

Chapter 6

Dependency Analysis and Attribute Reduction in the Probabilistic Approach to Rough Sets

Wojciech Ziarko

Abstract Two probabilistic approaches to rough sets are discussed in this chapter: the variable precision rough set model and the Bayesian rough set model, as they apply to data dependencies detection, analysis and their representation. The focus is on the analysis of data co-occurrence-based dependencies appearing in classification tables and probabilistic decision tables acquired from data. In particular, the notion of attribute reduct, in the framework of probabilistic approach, is of interest in the chapter. The reduct allows for information-preserving elimination of redundant attributes from classification tables and probabilistic decision tables. The chapter includes two efficient reduct computation algorithms.

Keywords Variable precision rough set model · Bayesian rough set model · Dependency analysis · Reduct

6.1 Introduction

The chapter reviews the basics of the variable precision rough set [2, 3, 7, 13, 15, 26, 28, 30, 32, 34, 35] and the Bayesian rough set [13] approaches to data dependencies detection, analysis and their optimal representation. The variable precision rough set and the Bayesian rough set theories are extensions of the rough set theory, as introduced by Pawlak [10, 11]. They are among many extensions and generalizations of the rough set approach, which inspired significant research interest worldwide (see, for example [5, 12, 17, 18, 22]). The primary motivation behind the research aimed at extending rough set approach is the imperfections of gathered practical application data. In particular, application data often suffer from presence of measurement noise, leading to lack of consistency and resulting difficulty to form data classifications and set approximations of the rough set model. In addition, the data often are real-valued, for example in pattern recognition or control applications, requiring

W. Ziarko (✉)

Department of Computer Science, University of Regina, Regina S4S 0A2, Canada
e-mail: ziarko@sasktel.net

initial preprocessing via a discretization procedure to make it applicable to rough set methodology. This pre-processing however leads to a loss of information and introduces a subjective factor into the method.

The variable precision and Bayesian rough set models are focused on the recognition and modelling of set overlap-based, also referred to as probabilistic, relationships between sets, which are most useful when dealing with noisy data. In this approach, the set-overlap relationships are used to construct approximations of undefinable sets [11]. The primary application of the approach is to the analysis of data co-occurrence-based dependencies in classification tables and probabilistic decision tables derived from data, as discussed in the following sections. Both, the probabilistic decision tables and classification tables are normally “learned” from data to represent some inter-data item connections, typically for the purpose of their analysis or data value prediction. The probabilistic decision tables can also be used as a basis of generalized probabilistic rule induction algorithms [29], but this topic is outside the scope of this chapter.

In practical applications of the data-acquired decision tables, one of the main issues is the identification of a minimal subset of attributes, which are discrete functions of measured features, to represent an identified data dependency without any loss, or with minimal loss, of information. The original general idea of attribute reduct, as introduced by Pawlak [10, 11], is applicable here. However, the original specific notion of reduct is applicable only to functional, or partial functional, data dependencies. In this chapter, we discuss an extended notion of reduct, as defined in the contexts of variable precision and Bayesian rough set models. The notion of reduct in these contexts allows for information-preserving identification of minimal subsets of attributes, in the presence of probabilistic dependencies between attributes.

The chapter is organized as follows. In the next section, we review the fundamentals of the variable precision rough set approach, which include the introduction of set approximations and the presentation of the basics of the related Bayesian rough set model. In Sect. 6.3, we discuss different kinds of probabilistic dependencies occurring between a “target set” and a partition of the universe of interest. The partition is assumed to represent our classification knowledge. The target set is our learning goal, whose approximate classification in terms of the classification knowledge we are trying to learn. The dependencies in question reflect our overall ability to create such a classification. In Sect. 6.4, the probabilistic attribute value-based decision tables are introduced, along with related classification tables. Both kinds of these tables represent our classification knowledge with respect to the target set.

The probabilistic decision tables additionally represent rough approximations of the target set, as defined in the framework of the variable precision rough set theory. The inter-attribute dependencies occurring in both, the probabilistic decision tables and classification tables, are subject of Sect. 6.5. All the discussed dependencies are of probabilistic nature and are either defined in the contexts of variable precision or Bayesian rough set models. They generalize and expand the attribute dependencies introduced by Pawlak in the original rough set theory [11]. Attribute reduction with respect to introduced dependencies is a subject of Sect. 6.6. The monotonicity property of the introduced λ —*dependency* measure allows for a definition of the notion of

information-preserving reduct with respect to this dependency. Couple of efficient, linear-time algorithms for computing single attribute reducts, either in classification tables or probabilistic decision tables, are presented. The ability to compute reducts allows us also to determine the importance, or significance of attributes. This is the subject of Sect. 6.7. Finally, in Sect. 6.8, we discuss the concept of generalized core attributes, the extension of the original core attributes introduced by Pawlak [10, 11]. The core attributes are the fundamental ones, which are preserved in every attribute reduction.

6.2 Variable Precision Rough Sets

In the rough set approach to data analysis, the crucial aspect is the existence of an ability, or knowledge, to form the *prior* classification of the universe of objects of interest into distinct classes. This ability, or *classification knowledge*, is usually associated with an external agent, such as medical professional for example, who is assumed to *know how* to classify objects (for example patients) into categories (for example, into health condition groups). However, in automated systems such an expert typically is not available. Instead, the system has to rely on measurements taken by system sensors (for example, temperature, blood pressure etc.) to perform the classification. In the rough set approach, the measurements are converted into discrete features called *attribute values*, which are then used to classify objects. We elaborate in detail about the attribute value-based classifications in Sect. 6.4.

The general variable precision rough set (VPRS) model does not make any assumptions how the prior classification was performed. It just assumes that some kind of prior knowledge exists and is represented in mathematical form by an equivalence relation, referred to as an indiscernibility relation IND on the universe U , $IND \subseteq U \times U$. The relation is assumed to have a finite number of equivalence classes, i.e. classification categories, called elementary sets. It should be noted that the assumption of finite number of classes may not be satisfied in general, but in attribute-value systems, which are the focus of this chapter, it is always the case. The collection of elementary sets of the IND relation will be denoted as IND^* . The pair (U, IND) is called an approximation space.

Let X be an arbitrary subset, referred to as the *target set*, of the universe U , $X \subseteq U$. In practice, the universe is a finite non-empty collection of objects of interest, such as medical patients, and the target set is our “goal” class, for example representing the class of patients suffering from a specific disease. Our objective is to create a system which would allow us to classify arbitrary objects into the “goal” class, or its complement, with an error rate which we would consider acceptable in the context of our criteria (which are domain-specific and, consequently, outside of the rough set model), but lower, on average, than in the case of random classification. For example, the objective may be to predict (diagnose) the presence, or absence, of a specific disease based on the results of medical tests, which are supposed to increase the accuracy of such predictions (if tests are properly designed) in comparison to predictions based solely on the frequency of occurrence of the disease in the population.

In the VPRS approach, each equivalence class E of the indiscernibility relation IND is assigned two measures which are: the relative “size” of the class E within universe U , referred to as the probability $P(E)$ of E , and the relative “size” of the target set X within an elementary set E , referred to as the conditional probability $P(X|E)$. The conditional probability, in this context, is just a measure of the degree of overlap between the target set X and the elementary set E . These two measures can be approximated from data respectively by:

$$P(E) = \frac{card(E)}{card(U)} \quad (6.1)$$

and

$$P(X|E) = \frac{card(X \cap E)}{card(E)} \quad (6.2)$$

where $card$ denotes set cardinality.

The target set X may be undefinable [11], which informally means that, in general, it cannot be expressed as a set union of some elementary sets forming our classification knowledge. That is, in general, the set definability criterion:

$$X = \cup\{E \in IND^* : E \subseteq X\} \quad (6.3)$$

is not satisfied.

This lack of definability is more common than not in applications. The original rough set theory, as introduced by Pawlak [10, 11], deals with this problem via the notions of lower and upper set approximations. However, in many applications, when the target set is not definable, this approach is not sufficient due to the absence of numeric assessments of the degree of association of elementary sets with the target set X .

The VPRS approach extends the rough set model to make it more flexible, by replacing the full inclusion relation with the overlap relation in the definitions of set approximations. Two precision control parameters called lower limit l and upper limit u are used in the definition of lower approximation of the target set X , or its complement. In this way, one can control the process of computation of approximations of the target set to identify such approximations which satisfy user-imposed criteria, such as for example, characterizing classes of patients with an elevated (or reduced) risk of a disease.

6.2.1 Set Approximations in the VPRS Approach

The approximations of the target set in the VPRS approach are defined in terms of unions of some elementary sets, as controlled by lower limit l and upper limit u precision parameters.

The notion of prior probability $P(X)$ plays also an essential role in the definitions of approximations, also called *approximation regions*: it represents the likelihood that a random object $e \in U$ is a member of the target set X in the absence of any classification knowledge about the object. If the classification knowledge is available, as represented by the equivalence relation IND , the likelihood of membership in the set X of objects belonging to different elementary sets can either increase, or decrease, or stay approximately the same as the prior probability $P(X)$. These variations in the set X membership likelihood across different elementary sets are reflected in the definitions of set approximation regions, which characterize areas of the universe U with significantly increased, significantly decreased, or approximately unchanged target set X membership probability.

Each elementary set is classified either into one of approximation regions of the set X , i.e. a positive region POS_u , a negative region NEG_l , or a boundary region $BND_{l,u}$. The upper limit u defines the positive region, or lower approximation, of the target set X , with the constraint $0 < P(X) < u \leq 1$. It represents the least acceptable degree of the conditional probability $P(X|E)$, or the set overlap degree, to include the elementary set E in the positive region. The positive region, or the lower approximation of the target set X , denoted as POS_u , is a collection of objects for which the probability of membership in the target set X is significantly higher than the prior probability $P(X)$, where the term *significantly higher* is precisely specified by the parameter u (as defined by some external criteria):

$$POS_u(X) = \cup\{E : P(X|E) \geq u\}. \quad (6.4)$$

The lower limit l defines the negative region of the target set X , with the constraint $0 \leq l < P(X) < 1$. It is the highest acceptable degree of the conditional probability $P(X|E)$ to include the elementary set E in the negative region. The negative region of the target set X , denoted as NEG_l is a collection of objects for which the probability of membership in the target set X is significantly lower than the prior probability $P(X)$, where the term *significantly lower* is precisely specified by the parameter l (as defined by some external, application-related, criteria):

$$NEG_l(X) = \cup\{E : P(X|E) \leq l\}. \quad (6.5)$$

The boundary region denoted as $BND_{l,u}$, is a collection of remaining objects which cannot be classified with sufficient certainty into either positive or negative regions. For the boundary area objects, the probability of membership in the target set X is not significantly different from the prior probability $P(X)$, that is:

$$BND_{l,u}(X) = \cup\{E : l < P(X|E) < u\}. \quad (6.6)$$

Regardless of the choice of lower and upper limit control parameters, the positive and negative approximation regions are subsets of *absolute approximation regions*, as described in the next subsection.

In the Pawlak’s rough set model [11], the notion of upper approximation of a set is defined as a union of all elementary sets which have non-empty intersection with the set. The generalized definition of upper approximation $UPP_l(X)$ in the VPRS approach, as in the original rough set model, is a set union of the positive region and of the boundary region giving:

$$UPP_l(X) = \cup\{E : P(X|E) > l\}. \tag{6.7}$$

Note that the generalized definition coincides with the Pawlak’s definition of upper approximation when $l = 0$. In addition, when $u = 1$, it can be easily demonstrated that the VPRSM definitions of positive, negative and boundary regions, become equivalent to the original rough set model’s definitions of lower approximation, negative and boundary regions [11].

One can also note that, in general, as opposed to Pawlak’s rough sets, it is not true that $POS_u(X) \subseteq X$ and it is not true that $X \subseteq UPP_l(X)$. Consequently, the rough set cannot be defined in the VPRSM as a pair consisting of upper and lower approximation, as it is done in Pawlak’s rough sets [11].

A frequently asked question is to how to set, or tune, the values of the precision control parameters l and u . The author’s point of view is that apart from the general constraint $0 \leq l < P(X) < u \leq 1$, the settings of the parameters are entirely dependent on the requirements of a practical application, while being likely subjective or obtained via the cost-benefit analysis [27].

6.2.2 Absolute Set Approximation Regions

To describe the areas of the universe characterized by an unconstrained increase, or decrease of the set X membership probability, the following definitions of absolute approximation regions are applicable. In this case, no parameters to specify “sufficiently” high increase, or decrease of the set membership probability in those areas are used. We call these areas *absolute approximation regions*.

The *absolute boundary region* of the target set X is a definable region of the universe U consisting of those elementary sets which are characterized by the unchanged probability of membership in the target set $X \subseteq U$, that is:

$$BND^*(X) = \cup\{E : P(X|E) = P(X)\}. \tag{6.8}$$

As it can be easily verified, in the absolute boundary region, each elementary set E is probabilistically independent from the target set X , i.e. $P(X \cap Y) = P(X)P(Y)$. Consequently, the whole boundary region is independent from the target set X . In other words, the objects in the absolute boundary regions can be considered entirely unrelated with the target set.

The region of the universe U that is characterized by an increased probabilistic connection with the target set $X \subseteq U$, relative to the prior probability $P(X)$, is called the *absolute positive region* of the set X , denoted as $POS^*(X)$:

$$POS^*(X) = \cup\{E : P(X|E) > P(X)\}. \quad (6.9)$$

In the absolute positive region of X , the likelihood of an object belonging to the target set is higher than in the whole universe U , but in practice that increase may be not sufficient from an application perspective.

Similarly, the *absolute negative region*, $NEG^*(X)$, of the target set X is an area of the universe U characterized by reduced likelihood of an object being a member of the target set X :

$$NEG^*(X) = \cup\{E : P(X|E) < P(X)\}. \quad (6.10)$$

The above definitions provide the basis of the Bayesian rough set model [13, 30].

6.3 Dependencies in Approximation Spaces

The probabilistic connections between elementary sets and the target set, and between definable sets and the target set in the approximation spaces can be quantified by using different dependency measures [24, 33]. Some of these measures are reviewed below.

6.3.1 Absolute Certainty Gain

Absolute certainty gain, denoted as $gabs$, evaluates the degree of one-directional dependency between any two sets. In the simplest case, it is a single-directional dependency measure representing the degree of change of the probability of membership in the set X for an object belonging to the elementary set E . The absolute certainty gain is defined by:

$$gabs(X|E) = |P(X|E) - P(X)|, \quad (6.11)$$

where $|\cdot|$ is the absolute value function.

The above definition can be extended to any definable set Y . The absolute certainty gain between the subsets X and Y can be computed directly from the available probabilistic knowledge based on the formula below, where the summation is over all elementary sets forming the definable set Y :

$$gabs(X|Y) = \frac{|\sum_{E \subseteq Y} P(E)P(X|E) - P(X)\sum_{E \subseteq Y} P(E)|}{\sum_{E \subseteq Y} P(E)}. \quad (6.12)$$

6.3.2 Absolute Dependency Gain

Another dependency measure is an *absolute dependency gain*, which is a bi-directional dependency measure used to evaluate the degree of the two-way connection between any two sets. Given two arbitrary subsets X and Y of the universe U , the absolute dependency gain, denoted as $dabs(X, Y)$, is defined by:

$$dabs(X, Y) = |P(X \cap Y) - P(X)P(Y)|. \quad (6.13)$$

The absolute dependency gain reflects the degree of probabilistic dependency between sets X and Y by quantifying the amount of deviation from the probabilistic independence between sets X and Y , as represented by the product $P(X)P(Y)$.

Similar to the absolute certainty gain, in an approximation space (U, IND) , if a subset Y is definable, then the absolute dependency gain between the subsets X and Y can be computed directly from the available probabilistic knowledge based on the following formula:

$$dabs(X, Y) = |\sum_{E \subseteq Y} P(E)P(X|E) - P(X)\sum_{E \subseteq Y} P(E)|. \quad (6.14)$$

The absolute boundary region of the target set X can alternatively be expressed by the absolute dependency gain as:

$$BND^*(X) = \cup\{E : dabs(X, E) = 0\}. \quad (6.15)$$

In other words, the absolute boundary region is an area with no dependency gain.

6.3.3 Average Dependency Gain

The average, or expected gain function, denoted as $egabs(X|IND)$, is a measure of the degree of probabilistic dependency between classification represented by the indiscernibility relation IND and the classification $(X, \neg X)$ of the universe U induced by the target set X , and its complement $\neg X$. It is a measure of dependency between two partitions of the universe U :

$$egabs(X|IND) = \sum_{E \in IND^*} |P(X \cap E) - P(X)P(E)| = \sum_{E \in IND^*} dabs(X, E). \quad (6.16)$$

When the dependency is functional, i.e. when set X is definable in Pawlak's sense [11], we have:

$$egabs(X|IND) = \sum_{E \in IND^*} |P(X \cap E) - P(X)P(E)| \quad (6.17)$$

that is:

$$egabs(X|IND) = \sum_{E \in IND^*} P(E)(1 - P(X)) = 1 - P(X) = P(\neg X). \quad (6.18)$$

Similarly, $egabs(\neg X|IND) = P(X)$ in the functional case.

In the case when $egabs = 0$, $P(X \cap E) = P(X)P(E)$, for every elementary set E . This means that for every elementary set E , $P(X|E) = P(X)$ and $P(E|X) = P(E)$. This is equivalent to saying that all elementary sets are probabilistically independent from the target set X . In practical terms, it means that the occurrence of an object belonging to any of the elementary sets does not affect in any way our ability to guess whether the object is the member of the set X , or of its complement $\neg X$.

6.4 Probabilistic Decision Tables

Probabilistic decision tables describe classes of approximation space and their probabilistic relations with a target set. They are composed of combinations of attribute values, probability values and approximation region designations.

6.4.1 Attributes

In many applications, the information about objects is expressed in terms of values of observations or measurements, often real-valued, referred to as *features*. For the purpose of rough set-based analysis and classifier construction, the feature values are typically mapped into finite-valued numeric or symbolic domains to form composite mappings, referred to as *attributes*. A common kind of mapping is dividing the range of values of a feature into a number of suitably chosen disjoint subranges via a discretization procedure (see, for example, [9]). Formally, an attribute a is a function on the universe U , $a : U \rightarrow a(U) \subseteq V_a$, where V_a is a finite set of values called the *domain* of the attribute a .

Based on combinations of attributes and their values, a structure of approximation space can be created and analyzed using general notions and results of rough set theory and of the VPRSM. Each attribute defines a classification of the universe U into classes corresponding to different values of the attribute. Each attribute value $v \in a(U)$, corresponds to a set of objects $E_v^a \subseteq U$ such that $E_v^a = a^{-1}(v) = \{e \in U : a(e) = v\}$. The classes E_v^a , referred to as *a-elementary sets*, form a partition of U . The equivalence relation corresponding to this partition will be denoted as IND_a . Similarly, an equivalence relation IND_B , and the corresponding approximation space, can be defined on the basis of any non-empty set of attributes B .

6.4.2 Decision Tables

A knowledge representation system [11] is a pair (U, A) , where U is a universe and A is a nonempty and finite set of attributes defined on U . In the context of rough set approach, decision tables are constructed in terms of knowledge representation systems as follows.

Let $C, D \subset A$ be two disjoint subsets of attributes, called condition and decision attributes, respectively. The condition attributes generate the partitioning of the universe U into classes of objects having identical values of attributes belonging to C , thus forming the structure of approximation space on U . The corresponding collection of elementary sets of this approximation space is denoted by U/C . Similarly, the decision attributes D induce a structure of approximation space on U , with U/D denoting its elementary sets. The knowledge representation system with defined condition and decision attributes is called a decision table [11]. Decision tables fall into two broad groups: *deterministic decision tables* and *non-deterministic decision tables*.

Deterministic decision tables describe the functional relation between a set of observations (inputs, conditions) and the corresponding decisions (outcomes). In practice, deterministic decision knowledge is not always available. When only some, but not all, decisions can uniquely be determined by combinations of attribute values, the decision table is called non-deterministic. In a non-deterministic decision table, the relationship between conditions and decisions is only partially functional.

Compared to the previous two types of decision tables, which are based on the original rough set theory, a *probabilistic decision table* is developed within the framework of the variable precision rough set theory. It contains some built-in probabilistic measures to help in the process of decision making or prediction in non-deterministic cases.

When defining the probabilistic decision tables, we focus on elementary sets (our target sets) of the decision attribute D , $X \in U/D$, of the partition generated by the decision attributes.

For a given target set X , the probabilistic decision table can be defined as a mapping associating each combination of condition attribute values, corresponding to an elementary set $E \in U/C$, with a triple of values representing:

1. the unique designation of the rough approximation region (positive, negative, or boundary region),
2. the respective values of the elementary set probability $P(E)$, and
3. the conditional probability $P(X|E)$.

In practice, when deriving a probabilistic decision table, the measures of $P(E)$ and $P(X|E)$ are usually computed based on available data. An example probabilistic decision table is shown in Table 6.1. It should be noted at this point, that while probabilistic decision tables are containing information about set approximation regions of the variable precision rough set model, and consequently depend on the settings of the parameters l and u , similar decision tables can be constructed based on Bayesian

Table 6.1 Probabilistic decision table

E_i	a_1	a_2	a_3	Region	$P(E_i)$	$P(X E_i)$
E_0	1	1	1	BND	0.0520	0.78
E_1	1	1	0	NEG	0.1354	0.02
E_2	1	0	1	POS	0.1562	0.99
E_3	1	0	0	BND	0.1562	0.36
E_4	0	1	1	NEG	0.1406	0.11
E_5	0	1	0	BND	0.1093	0.41
E_6	0	0	1	NEG	0.1562	0.27
E_7	0	0	0	POS	0.0941	0.85

rough set model, using absolute approximation regions. Another related issue is that the probabilistic decision tables can be structured into parent-child linear hierarchies, in which a parent boundary region provides a basis to form an approximation space for the child decision table [31]. In this way, the exponential growth of decision tables caused by the increase in the number of attributes can be effectively controlled without reducing the quality of rough approximations.

6.4.3 Classification Tables

An intermediate step leading to the probabilistic decision table is the creation of the *classification table*, as illustrated in Table 6.2. The classification table associates combinations of condition attribute values, for each elementary set $E \in U/C$, with a pair of corresponding $P(E)$ and $P(X|E)$ probability measures. In the example Table 6.2, the partitioning of U is obtained in terms of conditional attributes $C = \{a_1, a_2, a_3\}$, with the connected probabilistic measures. The information contained in the classification table can then be used to build rough approximations of any target set $X \in U/D$, based on pre-set values of the precision control lower and upper limit parameters l and u .

Table 6.2 Classification table

E_i	a_1	a_2	a_3	$P(E_i)$	$P(X E_i)$
E_0	1	1	1	0.0520	0.78
E_1	1	1	0	0.1354	0.02
E_2	1	0	1	0.1562	0.99
E_3	1	0	0	0.1562	0.36
E_4	0	1	1	0.1406	0.11
E_5	0	1	0	0.1093	0.41
E_6	0	0	1	0.1562	0.27
E_7	0	0	0	0.0941	0.85

Once the approximation region of each elementary set E was determined, the classification table can be converted into a probabilistic decision table. The creation of the probabilistic decision table involves adding an extra column, technically of a new decision attribute called *Region*, to mark the approximation region designation of each elementary set. The decision table created in that way is fully deterministic with respect to the new *Region* decision attribute which is representing the corresponding three approximation regions: *POS*, *NEG* and *BND*. This is illustrated in the example probabilistic decision Table 6.1, derived from the classification Table 6.2, with $l = 0.3$ and $u = 0.8$.

6.5 Dependencies in Decision Tables

In this section, dependencies between attributes occurring in classification tables and probabilistic decision tables are discussed. Specifically, our interest is in the dependencies occurring between condition attributes C , or their subset, and the two-class classification $(X, \neg X)$ formed by the target set X and its complement $\neg X$. This classification is numerically represented in both classification and probabilistic decision tables, by values of the conditional probability $P(X|E)$. Technically, the columns $P(E)$ and $P(X|E)$ can be treated as extra “attributes” associating some real values with elementary sets of the classification generated by condition attributes. In particular, the attribute $P(X|E)$ describes the distribution of the degrees of association across different elementary sets E with the target set X . Consequently, it can be used, in conjunction with the attribute $P(E)$, for computing the overall degree of association of the set of condition attributes, or of its subset, with the binary classification of the universe U , as defined by the target set X and its complement $\neg X$.

In our research, we identified two dependencies, called γ —*dependency* and λ —*dependency*, which provide useful measures for evaluating probabilistic decision tables. They also provide criteria for decision table optimization through reduction of redundant condition attributes.

6.5.1 Functional and Partial Functional Dependencies

Functional dependencies and partially functional dependencies between attributes of decision tables were originally explored in [11]. We will refer to them as γ —*dependencies*. They capture the quality of approximation of the target set $X \in U/D$ in terms of the elementary sets of the approximation space induced by condition attributes. We generalize them within the framework of the VPRS model by defining the γ —*dependencies* [33] as a relative size of the positive region of the two class partition $(X, \neg X)$, subject to prior setting of the values of the control parameters l and u :

$$\gamma_{l,u}(X|C) = P(POS_u(X|C) \cup NEG_l(X|C)), \quad (6.19)$$

where $POS_u(X|C)$ and $NEG_l(X|C)$, respectively are positive and negative regions of X in the approximation space induced on U by the set of condition attributes C . This dependency measure reflects the proportion of objects in the universe U that can be classified as members of the target set X , or a complement of the target set X , with sufficient certainty, as given by the parameters l and u .

The $\gamma_{l,u}(X|C)$ measure was inspired by the partial functional dependency measure $\gamma(D|C)$ introduced by Pawlak [11], which is given as a fraction of objects of the universe U that can be uniquely classified, based on their condition attributes value combinations, as members of some classes of the decision attribute D . More precisely, in the VPRS model terms:

$$\gamma(D|C) = \sum_{F \in U/D} P(POS_1(F|C)). \quad (6.20)$$

The above measures play useful role in decision table analysis and reduction of condition attributes.

6.5.2 λ —Dependency Measure

Another kind of dependency, unrelated to the the γ —dependencies measure and conveying different kind of information, is a parametric λ —dependency, denoted as $\lambda_{l,u}(X|C)$ [33]. It captures the average, or expected degree of the probabilistic connection between elementary sets E ($E \in U/C$) and the binary classification $(X, \neg X)$ corresponding to the target set X and its complement $\neg X$. The dependency is defined as a normalized expected degree of deviation of the conditional probability $P(X|E)$ from the prior probability $P(X)$:

$$\lambda_{l,u}(X|C) = \frac{\sum_{E \subseteq POS_u(X|C) \cup NEG_l(X|C)} P(E)|P(X|E) - P(X)|}{2P(X)(1 - P(X))}, \quad (6.21)$$

where $2P(X)(1 - P(X))$ is a normalization factor equal to the theoretically maximum value of the numerator summation, achievable only when X is definable in Pawlak's rough set's sense, independent of settings of the parameters l and u . The higher the deviation, the stronger the probabilistic connection between conditional attributes C and the decision partition $(X, \neg X)$, and vice versa, with the total probabilistic independence occurring at $\lambda_{l,u}(X|C) = 0$.

In the framework of the Bayesian rough set model, the parametric λ —dependency reduces to non-parametric λ —dependency defined as:

$$\lambda(X|C) = \frac{\sum_{E \in U/C} P(E)|P(X|E) - P(X)|}{2P(X)(1 - P(X))}. \quad (6.22)$$

The non-parametric λ —*dependency* $\lambda(X|C)$ is a normalized expected degree of deviation of the conditional probability $P(X|E)$ from the prior probability $P(X)$. The main practical advantage of the non-parametric λ —*dependency* is the absence of any external parameters, which may be difficult to obtain, to compute the dependency. Another useful advantage is its *monotonicity* with respect to condition attributes, as explained in the next section.

6.6 λ —*Dependency-Based Reduct*

The application of idea of *reduct*, introduced by Pawlak [10, 11], allows for optimization of representation of classification knowledge by providing a technique for removal of redundant attributes. The concept of reduct generated considerable amount of research interest, primarily as a method for feature selection [1, 2, 6, 8, 12–14, 16, 19–21, 23–25]. The general notion of reduct is applicable to the optimization of classification tables and probabilistic decision tables. The following theorem [13] demonstrates that the λ —*dependency* measure is *monotonic*, which means that expanding the set of condition attributes $B \subseteq C$ will not result in the decrease of the dependency level $\lambda(X|B)$.

Theorem 1 *Let $B \subseteq C$ be a subset of condition attributes on U and let “ a ” be any condition attribute. Then the following relation holds:*

$$\lambda(X|B) \leq \lambda(X|B \cup \{a\}). \quad (6.23)$$

As a consequence of the Theorem, the notion of the *probabilistic reduct* of attributes $RED \subseteq C$ can be defined as a minimal subset of attributes preserving the λ —*dependency* with the target classification $(X, \neg X)$.

The reduct satisfies the following two important properties:

$$\lambda(X|RED) = \lambda(X|C) \quad (6.24)$$

and for any attribute $a \in RED$:

$$\lambda(X|RED - \{a\}) < \lambda(X|RED). \quad (6.25)$$

The probabilistic reducts, called λ —*reducts*, can be computed using any methods available for reduct computation in the framework of the Pawlak’s original rough set approach, and in particular, a single λ —*reduct* can be easily computed from a classification table using the following λ —*Reduction* algorithm:

Algorithm 1 λ —Reduction:**Step 1:** Let *Initial Dependency* $\leftarrow \lambda(X|C)$;**Step 2:** Arrange condition attributes $a \in C$ in descending order based on the degree of λ —dependency measure $\lambda(X|\{a\})$;**Step 3:** Starting with the attribute with the lowest λ —dependency degree and proceeding in ascending order, perform the following two steps for all condition attributes:**Step 3.1:** Test the condition *Initial Dependency* $= \lambda(X|C - \{a\})$;**Step 3.2:** If *Initial Dependency* $= \lambda(X|C - \{a\})$ then eliminate the attribute a from the set of condition attributes C ;**Step 4:** The remaining set of condition attributes at the end of the process is a λ —reduct of the initial collection of condition attributes.

In the above algorithm, the condition attributes with the weakest connection with the target classification are eliminated first. Although this technique does not guarantee finding the shortest reduct, it appears to be a reasonable heuristic to find best attributes in the reduct. It should also be noted that the λ —reduct, in general, does not preserve the approximation regions of a target set X . This means that after computing the λ —reduct of a condition attributes, the approximation regions of a probabilistic decision table have to be re-computed again.

If the preservation of the approximation regions of a probabilistic decision table is of interest, the reduction of condition attributes can be conducted using γ —dependencies measure (Eq. 6.19), which is also monotonic. In this case, any reduct, referred to as γ —reduct, of condition attributes preserving the functional dependency between the condition attributes and the attribute *Region* indicating the approximation region of each elementary set, can be computed. A single γ —reduct can be identified using a variant of λ —Reduction algorithm, referred to as γ —Reduction algorithm:

Algorithm 2 γ —Reduction:**Step 1** Let *Initial Dependency* $\leftarrow 1$;**Step 2** Arrange condition attributes $a \in C$ in descending order based on the degree of λ —dependency measure $\lambda(X|\{a\})$;**Step 3** Starting with the attribute with the lowest λ —dependency degree and proceeding in ascending order, perform the following two steps for all condition attributes:**Step 3.1** Test the condition *Initial Dependency* $= \gamma(\text{Region}|C - \{a\})$;**Step 3.2** If *Initial Dependency* $= \gamma(\text{Region}|C - \{a\})$ then eliminate the attribute a from the set of condition attributes C ;**Step 4** The remaining set of condition attributes at the end of the process equals to a γ —reduct of the initial collection of condition attributes of a probabilistic decision table.

6.7 Probabilistic Decision Rules

Once the attribute reduct was computed, corresponding classification and decision tables can be formed based on the reduced set of condition attributes. Each row of either of such tables is a probabilistic decision rule with probabilistic “confidence factor” given by $P(X|E_i)$ attached to it. The “strength” of such a rule is given by the fraction of “supporting” cases, that is, $P(E_i)$. For example, the row for the elementary set E_2 of the classification Table 6.2, can be interpreted as a rule:

if $(a_1 = 1) \wedge (a_2 = 0) \wedge (a_3 = 1)$ **then** X with *confidence* = 0.99 and *strength* = 0.1562.

The rule of this kind gives the likelihood that a new object matching the rule’s preconditions will belong to the target set X .

Similarly, the probabilistic rules can be computed from probabilistic decision tables. In this case, the target set X is replaced by either positive, negative or boundary regions. For example, the row for the elementary set E_2 of the classification Table 6.2, can be interpreted as a rule:

if $(a_1 = 1) \wedge (a_2 = 0) \wedge (a_3 = 1)$ **then** POS with *confidence* = 0.99 and *strength* = 0.1562.

This rule specifies the likelihood that a new object matching the rule’s preconditions will belong to the positive region of the target set X . Clearly, these rules are dependent on the settings of the precision parameters l and u .

If required, the rules based on the probabilistic decision tables can be further simplified (or “generalized”, using machine learning terminology) by removing some unnecessary attribute-value pairs from their preconditions, without affecting their confidence factors. This objective can be accomplished by computing a *value reduct* of attributes [11]. Value reduct was used in some machine learning algorithms based on the rough set theory [31]. However, we will not elaborate more about this comprehensive topic in this chapter as it deserves another chapter of its own.

6.8 Significance of λ —Reduct Attributes

The λ —Reduct provides a method for computing fundamental factors of the λ —dependency.

The attributes appearing in a λ —reduct can be evaluated with respect to their contribution to the dependency with the target classification by adopting the notion of a *significance factor*. The significance factor $sig_{RED}(a)$ of an attribute $a \in RED$ is a relative decrease of the dependency $\lambda(X|RED)$ caused by removal of the attribute “ a ” from the reduct:

$$sig_{RED}(a) = \frac{\lambda(X|RED) - \lambda(X|RED - \{a\})}{\lambda(X|RED)}. \quad (6.26)$$

Similarly, the significance of attributes in a probabilistic decision table can be assessed within any γ —reduct, using the approach given above.

6.9 λ —Core Collection of Attributes

As in the original rough set approach [11], one can easily identify the set of most essential condition attributes with respect to the λ —dependency. These attributes, called the λ —core, are the ones which would never be eliminated in the process of any λ —Reduct computation. They are included in all λ —reducts i.e. their collection is equal to the intersection of all λ —reducts.

Any core attribute $\{a\}$ satisfies the following inequality:

$$\lambda(X|C) > \lambda(X|C - \{a\}). \quad (6.27)$$

The above inequality demonstrates that there is no need to compute all λ —reducts, which is NP-hard, to identify the λ —core as the core attributes can be found by simple linear testing procedure.

As in the case of λ —core attributes, γ —core attributes can also be computed in a probabilistic decision table with respect to the dependency $\gamma(\text{Region}|C)$ by testing the effect of removal of each condition attribute.

6.10 Final Remarks

The chapter reviews results of our long-term research on data dependencies, within the frameworks of the variable precision and Bayesian rough set models, occurring in approximation spaces and in both, classification and decision tables. These probabilistic dependencies are defined based on the degrees of overlap between sets. The primary dependency measures discussed in the chapter are γ —dependency and λ —dependency. They generalize and expand the attribute functional and partial functional dependency measures introduced by Pawlak [10, 11]. The applicability of the measures to creation, analysis and optimization of classification and decision tables, via the concept of attribute reduct, was also discussed and two reduct computation algorithms were presented. The variable precision rough set approach was used in many applications since its introduction in 1990s. To our best knowledge, the most comprehensive application, involving the use of hierarchies of probabilistic decision tables and the attribute dependency measures presented in this chapter, were the experiments with face recognition [4]. It is our belief that the theory and methods presented in the chapter will find additional useful applications in areas dealing with large amounts of data such as, for example, in medicine, pattern classification, market analysis and prediction, machine learning and data mining in general, just to mention a few areas where in our opinion this theory is applicable.

Acknowledgments Thanks are due to anonymous referees for their detailed and inspiring comments. The research reported in the chapter was supported by research grants from Natural Sciences and Engineering Research Council of Canada.

References

1. Bac, L., Tuan, N.: Using rough set in feature selection and reduction in face recognition problem. In: Proceedings of the 9th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining PAKDD. Lecture Notes in Artificial Intelligence, vol. 3518, pp. 226–233 (2005)
2. Beynon, M.: Reducts within the variable precision rough sets model: a further investigation. *Eur. J. Oper. Res.* **134**(3), 592–605 (2001)
3. Beynon, M., Peel, M.: Variable precision rough set theory and data discretization: an application to corporate failure prediction. *Int. J. Manag. Sci.* **29**, 561–576 (2001)
4. Chen, X., Ziarko, W.: Rough set-based incremental learning approach to face recognition. In: Proceedings of the International Conference on Rough Sets and Current Trends in Computing. Lecture Notes in Artificial Intelligence, vol. 6086, pp. 356–365 (2010)
5. Greco, S., Matarazzo, B., Slowinski, R.: Multicriteria classification by dominance-based rough set approach. In: Kloesgen, W., Zytkow, J. (eds.) *Handbook of Data Mining and Knowledge Discovery*, chap. C5.1.9. Oxford University Press, New York (2002)
6. Inuiguchi, M., Yoshioka, Y., Kusunoki, Y.: Variable-precision dominance-based rough set approach and attribute reduction. *Int. J. Approx. Reason.* **50**, 1199–1214 (2009)
7. Katzberg, J., Ziarko, W.: Variable precision rough sets with asymmetric bounds. In: Ziarko, W. (ed.) *Proceedings of the International Workshop on Rough Sets, Fuzzy Sets and Knowledge Discovery RSKD*, pp. 167–177. Springer, London (1994)
8. Mi, J., Leung, Y., Wu, W.: Approaches to attribute reduction in concepts lattices induced by axialities. *Knowl. Based Syst.* **23**(6), 504–511 (2010)
9. Nguyen, H.: On exploring soft discretization of continuous attributes. In: Pal, S.K., Polkowski, L., Skowron, A. (eds.) *Rough-Neural Computing: Techniques for Computing with Words, Cognitive Technologies*, pp. 333–350. Springer (2003)
10. Pawlak, Z.: Rough sets. *Int. J. Comput. Inf. Sci.* **11**, 341–356 (1982)
11. Pawlak, Z.: *Rough Sets: Theoretical Aspects of Reasoning About Data*. Kluwer, The Netherlands (1991)
12. Peters, J.F., Ramanna, S.: Feature selection: near set approach. In: Proceedings of the 3rd ECML/PKDD International Workshop on Mining Complex Data MCD, pp. 57–71 (2007)
13. Slezak, D., Ziarko, W.: Attribute reduction in the Bayesian version of variable precision rough set model. *Electron. Notes Theor. Comput. Sci.* **82**(4), 263–273 (2003)
14. Swiniarski, R., Skowron, A.: Rough set methods in feature selection and recognition. *Pattern Recognit. Lett.* **24**(6), 833–849 (2003)
15. Wei, L., Zhang, W.: Probabilistic rough sets characterized by fuzzy sets. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* **12**, 47–60 (2004)
16. Xia Wang, X., Zhang, W.: Relations of attribute reduction between object and property oriented concept lattices. *Knowl. Based Syst.* **21**(5), 398–403 (2008)
17. Yao, Y.: Decision theoretic rough set models, rough sets and knowledge. In: Proceedings of the 2nd International Conference on Rough Sets and Knowledge Technology RSKT. Lecture Notes in Artificial Intelligence, vol. 4481, pp. 1–12 (2007)
18. Yao, Y., Lin, T.: Generalization of rough sets using modal logic. *Intell. Autom. Soft Comput.* **2**(2), 103–120 (1996)
19. Yao, Y., Zhao, Y.: Discernibility matrix simplification for constructing attribute reducts. *Inf. Sci.* **179**(5), 867–882 (2009)
20. Yao, Y., Zhao, Y., Wang, J.: On reduct construction algorithms. In: Proceedings of the 1st International Conference on Rough Sets and Knowledge Technology RSKT. Lecture Notes in Artificial Intelligence, vol. 4062, pp. 297–304 (2006)
21. Zhang, W., Mi, J., Wu, W.: Approaches to knowledge reductions in inconsistent systems. *Int. J. Intell. Syst.* **18**(9), 989–1000 (2003)
22. Zhang, H., Leung, Y., Zhou, L.: Variable precision-dominance based rough set approach to interval-valued information systems. *Inf. Sci.* **244**, 75–272 (2013)

23. Zhang, J., Wang, J., Li, D., He, H., Sun, J.: A new heuristic reduct algorithm based on rough sets theory. In: Proceedings of the 4th International Conference on Advances in Web-Age Information Management WAIM. Lecture Notes on Computer Science, vol. 2762, pp. 247–253 (2003)
24. Zhao, Y., Luo, F., Wong, S., Yao, Y.: A general definition of an attribute reduct. In: Proceedings of the 2nd International Conference on Rough Sets and Knowledge Technology RSKT, Lecture Notes in Artificial Intelligence, vol. 4481, pp. 101–108 (2007)
25. Zhong, N., Dong, J.: Using rough sets with heuristics for feature selection. *J. Intell. Inf. Syst.* **16**, 199–214 (2001)
26. Ziarko, W.: Variable precision rough sets model. *J. Comput. Syst. Sci.* **46**(1), 39–59 (1993)
27. Ziarko, W.: Decision making with probabilistic decision tables. In: Proceedings of the 7th International Workshop on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing RSFDGrC. Lecture Notes on Artificial Intelligence, pp. 463–471. Springer, Yamaguchi (1999)
28. Ziarko, W.: Probabilistic decision tables in the variable precision rough set model. *Comput. Intell.* **17**(3), 593–603 (2002)
29. Ziarko, W.: Rough set approaches for discovery of rules and attribute dependencies. In: Kloesgen, W., Zytkow, J. (eds.) *Handbook of Data Mining and Knowledge Discovery*, pp. 328–339. Oxford University Press, New York (2002)
30. Ziarko, W.: Set approximation quality measures in the variable precision rough set model. In: Proceedings of the 2nd International Conference on Hybrid Intelligent Systems HIS. *Soft Computing Systems, Management and Applications*, vol. 87, pp. 442–452. IOS Press (2002)
31. Ziarko, W.: Acquisition of hierarchy—structured probabilistic decision tables, and rules from data. *Expert Syst., Int. J. Knowl. Eng. Neural Netw.* **20**(5), 10–305 (2003)
32. Ziarko, W.: Probabilistic rough sets. In: Proceedings of the 10th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing RSFDGrC. Lecture Notes in Computer Science, vol. 3641, pp. 283–293 (2005)
33. Ziarko, W.: Partition dependencies in hierarchies of probabilistic decision tables. In: Proceedings of the 1st International Conference on Rough Sets and Knowledge Technology RSKT. Lecture Notes in Artificial Intelligence, vol. 4062, pp. 42–49 (2006)
34. Ziarko, W.: Probabilistic approach to rough sets. *Int. J. Approx. Reason.* **49**(2), 272–284 (2008)
35. Ziarko, W.: Probabilistic Dependencies in Linear Hierarchies of Decision Tables. *Transactions on Rough Sets 9*, vol. 5390, pp. 444–454 (2008)