Generalized Probabilistic Approximations

Jerzy W. Grzymala-Busse[1,2]

[1] Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS 66045, USA
[2] Institute of Computer Science, Polish Academy of Sciences, 01–237 Warsaw, Poland
jerzy@ku.edu

Abstract. We study generalized probabilistic approximations, defined using both rough set theory and probability theory. The main objective is to study, for a given subset of the universe U, all such probabilistic approximations, i.e., for all parameter values. For an approximation space (U, R), where R is an equivalence relation, there is only one type of such probabilistic approximations. For an approximation space (U, R), where R is an arbitrary binary relation, three types of probabilistic approximations are introduced in this paper: singleton, subset and concept. We show that for a given concept the number of probabilistic approximations of given type is not greater than the cardinality of U. Additionally, we show that singleton probabilistic approximations are not useful for data mining, since such approximations, in general, are not even locally definable.

Keywords: Probabilistic approximations, parameterized approximations, generalization of probabilistic approximations, singleton, subset and concept probabilistic approximations.

1 Introduction

The entire rough set theory is based on ideas of the lower and upper approximations. Complete data sets, presented as decision tables, are well described by an indiscernibility relation, yet another fundamental idea of rough set theory. The indiscernibility relation is an equivalence relation. Standard lower and upper approximations were extended, using probability theory, to probabilistic (parameterized) approximations. Such approximations were studied, among others, in [1,2,3,4,5,6,7]. The parameter, called a threshold and associated with the probabilistic approximation, may be interpreted as a probability. The threshold is, in general, a real number.

So far probabilistic approximations were usually defined as lower and upper approximations. As it was observed in [8], the only difference between so called lower and upper probabilistic approximations is in the choice of the value of the threshold.

Due to the fact that we explore the set of all probabilistic approximations of a given type, the distinction between lower and upper approximations is blurred.

J.F. Peters et al. (Eds.): Transactions on Rough Sets XVI, LNCS 7736, pp. 1–16, 2013.

Therefore, we will define only one kind of probabilistic approximations for an approximation space (U, R), where U is a finite set and R is an equivalence relation on U.

This paper, for a given decision table and a subset of the universe explores the set of all probabilistic approximations. It is shown that the number of all distinct probabilistic approximations is quite limited.

Additionally, this paper generalizes the usual three types of approximations: singleton, subset and concept, used for approximation spaces (U, R), where R is an arbitrary binary relation. Similarly as for singleton standard approximations, a singleton probabilistic approximation of a subset X of the universe U is, in general, not definable. There are two types of definability, local and global. If the set X is globally definable, it is locally definable, the converse is, in general, not true. Sets that is the singleton probabilistic approximation of X are, in general, not even locally definable. The idea of probabilistic approximations is applied to incomplete data sets. It is well known [9,10] that incomplete data sets, i.e., data sets with missing attribute values, are described by characteristic relations, which are reflexive but, in general, neither symmetric nor transitive.

A preliminary version of this paper was prepared for the 6-th International Conference on Rough Sets and Knowledge Technology, Banff, Canada, October 9–12, 2011 [11].

2 Equivalence Relations

In this section we will discuss data sets without missing attribute values, i.e., complete. Complete data sets are describable by equivalence relations. Then we will discuss all probabilistic partitions defined over a space approximation (U, R), where U is a finite set and R is an equivalence relation.

2.1 Complete Data

Many real-life data sets have conflicting cases, characterized by identical values for all attributes but belonging to different concepts (classes). Data sets with conflicting cases are called inconsistent. An example of the inconsistent data set is presented in Table 1. The data set presented in Table 1 is inconsistent since it contains conflicting cases: the cases 2 and 4 are in conflict with the case 3 and the case 6 is in conflict with case 8.

In Table 1, the set A of all attributes consists of three variables *Temperature*, *Headache* and *Cough*. A *concept* is a set of all cases with the same decision value. There are two concepts in Table 1, the first one contains cases 1, 2, 4 and 6 and is characterized by the decision value *no* of decision *Flu*. The other concept contains cases 3, 5, 7 and 8 and is characterized by the decision value *yes*.

The fact that an attribute a has the value v for the case x will be denoted by $a(x) = v$. The set of all cases will be denoted by U. In Table 1, $U = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

Table 1. An inconsistent data set

	Attributes			Decision
Case	Temperature	Headache	Cough	Flu
1	normal	no	yes	no
2	normal	no	no	no
3	normal	no	no	yes
4	normal	no	no	no
5	high	yes	no	yes
6	high	yes	yes	no
7	high	no	yes	yes
8	high	yes	yes	yes

For an attribute-value pair $(a,\ v) = t$, a *block* of t, denoted by $[t]$, is a set of all cases from U such that for attribute a have value v. An *indiscernibility relation* R on U is defined for all $x, y \in U$ by

$$xRy \text{ if and only if } a(x) = a(y) \text{ for all } a \in A.$$

Equivalence classes of R are called *elementary sets* of R. An equivalence class of R containing x is denoted $[x]$. Any finite union of elementary sets is called a *definable set* [12]. Let X be a concept. In general, X is not a definable set. However, set X may be approximated by two definable sets, the first one is called a *lower approximation* of X, denoted by $\underline{appr}(X)$ and defined as follows

$$\cup \ \{[x] \mid x \in U,\ \ [x] \subseteq X\},$$

The second set is called an *upper approximation* of X, denoted by $\overline{appr}(X)$ and defined as follows

$$\cup \ \{[x] \mid x \in U,\ \ [x] \cap X \neq \emptyset\}.$$

For example, for the concept $[(Flu,\ no)] = \{1,\ 2,\ 4,\ 6\}$,

$$\underline{appr}(\{1, 2, 4, 6\}) = \{1\},$$

and

$$\overline{appr}(\{1, 2, 4, 6\}) = \{1, 2, 3, 4, 6, 8\}.$$

2.2 Probabilistic Approximations

Let (U, R) be an approximation space, where R is an equivalence relation on U. A probabilistic approximation of the set X with the threshold α, $0 < \alpha \leq 1$, is denoted by $appr_\alpha(X)$ and defined as follows

$$\cup\{[x] \mid x \in U,\ Pr(X|[x]) \geq \alpha\},$$

where $[x]$ is an elementary set of R and $Pr(X|[x]) = \frac{|X \cap [x]|}{|[x]|}$ is the conditional probability of X given $[x]$.

Obviously, the equivalence relation R uniquely defines a partition on U defined as the family of all elementary sets of R. Such a partition will be denoted by R^*. For Table 1, $R^* = \{\{1\}, \{2, 3, 4\}, \{5\}, \{6, 8\}, \{7\}\}$.

Table 2. Conditional probabilities

$[x]$	{1}	{2, 3, 4}	{5}	{6, 8}	{7}
$Pr(\{1, 2, 4, 6\} \mid [x])$	1	0.667	0	0.5	0

The number of distinct probabilistic approximations of the concept X is smaller than or equal to the number n of distinct thresholds α from the definition of a probabilistic approximation. The number n is equal to the number of distinct positive conditional probabilities $Pr(X|[x])$, where $x \in U$. Additionally, the number n is smaller than or equal to the number m of elementary sets $[x]$ of R. Finally, $m \leq |U|$. Thus the number of distinct probabilistic approximations of the given concept is smaller than or equal to the cardinality of U.

Table 2 shows conditional probabilities for all members of R^*. In Table 2 there are three positive conditional probabilities: 0.5, 0.667 and 1. Therefore there are only three probabilistic approximations:

$appr_{0.5}(\{1, 2, 4, 6\}) = \{1, 2, 3, 4, 6, 8\}$,

$appr_{0.667}(\{1, 2, 4, 6\}) = \{1, 2, 3, 4\}$,

and

$appr_1(\{1, 2, 4, 6\}) = \{1\}$.

Obviously, for the concept X, the probabilistic approximation of X computed for the threshold equal to the smallest positive conditional probability $Pr(X \mid [x])$ is equal to the upper approximation of X. Additionally, the probabilistic approximation of X computed for the threshold equal to 1 is equal to the lower approximation of X.

2.3 Rule Induction

In this subsection we assume that R is an equivalence relation. We will discuss how the existing rough set based data mining systems, such as LERS (Learning from Examples based on Rough Sets), may be used to induce rules using probabilistic approximations. As we will show, all what is necessary is, for every concept, to modify the input data set, run LERS, and then edit the induced rule set. We will illustrate this procedure by inducing a rule set for Table 1 and the concept $[(Flu, no)] = \{1, 2, 4, 6\}$ using the probabilistic approximation $appr_{0.667}(\{1, 2, 4, 6\}) = \{1, 2, 3, 4\}$. First, a new data set should be created in which for all cases that are members of the set $appr_{0.667}(\{1, 2, 4, 6\})$ the decision values are copied from the original data set (Table 1) and for all remaining cases, not being in the set $appr_{0.667}(\{1, 2, 4, 6\})$, a new decision value is introduced, say SPECIAL. Thus a new data set is created, see Table 3.

Table 3. A new data set

	Attributes			Decision
Case	Temperature	Headache	Cough	Flu
1	normal	no	yes	no
2	normal	no	no	no
3	normal	no	no	yes
4	normal	no	no	no
5	high	yes	no	SPECIAL
6	high	yes	yes	SPECIAL
7	high	no	yes	SPECIAL
8	high	yes	yes	SPECIAL

The LERS data mining system may be used to induce certain rules (from lower approximations) or possible rules (from upper approximations) [13]. Any rule r is characterized by the conditional probability $Pr(X|Y)$, where X is the concept and Y is the domain of r (the set of all cases described by the rule conditions). For a certain rule r, by definition, $Pr(X|Y) = 1$. We are interested in inducing probabilistic rules, with $Pr(X|Y > 0$, so we need to induce possible rules. For example, for Table 3,

$$\underline{appr}([(Flu, no)]) = \underline{appr}(\{1, 2, 4\}) = \{1\},$$

and

$$\overline{appr}([(Flu, no)]) = \overline{appr}(\{1, 2, 4\}) = \{1, 2, 3, 4\}.$$

Therefore, certain rules for [(*Flu*, *no*)] will describe only the set {1}, while possible rules for the same concept will describe all cases from the set {1, 2, 3, 4}, so the obvious choice is to use possible rules.

The data set presented in Table 3 should be inputted to the LERS system, where first the ordinary upper approximations of all concepts, [(*Flu*, *no*)], [(*Flu*, *yes*)] and [(*Flu*, *SPECIAL*)] are computed and then the MLEM2 algorithm [14] is applied. For Table 3, the MLEM2 algorithm will return the following rule set

1, 3, 4
(Temperature, normal) -> (Flu, no),
2, 1, 3
(Temperature, normal) & (Cough, no) -> (Flu, yes),
1, 4, 4
(Temperature, high) -> (Flu, SPECIAL).

Rules are presented in the LERS format, every rule is associated with three numbers: the total number of attribute-value pairs on the left-hand side of the rule, the total number of cases correctly classified by the rule during training, and the total number of training cases matching the left-hand side of the rule, i.e., the rule domain size. These numbers are computed by comparing induced rules with Table 3. In this rule set only rules describing the concept [(*Flu*, *no*)] are useful, the remaining rules should be deleted. Hence, only one rule is useful

1, 3, 4
(Temperature, normal) -> (Flu, no).

This rule describes the set {1, 2, 3, 4}, three cases (1, 2, and 4) truly belong to the concept.

For the second concept from Table 1, [(*Flu*, *yes*)] = {3, 5, 7, 8}, and for the following probabilistic approximation

$$appr_{0.667}(\{3, 5, 7, 8\}) = \{5, 7\},$$

the corresponding rule set may be induced from the data set presented in Table 4.

Table 4. A new data set

	Attributes			Decision
Case	Temperature	Headache	Cough	Flu
1	normal	no	yes	SPECIAL
2	normal	no	no	SPECIAL
3	normal	no	no	SPECIAL
4	normal	no	no	SPECIAL
5	high	yes	no	yes
6	high	yes	yes	SPECIAL
7	high	no	yes	yes
8	high	yes	yes	SPECIAL

The MLEM2 algorithm returns the following rule set

1, 4, 4
(Temperature, normal) -> (Flu, SPECIAL),
2, 2, 2
(Headache, yes) & (Cough, yes) -> (Flu, SPECIAL),
2, 1, 1
(Headache, yes) & (Cough, no) -> (Flu, yes),
2, 1, 1
(Temperature, high) & (Headache, no) -> (Flu, yes).

Among these four rules only the following two rules

2, 1, 1
(Headache, yes) & (Cough, no) -> (Flu, yes),
2, 1, 1
(Temperature, high) & (Headache, no) -> (Flu, yes).

describe the concept [(*Flu*, *yes*)]. Finally, the following rule set

1, 3, 4
(Temperature, normal) -> (Flu, no),
2, 1, 1
(Headache, yes) & (Cough, no) -> (Flu, yes),
2, 1, 1
(Temperature, high) & (Headache, no) -> (Flu, yes).

describes both concepts of the probabilistic approximations associated with the parameter $\alpha = 0.667$.

3 Arbitrary Binary Relations

In this section, first we will study approximations defined on the approximations space $A = (U, R)$ where U is a finite nonempty set and R is an arbitrary binary relation. Then we will extend corresponding definitions to generalized probabilistic approximations.

3.1 Non-parameterized Approximations

First we will quote some definitions from [15]. Let x be a member of U. The R-*successor* set of x, denoted by $R_s(x)$, is defined as follows

$$R_s(x) = \{y \mid xRy\}.$$

The R-*predecessor* set of x, denoted by $R_p(x)$, is defined as follows

$$R_p(x) = \{y \mid yRx\}.$$

For the rest of the paper we will discuss only R-successor sets and corresponding approximations.

Let X be a subset of U. The R-*singleton lower approximation* of X, denoted by $\underline{appr}^{singleton}(X)$, is defined as follows

$$\{x \mid x \in U, R_s(x) \subseteq X\}.$$

The singleton lower approximations were studied in many papers, see, e.g., [9,10,16,17,18,19,20,21,22,23].

The R-*singleton upper approximation* of X, denoted by $\overline{appr}^{singleton}(X)$, is defined as follows

$$\{x \mid x \in U, R_s(x) \cap X \neq \emptyset\}.$$

The singleton upper approximations, like singleton lower approximations, were also studied in many papers, e.g., [9,10,16,17,20,21,22,23].

The R-*subset lower approximation* of X, denoted by $\underline{appr}^{subset}(X)$, is defined as follows

$$\cup \ \{R_s(x) \mid x \in U, R_s(x) \subseteq X\}.$$

The subset lower approximations were introduced in [9,10].

The R-*subset upper approximation* of X, denoted by $\overline{appr}^{subset}(X)$, is defined as follows

$$\cup \ \{R_s(x) \mid x \in U, R_s(x) \cap X \neq \emptyset\}.$$

The subset upper approximations were introduced in [9,10].

The R-*concept lower approximation* of X, denoted by $\underline{appr}^{concept}(X)$, is defined as follows

$$\cup \ \{R_s(x) \mid x \in X, R_s(x) \subseteq X\}.$$

The concept lower approximations were introduced in [9,10].

The R-*concept successor upper approximation* of X, denoted by $\overline{appr}^{concept}(X)$, is defined as follows

$$\cup \ \{R_s(x) \mid x \in X, R_s(x) \cap X \neq \emptyset\} = \cup \ \{R_s(x) \mid x \in X\}.$$

The concept upper approximations were studied in [9,10,19].

3.2 Probabilistic Approximations

By analogy with standard approximations defined for arbitrary binary relations, we will introduce three kinds of probabilistic approximations for such relations: singleton, subset and concept.

The singleton probabilistic approximation of X with the threshold α, $0 < \alpha \leq 1$, denoted by $appr_{\alpha}^{singleton}(X)$, is defined as follows

$$\{x \mid x \in U, \ Pr(X|R_s(x)) \geq \alpha\},$$

where $Pr(X|R_s(x)) = \frac{|X \cap R_s(x)|}{|R_s(x)|}$ is the conditional probability of X given $R_s(x)$.

A subset probabilistic approximation of the set X with the threshold α, $0 < \alpha \leq 1$, denoted by $appr_{\alpha}^{subset}(X)$, is defined as follows

$$\cup \{R_s(x) \mid x \in U, \ Pr(X|R_s(x)) \geq \alpha\},$$

where $Pr(X|R_s(x)) = \frac{|X \cap R_s(x)|}{|R_s(x)|}$ is the conditional probability of X given $R_s(x)$.

A concept probabilistic approximation of the set X with the threshold α, $0 < \alpha \leq 1$, denoted by $appr_{\alpha}^{concept}(X)$, is defined as follows

$$\cup\{R_s(x) \mid x \in X,\ Pr(X|R_s(x)) \geq \alpha\},$$

where $Pr(X|R_s(x)) = \frac{|X \cap R_s(x)|}{|R_s(x)|}$ is the conditional probability of X given $R_s(x)$.

It is not difficult to see that the number of different probabilistic approximations of a given type (singleton, subset or concept) is not greater than the cardinality of U.

Obviously, for the concept X, the probabilistic approximation of a given type (singleton, subset or concept) of X computed for the threshold equal to the smallest positive conditional probability $Pr(X \mid [x])$ is equal to the standard upper approximation of X of the same type. Additionally, the probabilistic approximation of a given type of X computed for the threshold equal to 1 is equal to the standard lower approximation of X of the same type.

3.3 Incomplete Data Sets

It is well-known that any incomplete data set is described by a *characteristic relation* R, a generalization of the indiscernibility relation. The characteristic relation is reflexive but, in general, is neither symmetric nor transitive. For incomplete data sets R-definable sets are called *characteristic sets*, a generalization of elementary sets.

We distinguish between two types of missing attribute values: *lost* (e.g., the value was erased) and *"do not care" conditions* (such a value may be any value of the attribute), see [9,10].

An example of incomplete data set is presented in Table 5.

For incomplete decision tables the definition of a block of an attribute-value pair must be modified in the following way:

Table 5. An incomplete data set

	Attributes			Decision
Case	Temperature	Headache	Cough	Flu
1	normal	no	*	no
2	?	no	no	no
3	normal	*	no	yes
4	normal	no	?	no
5	high	yes	*	yes
6	high	yes	yes	no
7	high	?	yes	yes
8	high	yes	yes	yes

– If for an attribute a there exists a case x such that $a(x) =?$, i.e., the corresponding value is lost, then the case x should not be included in any blocks $[(a, v)]$ for all values v of attribute a,
– If for an attribute a there exists a case x such that the corresponding value is a "do not care" condition, i.e., $a(x) = *$, then the case x should be included in blocks $[(a, v)]$ for all specified values v of attribute a.

For a case $x \in U$ the *characteristic set* $K_B(x)$ is defined as the intersection of the sets $K(x, a)$, for all $a \in B$, where the set $K(x, a)$ is defined in the following way:

– If $a(x)$ is specified, then $K(x, a)$ is the block $[(a, a(x))]$ of attribute a and its value $a(x)$,
– If $a(x) =?$ or $a(x) = *$ then the set $K(x, a) = U$.

The characteristic set $K_B(x)$ may be interpreted as the set of cases that are indistinguishable from x using all attributes from B and using a given interpretation of missing attribute values.

For the data set from Table 5, the set of blocks of attribute-value pairs is
$[(Temperature, normal)] = \{1, 3, 4\}$,
$[(Temperature, high)] = \{5, 6, 7, 8\}$,
$[(Headache, no)] = \{1, 2, 3, 4\}$,
$[(Headache, yes)] = \{3, 5, 6, 8\}$,
$[(Cough, no)] = \{1, 2, 3, 5\}$,
$[(Cough, yes)] = \{1, 5, 6, 7, 8\}$.

The corresponding characteristic sets are

$K_A(1) = K_A(4) = \{1, 3, 4\}$,
$K_A(2) = \{1, 2, 3\}$,
$K_A(3) = \{1, 3\}$,
$K_A(5) = K_A(6) = K_A(8) = \{5, 6, 8\}$,
$K_A(7) = \{5, 6, 7, 8\}$.

Conditional probabilities of the concept {1, 2, 4, 6} given a characteristic set $K_A(x)$ are presented in Table 6.

For Table 5, all probabilistic approximations (singleton, subset and concept) are

Table 6. Conditional probabilities

$K_A(x)$	{1, 3, 4}	{1, 2, 3}	{1, 3}	{5, 6, 8}	{5, 6, 7, 8}
$Pr(\{1, 2, 4, 6\} \mid K_A(x))$	0.667	0.667	0.5	0.333	0.25

$$appr_{0.25}^{singleton}(\{1, 2, 4, 6\}) = U,$$

$$appr_{0.333}^{singleton}(\{1, 2, 4, 6\}) = \{1, 2, 3, 4, 5, 6, 8\},$$

$$appr_{0.5}^{singleton}(\{1, 2, 4, 6\}) = \{1, 2, 3, 4\},$$

$$appr_{0.667}^{singleton}(\{1, 2, 4, 6\}) = \{1, 2, 4\},$$

$$appr_{1}^{singleton}(\{1, 2, 4, 6\}) = \emptyset,$$

$$appr_{0.25}^{subset}(\{1, 2, 4, 6\}) = U,$$

$$appr_{0.333}^{subset}(\{1, 2, 4, 6\}) = \{1, 2, 3, 4, 5, 6, 8\},$$

$$appr_{0.5}^{subset}(\{1, 2, 4, 6\}) = \{1, 2, 3, 4\},$$

$$appr_{0.667}^{subset}(\{1, 2, 4, 6\}) = \{1, 2, 3, 4\},$$

$$appr_{1}^{subset}(\{1, 2, 4, 6\}) = \emptyset,$$

$$appr_{0.25}^{concept}(\{1, 2, 4, 6\}) = \{1, 2, 3, 4, 5, 6, 8\},$$

$$appr_{0.333}^{concept}(\{1, 2, 4, 6\}) = \{1, 2, 3, 4, 5, 6, 8\},$$

$$appr_{0.5}^{concept}(\{1, 2, 4, 6\}) = \{1, 2, 3, 4\},$$

$$appr_{0.667}^{concept}(\{1, 2, 4, 6\}) = \{1, 2, 3, 4\},$$

$$appr_{1}^{concept}(\{1, 2, 4, 6\}) = \emptyset.$$

3.4 Definability

Definability for completely specified decision tables should be modified to fit into incomplete decision tables. For incomplete decision tables, a union of some intersections of attribute-value pair blocks, where such attributes are members of B and are distinct, will be called B-*locally definable* sets. A union of characteristic sets $K_B(x)$, where $x \in X \subseteq U$ will be called a B-*globally definable* set. Any set X that is B-globally definable is B-locally definable, the converse is not true. For example, the set $\{1\}$ is A-locally definable since $\{1\} = [(Temperature, normal)] \cap [(Cough, yes)]$. However, the set $\{1\}$ is not A-globally definable. On the other hand, the set $\{1, 2, 4\} = appr_{0.667}^{singleton}(\{1, 2, 4, 6\})$ is not

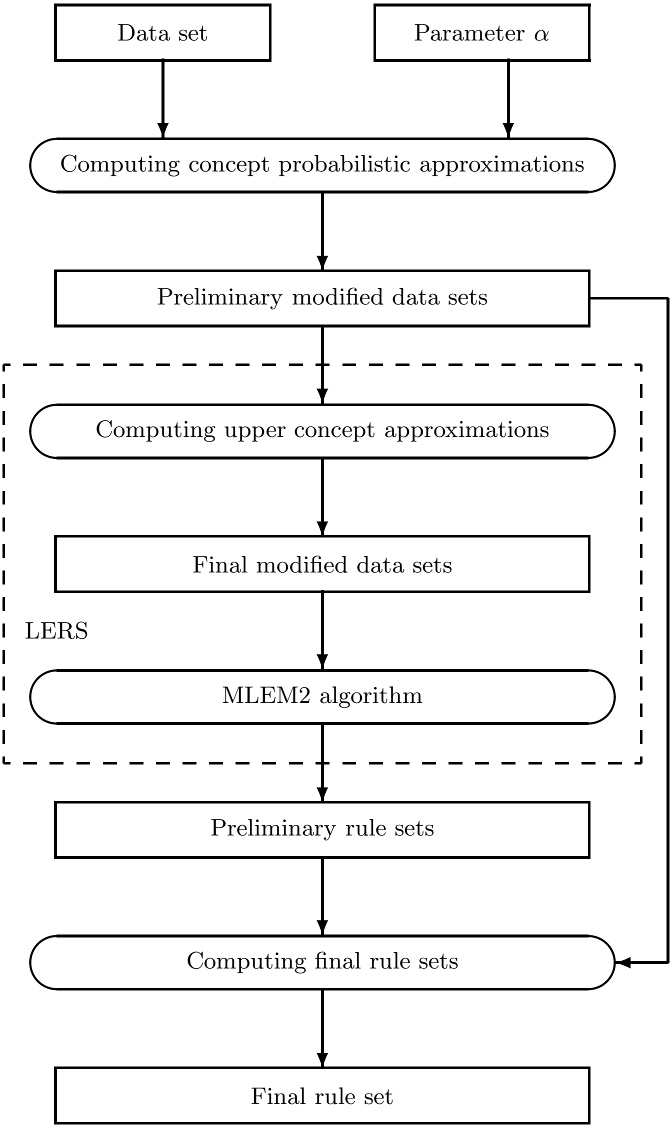

Fig. 1. Rule induction from probabilistic approximations using the LERS data mining system

even locally definable since in all blocks of attribute-value pairs containing the case 4 contain also the case 3 as well. Obviously, if a set is not B-locally definable then it cannot be expressed by rule sets using attributes from B. This is why it is so important to distinguish between B-locally definable sets and those that are not B-locally definable. In general, subset and concept probabilistic approximations are globally definable while singleton probabilistic approximations are not even locally definable.

3.5 Rule Induction

We will study how to adapt the LERS data mining system for rule induction from probabilistic approximations of the given concept. We will use a similar technique as in Subsection 3.3, i.e., for a concept and the probabilistic approximation of the concept we will create a new decision table. However, we have more choices since we may use a few different types of approximations.

Let us say that we want to induce rules for the concept [(*Flu*, *no*)] and the concept probabilistic approximation with the parameter $\alpha = 0.5$. The preliminary modified data set, constructed in the same way as described in Subsection 2.3, is presented in Table 7.

Table 7. A preliminary modified data set

	Attributes			Decision
Case	Temperature	Headache	Cough	Flu
1	normal	no	*	no
2	?	no	no	no
3	normal	*	no	yes
4	normal	no	?	no
5	high	yes	*	SPECIAL
6	high	yes	yes	SPECIAL
7	high	?	yes	SPECIAL
8	high	yes	yes	SPECIAL

This data set is inputted to the LERS data mining system, see Figure 1. The LERS system computes the upper concept approximation of the set $\{1, 2, 4, 6\}$, in our example it is $\{1, 2, 3, 4\}$, and the corresponding final modified data set. The MLEM2 algorithm induces the following preliminary rule set from the final modified data sets

1, 3, 4
(Headache, no) -> (Flu, no),
2, 1, 2
(Temperature, normal) & (Cough, no) -> (Flu, yes),
1, 4, 4
(Temperature, high) -> (Flu, SPECIAL).

where the three numbers that precede every rule are computed from Table 7. Obviously, only the first rule

1, 3, 4
(Headache, no) -> (Flu, no),

should be saved and the remaining two rules should be deleted in computing the final rule set.

Note that, in general, the result of computing the upper concept approximation by LERS results in the set

$$\cup \{K_A(y) \mid y \in \cup \{K_A(x) \mid x \in X, \; Pr(X|K_A(x)) \geq \alpha\}\}$$

which is a superset of the concept probabilistic approximation of X. For some data sets, for example for incomplete data sets with only lost values, both sets are identical. Nevertheless, in the preliminary rule set the three numbers that precede every rule are adjusted taking into account the preliminary modified data set. Thus during classification of unseen cases by the LERS classification system rules describe the original concept probabilistic approximation of the concept X.

4 Conclusions

In this paper we study a set of all probabilistic approximations, first for the approximation space (U, R), where U is a nonempty finite set and R is an equivalence relation, and then for the approximation space (U, R), where R is an arbitrary binary relation. For an arbitrary binary relation R standard definitions of singleton, subset and concept approximations are generalized to probabilistic approximations. It is shown that the set of such probabilistic approximations, even if R is an arbitrary binary relation, is finite and quite limited. Moreover, singleton probabilistic approximations of a subset X of the universe U is, in general, not even locally definable, so X is not expressible by a rule set. Therefore, singleton probabilistic approximations should not be used for data mining.

Acknowledgement. The author would like to thank the anonymous referees for all their valuable suggestions.

References

1. Pawlak, Z., Wong, S.K.M., Ziarko, W.: Rough sets: probabilistic versus deterministic approach. International Journal of Man-Machine Studies 29, 81–95 (1988)
2. Tsumoto, S., Tanaka, H.: PRIMEROSE: probabilistic rule induction method based on rough sets and resampling methods. Computational Intelligence 11, 389–405 (1995)
3. Yao, Y.: Decision-Theoretic Rough Set Models. In: Yao, J., Lingras, P., Wu, W.-Z., Szczuka, M.S., Cercone, N.J., Ślęzak, D. (eds.) RSKT 2007. LNCS (LNAI), vol. 4481, pp. 1–12. Springer, Heidelberg (2007)
4. Yao, Y.Y., Wong, S.K.M.: A decision theoretic framework for approximate concepts. International Journal of Man-Machine Studies 37, 793–809 (1992)
5. Yao, Y.Y., Wong, S.K.M., Lingras, P.: A decision-theoretic rough set model. In: Proceedings of the 5th International Symposium on Methodologies for Intelligent Systems, pp. 388–395 (1990)
6. Ziarko, W.: Variable precision rough set model. Journal of Computer and System Sciences 46(1), 39–59 (1993)
7. Ziarko, W.: Probabilistic approach to rough sets. International Journal of Approximate Reasoning 49, 272–284 (2008)
8. Grzymala-Busse, J.W., Marepally, S.R., Yao, Y.: An Empirical Comparison of Rule Sets Induced by LERS and Probabilistic Rough Classification. In: Szczuka, M., Kryszkiewicz, M., Ramanna, S., Jensen, R., Hu, Q. (eds.) RSCTC 2010. LNCS, vol. 6086, pp. 590–599. Springer, Heidelberg (2010)
9. Grzymala-Busse, J.W.: Rough set strategies to data with missing attribute values. In: Workshop Notes, Foundations and New Directions of Data Mining, in Conjunction with the 3rd International Conference on Data Mining, pp. 56–63 (2003)
10. Grzymała-Busse, J.W.: Data with Missing Attribute Values: Generalization of Indiscernibility Relation and Rule Induction. In: Peters, J.F., Skowron, A., Grzymała-Busse, J.W., Kostek, B.z., Swiniarski, R., Szczuka, M.S. (eds.) Transactions on Rough Sets I. LNCS, vol. 3100, pp. 78–95. Springer, Heidelberg (2004)
11. Grzymała-Busse, J.W.: Generalized Parameterized Approximations. In: Yao, J., Ramanna, S., Wang, G., Suraj, Z. (eds.) RSKT 2011. LNCS, vol. 6954, pp. 136–145. Springer, Heidelberg (2011)
12. Pawlak, Z.: Rough Sets. Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1991)
13. Grzymala-Busse, J.W.: A new version of the rule induction system LERS. Fundamenta Informaticae 31, 27–39 (1997)
14. Grzymala-Busse, J.W.: MLEM2: A new algorithm for rule induction from imperfect data. In: Proceedings of the 9th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, pp. 243–250 (2002)
15. Grzymala-Busse, J.W., Rząsa, W.: Definability and Other Properties of Approximations for Generalized Indiscernibility Relations. In: Peters, J.F., Skowron, A. (eds.) Transactions on Rough Sets XI. LNCS, vol. 5946, pp. 14–39. Springer, Heidelberg (2010)
16. Kryszkiewicz, M.: Rough set approach to incomplete information systems. In: Proceedings of the Second Annual Joint Conference on Information Sciences, pp. 194–197 (1995)
17. Kryszkiewicz, M.: Rules in incomplete information systems. Information Sciences 113, 271–292 (1999)

18. Lin, T.Y.: Neighborhood systems and approximation in database and knowledge base systems. In: Proceedings of the ISMIS 1989, the Fourth International Symposium on Methodologies of Intelligent Systems, pp. 75–86 (1989)
19. Lin, T.Y.: Topological and fuzzy rough sets. In: Slowinski, R. (ed.) Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory, pp. 287–304. Kluwer Academic Publishers, Dordrecht (1992)
20. Slowinski, R., Vanderpooten, D.: A generalized definition of rough approximations based on similarity. IEEE Transactions on Knowledge and Data Engineering 12, 331–336 (2000)
21. Stefanowski, J., Tsoukiàs, A.: On the Extension of Rough Sets under Incomplete Information. In: Zhong, N., Skowron, A., Ohsuga, S. (eds.) RSFDGrC 1999. LNCS (LNAI), vol. 1711, pp. 73–82. Springer, Heidelberg (1999)
22. Stefanowski, J., Tsoukias, A.: Incomplete information tables and rough classification. Computational Intelligence 17(3), 545–566 (2001)
23. Yao, Y.Y.: Relational interpretations of neighborhood operators and rough set approximation operators. Information Sciences 111, 239–259 (1998)