Hierarchical Information Maps

Andrzej Skowron[1] and Piotr Synak[2]

[1] Institute of Mathematics, Warsaw University,
Banacha 2, 02-097 Warsaw, Poland
[2] Polish-Japanese Institute of Information Technology,
Koszykowa 86, 02-008 Warsaw, Poland

Abstract. We discuss the problems of spatio-temporal reasoning in the context of hierarchical information maps and approximate reasoning networks (AR networks). Hierarchical information maps are used for representation of domain knowledge about objects, their parts, and their dynamical changes. They are constructed out of information maps connected by some spatial relations. Each map describes changes (e.g., in time) of states corresponding to some parts of complex objects. We discuss the details of defining relations between levels of hierarchical information maps as well as between parts satisfying some additional constraints, e.g. spatial ones.

1 Introduction

One of the forms of data representation is an information system, where each investigated object is described by means of some attributes (features). Once some reflexive binary relation on a set of objects is given (e.g., a neighbourhood relation), one can consider new information systems with more complex objects that are clusters (clumps) of objects determined by this relation. In this case, the attributes reflect some more general properties of objects, i.e., properties of sets of objects. This approach is typical for time series analysis, where attributes (features) are defined on the basis of relevant windows [10]. The chosen neighbourhoods and their properties should make it possible to induce the high quality approximations of a given concept. Observe that there are two problems in this approach: discovery of relevant neighbourhoods of objects and their properties. These are key problems of spatio-temporal data mining [3].[1]

In this paper, we extend this approach to the case of information maps and hierarchical information maps, where unstructured objects are substituted by more complex information granules corresponding to structured objects evolving in time. The paper is a continuation of [15,16,7].

We emphasise that in the case of modelling of structured objects the information granulation, in passing from a lower level of a hierarchy (defined by the structure of an object) to a higher one, may be performed, e.g., by indiscernibility or similarity relation. Hierarchical information maps make it possible to model information granules relevant for the target tasks by taking into account the functionality that the information granules should possess.

[1] See [11] for recent issues on modelling of spatio-temporal data.

D. Ślęzak et al. (Eds.): RSFDGrC 2005, LNAI 3641, pp. 622–631, 2005.

2 Preliminaries

In the paper, we use the notation of rough set theory [6,4]. In particular, by $\mathbb{A} = (U, A)$ we denote an *information system* with the universe U of *objects* and the attribute set A. Each *attribute* $a \in A$ is a function $a : U \rightarrow V_a$, where V_a is the *value set* of a. For a given set of attributes $B \subseteq A$, we define the *indiscernibility relation* $IND(B)$ on the universe U that partitions U into classes of indiscernible objects. We say that objects x and y are *indiscernible* with respect to B if and only if $a(x) = a(y)$ for each $a \in B$.

Decision tables are denoted by $\mathbb{A} = (U, A, d)$, where $d \notin A$ is the *decision* attribute. The decision attribute d defines partition of the universe U into *decision classes*. An object x is *inconsistent* if there exists an object y such that $xIND(A)y$, but x and y belong to different decision classes, i.e., $d(x) \neq d(y)$. The *positive region* of a decision table $\mathbb{A}$ (denoted by $POS(\mathbb{A})$) is the set of all consistent objects.

Any pair $(\mathbb{A}, \mathbb{R})$, where $\mathbb{A} = (U, A, d)$ is a decision table and $\mathbb{R}$ is a set of binary and reflexive relations over $U \times U$, is called a *relational decision table*. For any $R \in \mathbb{R}$ by $R(x)$ we denote the *neighbourhood* of an object x, i.e., the set $\{y \in U : xRy\}$. One can consider a new decision table $\mathbb{A}_R = (U_R, A_R, d_R)$ obtained from $(\mathbb{A}, \mathbb{R})$, where $U_R = \{(x, R(x)) : x \in U\}$ is a family of object neighbourhoods, A_R is a set of attributes describing properties of objects and their neighbourhoods, and, e.g., $d_R((x, R(x))) = d(x)$. In this way, one can consider attributes whose values depend on the context in which objects occur, i.e., on neighbourhoods of objects rather than on objects only. This approach is typical for time series analysis, where attributes (features) are defined on the basis of relevant windows [2,1,10]. It is also used in multi-criteria decision making (see, e.g., [17]). The chosen neighbourhoods and their properties should make it possible to induce high quality approximations of a given target concept. Observe that there are two problems in this approach: discovery of relevant neighbourhoods of objects and properties of such neighbourhoods defined by means of some new attributes. The former problem is related to the selection of $\mathbb{R}$ as well as $R \in \mathbb{R}$ for any object, whereas the latter is based on discovery of a relevant language of formulas expressing properties of neighbourhoods and next on the selection of relevant formulas from this language. Discovery of relevant neighbourhoods and their properties for proper object approximation is a key problem of spatio-temporal data mining [3]. From such a decision table there can be derived concept approximation classifiers by using strategies developed in rough sets or other areas like machine learning and pattern recognition.

3 Information Maps

3.1 Basic Definitions

Information maps [14,16] are usually generated from experimental data (e.g., information systems or decision tables) and are defined by some binary (transition) relations on the set of states. In this context a state consists of an information

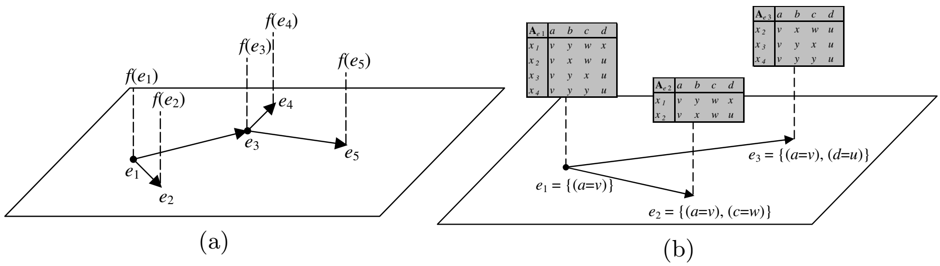

Fig. 1. (a) An information map; (b) An information map of an information system

label and the corresponding information extracted from a given data set. This kind of structure provides basic models over which one can search for relevant patterns for many data mining problems [14,16].

An *information map* $\mathcal{A}$ is a quadruple

$$\mathcal{A} = (E, \leq, I, f), \tag{1}$$

where E is a finite set of *information labels*, $\leq \subseteq E \times E$ is a binary *transition relation* on information labels, I is an *information set* and $f : E \to I$ is an *information function* associating any information label with the corresponding information. In Fig. 1a, we present an example of information map, where $E = \{e_1, e_2, e_3, e_4, e_5\}$, $I = \{f(e_1), f(e_2), f(e_3), f(e_4), f(e_5)\}$, and the transition relation $\leq$ is a partial order on E.

A *state* is any pair $(e, f(e))$, where $e \in E$. The set $\{(e, f(e)) : e \in E\}$ of all states of $\mathcal{A}$ is denoted by $S_{\mathcal{A}}$. The transition relation on information labels can be extended to the relation on states, e.g., in the following way: $(e_1, i_1) \leq (e_2, i_2)$ if and only if $e_1 \leq e_2$. A *path* in $\mathcal{A}$ is any sequence $s_0 s_1 s_2 \ldots$ of states such that $s_i \leq s_{i+1}$ for every $i \geq 0$, and if $s_i \leq s \leq s_{i+1}$ then $s = s_i$ or $s = s_{i+1}$.

3.2 Information Maps of Data Tables

Any information system $\mathbb{A} = (U, A)$ defines its information map as a graph consisting of nodes that are elementary patterns generated by $\mathbb{A}$, where an *elementary pattern* (or *information signature*) $Inf_B(x)$ is a set $\{(a, a(x)) : a \in B\}$ of attribute-value pairs over $B \subseteq A$ consistent with a given object $x \in U$. Thus, the set of labels E is equal to the set $INF(A) = \{Inf_B(x) : x \in U, B \subseteq A\}$ of all elementary patterns of $\mathbb{A}$. The relation $\leq$ is defined in a straightforward way, i.e., for $e_1, e_2 \in INF(A)$, $e_1 \leq e_2$ if and only if $e_1 \subseteq e_2$. Hence, relation $\leq$ is a partial order on E. Finally, the information set I is equal to $\{\mathbb{A}_e : e \in INF(A)\}$, where $\mathbb{A}_e$ is a sub-system of $\mathbb{A}$ with the universe U_e equal to the set $\{x \in U : \forall (a, t) \in e \; a(x) = t\}$. Attributes in $\mathbb{A}_e$ are attributes from $\mathbb{A}$ restricted to U_e. The information function f mapping $INF(A)$ into I is defined by $f(e) = \mathbb{A}_e$ for any $e \in INF(A)$ (see Fig. 1b).

One can consider other information functions for information maps over data tables. Such a function can be a kind of "view" of dependencies in the data table.

Then, for example, $f(e)$ can be equal to the set of all dependencies in $\mathbb{A}_e$ that have sufficient support and confidence.

3.3 Decision Tables over Information Maps

One of the typical schemes of object classification is based on the analysis of decision tables. From the given information about an object (object pattern), we try to classify it relative to a proper decision class. In many cases this scheme needs to be extended because the context of the information should be considered together with the information itself. This means that instead of a single information signature relative to the investigated object x, we also have to examine some other objects that are in some relation to x. Properties of those objects can be important in order to extend information about x by information about the context in which x occurs. In a more complex case, we can consider states of objects and relations between such states. Temporal relations between states, in the case of objects changing in time, provide another possible source of information about the context in which objects occur.

Thus, the scheme of object classification can be as follows. We are given a decision table. Next, it can be extended by some relations on objects (or values of attributes) to a relational decision table defining some neighbourhoods of objects (possibly overlapping each other). Thus, we construct a new decision table, where objects are pairs (*object*, *object_neighbourhood*), and attributes describe properties of the objects in the context of their neighbourhoods.

In the case of information maps, the above idea is generalised to more complex information granules that are pairs (*state*, *state_neighbourhood*), where *state* is a state of a given information map $\mathcal{A}$ and *state_neighbourhood* is the neighbourhood of this state in $\mathcal{A}$. A state can be identified by some information about an object and it determines some set of objects (a sub-table), e.g., set of objects indiscernible by means of some attributes. Thus, *state_neighbourhood* is a much more complex structure than *object_neighbourhood* in the previous case, because it is a set (defined by transition relation) of sub-tables satisfying some constraints. Also the attributes of the constructed decision table are more complex because they express properties of complex neighbourhoods. The decision attribute is complex as well because it classifies a state, which is a complex object (in our example – a sub-table). Thus, for a given state s, we can consider, e.g., the distribution of objects corresponding to s in decision classes as the value of decision for s.

4 Hierarchical Information Maps

4.1 Spatio-temporal Modelling of Objects

Let us discuss in more detail the possibilities of modelling of objects evaluated over time. In the simplest case, we can consider separate series of observations: one series corresponds to one object (see Fig. 2a). Each of the series of observations can be modelled, e.g., by an information map (see Section 3), where labels

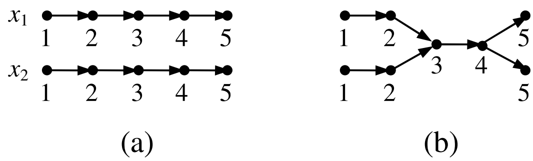

Fig. 2. States of objects evaluating in time

correspond to time indices and information to object signatures (for details see [16]). To make the modelling more general, one can combine different series to one more complex information map by joining those states that carry the same information. In this case, we lose some information about the observed objects, however, the model is more general, hopefully still relevant, and applicable to a potentially larger number of cases (see Fig. 2b).

Another possibility is to construct an information map where states denote all the possible states of observed objects (defined by means of some properties, e.g., "moving car", "stopped car") and the transition relation describes the possible next (previous) states if some temporal relation is additionally satisfied. The main difference here is that the states are not labelled by time indices but by some properties of objects. Thus, the space of states can potentially be reduced to a significant degree.

Yet another case of perceiving objects is when we consider their structure. Structured (complex) objects can consist of some parts constrained by some relations of different nature, e.g., spatial relations. The parts can be built from some simpler parts and therefore the structure can be hierarchical with many different levels. The relation object-part corresponds in most cases to some spatial relation. These problems are considered in rough-mereological approach [9].

The combination of the last two cases, i.e., structured objects evaluating in time, gives spatio-temporal objects. For modelling of such objects we can use hierarchical information maps. Each level of such a map models temporal behaviour of the corresponding parts. The levels are connected by spatial relation, e.g., object-part relation relative to the actual context (state of a complex object and states of its parts) (see Fig. 3). The hierarchical information maps are presented in more detail in the following section.

Especially interesting in modelling of object changes are rules that describe how changes of some features (attributes) influence changes of some other ones. Let us consider an example related to information maps. Assume that with any label e there is associated an information $f(e)$ which is a pair $(T_1(e), T_2(e))$ of theories representing some view on knowledge represented in $\mathbb{A}_e$ consisting of the set of dependencies between conditional and decision attributes in the data table $\mathbb{A}_e$, respectively. Such a view can consist of association rules with sufficient support and confidence. Assume that e' is another label (e.g., an extension of e). Then, one can consider rules making it possible to predict differences between $T_2(e)$ and $T_2(e')$ on the basis of differences between $T_1(e)$ and $T_1(e')$. Such rules are interesting on different levels of hierarchical modelling for spatio-temporal

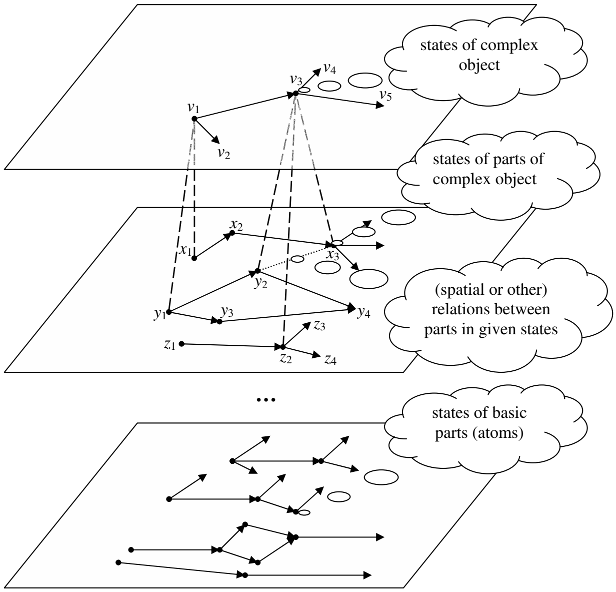

Fig. 3. An example of hierarchical information map

reasoning. Moreover, the laws for predicting changes in decisions quite often require to discover relevant trends of conditional attribute changes (e.g., over some period of time) from data. We plan to develop algorithmic tools for discovery of such laws (dependencies) supported by hierarchical modelling. Observe that in searching for these laws one should, in particular, discover relevant "views" of sets of dependencies and measures of differences.

4.2 Hierarchical Information Maps

One possibility of modelling structured objects evaluated over time is to use some multi-level relational structure. A hierarchical information map is an example of such a structure. It consists of several levels, each modelling temporal behaviour of parts from the same level of the object's structure. Every part of a complex (structured) object defines its own space of states together with the corresponding transitions. Thus, on each level we keep several graphs – one graph for one part. The edges of these graphs are labelled with some temporal relations, however, they are defined for particular parts. The lowest level of the map corresponds to elementary (atomic) parts.

We connect the nodes of graphs from adjacent levels by some spatial relation, defining schemes of constructing a more complex object in a given state from its parts (which are also in some states). An example is presented in Fig. 3. A complex object in state v_1 consists of two parts that are in states x_1 and y_1. The same object in state v_3 consists of three parts in states x_3, y_2, and z_2, respectively. With each non-atomic part in some state x_i at any level, we can associate a decision table containing, e.g., information about historical observations of this part in x_i. The rows (objects) of such a system correspond to different observations.

In a more general case, there can be also given some other relations defined between parts from the same level, e.g., spatial or temporal, reflecting some constraints which should be satisfied by parts in given states in order to reason about more complex object (see Fig. 3). For example, the state of an object can change from safe to unsafe if its parts are in some particular states and, additionally, if they are too close each other. Thus, while modelling complex objects we have to also take into account such relations.

We propose to use labelling of relations linking levels of hierarchical information maps. A label can reflect the fact that some parts satisfy some additional constraint R, or do not satisfy R, or, e.g., do not satisfy any additional constraint at all. In Fig. 4 we can see a part of hierarchical information map where two parts x and y constitute a more complex object $x \oplus y$. There are two additional constraints defined: relations R and S, denoted by dashed and dotted line respectively. From the map it follows that the state of the complex object $x \oplus y$ can depend on the states of parts x and y as well as satisfaction of R or S.

A very important problem is how to check that some complex relation is satisfied or not. Some simple constrains can be checked directly by using some predefined formulas. For example, we can consider a spatial relation "too close" reflecting the fact that two cars are too close each other. Assuming that measurements include location of the cars, we can directly compute the distance and check whether the relation is satisfied or not.

In a more general case we are unable to check satisfiability of relations directly and have to learn it from historical observations of objects. For this purpose, we propose to construct relevant decision tables and to induce classifiers. Let $A = (a_1, \ldots, a_n)$, $B = (b_1, \ldots, b_m)$ be sets of attributes describing parts x and y respectively, and let R be a binary relation that we want to learn. We can construct a decision table $\mathbb{A}_R = (X \times Y, A \cup B, d_R)$, where X and Y are all historical observations of parts x and y respectively; each pair of observations $(x_i, y_j) \in X \times Y$ is described by a vector $(a_1(x_i), \ldots, a_n(x_i), b_1(y_j), \ldots, b_m(y_j))$; and d_R is a binary decision attribute taking value 1 if given observation of

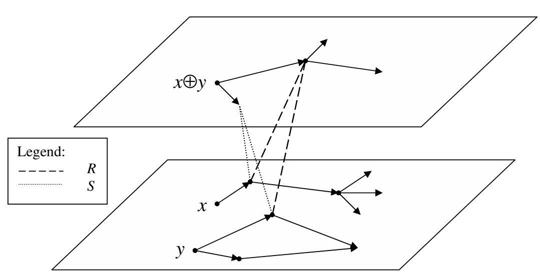

Fig. 4. Satisfaction of spatial or spatio-temporal constraints R and S

parts x and y satisfies the relation R, and 0 otherwise. One can also consider extraction of new features from $A \cup B$ to make the approximation of R more precise. Therefore, for each non-trivial constraint related to parts on a certain level of information map we need to build separate classifier.

In the case of spatio-temporal constraints we may be required to define more complex decision tables for classifier's induction. First of all, we may need to observe a particular object in time, e.g., in some time window. Then, the set of attributes has to be extended by features describing dynamical properties of observed objects. Secondly, new features may have to be extracted. For example, basing on positions of two parts we can extract a new feature describing distance between them by using some specialised metric.

The presented structure – multi-level hierarchical information maps – consists of several information maps that are linked together by some relations on the sets of states. It is important to note that in modelling of such maps we express properties of states and relations between them using the language of domain knowledge (e.g., a simplified natural language). Next, using hierarchical information maps and experimental data one can search for AR networks (see [15,16]), representing relevant patterns for approximation of complex concepts that appear on different levels of maps. Such AR networks are constructed along the derivations performed in domain knowledge using the representation in hierarchical information maps.

4.3 Constructing Higher Levels of Hierarchical Maps by Information Granulation

In this section we discuss an important role which the relational structure granulation [13,8] plays in searching for relevant patterns in approximate reasoning, e.g., approximation patterns (see Fig. 5). For any object x, there is defined a neighbourhood $I(x)$ specified by the value of the uncertainty function from an approximation space (see [12]). From these neighbourhoods some other, more relevant ones (e.g., for the considered concept approximation), should be found. Such neighbourhoods can be extracted by searching in a space of neighbourhoods generated from values of the uncertainty function by applying to them some operations like generalisation operations, set theoretical operations (union, intersection), clustering, and operations on neighbourhoods defined by functions and relations in the underlying relational structure.[2] Fig. 5 illustrates an exemplary scheme of searching for neighbourhoods (patterns, clusters) relevant for concept approximation. In this example f denotes a function with two arguments from the underlying relational structure. Due to the uncertainty, we cannot perceive objects exactly but only by using available neighbourhoods defined by the uncertainty function from an approximation space. Hence, instead of the value $f(x, y)$ for a given pair of objects (x, y), one should consider a family of neighbourhoods $\mathcal{F} = \{I(f(x', y')) : (x', y') \in I(x) \times I(y)\}$. From this family $\mathcal{F}$, a subfamily $\mathcal{F}'$ of neighbourhoods can be chosen which consists of neighbourhoods

[2] Relations from such a structure may define relations between objects or their parts.

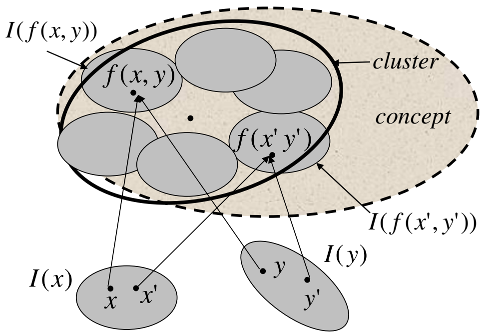

Fig. 5. Relational structure granulation

with some properties relevant for approximation. Next, a subfamily $\mathcal{F}'$ can be, e.g., generalised to clusters that are relevant for the concept approximation, i.e., clusters sufficiently included into the approximated concept (see Fig. 5). The inclusion degrees can be measured by granulation of the inclusion function from the relational structure.

Using information granulation one can construct from a given information map a new one at the higher level which is simpler (more compact) but still sufficient for approximation of complex concepts with a satisfactory quality.

5 Conclusions

In the paper, we have discussed some problems related to hierarchical approximation of spatio-temporal knowledge by means of hierarchical information maps. They can help to discover AR networks representing relevant spatio-temporal patterns from data and soft domain knowledge.

The levels of hierarchical information maps are connected by some spatial or spatio-temporal relations. Satisfaction of different constraints may lead to connecting of the same states from one level to different state in the upper level. We have also discussed the problem of learning such constraints.

Acknowledgements. The research has been supported by the grant 3 T11C 002 26 from Ministry of Scientific Research and Information Technology of the Republic of Poland, and by the Research Center at the Polish-Japanese Institute of Information Technology, Warsaw, Poland.

References

1. J. P. Caraça-Valente and I. López-Chavarrías. Discovering similar patterns in time series. In R. Ramakrishnan, S. Stolfo, R. Bayardo, and I. Parsa, editors, *Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD*, pp. 497–505, Boston, MA, August 20-23 2000. ACM Press.

2. G. Das, K.-I. Lin, H. Mannila, G. Renganathan, and P. Smyth. Rule discovery from time series. In R. Agrawal, P. E. Stolorz, and G. Piatetsky-Shapiro, editors, *Fourth International Conference on Knowledge Discovery and Data Mining KDD*, pp. 16–22, New York, NY, August 27-31 1998. AAAI Press.
3. W. Kloesgen and J. Żytkow, editors. *Handbook of Knowledge Discovery and Data Mining*. Oxford University Press, 2002.
4. J. Komorowski, L. Polkowski, and A. Skowron. Rough sets: A tutorial. In S. K. Pal and A. Skowron, editors, *Rough Fuzzy Hybridization: A New Trend in Decision-Making*, pp. 3–98. Springer-Verlag, Singapore, 1999.
5. S. K. Pal, L. Polkowski, and A. Skowron, editors. *Rough-Neural Computing: Techniques for Computing with Words*. Cognitive Technologies. Springer Verlag, Heidelberg, Germany, 2004.
6. Z. Pawlak. Rough sets. *International Journal of Computer and Information Sciences*, 11:341–356, 1982.
7. J. F. Peters, A. Skowron, J. Stepaniuk, and S. Ramanna. Towards an ontology of approximate reason. *Fundamenta Informaticae*, 51(1-2):157–173, 2002.
8. J. F. Peters, A. Skowron, P. Synak, and S. Ramanna. Rough sets and information granulation. In T. Bilgic, D. Baets, and O. Kaynak, eds., *LNAI* 2715, pp. 370–377, 2003. Springer-Verlag.
9. L. Polkowski and A. Skowron. Rough mereology: A new paradigm for approximate reasoning. *International Journal of Approximate Reasoning*, 15(4):333–365, 1996.
10. J. F. Roddick, K. Hornsby, and M. Spiliopoulou. YABTSSTDMR - yet another bibliography of temporal, spatial and spatio-temporal data mining research. In K. P. Unnikrishnan and R. Uthurusamy, editors, *SIGKDD Temporal Data Mining Workshop*, pp. 167–175, San Francisco, CA, 2001. ACM Press.
11. J. F. Roddick, E. Hoel, M. J. Egenhofer, D. Papadias, and B. Salzberg. Spatial, Temporal and Spatio-Temporal Databases - Hot Issues and Directions for PhD Research. *ACM SIGMOD Record*, 33(2):126–131, 2004.
12. A. Skowron and J. Stepaniuk. Tolerance approximation spaces. *Fundamenta Informaticae*, 27(2-3):245–253, 1996.
13. A. Skowron and J. Stepaniuk. Information granules and rough-neural computing. In Pal et al. [5], pp. 43–84.
14. A. Skowron and P. Synak. Patterns in information maps. In J. Alpigini, J. Peters, A. Skowron, and N. Zhong, eds., *LNAI* 2475, pp. 453–460, 2002. Springer-Verlag.
15. A. Skowron and P. Synak. Complex patterns. *Fundamenta Informaticae*, 60(1-4):351–366, 2004.
16. A. Skowron and P. Synak. Reasoning in information maps. *Fundamenta Informaticae*, 59(2-3):241–259, 2004.
17. R. Słowiński, S. Greco, and B. Matarazzo. Rough set analysis of preference-ordered data. In J. Alpigini, J. Peters, A. Skowron, and N. Zhong, eds., *LNAI* 2475, pp. 44–59, 2002. Springer-Verlag.