From Deterministic to Probabilistic Rough Sets

Wojciech Ziarko

University of Regina
Computer Science
Saskatchewan, Canada

ziarko@sasktel.net

May 14, 2024

Overview

Pawlak Rough Sets: Approximation Space

- Introduced by late Professor Zdzislaw Pawlak
- Collection of objects or observations of interest: universe U, assumed finite, but can be generalized to infinite
- Classification knowledge, an ability to define categories (not necessarily data-based), represented by an equivalence *indiscernibility* relation: $R \subseteq U \times U$
- Elementary sets collection, assumed finite: $R^* = \{E_1, E_2, ..., E_n\}$
- Approximation space: (U, R)

Pawlak Rough Sets: Undefinable Sets

- It is not possible, in general, to form precise discriminating definition, in terms of the available classification knowledge, of an arbitrary set $X \subseteq U$
- Some sets (sometimes called concepts) can never be defined, or learned, with a given classification knowledge
- At the best, only approximate definitions can be created, or learned
- Sets for which discriminating definitions do not exist are called *undefinable*, or *rough*

Pawlak Rough Sets: Rough Approximations of Sets

- Lower approximation:

$$\underline{R}(X) = \cup\{E \in R^* : E \subseteq X\}$$

- Upper approximation:

$$\overline{R} = \cup\{E \in R^* : E \cap X \neq \emptyset\}$$

Disjoint approximation regions:

- Positive region:

$$POS(X) = \cup\{E \in R^* : E \subseteq X\}$$

- Negative region:

$$NEG(X) = \cup\{E \in R^* : E \cap X = \emptyset\}$$

- Boundary region:

$$BND(X) = \cup\{E \in R^* : E \cap X \neq \emptyset \wedge E \not\subseteq X\}$$

Pawlak Rough Sets: Rough Approximations of Sets

- If $BND(X) = \emptyset$ then X is definable
- Elementary sets and approximation are definable and disjoint
- A rough set X is approximately defined by specifying definitions of its approximation regions $POS(X)$, $NEG(X)$ and $BND(X)$
- This can be done in a tabular form by creating a *rough decision table*
- The approximation space is defined via classification of objects, collected in a *classification table*, based on identity of values of their *attributes*

Classification Table: Representation of Instances

Obj	HRate	BP	Temp	Treatm	**Response**
e_1	High	High	High	1	**Positive**
e_2	High	High	High	1	**Negative**
e_3	High	High	Normal	2	**Negative**
e_4	High	Normal	High	1	**Positive**
e_5	High	Normal	High	1	**Positive**
e_6	Normal	Normal	Normal	3	**Positive**
e_7	Normal	Normal	Normal	3	**Negative**
e_8	Low	High	High	1	**Negative**
e_9	Low	High	High	1	**Negative**
e_{10}	Low	High	Normal	2	**Positive**
e_{11}	Low	High	Normal	2	**Negative**
e_{12}	High	Normal	High	2	**Negative**
e_{13}	Normal	Normal	Normal	1	**Positive**

Table: Classification Table of medical records

Rough Decision Table: Representation of Classes

HRate	*BP*	*Temp*	*Treatment*	*Appr Region*
High	High	High	1	**BND**
High	High	Normal	2	**NEG**
High	Normal	High	1	**POS**
Normal	Normal	Normal	3	**BND**
Low	High	High	1	**NEG**
Low	High	Normal	2	**BND**
High	Normal	High	2	**NEG**
Normal	Normal	Normal	1	**POS**

Table: Rough Decision Table representation of the rough set: **Response=Positive**

Acquisition of Rough Decision Tables

- From data, based on analysis and pre-processing of existing data: data mining, most common
- Based on prior human expert knowledge: expert specifies the classes of rough decision table
- Through learning from individual observations using pre-selected training data: objects, cases, instances
- The decision tables acquired from data are likely to be incomplete, due to the nature of learning from data

Analysis and Processing of Rough Decision Tables

- Analysis of dependencies occurring in the decision table: functional, partial functional
- Reduction - elimination of redundant or unrelated parts of the decision, such as:
  - Elimination of redundant columns (attribute reducts)
  - Elimination of redundant values (value reducts)
- Significance analysis of individual attributes
- Formation of minimal length, that is, most generalized predictive rules

Typical Issues with Applications to Real-World Data

- Imperfections of practical application data
- Presence of measurement noise
- Lack of consistency
- Extensive boundary regions
- Inter-data relationships are often probabilistic in nature, rather than deterministic
- Difficulty in creating any deterministic predictive rules or models from data

Variable Precision Approach to Rough Sets

- An attempt to create "softer" rough sets, more applicable to real world problems and imperfect data
- To utilize frequency distribution info in data when creating decision tables and rules (*probabilistic knowledge*)
- To allow for use of subjective probabilities obtained from human experts
- To enhance the scope of applications of rough set theory fundamental ideas

Single Parameter Variable Precision Model of Rough Sets

- Misclassification degree of set X with respect to Y:

$$c(X, Y) = 1 - \frac{card(X \cap Y)}{card(X)} = \frac{card(X \cap -Y)}{card(X)} = P(\neg Y|X)$$

$$\text{if } card(X) > 0 \text{ and } c(X, Y) = 0 \text{ if } card(X) = 0$$

- The partial majority inclusion of the set X within Y:

$$Y \supseteq^{\beta} X \text{ if and only if } c(X, Y) \leq \beta \text{ with } 0 \leq \beta < 0.5$$

Example:

Let $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$, $X = \{x_1, x_2, x_3, x_4\}$ and $Y = \{x_1, x_2, x_3, x_8\}$. Then $\neg Y = \{x_4, x_5, x_6, x_7\}$ giving $Y \supseteq^{0.25} X$

Set Approximations

- Lower approximation

$$\underline{R}_{\beta}(X) = \bigcup\{E \in R^* : c(E, X) \leq \beta\} = \bigcup\{E \in R^* : X \supseteq^{\beta} E\}$$

- Upper approximation

$$\overline{R}_{\beta}(X) = \bigcup\{E \in R^* : c(E, X) < 1 - \beta\}$$

- Negative region

$$NEGR_{\beta}(X) = \bigcup\{E \in R^* : c(E, X) \geq 1 - \beta\}$$

- Boundary region

$$BNR_{\beta}(X) = \bigcup\{E \in R^* : \beta < c(E, X) < 1 - \beta\}$$

Relationship to Pawlak Rough Sets

- $$\underline{R}_0(X) = \underline{R}(X)$$
- $$\overline{R}_0(X) = \overline{R}(X)$$
- $$BNR_0(X) = BNR(X)$$
- $$NEGR_0(X) = NEGR(X)$$

Variable Precision Rough Sets: Asymetric Bounds

- Two parameters used to control approximation regions
- Lower and upper limit $0 \leq l < u \leq 1$
- Lower approximation

$$\underline{R}_l(X) = \bigcup \{E \in R^* : c(E, X) \leq l\}$$

- Upper approximation

$$\overline{R}_u(X) = \bigcup \{E \in R^* : c(E, X) < u\}$$

- Negative region

$$NEGR_u(X) = \bigcup \{E \in R^* : c(E, X) \geq u\}$$

- Boundary region

$$BNR_{l,u}(X) = \bigcup \{E \in R^* : l < c(E, X) < u\}$$

When $l = 0$ and $u = 1$, the above definitions reduce to the original rough sets approximation regions (check)

Example: A Classification Table of Finishing Mill Data

Situation	Width	Gauge	Result
e_1	Wide	Heavy	**Good**
e_2	Wide	Heavy	**Bad**
e_3	Wide	Medium	**Bad**
e_4	Wide	Medium	**Good**
e_5	Wide	Thin	**Bad**
e_6	Wide	Thin	**Bad**
e_7	Narrow	Heavy	**Bad**
e_8	Narrow	Heavy	**Good**
e_9	Narrow	Medium	**Good**
e_{10}	Narrow	Medium	**Good**
e_{11}	Narrow	Thin	**Good**
e_{12}	Narrow	Thin	**Good**
e_{13}	Wide	Heavy	**Good**
e_{14}	Wide	Heavy	**Good**
e_{15}	Wide	Medium	**Good**
e_{16}	Wide	Thin	**Good**
e_{17}	Wide	Thin	**Good**
e_{18}	Narrow	Heavy	**Good**
e_{19}	Narrow	Heavy	**Good**
e_{20}	Narrow	Medium	**Bad**
e_{21}	Narrow	Thin	**Bad**
e_{22}	Narrow	Thin	**Bad**

Condition attributes: C={Width, Gauge}

Approximation Space Based on *Width* and *Gauge*

Condition Classes Forming Approximation Space:

- $(Width := Wide) AND (Gauge := Heavy)$ with $C_1 = \{1, 2, 13, 14\}$,
- $Width := Wide) AND (Gauge := Medium)$ with $C_2 = \{3, 4, 15\}$,
- $(Width := Wide) AND (Gauge := Thin)$ with $C_3 = \{5, 6, 16, 17\}$,
- $(Width := Narrow) AND (Gauge := Heavy)$ with $C_4 = \{7, 8, 18, 19\}$,
- $(Width := Narrow) AND (Gauge := Medium)$ with $C_5 = \{9, 10, 20\}$,
  $(Width := Narrow) AND (Gauge := Thin)$ with $C_6 = \{11, 12, 21, 22\}$.

Decision Classes:

- $(Result := Good)$ with
  $D_1 = \{1, 4, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19\}$
- $(Result :=$ **Bad**$)$ with $D_2 = \{2, 3, 5, 6, 7, 20, 21, 22\}$.

Approximations of $(Result := \textbf{Bad})$

$\underline{C}(D_2) = \emptyset$ then no deterministic decision rule exists for $(Result := \textbf{Bad})$

$$c(C_1, D_2) = 0.75,\ c(C_2, D_2) = 0.67,\ c(C_3, D_2) = 0.5$$
$$c(C_4, D_2) = 0.75,\ c(C_5, D_2) = 0.67,\ c(C_6, D_2) = 0.5$$

$$\underline{C}_{0.6}(D_2) = \bigcup\{C_i \in C^* : c(C_i, D_2) \leq 0.6\} = \bigcup\{C_3, C_6\}$$

The above means that non-deterministic rule exists for $(Result := \textbf{Bad})$

$$BNC_{0.6,0.7}(D_2) = \bigcup\{C_i \in C^* : 0.6 < c(C_i, D_2) < 0.7\} = \bigcup\{C_2, C_5\}$$

$$\overline{C}_{0.7}(D_2) = \bigcup\{C_i \in C^* : c(C_i, D_2) < 0.7\} = \bigcup\{C_2, C_3, C_5, C_6\}$$

$$NEGC_{0.7}(D_2) = \bigcup\{C_i \in C^* : c(C_i, D_2) \geq 0.7\} = \bigcup\{C_1, C_4\}$$

Probabilistic Rough Sets

- Rough set X *prior probability* , representing the probability of occurrence of X, in the absence of any other information:

$$P(X) = \frac{card(X)}{card(U)} \tag{1}$$

  with the assumption $0 < P(X) < 1$
- Elementary set E prior probability:

$$P(E) = \frac{card(E)}{card(U)} \tag{2}$$

  with the assumption $0 < P(E) < 1$
- Conditional occurrence probability of the rough set X within elementary set E:

$$P(X|E) = \frac{card(X \cap E)}{card(E)}, \tag{3}$$

- if U if infinite, the set measure theory can be used to substitute for cardinalities

Probabilistic Approximation Regions 1

- Based on the degree deviation of conditional probabilities from prior probability rather than presence or absence of a set inclusion
- Approximation regions characterize areas with *significantly* increased, *significantly* decreased, or approximately unchanged rough set X occurrence probability in relation to prior probability
- The approximation regions are defined in terms of two *precision control* parameters:
- The upper limit u and the lower limit l, with the constraint:

$$0 \leq l < P(X) < u \leq 1 \tag{4}$$

- The upper and lower limit parameters give precise meaning of the terms *significantly* decreased, or *significantly* increased rough set X occurrence probability

Probabilistic Approximation Regions 2

- Positive region:

$$POS_u(X) = \cup\{E : 0 < P(X) < u \leq P(X|E) \leq 1\}. \tag{5}$$

- Negative region:

$$NEG_l(X) = \cup\{E : 0 \leq P(X|E) \leq l < P(X) < 1\}. \tag{6}$$

- Boundary region

$$BND_{l,u}(X) = \cup\{E : l < P(X|E) < u\}. \tag{7}$$

- When $l = 0$ and $u = 1$, the above definitions reduce to the original rough sets approximation regions

Probabilistic Rough Sets in the Limit: Bayesian Rough Sets

- The lower and upper limit parameters are constrained by

$$0 \leq l < P(X) < u \leq 1 \tag{8}$$

- When both l and u approach $P(X)$, $l \rightarrow P(X)$ and $P(X) \leftarrow u$ then:

$$POS_u(X) \rightarrow POS^*(X) = \cup\{E : 0 < P(X) < P(X|E) \leq 1\} \tag{9}$$

$$NEG_u(X) \rightarrow NEG^*(X) = \cup\{E : 0 \leq P(X|E) < P(X) < 1\} \tag{10}$$

$$BND_{l,u}(X) \rightarrow BND^*(X) = \cup\{E : 0 < P(X|E) = P(X) < 1\}. \tag{11}$$

- The limit approximation regions are called Absolute Approximation Regions
- They provide the basis of Bayesian Rough Set model

Probabilistic Approximations: Example

- A rough set X is probabilistically defined by specifying definitions of its approximation regions elementary sets and their probabilistic relations to the rough set
- This can be done in a tabular form by creating a *probabilistic decision table*

Classification Table

Obj	HRate	BP	Temp	Treatm	**Response**
e_1	High	High	High	1	**Positive**
e_2	High	High	High	1	**Negative**
e_3	High	High	Normal	2	**Negative**
e_4	High	Normal	High	1	**Positive**
e_5	High	Normal	High	1	**Positive**
e_6	Normal	Normal	Normal	3	**Positive**
e_7	Normal	Normal	Normal	3	**Negative**
e_8	Low	High	High	1	**Negative**
e_9	Low	High	High	1	**Negative**
e_{10}	Low	High	Normal	2	**Positive**
e_{11}	Low	High	Normal	2	**Negative**
e_{12}	High	Normal	High	2	**Negative**
e_{13}	Normal	Normal	Normal	1	**Positive**

Table: Classification Table of medical records

Probabilistic Decision Tables

HRate	*BP*	*Temp*	*Treatment*	*Region*	$P(E_i)$	$P(X\|E_i)$
High	High	High	1	BND	0.0520	0.78
High	High	Normal	2	NEG	0.1354	0.02
High	Normal	High	1	POS	0.1562	0.99
Normal	Normal	Normal	3	BND	0.1562	0.36
Low	High	High	1	NEG	0.1406	0.11
Low	High	Normal	2	BND	0.1093	0.41
High	Normal	High	2	NEG	0.1562	0.27
Normal	Normal	Normal	1	POS	0.0941	0.85

Table: Probabilistic decision table with $l = 0.3$, $u = 0.8$, $P(X) = 0.4366$ of $Response :=$ **Positive**

Probabilistic Rules From Decision Tables

- The *Probabilistic rule* $r_{X|Y}$ is an expression: $des(Y) \rightarrow s(X)$
- Y is a definable set with a defining description $des(Y)$
- The description $des(Y)$ is a conjunction of attribute-value pairs
- X is a rough set referenced by $s(X)$
  - Three kinds of rules:
  - Positive rule: $Y \subseteq POS_u(X)$
  - Negative rule: $Y \subseteq NEG_l(X)$
  - Boundary rule: $Y \subseteq BND_{l,u}(X)$
- Rules corresponding to elementary sets E are *elementary rules*
- In practice, minimal length rules are of most interest

Minimal Probabilistic Rules

HRate	*BP*	*Temp*	*Treatment*	*Region*	$P(E_i)$	$P(X\|E_i)$
High	High	High	1	BND	0.0520	0.78
High	High	Normal	2	NEG	0.1354	0.02
High	Normal	High	1	POS	0.1562	0.99
Normal	Normal	Normal	3	BND	0.1562	0.36
Low	High	High	1	NEG	0.1406	0.11
Low	High	Normal	2	BND	0.1093	0.41
High	Normal	High	2	NEG	0.1562	0.27
Normal	Normal	Normal	1	POS	0.0941	0.85

- Correspond to *value reducts* of rough set theory
- Are maximally general (strong data support)
- Example: $BP = Normal \wedge Treatment = 1 \rightarrow PositiveResponse$

Rule Evaluative Parameters

- The rule $r_{X|Y}$ *certainty* parameter defined as the conditional probability $cert(r_{X|Y}) = P(X|Y)$
- The rule $r_{X|Y}$ *generality* (*support*,*strength*) parameter defined as the probability $gen(r_{X|Y}) = P(Y)$
- The rule *certainty gain* parameter $gain(r_{X|Y}) = |P(X|Y) - P(X)|$
- For all positive rules: $u \leq cert(r_{X|Y})$
- For all negative rules: $cert(r_{X|Y}) \leq l$

Rule Evaluative Measures

- The rules and the evaluative measures can be computed from the probabilistic decision table
- Rule certainty:

$$cert(r_{X|Y}) = \frac{\sum_{E \subseteq Y} P(E)P(X|E)}{\sum_{E \subseteq Y} P(E)}\}$$

- Rule generality:

$$gen(r_{X|Y}) = \sum_{E \subseteq Y} P(E)\}$$

- Certainty gain:

$$gain(r_{X|Y}) = \frac{|\sum_{E \subseteq Y} P(E)P(X|E) - P(X)\sum_{E \subseteq Y} P(E)|}{\sum_{E \subseteq Y} P(E)}\}$$

Evaluation of Rules: Example

HRate	*BP*	*Temp*	*Treatment*	*Region*	$P(E_i)$	$P(X\|E_i)$
High	High	High	1	BND	0.0520	0.78
High	High	Normal	2	NEG	0.1354	0.02
High	Normal	High	1	POS	0.1562	0.99
Normal	Normal	Normal	3	BND	0.1562	0.36
Low	High	High	1	NEG	0.1406	0.11
Low	High	Normal	2	BND	0.1093	0.41
High	Normal	High	2	NEG	0.1562	0.27
Normal	Normal	Normal	1	POS	0.0941	0.85

- $P(X) = 0.4366$
- $BP = Normal \wedge Treatment = 1 \rightarrow PositiveResponse$
  $Generality = 0.2503, \ Certainty = 0.9374, \ CertaintyGain = 0.5008$

Evaluation of Decision Tables: Dependency Measures

- The focus is on partial functional and probabilistic dependencies
- They capture the quality of approximation of the rough set X in terms of the elementary sets of the approximation space
- Parametric partial functional measure, $\gamma_{l,u}(X|C)$, is a variable precision generalization of Pawlak's partial functional dependency measure:

$$\gamma_{l,u}(X|C) = P(POS_u(X|C) \cup NEG_l(X|C))$$

  where C is a set of condition attributes.
- It represents the relative degree of accuracy of approximation of a rough set X
- It can be computed directly from probabilistic decision table by:

$$\gamma_{l,u}(X|C) = \sum_{E \subseteq POS_u(X|C) \cup NEG_l(X|C)} P(E)$$

Partial Functional Dependencies in Decision Tables

- Parametric partial functional measure, $\gamma_{l,u}(X|C)$ becomes partial functional dependency measure $\gamma(X|C)$ when $l = 0$ and $u = 1$:

$$\gamma(X|C) = \gamma_{0,1}(X|C) = \sum_{E \subseteq POS_1(X|C) \cup NEG_0(X|C)} P(E)$$

- Approximation regions *POS*,*BND* and *NEG* are definable and disjoint
- The dependency between condition attributes C and the attribute *Region* is functional

Dependency Monotonicity

- Monotonicity is a property of a dependency which ensures that the dependency will not decrease with the addition of an attribute
- It allows for efficient computation (linear in the number of attributes) of a minimal subset of attributes preserving the dependency, called *attribute reduct*

Theorem (Monotonicity of Partial Functional Dependency)

Let $B \subset C$ be a subset of condition attributes on U and let "a" be any condition attribute. Then the following relation holds:

$$\gamma(X|B) \leq \gamma(X|B \cup \{a\})$$

Parametric Probabilistic Dependencies

- The probabilistic dependency represents a degree of probabilistic dependency between classification formed by condition attributes and the classification $\{X, \neg X\}$
- The parametric probabilistic dependency, $\lambda - \textit{dependency}$, is a normalized expected maximum degree of deviation of the probability of the rough set $P(X|E)$, or $P(\neg X|E)$, from the prior probability of the rough set $P(X)$, or from the prior probability of the rough set $P(\neg X)$, respectively:

$$\lambda_{l,u}(X|C) = \frac{\sum_{E \subseteq POS_u(X|C) \cup NEG_l(X|C)} P(E)|P(X|E) - P(X)|}{P(POS_u(X|C))P(\neg X) + P(NEG_l(X|C))P(X)}$$

- The higher deviation reflects stronger probabilistic connection between elementary set E and the rough set X
- If all elementary sets E are probabilistically independent from rough set X, that is, if $P(X \cap E) = P(X)P(E)$, then $\lambda_{l,u}(X|C) = 0$

Probabilistic Dependencies in Bayesian Model

- To evaluate and optimize probabilistic decision tables derived in the framework of Bayesian Rough Set model, the non-parametric probabilistic dependency (probabilistic dependency), $\lambda - \mathit{Dependency}$, can be used:

$$\lambda(X|C) = \frac{\sum_{E \in U/C} P(E)|P(X|E) - P(X)|}{2P(X)(1 - P(X))} \tag{12}$$

- It represents the average degree of probabilistic dependency between elementary sets corresponding to combinations of attribute values and rough set X
- If all elementary sets E are probabilistically independent from rough set X, then $\lambda(X|C) = 0$
- It can be computed directly from the probabilistic decision table
- It is monotonic

Mnotonicity of Probabilistic Dependency

Theorem (Monotonicity of Probabilistic Dependency)

Let $B \subseteq C$ be a subset of condition attributes on U and let "a" be any condition attribute. Then the following relation holds:

$$\lambda(X|B) \leq \lambda(X|B \cup \{a\}) \tag{13}$$

- It allows for efficient procedure for decision table optimization and analysis of significance of attributes

Probabilistic Decision Table Reduction 1

- Minimal subset of attributes $RED \subseteq C$ preserving $\gamma(Region|C) = 1$ is called *functional reduct*:

$$\gamma(Region|RED) = 1 \tag{14}$$

and for any attribute $a \in RED$:

$$\gamma(Region|RED - \{a\}) < 1 \tag{15}$$

- After reduction, some elementary sets may combine requiring re-computation of $P(E)$ and $P(X|E)$

Probabilistic Decision Table Reduction Example

HRate	*BP*	*Temp*	*Treatment*	*Region*	$P(E_i)$	$P(X\|E_i)$
High	High	High	1	BND	0.0520	0.78
High	High	Normal	2	NEG	0.1354	0.02
High	Normal	High	1	POS	0.1562	0.99
Normal	Normal	Normal	3	BND	0.1562	0.36
Low	High	High	1	NEG	0.1406	0.11
Low	High	Normal	2	BND	0.1093	0.41
High	Normal	High	2	NEG	0.1562	0.27
Normal	Normal	Normal	1	POS	0.0941	0.85

- The functional reduct is *(HRate, BP, Treatments)*
- The attribute *Temp* can be eliminated

Reduced Probabilistic Decision Table Example

HRate	*BP*	*Treatment*	*Region*	$P(E_i)$	$P(X\|E_i)$
High	High	1	BND	0.0520	0.78
High	High	2	NEG	0.1354	0.02
High	Normal	1	POS	0.1562	0.99
Normal	Normal	3	BND	0.1562	0.36
Low	High	1	NEG	0.1406	0.11
Low	High	2	BND	0.1093	0.41
High	Normal	2	NEG	0.1562	0.27
Normal	Normal	1	POS	0.0941	0.85

Probabilistic Reduct

- Minimal subset of attributes preserving either the non-parametric $\gamma - dependency$ or $\lambda - dependency$), called *reduct*, or *probabilistic reduct*, respectively, *RED*):

$$\lambda(X|RED) = \lambda(X|C) \tag{16}$$

  and for any attribute $a \in RED$:

$$\lambda(X|RED - \{a\}) < \lambda(X|RED) \tag{17}$$

- Similarly for the partial functional dependencies

Significance and Core of Attributes

- Determining the contribution of individual reduct *RED* attributes to the dependency in question via their *significance* analysis
- Can be done by evaluating the relative decrease of dependency due to removal of an attribute *a* from the reduct *RED*:

$$sig_{RED}(a) = \frac{\lambda(X|RED) - \lambda(X|RED - \{a\})}{\lambda(X|RED)} > 0 \qquad (18)$$

- Finding the subset of most essential attributes with respect to the probabilistic dependency, the ones contained in ALL reducts called *core* attributes
- Any core attribute $\{a\}$ satisfies the following inequality:

$$\lambda(X|C) > \lambda(X|C - \{a\}). \qquad (19)$$

- Similarly for the non-parametric partial functional dependency

Issues and Directions

- The problem of converting precise numeric values into general qualitative values
- Value Range-based discretization is not good - possible formation of spurious new combinations
- Development of methods for avoiding *overfitting* and for reliable estimates of probabilities
- Reduction of combinatorial complexity: minimizing the number of attributes and their values
- Use of classification hierarchy
- Dealing with incomplete observations
- Developing applications: finance, medicine, pharmacy, control, pattern classification

Thank You!