Reasoning Based on Information Changes in 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 basic concepts for approximate reasoning about information changes. Any rule for reasoning about information changes specifies how changes of information granules from the rule premise influence changes of information granules from the rule conclusion. Changes in information granules can be measured, e.g., using expressions analogous to derivatives. We illustrate our approach by means of information maps and information granules defined in such maps.

1 Introduction

We discuss basic concepts for reasoning about information changes related, e.g., to changes in time or space [11]. Our approach has roots in rough mereology [12, 9,10]. The basic concepts used are information granules, measures of inclusion and closeness of information granules and rules for reasoning about changes in information granules. Information granules and their relevant changes used in reasoning are assumed to be extracted from the underlying information maps.

In [13] we have shown that patterns defined over information maps are relevant in data mining [1,4]. Any map consists of a transition relation on a set of states, i.e., pairs $(label, information(label))$. Any label describes a context in which information assigned to the label is relevant. Exemplary information maps can be extracted from decision systems [8,6]. In this case one can take attribute–value vectors as labels. The corresponding information is a subsystem of a given decision system consisting of all objects consistent with the label. Patterns over information maps describe sets of states expressible by means of temporal formulas [3,2].

We investigate expressibility of information changes in comparison with the changes of attribute value vectors in a family of descending neighborhoods of given label e_0. We have found an analogy to derivative of information function f, mapping a set of labels onto an information set. Any rule for reasoning about information changes specifies how changes of information granules from the rule premise influence changes of information granules from the rule conclusion. Changes in information granules can be measured, e.g., using expressions analogous to derivatives. We illustrate our approach by means of information maps and information granules defined in such maps. The presented approach can be extended to hierarchical information maps.

G. Wang et al. (Eds.): RSFDGrC 2003, LNAI 2639, pp. 229–236, 2003.

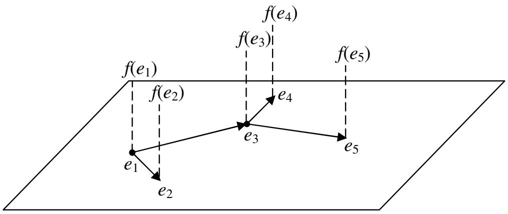

Fig. 1. Information map.

2 Preliminaries

Let us recall notation [6]. In particular $\mathbb{A} = (U, A)$ denotes an *information system* [8] with the universe U of *objects* and the attribute set A. Each *attribute* $a \in A$ is a function $a : U \to V_a$, 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 universe U that splits U into indiscernibility classes, i.e., sets $\{y \in U : a(x) = a(y)$ for any $a \in B\}$ for any $x \in U$. *Decision tables* are denoted by $\mathbb{A} = (U, A, d)$ where d is the *decision* attribute. The decision attribute d defines a 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 it belongs to a different decision class than x, i.e., $d(x) \neq d(y)$. A *positive region* of the decision table $\mathbb{A}$ (denoted by $POS(\mathbb{A})$) is the set of all consistent objects.

3 Information Maps

Information maps [13] are usually generated from experimental data, like information systems or decision tables, and are defined by means of some binary (transition) relations on set of states. Any state consists of information label and information extracted from a given data set corresponding to the information label. We present examples explaining the meaning of information labels, information related to such labels and transition relations (in many cases partial orders) on states. Such structures are basic models over which one can search for relevant patterns for many data mining problems [13].

An *information map* $\mathcal{A}$ is a quadruple $(E, \leq, I, f)$, where E is a finite set of *information labels*, a *transition relation* $\leq \subseteq E \times E$ is a binary relation on information labels, I is an *information set* and $f : E \to I$ is an *information function* associating the corresponding information to any information label. In Figure 1 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 $\leq$ 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 relation on states, e.g., in the following way: $(e_1, i_1) \leq (e_2, i_2)$ iff $e_1 \leq e_2$. A *path* in $\mathcal{A}$ is any sequence $s_0 s_1 s_2 \ldots$ of states, such that (i) $s_i \leq s_{i+1}$ and (ii) if $s_i \leq s \leq s_{i+1}$ then $s = s_i$ or $s = s_{i+1}$, for any $i \geq 0$. A *property* of $\mathcal{A}$

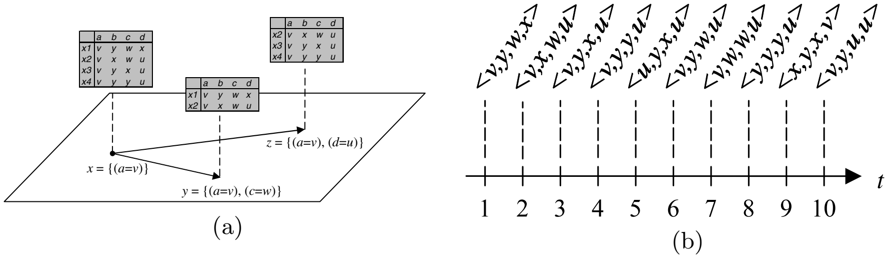

Fig. 2. Information map of (a) information system; (b) temporal information system.

is any subset of $S_{\mathcal{A}}$. Let F be a set of formulas of given language. Property φ is *expressible* in F iff $\varphi = \|\alpha\|$ for some $\alpha \in F$ ($\|\alpha\|$ denotes the semantics of α).

We present two examples of information maps – more can be found in [13]. Any information system $\mathbb{A} = (U, A)$ defines its information map defined by (i) the set of labels $E = INF(A) = \{Inf_B(x) : x \in U, B \subseteq A\}$ where $Inf_B(x) = \{(a, a(x)) : a \in B\}$, (ii) the relation (being a partial order on E) $\leq$ defined by $u \leq v$ iff $u \subseteq v$ for any $u, v \in INF(A)$, (iii) the information set $I = \{\mathbb{A}_v : v \in INF(A)\}$ where $\mathbb{A}_v = (U_v, A_v)$, $U_v = \{x \in U : \forall (a,t) \in v \; a(x) = t\}$ and attributes from A_v are attributes from A restricted to U_v, (iv) the information function f mapping $INF(A)$ into I is defined by $f(v) = \mathbb{A}_v$ for any $v \in INF(A)$. In Figure 2 (a) three information vectors x $= \{(a, v)\}$, y $= \{(a, v), (c, w)\}$ and z $= \{(a, v), (d, u)\}$ are shown satisfying conditions x $\leq$ y, x $\leq$ z. Our second example is related to temporal information systems. A *temporal information system* [14] is an information system $\mathbb{A}$=($\{x_t\}_{t \in E \subseteq \mathbb{N}}$, A) with linearly ordered universe by $x_t \leq x_{t'}$ iff $t \leq t'$. Patterns in such systems are widely studied in data mining (see, e.g., [7,11]). Any temporal information system $\mathbb{A}$ defines in a natural way its information map. The information label set E is the set of all possible time units and let the relation $\leq$ be the natural order on $\mathbb{N}$ restricted to E. The information function f maps any given unit of time t into information corresponding to an object of U related to t, i.e., $f(t) = Inf_A(x_t)$. In this case the map reflects temporal order of attribute value vectors ordered in time. An example of such information map is presented in Figure 2 (b). The transition relation in the first example defines information changes in space (assuming the objects are spatial) while the in second example the transition relation concerns of time changes only. A combination of such transition relations may lead to a spatio-temporal information map.

4 Exemplary Problem

Information maps can be used to formulate numerous problems relevant for data mining. We present one example related to a specific kind of information maps – maps for information systems and decision tables. For more examples the reader is referred to [13]. Solutions of considered problems can be obtained by searching for *good* patterns in a relevant language. Such patterns express (temporal and/or spatial) properties of given information system.

Let us consider an example.

Problem. For a given information map $\mathcal{A}$ of a given information system (or decision table) $\mathbb{A} = (U, A)$, find the minimal element e of E with respect to partial order $\leq$, such that the set of subtables $S(e) = \{f(e') : e \leq e'\}$ satisfies given constraints (expressible in a fixed temporal logic).

One can choose such constraints in the following way. We look for such states that the set of states reachable from them is sufficiently large and has the following property: any two states $s_1 = (e_1, f(e_1))$, $s_2 = (e_2, f(e_2))$ reachable from the state $s = (e, f(e))$ (i.e., $s \leq s_1$ and $s \leq s_2$) consist of decision subtables $f(e_1), f(e_2)$ with close positive regions. The closeness of positive regions can be defined by means of closeness measures of sets. Other possible choices can be made using entropy or association rules parameterised by thresholds (support and confidence [1]) instead of positive regions.

In the new setting one can consider a generalization of association rules [1] by considering implications $\alpha \Rightarrow \beta$ where α, β are some temporal formulas from a fixed temporal logic interpreted in an information map. The support and confidence can be defined in an analogous way as in the case of the standard association rules taking into account the specificity of states in information maps. For example, one can search for such a pattern α of subtables that if a state s is satisfying α then with certainty defined by the confidence coefficient this state has also property β (e.g., β means any path starting from such a state s consists of a subtable with a sufficiently small entropy). At the same time the set of states satisfying α and β should be sufficiently large to a degree defined by the support coefficient. The temporal structure of association rules depends on application.

5 Reasoning Based on Information Changes

Discovery of relevant information changes in spatio-temporal information granules is a challenge. We discuss approximate reasoning based on information changes using terminology from granular computing and rough mereology [12, 9,10]. In case of information maps one can talk about different levels of information granules, e.g., elementary granules like labels, as well as about complex granules being labels' neighborhoods or sets of all labels reachable from a given one by applying transition relation. In all such cases we would like to perform reasoning based on information changes in such granules.

Let $F : G \to G$ where G is a set of granules. For a given granule $e_0 \in G$ we can investigate how values of a function F change compared to changes of an argument, for a sequence $\{e_k\}$ of granules becoming sufficiently close to e_0 if k increases. To compare granules we can use, e.g., a closeness function $cl : G \times G \to [0, 1]$, so the presented idea may be expressed by the formula:

$$\frac{cl(F(e_0), F(e_k))}{cl(e_0, e_k)} \tag{1}$$

If there exists limit of (1) for $k \to \infty$ we can talk about the derivative of F in point e_0. However, in case when G is a finite set such a limit may be trivial and have no practical sense. Therefore, we are looking for minimal k_0 such that for any greater index the value of expression (1) does not change significantly.

In the next section we present an example of application of above schema for information maps: e_0 is an elementary granule, i.e., label; $\{e_k\}_{k=1,2,...}$ is a descending family of neighborhoods of e_0; F is an information function f. Notion of derivative in above sense is important for many applications. For example, its existence may be used to approximate of information changes on states reachable from two close states.

5.1 Derivatives in Information Maps

Let $\mathcal{A} = (E, \leq, I, f)$ be the information map of an information system $\mathbb{A} = (U, A, d)$. Labels from E are elementary patterns from $INF_{\mathbb{A}}(A)$. For a given label e_0, we estimate changes of information function f on a descending sequence of neighborhoods of e_0, compared to the change of its argument (label). Assuming a closeness measure $cl : E \times E \to [0, 1]$ is given we define $k-th$ neighborhood of the label e_0 by

$$N_k(e_0) = \{e : cl(e, e_0) \geq 1 - \frac{1}{k}\}. \tag{2}$$

$cl(e_0, N_k(e_0))$ is defined by means of closeness degrees between e_0 and all labels from $N_k(e_0)$. Analogously, we define neighborhoods of the information corresponding to e_0 as a set of information corresponding to all labels from $N_k(e_0)$:

$$N_k(f(e_0)) = \{f(e) : cl(e, e_0) \geq 1 - \frac{1}{k}\} \tag{3}$$

Finally, the measure of information change with respect to the argument (label) change within some neighborhood $N_k(e_0)$ of label e_0 can be expressed in terms of closeness functions, i.e., by the following formula:

$$\frac{cl(f(e_0), N_k(f(e_0)))}{cl(e_0, N_k(e_0))}. \tag{4}$$

Let us see, that because the number of labels is finite there exists $l \geq k$ such that corresponding neighborhoods are trivial, i.e., $N_l(e_0), N_{l+1}(e_0), \ldots$ are one-element sets $\{e_0\}$. Analogously, neighborhoods defined by $f(e_0)$, i.e., $N_l(f(e_0))$, $N_{l+1}(f(e_0)), \ldots$, are one-element sets $\{f(e_0)\}$. Thus, the limit (in classical sense) of (4), for $k \to \infty$ is 1. Instead, we are looking for the minimal index k_0 such that for all neighborhoods, starting from k_0, if labels are close enough, so is the corresponding information. Because $\{N_k(e_0)\}_{k=1,2,...}$ is a descending family of neighborhoods of e_0 (i.e., $N_1(e_0) \supseteq N_2(e_0) \supseteq \ldots$) we can define the following limit, for some threshold parameter $\delta \geq 0$:

$$\delta - \lim_{k\to\infty} \{N_k(e_0)\}. \tag{5}$$

The neighborhood $N_{k_0}(e_0)$ is a δ-limit of $\{N_k(e_0)\}_{k=1,2,...}$ if and only if k_0 is the smallest number such that for each $k \geq k_0$

$$\left| \frac{cl(f(e_0), N_k(f(e_0)))}{cl(e_0, N_k(e_0))} - \frac{cl(f(e_0), N_{k_0}(f(e_0)))}{cl(e_0, N_{k_0}(e_0))} \right| < \delta. \tag{6}$$

The number $\frac{cl(f(e_0), N_{k_0}(f(e_0)))}{cl(e_0, N_{k_0}(e_0))}$ can be interpreted as derivative of an information function f in point e_0 with precision δ. Such notion can be used, e.g., to measure closeness of states reachable from close sets.

5.2 Examples

Let us consider two examples of possible interpretations of presented ideas. In the first one the closeness between labels and defined label's neighborhood is based on label's syntax. By close labels we understand those described by the same set of attributes assuming the corresponding descriptors are close each other. To talk about a closeness between descriptors we assume they are numerical ones. Let $e_1 = \{(a, a_1), (b, b_1), \ldots\}$, $e_2 = \{(a, a_2), (b, b_2), \ldots\}$ and let $Attr(e)$ be a set of all attributes occurring in label e. Then the closeness between two descriptors based on the same attribute we define by

$$cl((a, a_1), (a, a_2)) = 1 - \frac{|a_1 - a_2|}{|V_a|} \tag{7}$$

where $|V_a|$ is the length of interval being the domain (value set) of attribute a.

Let $Attr(e_1) = Attr(e_2)$. We define closeness between e_1 and e_2 by

$$cl(e_1, e_2) = \frac{1}{card(Attr(e_1))} \cdot \sum_{a \in Attr(e_1)} cl(e_1(a), e_2(a)) \tag{8}$$

where $e(a)$ is a descriptor of label e based on attribute a. In case when $Attr(e_1) \neq Attr(e_2)$ we define closeness of such labels as 0. Presented definition expresses the idea that close labels are based on the same attributes and have close values. That means that close labels are only those from the same level of information map and they form an anti-chain of $\leq$. Also the neighborhoods are defined in this manner.

It is easy to see that the distance based on such a definition of the closeness function, i.e., $d(e_1, e_2) = 1 - cl(e_1, e_2)$, is a metric.

The closeness of a label e_0 and its neighborhood we define as an average closeness between e_0 and all of the labels from $N_k(e_0)$:

$$cl(e_0, N_k(e_0)) = \frac{1}{card(N_k(e_0))} \cdot \sum_{e \in N_k(e_0)} cl(e, e_0). \tag{9}$$

Now, let us define the closeness between elements of an information set. It can be expressed in terms of a chosen property of information systems determined by labels. For example, for a given label e, we can consider the relative size of positive region $POS(\mathbb{A}_e)$ where $\mathbb{A}_e$ is the decision table corresponding to e:

$$cl(f(e), f(e_0)) = 1 - \left| \frac{card(POS(\mathbb{A}_e))}{card(U_e)} - \frac{card(POS(\mathbb{A}_{e_0}))}{card(U_{e_0})} \right| \tag{10}$$

Thus, by close information systems we understand those having close relative positive regions.

The closeness of an information system $\mathbb{A}_{e_0}$ and its neighborhood we define as an average of closeness between $\mathbb{A}_{e_0}$ and all systems from $N_k(f(e_0))$:

$$cl(f(e_0), N_k(f(e_0))) = \frac{1}{card(N_k(e_0))} \cdot \sum_{e \in N_k(e_0)} cl(f(e), f(e_0)). \tag{11}$$

In the second example the closeness between labels is defined on the syntax level as well (see (8)) – so is the definition of label's neighborhood. However, to define the closeness between a label and its neighborhood we use labels' semantics, i.e., corresponding sets of objects. It is worth to notice that if label $e \neq e_0$ belongs to the neighborhood of e_0 then universes of corresponding information systems $\mathbb{A}_e$ and $\mathbb{A}_{e_0}$, respectively, are disjoint.

$$cl(e_0, N_k(e_0)) = 1 - \frac{\sup_{e \in N_k(e_0)} |card(U_e) - card(U_{e_0})|}{\sup_{e \in N_k(e_0)} card(U_e)} \tag{12}$$

The closeness between an information related to a label e_0 and its neighborhood we define again in terms of the size of positive region. However, this time we measure the degree of its change when we extend the universe of an information system $\mathbb{A}_{e_0}$ using universes from the neighborhood:

$$cl(f(e_0), N_k(f(e_0))) =$$

$$\sup_{e \in N_k(e_0)/e_0} \{1 - \left| \frac{card(POS(\mathbb{A}_{e_0} \cup \mathbb{A}_e))}{card(U_{e_0} \cup U_e)} - \frac{card(POS(\mathbb{A}_{e_0}))}{card(U_{e_0})} \right| \} \tag{13}$$

Let us see that now the definition of the derivative of an information function f in point e_0 is defined as quotient where the numerator expresses degree of change of the information (value) and the denominator degree of change of the label (argument).

6 Conclusions and Directions for Further Research

We have discussed foundations for reasoning about information changes. An illustrative example was used to present our approach. The approach can be extended to a more general case assuming changes are measured by means of information granules extracted from underlying relational structure (with many transition relations and functions having spatial, temporal or spatio–temporal nature). Such an approach can be based on hierarchical information maps defined over a given relational structure. Information granules over hierarchical information maps are sets defined by means of formulas specifying properties of relational structures. More complex information granules and closeness measures between them are defined recursively using a given relational structure and previously defined information granules as well as closeness measures [12]. This more general case will be studied in next papers.

Acknowledgements. The research has been supported by the State Committee for Scientific Research of the Republic of Poland (KBN) research grant 8 T11C 025 19. The research of Andrzej Skowron has been also supported by the Wallenberg Foundation grant.

References

1. Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., Verkamo, I.: Fast Discovery of Association Rules, *Proc. of the Advances in Knowledge Discovery and Data Mining.* AAAI Press/The MIT Press, 1996, 307–328.
2. Bolc, L., Szałas, A. (Eds.): *Time and logic*, UCL Press, London 1995.
3. Clark, E.M., Emerson, E.A., Sistla, A.P.: Automatic verification of finite state concurrent systems using temporal logic specifications: A practical approach, *ACM Transactions on Programming Languages and Systems* **8**(2), 1986, 244–263.
4. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (Eds.): *Advances in knowledge discovery and data mining*, AAAI/MIT Press, 1996.
5. Gabbay, D.M., Hogger, C.J., Robinson, J.A. (Eds.): *Handbook of logic in artificial intelligence and logic programming, Epistemic and temporal reasoning* **4**, Oxford University Press, Oxford 1995.
6. Komorowski, J., Pawlak, Z., Polkowski, L., and Skowron, A.: Rough sets: A tutorial. In Pal, S.K., Skowron, A. (Eds.): *Rough fuzzy hybridization: A new trend in decision–making*, Springer-Verlag, Singapore 1999, 3–98.
7. Mannila, H., Toivonen, H.,Verkamo, A.I.: Discovery of frequent episodes in event sequences. *Report C-1997-15*, University of Helsinki, Finland 1997.
8. Pawlak, Z.: *Rough sets – Theoretical aspects of reasoning about data*, Kluwer, 1991.
9. Polkowski, L., Skowron, A.: Rough mereology: a new paradigm for approximate reasoning, *International Journal of Approximate Reasoning*, **15**(4), 1996, 333–365.
10. Polkowski, L., Skowron, A.: Towards an adaptive calculus of granules. In Zadeh, L.A., Kacprzyk, J., (Eds.): *Computing with Words in Information/Intelligent Systems* **1**, Physica-Verlag, Heidelberg 1999, 201–228.
11. Roddick, J. F., Hornsby, K., and Spiliopoulou, M.: YABTSSTDMR - yet another bibliography of temporal, spatial and spatio-temporal data mining research. In Unnikrishnan, K. P. and Uthurusamy, R. (Eds.): *SIGKDD Temporal Data Mining Workshop*, ACM Press, San Francisco 2001, 167–175.
12. Skowron, A.: Toward intelligent systems: Calculi of information granules. *Bulletin of the International Rough Set Society* **5**(1–2), 2001, 9–30.
13. Skowron, A. and Synak, P.: Patterns in information maps. LNAI **2475**, Springer-Verlag, Heidelberg 2002, 453–460.
14. Synak, P.: Temporal templates and analysis of time related data, LNAI **2005**, Springer-Verlag, Heidelberg 2000, 420–427.