Rough Neurocomputing Based on Hierarchical Classifiers

Andrzej Skowron[1], Jarosław Stepaniuk[2], and James F. Peters[3]

[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] Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, Manitoba R3T 5V6, Canada
jfpeters@ee.umanitoba.ca

Abstract. In the paper we discuss parameterized approximation spaces relevant for rough neurocomputing. We propose to use standards defined by classifiers in approximate reasoning. In particular, such standards are used for extraction rules of approximate reasoning (called productions) from data and next for deriving approximate reasoning schemes.

1 Introduction

Information sources provide us with granules of information that must be transformed, analyzed and built into structures that support problem-solving. Lotfi A. Zadeh has recently pointed out the need to develop a new research branch called Computing with Words (see, e.g., [13,14,15]). One way to achieve Computing with Words is through rough neurocomputing (see, e.g., [3,4,9,10]) based on granular computing (GC) (see, e.g., [10]) and on rough neural networks performing computations on information granules rather than on vectors of real numbers (i.e., weights). GC is based on information granule calculi [7]. One of the main goals of information granule calculi is to develop algorithmic methods for construction of complex information granules from elementary ones by means of available operations and inclusion (closeness) measures. These constructions can also be interpreted as approximate reasoning schemes (AR-schemes, for short) (see, e.g., [6,8,10]). Such schemes in distributed environments can be extended by adding interfaces created by approximation spaces. They make it possible to induce approximations of concepts (or information about relations among them) exchanged between agents. In the paper, we introduce parameterized approximation spaces as one of the basic concepts of the rough neurocomputing paradigm. Moreover, we propose to use patterns defined by classifier approximations as

J.J. Alpigini et al. (Eds.): RSCTC 2002, LNAI 2475, pp. 316–323, 2002.

standards in approximate reasoning. Such standards are next used for extraction from data rules of approximate reasoning, called productions. AR-schemes can be derived using productions (see, e.g. [3,4,10]).

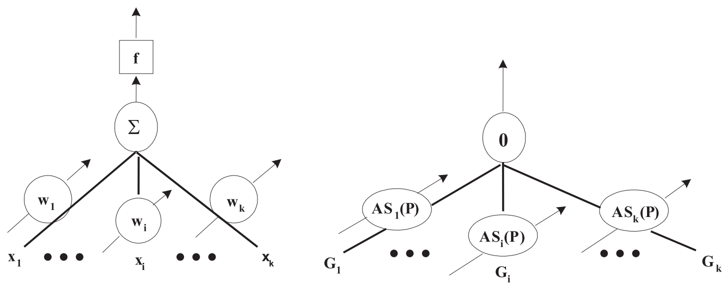

Fig. 1. Classical neuron and rough neuron

A parameterized approximation space can be treated as an analogy to a neural network weight (see Fig. 1). In Fig. 1, $w_1, \cdots, w_k, \Sigma, f$ denote weights, aggregation operator, and activation function of a classical neuron, respectively, while $AS_1(P), \cdots, AS_k(P)$ denote parameterized approximations spaces where agents process input granules $G_1, \cdots, G_k$ and O denotes an operation (usually parameterized) that produces the output of a granular network. The parameters P of approximation spaces should be learned to induce the relevant information granules.

2 Information Granule Systems and Parameterized Approximation Spaces

In this section, we present a basic notion for our approach, i.e., information granule system. Any information granule system is any tuple

$$S = (G, R, Sem) \tag{1}$$

where

1. G is a finite set of parameterized constructs (e.g., formulas) called information granules;
2. R is a finite (parameterized) relational structure;
3. Sem is a semantics of G in R.

For any information granule system two more components are fixed:

1. A finite set H of granule inclusion degrees with a partial order relation $<$ which defines on H a structure used to compare the inclusion degrees; we assume that H consists of the lowest degree 0 and the largest degree 1;

2. A binary relation $\nu_p \subseteq G \times G$ *to be a part to a degree at least* $p \in H$ between information granules from G, called *rough inclusion.* (Instead of $\nu_p(g, g')$ we also write $\nu(g, g') \geq p$.)

Components of an information granules system are parameterized. This means that we deal with parameterized formulas and a parameterized relational system. The parameters are tuned to make it possible to construct finally relevant information granules, i.e., granules satisfying a given specification or/ and some optimization criteria.

There are two kinds of computations on information granules. These are computations on information granule systems and computations on information granules in such systems, respectively. The purpose of the first type of computation is the relevant information granule systems defining parameterized approximation spaces for concept approximations used on different levels of target information granule constructions and the purpose of the second types of computation is to construct information granules over such information granule systems to obtain target information granules, e.g., satisfying a given specification (at least to a satisfactory degree).

Examples of complex granules are tolerance granules created by means of similarity (tolerance) relation between elementary granules, decision rules, sets of decision rules, sets of decision rules with guards, information systems or decision tables (see, e.g., [8], [11], [10]). The most interesting class of information granules are information granules approximating concepts specified in natural language by means of experimental data tables and background knowledge.

One can consider as an example of the set H of granule inclusion degrees the set of binary sequences of a fixed length with the relation ν to be a part defined by the lexicographical order. This degree structure can be used to measure the inclusion degree between granule sequences or to measure the matching degree between granules representing classified objects and granules describing the left hand sides of decision rules in simple classifiers (see, e.g., [9]). However, one can consider more complex degree granules by taking as degree of inclusion of granule g_1 in granule g_2 the granule being a collection of common parts of these two granules g_1 and g_2.

New information granules can be defined by means of operations performed on already constructed information granules. Examples of such operations are set theoretical operations (defined by propositional connectives). However, there are other operations widely used in machine learning or pattern recognition [2] for construction of classifiers. These are the $Match$ and $Conflict\_res$ operations [9]. We will discuss such operations in the following section. It is worthwhile mentioning yet another important class of operations, namely, operations defined by data tables called decision tables [11]. From these decision tables, decision rules specifying operations can be induced. More complex operations on information granules are so called transducers [1]. They have been introduced to use background knowledge (not necessarily in the form of data tables) in construction of new granules. One can consider theories or their clusters as information granules. Reasoning schemes in natural language define the most important class of operations on information granules to be investigated. One of the basic problems for such operations and schemes of reasoning is how to approximate them by available information granules, e.g., constructed from sensor measurements.

In an information granule system, the relation ν_p to be a part to a degree at least p has a special role. It satisfies some additional natural axioms and additionally some axioms of mereology [7]. It can be shown that the rough mereological approach built on the basis of the relation to be a part to a degree generalizes the rough set and fuzzy set approaches. Moreover, such relations can be used to define other basic concepts like closeness of information granules, their semantics, indiscernibility and discernibility of objects, information granule approximation and approximation spaces, perception structure of information granules as well as the notion of ontology approximation. One can observe that the relation to be a part to a degree can be used to define operations on information granules corresponding to generalization of already defined information granules. For details the reader is referred to [4].

Let us finally note that new information granule systems can be defined using already constructed information granule systems. This leads to a hierarchy of information granule systems.

3 Classifiers as Information Granules

An important class of information granules create classifiers. The classifier construction from data table $DT = (U, A, d)$ can be described as follows:

1. First, one can construct granules G_j corresponding to each particular decision $j = 1, \ldots, r$ by taking a collection $\{g_{ij} : i = 1, \ldots, k_j\}$ of left hand sides of decision rules for a given decision.
2. Let E be a set of elementary granules (e.g., defined by conjunction of descriptors) over $IS = (U, A)$. We can now consider a granule denoted by
$$Match(e, G_1, \ldots, G_r)$$
for any elementary granules $e \in E$ described by a collection of coefficients ε_{ij} where $\varepsilon_{ij} = 1$ if the set of objects defined by e in IS is included in the meaning of g_{ij} in IS, i.e., $Sem_{IS}(e) \subseteq Sem_{IS}(g_{ij})$; and 0, otherwise. Hence, the coefficient ε_{ij} is equal to 1 if and only if the granule e matches in IS the granule g_{ij}.
3. Let us now denote by $Conflict\_res$ an operation (resolving conflict between decision rules recognizing elementary granules) defined on granules of the form $Match(e, G_1, \ldots, G_r)$ with values in the set of possible decisions $1, \ldots, r$. Hence,
$$Conflict\_res(Match(e, G_1, \ldots, G_r))$$
is equal to the decision predicted by the classifier
$$Conflict\_res(Match(\bullet, G_1, \ldots, G_r))$$
on the input granule e.

Hence, classifiers are special cases of information granules. Parameters to be tuned are voting strategies, matching strategies of objects against rules as well as other parameters like closeness of granules in the target granule.

The classifier construction is illustrated in Fig. 2 where three sets of decision rules are presented for the decision values $1, 2, 3$, respectively. Hence, we have $r = 3$. In figure to omit too many indices we write α_i instead of g_{i1}, β_i instead of g_{i2}, and γ_i instead of g_{i3}, respectively. Moreover, $\varepsilon_1, \varepsilon_2, \varepsilon_3$, denote $\varepsilon_{1,1}, \varepsilon_{2,1}, \varepsilon_{3,1}$; $\varepsilon_4, \varepsilon_5, \varepsilon_6, \varepsilon_7$ denote $\varepsilon_{1,2}, \varepsilon_{2,2}, \varepsilon_{3,2}, \varepsilon_{4,2}$; and $\varepsilon_8, \varepsilon_9$ denote $\varepsilon_{1,3}, \varepsilon_{2,3}$, respectively.

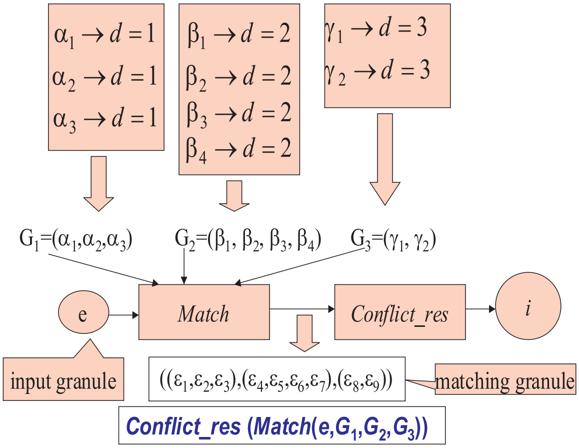

Fig. 2. Classifiers as Information Granules

The reader can now easily describe more complex classifiers by means of information granules. For example, one can consider soft instead of crisp inclusion between elementary information granules representing classified objects and the left hand sides of decision rules or soft matching between recognized objects and left hand sides of decision rules.

4 Approximation Spaces in Rough Neurocomputing

In this section we would like to look more deeply on the structure of approximation spaces in the framework of information granule systems.

Such information granule systems are satisfying some conditions related to their information granules, relational structure as well as semantics. These conditions are the following ones:

1. Semantics consists of two parts, namely relational structure R and its extension R^*.

2. Different types of information granules can be identified: (i) object granules (denoted by x), (ii) neighborhood granules (denoted by n with subscripts), (iii) pattern granules (denoted by pat), and (iv) decision class granules (denoted by c).
3. There are decision class granules $c_1, \cdots, c_r$ with semantics in R^* defined by a partition of object granules into r decision classes. However, only the restrictions of these collections to the object granules from R are given.
4. For any object granule x there is a uniquely defined neighborhood granule n_x.
5. For any class granule c there is constructed a collection granule $\{(pat, p) : \nu_p^R(pat, c)\}$ of pattern granules labeled by maximal degrees to which pat is included in c (in R).
6. For any neighborhood granule n_x there is distinguished a collection granule $\{(pat, p) : \nu_p^R(n_x, pat)\}$ of pattern granules labeled by maximal degrees to which n_x is at least included in pat (in R).
7. There is a class of $Classifier$ functions transforming collection granules (corresponding to a given object x) described in two previous steps into the power-set of $\{1, \cdots, r\}$. One can assume object granules to be the only arguments of $Classifier$ functions if other arguments are fixed.

The classification problem is to find a $Classifier$ function defining a partition of object granules in R^* as close as possible to the partition defined by decision classes.

Any such $Classifier$ defines the lower and the upper approximations of family of decision classes c_i over $i \in I$ where I is a non-empty subset of $\{1, \cdots, r\}$ by

$$\underline{Classifier}(\{c_i\}_{i \in I}) = \{x \in \bigcup_{i \in I} c_i : \emptyset \neq Classifier(x) \subseteq I\}$$

$$\overline{Classifier}(\{c_i\}_{i \in I}) = \{x \in U^* : Classifier(x) \cap I \neq \emptyset\}.$$

The positive region of $Classifier$ is defined by

$$POS\,(Classifier) = \underline{Classifier}(\{c_1\}) \cup \cdots \cup \underline{Classifier}(\{c_r\}).$$

The closeness of the partition defined by the constructed $Classifier$ and the partition in R^* defined by decision classes can be measured, e.g., using ratio of the positive region size of $Classifier$ to the size of the object universe. The quality of $Classifier$ can be defined taking, as usual, only into account objects from $U^* - U$:

$$quality(Classifier) = \frac{card(POS\,(Classifier) \cap (U^* - U))}{card((U^* - U))}.$$

One can consider neural networks as a special case of the above classifiers.

Approximation spaces have many parameters to be tuned to construct the approximation of granules of high quality.

5 Standards, Productions, and *AR*-Schemes

AR-schemes have been proposed as schemes of approximate reasoning in rough neurocomputing (see, e.g., [4,6,9,10]). The main idea is that the deviation of objects from some distinguished information granules, called standards or prototypes, can be controlled in appropriately tuned approximate reasoning. Several possible standard types can be chosen. Some of them are discussed in the literature (see, e.g., [12]). We propose to use standards defined by classifiers. Such standards correspond to lower approximations of decision classes or (definable parts of) boundary regions between them.

Rules for approximate reasoning, called productions, are extracted from data (for details see [4,9,10]). Any production has some premisses and conclusion. In the considered case each premiss and each conclusion consists of a pair $(classifier, standard)$. This idea in hybridization with rough-fuzzy information granules (see, e.g., [10]) seems to be especially interesting. The main reasons are:

- standards are values of classifiers defining approximations of cut differences and boundary regions between cuts [10],
- there is a natural order on such standards defined by classifiers.

We assume productions satisfy a *monotonicity* property. To explain this property, let us consider a production with two premisses:

$$\textbf{if} \quad (C_1, stand_1) \textbf{ and } \quad (C_2, stand_2) \textbf{ then } (C, stand)$$

In this production, classifiers C_1, C_2, C are labelled by standards $stand_1$, $stand_2$, $stand$. The intended meaning of such a production is that if input patterns characterizing a given object are classified at least by C_1 and C_2 to $stand_1$ and $stand_2$, then the composition of such patterns characterizing the object is classified to at least $stand$ by the classifier C.

From productions extracted from data, it is possible to derive productions robust with respect to deviations of premises and from such productions *AR*-schemes (see, e.g., [9,10]).

6 Conclusion

Parameterized Approximation Spaces Are Basic constructs in the rough neurocomputing paradigm. They can be treated as target information granule systems. Such systems are making possible to perform efficient searches for relevant information granules for concepts approximations. The concept of approximation known from rough set theory [5] has been modified to capture approximate reasoning aspects. We have proposed to use in rough neurocomputing standards defined by classifiers. Such standards can next be used in the process of extracting of productions from data and for deriving *AR*-schemes.

Acknowledgements. The research of Andrzej Skowron and Jarosław Stepaniuk has been supported by the State Committee for Scientific Research of the Republic of Poland (KBN) research grants 8 T11C 025 19, 8 T11C 003 23. Moreover, the research of Andrzej Skowron has been partially supported by the Wallenberg Foundation grant. The research of James Peters has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) research grant 185986.

References

1. Doherty, P., Łukaszewicz, W., Skowron, A., Szałas, A.: *Combining rough and crisp knowledge in deductive databases.* (to appear in [4])
2. Mitchell, T.M.: *Machine Learning.* Mc Graw-Hill, Portland (1997)
3. Pal, S.K., Pedrycz, W., Skowron, A., Swiniarski, R. (Eds.): Rough-Neuro Computing. *Neurocomputing: An International Journal* (special issue) **36** (2001)
4. Pal, S.K., Polkowski, L., Skowron, A. (Eds.): *Rough-Neuro Computing: Techniques for Computing with Words.* Springer-Verlag, Berlin (2002) (to appear).
5. Pawlak, Z.: *Rough Sets–Theoretical Aspects of Reasoning about Data.* Kluwer Academic Publishers, Dordrecht (1991)
6. Peters, J.F., Skowron, A., Stepaniuk, J., Ramanna, S.: Towards an ontology of approximate reason. *Fundamenta Informaticae*, **51**(1-2) (2002) (to appear)
7. Polkowski, L., Skowron, A.: Rough mereology: a new paradigm for approximate reasoning. *International J. Approximate Reasoning* **15**(4) (1996) 333–365
8. Polkowski, L., Skowron, A.: Towards adaptive calculus of granules. In [16] (1999) 201–227
9. Polkowski, L., Skowron, A.: Rough-neuro computing. *Lecture Notes in Artificial Intelligence* **2005**, Springer-Verlag, Berlin (2002) 57–64
10. Skowron, A.: Toward intelligent systems: Calculi of information granules. *Bulletin of the International Rough Set Society* **5**(1-2) (2001) 9–30
11. Skowron, A., Stepaniuk, J.: Information granules: Towards foundations of granular computing. *International Journal of Intelligent Systems* **16**(1) (2001) 57–86
12. Skowron, A., Stepaniuk, J.: Information Granules and Rough-Neuro Computing. (to appear in [4])
13. Zadeh, L.A.: Fuzzy logic = computing with words. *IEEE Trans. on Fuzzy Systems* **4** (1996) 103–111
14. Zadeh, L.A.: Toward a theory of fuzzy information granulation and its certainty in human reasoning and fuzzy logic. *Fuzzy Sets and Systems* **90** (1997) 111–127
15. Zadeh, L.A.: A new direction in AI: Toward a computational theory of perceptions. *AI Magazine* **22**(1) (2001) 73–84
16. Zadeh, L.A., Kacprzyk, J. (Eds.): *Computing with Words in Information/Intelligent Systems* **1–2**. Physica-Verlag, Heidelberg (1999)