Available online at www.sciencedirect.com

SCIENCE DIRECT®

International Journal of Approximate Reasoning 40 (2005) 81–91

ELSEVIER

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

www.elsevier.com/locate/ijar

The investigation of the Bayesian rough set model

Dominik Ślęzak, Wojciech Ziarko *

Department of Computer Science, University of Regina, Regina, SK, Canada S4S 0A2

Received 1 July 2004; accepted 1 November 2004
Available online 6 January 2005

Abstract

The original Rough Set model is concerned primarily with algebraic properties of approximately defined sets. The Variable Precision Rough Set (VPRS) model extends the basic rough set theory to incorporate probabilistic information. The article presents a non-parametric modification of the VPRS model called the Bayesian Rough Set (BRS) model, where the set approximations are defined by using the prior probability as a reference. Mathematical properties of BRS are investigated. It is shown that the quality of BRS models can be evaluated using probabilistic gain function, which is suitable for identification and elimination of redundant attributes.

1. Introduction

Most of practical data mining problems require identification of probabilistic patterns in data, typically in the form of probabilistic rules. To compute probabilistic rules using the rough set theory [4], the original Rough Set (RS) model (Pawlak's model) has to be "softened" to allow for some degree of uncertainty in approximating the target events. Several probabilistic extensions of the RS model have been proposed in the past. In particular, the Variable Precision Rough Set (VPRS) model

* Corresponding author. Tel.: +1 306 585 5213; fax: +1 306 585 4745.
E-mail address: ziarko@cs.uregina.ca (W. Ziarko).

0888-613X/$ - see front matter
doi:10.1016/j.ijar.2004.11.004

[17,1,19] was used as a basis of many algorithms for computation of probabilistic rules from data.

The VPRS model is parametric—definitions of positive and negative regions depend on the settings of permissible levels of uncertainty associated with each of the approximation regions. In some applications, however, it is not clear how to define the parameters. Also, using parameters is sometimes not required, as the general objective can be to increase certainty of a prediction that an event of interest would occur, or would not occur, rather than to find high probability rules, which might be impossible to get.

For example, in medical domain, the results of medical tests might indicate increased/decreased chances of a specific disease. Without the tests, its chances would be given by the prior probability of its occurrence in the general population. In such cases, the prior probability can be used as a benchmark value against which the quality of available information about domain objects can be measured. There are three possible scenarios in that respect: (i) the acquired information increases our perception of the chance that the event would happen; (ii) the acquired information decreases our perception of the chance that the event would not happen; (iii) the acquired information has no effect on our perception of the chance that the event would or would not happen.

In the latter scenario, the information is totally unrelated to the event, effectively forcing us to accept the prior probability as the only estimate of the chances that the event of interest would occur. Positive or negative deviation from the prior probability of an event is an improvement in our ability to better assess the chances of its occurrence or non-occurrence, respectively. In the context of rough set theory, the universe of interest can thus be divided into three regions: (i) the positive region where the probability of a target event is higher than its prior probability; (ii) the negative region where the probability of a target event is lower than its prior probability; (iii) the boundary region where it is equal, or approximately equal to the prior probability.

Such a categorization results in a new approach to VPRS model and rough set theory in general, referred to as to the *Bayesian Rough Set* (BRS) model. The name of the proposed model emphasizes its connections with fundamental ideas of Bayesian reasoning [12]. Its main feature, as opposed to the original VPRS approach, is the absence of parameters, which makes it appropriate for applications concerned with achieving *any* certainty gain in decision-making processes, rather than meeting specific certainty goals.

The article is organized as follows. Section 2 provides basics of data-based probabilistic calculus. Section 3 outlines the basics of the original theory of rough sets. Section 4 presents VPRS extension of the theory of rough sets. Section 5 introduces fundamental notions and properties of BRS model. Section 6 investigates the connections of the BRS model with Bayesian reasoning. Section 7 is concerned with information quality measures called certainty gain measures. Section 8 deals with some issues related to the attribute reduction in the context of BRS. Finally, Section 9 addresses directions for further research.

2. Probabilistic framework

Let U denote a universe of objects, infinite in general. We assume the existence of *probabilistic measure* P over *σ-algebra* $\mathscr{M}(U)$ of *measurable subsets* of U. We assume that all subsets (events) $X \in \mathscr{M}(U)$ under consideration are likely to occur but their occurrence not certain, that is $0 < P(X) < 1$.

We also assume the existence of an *indiscernibility* equivalence relation on U, with possibly infinite but countable family of measurable mutually disjoint classes (elementary sets) $\mathscr{E} \subseteq \mathscr{M}(U)$ such that $0 < P(E) < 1$, $E \in \mathscr{E}$, and $\sum_{E \in \mathscr{E}} P(E) = 1$. We will refer to this relation as to the IND-relation.

In practice, the elementary sets are obtained by grouping together objects having (almost) identical values of a selected set of features (attributes). In real-life applications, it is normally realistically assumed that $\mathscr{E}$ is finite. However, the results presented in this article hold also for IND-relations $\mathscr{E} \subseteq \mathscr{M}(U)$ with infinite, but countable collections of equivalence classes (elementary sets).

Each elementary set $E \in \mathscr{E}$ is associated with the conditional probabilities $P(X|E) = P(X \cap E)/P(E)$ and $P(E|X) = P(X \cap E)/P(X)$. The values of these probabilities are normally estimated based on a finite random sample $U_f \subseteq U$, using the following formulas:

$$
\begin{aligned}
\widetilde{P}(X) &= \frac{\mathrm{card}(X \cap U_f)}{\mathrm{card}(U_f)} \quad \widetilde{P}(X|E) = \begin{cases} \dfrac{\mathrm{card}(X \cap E \cap U_f)}{\mathrm{card}(E \cap U_f)} \Longleftrightarrow E \cap U_f \neq \emptyset \\ \text{undefined} \Longleftrightarrow E \cap U_f = \emptyset \end{cases} \\
\widetilde{P}(E) &= \frac{\mathrm{card}(E \cap U_f)}{\mathrm{card}(U_f)} \quad \widetilde{P}(E|X) = \begin{cases} \dfrac{\mathrm{card}(X \cap E \cap U_f)}{\mathrm{card}(X \cap U_f)} \Longleftrightarrow X \cap U_f \neq \emptyset \\ \text{undefined} \Longleftrightarrow X \cap U_f = \emptyset \end{cases}
\end{aligned} \tag{1}
$$

where card(∗) denotes cardinality of a set. Any other statistical estimates, depending on the type of a source of information, such as probability density functions etc. can also be used. In this article, while introducing and investigating the new model of rough sets, we do not assume any specific kind of probability estimation technique. Nevertheless, we refer to estimators (1) in examples, as they are most commonly used in rough set-related applications.

3. Rough set framework

In this section, we restate the basic Pawlak's model in probabilistic terms. It provides definitions of positive, negative and boundary approximation regions of events estimated using finite samples $U_f \subseteq U$.

The positive region $\widetilde{\mathrm{POS}}(X)$ of the event $X \subseteq U$ is an area of U_f where the occurrence of $X \cap U_f$ is certain. That is

$$
\widetilde{\mathrm{POS}}(X) = \bigcup \{E \cap U_f : E \cap U_f \subseteq X \cap U_f\} = \bigcup \{E \cap U_f : \widetilde{P}(X|E) = 1\} \tag{2}
$$

The negative region $\widetilde{\mathrm{NEG}}(X)$ is a mirror image of $\widetilde{\mathrm{POS}}(X)$. It covers an area of U_{f} where the occurrence of X is unlikely. That is

$$\widetilde{\mathrm{NEG}}(X) = \bigcup\{E \cap U_{\mathrm{f}} : E \cap X \cap U_{\mathrm{f}} = \emptyset\} = \bigcup\{E \cap U_{\mathrm{f}} : \widetilde{P}(X|E) = 0\} \tag{3}$$

The boundary region defines an area of U_{f} where the occurrence of the event X is possible but not certain. That is

$$\widetilde{\mathrm{BND}}(X) = \bigcup\{E \cap U_{\mathrm{f}} : 0 < \widetilde{P}(X|E) < 1\} \tag{4}$$

If the boundary of X is empty, then the set X is said to be definable. Otherwise we refer to X as a rough set [4]. One can see that $\widetilde{\mathrm{POS}}(X) = \widetilde{\mathrm{NEG}}(\neg X)$ and $\widetilde{\mathrm{BND}}(X) = \widetilde{\mathrm{BND}}(\neg X)$ for the complementary target event $\neg X = U \setminus X$. Therefore, X is definable, if and only if $\neg X$ is definable.

In (2)–(4) we use notation "$\sim$" to emphasize that the respective rough set regions are based on the estimated probabilities, using (1) in this particular case. In the rough set related literature there are many approaches to expressing probabilistic information (cf. [2,3,5,15]). Generally, one could use any type of estimation or simply refer to the *true* rough approximation regions

$$\begin{aligned} \mathrm{POS}(X) &= \bigcup\{E \in \mathscr{E} : P(X|E) = 1\} \\ \mathrm{NEG}(X) &= \bigcup\{E \in \mathscr{E} : P(X|E) = 0\} \\ \mathrm{BND}(X) &= \bigcup\{E \in \mathscr{E} : 0 < P(X|E) < 1\} \end{aligned} \tag{5}$$

where $P(X|E)$ are the *true* probabilities defined in σ-algebra $\mathscr{M}(U)$. A statistical problem is how accurate are "estimations" $\widetilde{\mathrm{POS}}(X)$, $\widetilde{\mathrm{NEG}}(X)$, $\widetilde{\mathrm{BND}}(X)$ with regards to the *true* regions defined by (5). This issue is beyond the scope of this article. Consequently, we present our approach without distinguishing between estimates and actual probabilities, using a unified notation P and skipping the symbol "$\sim$". All presented results hold both for finite samples and infinite universes, under assumptions introduced in Section 2.

4. Variable precision rough set model

The VPRS model [17] aims at increasing the discriminatory capabilities of the rough set approach by using parameter-controlled grades of conditional probabilities. The asymmetric VPRS generalization [1,19] is based on the lower and upper limit certainty thresholds l and u when defining approximation regions, satisfying $0 \leqslant l < P(X) < u \leqslant 1$.

The u-positive region $\mathrm{POS}_u(X)$ is controlled by the upper limit parameter u, which reflects the least acceptable degree of the conditional probability $P(X|E)$ to include elementary set E in $\mathrm{POS}_u(X)$. That is

$$\mathrm{POS}_u(X) = \bigcup\{E \in \mathscr{E} : P(X|E) \geqslant u\} \tag{6}$$

The l-negative region $\mathrm{NEG}_l(X)$ is controlled by the lower limit l, such that $0 \leqslant l < P(X)$. $\mathrm{NEG}_l(X)$ is an area where the occurrence of X is significantly—with respect to l—less likely than random guess $P(X)$. That is

$$\mathrm{NEG}_l(X) = \bigcup\{E \in \mathscr{E} : P(X|E) \leqslant l\} \tag{7}$$

The l-negative region $\mathrm{NEG}_l(X)$ can be expressed as the $(1 - l)$-positive region $\mathrm{POS}_{(1-l)}(\neg X)$ for $\neg X = U \setminus X$. Therefore, we can talk about a complete duality of positive and negative regions in the VPRS model. One can also consider the (l,u)-boundary region, which is a ''grey'' area where there is no sufficient probabilistic bias towards neither X nor $\neg X$. That is

$$\mathrm{BND}_{l,u}(X) = \bigcup\{E \in \mathscr{E} : l < P(X|E) < u\} \tag{8}$$

The VPRS model's ability to flexibly control approximation regions' definitions allows for capturing probabilistic relations existing in data. The original rough set model is a special case of VPRS, for $l = 0$ and $u = 1$. That is

$$\mathrm{POS}(X) = \mathrm{POS}_1(X), \;\; \mathrm{NEG}(X) = \mathrm{NEG}_0(X), \;\; \mathrm{BND}(X) = \mathrm{BND}_{0,1}(X)$$

Usually, however, more interesting results are expected for non-trivial settings $0 < l < P(X) < u < 1$, where l and u are appropriately tuned [18].

5. Bayesian rough set model

In some applications, for example in stock market, medical diagnosis etc., the objective is to achieve *some* certainty prediction improvement rather than trying to produce rules satisfying preset certainty requirements. Then, it is more appropriate not to use any parameters to control model derivation. In what follows, we present and investigate a modification of VPRS model, which allows for derivation of parameter-free predictive models. We call it the Bayesian Rough Set (BRS) model because of its connections with Bayesian reasoning.

The BRS positive region $\mathrm{POS}^*(X)$ defines an area of the universe where the probability of X is higher than the prior probability. It is an area of certainty improvement or gain with respect to predicting the occurrence of X. That is

$$\mathrm{POS}^*(X) = \bigcup\{E : P(X|E) > P(X)\} \tag{9}$$

The BRS negative region $\mathrm{NEG}^*(X)$ defines an area of the universe where the probability of X is lower than the prior probability. It is an area of certainty loss with respect to predicting the occurrence of X. That is

$$\mathrm{NEG}^*(X) = \bigcup\{E : P(X|E) < P(X)\} \tag{10}$$

The BRS boundary region is an area characterized by the lack of certainty improvement with respect to predicting neither X nor $\neg X$. That is

$$\mathrm{BND}^*(X) = \bigcup\{E : P(X|E) = P(X)\} \tag{11}$$

Information defining the boundary area is totally unrelated to X, which results in the same probabilistic distribution of objects belonging to X. In other words, the target event X is independent from all the elementary events in BND*(X), that is, for all $E \subseteq$ BND*(X) we have $P(X \cap E) = P(X)P(E)$.

One can compare properties of BRS with the classical rough set model. The following result provides us with the same duality properties as before.

Proposition 5.1. *For* $X \in \mathscr{M}(U)$ *and IND-relation* $\mathscr{E} \subseteq \mathscr{M}(U)$, $\mathrm{POS}^*(X) = \mathrm{NEG}^*(\neg X)$, $\mathrm{NEG}^*(X) = \mathrm{POS}^*(\neg X)$, *and* $\mathrm{BND}^*(X) = \mathrm{BND}^*(\neg X)$ *holds.*

Proof. Equalities with POS* and NEG* follow from equivalence $P(X|E) > P(X) \iff P(\neg X|E) = 1 - P(X|E) < 1 - P(X) = P(\neg X)$. For *BND** analogous equivalence with equalities instead of inequalities is applied. □

6. Connections with Bayesian reasoning

The nature of Bayesian reasoning is to combine the prior knowledge with the data-driven inverse probabilities to achieve the posterior probabilities [12]. The posterior probabilities are then compared to the prior ones to evaluate the obtained information. The following equation, which involves the posterior, prior and inverse *odd ratios*, helps in understanding this approach

$$\frac{P(X|E)}{P(\neg X|E)} = \frac{P(X)}{P(\neg X)} \cdot \frac{P(E|X)}{P(E|\neg X)} \tag{12}$$

Consequently, we could try to compare the inverse probabilities $P(E|X)$ and $P(E|\neg X)$ instead of comparing $P(X|E)$ with $P(X)$ and $P(\neg X|E)$ with $P(\neg X)$. The following result formalizes this intuition. The idea of defining rough set regions based on the inverse probabilities is further developed in [8,9].

Proposition 6.1. *For* $X \in \mathscr{M}(U)$ *and IND-relation* $\mathscr{E} \subseteq \mathscr{M}(U)$ *we have*

$$\mathrm{POS}^*(X) = \bigcup\{E \in \mathscr{E} : P(E|X) > P(E|\neg X)\}$$

$$\mathrm{NEG}^*(X) = \bigcup\{E \in \mathscr{E} : P(E|X) < P(E|\neg X)\}$$

$$\mathrm{BND}^*(X) = \bigcup\{E \in \mathscr{E} : P(E|X) = P(E|\neg X)\}$$

Proof. Let us consider the first case. The others are analogous. We have to show $P(X|E) > P(X) \iff P(E|X) > P(E|\neg X)$. Assume $P(X|E) > P(X)$. It implies $P(\neg X|E) = 1 - P(X|E) < 1 - P(X) = P(\neg X)$. Therefore, we get $(P(X|E)/P(X))/(P(\neg X|E)/P(\neg X)) > 1$, that is, using identity (12), $P(E|X)/P(E|\neg X) > 1$. Now, by contradiction, assume $P(X|E) \leqslant P(X)$. It implies $P(E|X)/P(E|\neg X) \leqslant 1$ in the same way as above. □

7. Certainty gain

The objective of predictive models is to increase the degree of certainty of decision making. In this section, we apply the certainty gain measure [10,19] to evaluation of the BRS regions. The *local gain* measure $g(X|E)$ is associated with every elementary set $E \in \mathscr{E}$ by

$$g(X|E) = P(X|E)/P(X) - 1 \tag{13}$$

It reflects the degree of certainty increase/decrease relative to the value of the prior probability $P(X)$.

Proposition 7.1. *For any $X \in \mathscr{M}(U)$ and IND-relation $\mathscr{E} \subseteq \mathscr{M}(U)$ we have*

$$g(X|E) > 0 \iff E \subseteq \mathrm{POS}^*(X) \iff E \subseteq \mathrm{NEG}^*(\neg X) \iff g(\neg X|E) < 0$$

$$g(X|E) < 0 \iff E \subseteq \mathrm{NEG}^*(X) \iff E \subseteq \mathrm{POS}^*(\neg X) \iff g(\neg X|E) > 0$$

$$g(X|E) = 0 \iff E \subseteq \mathrm{BND}^*(X) \iff E \subseteq \mathrm{BND}^*(\neg X) \iff g(\neg X|E) = 0$$

Proof. The proof is directly derivable from (13) and Proposition 5.1. □

Proposition 7.1 establishes a relationship between the gain function and BRS regions. Proposition 7.2 demonstrates the alternative representation.

Proposition 7.2. *For $X \in \mathscr{M}(U)$ and IND-relation $\mathscr{E} \subseteq \mathscr{M}(U)$ we have*

$$\mathrm{POS}^*(X) = \bigcup\{E \in \mathscr{E} : g(X|E) > g(\neg X|E)\}$$

$$\mathrm{NEG}^*(X) = \bigcup\{E \in \mathscr{E} : g(X|E) < g(\neg X|E)\}$$

$$\mathrm{BND}^*(X) = \bigcup\{E \in \mathscr{E} : g(X|E) = g(\neg X|E)\}$$

Proof. The proof is directly derivable from Proposition 7.1. □

Based on the local gain function, the *local relative gain* function is defined as

$$r(X|E) = \max\{g(X|E), g(\neg X|E)\} \tag{14}$$

It represents relative improvement in the prediction accuracy when predicting either X or $\neg X$, depending on the effect of the new information. The effect may be "positive", that is it can lead to higher chances of X, or "negative", leading to higher chances of $\neg X$.

The local relative gain can be applied to measure the average certainty gain over all elementary sets. It leads to the *global relative gain* defined as

$$R(X) = \sum_{E \in \mathscr{E}} P(E) r(X|E) \tag{15}$$

The following result provides helpful formulas for computation of $R(X)$.

Proposition 7.3. *For $X \in \mathscr{M}(U)$ and IND-relation $\mathscr{E} \subseteq \mathscr{M}(U)$, we have*

$$\begin{aligned} R(X) &= \sum_{E \in \mathscr{E}} \max\{P(E|X), P(E|\neg X)\} - 1 \\ &= P(\mathrm{POS}^*(X)|X) + P(\neg \mathrm{POS}^*(X)|\neg X) - 1 \end{aligned}$$

where $P(\mathrm{POS}^*(X)|X)$ is the probability of belonging to $\mathrm{POS}^*(X)$ conditioned by belonging to X, and $P(\neg \mathrm{POS}^*(X)|\neg X)$ is the probability of not belonging to $\mathrm{POS}^*(X)$ conditioned by not belonging to X.

Proof. Let us note that $r(X|E) = \max\{P(X|E)/P(X), P(\neg X|E)/P(\neg X)\} - 1$ and further that $P(E)r(X|E) = \max\{P(E|X), P(E|\neg X)\} - P(E)$, using the Bayes rule. Therefore $R(X)$ sums up to the first above form. Further, $\sum_{E \in \mathscr{E}} \max\{P(E|X), P(E|\neg X)\}$ equals to $\sum_{E \in \mathscr{E}: P(E|X) > P(E|\neg X)} P(E|X)$ plus $\sum_{E \in \mathscr{E}: P(E|X) \leqslant P(E|\neg X)} P(E|\neg X)$. The constraints for $E \in \mathscr{E}$ can be rewritten as $E \subseteq \mathrm{POS}^*(X)$ and $E \subseteq \mathrm{NEG}^*(X) \cup \mathrm{BND}^*(X)$. Therefore, the above sums equal to $P(\mathrm{POS}^*(X)|X)$ and $P(\neg \mathrm{POS}^*(X)|\neg X)$, respectively. □

The following result establishes a link between the global relative gain and the Pawlak's classical notion of definability [4].

Theorem 7.4. *For any $X \in \mathscr{M}(U)$ and IND-relation $\mathscr{E} \subseteq \mathscr{M}(U)$, we have inequalities $0 \leqslant R(X) \leqslant 1$. Moreover, the following properties hold:*

$$\begin{aligned} R(X) = 0 &\iff P(\mathrm{BND}^*(X)) = 1 \\ R(X) = 1 &\iff P(\mathrm{BND}(X)) = 0 \end{aligned}$$

Proof. According to Proposition 7.1 there is $r(X|E) \geqslant 0$, where equality holds if and only if $P(X|E) = P(X)$. Hence, we have $R(X) \geqslant 0$ with equality holding if and only if $E \subseteq \mathrm{BND}^*(X)$ for all $E \in \mathscr{E}$, i.e. $P(\mathrm{BND}^*(X)) = 1$.

According to Proposition 7.3 there is $R(X) \leqslant 1$, where equality holds if and only if $P(\mathrm{POS}^*(X)|X) = P(\neg \mathrm{POS}^*(X)|\neg X) = 1$. This can be rewritten as $P(X \setminus \mathrm{POS}^*(X)) = P(\neg X \setminus \neg \mathrm{POS}^*(X)) = 0$, or as $P(\mathrm{NEG}^*(X) \cup \mathrm{BND}^*(X) \setminus \neg X) = 0$ and $P(\mathrm{POS}^*(X) \setminus X) = 0$. The first of these equalities holds if and only if every $E \subseteq \mathrm{NEG}^*(X) \cup \mathrm{BND}^*(X)$ satisfies $P(E \setminus \neg X) = 0$, that is $P(\neg X|E) = 1$. The second equality holds if and only if every $E \subseteq \mathrm{POS}^*(X)$ satisfies $P(E \setminus X) = 0$, that is $P(X|E) = 1$. These facts can be expressed as implications $P(X|E) \leqslant P(X) \Rightarrow P(X|E) = 0$ and

$P(X|E) > P(X) \Rightarrow P(X|E) = 1$. It means that every $E \in \mathscr{E}$ must satisfy $P(X|E) = 0$ or $P(X|E) = 1$, consequently there are no $E \in \mathscr{E}$ in BND(X). □

8. Attribute reduction

One of the major applications of rough set theory is attribute reduction (cf. [4,6,7,16]), that is elimination of attributes considered as redundant while preserving quality of information. We will explore in this section the attribute reduction issue in the context of BRS using global relative gain function as measure of information quality.

Let us assume that every element E of IND-relation groups together objects $e \in U$ with identical values over a set of attributes A. We will denote by U/A the collection of elementary sets corresponding to this relation. Any alternative classification of U in terms of any subset of attributes $B \subseteq A$ will be denoted by U/B. To distinguish between U/A and U/B, for any $B \subseteq A$, we index all respective symbolic names with attribute set symbols. For example $\mathrm{POS}^*_B(X)$ means a positive BRS region of X obtained using classification U/B.

Global relative gain of X based on U/B will be denoted as $R_B(X)$. We are interested in comparing the values of $R_A(X)$ with $R_B(X)$ for subsets $B \subseteq A$. We adapt the notion of a rough set reduct and say that $B \subseteq A$ is an *R-reduct* for X, if and only if it satisfies the equality $R_B(X) = R_A(X)$, that is, it preserves the value of the global gain function, and none of its proper subsets does it. The following result is crucial for such a definition.

Theorem 8.1. *Let $B \subseteq A$ and $X \subseteq U$ be given. We have $R_B(X) \leqslant R_A(X)$ where equality holds, if and only if*

$$\mathrm{POS}^*_A(X) \subseteq \mathrm{POS}^*_B(X) \quad \text{and} \quad \mathrm{POS}^*_A(\neg X) \subseteq \mathrm{POS}^*_B(\neg X) \tag{16}$$

Proof. Any assertion $F \in U/B$ can be expressed as $F = \cup\{E \in U/A : E \subseteq F\}$. To prove $R_B(X) \leqslant R_A(X)$ it suffices to demonstrate that the following holds:

$$\max\{P(F|X), P(F|\neg X)\} \leqslant \sum_{E \in U/A : E \subseteq F} \max\{P(E|X), P(E|\neg X)\} \tag{17}$$

Clearly we have equalities $P(F|X) = \sum_{E \in U/A:E \subseteq F} P(E|X)$ and $P(F|\neg X) = \sum_{E \in U/A:E \subseteq F} P(E|\neg X)$. Hence, there is either $\max\{P(F|X), P(F|\neg X)\} = \sum_{E \in U/A:E \subseteq F} P(E|X)$ or $\max\{P(F|X), P(F|\neg X)\} = \sum_{E \in U/A:E \subseteq F} P(E|\neg X)$. We have also $\sum_{E \in U/A:E \subseteq F} P(E|X) \leqslant \sum_{E \in U/A:E \subseteq F} \max\{P(E|X), P(E|\neg X)\}$ and $\sum_{E \in U/A:E \subseteq F} P(E|\neg X) \leqslant \sum_{E \in U/A:E \subseteq F} \max\{P(E|X), P(E|\neg X)\}$. It implies inequality (17) and consequently $R_B(X) \leqslant R_A(X)$.

To prove that (16) is equivalent to $R_B(X) = R_A(X)$ let us note that equality in (17) holds, if and only if we have $\forall_{E \in U/A:E \subseteq F} P(E|X) \geqslant P(E|\neg X)$ or $\forall_{E \in U/A:E \subseteq F} P(E|X) \leqslant$

$P(E|\neg X)$. Consequently, equality in (17) holds, if and only if both implications $P(F|X) \geqslant P(F|\neg X) \Rightarrow \forall_{E\in U/A:E\subseteq F} P(E|X) \geqslant P(E|\neg X)$ and $P(F|X) \leqslant P(F|\neg X) \Rightarrow \forall_{E\in U/A:E\subseteq F} P(E|X) \leqslant P(E|\neg X)$ are satisfied. The first of them means that if $F \subseteq \mathrm{POS}^*_B(X) \cup \mathrm{BND}^*_B(X)$, then every $E \subseteq F$, $E \in U/A$, is included in $\mathrm{POS}^*_A(X) \cup \mathrm{BND}^*_A(X)$. The second implication means that if $F \subseteq \mathrm{NEG}^*_B(X) \cup \mathrm{BND}^*_B(X)$, then every $E \subseteq F$ is included in $\mathrm{NEG}^*_A(X) \cup \mathrm{BND}^*_A(X)$. Consequently, we obtain inclusions $\mathrm{POS}^*_B(X) \cup \mathrm{BND}^*_B(X) \subseteq \mathrm{POS}^*_A(X) \cup \mathrm{BND}^*_A(X)$ and $\mathrm{NEG}^*_B(X) \cup \mathrm{BND}^*_B(X) \subseteq \mathrm{NEG}^*_A(X) \cup \mathrm{BND}^*_A(X)$, equivalent to the inclusions (16). □

Theorem 8.1 indicates that the information gain will not increase when we replace the classification U/A by potentially less accurate classification U/B. Hence, it makes sense to investigate conditions for keeping that information at the appropriate level, such as in case of R-reducts. Searching for R-reducts is comparable to other rough set-based feature reduction techniques. We can use similar search heuristics, like for instance the method proposed in [6] based on *discernibility matrices*. To keep the value of R unchanged, we should check subsets $B \subseteq A$ to *discern* between elementary sets $E \in U/A$ belonging to $\mathrm{POS}^*_A(X)$ and $\mathrm{NEG}^*_A(X)$, which is equivalent to maintaining inclusions (16).

Another possibility is to calculate the gains directly from data using the sort operations and adapting the order-based genetic algorithms developed in [16]. This approach was extended into the case of *approximate* R-reducts satisfying inequality $R_B(X) \geqslant (1-\varepsilon)R_A(X)$, for a preset threshold $\varepsilon \in [0,1)$, and successfully applied in the medical domain [13,14]. The correspondence between approximate R-reducts and the BRS-like models was reported in [11].

9. Summary and conclusions

The objective of this article is a presentation and elementary investigation of a modification of VPRS, Bayesian rough set model, where the approximation regions are defined using prior probability of a set as a reference. The global relative gain function is used as the model's information quality measure. The measure captures the relative degree of increase of average certainty of predictions based solely on information represented by systems' attributes.

Presented approach appears to be well suited for data mining applications where the acquisition of probabilistic, rather than deterministic, predictive models is of primary importance. Further research is planned to evaluate the BRS model in comparison to the VPRS and original Pawlak's approaches.

Acknowledgements

The research reported in this article was supported in part by research grants from Natural Sciences and Engineering Research Council of Canada awarded to both authors. The comments and suggestions of Professor T.Y. Lin are gratefully appreciated.

References

[1] J. Katzberg, W. Ziarko, Variable precision rough sets with asymmetric bounds, in: Proc. of the Int. Workshop on Rough Sets and Knowledge Discovery (RSKD'93), Banff, Canada, 12–15 October 1993, pp. 163–191.

[2] T.Y. Lin, Sets with partial memberships: A rough sets view of fuzzy sets, in: Proc. of the World Congress of Computational Intelligence (WCCI'98), Anchorage, Alaska, 4–9 May 1998, pp. 785–790.

[3] T.Y. Lin, Granular fuzzy sets: A view from rough set and probability theories, Int. J. Fuzzy Syst. 3 (2) (2001) 373–381.

[4] Z. Pawlak, Rough Sets—Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, 1991.

[5] Z. Pawlak, A. Skowron, Rough membership functions, in: R.R. Yaeger, M. Fedrizzi, J. Kacprzyk (Eds.), Advances in the Dempster Shafer Theory of Evidence, John Wiley & Sons, 1994, pp. 251–271.

[6] A. Skowron, C. Rauszer, The discernibility matrices and functions in information systems, in: R. Slowinski (Ed.), Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory, Kluwer Academic Publishers, 1992, pp. 311–362.

[7] D. Ślęzak, Approximate entropy reducts, Fundamenta Informaticae 53 (3–4) (2002) 365–390.

[8] D. Ślęzak, The rough Bayesian model for distributed decision systems, in: Proc. of the 4th Int. Conf. on Rough Sets and Current Trends in Computing (RSCTC'2004), Uppsala, Sweden, 1–5 June, Springer Verlag, LNAI 3066, 2004, pp. 384–393.

[9] D. Ślęzak, Rough sets and Bayes factor, in: J. Peters, A. Skowron (Eds.), LNCS Transactions on Rough Sets III, Springer, Heidelberg, LNCS 3400, 2005, to appear.

[10] D. Ślęzak, W. Ziarko, Bayesian rough set model, in: Proc. of the Int. Workshop on Foundation of Data Mining (FDM'2002), December 9, Maebashi, Japan, 2002, pp. 131–135.

[11] D. Ślęzak, W. Ziarko, Attribute reduction in the Bayesian version of variable precision rough set model, in: Proc. of the Int. Workshop on Rough Sets and Knowledge Discovery (RSKD'2003), Elsevier, ENTCS 82/4, 2003.

[12] R. Swinburne (Ed.), Bayes's Theorem, Proc. of the British Academy, 113, 2003.

[13] S. Widz, K. Revett, D. Ślęzak, Application of rough set based dynamic parameter optimization to MRI segmentation, in: Proc. of 23rd Int. Conf. of the North American Fuzzy Information Processing Society (NAFIPS'2004), Banff, Canada, 27–30 June 2004.

[14] S. Widz, D. Ślęzak, K. Revett, An automated multispectral MRI segmentation algorithm using approximate reducts, in: Proc. of the 4th Int. Conf. on Rough Sets and Current Trends in Computing (RSCTC'2004), Uppsala, Sweden, June 1–5, Springer Verlag, LNAI 3066, 2004, pp. 815–824.

[15] S.K.M. Wong, W. Ziarko, Comparison of the probabilistic approximate classification and the fuzzy set model, Int. J. Fuzzy Sets Syst. 21 (1986) 357–362.

[16] J. Wróblewski, Theoretical foundations of order-based genetic algorithms, Fundamenta Informaticae 28 (3–4) (1996) 423–430.

[17] W. Ziarko, Variable precision rough sets model, J. Comput. Syst. Sci. 46 (1) (1993) 39–59.

[18] W. Ziarko, Decision making with probabilistic decision tables, in: Proc. of the 7th Int. Workshop on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC'99), Yamaguchi, Japan, Springer Verlag, LNAI 1711, 1999, pp. 463–471.

[19] W. Ziarko, Set approximation quality measures in the variable precision rough set model, Proc. of the 2nd Int. Conf. on Hybrid Intelligent Systems (HIS'02), Soft Computing Systems, 87, 2002, pp. 442–452.