Title	Models and Algorithms for Event Mining
Author(s)	Granat, Janusz
Citation	
Issue Date	2007-11
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/4107
Rights	
Description	The original publication is available at JAIST Press http://www.jaist.ac.jp/library/jaist-press/index.html, Proceedings of KSS'2007 : The Eighth International Symposium on Knowledge and Systems Sciences : November 5-7, 2007, [Ishikawa High-Tech Conference Center, Nomi, Ishikawa, JAPAN], Organized by: Japan Advanced Institute of Science and Technology

Models and Algorithms for Event Mining

Janusz Granat[†‡]
[†]National Institute of Telecommunications
Szachowa 1, 04-894 Warsaw
[‡]Institute of Control and Computation Engineering
Warsaw University of Technology
00-665 Warsaw, Poland
J.Granat@itl.waw.pl

Abstract

The information about events is crucial in real-time decision analysis and support. Historically, diverse approaches have been developed to build models and algorithms that relied on the concept of events. The related research areas concern data mining, hybrid dynamic systems, stochastic systems, discrete event simulation etc. An integration of results of the research obtained in those diverse areas might result in an unified methodology that would advance future event-based models and algorithms. In this paper we will show an outline of such unified approach to event-based modelling and analysis.

Keywords: events, event mining, event-based modelling, real-time decision support systems

1 Introduction

There exist diverse modelling approaches that deal with events [10; 11; 7; 2; 9; 8; 2]. However, an integrated approach to events in modelling and analysis of decision problems is lacking. In this paper, we try to develop such an unified approach to event-based modelling and analysis. We are motivated by the need of better understanding the possible futures on one hand and of predicting the most probable future on the other hand. There is increasing number of publications in this area.

1.1 Definition of an event

Definition 1.1 *According to Oxford dictionary the event is "a thing that happens or takes place"*

However, for event-based modelling diverse more technical definitions can be used. For real time decision analysis and support, we introduce here a definition where the event might be described by a set of attributes including attributes related to time:

Definition 1.2 *Let* $A_{e_i} = \{a_{1,e_i}, a_{2,e_i} \ldots, a_{m,e_i}\}$ *denote the set of attributes of an event* e_i *and* $V_{a_{j,e_i}}$ *be the domain of an attribute,* $a_{j,e_i} \in A_{e_i}$*. An event is defined as* $(m+2)-$ *tuple*

$$(a_{1,e_i}, a_{2,e_i} \ldots, a_{m_{e_i},e_i}, t_s, t_e, \Delta t),$$

where $a_{j,e_i} \in V_{a_{j,e_i}}$*,* t_s *is a time of occurrence of the event,* Δt *is the duration of the event,* t_e *- the ending time of the event.*

If $\Delta t = 0$ then we have a point event and for $\Delta t \neq 0$ we have an interval event. If any of attribute is important for analysis we will show this attribute in the brackets e.g. if we want to analyse only the time when the event has happened we write $e_i(t_i)$. Sometimes in the analysis the event is characterised by two attributes $e_i = (E_i, t_i)$ where E_i is a type of event and t_i is the time when event happens.

The same event e_i might be defined by several sets of attributes: $\{A_{1,e_i}, ...A_{k,e_i}\}$. The selection of the set of attributes might depend of the purpose and type of analysis.

1.2 Sets of events

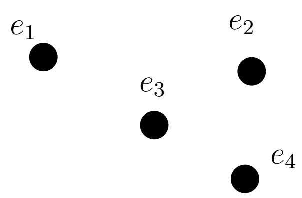

Figure 1. A set of events

We can define:

- A set of *event types*

$$\mathcal{E}^{\mathcal{T}} = E_1 \cup E_2 \ldots \cup E_n$$

For example, event types specified in next items belong to the set of event types. Further enhanced with a structure – e.g., hierarchy – of event types, with cross-relations between them, and with an explication of assumptions usually accepted implicitly (with an analysis of the meta-level of concepts), this set becomes an ontology of events. The entire paper is actually an attempt to construct such an ontology.

- *A finite set of events*

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

where $e_i \in E_j \subset \mathcal{E}^{\mathcal{T}}$.

- *Past and the future events*

$$\mathcal{E} = \mathcal{E}^P_{t_o} \cup \mathcal{E}^F_{t_o}.$$

where events in the set $\mathcal{E}^P_{t_o}$ happens before specified time t_o and other predicted events in the set $\mathcal{E}^F_{t_o}$.

- *Observable and non-observable events*

$$\mathcal{E} = \mathcal{E}^O \cup \mathcal{E}^N.$$

The types of events discussed here are not exclusive: a past event can be observable – i.e., directly measured or otherwise directly observed – or non-observable – i.e., only inferred indirectly from other measurements or observations. An example of observable event is a breakdown of a communication connection, if we measure the throughput in this connection directly; the same event becomes non-observable, if we do not measure the throughput in this connection directly, but infer the event from measurements in other communication connections.

- *Rare and frequent events*

$$\mathcal{E} = \mathcal{E}^{RA} \cup \mathcal{E}^{FR}.$$

The distinction of *rare events* is relative – depends upon the definition of what is rare – but important from the point of view of model selection, since it characterizes the amount of data that can be used for event analysis.

1.3 Structural types of events

Sets of events can be organised in various structures or structural types, such as sequences, temporal sequences, graphs, spatio-temporal structures, etc.

1.3.1 Sequences of events

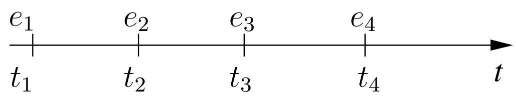

Figure 2. A sequence of events

A *temporal sequence* of events is defined as

$$\mathcal{S}_{temp} = (e_1(t_1), e_2(t_2) \dots, e_n(t_n))$$

where
$t_i <= t_{i+1}$ for each $i = 1, \dots, n-1$

In a *simple sequence* of events we do not consider the exact time of event occurrences, only the order of events is important:

$$\mathcal{S}_{seq} = \{e_1, e_2, \dots, e_n\}$$

1.3.2 A stream of events

A stream of events means that the input data to be analysed are not available all at once, but arrive as continuous temporal sequences of events. The events stream is defined as

$$\mathcal{S}_{stream} = (\dots, e_1(t_1), e_2(t_2) \dots, e_n(t_n), \dots)$$

where
$t_i <= t_{i+1}$ for each $i = 1, \dots, n-1$

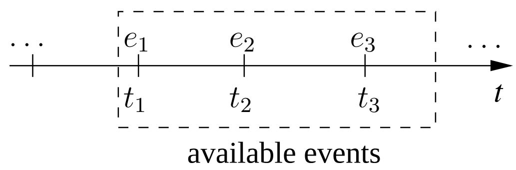

Figure 3. The events stream

1.3.3 Spatio-temporal events

In some problems, event location might be also important.

We can define the set of events locations:

$$\mathcal{L} = \{l_1, l_2, \dots, l_n\}$$

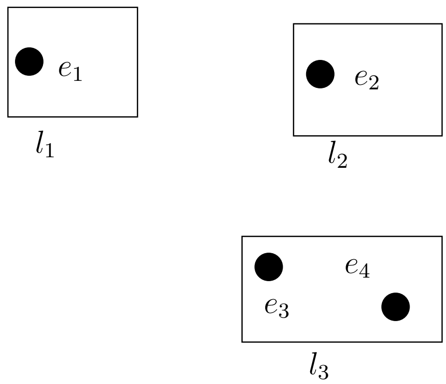

Figure 4. Spatio-temporal events

and then we can define a temporal sequence of events at location l_i as

$$\mathcal{S}_{temp,l_i} = (e_1(t_1, l_i), e_2(t_2, l_i) \ldots, e_n(t_n, l_i))$$

where
$t_i <= t_{i+1}$ for each $i = 1, \ldots, n-1$. In a similar way we can define a simple sequence or a stream of events at a given location.

1.3.4 A graph of events

A graph $G = (V, Ed)$ is a set of vertices V and the set of edges Ed, $Ed \subset (V \times V)$. The set of vertices can be defined to be equivalent to a set of events $\mathcal{E}$. The edges are then equivalent to relations between events.

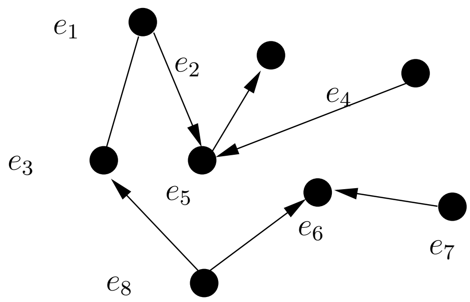

Figure 5. A graph of events

Additionally, we can introduce diverse conceptual levels. It is useful to analyse the relationship between events on various conceptual levels. In this case, the set of vertices can be subdivided to belong to several conceptual levels $V = V_1, V_2, \ldots, V_n$.

2 Stochastic models of events

There is a long tradition of modelling events by stochastic processes. Here, we will present some

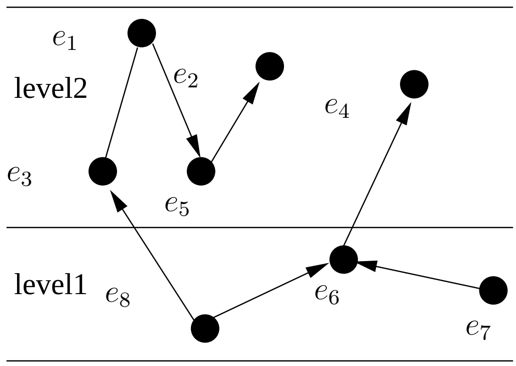

Figure 6. Levels of events

basic formulations.

- Survival processes

  Survival analysis is also known as *time to event analysis:* e.g. time until a response, time until failure. We can model this by defining a starting event type $E_1(t_1)$ and ending event type $E_2(t_2)$. The time between these two events are $T = t_2 - t_1$ is a random variable and t is realization of T. S(t), called the survival probability, is defined as the probability of surviving beyond time interval t, i.e., the probability that the event occurs after t:

$$S(t) = P(T > t)$$

$$S(t) = \int_t^{\infty} f(x)dx$$

We can also define *the hazard function* $h(t)$, also called *conditional failure function*:

$$h(t) = \lim_{\Delta t \mapsto 0} \frac{P(t \leq T \leq t + \Delta t \| T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}$$

The function $h(t)$ can be interpreted as an "instantaneous" probability that the failure event occurs at time t, given that no failure occurred before t. The survival probability $S(t)$ can be also expressed as follows:

$$S(t) = \exp[-\int_0^t h(x)dx]$$

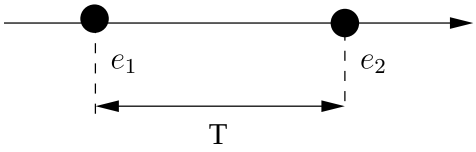

Figure 7. The analysis of time to event

- Recurrent events
  In this case, only one type of events is considered and the times of events are interpreted as *arrival times*; arrival times or the times between events are recorded, the latter called *inter-arrival times*. A binary indicator at any time point can be used, equal to 1 if the event occurs and 0 otherwise. In this case, the process can also be called a *binary point process*. Such temporal sequence of events can be analysed in various ways, using:

  1. The frequency of occurrence of events;
  2. The intervals of times between events; This case is sometimes called a *renewal process*; this name comes from industry, where certain machines or parts must be replaced or renewed at varying intervals of time. But models for renewal processes have much wider application, e.g.

  For recurrent events, it is often especially important to define what is the zero time point ($t = 0$, or any other t_0). It is usually convenient to start the process from the time of the first event considered (although the concept of the first event is obviously also relative).

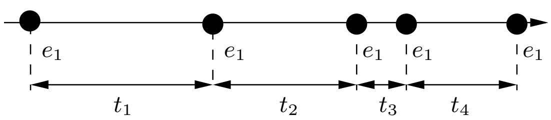

Figure 8. The analysis of recurrent events

In this case, we can define a random variable $N(t)$ representing a counting process in the time interval $[t_0, t]$, or the number of occurrences of events in the given time interval. The Poisson counting process $\{N(t) : t \geq t_0\}$ has the following properties [3]:

1. $Pr\{N(t_0) = 0\} = 1$
2. The increment $N_{s,t} = N(t) - N(s)(t_0 \geq s < t)$ has the Poisson distribution with the mean parameter $\Lambda_t - \Lambda_s$, for some positive and increasing function in t
3. $\{N_t : t \geq t_0\}$ is a process of independent increments, $N(t_1) - N(t_0), \ldots, N(t_n) - N(t_{n-1})$ are mutually independent

For this counting process $\{N(t) : t \geq t_0\}$ we can define an intensity function as:

$$\lambda(t) = \lim_{\Delta \mapsto 0} \frac{1}{\Delta} Pr\{N(t+\Delta) - N(t) = 1 | H(t)\}$$

where $H(t)$ is the history of the process up to t

$$H(t) = \{N(u) : t_0 \leq u \leq t\}$$

$$\Lambda(t) = \int_{t0}^{t} \lambda(t) dt$$

- Until now we have assumed that events are independent of each other. However, it is often useful to consider the case when an event is dependent on previous event. In such a case we can model the events by a hidden Markov model.

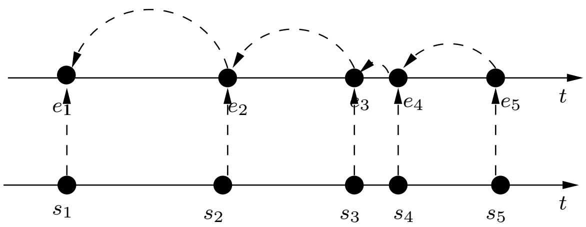

Figure 9. A dependence between events

3 Observations and events in a system

In the previous sections we have assumed that we have direct information about the occurrence of events and we are analysing various properties of events. Here, we shall introduce an additional term: an observation of the system [4]. The common theoretical framework that allows to analyse observations and events jointly is the concept of an information system.

3.1 Observations

Information system $\mathcal{O}$ of the set of observations can be defined as follows:

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

gdzie:

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 a set values of observations $o \in O$, called the domain of o,
ρ – is an information function:
$\rho : O \times T \rightarrow V$.

We assume that we will have the set of observations:

$$\mathbf{O} = (o_1, o_2, \ldots, o_l) \quad (2)$$

$$T = (t_1, t_2, \ldots, t_n)$$

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

3.2 Observations pattern

Observations pattern p_i is distinguishable sequence of observations

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

Pattern p_i might be described by a set of parameters, etc.

Set of patterns:

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

3.3 Event mining models

There are various tasks of modelling that considers events. In this paper we will focus on relations between events and changes of observations.

The objectives of modelling might be as follows:

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

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

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

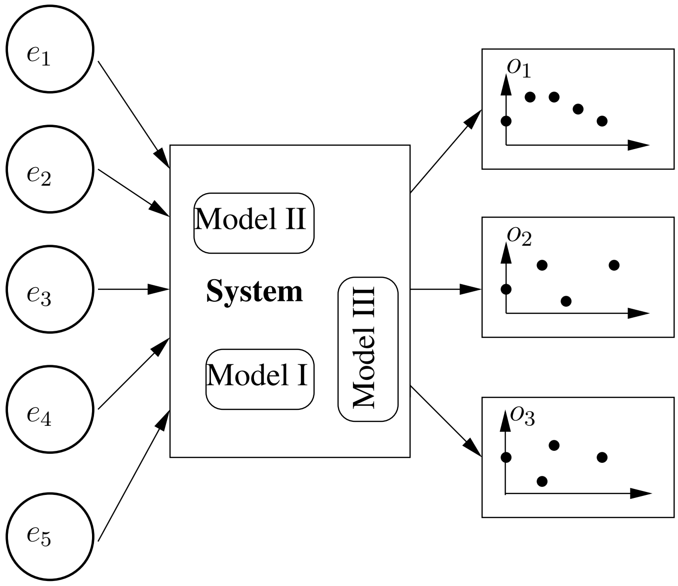

Figure 10. Relations: system, models, events, obserwations

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

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

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

3. Prediction of changes of observations after an event occurs

   if $e_1, e_2, \ldots, e_k \in \mathcal{E}_p$ **then** *there will occur a pattern of changes of observations o*

Table 1. Event mining table

time	events	observations			
t_1	$\emptyset$	o_{1,t_1}	o_{2,t_1}	$\ldots$	o_{n,t_1}
t_2	$\emptyset$	o_{1,t_2}	o_{2,t_2}	$\ldots$	o_{n,t_2}
t_3	E_{t_3}	o_{1,t_3}	o_{2,t_3}	$\ldots$	o_{n,t_3}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$

E_{t_i} - the set of events that occur at time t_i

The table have column event where there is information about a set of events E_{t_i} that occur at time t_i. Sometimes it is difficult to determine the exact time of event. In this paper, however, we consider only the cases when the times of events are well determined.

4 Conclusions

The paper is an attempt to construct an ontology focused on description and modeling of various relations of the behavior of the system and internal as well as external events that affect behavior of the system. The integration of research results obtained in diverse areas is essential for building such ontology. Events are one of the phenomenon that contribute to the enhancement of capabilities of the future knowledge management systems. The selected applications of event-based modeling can be found in [5; 6].

Acknowledgment

This paper was partially supported by 21st COE (Center of Excellence) Program Technology Creation Based on Knowledge Science of JAIST, funded by Ministry of Education, Culture, Sports, Science and Technology (MEXT, Japan).

References

[1] Baccelli F. and G. Cohen and G.J.Olsder and J.-P. Quadrat, Synchronization and Linearity. An Algebra for Discrete Event Systems, Wiley, 1992.

[2] Basseville. M. and I.V. Nikiforov, Detection of Abrupt Changes:Theory and Application, Prentice-Hall, Inc., 1993.

[3] J.E.Gentle, Hrdle W, Mori Y (eds), Handbook of Computational Statistics Concepts and Methods, Berlin: Springer-Verlag, 2004.

[4] J.Granat, Events driven approach for supporting real-time management of complex systems, Proceedings of the First Word Congress of the International Federation for Systems Research (IFSR), Kobe, Japan, pp. 306-309, 2005.

[5] 13] Granat J., Event mining based on observations of the system (Journal of Telecommunications and Information Technology) National Institute of Telecommunications, no. 3, pp. 87-90, 2005.

[6] Granat J., A framework for event based modeling and analysis (Journal of Telecommunications and Information Technology) National Institute of Telecommunications, no. 4, pp. 88-90, 2006.

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

[8] Nielsen, M. and G. Plotkin and G. Winskel, Petri Nets, Event structures and Domains, Theoretical Computer Science 13, pp. 85–108, 1981.

[9] Winskel, G., Events in Computation, PhD thesis, University of Edinburgh, 1980.

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

[11] Perrochon, L., Mann, W., Kasriel, S. and Luckham, D. C., Event mining with event processing networks, Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 474–478, 1999.