

Approximation Spaces in Machine Learning and Pattern Recognition

Andrzej Skowron¹, Jarosław Stepaniuk², and Roman Swiniarski^{3,4}

¹ Institute of Mathematics, Warsaw University,
Banacha 2, 02-097 Warsaw, Poland
skowron@mimuw.edu.pl

² Department of Computer Science, Białystok University of Technology,
Wiejska 45a, 15-351 Białystok, Poland
jstepan@ii.pb.bialystok.pl

³ Institute of Computer Science, Polish Academy of Sciences,
Ordona 21, 01-237 Warsaw, Poland

⁴ Department of Computer Science, San Diego State University,
5500 Campanile Drive San Diego, CA 92182, USA
rswiniar@sciences.sdsu.edu

Abstract. Approximation spaces are fundamental for the rough set approach. We discuss their application in machine learning and pattern recognition.

Keywords: Rough sets, approximation spaces, concept approximation.

1 Introduction

Approximation spaces are fundamental structures for the rough set approach [3,9]. In this note we discuss a generalization of approximation spaces. We show how the rough set approach can be used for approximation of concepts assuming that partial information on approximation spaces is available only. We show that searching for relevant approximation spaces is one of the basic task in machine learning and pattern recognition.

2 Approximation Spaces

In this section we recall the definition of an approximation space from [6].

Definition 1. A parameterized approximation space is a system

$AS_{\#, \$} = (U, I_{\#}, \nu_{\$})$, where

- U is a non-empty set of objects,
- $I_{\#} : U \rightarrow P(U)$ is an uncertainty function, where $P(U)$ denotes the power set of U ,
- $\nu_{\$} : P(U) \times P(U) \rightarrow [0, 1]$ is a rough inclusion function,

and $\#, \$$ denote vectors of parameters (the indexes $\#, \$$ will be omitted if it does not lead to misunderstanding).

2.1 Uncertainty Function

The uncertainty function defines for every object x , a set of objects described similarly to x . The set $I(x)$ is called the neighborhood of x (see, e.g., [3,6]).

We assume that the values of uncertainty function are defined using a *sensory environment*, i.e., a pair $(L, \|\cdot\|_U)$, where L is a set of formulas, called the *sensory formulas*, and $\|\cdot\|_U : L \rightarrow P(U)$ is the *sensor semantics*. We assume that for any sensory formula α and any object $x \in U$ the information if $x \in \|\alpha\|_U$ holds is available. The set $\{\alpha : x \in \|\alpha\|_U\}$ is called the *signature of x in AS* and is denoted by $Inf_{AS}(x)$. For any $x \in U$ the set $\mathcal{N}_{AS}(x)$ of *neighborhoods of x in AS* is defined by $\{\|\alpha\|_U : x \in \|\alpha\|_U\}$ and from this set the neighborhood $I(x)$ is constructed. For example, $I(x)$ is defined by selecting an element from the set $\{\|\alpha\|_U : x \in \|\alpha\|_U\}$ or by $I(x) = \bigcap \mathcal{N}_{AS}(x)$. Observe that any sensory environment $(L, \|\cdot\|_U)$ can be treated as a parameter of I from the vector $\#$ (see Def. 1).

Let us consider two examples. Any decision table $DT = (U, A, d)$ [3] defines an approximation space $AS_{DT} = (U, I_A, \nu_{SRI})$, where, as we will see, $I_A(x) = \{y \in U : a(y) = a(x) \text{ for all } a \in A\}$. Any sensory formula is a descriptor, i.e., a formula of the form $a = v$ where $a \in A$ and $v \in V_a$ with the standard semantics $\|a = v\|_U = \{x \in U : a(x) = v\}$. Then, for any $x \in U$ its signature $Inf_{AS_{DT}}(x)$ is equal to $\{a = a(x) : a \in A\}$ and the neighborhood $I_A(x)$ is equal to $\bigcap \mathcal{N}_{AS_{DT}}(x)$. Another example can be obtained assuming that for any $a \in A$ there is given a tolerance relation $\tau_a \subseteq V_a \times V_a$ (see, e.g., [6]). Let $\tau = \{\tau_a\}_{a \in A}$. Then, one can consider a tolerance decision table $DT_\tau = (U, A, d, \tau)$ with tolerance descriptors $a =_{\tau_a} v$ and their semantics $\|a =_{\tau_a} v\|_U = \{x \in U : v\tau_a a(x)\}$. Any such tolerance decision table $DT_\tau = (U, A, d, \tau)$ defines the approximation space $AS_{DT_\tau} = (U, I_A, \nu_{SRI})$ with the signature $Inf_{AS_{DT_\tau}}(x) = \{a =_{\tau_a} a(x) : a \in A\}$ and the neighborhood $I_A(x) = \bigcap \mathcal{N}_{AS_{DT_\tau}}(x)$ for any $x \in U$.

The fusion of $\mathcal{N}_{AS_{DT_\tau}}(x)$ for computing the neighborhood of x can have many different forms, the intersection is only an example. One can also consider some more general uncertainty functions, e.g., with values in $P^2(U)$ [9]. For example, to compute the value $I(x)$ first some subfamilies of $\mathcal{N}_{AS}(x)$ can be selected and next the family consisting of intersection of each such subfamily is taken as the value of $I(x)$.

Note, that any sensory environment $(L, \|\cdot\|_U)$ defines an information system with the universe U of objects. Any row of such an information system for an object x consists of information if $x \in \|\alpha\|_U$ holds, for any sensory formula α . Let us also observe that in our examples we have used a simple sensory language defined by descriptors of the form $a = v$. One can consider a more general approach by taking, instead of the simple structure $(V_a, =)$, some other relational structures R_a with the carrier V_a for $a \in A$ and a signature τ . Then any formula (with one free variable) from a sensory language with the signature τ that is interpreted in R_a defines a subset $V \subseteq V_a$ and induces on the universe of objects a neighborhood consisting of all objects having values of the attribute a in the set V . Note, that this is the basic step in hierarchical modelling [8].

2.2 Rough Inclusion Function

One can consider general constraints which the rough inclusion functions should satisfy. Searching for such constraints initiated investigations resulting in creation and development of rough mereology (see, e.g., [5,4] and the bibliography in [4]). In this subsection we present only some examples of rough inclusion functions.

The rough inclusion function $\nu_{\S} : P(U) \times P(U) \rightarrow [0, 1]$ defines the degree of inclusion of X in Y , where $X, Y \subseteq U$.

In the simplest case it can be defined by (see, e.g., [6,3]):

$$\nu_{SRI}(X, Y) = \begin{cases} \frac{card(X \cap Y)}{card(X)} & \text{if } X \neq \emptyset \\ 1 & \text{if } X = \emptyset. \end{cases}$$

This measure is widely used by the data mining and rough set communities. It is worth mentioning that Jan Lukasiewicz [2] was the first one who used this idea to estimate the probability of implications. However, rough inclusion can have a much more general form than inclusion of sets to a degree (see, e.g., [5,4,9]).

Another example of rough inclusion function ν_t can be defined using the standard rough inclusion and a threshold $t \in (0, 0.5)$ using the following formula:

$$\nu_t(X, Y) = \begin{cases} 1 & \text{if } \nu_{SRI}(X, Y) \geq 1 - t \\ \frac{\nu_{SRI}(X, Y) - t}{1 - 2t} & \text{if } t \leq \nu_{SRI}(X, Y) < 1 - t \\ 0 & \text{if } \nu_{SRI}(X, Y) \leq t. \end{cases}$$

The rough inclusion function ν_t is used in the variable precision rough set approach [10].

Another example of rough inclusion is used for function approximation [9].

Usually, there are several parameters that are tuned in searching for relevant rough inclusion function. Such parameters are listed in the vector $\#$. An example of such parameter is the threshold mentioned for the rough inclusion function used in the variable precision rough set model. We would like to mention some other important parameters. Among them are pairs $(L^*, \|\cdot\|_U^*)$ where L^* is an extension of L and $\|\cdot\|_U^*$ is an extension of $\|\cdot\|_U$, where $(L, \|\cdot\|_U)$ is a sensory environment. For example, if L consists of sensory formulas $a = v$ for $a \in A$ and $v \in V_a$ than one can take as L^* the set of descriptor conjunctions. For rule based classifiers we search in such a set of formulas for relevant patterns for decision classes. We present more detail in the following section.

2.3 Lower and Upper Approximations

The lower and the upper approximations of subsets of U are defined as follows.

Definition 2. For any approximation space $AS_{\#, \S} = (U, I_{\#}, \nu_{\S})$ and any subset $X \subseteq U$, the lower and upper approximations are defined by

$$\begin{aligned} LOW(AS_{\#, \S}, X) &= \{x \in U : \nu_{\S}(I_{\#}(x), X) = 1\}, \\ UPP(AS_{\#, \S}, X) &= \{x \in U : \nu_{\S}(I_{\#}(x), X) > 0\}, \text{ respectively.} \end{aligned}$$

The lower approximation of a set X with respect to the approximation space $AS_{\#,s}$ is the set of all objects, which can be classified with certainty as objects of X with respect to $AS_{\#,s}$. The upper approximation of a set X with respect to the approximation space $AS_{\#,s}$ is the set of all objects which can be possibly classified as objects of X with respect to $AS_{\#,s}$.

Several known approaches to concept approximations can be covered using the discussed here approximation spaces, e.g., the approach given in [3], approximations based on the variable precision rough set model [10] or tolerance (similarity) rough set approximations (see, e.g., [6] and references in [6]).

In machine learning and pattern recognition many classification methods for concept approximation have been developed. They make it possible to decide for a given object if it belongs to the approximated concept or not. The classification methods yield the decisions using only partial information about approximated concepts. This fact is reflected in the rough set approach by assumption that concept approximations should be defined using partial information about approximation spaces only. To decide if a given object belongs to the (lower or upper) approximation of a given concept the rough inclusion function values are needed. In the next section we show how such values necessary for classification making are estimated on the basis of available partial information about approximation spaces.

3 Concept Approximation by Partial Information About Approximation Spaces

In machine learning and pattern recognition [1] we often search for approximation of a concept $C \subset U^*$ in approximation space $AS^* = (U^*, I^*, \nu^*)$ having only a partial information about AS^* and C , i.e., information restricted to a sample $U \subset U^*$. Let us denote the restriction of AS^* to U by $AS = (U, I, \nu)$, i.e., $I(x) = I^*(x) \cap U$, $\nu(X, Y) = \nu^*(X, Y)$ for $x \in U$, $X, Y \subseteq U$.

To decide if a given object x belongs to the lower approximation or the upper approximation of $C \subset U^*$ it is necessary to know the value $\nu^*(I^*(x), C)$. However, in case a partial information about the approximation space AS^* is available only, then an estimation of such a value rather than its exact value can be induced. In machine learning, pattern recognition or data mining different heuristics are used for estimation of the values of ν^* . Using different heuristic strategies values of another function ν' are computed and they are used for estimation of values of ν^* . Then, the function ν' is used for deciding if objects belong to C or not. Hence, we define an approximation of C in the approximation space $AS' = (U^*, I^*, \nu')$ rather than in AS^* . Usually, it is required that the approximations of $C \cap U$ in AS and AS' are close (or the same). If a new portion of objects extending the sample U to U_1 is received, then the closeness of approximations of C in the new approximation space $AS_1 = (U_1, I_1, \nu_1)$ (where I_1, ν_1 are obtained by restriction of I^*, ν^* to U_1) with approximations over AS' restricted to U_1 is verified. If the approximations are not close enough the definition of ν' is modified using new information about the extended sample. In

this way we gradually improve the quality of approximation of C on larger parts of the universe U^* .

Now, we would like to explain in more detail a method for estimation of values $\nu^*(I^*(x), C)$. Let us consider an illustrative example. In the example we follow a method often used in rule based classifiers [9]. The method is based on the following steps. First, a set of patterns, that are used as left hand sides of decision rules, is induced. Each pattern describes a set of objects in U^* with a satisfactory degree of inclusion to one of decision classes (C or $U^* - C$ for the binary decision). Next, for any object the set of all such patterns that are matched to a satisfactory degree by the given object is extracted. Finally, it is applied a conflict resolution strategy (e.g., voting) for resolving conflicts between votes for different decisions by the matched patterns.

We now present an illustrative example to describe this process more formally in the framework of approximation spaces. First, we assume that among parameters of rough inclusion functions are pairs $(PAT, \|\cdot\|_{U^*})$, where PAT is a set of descriptor conjunctions over a set of condition attributes and $\|\cdot\|_{U^*} : PAT \rightarrow P(U^*)$ is the semantics of patterns in U^* . Using such parameters we estimate the value $\nu^*(\|pat\|_{U^*}, C)$ by $\nu(\|pat\|_U, C \cap U)$ for any $pat \in PAT$ and we obtain, for a given threshold $deg \in [0, 1]$, the set S_1 of all patterns pat such that $\nu(\|pat\|_U, C \cap U) \geq deg$, i.e., consisting of patterns “for” the concept C . In an analogous way we obtain the set S_2 of all patterns pat satisfying $\nu^*(\|pat\|_{U^*}, U^* - C) \geq deg$. S_2 consists of patterns “for” the complement $U^* - C$ of the concept C . Next, we estimate $\nu^*(I^*(x), \|pat\|_{U^*})$ for $pat \in S_i$, for $i = 1, 2$. To do this we use our assumption on computing $I^*(x)$ for $x \in U^*$. We assume that the sensory formulas from L are descriptors $a = v$ over the condition attributes from a given set of condition attributes A with the semantics in U^* defined by $\|a = v\|_{U^*} = \{x \in U^* : a(x) = v\}$ for $a \in A$ and $v \in V_a$, where V_a is the value set of a . We also have $I^*(x) = \{y \in U^* : Inf_A(x) = Inf_A(y)\}$, where $Inf_A(x) = \{(a, a(x)) : a \in A\}$. Often, we estimate $\nu^*(I^*(x), \|pat\|_{U^*})$ using a matching strategy based on similarity of the syntactic description of x by $Inf_A(x)$ and the pattern pat . In this way we obtain for a given x the set S'_i of all patterns $pat \in S_i$ (for $i = 1, 2$) such that $\nu(I^*(x) \cap U, \|pat\|_U) \geq deg_1$ where $deg_1 \in [0, 1]$ is a given threshold. Finally, the estimation $\nu'(I^*(x), C)$ of the value $\nu^*(I^*(x), C)$ is obtained by application to the sets S'_1, S'_2 a conflict resolution strategy for resolving conflicts between patterns “for” and “against” the membership of x to C .

Usually, the function ν' is parameterized, e.g., by a threshold to which at least the patterns should be included into the decision classes. Also the discussed sets of patterns are among parameters of ν' tuned in the process of rule based classifier construction. Moreover, matching strategies used for estimation of matching degrees are usually parameterized and such parameters are also among tuned parameters of ν' . In machine learning, pattern recognition and data mining many different searching techniques have been developed for inducing concept approximations of the high quality. Among such components are relevant features, patterns, measures of closeness, model quality measures.

The approximation spaces defined above have been generalized in [9] to approximation spaces consisting of information granules.

4 Conclusions

In the paper we have discussed a role of approximation spaces in the rough set framework for machine learning and pattern recognition. We emphasized that approximation spaces are fundamental objects constructed for concept approximation. In our project we are developing evolutionary strategies searching for relevant approximation spaces for a given ontology [7] of concepts. We also investigate properties of evolutionary strategies for constructing sequences of approximation spaces in adaptive approximation of concepts.

Acknowledgements

The research has been supported by the grants 3 T11C 002 26 and 4 T11C 014 25 from Ministry of Scientific Research and Information Technology of the Republic of Poland.

References

1. Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning*. Springer-Verlag, Heidelberg, 2003.
2. Łukasiewicz, J.: Die logischen Grundlagen der Wahrscheinlichkeitsrechnung, Kraków 1913. In: Borkowski, L. (ed.), *Jan Łukasiewicz - Selected Works*. North Holland, Amsterdam, Polish Scientific Publishers, Warsaw, 1970.
3. Pawlak, Z.: *Rough Sets. Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht, 1991.
4. Polkowski, L.: *Rough Sets: Mathematical Foundations*. Advances in Soft Computing, Physica-Verlag, Heidelberg, 2002.
5. Polkowski, L., Skowron, A.: Rough mereology: A new paradigm for approximate reasoning. *Journal of Approximate Reasoning* **15**(4), 1996, 333–365.
6. Skowron, A., Stepaniuk, J.: Tolerance approximation spaces. *Fundamenta Informaticae* **27**, 1996, 245–253.
7. Staab, S., Studer, R., (Eds.): *Handbook on Ontologies*. International Handbooks on Information Systems, Springer, Heidelberg, 2004.
8. Skowron, A., Synak, P., Complex patterns. *Fundamenta Informaticae* **60**(1-4), 2004, 351–366.
9. Skowron, A., Swiniarski, R., Synak, P: Approximation spaces and information granulation. *Transactions on Rough Sets III: LNCS Journal Subline*, LNCS **3400**, Springer, Heidelberg, 2005, 175–189.
10. Ziarko, W., Variable precision rough set model, *Journal of Computer and System Sciences* **46**, 1993, 39–59.