

A study of the application of computational intelligence and machine learning techniques in business process mining

Ana Rocío Cárdenas Maita - *USP (Brasil) and UNSAAC (Perú)*

1 Abstract

Mining process is a emerging research area that combines data mining and machine learning, on one hand, and business process modeling and analysis, on the other hand. Mining process aims at discovering, monitoring and improving business processes by extracting real knowledge from event logs produced by the information systems used by organizations. This work aims to assess the application of computational intelligence and machine learning techniques in process mining context. The main focus of the study was to identify why the computational intelligence and machine learning techniques are not being widely used in process mining field and identify the main reasons for this phenomenon.

2 Introduction and background

Business Process Management (BPM) involves managing the entire life-cycle of business processes, with the aid of different techniques, methods and supporting tools. In the last stage of the BPM life-cycle – assessment and improvement of process models– the history of the execution and monitoring of the instances of a business process can be assessed so that the process can be optimized [1].

Data mining entails identifying patterns and relationships hidden in a large amount of data; it allows new information and useful knowledge to be extracted so that it can support the decision-making process in an appropriate way [2]. Moreover, data mining can be seen as a stage in a larger process called Knowledge Discovery from Data (KDD) [2]. In KDD, data from a specific context is collected, transformed and arranged before undergoing the mining process. After the mining, the results must be arranged in a structure that is accessible to direct human interpretation.

The combination of BPM and data mining have established a new research field, known as process mining [3]. This entails undertaking data mining tasks with data from the BPM life-cycle. Its goal is to extract knowledge about events/data from the work carried out in the different stages of BPM, and thus seek to improve business processes, by discovering links between variables and behavioral patterns [3]. Its goal is to scan event logs – available in process-aware information systems – to extract knowledge that allows process discovery, monitoring and improvement to be addressed [3].

In this scenery the goal of this work was understand in width the process mining area; including the main features that have been found in terms of: types of process mining, data mining tasks used, and techniques applied to solving such tasks.

3 Research problem and motivation

Through an initial exploratory assessment, it is found that, although process mining is a BPM junction and data mining, many techniques and algorithms widely and satisfactorily used in

the general context of data mining are rarely used in the specific context of process mining, more specifically, computational intelligence techniques as Artificial Neural Networks or Genetic Algorithms. Unlike of dataset commonly used in data mining, in the context of process mining datasets are generate by logs that represents the execution flow of unstructured processes. This processes has a weak causal dependence on its activities, i.e., these flows largely depend on the occasional decisions made by their participants, which makes the execution of the instances different from each other. This high degree of irregular behavior leads to considerable complexity and represents a challenge to process mining.

4 Technical contribution

Results are valuable for scientists and practitioners related to process mining. With the aim of contextualizing this research field, we conducted an extensive systematic mapping study on process mining, which involved assessing 705 papers published over a ten-year period [4]. The main purpose was to discuss the types of process mining and explore how data mining tasks and techniques have been applied. According to the systematic mapping produced by our research group [4], the majority of research being conducted on process mining (71%) has addressed the question of process discovery. This can be regarded as natural since process conformance and enhancement rely on a process model. Unstructured processes, for example, are still poorly explored and difficult to solve when it comes to discovering process models that represent processes in an abstract light that are useful and easy for people to understand.

According to the systematic mapping conducted by our research group [4], most of the techniques being applied in process mining (81%) are traditional; only 19% of them are from the areas of computational intelligence or machine learning. Regarding artificial neural networks, for example, another study from our research group showed that only 11 studies related to process mining applied this technique until 2014 [5]. This can be demonstrated by a specific example we came across: an unstructured process followed by distance learning students in a Learning Management System (LMS), which has been explored in some experiments explained in [6]. Our grand aim is to discover process models that, on the one hand, avoid an excessive amount of useless information for the user and, on the other hand, reveal control flow structures that aggregate significant knowledge.

References

- [1] M. Weske, *Business Process Management: Concepts, Languages, Architectures*, 2nd ed. Springer, 2012.
- [2] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Waltham: Morgan Kaufmann, 2011.
- [3] W. M. P. Aalst, *Process Mining – Data Science in Action*, 1st ed. Springer, 2016.
- [4] A. R. C. Maita, L. C. Martins, C. R. L. Paz, L. Rafferty, P. C. K. Hung, S. M. Peres, and M. Fantinato, “A systematic mapping study of process mining,” *Enterprise Information Systems*, pp. 1–45, 2017, published online first.
- [5] A. R. C. Maita, L. C. Martins, C. R. L. Paz, S. M. Peres, and M. Fantinato, “Process mining through artificial neural networks and support vector machines: A systematic literature review,” *Business Process Management J.*, pp. 1391–1415, 2016.
- [6] A. R. C. Maita, M. Fantinato, S. M. Peres, L. H. Thom, and P. C. K. Hung, “Mining unstructured processes: An exploratory study on a distance learning domain,” in *Proc. of 30th Int. Joint Conf. on Neural Networks (IJCNN)*. IEEE, 2017, pp. 3240–3247.