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

Ana Rocío Cárdenas Maita
Universidade de São Paulo (Brazil)

ABSTRACT

Process mining 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. 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. The stage of experiments in this study was carried out based on an unstructured process related to a distance learning supported by a Learning Management System (LMS).

BACKGROUND

Business Process Management (BPM) involves managing the entire life-cycle of business processes. In the last stage of that cycle 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 can support the decision-making process in an appropriate way [2]. The combination of BPM and data mining have established a new research field, known as process mining [3]. 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]. The event logs must go through a preprocessing and modeling stage before applying an ML technique.

CaseId	Time	User	Activity
8826	06/18/2014 18:17	Student B	Research project
8826	06/18/2014 18:17	Student B	Portfolio
8827	06/26/2014 10:32	Student B	The teaching profession|inclusive/special education
8827	06/26/2014 10:32	Student B	The teaching profession|inclusive/special education
8827	06/26/2014 10:50	Student B	The teaching profession|inclusive/special education
8827	06/26/2014 10:50	Student B	Portfolio
8827	06/26/2014 10:50	Student B	The teaching profession|inclusive/special education
8827	06/26/2014 10:51	Student B	Video class caption
8827	06/26/2014 10:51	Student B	Video class caption
8827	06/26/2014 10:51	Student B	The teaching profession
8827	06/26/2014 10:51	Student B	Inclusive/special education
8827	06/26/2014 11:15	Student B	The teaching profession|inclusive/special education
8827	06/26/2014 11:16	Student B	Video class caption
8827	06/26/2014 11:57	Student B	Video class caption
8827	06/26/2014 12:34	Student B	The teaching profession

Fig. 1. Preprocessed log's excerpt used in experiments [6]

RESEARCH PROBLEM AND MOTIVATION

Although process mining has been considerably improved in recent years, there are some types of processes for which no satisfactory mining techniques has been proposed. However, techniques widely applied in data mining such as Computational Intelligence (CI) or Machine Learning (ML) techniques, have been fairly neglected by the process mining field [4]; as you can see in the Fig. 2.

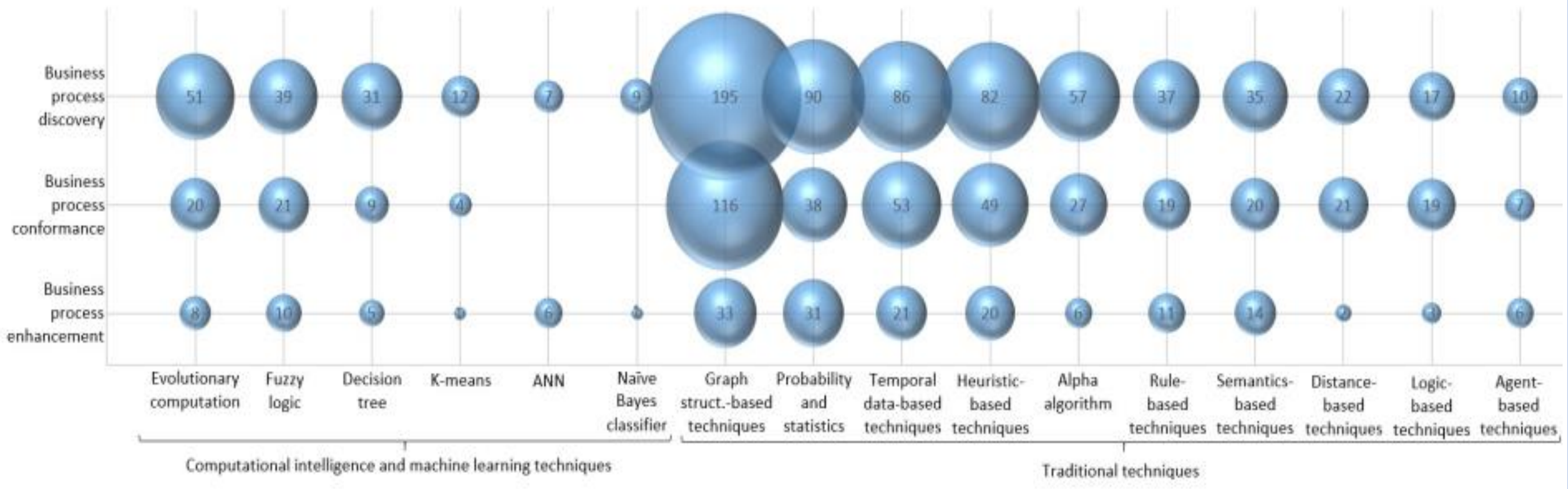

Fig. 2. Cross-distribution among process mining types and data mining techniques [4]

A dataset commonly used in the context of process mining 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.

RESULTS AND CONTRIBUTIONS

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].

A systematic mapping indicated ProM and Disco as the most commonly used process mining tools[4]. Regarding the process mining types to be used in this part of the study, we selected discovery and conformance. On the side of ML we found the categorical prediction task as the most frequently used in process mining. It was: defined as a binary problem, based on the "pass" or "fail" result, aimed to get a model able to predict the outcome for new students.

User	Category	1-2-3-4-5-6-7-9-10	2-3-4-6-7-9-10-11	2-3-4-6-10-12	2-3-5-6-10	2-3	2-3-5-6-9-10-11	2-3-6-10-11	2-4-5-6-7-9-10	2-3-4	(...)
Student 1	Passed	1	0	0	0	12	0	0	0	13	
Student 2	Passed	2	0	0	0	9	0	0	0	18	
Student 3	Passed	2	0	0	0	35	0	0	0	0	
Student 4	Passed	0	1	0	0	23	0	0	0	12	
Student 5	Passed	0	0	0	0	6	0	1	0	0	
Student 6	Passed	3	0	0	0	21	0	0	0	0	
Student 7	Passed	3	0	0	0	82	0	0	0	0	
Student 8	Failed	1	0	0	0	1	0	0	1	0	
Student 9	Passed	1	0	0	0	32	1	0	1	0	
Student 10	Passed	0	0	0	0	20	0	0	0	11	
Student 11	Passed	0	0	0	0	19	0	0	0	23	
Student 12	Passed	0	0	0	0	25	0	0	0	0	
Student 13	Passed	1	0	0	0	2	0	0	0	2	
Student 14	Passed	1	0	0	0	5	0	0	0	1	
Student 15	Passed	0	0	0	0	22	0	0	0	0	

Fig. 3. Example of the dataset A

Measure	Resubstitution	Cross validat.
# of correctly classified instances	274	263
# of incorrectly classified instances	17	28
# of true positives (successes in "pass")	261	248
# of true negatives (successes in "fail")	13	15
# of false negatives (errors in "pass")	0	13
# of false positives (errors in "fail")	17	15
Precision (or positive predictive)	0.94	0.94
Negative predictive	1.00	0.54
Recall (or true positives rate)	1.00	0.95
Specificity (true negative rate)	0.43	0.50

Fig. 4. Classifier´s evaluation measures considering two testing strategies

In addition, for classical process mining results, Disco was greatest benefit, comparing to ProM, is in terms of the provided results' usability and readability.

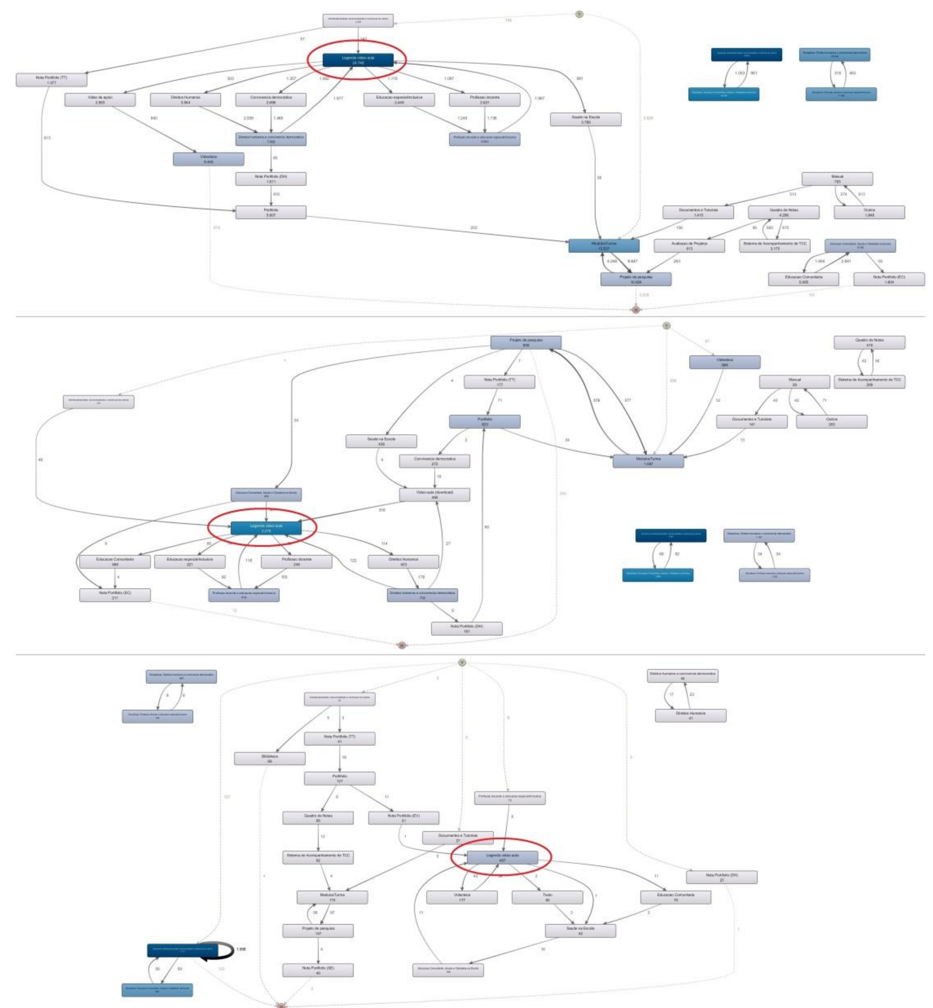

Fig. 5. Overview of the discovered process models with Disco

Actually, our research group is focusing in explore the application in process mining of other CI/ML techniques such as genetic algorithms to discovery business process.

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] Maita, A. R. C., Martins, L. C., Lopez Paz, C. R., Rafferty, L., Hung, P. C., Peres, S. M., & Fantinato, M. (2018). A systematic mapping study of process mining. Enterprise Information Systems, 12(5), 505-549.

[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 literatura review," Business Process Management J., 1391–1415, 2016.

[6] A. R. C. Maita, M. Fantinato, S. M. Peres, L. H. Thom, and P. C. Hung, "Mining unstructured processes: An exploratory study on a distance learning domain," in Proc. Of Int. Joint Conf. on Neural Networks (IJCNN). IEEE, 2017, 3240–3247.

CONTACT

ar.cardenasmaita@usp.br / ar.cardenasmaita@gmail.com
https://www.linkedin.com/in/ana-rocio-cardenas-maita-92a67bb7
http://www.aitec-cusco.org/