Paper

Event mining based on observations of the system

Janusz Granat

Abstract—Event mining is becoming a challenging area of research. Events in system analysis is not a new concept. It has been used in Petri nets, stochastic modeling, etc. However, there are new opportunities that come from the lage amount of data that is stored in various databases. In this paper we will focus on formulating the event mining tasks that consider observations of the system as well as internal and external events.

***Keywords—** event mining, temporal data mining, telecommunications.*

1. Introduction

Data mining have many industrial and scientific applications. However, the existing algorithms consider limited information about the events. Recently, an increased importance of events in modeling and understanding complex systems can be observed. D. Luckham [5] provides us with a framework for thinking about complex events and for designing systems that use such events (see also [6]). The event mining is new and challenging area of research and applications. Events are especially challenging for real-time analysis. Gartner Inc., a technology research and advisory firm, defines real-time as the complete compression of lag between the detection of an event, the reporting of that event, the decision-making, and the response. They further observe that the real-time enterprise (RTE) is an enterprise that competes by using up-to-date information to progressively remove delays to the management and execution of its critical business processes. Therefore, real-time computing might be the focal point of IT departments because it allows companies to provide on-line information for effective decision making.

A key to understanding events is knowing what caused them and having that knowledge at the time the events happen. Another issue is the knowledge about the consequences of events. The ability to track event and consequences is an essential step toward on-line decision support and an important challenge for new algorithms for event mining.

Many existing enterprise systems are distributed and event-driven. Events might be described by structured and unstructured information. The structured information is well recognized and is stored in databases. However, the organizations are working on improvement of the analysis of the external environment and influence of this environment on the performance of the organization. Environmental scanning is a new term and it means the acquisition and use of the information about events, trends, and relationships in an external environment. Therefore, the methods of dealing with unstructured information about events are especially important. The example of an event detection from online news documents is presented in [7].

Event mining might have various applications. On the business level it is the business activity monitoring (BAM). BAM is defined (by Gartner Inc.) as a concept that provides a real-time access to critical business performance indicators to improve the speed and effectiveness of business operations. BAM involves alerts, triggers, sensors, and agents that determine a transaction or event that is meaningful. Another group of applications are on the level of the network infrastructure of the company. Computer networks produce a large amount of event-based data that can be collected for network analysis. These data include alerts from firewalls and intrusion detection systems (IDS), log files of various software systems, routing information from the Internet and so on. An example of use of the concept of events for network analysis can be found in [4].

2. Basic definitions

We have a system which is influenced by internal and external events and the behavior of the system can be monitored (see Fig. 1).

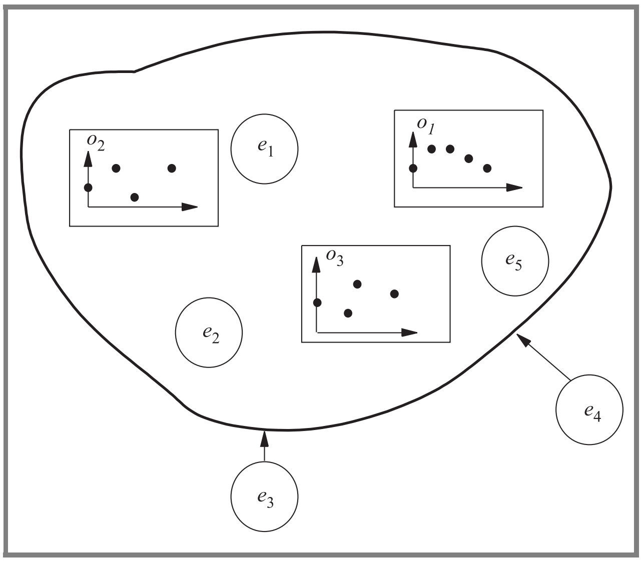

Fig. 1. The events and observations.

We define here the basic terms:

- an event,
- observations,
- observations pattern,
- event mining models.

2.1. An event

Definition 1: An event e_i is something that hapen in the system or its enviroment and can be described by a set of parameters.
Let us consider the finite set of events

$$\mathcal{E} = \{e_1, e_2, \dots, e_n\}.$$

We can distinguish past and future events:
The set of past events:

$$\mathcal{E}_p = \{e_1, e_2, \dots, e_e\}.$$

The set of future events:

$$\mathcal{E}_f = \{e_{e+1}, e_2, \dots, e_n\},$$

$$\mathcal{E} = \mathcal{E}_p \cup \mathcal{E}_f.$$

A formal definition of the event.

Definition 2: Let $A_{e_i} = \{a_{1,e_i}, a_{2,e_i} \dots, a_{m,e_i}\}$ be the set of attributes for an event e_i and $V_{a_{j,e_i}}$ be the domain of attribute $a_{j,e_i} \in A_{e_i}$. An event is defined as $(m+2)-$ tuple $(a_{1,e_i}, a_{2,e_i} \dots, a_{m_{e_i},e_i}, t, \Delta t)$, where $a_{j,e_i} \in V_{a_{j,e_i}}$, t is the time of occurence of the event, Δt is the duration of the event.

If $\Delta t = 0$ then then we have a point event and for $\Delta t \neq 0$ we have an interval event. Each event might have different number of attributes and attributes of different types (numerical, textual, etc.) Attributes for interval event might change over time. We assume in this paper that attributes do not change at the time interval Δt.

2.2. Observations

Events might influence the behavior of the system.
An information system $\mathcal{O}$ of a set of observations can be defined as follows:

$$\mathcal{O} = (O, V, \rho, T, R), \qquad (1)$$

where: T – is a nonempty set whose elements t are called moments of time, R – is an order on the set T (here we assume linear order), O – is finite and nonempty set of observations, $V = \bigcup_{o \in O} V_o$, V_o is the set of values of observation $o \in O$, called the domain of o, ρ – is an information function: $\rho : O \times T \to V$.

We assume that we will have the set of observations:

$$\mathbf{O} = (o_1, o_2, \dots, o_l), \qquad (2)$$

$$T = (t_1, t_2, \dots, t_n).$$

For simplicity instead of $\rho(o_i, t)$ we will use $o_{i,t}$.

2.3. Observations pattern

Observations pattern p_i is a distinguishable sequence of observations:

$$(\hat{o}_1(t), \hat{o}_2(t), \dots, \hat{o}_n(t)) \Rightarrow p_i \quad t \in \Delta t.$$

Pattern p_i might be described by a set of parameters, etc.
We can have set of patterns:

$$\mathcal{P} = (p_1, p_2, \dots, p_p).$$

2.4. Event mining models

There are various tasks of modeling that consider events. In this paper we will focus on relations of events and changes of observations (Fig. 2).

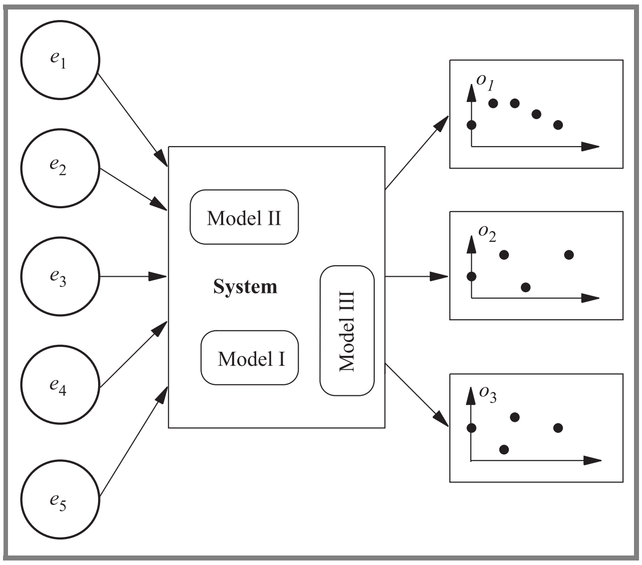

Fig. 2. Relations: system, models, events, observations.

The objectives of modeling:

1. For significant changes of observations find events that are the reasons of these changes

   if *change_detection_after_event*$(\mathbf{O}, w_s)$ **then** *the reasons are the events:* $e_1, e_2, \dots, e_k \in \mathcal{E}_p$,

   w_s is an observation window after the event that occure at time t_i.

2. Prediction of future events by analysing the changes of observations

   if *change_before_event*$(\mathbf{O}, w_p)$ **then** *there is a probability of future events* $e_1, e_2, \dots, e_k \in \mathcal{E}_f$,

   w_p is an observation window before the event that occure at time t_i.

3. Prediction of changes of observations after the event occurs

   if $e_1, e_2, \dots, e_k \in \mathcal{E}_p$ **then** *there will occure pattern* p_i *of changes of observations* $\mathbf{O}$.

The Table 1 have a colum "events", where there is an information about a set of events E_{t_i} that occure at time t_i. Sometimes it is difficult to dettermine the exact time of event. In this paper we will not consider such cases.

Table 1
Event mining data

Time	Events	Observations			
t_1	$\emptyset$	o_{1,t_1}	o_{2,t_1}	...	o_{n,t_1}
t_2	$\emptyset$	o_{1,t_2}	o_{2,t_2}	...	o_{n,t_2}
t_3	E_{t_3}	o_{1,t_3}	o_{2,t_3}	...	o_{n,t_3}
...	...	...	...	...	...
E_{t_i} – the set of events that occure at time t_i					

3. The faults of the network

In this section we will present a simple illustrative example. We are considering a small network with several ATM switches [2] (see Fig. 3). We have two observations of the network o_1 (Fig. 4) and o_2 (Fig. 5) and we know that two events (faults of one connection) occured at time 15:01:30 and 15:19:45. The data is shown in Table 2.

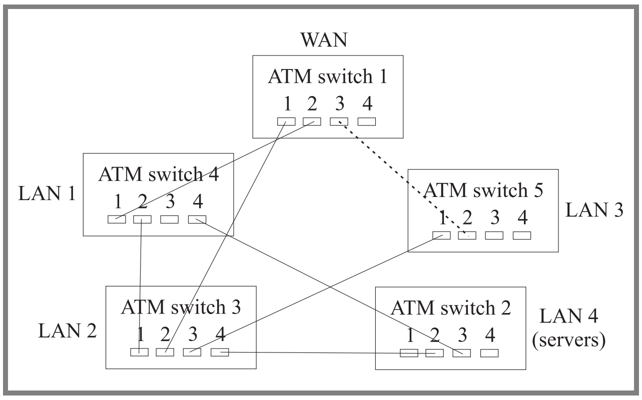

Fig. 3. The network.

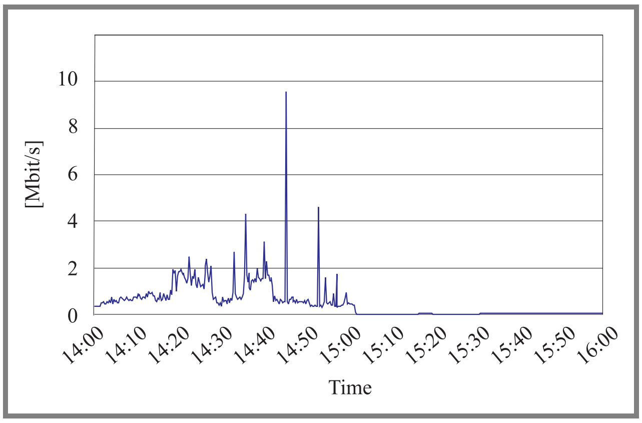

Fig. 4. The observation o_1.

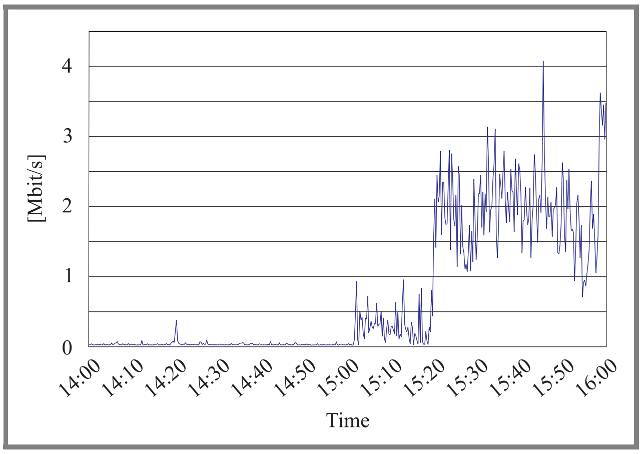

Fig. 5. The observation o_2.

Table 2
Event mining table (training data)

Time	Events	o_1	o_2
...	...	...	...
15:01:30	e_1	419 103	94 570
15:01:45	$\emptyset$	0	433 335
15:02:00	$\emptyset$	0	931 090
...	...	...	...
15:19:45	e_2	23 783	1 563 489
15:20:00	$\emptyset$	20 038	2 108 248
...	...	...	...

We have the following two apriori known events:

$e_1 = (t = 15{:}01{:}30,\ a =$ "fault of the connection of port 3 (switch 1) to port 2 (switch 5)"),

$e_2 = (t = 15{:}15{:}45,\ a =$ "fault of the connection of port 1 (switch 4) to port 2 (switch 1)").

We can observed significant changes of observations after the occurence of events. These changes can be easily detected by statistical algorithms (see [1, 3]). However, the correlation of patterns of changes of observations and events requires an representative training set of cases of faults and set of observation data that fully cover behavior of the system.

4. Conclusions

The observations of the system give an important information about the internal state of the system. However, finding the relations between observations of the system and internal or external events gives the possibility of finding the reasons of changing the behavior of the system. This paper shows mathematical formulation of event mining tasks basd on the set of observation of the system.

References

[1] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. New York: Englewood Cliffs, 1993.

[2] J. Granat, P. Celej, C. Głowiński, and P. Białoń, "Design and management of IP networks, analitycal modules in management of IP networks". Rep., National Institute of Telecommunications, Warsaw, 2001 (in Polish).

[3] J. Granat, P. Celej, M. Kaska, M. Majdan, P. Rzepakowski, and J. Sobieszek, "An environment for monitoring, data mining and modeling with on-line processing of lagrge volumes of data". Rep., National Institute of Telecommunications, Warsaw, 2004 (in Polish).

[4] G. Jiang and G. Cybenko, "Temporal and spatial distributed event correlation for network security", in *Amer. Contr. Conf.*, Boston, USA, 2004.

[5] D. Luckham, *The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems*. Boston: Addison-Wesley, 2002.

[6] L. Perrochon, W. Mann, S. Kasriel, and D. C. Luckham, "Event mining with event processing networks", in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, *Lecture Notes in Computer Sciences*. London: Springer-Verlag, 1999, vol. 1574, pp. 474–478.

[7] C. Wei and Y. H. Lee, "Event detection from online news documents for supporting environmental scanning", *Decis. Supp. Syst.*, vol. 36, pp. 385–401, 2004.

Janusz Granat received his M.Sc. in control engineering (1996) and his Ph.D. (1997) in computer science from the Warsaw University of Technology, Poland. He holds a position as an Assistant Professor at the Warsaw University of Technology, and is the leader of a research group on applications of decision support systems at the National Institute of Telecommunications in Warsaw. He lectured decision support systems and various subjects in computer science. His scientific interests include data mining, modeling and decision support systems, information systems for IT management. Since 1988 he has been cooperating with IIASA. He contributed to the development of decision support systems of DIDAS family and the ISAAP module for specifying user preferences. He has been involved in various projects related to data warehousing and data mining for telecommunication operators. He was also involved in EU MiningMart project.
e-mail: J.Granat@itl.waw.pl
National Institute of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

Institute of Control and Computation Engineering
Warsaw University Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland