Artificial Intelligence in Medicine 54 (2012) 15–27

Contents lists available at SciVerse ScienceDirect

Artificial Intelligence in Medicine

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

Similarity metrics for surgical process models

Thomas Neumuth [a,*], Frank Loebe [b,c], Pierre Jannin [d,e,f]

[a] Innovation Center Computer Assisted Surgery (ICCAS), University of Leipzig, Semmelweisstr. 14, D-04103 Leipzig, Germany
[b] Department of Computer Science, University of Leipzig, Johannisgasse 26, D-04103 Leipzig, Germany
[c] Institute of Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig, Haertelstr. 16-18, D-04107 Leipzig, Germany
[d] INSERM, U746, Faculté de Médecine, 2, Avenue du Pr. Léon Bernard, CS 34317, 35043 Rennes Cedex, France
[e] INRIA, VisAGeS Unit/Project, 2, Avenue du Pr. Léon Bernard, CS 34317, 35043 Rennes Cedex, France
[f] University of Rennes I, CNRS, 2, Avenue du Pr. Léon Bernard, CS 34317, 35043 Rennes Cedex, France

A R T I C L E   I N F O

Article history:
Received 7 May 2010
Received in revised form 18 August 2011
Accepted 4 October 2011

Keywords:
Medical Informatics
Weights and measures
Workflow
Decision making, computer-assisted
Surgery, computer-assisted
Surgical process model

A B S T R A C T

Objective: The objective of this work is to introduce a set of similarity metrics for comparing surgical process models (SPMs). SPMs are progression models of surgical interventions that support quantitative analyses of surgical activities, supporting systems engineering or process optimization.
Methods and materials: Five different similarity metrics are presented and proven. These metrics deal with several dimensions of process compliance in surgery, including granularity, content, time, order, and frequency of surgical activities. The metrics were experimentally validated using 20 clinical data sets each for cataract interventions, craniotomy interventions, and supratentorial tumor resections. The clinical data sets were controllably modified in simulations, which were iterated ten times, resulting in a total of 600 simulated data sets. The simulated data sets were subsequently compared to the original data sets to empirically assess the predictive validity of the metrics.
Results: We show that the results of the metrics for the surgical process models correlate significantly ($p < 0.001$) with the induced modifications and that all metrics meet predictive validity. The clinical use of the metrics was exemplarily, as demonstrated by assessment of the learning curves of observers during surgical process model acquisition.
Conclusion: Measuring similarity between surgical processes is a complex task. However, metrics for computing the similarity between surgical process models are needed in many uses in the field of medical engineering. These metrics are essential whenever two SPMs need to be compared, such as during the evaluation of technical systems, the education of observers, or the determination of surgical strategies. These metrics are key figures that provide a solid base for medical decisions, such as during validation of sensor systems for use in operating rooms in the future.

1. Introduction

1.1. Objectives and motivation

The objective of this work is to introduce a set of similarity measures for comparing SPMs and to describe an approach for experimental validation of these measures.

Surgical process models (SPMs) are progression models of surgical interventions that are used in a variety of cases, including the optimization and evaluation of computer-assisted surgery systems and requirements engineering. Because surgical interventions, which can be more abstractly defined as surgical processes, cannot be analyzed directly, SPMs are used as computable process models that allow for quantitative analysis. As formal representations of surgeons' activities collected from patient data, SPMs have a great deal of potential in surgical education and training.

Similarity metrics are required for quantitatively describing the similarities among multiple SPMs. Such comparisons can serve as an advanced method for representing surgical treatment strategies [1,2], a retrospective evaluation of surgical assist systems [3], assessment of a surgeon's expertise [4], the basis for requirements engineering [5], or the evaluation of the accuracy of SPM acquisition systems, as shown in this paper with human observers [6,7].

The research question is "How can we quantify the similarity between two surgical process models?". To answer this question, we introduced a measurement system that aims to study different SPM dimensions, such as granularity, content, time, order, and frequency. For each of the metrics, mathematical proofs and an experimental validation were performed. Validation studies were based on simulations of real clinical cases from three different types of surgeries from different surgical disciplines. In addition to

* Corresponding author. Tel.: +49 341 97 12010; fax: +49 341 97 12009.
E-mail addresses: thomas.neumuth@iccas.de, thomas.neumuth@medizin.uni-leipzig.de (T. Neumuth).

0933-3657/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.artmed.2011.10.001

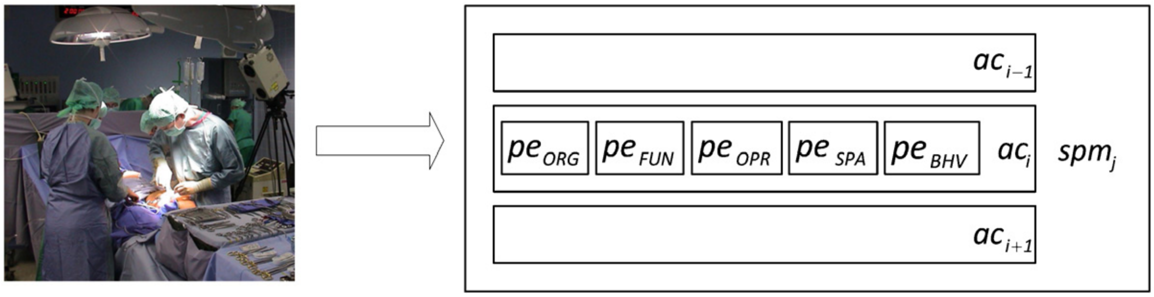

Fig. 1. General principle of using SPMs to present surgical work steps: an SPM spm_j consists of activities ac_i, and each activity consists of several perspectives $pe_p \in PE_p^{val} (p \in PE)$.

experimental validation, we demonstrated the value of our similarity measures by using them to report on the learning progress of clinical observers that recorded the SPMs. These topics have not yet been considered in the literature and may serve as a basis for future work in this field and in affiliated sectors of research.

Modeling surgical processes is a complex task. Established sources of intraoperative knowledge that may provide information about surgical processes, primarily surgical textbooks or clinical guidelines [8,9], have major constraints that engender the use of SPMs. They are either designed for representing expert knowledge in a top-down-modeling approach, but they do not cover concepts that are necessary for quantification, such as temporal constraints, or they are not able to cope with high inter-patient and inter-surgeon variability in the surgical processes. Because no suitable models exist, no metrics exist.

Approaches for modeling surgical processes have gained recent interest in the literature. Entire SPMs have been modeled by several groups in several application contexts, such as surgical education and surgeon training [10,11], image-guided surgery [12–14], context-driven user interfaces [15,16], the evaluation of surgical instruments [17,18], the performance of requirements analyses [5], and other assessments of surgical strategies or auxiliaries [1,2,19,20]. However, none of the approaches dealt with SPM metrics. None of these models used similarity metrics to compare different surgical processes, even though some of them used statistical approaches that are indirectly related to similarity metrics, such as hierarchical clustering [14]. Recently, similarity measures have been introduced by Combi et al. [21] in the context of clinical workflows. However, the similarity measures were restricted to temporal information and did not consider other dimensions, such as granularity.

In the framework of information systems theory, several studies have been presented in recent years. Bae et al. [22] presented the measurements of similarities between binary trees in business process models. They introduced δ-comparability and a structural comparison of process blocks. In another work, similarities in processes were assessed by subtracting network matrices [23,24]. Furthermore, van Dongen et al. [25] derived predecessor–successor relations as “causal footprints” from event-driven process chains (EPCs) and introduced similarity measures for these relationships. The measures were judged by facial validity.

van der Aalst introduced number-based fitness measures between traces of event logs for EPCs [26] and, in later works, equivalence structures for Petri nets [27,28]. A review of metrics related to process mining can be found elsewhere [29]. However, none of these existing metrics is focused on surgical process models. Furthermore, they do not consider inputs that differ because of changes in patient-specific properties or treatments that differ because of variation in surgical experience or available surgical technology. Finally, a metric for SPMs must be clinically meaningful because the results need to be used by the surgeon as a clinical end user [30], who must interpret the SPM results with a clinical perspective.

In this paper, we start with a brief introduction of contextual terms and definitions and then present the metrics, first on a general level and then on a formal level. Subsequently, metric properties are mathematically proven and experimentally validated using 20 clinical data sets each from cataract interventions, craniotomy interventions, and supratentorial tumor resections. Finally, we demonstrate the clinical utility of the metrics by assessing the learning curves of observers during surgical process modeling acquisition.

1.2. Contextual terms and definitions

A surgical process model represents a surgical process (SP) in the real world as a set of eventualities, which is a general term for (parts of) processes and processual entities [6,31]. Here, we focus on surgical work steps in SPs and define their representations in SPMs as *activities*. Thus, each activity in an SPM is associated with a surgical work step in the underlying SP. When aligning our terminology to that of the Workflow Management Coalition [31], our “activities” correspond to their “activity instances” [32].

Surgical processes that are performed with the same surgical objectives and the same strategies have high variability. This variability is caused by the use of different surgical technologies or by patient-specific properties in anatomy and pathology. To represent this variability, we introduce the concept of perspectives into our SPMs. Jablonski and Bussler [33] introduced the use of perspectives to differentiate among several aspects of activities for workflow management systems. We use perspectives in our application context to decompose activities into more fine-grained entities. Five different perspectives are distinguished (cf. also Fig. 1):

- the functional perspective (*FUN*) describes what is done in a work step,
- the organizational perspective (*ORG*) describes who performs the work step,
- the operational perspective (*OPR*) describes which instruments or devices are used to perform the work step,
- the spatial perspective (*SPA*) describes the location on the patient’s body where a work step is performed, and
- the behavioral perspective (*BHV*) describes time information.

The perspectives express who was doing what, where, how and when (cf. Table 1) for each work step in the surgical process. A number of activities together constitute the surgical process model.

Formally, SPMs, activities and perspectives are captured as follows, starting from the set of perspectives relevant to our purposes. That is, $PE = \{FUN, ORG, OPR, SPA, BHV\}$, where BHV is called the behavioral perspective and all remaining perspectives are defined as nominal perspectives, i.e., those in $PE \backslash \{BHV\}$. Each perspective is associated with a set of possible values, denoted by PE_{OPR}^{val} for each perspective $p \in PE$. These value sets are required to be mutually pairwise disjoint. Perspectives span the space of possible activities