Journal of Biomedical Informatics 83 (2018) 10–24

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Leveraging semantic labels for multi-level abstraction in medical process mining and trace comparison

Giorgio Leonardi[a], Manuel Striani[b], Silvana Quaglini[c], Anna Cavallini[d,1], Stefania Montani[a,*]

[a] DISIT, Computer Science Institute, Università del Piemonte Orientale, Viale Michel 11, I-15121 Alessandria, Italy
[b] Department of Computer Science, Università di Torino, Italy
[c] Department of Electrical, Computer and Biomedical Engineering, Università di Pavia, Via Ferrata 1, I-27100 Pavia, Italy
[d] Istituto di Ricovero e Cura a Carattere Scientifico Fondazione "C. Mondino", Via Mondino 2, I-27100 Pavia, Italy

ARTICLE INFO

Keywords:
Abstraction
Trace comparison
Semantic process mining
Stroke management

ABSTRACT

Many medical information systems record data about the executed process instances in the form of an *event log*. In this paper, we present a framework, able to convert actions in the event log into higher level concepts, at different levels of abstraction, on the basis of domain knowledge. Abstracted traces are then provided as an input to trace comparison and semantic process discovery. Our abstraction mechanism is able to manage non trivial situations, such as interleaved actions or delays between two actions that abstract to the same concept. Trace comparison resorts to a similarity metric able to take into account abstraction phase penalties, and to deal with quantitative and qualitative temporal constraints in abstracted traces. As for process discovery, we rely on classical algorithms embedded in the framework ProM, made *semantic* by the capability of abstracting the actions on the basis of their conceptual meaning. The approach has been tested in stroke care, where we adopted abstraction and trace comparison to cluster event logs of different stroke units, to highlight (in)correct behavior, abstracting from details. We also provide process discovery results, showing how the abstraction mechanism allows to obtain stroke process models more easily interpretable by neurologists.

1. Introduction

Today's medical information systems log enormous amounts of data, including details about the actions that have been executed at a given organization. Such data collections can be provided in the form of *event logs* [1], which maintain the sequences (*traces* [1] henceforth) of actions that have been completed, identified by their *names*, together with their *timestamps* and possible additional data elements.

Event logs are exploited by **process mining** [1], a term that encopasses a family of a posteriori analysis techniques. Basically, event logs can be used to feed and run four types of process mining tasks [2,3,1]:

- *discovery*. A discovery technique takes as input the event log and produces a process model. Discovery is the most relevant and widely used process mining activity;
- *conformance*. A conformance technique takes as input the event log and an existing process model: the model is compared to the log of the same process, to measures the alignment (i.e., the conformance) between the model and reality;
- *enhancement*. An enhancement technique takes as input the event log and an existing process model as well. It aims at improving the existing model using information about the actual process recorded in the log; in fact, if needed, the a priori model can be changed, to better mirror the log data;
- *operational support*. An operational support technique does not operate off-line, as in the previous cases. On the contrary, it is used to influence the running process instance by checking, predicting, or recommending actions to be executed. It typically resorts to the comparison and/or analysis of past traces similar to the running one.

The action *names* maintained in the traces in the event log are strings without any semantics, so that identical actions, labeled by synonyms, will be considered as different, or actions that are special cases of other actions will be processed as unrelated by the process mining techniques illustrated above.

On the other hand, the capability of relating *semantic structures* such as ontologies to actions in the log can enable all such tasks to work at *different levels of abstraction* (i.e., at the level of instances and/or concepts) and, therefore, to mask irrelevant details, to promote reuse, and,

* Corresponding author.
E-mail address: stefania.montani@uniupo.it (S. Montani).
[1] On behalf of the Stroke Unit Network (SUN) collaborating centers.

https://doi.org/10.1016/j.jbi.2018.05.012
Received 28 November 2017; Received in revised form 16 May 2018; Accepted 19 May 2018
Available online 21 May 2018
1532-0464/

Fig. 1. Traces showing the treatment of two stroke patients in two different hospitals.

in general, to make process mining much more flexible and reliable.

In fact, it has been observed that human readers are limited in their cognitive capabilities to make sense of large and complex process models [4,5], while it would be often sufficient to gain a quick overview of the process, in order to familiarize with it in a short amount of time. One well-known way to address this issue is by applying *process model abstraction* [6], thus retaining the essential properties of the model on a particular level of analysis, while simultaneously hiding insignificant details for that level. Of course, deeper investigations can still be conducted, subsequently, on the detailed (ground) process model.

Interestingly, *semantic process mining*, defined as the integration of semantic processing capabilities into classical process mining techniques, has been proposed in the literature since the first decade of this century (see, e.g., [7,8], and Section 6). However, while more work has been done in the field of semantic conformance checking [7,9], to the best of our knowledge semantic process discovery and semantic-based trace analysis for operational support need to be further investigated.

In this paper, we present a **semantic-based, multi-level abstraction mechanism**, able to operate on event log traces. In our approach:

- actions in the event log are related to the medical goals they are aimed to fulfill, by means of an ***ontology***;
- a ***rule base*** is exploited, in order to identify which of the multiple goals of an action in the ontology should be considered as the correct abstraction of the action itself.

The abstraction mechanism is then provided as an input to further analysis mechanisms, namely **trace comparison** and **process discovery**.

The methodological approach has been tested in the field of stroke patient management, where we have adopted our framework for trace clustering and process model discovery. In the first case, we have verified that it is possible to obtain more homogeneous clusters, abstracting from details such as local resource constraints or local protocols, but still preserving the capability of isolating outlying situations; in the second experiment, we have mined more readable process models, where unnecessary details are hidden, but key behaviors are clear.

The paper is organized as follows. Section 2 introduces a motivating example in the field of stroke care. Section 3 presents the terminology that will be adopted in the paper. Section 4 presents methodological and technical details of the framework. Section 5 describes experimental results. Section 6 addresses comparisons with related works and discusses the novelty of the approach. Finally, Section 7 is devoted to conclusions and future research directions.

2. The need for multi-level abstraction in stroke care

In this section, we present our methodology in the context of a real-life scenario, where we aim at comparing two process traces executed at two different hospitals for the treatment of patients with ischemic stroke. The traces in Fig. 1 show the sequence of actions performed by hospital T1 and T2, respectively, to treat two patients admitted in their emergency department and potentially affected by ischemic stroke in the acute phase. In hospital T1, the patient firstly undergoes a contrast-enhanced Computed Tomography scan (CAT-C), then a Diabetologist Counseling (DC), then a Magnetic Resonance (MRI). After these examinations, the patient is treated with an Endovascular Procedure (EP) and with the administration of Anti-Aggregants (AAG). Hospital T2, instead, performs a different sequence on its patient: it starts with an MRI, then it continues by performing the CAT-C, then treating the patient with thrombolysis (tissue plasminogen activator – TPA) and finally by administering an Anticoagulant Oral Therapy (AOT).

Apparently, the traces in Fig. 1 describe two very different behaviors, considering both the actions and their order. However, the Italian guidelines for the treatment of stroke (ISO-SPREAD guidelines[2]), offer the domain knowledge for a more abstract interpretation.

According to the guidelines, in the acute phase some of the goals to be pursued are the identification of the pathogenetic mechanism of the ischemic stroke and its localization, the reduction of the brain damage through recanalization therapies, if possible, and the secondary prevention to lower the risk of recurrence. The pathogenetic mechanism and localization of stroke can be investigated through neuro-imaging tests such as a CAT-C scan or an MRI, potentially executed both, in any order, for a deeper analysis. The main therapy to reduce brain damage is the TPA drug, which tries to dissolve the thrombus and to restore the brain circulation in order to mitigate damages from ischemia, but mechanical revascularization through endovascular procedures could be performed, if necessary and if the patient is not eligible for TPA. Early relapse prevention can be obtained through the administration of antithrombotic drugs, mainly antiaggregants or anticoagulants.

Applying this knowledge to the traces in Fig. 1, it is possible to obtain a higher level interpretation of the operations performed by the two hospitals, by means of abstracted traces that show the therapeutic and diagnostic goals instead of the ground actions. This "bird's eye view" allows to perform a comparison that ignores unnecessary details. In trace T1, CAT-C and MRI are merged into a single macro-action PE (Parenchima Examination); EP abstracts to a macro-action RT (Recanalization Therapy); AAG becomes an ERP (Early Relapse Prevention) macro-action. It is worth noting that in the trace T1, a DC is also executed. This action, however, is not part of any of the main goals to be achieved in the management of the acute phase, therefore it remains an isolated action, performed during PE. Regarding T2, MRI and CAT-C are abstracted as PE; TPA is abstracted as RT, while AOT is abstracted as ERP. The result of the abstraction process on the traces T1 and T2 is shown in Fig. 2.

Comparing T1 and T2 at this higher level of abstraction, we can observe that they are indeed very similar: ground differences do exist, but they can be interpreted as minor variations of two process instances, which share the same goals, and are thus basically compliant with the guideline indications.

Interestingly, the abstraction technique is not only able to shed the light on the compatibility of apparently different behaviors, but is also able to preserve significant differences. For example, the abstracted traces in Fig. 2 show even more clearly than the ones in Fig. 1, that hospital T1 performs a single DC action not in line with the goals suggested by guideline, while hospital T2 is compliant with the guideline recommendations. This situation suggests an investigation of the causes for this non-compliance to be carried out by the T1 managers, who can conclude, for example, that the additional DC is driven by an excess of precaution: it could be considered as a waste of resources and time, if not even a damage for the patient, since other

[2] http://www.iso-spread.it/index.php, last accessed on 11/09/2017.