with other mathematics disciplines. In [5], the author studied the ordered set of rough set theory and proved that the relations are not necessarily reflexive, symmetric or transitive. Next, as defined, with $X \subseteq U$ and $R \in \mathrm{IND}(K)$,

$$x = \underline{R}X \quad \text{if and only if } [x]_R \subseteq X \tag{1}$$

$$x = \overline{R}X \quad \text{if and only if } [x]_R \cap X \neq \emptyset \tag{2}$$

called the *R-lower approximation* and *R-upper approximation* of X respectively. Also let $\mathrm{POS}_R(X) = \underline{R}X$ denote the R-positive region of X, $NEG_R(X) = U - \overline{R}X$ denote the R-negative region of X and $\mathrm{BN}_R(X) = \overline{R}X - \underline{R}X$ denote the R-borderline region of X.

The degree of completeness can also be characterized by the *accuracy measure*, in which *card R* represents the cardinality of set R as follows:

$$\alpha_R(X) = \frac{\mathrm{card}\underline{R}}{\mathrm{card}\overline{R}} \quad \text{where } X \neq \emptyset. \tag{3}$$

Accuracy measures try to express the degree of completeness of knowledge. Eq. (3) is able to capture how large the boundary region of the data sets is; however, we cannot easily capture the structure of the knowledge. A fundamental advantage of rough set theory is the ability to handle a category that cannot be sharply defined given a knowledge base. Characteristics of the potential data sets can be measured through the rough sets framework. We can measure inexactness and express topological characterization of imprecision with:

(1) If $\underline{R}X \neq \emptyset$ and $\overline{R}X \neq U$, then X is *roughly R-definable*.
(2) If $\underline{R}X = \emptyset$ and $\overline{R}X \neq U$, then X is *internally R-undefinable*.
(3) If $\underline{R}X \neq \emptyset$ and $\overline{R}X = U$, then X is *externally R-undefinable*.
(4) If $\underline{R}X = \emptyset$ and $\overline{R}X = U$, then X is *totally R-undefinable*.

With Eq. (3) and classifications above we can characterize rough sets by the size of the boundary region and structure. Rough sets are treated as a special case of relative sets and integrated with the notion of Belnap's logic [6].

3. Medical science

Traditional medical data analysis tends to employ analysts who are familiar with particular data and use statistical techniques to provide reports. This approach is no longer viable. We extend some uniqueness properties in medical data in [7]. The salient points for rough sets are:

Sensitivity and specificity analysis: Most diagnoses and treatments in medical science are imprecise and accompanied by rates of error. The authors reveal the meaninglessness of sensitivity, specificity and accuracy measures used to evaluate data mining applications. Use of rough sets is able to circumvent this limitation with its ability to handle imprecise and uncertain data.

Poor physical formulae or equations for characterizing medical data: Other physical sciences mainly observe and collect data that can be fit into formulae reasonably and solved for the characteristics or relationship of that data.