A View on Rough Set Concept Approximations

Jan Bazan[1], Nguyen Hung Son[2], Andrzej Skowron[2], and Marcin Szczuka[2]

[1] Institute of Mathematics,University of Rzeszów
Rejtana 16A, 35-959 Rzeszów, Poland
[2] Institute of Mathematics, Warsaw University
Banacha 2, 02-097 Warsaw, Poland
{bazan,son,skowron,szczuka}@mimuw.edu.pl

Abstract. The concept of approximation is one of the most fundamental in rough set theory. In this work we examine this basic notion as well as its extensions and modifications. The goal is to construct a parameterised approximation mechanism making it possible to develop multi-stage multi-level concept hierarchies that are capable of maintaining acceptable level of imprecision from input to output.

1 Introduction

The notion of concept approximation is a focal point of many approaches to data analysis based on rough set theory. The original concept of indiscernibility and approximation, as introduced by Pawlak in [5], is meant to provide a way of dealing with inconsistency and incompleteness in data. Elegantly and simply devised, the rough set approximations proved to be a useful tool in supporting data-related tasks such as classification, decision making, and description.

In the majority of rough set applications the approximations are used only at some initial stage of inductive learning. However, in most cases the final system is based on extended representation. Majority of the existing solutions (see [2,3, 15]) make use of decision rules derived from data and accompanied with a *recipé* for their usage. We show that approximations of concepts retrieved from such a system can be enriched by using some patterns defined by rule based classifiers or other classifiers. Such approach is much more flexible and better suited for multi-stage reasoning schemes (see [11]) that are discussed further in the paper.

The importance of proper approximation choice becomes much more crucial when it comes to construction of compound decision support and classification systems. These systems create higher level concepts using as building blocks not only attribute values but also previously derived, more primitive concepts. The lower level concepts may be created with use of approximate techniques as well. Therefore, higher level concepts incorporate not only the imprecision resulting from the way they are being constructed but also the imprecision inherited from the lower level concepts. The crucial point is to assure the parameterised space of possible approximations. Then, by proper tuning of parameters we may control the proliferation of imprecision. In rough set terms this corresponds to setting size and shape of boundary region.

G. Wang et al. (Eds.): RSFDGrC 2003, LNAI 2639, pp. 181–188, 2003.

To illustrate the problems that require multi-layered approximation schemes and compound concept approximation let us bring two examples related to RoboCup [14] and WITAS project [11,17]. The RoboCup [14] is an initiative aimed at fostering research in the field of cooperative autonomous robotics by construction of programs for control of robots which play football. The aim of WITAS project is to construct an unmanned aerial vehicle (autonomous helicopter) capable of recognising the road situation underneath and take an appropriate action. The aircraft is equipped with several sensors, most importantly - a video camera. In both examples, the decision making process should not only take into account simple observations, but also the higher level concepts such as situation of other objects on the field.

The paper presents the topics sketched above in the following way. First, the concept of rough set approximation is introduced and discussed. After that we present and discuss the approximations induced by rule sets and concept assessment schemes accompanying them. Finally we present how the proposed approach deals with the issue of construction of compound concept hierarchies while preserving approximation quality. We also make connections with the concept of parameterised approximation space [11] and ideas connected to the approach known as *rough mereology* [8,9].

2 Rough Set Preliminaries

2.1 Information Systems

An *information system* [5] is a pair $\mathbb{S} = (U, A)$, where U is a non-empty, finite set of *objects* and A is a non-empty, finite set, of *attributes*. Each $a \in A$ corresponds to the function $a : U \to V_a$, where V_a is called the *value set* of a. We associate with any non-empty set of attributes $B \subseteq A$ *the* B*-information signature* for any object $x \in U$ by $inf_B(x) = \{(a, a(x)) : a \in B\}$. The set $\{inf_A(x) : x \in U\}$ is called the A*-information set* and it is denoted by $INF(\mathbb{S})$.

In supervised learning problems, objects from training set are pre-classified into several *categories* or *classes*. To deal with such type of data we use *decision systems* of the form $\mathbb{S} = (U, A, dec)$, where $dec \notin A$ is a distinguished attribute called *decision* and elements of attribute set A are called *conditions*. In practice, decision systems contain description of a finite sample U of objects from larger (may be infinite) universe $\mathcal{U}$. Conditions are such attributes that their values are known for all objects from $\mathcal{U}$, but decision is a function defined on the objects from the sample U only. Usually decision attribute is a characteristic function of an unknown concept or several concepts on a sample of objects. Without loss of generality one can assume that the domain V_{dec} of the decision dec is equal to $\{1, \ldots, d\}$. The decision dec determines a partition $\{CLASS_1, \ldots, CLASS_d\}$ of the universe U, where $CLASS_k = \{x \in U : dec(x) = k\}$ is called the k-th *decision class of* $\mathbb{S}$ for $1 \leq k \leq d$. By *class distribution* of any set $X \subseteq U$ we denote the vector $ClassDist(X) = \langle n_1, \ldots, n_d \rangle$, where $n_k = card(X \cap CLASS_k)$ is the number of objects from X belonging to the k-th decision class.

2.2 Concept Approximation

In many real life situations, we are not able to give an exact definition of the concept. Such uncertain situations are caused by either the lack of information about the concept or by the richness of natural language. There are different approaches to deal with uncertain and vague concepts like multi-valued logics, fuzzy set theory, and rough set theory. Using these approaches, concepts are defined by "multi-value membership function" instead of classical "binary (crisp) membership relations" (set characteristic functions). In particular, what we want to underline in this paper, rough set approach offers a way for establishing membership functions that is data-grounded and significantly different from others.

In rough set methods, it is assumed that there exists a concept X defined over the huge universe $\mathcal{U}$ of objects ($X \subseteq \mathcal{U}$). The problem is to find a description of the concept X, which can be expressed in a predefined descriptive language, which is a set of formulas that are interpretable as subsets of $\mathcal{U}$. In general, the problem is to find a description of a concept X in a language $\mathcal{L}_2$ (e.g., consisting of boolean formulas defined over subset of attributes) assuming the concept is definable in another language $\mathcal{L}_1$ (e.g., natural language, or defined by a set of attributes).

Usually, the concept X is specified partially, i.e., value of characteristic function of X is given only on a small subset $U \subseteq \mathcal{U}$ called training sample. Such information makes it possible to search for patterns in a given language defining on the training sample sets included (or sufficiently included) in a given concept. Observe that the approximations of a concept can not be defined uniquely from a given sample of objects. The approximations of the whole concept X are obtained by induction from given information on a sample U of objects (containing some positive examples $X \cap U$ and negative examples $\overline{X} \cap U$). Hence, the quality of such approximations should be verified on new testing objects. Thus we propose to search for concept approximations gradually. Parameterised patterns defined by rough membership functions related to classifiers help to discover relevant patterns on the object universe extended by adding new testing objects. In the paper we present illustrative examples of such parameterised patterns. By tuning parameters of such patterns one can obtain patterns relevant for concept approximation of the extended training sample by testing objects from U^* where $U \subseteq U^* \subseteq \mathcal{U}$.

Due to bounds on expressiveness of language $\mathcal{L}$ in the universe $\mathcal{U}$, we are forced to find some approximated rather than exact description of a given concept. Rough set methodology for approximation of a concept $X \subseteq \mathcal{U}$, assuming X and $\mathcal{U} - X$ are known only on a sample $U \subseteq \mathcal{U}$, can be based on finding pairs $\mathbb{P} = (\mathbf{L}, \mathbf{U})$ of object sets in $\mathcal{U}$ satisfying the following conditions:

1. $\mathbf{L}, \mathbf{U}, \mathcal{U} \setminus \mathbf{L}, \mathcal{U} \setminus \mathbf{U}$ are subsets of $\mathcal{U}$ expressible in the language $\mathcal{L}$;
2. $\mathbf{L} \cap U \subseteq X \cap U \subseteq \mathbf{U} \cap U$;
3. the set $\mathbf{L}$ ($\mathbf{U}$) is maximal (minimal) in the family of sets definable in $\mathcal{L}$ satisfying (2).

The sets $\mathbf{L}$ and $\mathbf{U}$ are called the *lower approximation* and the *upper approximation* of the concept $X \subseteq \mathcal{U}$ (*generated by its sample on* U), respectively. The set

$\mathbf{BN} = \mathbf{U} \setminus \mathbf{L}$ is called the *boundary region of approximation* of X. The set X is called *rough* with respect to its approximations $(\mathbf{L}, \mathbf{U})$ if $\mathbf{L} \neq \mathbf{U}$, otherwise X is called *crisp* in $\mathcal{U}$. In practical applications the last constraint in the above definition can be hard to satisfy. Hence, by using some heuristics sub-optimal instead of maximal or minimal sets are constructed. The rough approximation of concept can be also defined by means of rough membership function.

A function $f : \mathcal{U} \rightarrow [0, 1]$ is called *a rough membership function* of the concept $X \subseteq \mathcal{U}$ approximated by $(\mathbf{L}, \mathbf{U})$ (assuming X and $\mathcal{U} - X$ are known only on a sample $U \subseteq \mathcal{U}$) if and only if $\mathbf{L} = \mathbf{L_f} = \{x \in \mathcal{U} : f(x) = 1\}$ and $\mathbf{U} = \mathbf{U}_f = \{x \in \mathcal{U} : f(x) > 0\}$.

Rough set approximations [5,6] are fundamental and widely used in many reasoning methods under uncertainty (caused, e.g., by the lack of some attributes). For a given information system $\mathbb{S} = (U, A)$ and an attribute set $B \subseteq A$, one can define a *B-indiscernibility relation* $IND(B)$ assuming $IND(B) = \{(x, y) \in U \times U : inf_B(x) = inf_B(y)\}$. Its equivalence classes are defined by $[x]_{IND(B)} = \{u \in U : (x, u) \in IND(B)\}$ for any object $x \in U$. The problem is to define a concept $X \subseteq U$, assuming that only some attributes from $B \subseteq A$ are given. This problem is often specified by a decision system $\mathbb{S}_1 = (U, B, dec_X)$, where $dec_X(u) = 1$ for $u \in X$, and $dec_X(u) = 0$ for $u \notin X$. Attributes from B determine the rough membership function $\mu_X^B : U \rightarrow [0, 1]$ for the concept X by $\mu_X^B(x) = card(X \cap [x]_{IND(B)})/card([x]_{IND(B)})$. This function, according to the rough membership function definition, yields rough approximations of the concept X by using indiscernibility classes:

$$\mathbf{L}_B(X) = \mathbf{L}_{\mu_X^B} = \left\{x \in U : \mu_X^B(x) = 1\right\} = \left\{x \in U : [x]_{IND(B)} \subseteq X\right\}$$

$$\mathbf{U}_B(X) = \mathbf{U}_{\mu_X^B} = \left\{x \in U : \mu_X^B(x) > 0\right\} = \left\{x \in U : [x]_{IND(B)} \cap X \neq \varnothing\right\}$$

called the *B-lower* and the *B-upper approximation* of X in $\mathbb{S}$, respectively. The set $\mathbf{BN}_B(X) = \mathbf{U}_B(X) \setminus \mathbf{L}_B(X)$ is called *B-boundary region* of the concept X.

Observe, in such definition of approximation we assume *closed world assumption*, i.e., the concept approximation problem is related to objects from the information system $\mathbb{S}$ only. In inductive learning, it is necessary to extend the rough set based approximations for objects outside U. Unfortunately, the lack of generalisation in the process of attribute-based approximations implies that there may exist objects $x \in \mathcal{U} \setminus U$ satisfying $[x]_{IND(B)} \cap U = \varnothing$. Hence, for such objects we are unable to make any decision. In the following sections we discuss some extensions of approximations on supersets of U which are less sensitive to the above mentioned problems.

3 Case-Based Approximations

In case-based reasoning methods, like k-NN (k-Nearest-Neighbour) method, it is necessary to define a distance function between objects $\delta : \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{R}^+$. The problem searching for a relevant distance function for the given data set is not trivial, but at this point, let us assume that such function has been already defined.

In k-NN classification method, the decision for a new object $x \in \mathcal{U} \setminus U$ is made on the basis of the set $NN(x;k) := \{x_1, \ldots, x_k\} \subseteq U$ with k objects from U which are nearest to x with respect to the distance function δ. Usually k is a parameter which can be set up by expert or constructed from experiment data. The k-NN classifiers often use voting algorithm for decision making, i.e. the decision value for new object x can be predicted by: $dec(x) = Voting(\langle n_1, \ldots, n_d \rangle)$ where $ClassDist(NN(x;k)) = \langle n_1, \ldots, n_d \rangle$ is the class distribution of the set $NN(x,k)$ satisfying $n_1 + \ldots + n_d = k$. The voting function can return the most frequent decision value occurring in $NN(x,k)$. In case of imbalanced data, the vector $\langle n_1, \ldots, n_d \rangle$ can be scaled w.r.t. global class distribution first, and after that the voting algorithm can be employed.

Now, we are going to present the rough approximation based on the sets $NN(x;k)$. Let us define a family of functions defined by

$$\mu^{t_1,t_2}_{CLASS_i}(x) = \begin{cases} 1 & \text{if } n_i \geq t_2 \\ \frac{n_i - t_1}{t2 - t1} & \text{if } n_i \in (t_1, t_2) \\ 0 & \text{if } n_i \leq t_1 \end{cases}$$

where $t_1 < t_2 < k$, n_i is the i-th coordinate in the class distribution of $NN(x;k)$. Any such function defines patterns described by means of the following formulas: $\mu^{t_1,t_2}_{CLASS_i}(x) \circ c$, where $\circ \in \{=, \geq, \leq, <, >\}$ and $c \in \{0, 1, \frac{n_i - t_1}{t2 - t1}\}$. One can tune parameters of such formulas to obtain new relevant patterns for the concept approximation on the considered extension of the universe by testing objects.

4 Rule-Based Approximations

Let $\mathbb{S} = (U, A, dec)$ be a decision system. Any implication of the form $(a_{i_1} = v_1) \wedge \ldots \wedge (a_{i_m} = v_m) \Rightarrow (dec = k)$ where $a_{i_j} \in A$ and $v_j \in V_{a_{i_j}}$, is called a *decision rule* for the k-th decision class. Any decision rule $\mathbf{r}$ of the above form can be characterised by following parameters:

- $length(\mathbf{r})$ = the number of descriptors in the premise of $\mathbf{r}$;
- $[\mathbf{r}]$ = *carrier of* $\mathbf{r}$, i.e., the set of objects satisfying the premise of $\mathbf{r}$.
- $support(\mathbf{r})$ = number of objects satisfying the premise of $\mathbf{r}$;
- $confidence(\mathbf{r})$ = the measure of truth of the decision rule = $\frac{card([\mathbf{r}] \cap CLASS_k)}{card([\mathbf{r}])}$.

In data mining, we are interested in searching for *short, strong* decision rules with *high confidence*. The linguistic features like "short", "strong" or "high confidence" of decision rules can be formulated by means of their length, support or confidence. Many decision rule generation methods have been developed using rough set theory (see e.g., [1,2,10,15]).

The rule based classification methods work in three phases:

1. Learning phase: generates a set of decision rules $RULES(\mathbb{S})$ (satisfying some predefined conditions) from a given decision system $\mathbb{S}$.

2. Rule selection phase: selects from $RULES(\mathbb{S})$ the set of such rules that can be supported by x, where $x \in \mathcal{U}$ is a testing object. We denote this set by $MatchRules(\mathbb{S}, x)$.
3. Decision making phase: makes a decision for x using some voting algorithm for decision rules from $MatchRules(\mathbb{S}, x)$

A rule based classifier works as follows. Suppose we would like to decide if a given object $x \in \mathcal{U}$ belongs to the i-th decision class. Let $MatchRules(\mathbb{S}, x) = \mathbf{R}_{yes} \cup \mathbf{R}_{no}$, where $\mathbf{R}_{yes}$ is the set of all decision rules for i-th class matched by x and $\mathbf{R}_{no}$ is the set of decision rules for other classes matched by x. We assign two real values w_{yes}, w_{no} defined by

$$w_{yes} = \sum_{\mathbf{r} \in \mathbf{R}_{yes}} strength(\mathbf{r}) \qquad w_{no} = \sum_{\mathbf{r} \in \mathbf{R}_{no}} strength(\mathbf{r})$$

where w_{yes}, w_{no} are called *for* and *against* weights of the object x, and $strength(\mathbf{r})$ is a normalised function depending on $length(\mathbf{r})$, $support(\mathbf{r})$, $confidence(\mathbf{r})$ and some global information about the decision system $\mathbb{S}$ like decision system size, global class distribution, etc. (see [1]). Using some relationships between w_{yes}, w_{no} the classifier is predicting the decision.

Notice, that in such approach any classifier can be identified with a membership function. We can define rule-based classifiers by a parameterised function $\mu_{CLASS_k}(x)$ of the following form:

IF $\max(w_{yes}, w_{no}) < \omega$ **THEN** $\mu_{CLASS_k}(x) = 0$

ELSE $$\mu_{CLASS_k}(x) = \begin{cases} 1 & \text{if } w_{yes} - w_{no} \geq \theta \\ \frac{\theta + (w_{yes} - w_{no})}{2\theta} & \text{if } |w_{yes} - w_{no}| < \theta \\ 0 & \text{if } w_{yes} - w_{no} \leq -\theta \end{cases}$$

where ω, θ are parameters that allow to search for new relevant patterns (pieces of concept description) for the concept approximation (on the extension of the initial training sample by testing objects).

5 Approximations of Compound Objects

As mentioned before, here we are not only concerned with approximation of concepts that are described with simple attributes but also with higher level concepts established from already existing ones. The idea is to use approximations in the way which gives us the ability to control the level of approximation quality. To simplify notation let us assume that we have two concepts C_1 and C_2 that are given by means of rule-based approximations derived from decision systems $\mathbb{S}_{C_1} = (U, A_{C_1}, dec_{C_1})$ and $\mathbb{S}_{C_2} = (U, A_{C_2}, dec_{C_2})$.

Hence, we are given two sets of patterns for approximation of concepts C_1 and C_2 (see Section 4). They can be obtained by tuning parameters $\{w_{yes}^{C_1}, w_{no}^{C_1}, \omega^{C_1}, \theta^{C_1}\}$ and $\{w_{yes}^{C_2}, w_{no}^{C_2}, \omega^{C_2}, \theta^{C_2}\}$ discussed previously. We want to establish a relevant set of patterns and parameters for the target concept C, i.e. $\{w_{yes}^{C}, w_{no}^{C}, \omega^{C}, \theta^{C}\}$.

The issue is to define a decision system from which we can derive rules determining approximations of C. Let us recall that both simpler concepts C_1, C_2 and target concept C are defined over the same universe $\mathcal{U}$ and are specified on a sample $U \subseteq \mathcal{U}$. To complete the construction of $\mathbb{S}_C = (U, A_C, dec_C)$ we need to specify $A_C \cup \{dec_C\}$. The decision attribute is known for an arbitrary object $x \in U$ and conditional attributes in (the simplest case of) our proposal are either rough memberships for simpler concepts ($A = \{\mu_{C_1}(x), \mu_{C_2}(x)\}$) or weights for simpler concepts ($A = \{w_{yes}^{C_1}, w_{no}^{C_1}, w_{yes}^{C_2}, w_{no}^{C_2}\}$). In the former case we concentrate on the degree of inclusion while in the later case we take into account the relationships with positive and negative information generated during object classification.

The rules used for C construction are making use of attributes that are in fact classifiers themselves. Therefore, it is worth to stratify and interpret attribute domains for attributes in A_C. Instead of using just a value of membership function, weight or their combination, we would prefer to use linguistic statements such as *the likeliness of the occurrence of* C_1 *is low.* Hence, we have to map the attribute value sets onto some limited family of subsets. It is quite natural to introduce (linearly ordered) ranges of values, e.g., $\{negative, low, medium, high, positive\}$, what yields fuzzy-like layout of attributes. One may also consider the case when these subsets overlap. Then, there may be more linguistic values related to attributes. Stratification of attribute values and introduction of linguistic variable attached to the strata provides a way for representing knowledge in more human-readable format since for new object $x^* \in \mathcal{U} \setminus U$ to be classified we may use rules like:

If *compliance of* x^* *with* C_1 *is* *high* *or* *medium* ***and*** *compliance of* x^* *with* C_2 *is* *high* ***then*** $x^* \in C$.

Another advantage of imposing the division of attribute value sets lays in extended control over flexibility and validity of system constructed in this way. We gain the ability of making system more stable and inductively correct and we control the general layout of boundary regions that contribute to construction of the target concept. The process of setting the intervals for attribute values may be performed by hand or with use of automated methods for interval construction e.g., clustering, template analysis, and discretisation. For some discussion of this approach, related to *rough neurocomputing* and *computing with words* see [11].

6 Conclusions and Further Directions

In the paper we have presented a collection of basic ideas that redefine the view on the approximation of concepts in rough set framework. This is very initial work and there is still much more to be done. Some of the techniques mentioned in the paper are already implemented and may be tested (see e.g., [15]).

Still, there are much more issues that accompany the task of proper approximation construction. One of them is related to incremental construction (incremental learning) of approximations. It is a big challenge to devise the method that will be capable of learning an approximation from the sample U

and then, given enriched (finite) sample $U^* \supset U$, extend the approximation with few simple steps, without fundamental reconstruction.

Another interesting topic is the investigation of partially defined rough membership functions. In this paper the set of objects, which are not belonging to domain of rough membership function was treated as a part of boundary region (see Section 4). This solution was caused by the fact that rough set methods are based on 3-valued logics, and we are convinced that it can be improved by introducing 4-value logics [16]. We will dwell on this idea in next papers.

Acknowledgements. This work was supported by KBN grant 8T11C02519 and by the Wallenberg Foundation grant in frame of the WITAS project.

References

1. Bazan J., A comparison of dynamic non-dynamic rough set methods for extracting laws from decision tables. In: [7], pp. 321–365.
2. Grzymała-Busse J., A new version of the rule induction system LERS. Fundamenta Informaticae, Vol. 31(1), 1997, pp. 27–39.
3. Ohrn A., Komorowski J., Skowron A., Synak P., The ROSETTA Software System. In [7] pp. 572–576.
4. Nguyen H.S., Skowron A., Szczuka M., Situation Identification by Unmanned Aerial Vehicle. Proceedings of CS&P 2000, Informatik Berichte, Humboldt-Univerität zu Berlin, Berlin, 2000, pp. 177–188.
5. Pawlak Z., Rough sets: Theoretical aspects of reasoning about data, Kluwer, Dordrecht, 1991.
6. Pawlak Z., Skowron A., Rough membership functions. In Yager R., Fedrizzi M., Kacprzyk J. (eds)., Advances in the Dempster–Shafer Theory of Evidence, Wiley, New York, 1994, pp. 251–271.
7. Polkowski L., Skowron A. (eds.), Rough Sets in Knowledge Discovery vol. 1–2, Physica-Verlag, Heidelberg, 1998.
8. Polkowski L. , Skowron A., Rough mereology: A new paradigm for approximate reasoning. Int. Journal of Approximate Reasoning vol. 15(4), 1996, pp. 333–365.
9. Polkowski L., Skowron A., Towards an adaptive calculus of granules. In Zadeh L. A., Kacprzyk J.(eds.), Computing with Words in Information/Intelligent Systems vol. 1, Physica-Verlag, Heidelberg, 1999, pp. 201–228.
10. Skowron, A., Rauszer, C.: The discernibility matrices and functions in information systems. In Słowiński R. (ed.). Intelligent Decision Support – Handbook of Applications and Advances of the Rough Sets Theory, Kluwer Academic Publishers, Dordrecht, 1992, pp. 311–362.
11. Skowron A., Szczuka M., Approximate reasoning schemes: Classifiers for computing with words. In Proceedings of SMPS 2002, Advances in Soft Computing series, Physica-Verlag, Heidelberg, 2002, pp. 338–345.
12. Stefanowski J., On rough set based approaches to induction of decision rules. In [7] pp. 500–529.
13. Ziarko, W., Rough set as a methodology in Data Mining. In [7] pp. 554–576.
14. The RoboCup Homepage – `www.robocup.org`
15. The RSES Homepage – `logic.mimuw.edu.pl/~rses`
16. Vitoria A., Małuszyński: A logic programming framework for rough sets. LNAI Vol. 2475, Springer-Verlag, Heidelberg, 2002, pp. 205–212.
17. The WITAS Project Homepage – `www.ida.liu.se/ext/witas/`