Probabilistic Event Processing for Situational Awareness

Davy Preuveneers[1], Andreas D. Landmark[2], Leendert W.M. Wienhofen[2]

[1]IBBT-DistriNet, KU Leuven, Celestijnenlaan 200A, B-3001 Leuven, Belgium
[2]Department of Computer and Information Science, Norwegian University of Science and
Technology (NTNU), Sem Sælandsv 7-9, NO-7489 Trondheim, Norway

davy.preuveneers@cs.kuleuven.be
andreala@idi.ntnu.no
leendert.wienhofen@idi.ntnu.no

Abstract: Over the last century the continuous innovation of technology, coupled with a steady increase in the size of the healthcare organizations, has created a need for information systems supporting healthcare professionals with their daily tasks and decisions. Modern hospitals are full of technology producing electronic records of events and activities, with the opportunity of these events culminating in a wealth of information that these semi-autonomous experts can tap into to improve situational awareness, facilitate coordination and take better informed decisions. However, processing these footprints, contextualizing and inferring over them presents several interesting challenges to the current state of Complex Event Processing methods. This article looks at challenges presented by an information system for perioperative process support and how contextualization and adequate tool support can provide the essential backdrop for meaningful inference.

1 Background

Clinical work processes can resemble non-deterministic processes, at least when it comes to the core activities that determine clinical work. As problem solving activities, in which several actors participate, each from their own domain of knowledge, creating detailed models of clinical processes can be difficult. The exact flow of control (the trajectory of actions performed) is determined by the problem to be solved (e.g. the health issue of the patient) and the decisions that individual actors make, based on their assessment of this problem and their knowledge. This means that different problem solving activities can interfere and due to matters of urgency or lack of resources, the ordering of the activities will vary given the local circumstances.

The overall objective of COSTT [1] is to evaluate and design novel ICT support for a selection of clinical processes. Our belief is that efficient support can be achieved not through explicit control of the flow of work, but rather by providing all actors involved an easy accessible, comprehensible overview of the progress of a process and its current status. By making the process transparent to all those involved, the actors can coordinate their own

---

[1]http://www.costt.no/

work. So, coordination is facilitated but not dictated, which is often favorable in complex organizations [GM01].

The core concept in our visualization of progress and current status of clinical processes is the patient trajectory. We define a patient trajectory as a timeline-oriented representation of what actually has occurred and will happen with the patient during encounters with clinicians. Through inspecting a patient trajectory, a clinician can see how far the plan concerning a patient has progressed, and also whether there have been deviations from the original plan. Based on this information he can decide if he needs to make any adjustments to his own activities.

However, as shown by others [RRvdGK09, HTS06], stringent workflow systems and attempts at creating detailed guidelines or exhaustive process maps for healthcare often break down because of the apparent non-deterministic nature of healthcare processes. Additional work has also uncovered that some process variation can be beneficial for the treatment process and as such, should be supported rather than discouraged by information systems facilitating such work [BFS11].

Capturing these subtleties in software or tool support has shown to be difficult. Several strategies have been explored, often resulting in issues such as increasing the level of abstraction until the semantic value of the end result is diluted, or the numbers of disjoint but similar processes grows almost exponentially to cover all these eventualities. However, attempting to map any exception to any process of non-trivial size often descends into a chaotic representation of reality. Additionally the prospect of mapping out *the unknown* a priori, is daunting at best.

The goal of this paper is to address these challenges by introducing and motivating the concept of situational awareness through event-based information systems. The paper is organized as follows. In Section 2 we introduce the healthcare professionals as a social community with a strong desire to improve self-coordination and cooperation. We show the value of probabilistic event processing in Section 3, and present modeling and tool support in Section 4 demonstrating the importance of this concept. We end with conclusions and further work in Section 5.

2 Motivation: the Professional Healthcare Community

In 'social communities', being 'social' is the driving force behind most of the available platforms, and time is in general of less critical importance, nor is facilitating cooperation. In our domain, time and cooperation are key factors for achieving 'a good day' and hence providing good care. Keeping track of colleagues and what they are doing and how these activities might impact your own work is a difficult task. Operations and associated resource planning can only to a limited degree be captured in an explicit workflow. Due to the rapid changes in the domain (acute patients causing changing priorities and so on), continuous replanning is necessary. The overall activities such as start and finish are quite easy to conceptualize, though for coordination purposes it is also very relevant to know what happens between these markers. Knowing when an operation is expected to end en-

ables for example a coordinator to anticipate, prepare and reschedule following patients as needed.

Not unlike in less critical 'social communities', location and situation awareness can have positive and negative effects. In general people will rarely trust one system completely, yet better overview, provides a better starting point for self-coordination. Still, the 'terrain takes precedence over the map' when reality and plan diverge.

One of the fundamental assumptions in the COSTT-project was that given better transparency of the processes - or *situational awareness*, users of the system would be able to better self-coordinate. Situational awareness is a term used to describe at which level a person has perceived the current situation. In increasing levels of awareness [End95], the first level would be to perceive the current situation. The second level would be to not only perceive the elements in your immediacy, but also to comprehend the meaning of the events occurring. When the highest level of situational awareness is achieved, one is also able to project how the current situation might evolve based on comprehension of the current as well as knowledge about how the current situation *usually* evolves.

Using situational awareness as a concept to describe the comprehension of an environment and its impact on your own goals and objectives, it has also been shown that lack of or inadequate awareness is a contributory factor to 'human errors'. As such, building computer support to help increase and support situational awareness is an obvious extension to the already existing technology for distributing and increasing situational awareness .

3   Situational awareness with events

Situational awareness is a field of study concerned with perception of the environment critical to decision-makers in complex, dynamic areas such as aviation, military operations, and healthcare. The perception of environmental elements with respect to time and/or space, the comprehension of their meaning, and the projection of their status after some variable has changed, such as time. Situational awareness involves being aware of what is happening in the vicinity to understand how information, events, and one's own actions will impact goals and objectives.

Trying to create a software support system for such a process then entails support for aggregating and combining heterogeneous knowledge and events from a variety of sources.

3.1   Events and complex event patterns

In general it is meaningful to distinguish between (at least) two levels of event richness: Simple events and complex events. Simple events are single events that carry slivers of meaning in themselves, without much room for decomposition. Examples of simple events would be stock order placements, atomic bank account transactions, or stock trades (buy/sell-orders being matched). This is in contrast to complex events, which summarize,

represent, or denote a set of single events which combined would denote a 'pattern of events' of a specific process. An oft-cited example is the 1929 stock market crash [LS08]. Complex events are formally processes rather than events, but within the domain are often treated as events and subject to the same processing and rules as for simple events.

3.2 Contextualized semantics of non-deterministic event patterns

Tackling the challenge of the non-deterministic nature of healthcare processes is instrumental to realizing a system which can cope not only with the majority of regular cases - but also recognize the minority of cases with deviations in event value. From a practical point of view, the value of any support system is at its highest when it helps support the difficult minority rather than the more streamlined majority.
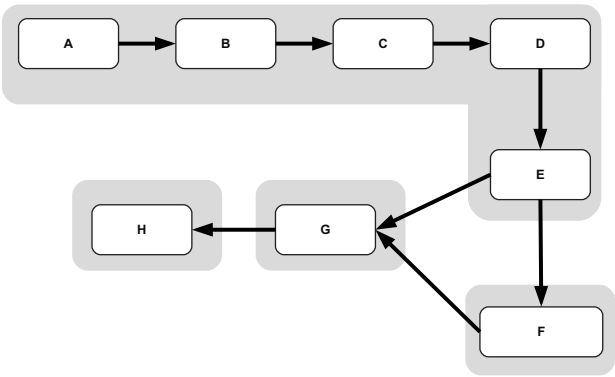

Figure 1: Example patient flow

Figure 1 shows a typical patient flow through several diagnostic activities. In a perfect world, these steps are taken in a pre-determined chronological order. However, in real life deviations to the regular patient flow order are possible. The allowed deviations are driven by dependencies in the information flow dependencies between the healthcare professionals (see Figure 2).

In the following sections, we will refer to these steps as the *situations* a patient can be in. Furthermore, each of these steps is characterized by events, which motivates our approach of situational awareness with events.

One of the main challenges in the perioperative domain is that the systems from which we harvest events were not designed to be part of a larger system and to a large degree live a life of their own. The COSTT-project assumes a bottom-up collection of events wherein we combine physical sensor sources such as indoor positioning systems with events culled from information systems such as planning- and recordkeeping-systems. While this allows for a broad-spectrum of information about the processes and activities in the hospital, it also pushes the issues of temporality, vagueness and uncertainty to the forefront.
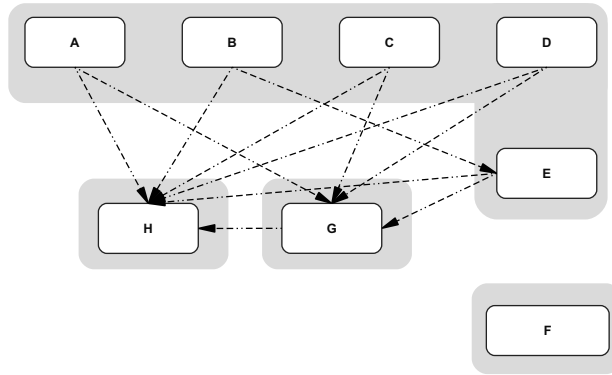
Figure 2: Information flow dependencies

Using sources that rely on manual input as the primary means of gathering information (such as most clinical recordkeeping systems) also means that the variation in the currency of information is, at times, extreme. On the other hand, physical sensors have predictable currency, but inherent shortcomings in sensitivity and specificity.
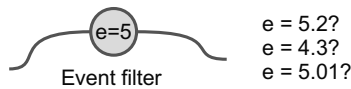

Figure 3: Acceptable deviations in event values

Filtering event values based on medical data can also prove difficult given that the test of clinical significance is often given in ranges rather than specific values. What is acceptable for one patient might be unacceptable to the next patient, even given the same diagnosis. This makes simple filtering as shown in Figure 3 difficult without a rich context for interpretation, which in this case could range from baseline values for this particular patient to encompassing significant parts of the patients medical history. Additionally, certain values have different ranges depending on how the reading was obtained (e.g. body temperature taken rectally, orally or axillary).
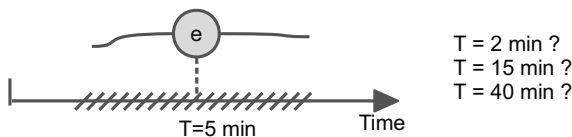

Figure 4: Acceptable deviations in event timestamps

Spatial and temporal relevance (see Figure 4) is just one aspect of the contextual relevance of an event.

Healthcare is rife with examples of procedures where the number of stages depends on the

results of prior steps. Education and experience can also influence whether or not certain stages are skipped or additional ones are performed. The effect from an information system point of view is that some events in a pattern become optional.
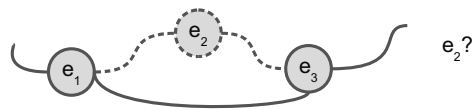

Figure 5: Optional or undetectable events

In terms of defining event patterns, one needs to define which events are optional - see Figure 5 for an illustration - where the non-occurrence of such an event will not make the pattern matching wrongly refute the correct pattern.
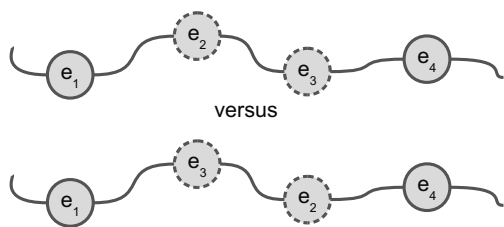

Figure 6: Acceptable deviations in event orderings

The order in which events occur often follows a predefined business logic or workflow path, which could be seen as an event pattern. The order in which events occur can differ based on the workflow path chosen. The ordering of events in the patient workflow may change due to resource constraints or interference with other patients. For example, whereas the logical consequence of events would be e_1, e_2, e_3, e_4, the order of events e_2 and e_3 for a particular patient might be altered, as depicted in Figure 6, if for example there is currently no free slot in the receiving department.

4 Situation Studio: Tool support for event-driven activity recognition

Based on one of the cases in the COSTT project, a pre-operation examination day (for details we refer to [WPL+11]), we designed a flexible event-based workflow system, called the 'Situation Studio'. It is used to model different situations, the probability of their occurrences, the variety of observable and non-observable events in each situation, and the possible partial ordering between situations.

Figure 7 illustrates the graphical user interface of the tool. It shows a concrete instance of the example patient trajectory as shown in Figure 1.
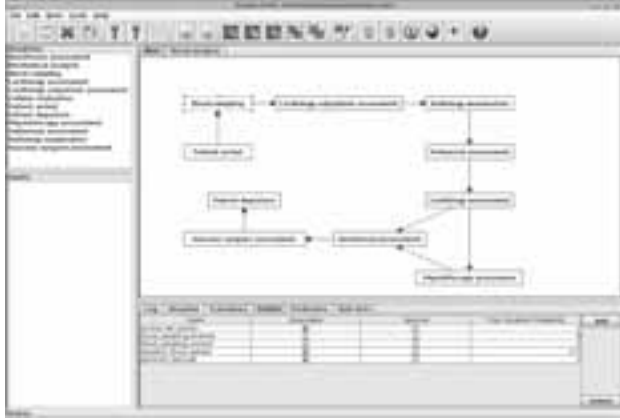
Figure 7: The Situation Studio

4.1 Characterizing situations through events

Each of the situations is characterized by events. For example, the *Blood sampling* situation is characterized by the following observable and mandatory system events:

- `access_lab_system`
- `generate_bar_code`
- `dispatch_blood_sample`

The order of events is typically as listed above. The *Blood sampling* situation is bounded in time by two non-observable events:

- `blood_sampling_started`
- `blood_sampling_finished`

These implicit events mark the beginning and the end of a situation. They can only be inferred from the occurrence of other observed or inferred events. Unless their occurrence matches one-on-one with an observable event, the exact timing of their occurrence is usually uncertain.

Non-observable events are triggered with *rule sets*, i.e. a series of *if-then-else* rules. For example, if the system detects a system event in the *Cardiology outpatient assessment*, we can infer that the previous *Blood sampling* situation has ended and we can trigger the `blood_sampling_finished` event.

With *predicates* we define when a situation can possibly emerge through declaring which (combination of) events should have already occurred, and which ones must not have occurred:

- `patient_id`   (this event must have occurred)

102

- `!blood_sampling_finished`   (this event must not have occurred)

For this scenario, when the patient arrives (i.e. he is in the *Patient arrival* situation), the system generates a `patient_id` event with all the personal details of the patient. As long as the patient is not registered, we are sure that the patient cannot be in the *Blood sampling* situation. By also specifying the `!blood_sampling_finished` predicate, we declare that a patient will never go back to this situation if he has completed this step before.

4.2   Probabilistic event processing: a pragmatic approach

With each situation, we associate a probability with each of its events to ascertain the possibility that the patient is still in this situation. For example, the *Cardiology outpatient assessment* situation is characterized by the following observable events:

- `access_epr`: the cardiologist opens the electronic patient record (EPR)
- `dicate_result`: the cardiologist dictates the results of the assessment into speech recognition software

However, healthcare specialists have different working habits. Some may only open the EPR while the patient is sitting in front of them, or dictate the results while the patient is still present, while other ones open all the patient files in the morning or dictate the results after the patient has left. Hence, the occurrence of a particular event is not a guarantee that the patient is still at this location. That is why we associate a prior probability of each event in each situation to characterize the possibility that the patient is at this location when this event occurs. These prior probabilities are derived through discussions with the medical stakeholders. For the *Blood sampling* situation this has led to a prior probability of 100% for the `access_lab_system` and `generate_barcode` events, and a prior probability of 70% for the `dispatch_blood_sample` event. This means that the patient is surely at this location when either of the two first events is recognized. However, there is a slight chance that the patient has already left when the last event is triggered.

Ideally, we would have used proven probabilistic reasoning techniques like Bayes' probability theory, Zadeh's fuzzy logic or Dempster-Shafer's evidence theory. We investigated each of these techniques but none of them turned out suitable due to pragmatic reasons, such as the maintenance of the knowledge for non-technical experts. With Bayes' theorem, we can compute the probability for a situation S given the events E knowing the probability of the events given the situation.

$$P(S|E) = P(S \cap E)/P(E) = P(E|S) * P(S)/P(E)$$

However, each situation is usually characterized by a set of events:

$$P(S|E_1, E_2, E_3, \ldots) = P(E_1, E_2, E_3, \ldots |S) * P(S)/P(E_1, E_2, E_3, \ldots)$$

This means that for any set of events we need to know their probability in every situation, and this is guess work without a proper data set from which we can obtain these probabilities.

Zadeh's fuzzy logic has the advantage that it allows you to express domain knowledge with linguistic terms rather than with crisp values. However, various arbitrary choices have to be made, such as the shape of each fuzzy variable (triangle, trapezoide, bell, . . .), the modeling of fuzzy sets and rules, as well as the defuzzification into crisp values. Figure 8
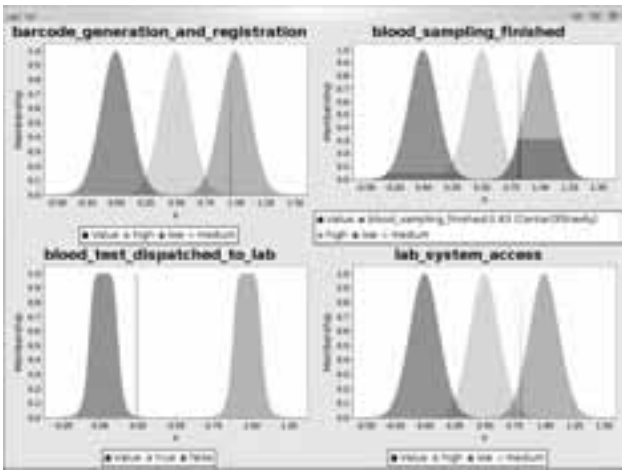

Figure 8: Fuzzy Logic for the Blood Sampling situation

illustrates this concern for inferring the `blood_sampling_finished` event based on the occurrences of the other observable events, based on fuzzy rules like the following:

IF (blood_test_dispatched_to_lab **IS** false) **THEN** blood_sampling_finished **IS** low;

IF (lab_system_access **IS** medium) **AND** (blood_test_dispatched_to_lab **IS NOT** true)
    THEN blood´sampling´finished **IS** low;

. . .

The evidence theory from Dempster-Share is a generalization of Bayes based on belief and plausibility, but without going into details, experiments with Dempster's combination rule of evidence have shown that it can sometimes lead to counter-intuitive results. Zadeh himself used the following example to illustrate this concern:

```
Doctor A:      99% brain tumor,    1% meningitis
Doctor B:      99% concussion,     1% meningitis
Dempster's combination rule:    100% meningitis
```

Obviously, this result is very counter-intuitive. Instead, we pursued a more pragmatic approach. Remember that situation *X* means that the patient is at location *X*. Various events pertain to a particular situation (e.g. `access_epr`, `change_ris`, `dictate_result`,

. . .). Due to the fact that events related to the situation can actually take place before, during or after events, we used prior probabilities to model these uncertainties:

```
P(access_ris | Radiology examination)    = 1.0
P(access_epr | Pulmonary examination)    = 0.6
P(dictate_result | Cardiology examination) = 0.6
```

If predicates of a situation are false, then then particular situation is impossible (likelihood is 0.0). For example, the *Cardiology assessment* cannot take place if the *Cardiology outpatient assessment* has not finished. If all the predicates are true, we compute the probability of the situation based on probability of the last correlated event, and infer the possibility of all the remaining situations. However, this may lead to some mathematical non-sense. Given the likelihoods of the following possible situations:

```
P(Cardiology outpatient assessment) = 0.7    // report_ready
P(Radiology examination) = 0.5               // access_epr
P(Pulmonary assessment) = 0.5                // access_epr
```

We see that the sum of the probabilities is not 1. The reason for this behavior is that the related events do not occur all at the same time. If P(X) would be 1.0, we would be absolutely sure that the patient is at that location. However, if it would be 0.95, then there is room for doubt. To solve this problem, we implemented a function $f(x_i)$ (with x_i being the values above) with the following properties:

- $\Sigma \, f(x_i) = 1.0$
- $f(1.0) = 1.0$ and $f(0.0) = 0.0$ *(What is absolutely true or false, remains so)*
- Partial ordering of x_i is the same as partial ordering of $f(x_i)$

The solution is a value z with $f(x_i) = (x_i)^z$ and z such that $\Sigma(x_i)^z = 1.0$. The value z is not easy to compute directly, so we use an iterative method to find the right value.

```
P(A) = 0.7                              f(P(A)) = 0.494
P(B) = 0.5         with z = 1.980       f(P(B)) = 0.253
P(C) = 0.5                              f(P(C)) = 0.253
```

or

```
P(A) = 0.99                             f(P(A)) = 0.948
P(B) = 0.5         with z = 5.265       f(P(B)) = 0.026
P(C) = 0.5                              f(P(C)) = 0.026
```

The property of the proposed function maintains the weight of the most likely situation while ensuring the transformed values add up to one.

4.3   Qualitative assessment

It is important to realize that the actual value of $(x_i)^z$ is meaningless. However, we used the transformed values in our tool as a baseline for a color coding for the likelihood of a

situation (e.g. red means *impossible*, green means *absolute certainty*, yellow means *possibility*, blue means *prior finished situation*). See Figure 9 for an illustration of a simulated event trace and the color coding of the different situations.

Figure 9: Visualization of a simulated event trace

While testing the Situation Studio, we found that cross-cutting work flows greatly impact the handling of events. For the blood sampling activity for example, for coordination purposes one only needs to know if the sample has been taken and if the patient is done with this activity. However, the outcome of the actual lab results is an input for later activities, but it does not impact the flow of the patient though the day.

We also compared the mathematical output and color coding with the experience of the patient coordinators, and while stepping through the trace of events the likelihood of the outcomes were similar to their expectations.

5 Conclusion

In this paper, we discussed clinical processes as non-deterministic, resembling stochastic processes with healthcare professionals being a social community with a strong desire to improve self-coordination and cooperation. We introduced and motivated the concept of situational awareness through event-based information systems, and presented preliminary tool support to capturing non-deterministic subtleties.

Note that no structured test of the system has been commenced, as one of the most important effects is understanding how putting events in a semi-structured work flow impacts the quality aspect. These effects have been the source for many discussions, which in turn resulted in iterations. This insight is valuable for understanding the effects of compiling events into complex events.

The Situation Studio will be further developed in the frame of the FP7 BUTLER project which investigates the challenges emerging from the recent ICT advances shaping up the

Internet-of-Things (IoT) that will connect sensors, actuators and smart portable devices into an ecosystem reaching an estimated 50 billion devices by 2015-2020. The sheer amount of information and the means to discover and benefit from it will be so vast that humans, no matter how technology-savvy, will not be able to handle it on their own. In BUTLER, the Situation Studio will be used to manage context-aware event processing and pattern recognition techniques to extract meaningful information and transform a stream of events into a representation of human intent for anticipating common user behavior.

6 Acknowledgments

References

[BFS11]   Berit Brattheim, Arild Faxvaag, and Andreas Seim. Process support for risk mitigation: a case study of variability and resilience in vascular surgery. *BMJ Quality & Safety*, February 2011.

[End95]   M.R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):32–64, 1995.

[GM01]   S. Glouberman and H. Mintzberg. Managing the care of health and the cure of disease-Part II: Integration. *Health care management review*, 26(1):70, 2001.

[HTS06]   Minmin Han, Thomas Thiery, and Xiping Song. Managing exceptions in the medical workflow systems. In *Proceedings of the 28th international conference on Software engineering*, ICSE '06, pages 741–750, New York, NY, USA, 2006. ACM.

[LS08]   David C. Luckham and Roy Schulte. Event Processing Glossary - Version 1.1. Online Resource. http://complexevents.com/2008/08/31/event-processing-glossary-version-11/, July 2008. Last visited: October 2009.

[RRvdGK09]   Hajo A. Reijers, Nick Russell, Simone van der Geer, and Gertruud A. M. Krekels. Workflow for Healthcare: A Methodology for Realizing Flexible Medical Treatment Processes. In *Business Process Management Workshops*, pages 593–604, 2009.

[WPL$^+$11] Leendert W. M. Wienhofen, Davy Preuveneers, Andreas D. Landmark, Pieter J. Toussaint, and Yolande Berbers. A notion of event quality for contextualized planning and decision support systems. In *Proceedings of the 7th international and interdisciplinary conference on Modeling and using context*, CONTEXT'11, pages 307–320, Berlin, Heidelberg, 2011. Springer-Verlag.