Course: CS 836

Name: Chao Zhang

Student#: 200383834

1. Reducts

• *Give a definition of a reduct of an information table.*

In a information table T, if a subset of attributes R $\subseteq$ AT satisfies the following conditions:

Existence: $DEF_{AT}(T)=DEF_{AT}(T)$

Sufficiency: $DEF_{R}(T)=DEF_{AT}(T)$

Minimization: $\forall a \in R\ (E_{R-\{a\}} \neq E_{AT})$

We call R an attribute reduct of AT

• *Give a definition of a reduct of a consistent decision table.*

Existence: $IND(C) \subseteq IND(D)$

Sufficiency: $IND(R) \subseteq IND(D)$

Minimization: $\forall a \in R\ (\neg(IND(R-\{a\}) \subseteq IND(D)))$

• *Give a definition of a reduct of a decision table. Show that your definition is applicable to both consistent and inconsistent decision table.*

In a decision table T with AT=C $\cup$ D, if a subset of condition attributes R$\subseteq$C satisfies the following conditions:

Existence: $POS(\pi_D|\pi_C) = POS(\pi_D|\pi_C)$

Sufficiency: $POS(\pi_D|\pi_R) = POS(\pi_D|\pi_C)$

Minimization: $\forall a \in R\ (\neg(POS(\pi_D|\pi_{R-\{a\}}) = POS(\pi_D|\pi_C)))$

We call R a relative attribute reduct of C.

For a consistent table which is $IND(C) \subseteq IND(D)$, according to the reduct R, we can get $IND(R) \subseteq IND(D)$ and $\forall a \in R\ (\neg(IND(R-\{a\}) \subseteq IND(D)))$. It is the same as $POS(\pi_D|\pi_R) = POS(\pi_D|\pi_C)$ and for every single attribute a in R is not superfluous. For an inconsistent table, the definition illustrates the partition induced by R with respect to D in positive region is equal to the partition induced by C with respect to D and for every single attribute a in R is not superfluous to give the same partition in positive region.

• *Give a general definition of a reduct to cover reducts in an information table, a consistent decision table, and a decision table.*

Generally in a table, we can consider two subsets of attributes $X, Y \subseteq AT$. For these two subsets are arbitrary and may have a non-empty overlap. Then the notion we have will be “a reduct of X with respect to Y”. Then we have following conditions:

Existence: $POS(\pi_Y|\pi_X) = POS(\pi_Y|\pi_X)$

Sufficiency: $POS(\pi_Y|\pi_X) = POS(\pi_Y|\pi_R)$

Minimization: $\forall a \in R\ (\neg(POS(\pi_Y|\pi_{R-\{a\}}) = POS(\pi_Y|\pi_C)))$

For a information table, these two subsets can be X=Y=AT.

For a decision table, these two subsets can be X is the set of condition attributes and Y is the set of decision attributes.

2. Probabilistic rough sets

• *Give a definition of probabilistic rough set approximations. Show that Pawlak rough set approximations are a special case.*

For a subset X of OB, its rough membership function is given by the conditional probability Pr(X|[x])=$\frac{|X \cap [x]|}{[x]}$. For general probabilistic approximations, we use a pair of parameters $\alpha,\beta \in [0,1]$ and $\alpha \geq \beta$ to make sure that the upper approximation is greater than the lower approximation. So the general probabilistic approximations can be defined:

$\underline{apr}(\alpha)=\{X \in OB \mid Pr(X|[x]) \geq \alpha\}$, $\overline{apr}(\beta)=\{X \in OB \mid Pr(X|[x])>\beta\}$

For Pawlak rough set approximations:

$\underline{apr}(\alpha)=\{X \in OB \mid Pr(X|[x]) \geq 1\}$, $\overline{apr}(\beta)=\{X \in OB \mid Pr(X|[x])>0\}$ where α=1 and β=0 in the general definition, so it is a special case.

• *Explain the motivations for introducing probabilistic rough sets.*

In the standard rough set model proposed by Pawlak, the lower and upper approximations are defined based on the two extreme cases according to the relationships between an equivalence class and a set. The lower approximation requires that the equivalence class is a subset of the set and the upper approximation requires the equivalence class must have a non-empty overlap with the set. A lack of consideration for the degree of their overlap unnecessarily limits the applications of rough sets and it is the main motivation for introducing probabilistic rough sets.

3. Decision-theoretic rough sets

• *Describe the Bayesian decision procedure.*

Let $\Omega=\{w_1,w_2\cdots w_s\}$ be a finite set of s states. Let $A=\{a_1,a_2\cdots a_m\}$ be a finite set of m possible actions. Let $Pr(w_j|x)$ be the conditional probability of an object x being in state w_j given that the object is described by x. Let $\lambda(a_i|w_j)$ denote the loss, or cost, for taking action a_i when the state is w_j . For an object with description x, suppose action a_i is taken. Since $Pr(w_j|x)$ is the probability that the true state is w_j given x, the expected loss associated with taking action a_i is given by: $R(a_i|x)=\sum_{j=1}^{s}\lambda(ai\mid wj)Pr(wj\mid x)$.

The quantity $R(a_i|x)$ is also called the conditional risk. Given a description x, a decision rule is a function $\tau(x)$ that specifies which action to take. That is, for every x, $\tau(x)$ takes one of the actions, $a_1,a_2\cdots a_{m.}$ The overall risk R is the expected loss associated with a given decision rule. Since $R(\tau(x)|x)$ is the conditional risk associated with action $\tau(x)$, the overall risk is defined by: $R=\sum_{x} R(\tau(x)\mid x)\Pr(x)$, where the summation is over the set of all possible descriptions of objects. If $\tau(x)$ is chosen so that $R(\tau(x)|x)$ is as small as possible for every x, the overall risk R is minimized. Thus, the Bayesian decision procedure can be formally stated as follows: for every x, compute the conditional risk $R(a_i|x)$ for $i=1,\cdots,m$ defined by $R(a_i|x)=\sum_{j=1}^{s}\lambda(ai\mid wj)Pr(wj\mid x)$ and select the action for which the conditional risk is minimum. If more than one action minimizes $R(a_i|x)$, a tie-breaking criterion can be used.

• *Derive the probabilistic rough set approximations by using the Bayesian decision procedure*

A decision-theoretic model formulates the construction of rough set approximations as a Bayesian decision problem with a set of two states and a set of three actions. The set of states is given by $\Omega=\{X, X_c\}$ indicating that an element is in X and not in X, respectively. Corresponding to the three regions, the set of actions is given by $A=\{a_P, a_B, a_N\}$, denoting the actions in classifying an object x, namely, deciding $x\in POS(X)$, deciding $x\in BND(X)$, and deciding $x\in NEG(X)$, respectively. The expected losses associated with taking different actions for objects in [x] can be expressed as:

$R(a_P|[x])=Pr(X|[x])\lambda_{PP}+Pr(X^c|[x])\lambda_{PN}$,

$R(a_N|[x])=Pr(X|[x])\lambda_{NP}+Pr(X^c|[x])\lambda_{NN}$,

$R(a_B|[x])=Pr(X|[x])\lambda_{BP}+Pr(X^c|[x])\lambda_{BN}$.

The Bayesian decision procedure leads to the following minimum-risk decision rules as follows:

(P) If $R(a_P|[x])\le R(a_B|[x])$ and $R(a_P|[x])\le R(a_N|[x])$, decide $x\in POS(X)$;

(B) If $R(a_B|[x]) \le R(a_P|[x])$ and $R(a_B|[x]) \le R(a_N|[x])$, decide $x \in BND(X)$;

(N) If $R(a_N|[x]) \le R(a_P|[x])$ and $R(a_N|[x]) \le R(a_B|[x])$, decide $x \in NEG(X)$.

Considering a special class of loss functions with $\lambda_{PP} \le \lambda_{BP} \le \lambda_{NP}$ and $\lambda_{NN} \le \lambda_{BN} \le \lambda_{PN}$, with this condition and the equation $Pr(X|[x])+Pr(X^c|[x])=1$, we can get:

(O) If $Pr(X|[x]) \ge \gamma$ and $Pr(X|[x]) \ge \alpha$, decide $x \in POS(X)$;

(B) If $\beta \le Pr(X|[x]) \le \alpha$, decide $x \in BND(X)$.

(N) If $Pr(X|[x]) \le \beta$ and $Pr(X|[x]) \le \gamma$, decide $x \in NEG(X)$;

where

$\alpha=(\lambda_{PN}-\lambda_{BN}) / ((\lambda_{PN}-\lambda_{BN})+(\lambda_{BP}-\lambda_{PP}))$

$\gamma=(\lambda_{PN}-\lambda_{NN}) / ((\lambda_{NP}-\lambda_{PP})+(\lambda_{PN}-\lambda_{NN}))$

$\beta=(\lambda_{BN}-\lambda_{NN}) / ((\lambda_{BN}-\lambda_{NN})+(\lambda_{NP}-\lambda_{BP}))$

Then we obtain the following condition on the loss function:

$((\lambda_{NP}-\lambda_{BP})/(\lambda_{BN}-\lambda_{NN})) \ge ((\lambda_{BP}-\lambda_{PP})/(\lambda_{BN}-\lambda_{NN}))$

It implies that $1 \ge \alpha \ge \gamma \ge \beta \ge 0$ and then we can get the simplified rules:

(P) If $Pr(X|[x]) \ge \alpha$, decide $x \in POS(X)$;

(N) If $Pr(X|[x]) \le \beta$, decide $x \in NEG(X)$;

(B) If $\beta < Pr(X|[x]) < \alpha$, decide $x \in BND(X)$.

So we can get the probabilistic approximations from the relationship between the three regions and approximations:

$\underline{apr}(\alpha)=\{X \in OB \mid Pr(X|[x]) \ge \alpha\}$, $\overline{apr}(\beta)=\{X \in OB \mid Pr(X|[x]) > \beta\}$.

4. Describe the main results of naive Bayesian rough sets.

Naive Bayesian rough set model is a practical method for estimating the conditional probability. First, we perform the logit transformation of the conditional probability:

$\text{logit}(\Pr(X|[x]))=\log\frac{\Pr(X|[x])}{1-\Pr(X|[x])}=\log\frac{\Pr(X|[x])}{\Pr(X^c|[x])}$, which is a monotonically increasing transformation of Pr(X|[x]). Then, we apply the Bayes' theorem: $\Pr(X|[x])=\frac{\Pr([x]|X)\Pr(X)}{\Pr([x])}$. Similarly, for X^c we also have: $\Pr(X^c|[x])=\frac{\Pr([x]|X^c)\Pr(X^c)}{\Pr([x])}$. Then using above three formulas, we can get

$$\text{logit}(\Pr(X|[x]))=\log O(X|[x])=\log\frac{\Pr(X|[x])}{1-\Pr(X|[x])}=\log\frac{\Pr(X|[x])}{\Pr(X^c|[x])}=$$

$$\log\frac{\Pr(X|[x])}{\Pr(X^c|[x])}=\log\frac{\Pr([x]|X)}{\Pr([x]|X^c)}\cdot\frac{\Pr(X)}{\Pr(X^c)}=\log\frac{\Pr([x]|X)}{\Pr([x]|X^c)}+\log O(X).$$

where O(X|[x]) and O(X) are the a posterior and the a prior odds, and $\frac{\Pr([x]|X)}{\Pr([x]|X^c)}$ is the likelihood ratio. A threshold value on the probability can be expressed as another threshold value on logarithm of the likelihood ratio. For the positive region, we have:

$$\Pr(X|[x])\geq\alpha\Leftrightarrow\log\frac{\Pr(X|[x])}{\Pr(X^c|[x])}\geq\log\frac{\alpha}{1-\alpha}\Leftrightarrow\log\frac{\Pr([x]|X)}{\Pr([x]|X^c)}\cdot\frac{\Pr(X)}{\Pr(X^c)}\geq\log\frac{\alpha}{1-\alpha}$$

$$\Leftrightarrow\log\frac{\Pr([x]|X)}{\Pr([x]|X^c)}\geq\log\frac{\alpha}{1-\alpha}+\frac{\Pr(X^c)}{\Pr(X)}=\alpha'$$

Similar expressions can be obtained for the negative and boundary regions. The three regions can now be written as:

$$\mathrm{POS}_{(\alpha,\beta)}(X)=\{x\in OB|\log\frac{\Pr([x]|X)}{\Pr([x]|X^c)}\geq\alpha'\},$$

$$\mathrm{NEG}_{(\alpha,\beta)}(X)=\{x\in OB|\log\frac{\Pr([x]|X)}{\Pr([x]|X^c)}\leq\beta'\},$$

$$\mathrm{BND}_{(\alpha,\beta)}(X)=\{x\in OB|\beta'<\log\frac{\Pr([x]|X)}{\Pr([x]|X^c)}<\alpha'\}.$$

where

$$\alpha'=\log\frac{\Pr(X^c)}{\Pr(X)}+\log\frac{\alpha}{1-\alpha}$$

$$\beta'=\log\frac{\Pr(X^c)}{\Pr(X)}+\log\frac{\beta}{1-\beta}$$