Chapter 8
Application of Rough Set Based Models in Medical Diagnosis

Balakrushna Tripathy
VIT University, India

ABSTRACT

Modeling intelligent system in the field of medical diagnosis is still a challenging work. Intelligent systems in medical diagnosis can be utilized as a supporting tool to the medical practitioner, mainly country like India with vast rural areas and absolute shortage of physicians. Intelligent systems in the field of medical diagnosis can also able to reduce cost and problems for the diagnosis like dynamic perturbations, shortage of physicians, etc. An intelligent system may be considered as an information system that provides answer to queries relating to the information stored in the Knowledge Base (KB), which is a repository of human knowledge. Rough set theory is an efficient model to capture uncertainty in data and the processing of data using rough set techniques is easy and convincing. Rule generation is an inherent component in rough set analysis. So, medical systems which have uncertainty inherent can be handled in a better way using rough sets and its variants. The objective of this chapter is to discuss on several such applications of rough set theory in medical diagnosis.

INTRODUCTION

Medical diagnosis is still considered an art, despite all standardization efforts. Medical diagnosis is the art of determining a patient’s pathological status from an available set of symptoms (findings). It is defined an art, because it is a complicated problem with many and manifold factors, and its solution comprises literally all of a human’s abilities including intuition and the subconscious. The process of medical diagnosis is composed of, evaluation of a given set of symptoms (findings) performing relevant pathological tests (patient’s test data), and ultimately identifying the diseases validating the particular findings. The functioning of the human body is characterized by the complex and extremely interactive chemistry of its organs and the psyche. This concerted effort results homeostasis and the equilibrium of all physiological quantities. This balance is maintained in a level within physiological bounds that varies from individual to individual. Due to internal or of external cause, deviations from it are indica-
tive of some kind of perturbation. The identification of the cause of these perturbations is the goal of medical diagnosis. Reaching a foolproof diagnosis is never an easy job for medical practitioners. Today in medical diagnosis it is often impossible to look inside a patient to determine the primary cause that led to the series of effects and reactions the patient complains about. Thus the diagnosis is based on indirect evidence, symptoms and the knowledge of the medical mechanisms that relate presumed causes to observed effects. The problems of diagnosis not only arise due to the incompleteness of knowledge, but also most immediate limitations of the theoretical and practical knowledge implications that lead from an initial cause to its observable effects.

**DEFINITIONS AND NOTATIONS**

In this section we provide some definitions and notations to be followed in this chapter. Imprecision in data has become a common feature and hence to handle them imprecise models have been developed in literature. In this sequel we have fuzzy sets, rough sets, intuitionistic fuzzy sets and soft sets along with their generalisations and modifications can be found. Our basic aim in this chapter is to deal with only rough sets. So, we start with the definition of rough sets and related concepts.

Let $U$ be a universe of discourse and $R$ be an equivalence relation over $U$. By $U/R$ we denote the family of all equivalence classes of $R$, referred to as categories or concepts of $R$ and the equivalence class of an element $x \in U$ is denoted by $[x]_R$. By a knowledge base, we understand a relational system $K = (U, P)$, where $U$ is as above and $P$ is a family of equivalence relations over $U$. For any subset $Q \neq \emptyset \subseteq P$, the intersection of all equivalence relations in $Q$ is denoted by $\text{IND}(Q)$ and is called the indiscernibility relation over $Q$.

**Definition 1:** Given any $X \subseteq U$ and $R \in \text{IND}(K)$, we associate two subsets $RX$ and $\overline{RX}$ called the $R$-lower and $R$-upper approximations of $X$ respectively and are defined as:

1. $\overline{RX} = \bigcup \{Y \subseteq U : Y \subseteq X \} \quad \text{and}$
2. $RX = \bigcup \{Y \subseteq U : Y \cap X \neq \emptyset \}$.

The $R$-boundary of $X$ is denoted by $BN_R(X)$ and is given by $BN_R(X) = \overline{RX} - RX$.

The elements of $RX$ are those elements of $U$, which can certainly be classified as elements of $X$, and the elements of $\overline{RX}$ are those elements of $U$, which can possibly be classified as elements of $X$, employing knowledge of $R$. We say that $X$ is rough with respect to $R$ if and only if $RX \neq \overline{RX}$, equivalently $BN_R(X) \neq \emptyset$. $X$ is said to be $R$-definable if and only if $RX = \overline{RX}$, or $BN_R(X) = \emptyset$.

**Definition 2:** An information system $I$ is a four tuple $<U, A, V, F>$, where $U = \{x_1, x_2, \ldots, x_n\}$ is a finite, non-empty set of objects, $A = \{A_1, A_2, \ldots, A_m\}$ is a set of attributes. $V = \bigcup_{i=1}^{m} V_{A_i}$, where $V_{A_i}$ is the domain of the attribute $A_i$, $i = 1, 2 \ldots m$. $F : U \times A \rightarrow V$ is the total decision function called the information function such that $F(x, A_i) \in V_{A_i}, x \in U$ and $A_i \in A_i$.