COURSE: CS 836

NAME: Chao Zhang

STUDENT#: 200383834

1. Discuss the concept of three-way decision in rough set theory (i.e., three types of rules).

The concept of three-way decisions was proposed and used to interpret rough set three regions. More specifically, the positive, negative and boundary regions are viewed, respectively, as the regions of acceptance, rejection, and non-commitment in a ternary classification. The positive and negative regions can be used to induce rules of acceptance and rejection; whenever it is impossible to make an acceptance or a rejection decision, the third non-commitment decision is made. It can be shown that, under certain conditions, probabilistic three-way decisions are superior to both Palwak three-way decisions and two-way decisions.

2. Describe the Pawlak three-step data analysis.

An information table provides all available information about a set of objects.We analyze attributes and objects based on the information functions in the table. Pawlak investigated three main tasks of rough set analysis and presented them in a sequential three steps. The first step analyzes attribute dependencies with an objective to simplify a table. The main tasks involving identifying superfluous attributes and finding a minimal subset of attributes that preserves the same information as the entire set of attributes for the purpose of classification. Such a minimal set of attributes is called an attribute reduct of the table or a relative attribute reduct of a classification table. There may exist more than one reduct for each table. With respect to a reduced table with a minimal set of attributes in a decision table, we can construct a set of decision rules. The left-hand-side of each decision rule is a conjunction of a set of attribute-value pairs. The second step analyzes dependencies of attribute values with an objective to simplify a decision rule. Similar to the notion of superfluous attribute in a table, there may exist superfluous attribute-value pairs in the left-hand-side of a decision rule. The main tasks of the second step are to identify superfluous attribute-value pairs and to derive a minimal set of attribute-value pairs for each decision rule. A minimal set of attribute-value pairs is called a relative attribute value-pair reduct. Again, there may exist more than one reduct. The result of the second step is a set of minimal decision rules. The third step analyzes dependencies of decision rules with an objective to simplifying a set of decision rules. There may exist superfluous rules in the set of decision rules obtained in the second step. By removing superfluous rules, one can obtain a minimal set of rules called a rule reduct.

3. Describe rough set approximations in an incomplete information table based on possible world semantics.

By using the family *CDEFI(OB)* instead of the family *CDEF(OB)* in an incomplete table, the meaning set of a formula becomes a conjunctively definable interval set. The family *CDEFI(OB)* is consequently used to define the structured positive and negative regions. We consider the component-wise inclusion relationships between a conjunctively definable interval set and *X* and *Xc*, that is, the set-theoretic inclusion between a set in the interval set and *X* and *Xc*. This leads to two types of structured positive and negative regions.

For a set of objects X in an incomplete table $\widetilde{T}$, we define two types of structured positive and negative regions of X as follows:

(1) $SPOS_*(X)=\{(p,\tilde{m}(p))\in CDEFI(OB) \mid \tilde{m}(p)\neq[\varnothing,\varnothing], \forall S\in\tilde{m}(p), S\subseteq X\}$,

$SPOS_*(X)=\{(p,\tilde{m}(p))\in CDEFI(OB) \mid \tilde{m}(p)\neq[\varnothing,\varnothing], \forall S\in\tilde{m}(p), S\subseteq X^c\}$;

(2) $SPOS^*(X)=\{(p,\tilde{m}(p))\in CDEFI(OB) \mid \exists S\in\tilde{m}(p), S\neq\varnothing, S\subseteq X\}$,

$SPOS^*(X)=\{(p,\tilde{m}(p))\in CDEFI(OB) \mid \exists S\in\tilde{m}(p), S\neq\varnothing, S\subseteq X^c\}$.

For a set of objects X in an incomplete table $\widetilde{T}$, its two types of structured positive and negative regions can be equivalently expressed as:

(1) $SPOS_*(X)=\{(p,\tilde{m}(p))\in CDEFI(OB) \mid \tilde{m}(p)\neq[\varnothing,\varnothing], \forall T\in COMP(\widetilde{T}),$

$m(p|T)\subseteq X\}$,

$$SPOS_*(X)=\{(p,\tilde{m}(p))\in CDEFI(OB)|\tilde{m}(p)\neq[\varnothing,\varnothing],\forall T\in COMP(\tilde{T}),$$

$$m(p|T)\subseteq X^c\};$$

(2) $$SPOS^*(X)=\{(p,\tilde{m}(p))\in CDEFI(OB)|\exists T\in COMP(\tilde{T}),\ m(p|T)\neq\varnothing,$$

$$m(p|T)\subseteq X\},$$

$$SPOS^*(X)=\{(p,\tilde{m}(p))\in CDEFI(OB)|\exists T\in COMP(\tilde{T}),\ m(p|T)\neq\varnothing,$$

$$m(p|T)\subseteq X^c\}.$$

For a set of objects X in an incomplete table $\tilde{T}$, its two types of structured positive and negative regions can be computed as:

(1) $$SPOS_*(X)=\{(p,[m_*(p),m^*(p)])\in CDEFI(OB)|m^*(p)\neq\varnothing,m^*(p)\subseteq X\},$$

$$SPOS_*(X)=\{(p,[m_*(p),m^*(p)])\in CDEFI(OB)|m^*(p)\neq\varnothing,m^*(p)\subseteq X^c\};$$

(2) $$SPOS^*(X)=\{(p,[m_*(p),m^*(p)])\in CDEFI(OB)|(m_*(p)\neq\varnothing\wedge m_*(p)\subseteq X)$$

$$\vee(m_*(p)=\varnothing\wedge m^*(p)\cap X\neq\varnothing)\},$$

$$SPOS^*(X)=\{(p,[m_*(p),m^*(p)])\in CDEFI(OB)|(m_*(p)\neq\varnothing\wedge m_*(p)\subseteq X)$$

$$\vee(m_*(p)=\varnothing\wedge m^*(p)\cap X^c\neq\varnothing)\},$$

4. Describe the main ideas of three-way decisions.

The main ideas of three-way decisions are described in terms of a ternary classification according to evaluations of a set of criteria. Suppose *U* is a finite nonempty set of objects or decision alternatives and *C* is a finite set of conditions. Each condition in *C* may be a criterion, an objective, or a constraint. For simplicity, we refer to conditions in *C* as criteria. Our decision task is to classify objects of U according to whether they satisfy the set of criteria. In widely used two-way decision models, it is assumed that an object either satisfies the criteria or does not satisfy the criteria. The set U is divided into two disjoint regions, namely, the positive region POS for objects satisfying the criteria and the negative region NEG for objects not satisfying the criteria. There are usually some classification errors associated with such a binary classification. Two main difficulties with two-way approaches are their stringent binary assumption of the satisfiability of objects and the requirement of a dichotomous classification. In many situations, it may happen that an object only satisfies the set of criteria to some degree. Even if an object may actually either satisfy or not satisfy the criteria, we may not be able to identify without uncertainty the subset of objects that satisfy the criteria due to uncertain or incomplete information. Consequently, we are only able to search for an approximate solution. Instead of making a binary decision, we use thresholds on the degrees of satisfiability to make one of three decisions: (a) accept an object as satisfying the set of criteria if its degree of satisfiability is at or above a certain level; (b) reject the object by treating it as not satisfying the criteria if its degree of satisfiability is at or below another level; and (c) neither accept nor reject the object but opt for a non-commitment. The third option may also be referred to as a deferment decision that requires further information or investigation.

5. Search the Web and give five examples in which three-way decision is used.

(1) *[Triage]* This is a classical example of three-way decisions that was used to prioritize treatment of the wounded during war. The basic version of triage divides a set of the wounded into three categories:

1. Those who are likely to live, regardless of what care they receive;
2. Those who are likely to die no matter what is done for them;
3. Those for whom immediate care will make a difference.

(2) *[Text Analysis]* Some of the tasks of text analysis are to identify a set of significant words and to determine their significance values. Following the pioneer work of Luhn, one can divide a set of words into three regions based on their frequencies. Given a pair of a high threshold u and a low threshold l, we can divide words into three regions. High-frequency words ranked before u and low-frequency words ranked after l are non-significant words, and medium-frequency words ranked between u and l are significant words. Once the three regions are constructed, one can design strategies to further process words in each region.

(3) *[Management of Students]* Many schools, colleges, and universities typically monitor student progress based on a three-way classification of students. For example, according to GPA (Grade Point Average), one can divide students into the groups of high-, medium-, and low- GPA students, respectively. A different action is taken for each group of students. A high-GPA student may be awarded a scholarship and a low-GPA student may receive a warning letter. It is surprising that we seldom question the rationales and impacts of these strategies, particularly, the strategy of doing nothing for medium-GPA students.

(4) *[Traffic Lights]* In our daily life, we use traffic lights based on a three-way classification of conditions. As we know, it has red light, green light and yellow light which are three conditions that satisfying the go, stop and the medium between go and stop which alerts drivers to make a decision.

(5) *[Marking System]* In many universities, professors or TA doing marking also uses a three-way decision. For the all correct part, it is in positive region, for the all wrong part, it is in negative region, and the medium which has correct part but not all correct is in boundary region.