Appendix A

Rough Set Theory

Rough sets theory is a new intelligent mathematical tool proposed by Pawlak [86-87] for use in reasoning from imprecise data. It is based on the concept of an upper and a lower approximation of a set, the approximation space and models of sets. Rough set theory is an extension of set theory for study of the intelligent systems characterized by insufficient and incomplete information. An undefinable subset is approximately represented by two definable subsets, called lower and upper approximations. Rough set theory is a good candidate for classification applications. Various efforts have been made to improve the efficiency and effectiveness of classification with rough sets. Here, we give a brief introduction to the rough sets as need arises for the rough sets based analysis in this work [86-87].

Rough set Concepts

Information system and decision system: an information system (IS) is an ordered pair \((U, A)\), where \(U=\{x_1, x_2, \ldots, x_n\}\) is a nonempty finite set of objects called the universe and \(A=\{a_1, a_2, \ldots, a_n\}\) is a nonempty set and the elements of \(A\), are called attributes. Decision System is an information system (IS) for which the attributes in \(A\) are further classified into disjoint sets of condition attributes \(C\) and decision attributes \(D\).

Discernibility matrix and indiscernibility relation: the discernibility matrix of \(A\) is the \(n \times n\) matrix with \((i,j)\)th entry is defined as follows:

\[
DM_{ij} = \{a \in A : a(x_i) \neq a(x_j)\}. \tag{A.1}
\]

Assume a subset of the set of attributes, \(B \subseteq A\), then two samples \(x\) and \(y\) in \(U\) are indiscernible with respect to \(B\) if and only if \(f(x, q) = f(y, q)\) \(\forall q \in B\). The indiscernibility relation for all \(B \subseteq A\) is defined as follows:
Ind(B) = \{(x,y) \in U \times U : (\forall a \in B) a(x) = a(y)\} \tag{A.2}

Let $U/\text{Ind}(B)$ be the set of all equivalence classes in the relation $\text{Ind}(B)$.

Owing to the imprecision existing in the real world data, there are always conflicting objects contained in a decision table. Here, conflicting objects refer to the two or more objects that are indiscernible by employing any set of condition attributes, but they belong to different decision classes. Such objects are called inconsistent; therefore, the decision table is called inconsistent decision table.

Lower and upper approximation: In the rough set theory, the approximations of sets are introduced to deal with inconsistency. A rough set approximates the traditional sets using a pair of sets named the lower and upper approximation of the set. The lower and upper approximations of a set $P \subseteq U$ are defined by Eqns. (A.3) and (A.4), respectively:

$$PY = \bigcup\{X : X \in U/\text{Ind}(P), X \subseteq Y\} \quad \tag{A.3}$$

$$PY = \bigcup\{X : X \in U/\text{Ind}(P), X \cup Y \neq \{\}\} \quad \tag{A.4}$$

Reduct: given a classification task mapping a set of variables $C$ to a set of labeling $D$, a reduct is defined as any $R \subseteq C$ such that $\gamma(C,D) = \gamma(R,D)$. The set of all reducts of $A$ is denoted by $\text{Red}(A)$.

Reduct set: given a classification task mapping a set of variables $C$ to a set of labeling $D$, a reduct set is defined with respect to the power set $P(C)$ as the set $\text{Red} = \{A \subseteq P(C) : \gamma(A,D) = \gamma(C,D)\}$. That is, the reduct set is the set of all possible reducts of the equivalence relation denoted by $C$ and $D$.

Minimal reduct: given a classification task mapping a set of variables $C$ to a set of labeling $D$, and $R$, is the reduct set for this problem space. A minimal reduct $R$ is the reduct such that $\|R\| \leq \|A\| \forall A \in \text{Red}$, $\forall A \in \text{Red}$. That is, the minimal reduct is the reduct of least cardinality for the equivalence relation denoted by $C$ and $D$.

Positive region: let $P$ and $Q$ be two equivalence relations in $U$, the important concept of positive region $\text{POS}_P(Q)$ is defined as:
\[ POS_P(Q) = \bigcup_{x \in Q} PX \]  

(A.5)

A positive region contains all patterns in \( U \) that can be classified in attribute set \( Q \) using the attribute set \( P \).

The degree of dependency: the degree of dependency \( \gamma(P, Q) \) of a set \( P \) of attributes with respect to a set \( Q \) of class labeling is defined as

\[ \gamma(P, Q) = \frac{|POS_P(Q)|}{|U|} \]  

(A.6)

where \(|S|\) denotes the cardinality of a set \( S \).

The degree of dependency provides a measure of how important \( P \) is in mapping the data sets examples into \( Q \). If \( \gamma(P, Q)=0 \), then classification \( Q \) is independent of the attributes in \( P \), hence the decision attributes are of no use to this classification. If \( \gamma(P, Q)=1 \), then \( Q \) is completely dependent on \( P \), hence the attributes are indispensable. Values \( 0<\gamma(P, Q)<1 \) denote partial dependency, which shows that only some of the attributes in \( P \) may be useful, or that the data set was flawed to begin with. In addition, the complement of \( \gamma(P, Q) \) gives a measure of the contradictions in the selected subset of the dataset.

It is now possible to define the significance of an attribute. This is done by calculating the change of dependency when removing the attribute from the set of considered conditional attributes.

Significance: given \( P, Q \) and an object \( x \in P \), the significant \( \sigma_x(P, Q) \) of \( x \) in the equivalence relation denoted by \( P \) and \( Q \) is \( \sigma_x(P, Q)=\gamma(P, Q)-\gamma(P-\{x\}, Q) \).

Now, attribute reduction involves removing attributes that have no significance to the classification at hand. It is obvious that a dataset may have more than one attribute reduct set.

Core: an attribute \( C_j \in C \) is a core attribute in \( C \) with respect to \( D \) and it is the intersection of all reducts. It is defined as follows:
\[ \text{Core}(C) = \bigcap_{R \in \text{Red}(B)} \text{Red}_i, \quad i = 1, 2, \ldots \] (A.7)

Suppose we have a decision table \( A = (U, A \cup \{d\}) \) where \( \text{card} \{V_a\} \) is high for some \( a \in A \). Then there is a very low chance that a new object is recognized by rules generated directly from this table, because the attribute value vector of a new object will not match any of these rules. Therefore for decision tables with real value attributes some discretization strategies are built to obtain a higher quality of classification rules. Some of the discretization methods used is Boolean reasoning, naïve’s algorithm, entropy algorithm, equal frequency binning etc.

There are two strategies of searching for semi-optimal set of cuts. The first, called local strategy after finding the best cut and dividing the object set into two subsets of objects, repeats this procedure for each object set separately until some stop condition holds. The quality of a cut, i.e. the number of objects discerned by a cut in the local strategy is computed locally on a subset of objects. In the second strategy, called global strategy, the quality of cuts is computed on whole set of objects. Usually, the local strategy is easier for realization than the global one. But the set of cuts obtained by the global strategy is smaller.