Handling Uncertain Rules in Composite Event Systems

Segev Wasserkrug

Technion, IBM Research
Haifa, Israel Haifa, Israel
segevw@il.ibm.com

Avigdor Gal

Technion –
Haifa, Israel
avigal@ie.technion.ac.il

Opher Etzion

IBM Research
Haifa, Israel
Opher@il.ibm.com

Abstract

In recent years, there has been an increased need for active systems - systems that are required to act automatically based on *events*, or changes in the environment. In many cases, the events to which the system should respond to, have to be inferred from other events based on complex temporal predicates. However, none of the existing composite event systems created to enable such inference can deal with cases in which an event cannot be inferred with absolute certainty based on the reported events. Therefore, in this paper, we describe how a deterministic event composition system can be extended to manage such uncertainty, and specify the principles of a formal framework for such inference. The contribution of this framework is twofold: It extends the semantics of event composition in a natural manner for probabilistic settings, and it enables the application of these extensions to the quantification of the occurrence probability of events.

Introduction

In recent years, there has been a growing need for the use of active systems, i.e. systems that are required to act automatically based on events. However, in many cases, the events of interest to which the system must respond are not generated by monitoring tools, but must be inferred from other events.

Although many *event composition* systems and prototypes have been defined to facilitate such inference, none of these mechanisms can take into account uncertainty in the inference process. Such uncertainty is inherent in many cases: For instance, it would be desirable for a banking system to detect all money laundering based solely on events indicating deposits and withdrawals. However, based on such events, the best that can be achieved is the calculation of some measure of likelihood regarding the occurrence of an actual money laundering event.

In this paper, we describe basic principles for extending deterministic event composition in a formal manner, to take uncertainty of this kind into account. To do this, we use probability theory as the uncertainty handling mechanism. To the best of our knowledge, this is the first work that enables the handling of uncertain rules in event composition languages, in a general and formal manner.

Related Works

Various systems enabling event composition exist in the literature (e.g. Snoop [Chakravarty 1994] and the Situation Manager Rule Language [Adi 2002]). A major shortcoming of all existing specification languages is that they are unable to handle uncertainty, in a general and formal manner, in the event inference process. Moreover, most of the existing systems do not support uncertainty management in any form.

Deterministic Event Composition Languages

Several composite event languages enable the specification of a set of rules, where each rule r is a tuple of the form $<sel_r, group_r, pattern_r, eventType_r, mappingExpressions_r>$ where:

- sel_r specifies event instance selection. This is a filter defined over individual event instances, limiting the events that may be considered for composition by the rule. Given a specific set of event instances, we will call the event instances selected by sel_r the *operands of rule r*.
- $group_r$ specifies event instance grouping. This will allow grouping together semantically related events, such as all events of type either *stockPurchase* or *stockSell* referring to the same stock.
- $pattern_r$ specifies an event instance pattern. This defines a temporal predicate over the evidence events.
- $eventType_r$ is the type of inferred event.
- $mappingExpressions_r$ defines how the attributes of the instance of the inferred event are calculated from the attributes of the operands.

The semantic meaning of such a rule is that event e_r of type *eventType* is inferred iff a set of event instances specified by sel_r have occurred, such that all of these events belong to the semantic group specified by $group_r$ and satisfy the predicate defined by $pattern_r$, irrespective of what other events have occurred. This means that only events of the types specified by sel_r can cause the inference of event e_r, irrespective of whatever other events have occurred.

Adding Uncertainty to Event Composition

In this section, we outline the general principles of a framework that enables handling uncertainty in event composition languages of the type described above. First, we extend such languages by adding to each rule definition a quantity $prob_r \in (0,1]$, which quantifies the probability of the event occurrence. Note that such a probability can also be assigned to deterministic rules, where $prob_r = 1$.

After extending the language as described above, the information the system has about the occurrence of each event will be represented by a *Random Variable* (RV). In the sequel, we shall somewhat abuse notation by referring both to the RV and the corresponding event by the same notation, where the meaning will be clear from the context.

Based on this representation, we now define the following notation: EH is the set of random variables representing the evidence held by the system regarding the events that have occurred up to the time rule *r* is considered for inference. $EH_{SEL,r}$ is the subset of random variables corresponding to events in EH that may be considered for selection in rule *r*. Note that $E \in EH_{SEL,r}$ iff E can be selected by sel_r. In addition, we shall define a new random variable $SEL_r(EH)$, and a function $pattern_r(x) \rightarrow \{true, false\}$, where x is a set of events. $SEL_r(EH)$ is a random variable whose values are all possible subsets of events which can be chosen, given that all that is known about the set of events which occurred before considering rule *r* is *EH*. $pattern_r(x)$ is a function whose domain is all the possible values of the random variable $SEL_r(EH)$, and $pattern_r(sel) = true$ iff the set of events sel satisfies the predicate defined by $pattern_r$.

Given the above notations and definitions, the semantics we define for the uncertain language can be formally represented by the following formulae:

(1) $\Pr(E_r \mid EH_{SEL,r}, EH) = \Pr(E_r \mid EH_{SEL,r})$

for all possible values of random variable E_r and the sets of random variables $EH_{SEL,r}$ and *EH*. This specifies that given the values of the random variables belonging to $EH_{SEL,r}$, the values of the remaining random variables of *EH* have no bearing on the occurrence of event E_r.

$$\begin{aligned} &\Pr(E_r \mid SEL_r(EH) = sel, EH) = \\ (2)\quad &\Pr(E_r \mid SEL_r(EH) = sel, EH_{SEL,r}) = \\ &\Pr(E_r \mid SEL_r(EH) = sel) \end{aligned}$$

for all values of random variable E_r, all specific values sel of random variable $SEL_r(EH)$, and all values of the sets of random variables $EH_{SEL,r}$ and *EH*. This specifies that given the events that should be selected by sel_r, the rest of the information contained in *EH* has no bearing on the occurrence of E_r. In addition, we have that:

$$(3)\quad \begin{aligned} &\Pr(E_r = true \mid \\ &\quad SEL_r(EH) = sel_r) = prob_r \text{ if } pattern_r(sel) = true \end{aligned}$$

that is, the quantity $prob_r$ defined in the rule is the probability that event E_r occurred given that the events *sel* selected by $SEL_r(EH)$ fulfill the predicate defined by $pattern_r$. Similarly:

$$(4)\quad \begin{aligned} &\Pr(E_r = true \mid \\ &\quad SEL_r(EH) = sel) = 0 \text{ if } pattern_r(sel) = false \end{aligned}$$

To calculate the probability distribution of the event occurrence, as events reach the system, a Bayesian network (see [Pearl 1988]) is dynamically constructed based on the probabilistic independencies represented by the above formulae (an example of such a network appears in Figure 1). Although the exact algorithm by which this network is constructed and updated is beyond the scope of this article, we stress that this Bayesian network is only used as a means to calculate the probability distribution over possible event occurrence at each point in time. It is not used as the sole means for representing the information the system has about the possible event occurrences.

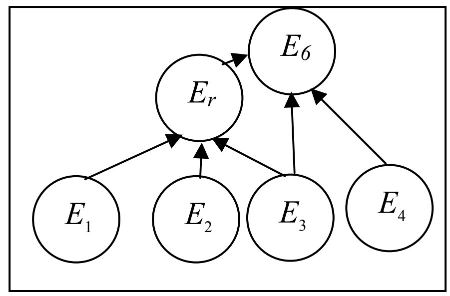

Figure 1: Bayesian network example

Summary and Future Work

In this paper, we described the principles by which existing composite event systems can be extended to take into account uncertain rules. This is the first research we are aware of that addresses uncertain rules in the context of event composition systems in a comprehensive and formal manner. However, much work still has to be carried out. Avenues for future research include the evaluation of other types of uncertainty and the optimization of the inference algorithm.

References

Adi, A.; and Etzion O. 2002. The Situation Manager Rule Language, *International Workshop on Rule Markup Languages for Business Rules on the Semantic Web*, 36-57 Springer-Verlag.

Chakravarthy S., 1994. Snoop: an expressive event specification language for active databases., *Data and Knowledge Engineering*, 14(1), 1-26.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.