Patterns in Information Maps

Andrzej Skowron[1] and Piotr Synak[2]

[1] Institute of Mathematics
Warsaw University
Banacha 2, 02-097 Warsaw, Poland
[2] Cognitron Technology Inc.
87 McCaul Str.
Toronto, ONT, M5T 2W7 Canada

Abstract. We discuss information maps and patterns defined over such maps. Any map is defined by some transition relation on states. Each state is a pair consisting of a label and information related to the label. We present several examples of information maps and patterns in such maps. In particular, temporal patterns investigated in data mining are special cases of such patterns in information maps. We also introduce association rules over information maps. Patterns over such maps can be represented by means of formulas of temporal logics. We discuss examples of problems of extracting such patterns from data.

1 Introduction

Patterns discussed in data mining [1], [5], [11] are special cases of patterns defined over information maps. Any such map consists of a relation on states, i.e., pairs $(label, information(label))$. Exemplary information maps can be extracted from decision systems [10], [7]. In this case one can take attribute value vectors as labels. The information corresponding to any label is a subsystem of a given decision system consisting of all objects consistent with the label. Patterns over information maps describe sets of states and can be expressed by means of temporal formulas [3], [2], [6]. For example, one can look for minimal (with respect to the length) labels with the following property: any state reachable (by means of a given transition relation) from the state s with such label consists of a subsystem of a given decision system with a required property (e.g., expressing that the entropy is not changing significantly in transition to any state reachable from the given state s comparing the entropy of the subsystem corresponding to s). In the paper we present several examples of information maps showing how they can be constructed from data represented, e.g., by information systems, decision tables, or ordered sequences of feature value vectors. Examples of patterns and generalized association rules over such maps are also presented. We discuss searching problems for optimal (in a given information map) association rules. Such rules can be obtained by tuning parameters of parameterized temporal formulas expressing the structure of association rules. Finally, we outline classification problems for states embedded in neighborhoods defined by information maps.

J.J. Alpigini et al. (Eds.): RSCTC 2002, LNAI 2475, pp. 453–460, 2002.

2 Preliminaries

We use in the paper standard notation of rough set theory (see, e.g., [7]). In particular by $\mathbb{A}$ we denote information system [10] with the universe U and the attribute set A. Decision systems are denoted by $\mathbb{A} = (U, A, d)$ where d is the decision attribute.

Temporal formulas can be used for expressing properties of states in information maps. We recall an example of temporal logic syntax and semantics [3], [2]. We begin from syntax of formulas. Let Var be a set of propositional variables. One can distinguish two kinds of formulas: state and path, and we define the set F of all formulas inductively:

S1. Every propositional variable from Var is a state formula,
S2. if α and β are state formulas, so are $\neg\alpha$ and $\alpha \wedge \beta$,
S3. $\mathsf{A}\alpha$ is a state formula, if α is a path formula,
P1. any state formula is also a path formula,
P2. if α and β are path formulas, so are $\neg\alpha$ and $\alpha \wedge \beta$,
P3. if α and β are path formulas, so is $\mathsf{U}(\alpha, \beta)$.

Both A and U are modal operators, where the former denotes "all paths" and the latter "until".

The semantics of temporal formulas is defined as follows. Assume S is a given set of *states* and let $R \subseteq S \times S$ be a *transition relation*. The pair (S, R) is called a *model*. Let Val be a *valuation function* $Val : Var \to 2^S$. Satisfaction of state formula $\alpha \in F$ in state $s \in S$ we denote by $s \models \alpha$ and define by

S1. $s \models \alpha$ iff $s \in Val(\alpha)$ for $\alpha \in Var$,
S2. $s \models (\neg\alpha)$ iff not $s \models \alpha$, $s \models \alpha \wedge \beta$ iff $s \models \alpha$ and $s \models \beta$,
S3. $s \models \mathsf{A}\alpha$ iff $\pi \models \alpha$ for every path π starting at s,

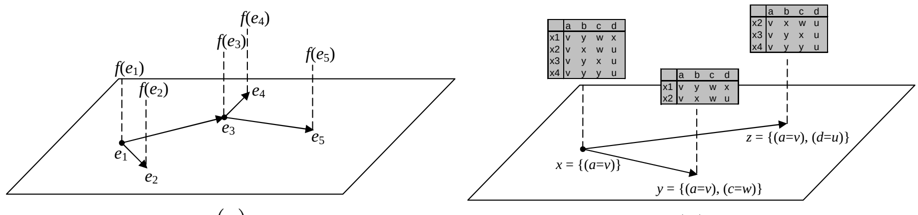
Let $\pi = s_0 s_1 s_2 \ldots$ be a path and let π_i denote the suffix $s_i s_{i+1} s_{i+2} \ldots$ of π. Satisfaction of path formula $\alpha \in F$ in path π we denote by $\pi \models \alpha$ and define by

P1. $\pi \models \alpha$ iff $s_0 \models \alpha$ for any state formula α,
P2. $\pi \models (\neg\alpha)$ iff not $\pi \models \alpha$, $\pi \models \alpha \wedge \beta$ iff $\pi \models \alpha$ and $\pi \models \beta$,
P3. $\pi \models \mathsf{U}(\alpha, \beta)$ iff $\pi_i \models \alpha$ and $\pi_j \models \beta$ for some $j \geq 0$ and all $0 \leq i < j$.

The semantics $\|\alpha\|_M$ (or $\|\alpha\|$, for short) of formula α in the model M is a set $\{s : s \models \alpha\}$ of all states in M in which α is satisfied.

There are numerous other temporal logic studied in literature [2], [6]. They are characterized by temporal operators defined on the basis of the introduced above operators or some new ones. Let us consider such exemplary past operator H_k^l (where $l \leq k$) with the following intended meaning: $s \models H_k^l\alpha$ if and only if for any path $\pi = s_1 \ldots s_m$ where $m \leq k$ and $s_m = s$ any l-window of π (i.e., subsequence $s_i s_{i+1} \ldots s_{i+l}$ of π) satisfies α. Analogously, one can define operator G_k^l related to the states reachable from s.

One can apply temporal formulas, like those defined above, to describe patterns expressing properties of states in information maps. Problems discussed in the paper are different from typical model checking problems. We consider



(a) (b)

Fig. 1. (a) Information map, (b) Information map of information system

problems assuming that a model (represented by means of information map) is given. Such models are generated from given information sources. The problems we consider are related to searching for patterns (which can be expressed by means of temporal formulas from a given set) having *required properties in a given model*. Examples presented in the following sections are explaining what we mean by the phrase *required properties in a given model*.

3 Information Maps

In this section we introduce the notion of an information map. Such maps 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. Presented examples explain the meaning of information labels, information related to such labels and transition relations (in many cases partial orders) on states. We show that such structures are basic models over which one can search for relevant patterns for many data mining problems.

An *information map* $\mathcal{A}$ is a quadruple $(E, \leq, I, f)$, where E is a finite set of *information labels*, *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 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 is extended to relation on states: $(e_1, i_1) \leq (e_2, i_2)$ 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 for every $i \geq 0$: (1) $s_i \leq s_{i+1}$; (2) if $s_i \leq s \leq s_{i+1}$ then $s = s_i$ or $s = s_{i+1}$.

A *property* of $\mathcal{A}$ is any subset of $S_{\mathcal{A}}$. Let F be a set of temporal formulas. We say that property φ is *expressible* in F if and only if $\varphi = \|\alpha\|$ for some $\alpha \in F$.

Any information system $\mathbb{A} = (U, A)$ defines its information map as a graph consisting of nodes being elementary patterns generated by $\mathbb{A}$, where an *elemen-*

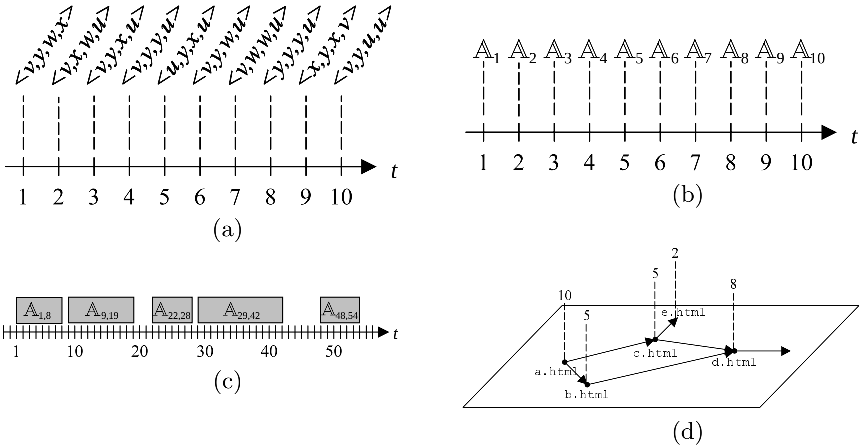


Fig. 2. Information map of (a) temporal information system, (b) information systems changing in time, (c) temporal patterns of temporal information system, (d) web pages

tary pattern $Inf_B(x)$ 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 then defined in a straightforward way, i.e. for $u, v \in INF(A)$ $u \leq v$ iff $u \subseteq v$. Hence, relation $\leq$ is a partial order on E. Finally, the information set I is equal to $\{\mathbb{A}_v : v \in INF(A)\}$ where $\mathbb{A}_v$ is a sub-system of $\mathbb{A}$ with the universe U_v equal to the set $\{x \in U : \forall (a,t) \in v$ we have $a(x) = t\}$. Attributes in $\mathbb{A}_v$ are attributes from $\mathbb{A}$ restricted to U_v. The information function f mapping $INF(A)$ into I is defined by $f(v) = \mathbb{A}_v$ for any $v \in INF(A)$. An example is presented in Figure 1b where three information vectors $x = \{(a, v)\}$, $y = \{(a, v), (c, w)\}$ and $z = \{(a, v), (d, u)\}$ are shown such that $x \leq y$, $x \leq z$.

One can investigate several properties of such system, e.g. related to distribution of values of some attribute. Let α_v be a formula, such that $(e, \mathbb{A}_e) \models \alpha_v$ has the following intended meaning: "at least 75% of objects of system $\mathbb{A}_e$ has value v on attribute d". In our example $\|\alpha\|_{\mathbb{A}} = \{(x, \mathbb{A}_x), (z, \mathbb{A}_z)\}$.

A *temporal information system* [13] is a 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., [8], [11]).

Any temporal information system $\mathbb{A}$ defines in a natural way its information map. Let the information label set E be 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 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 2a.

Let $\mathcal{F} = \{\mathbb{A}_t\}_{t \in E \subseteq \mathbb{N}}$ be a family of decision systems ordered in time. Assume $\mathbb{A}_t = (U_t, A_t, d_t)$, $A_t = \{a_1^{(t)}, \ldots, a_k^{(t)}\}$, and $U_t \subseteq U_{t'}$ for $t \leq t'$. Moreover, let $a_i^{(t)}(u) = a_i^{(t')}(u)$ and $d_t(u) = d_{t'}(u)$ for $i = 1, \ldots, k$, $t \leq t'$ and $u \in U_t$.

Such a family of decision systems defines the following $\mathcal{F}$-information map. E is the set of time stamps and each $\mathbb{A}_t$ corresponds to state of information system (knowledge) at a given time stamp. The transition relation $\leq$ is naturally derived from the order on $\mathbb{N}$ and the function f is a map from E onto $\mathcal{F}$. An example is presented in Figure 2b. For properties of such systems see [15].

One can consider properties of such information map invariant in time. Let us recall the decomposition problem of information systems changing in time [14]. Then the *time stability* property of a given cut c in $\mathcal{F}$-information map can be expressed by the following state expression: "c is the *optimal* cut for $\mathbb{A}_t$" for any $t \leq t_o$ where t_o is a given time threshold. A cut c on a given attribute is optimal for $\mathbb{A}_t$ if and only if it discerns the maximal number object pairs from different decision classes in $\mathbb{A}_t$.

Let $\mathbb{A} = (\{x_t\}_{t \in E \subseteq \mathbb{N}}, A)$ be a temporal information system. Let $\mathcal{F}_\tau$ be a family $\{\mathbb{A}_{t_i, t_i + \Delta_i}\}_{i=1,\ldots,k}$ of information systems ordered in time, such that $\mathbb{A}_{t_i, t_i+\Delta_i} = (\{x_{t_i}, x_{t_i+1}, \ldots, x_{t_i+\Delta_i}\}, A)$. Moreover, assume for any i a generalized template (see e.g. [9]) T_i of the quality at least τ occurs in the table $\mathbb{A}_{t_i, t_i+\Delta_i}$. Let $\leq$ be the natural order on $\mathbb{N}$ restricted to E. Finally, function f maps E onto $I{=}\mathcal{F}_\tau \cup \{\emptyset\}$ and is defined by (see Figure 2c):

$$f(t) = \begin{cases} \mathbb{A}_{t_i, t_i+\Delta_i} & \text{if exists } i \text{ such that } t_i \leq t \leq t_i + \Delta_i \\ \emptyset & \text{otherwise} \end{cases}$$

Any template in each of the systems $\mathbb{A}_{t_i, t_i+\Delta_i}$ is a pattern that has its occurrence time and the validity period. One can consider a collection of templates and characterize their relative occurrence in time by higher order patterns. For example, relative occurrence of templates T_1, T_2 can be of the form "the pattern T_1 is always followed by the pattern T_2" in a given information map [13].

4 Exemplary Problems

The notion of information map looks quite abstract, however, using maps one can formulate numerous problems relevant for data mining. In this section we formulate some examples. In general, solution of a given problem is based on searching for *good* patterns in relevant (temporal) language. Such patterns express (temporal) properties of given information system.

Let $\mathcal{A} = (E, \leq, I, f)$ be na information map of a given information system (or decision table). Let us observe that states in $\mathcal{A}$ consist of subtables defined by labels (elementary patterns) from a given information system. Hence, patterns (temporal formulas) describe properties of sets of subtables pointed by elementary patterns. Let us consider an example.

Problem 1. For information map $\mathcal{A}$ of a given decision table, find minimal, with respect to the partial order $\leq$, element e of E 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 are looking 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 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* relative positive regions [7]. The closeness of relative positive regions can be defined by means of closeness measures of sets. Other possible choices can be made using entropy or association rules parameterized by some values of 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 of the form $\alpha \Rightarrow \beta$ where α, β are some temporal formulas from a fixed (fragment of) 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. Let us consider an example of searching problem assuming an information map is given. The goal is to search for such 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., any path starting from such state s consists of a subtable with 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 the application (see below).

One can generalize a prediction problem considered for linear orders in data mining [5], [11]. The following example is included to illustrate this point of view.

Problem 2. Let $\mathcal{A} = (E, \leq, I, f)$ be an information map. Assume the relation $\leq$ is a partial order. Let us consider a searching problem for association rules of the form $H_k^l \alpha \Longrightarrow \circ\beta$ where $\circ$ is the next operator [2], β is a fixed formula and α is from a given set of formulas.

The intended meaning of such formula is the following: $s \models H_k^l \alpha \Longrightarrow \circ\beta$ if and only if any immediate successor of s has property β if the neighborhood of s defined by the operator H_k^l has property specified by α. Let us observe that the space of formulas from which we would like to extract relevant association rules is parameterized by non-negative integers k, l satisfying $l \leq k$ as well as α from a fixed set of temporal formulas.

For a given state s in an information map two neighborhoods are defined. The first called the *past neighborhood* $P(s)$ of s consists of some states from which s is reachable (by the transition relation of information map). The second one, the *future neighborhood* $F(s)$, consists of some states reachable from s. Let γ be a given formula representing a specified property of $F(s)$. For example, it can be of the form $G_n^2 \delta$ where δ is expressing that in any 2-window of states from $F(s)$ changes are not significant. Moreover, let L be a set of formulas expressing properties of 2-windows in $P(s)$. L can be chosen as a set of formulas specifying trends of changes of attribute values in labels of 2-windows.

Problem 3. For given δ, L, n and information map $\mathcal{A}$ find k and formula $\beta \in L$ such that $H_k^2 \beta \Longrightarrow G_n^2 \delta$ is an association rule of the required quality specified by the support and confidence coefficients.

An interesting task is to search for so called labels of changes. The 2-windows in past neighborhoods of states with such labels are showing significant changes while 2-windows in future neighborhoods are not showing significant changes of information.

One can consider more general structures than neighborhoods, for representing trends of pattern changes. We use terminology from granular computing (see, e.g., [12]).

An interesting class of rules (see also [4]) is defined by the following scheme of association rules:

$$\alpha(G) \wedge R(F_1(G), F_2(G)) \wedge \beta(F_1(G)) \Longrightarrow \gamma(F_2(G))$$

where $G, F_1(G), F_2(G)$ are information granules; $F_1(G), F_2(G)$ are parts of G; α, β, γ are given properties of granules and R is a binary constraint relation on granules.

Such rules have the following intended meaning. If the granule G has property α and its parts $F_1(G), F_2(G)$ are satisfying a given constraint R and $F_1(G)$ has property described by β then $F_2(G)$ satisfies γ.

Let us consider an example. Assume G is a sequence of information granules $(g_1, \ldots, g_k)$ being a sequence of states $((e_1, f(e_1)), \ldots, (e_k, f(e_k)))$, respectively in a map $\mathcal{A}$ of information system $\mathbb{A}$. Moreover, let $F_1(G), F_2(G)$ be granules $(e_1, \ldots, e_k)$, $(f(e_1), \ldots, f(e_k))$, respectively. The intended meaning of formulas α, β, γ, R are as follows:

- $\alpha(G)$ holds iff $\{g_1, \ldots, g_k\}$ is an anti-chain in $(E, \leq)$ in which all states are reachable from some state;
- $R(G, G')$ iff (G, G') is a state in $\mathcal{A}$;
- $\beta \in L$ where L is a set of formulas making possible to check if attribute values in $(e_1, \ldots, e_k)$ have some trends of changes, e.g., they are increasing, decreasing, changing significantly;
- $\gamma(f(e_1), \ldots, f(e_k))$ holds iff the cardinalities of relative positive regions of decision tables from $(f(e_1), \ldots, f(e_k))$ create a decreasing sequence.

Certainly, one can consider other semantics of the above formulas, for example, the following intended meaning of the formula γ: $\gamma(f(e_1), \ldots, f(e_k))$ if and only if a sequence of the most probable decision values of decision tables from $(f(e_1), \ldots, f(e_k))$ is increasing.

Finally, we obtain the following searching problem.

Problem 4. Assume $\mathcal{A}$ be an information map and consider the set of anti-chains each of them reachable from one state in $\mathcal{A}$. Find, for a given formula γ and a set of temporal formulas L, a formula $\beta \in L$ predicting trends of changing in $(f(e_1), \ldots, f(e_k))$ expressible by γ by means of trends of changes of attribute values in $(e_1, \ldots, e_k)$ expressible by β. The quality of extracted association rules can be measured by the required support and confidence coefficients.

5 Conclusions

We have introduced a class of patterns in information maps. Presented examples show that searching for such patterns can be important for data mining problems. In the future we plan to develop heuristics searching for solutions of the discussed problems and we would like to apply them to real-life problems. Boolean reasoning methods [7] can be used for extracting at least some of considered patterns. The details of methods based on Boolean reasoning for the temporal pattern extraction will be included in one of our next papers.

Acknowledgements. The research has been supported by the 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, CA (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. Dubois, D., Prade, H., Ughetto, L.: A new perspective on reasoning with fuzzy rules. In: Pal, N.R., Sugeno, M.: Lecture Notes in Artificial Intelligence **2275**, Springer-Verlag, Berlin (2002) 1–11
5. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (Eds.): *Advances in knowledge discovery and data mining.* AAAI/MIT Press (1996)
6. Gabbay, D.M., Hogger, C.J., Robinson, J.A. (eds.): *Handbook of logic in artificial intelligence and logic programming, Epistemic and temporal reasoning.* (vol. 4), Oxford University Press, Oxford (1995)
7. Komorowski, J., Pawlak, Z., Polkowski, L., 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
8. Mannila, H., Toivonen, H.,Verkamo, A.I.: Discovery of frequent episodes in event sequences. *Report C-1997-15*, University of Helsinki, Finland (1997)
9. Nguyen S.H.: Regularity Analysis And Its Applications In Data Mining. PhD Dissertation, Warsaw University, Poland (2000)
10. Pawlak, Z.: *Rough sets – Theoretical aspects of reasoning about data*, Kluwer Academic Publishers, Dordrecht (1991)
11. Roddick, J.F., Hornsby, K., Spiliopoulou, M.: An updated bibliography of temporal, spatial and spatio–temporal data mining research. Lecture Notes in Artificial Intelligence **2007**, Springer-Verlag, Berlin (2001) 147–163
12. Skowron, A.: Toward intelligent systems: Calculi of information granules. *Bulletin of the International Rough Set Society* **5**(1–2) (2001) 9–30
13. Synak, P.: Temporal templates and analysis of time related data. In: Ziarko, W. , Yao, Y.Y. (eds): Proc. of The Second International Conference on Rough Sets & Current Trends in Computing (RSCTC'00), Banff Kanada (2000)
14. Synak, P.: Decomposition of large databases growing in time. Proc. of Eigth International Conference on Information Processing and Management of Uncertainty on Knowledge Based Systems (IPMU'2000), Madrid Spain (2000) 234–239
15. Wakulicz-Deja, A.: Clasification of time-varying information systems. *Information Systems* (1991)