Surgical process modelling for procedural skill assessment

Germain Forestier[1] and Pierre Jannin[2]

[1] MIPS, University of Haute-Alsace, Mulhouse, France
[2] INSERM MediCIS, Unit U1099 LTSI, University of Rennes 1, Rennes, France

Abstract. Surgery is one of the riskiest and most important medical acts that is performed today. Desires to improve surgeon training and patient outcomes motivate a better understanding of surgical practices. To facilitate this, surgeons have started recording the activities that are performed during surgery. New data science methods have to be developed to be able to make the most of this extremely rich and complex data. The objective of this paper is to introduce how sequence analysis has been used recently to address the assessment of surgical activities. Understanding the ways in which surgeries are similar or different from each other is of major interest to automatically classify and analyse them. We also present the main challenges that this field would have to face in the coming years.

Keywords: surgical process modelling, sequence analysis, procedural knowledge assessment

1 Introduction

More than half a million surgeries are performed every day worldwide, which makes surgery one of the most important component of global health care. Competing demands are motivating a better understanding of surgical processes: the number of patients is continuously growing, surgical procedures are getting more complex, residents now have to be trained while performing less procedures, the surgical interventions have to be more and more justified and the procedures have to cost less money. Surgical Process Modelling (SPM) [12] is the general process that aims at understanding surgeries, in order to improve the quality of care, the training and assessment of surgeons. In order to better understand and analyse surgical behaviours, a special interest has been given to the recording of activities performed by surgeons. This data is complex and extremely challenging to analyse as it is composed of multidimensional sequences of the actions that are performed over the course of the surgery. As surgical interventions can last hours, each recording is composed of thousands of surgical activities. Consequently, new data mining and machine learning algorithms have to be designed to align with the specific characteristics of this data. Massive investments have been made to collect the data about the course of surgeries, especially in the

robotics field [2,20]. However, this data will be valuable and useful if and only if we can extract knowledge from it.

In recent years, *sequence analysis* became popular to automatically analyse sequences of surgical activities performed by surgeons. We refer to *sequence* as an ordered collection of objects or events. The application of sequence analysis to surgical activities has multiple objectives, like the automatic detection of the expertise of the surgeons (*e.g.*, junior vs. senior). The assumption is that the sequencing of the activities and their relative information (*e.g.*, duration, characteristics, etc.) are carrying enough information to precisely identify specific surgical behaviours. This process is particularly important to monitor skills acquisition, to provide feedback during learning phase and to better understand how surgical skills are transmitted. Relying on recordings instead of human observer to assess surgical skills allows to aim for the so called *objective evaluation* [22] and to automate the evaluation process [17]. However, all these objectives relate to the assessment of *procedural knowledge* which only partially covers the skill-set required to master surgery (see Table 1).

Technical Skills			Non-Technical Skills	
Motor Skills / Dexterity	Conceptual Knowledge	Procedural Knowledge	Cognitive Skills	Interpersonal skills

Table 1. Table of surgical skills [1].

In this article, we describe recent achievements in the field of sequence analysis for evaluating surgical procedural knowledge. We also present the most important challenges in this field for the coming years.

2 Sequence analysis

The majority of research efforts related to sequence analysis for surgical activities adopted the activity representation of triplet : *action*, *anatomical structure* and *instrument* [14]. Surgeries can be recorded either using sensor devices (*e.g.*, cameras) [3], or directly by an observer in the operating room (OR) using a dedicated software [18]. The first challenge to be tackled has been the design of measures allowing to compare two sequences of surgical activities in a consistent way. For a measure to be *consistent*, it has to provide a graduated evaluation of how similar two surgeries are. Only computing statistics on the number or activities or the mean duration was not satisfactory. Indeed, if the exact same activities were performed in a random order, the evaluation would have been the same. Thus, methods allowing to take into account the sequentiality of the activities were required. However, as sequences exhibit specific features (*e.g.*, not the same duration, break down in phases, etc.) new metrics had to be designed.

In one of the first work on this topic, Neumuth et al. [16] proposed five similarity metrics for comparing surgical process models (SPMs): granularity, content, temporal, transitional and transition frequency similarity. The different metrics

were tested with regards to the introduction of noise in the SPMs. However, not all of these metrics took into account the sequencing of the activities. To consider the sequencing of the activities and to reduce time differences, Forestier et al. [5] proposed to use Dynamic Time Warping (DTW) to compare sequences of activities. They used DTW alignment score as a way to assess the similarity between two sequences. This metric was then combined with hierarchical clustering in order to create groups of highly similar surgeries [6]. In the same spirit, Neumuth et al. [19] studied distance metrics for surgical process models. They compared four distance measures: the Jaccard distance, Levenshtein distance, Adjacency distance, and Graph matching distance. From their experiments, they concluded that Levenshtein and Adjacency distances were best suited to measure distances between surgical processes. In this work, they only considered the functional perspective (*i.e.*, the actions), but they mentioned the potential of using additional information like anatomical structure and instrument. Alternatively, Sugino et al. [21] proposed to use the Needleman-Wunsch (NW) algorithm to compare sequences of surgical activities. They also used the three components representation for the activities with an additional information of progress rate.

Multiple sequence alignment using NW was also explored by Bouarfa et al. [4] where they analysed a set of sequences of surgical tools used in laparoscopic surgery. In the following of these works, a new method called *Non-Linear Temporal Scaling* (NLTS) [7] was proposed to register multiple surgeries on their intrinsic timeline. Thus, it makes possible to compare multiple interventions and to highlight heterogeneities in surgical practice. Recent works on sequence analysis for surgery are focusing on the prediction of surgical activities [8,9].

3 Discussion

While the use of sequence analysis allowed to improve the information that can be extracted from SPM, several challenges still remain:

Vocabulary - The comparison of sequences of surgical activities is only relevant if the same vocabulary is used to record them (*i.e.*, list of terms for actions, anatomical structures and instruments). If the vocabularies are different, it will artificially increase the differences between the sequences. This problem was partially discussed in [6], where two acquisition sites had two different vocabularies. The proposed ad-hoc solution consisted in manually matching the vocabularies. A more robust solution lies in the development of ontologies dedicated to surgery and SPM [10] and standardized modelling languages [15].

Multi-dimensionality - As already mentioned, the surgical activities are generally composed of three components that have to be considered in the comparison. To simplify the analysis, the components are sometimes considered separately [7,19] or combined using a weighting scheme [5]. Furthermore, both hands of the surgeon are generally recorded [5,6,8] and the optimal and simultaneous use of these information is still an open question.

Handling time - Each activity, in addition to its three components, has a duration. The way this duration is used to compare sequences can greatly in-

fluence the results. It can be totally ignored if the method focuses only on the sequencing of the activities [16]. Its influence can also be reduced, for example by using DTW [5]. It can also be part of the analysis, by considering that two activities of different durations are different, for example with the Levenshtein distance [19] or the Needleman-Wunsch algorithm [21].

Benchmark - The lack of benchmarks available to compare the metrics prevent from a rigorous and precise comparison between the different propositions in the field. A repository of datasets should be available on the same model than UCR Archive for timeseries or the UCI Machine Learning Repository for classification and clustering problems. This kind of initiative is mandatory for the development of competitive methods in this field.

Clinical validation - The findings using sequence analysis for procedural knowledge still remain to be correlated with existing skills evaluation methodology like the objective structured assessments of technical skills (OSATS) [13]. A comparison between sequence analysis tools and skills evaluated using the OSATS would increase the value of automatic evaluation. As one of the goals is to perform an objective assessment [22] of procedural skills, the correlation with existing metrics would increase the impact of these methods.

Bias - As mentioned in the introduction, using sequence analysis only allows to evaluate procedural knowledge. However, many other factors can influence the way a surgery is performed: context of the OR, unexpected events, etc. Consequently, the differences observed between the sequences can be the consequence of many factors. The best way to obtain robust and relevant results will lie into the increase of data and involvement of multiple medical centres. Furthermore, sequences should be available with their meta-data providing rich information about the context of the acquisition.

4 Conclusion

In this paper, we presented how sequence analysis has been used in recent years for procedural skills evaluation. We reviewed the main existing approaches and described their most important features. Finally, we listed the main challenges that this field will have to address in the coming years.

References

1. Alderson, D.: Developing expertise in surgery. Medical teacher 32(10), 830–836 (2010)
2. Blum, T., Padoy, N., Feußner, H., Navab, N.: Workflow mining for visualization and analysis of surgeries. International Journal of Computer Assisted Radiology and Surgery 3(5), 379–386 (2008)
3. Bouarfa, L., Jonker, P., Dankelman, J.: Discovery of high-level tasks in the operating room. Journal of Biomedical Informatics 44(3), 455–462 (2011)
4. Bouarfa, L., Dankelman, J.: Workflow mining and outlier detection from clinical activity logs. Journal of biomedical informatics 45(6), 1185–1190 (2012)

5. Forestier, G., Lalys, F., Riffaud, L., Trelhu, B., Jannin, P.: Classification of surgical processes using dynamic time warping. Journal of Biomedical Informatics 45(2), 255–264 (2012)
6. Forestier, G., Lalys, F., Riffaud, R., Collins, L., Meixensberger, J., Wassef, S.N., Neumuth, T., Goulet, B., Jannin, P.: Multi-site study of surgical practice in neurosurgery based on surgical process models. Journal of Biomedical Informatics 46(5), 822–829 (2013)
7. Forestier, G., Petitjean, F., Riffaud, L., Jannin, P.: Non-linear temporal scaling of surgical processes. Artificial Intelligence in Medicine 62(3), 143–152 (2014)
8. Forestier, G., Petitjean, F., Riffaud, L., Jannin, P.: Optimal sub-sequence matching for the automatic prediction of surgical tasks. In: AIME 15th Conference on Artificial Intelligence in Medicine. vol. 9105, pp. 123–132. Springer (2015)
9. Franke S, Schreiber E, N.T.: Prediction of surgical activities for context-aware workflow management in the operating room. In: M2CAI-Modelling and Monitoring of Computer-Assisted Intervention (2015)
10. Gibaud, B., Penet, C., Jannin, P.: Ontospm: a core ontology of surgical procedure models. SURGETICA (2014)
11. Lalys, F., Bouget, D., Riffaud, L., Jannin, P.: Automatic knowledge-based recognition of low-level tasks in ophthalmological procedures. International Journal of Computer Assisted Radiology and Surgery 8(1), 39–49 (2013)
12. Lalys, F., Jannin, P.: Surgical Process Modelling: a review. International Journal of Computer Assisted Radiology and Surgery 9(3), 495–511 (2014)
13. Martin, J., Regehr, G., Reznick, R., MacRae, H., Murnaghan, J., Hutchison, C., Brown, M.: Objective structured assessment of technical skill (OSATS) for surgical residents. British Journal of Surgery 84(2), 273–278 (1997)
14. Mehta, N., Haluck, R., Frecker, M., Snyder, A.: Sequence and task analysis of instrument use in common laparoscopic procedures. Surgical endoscopy 16(2), 280–285 (2002)
15. Neumann, J., Vinz, S., Neumuth, T.: Surgical workflow and process modeling - an evaluation of modeling languages and process modeling tools. In: M2CAI-Modelling and Monitoring of Computer-Assisted Intervention (2015)
16. Neumuth, T., Loebe, F., Jannin, P.: Similarity metrics for surgical process models. Artificial Intelligence in Medicine 54(1), 15–27 (2012)
17. Reiley, C.E., Lin, H.C., Yuh, D.D., Hager, G.D.: Review of methods for objective surgical skill evaluation. Surgical endoscopy 25(2), 356–366 (2011)
18. Riffaud, L., Neumuth, T., Morandi, X., Trantakis, C., Meixensberger, J., Burgert, O., Trelhu, B., Jannin, P.: Recording of surgical processes: a study comparing senior and junior neurosurgeons during lumbar disc herniation surgery. Neurosurgery 67, 325–332 (2010)
19. Schumann, S., Bühligen, U., Neumuth, T., et al.: Distance measures for surgical process models. Methods Inf Med 52(5), 422–431 (2013)
20. Speidel, S., Delles, M., Gutt, C., Dillmann, R.: Tracking of instruments in minimally invasive surgery for surgical skill analysis. In: Medical Imaging and Augmented Reality, pp. 148–155. Springer Berlin Heidelberg (2006)
21. Sugino, T., Kawahira, H., Nakamura, R.: Surgical workflow modelling and monitoring. International Journal of Computer Assisted Radiology and Surgery 9(1), S117–S122 (2014)
22. Van Hove, P., Tuijthof, G., Verdaasdonk, E., Stassen, L., Dankelman, J.: Objective assessment of technical surgical skills. British Journal of Surgery 97(7), 972–987 (2010)