



Towards a Capitalization of Processes Analyzing Learning Interaction Traces

Alexis Lebis, Marie Lefevre, Vanda Luengo, Nathalie Guin

► To cite this version:

Alexis Lebis, Marie Lefevre, Vanda Luengo, Nathalie Guin. Towards a Capitalization of Processes Analyzing Learning Interaction Traces. 11th European Conference on Technology Enhanced Learning (EC-TEL 2016), Sep 2016, Lyon, France. pp.397-403, 10.1007/978-3-319-45153-4_33 . hal-01336851

HAL Id: hal-01336851

<https://hal.science/hal-01336851>

Submitted on 12 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Towards a Capitalization of Processes Analyzing Learning Interaction Traces

Alexis Lebis^{1,2}, Marie Lefevre², Vanda Luengo¹, and Nathalie Guin²

¹ Sorbonne Universités, UPMC Univ Paris 06, CNRS, LIP6 UMR 7606,
{alexis.lebis, vanda.luengo}@lip6.fr

² Université de Lyon, CNRS, Université Lyon 1, LIRIS, UMR5205
{alexis.lebis, marie.lefevre, nathalie.guin}@univ-lyon1.fr

Abstract. Analyzing data coming from e-learning environments can produce knowledge and potentially improve pedagogical efficiency. Nevertheless, TEL community faces heterogeneity concerning e-learning traces, analysis processes and tools leading these analyses. Therefore, analysis processes have to be redefined when their implementation context changes: they cannot be reused, shared nor easily improved. There is no capitalization and we consider this drawback as an obstacle for the whole community. In this paper, we propose an independent formalism to describe analysis processes of e-learning interaction traces, in order to capitalize them and avoid these technical dependencies. We discuss both this capitalization and its place and effects in the iterative learning analysis procedure.

Keywords: Learning analytics, capitalization, analysis process, interaction traces, operator, analysis tool, e-learning.

1 Introduction

The e-learning is defined by the use of digital environments that can be networked. Its aim is to reinforce the construction of knowledge by learners. These environments can produce data that relate the interaction of users among themselves (e.g. private messages, forum, etc.), with the system or even with resources. Thereafter, we name *traces* this data of learning interaction. These traces can be considered as knowledge warehouses since their analysis brings knowledge out of them. However, there is no solution to easily share, enrich or reuse such analysis processes of interaction traces nor the knowledge they are producing. Consequently, when the implementation context changes (e.g. analysis tool, formats of data used), analysis processes have to be reworked, frequently from scratch. Thus, in such a situation, TEL community cannot have an effective awareness of what is existing or what is redundant.

In this paper, we introduce our approach to bring capitalization for analysis processes. It relies on a formalism for describing analysis processes of learning interaction traces independently of analysis tools, which aims to avoid technical specificities. Moreover, we discuss the place of such analyses capitalization in the iterative learning analysis procedure, and the potential actors involved in it.

2 Related Works

Analysis Processes of Interaction Traces. An analysis process of traces is the use of operations, made by operators, over data in order to produce knowledge (e.g. indicators) addressing needs [1]. These analyses can be classified according to expected knowledge as descriptive (describe what happened), predictive (determine prospects), diagnostic (understand why something happened) or prescriptive (identify best decisions) [4]. Moreover, their inner steps have been widely covered too [4], [8], and three steps can be identified as recurrent across different fields: preprocessing, analysis of relevant data and post-processing. Other steps are more specific to the TEL field such as publication of results or reuse of data [8], giving us clues regarding capitalization needs.

Some works consider an analysis process as an organized and fixed combination of operators [6]. It can be seen as a "black box" and be reused in another analysis process [5]. This property led to methods that reinforce the importance of capitalization, such as the *discovery with models* method where previous developed models are used as components for other analyses [2].

Analysis Tools. TEL community has at its disposal a variety of cross-field analysis tools, like RapidMiner³ or R⁴, and specialized solutions. For instance, UnderTracks (UT) takes into consideration data and operators life cycle within analysis [6]. We can also cite Usage Tracking Language (UTL), which calculates and describes indicators by mapping data coming from heterogeneous traces into more generic ones expressed in XML [3]. All these tools can be classified into three categories [6]: data storage, data analysis (like R) and both data storage and analysis (like UT). Our work concerns only the tools designed for analyses.

Capitalization. Since analysis tools implement operators that are strongly dependent of data formalism in order to be computed, they are poorly permissive. As a result, some works suggest to work with a more generic data formalism before making any analysis, like Caliper Analytics⁵ or UTL. These tools map cross-origin data into a regulated formalism, allowing reproducible analyses. But the analysis capitalization is not guaranteed: they are done in a specific tool and produce specific formatted data. Such analysis processes cannot be shared and reused as they are in other tools.

For all we know, capitalizing analysis processes has not been worked out in TEL community. Despite the fact that some works aim to share results of analyses [8] or customised operators [6], they are mainly tool specific and there is no federation between these tools. *Ipsa facto*, TEL community is confronted to the difficulty of being aware of what already exists, involving re-implementation of pre-existing analyses. However, non TEL specific works go in this capitalization direction, like Predictive Model Markup Language (PMML)⁶. PMML aims to share predictive and machine learning models, trained or not, between free and non-free analysis tools: this is a clue for us that the need of capitalization is real.

³ <https://rapidminer.com/>

⁴ <http://www.revolutionanalytics.com/>

⁵ <https://www.imglobal.org/activity/caliperram>

⁶ <http://dmg.org/pmml/v4-1/GeneralStructure.html>

3 Preliminary Assumptions

As shown in the previous section, there is, to our knowledge, no effective and easy way to capitalize analysis processes of e-learning traces. Thus we focus on how analysis processes of traces and their inner components can be described in such a way that they are not related to a specific analysis tool. To do so, we based our work over three assumptions. Firstly, (A1) we assume that designing analysis processes is a cognitive task and is realized by manipulating the meaning of data instead of specific values of them. Indeed, Rosch expresses the fact that the cognition is made *via* categories playing the role of cognitive reference models instead of elementary instances [7]. Secondly, (A2) we assume that since this design is a cognition process, specificities of analysis tools are not taken into consideration. Thus, an analysis process can be regarded as a set of elementary operations. Finally, as our state of the art suggests it, (A3) an analysis process can be seen as a non linear ordered succession of operations, taking inputs and producing outputs: this brings up an important sequential property for ordering.

4 The capitalization of Analysis Processes

4.1 Where to Capitalize in the Iterative Learning Analysis Procedure?

We notice in the literature three main steps concerned by the capitalization of analysis processes. They are (S1) selection of relevant data and consideration of context constraints, (S2) preparation of the analysis and (S3) implementation of the analysis. From users' point of view, two roles are mainly involved in these steps: the e-learning tool expert and the analyst.

The step about the selection of relevant technical information (S1) is often implicit in the literature but since it requires practical knowledge, we consider it as an independent step [9, 10]. The e-learning tool expert is involved during this step. He/she has expertise about the technical context of needs, like the pedagogical domain, the pedagogical platform, learning traces produced, data subjects. Thus, this expert makes the needs more concrete by communicating these information and can eventually detect some inconsistencies or limits.

The preparation of the analysis step (S2) is realized by the analyst. This role is played for instance by data miners, statisticians or researchers. Thanks to his/her expertise in the analysis field and information obtained from S1, the analyst designs the analysis in order to address needs. This implies setting up its limits as well as its strategies, defining which data is pertinent or even how pedagogical domain specificities should be used. Hence, this is a complex step which requires a strong interactivity with experts in order to correctly understand and exploit the context of the analysis.

The outcome of this two previous steps is the analysis step (S3), nearly always realized by the analyst. Accordingly to our state of the art, many papers are concerned about analysis methods in several domains (e.g. EDM, LAK). In any case, the objective is to produce knowledge addressing needs (e.g. dropout rates

in a MOOC). S4 is strongly bound with S2 and S3, providing the possibilities to refine the supplied information, the analysis and even the needs.

Finally, we suggest that capitalizing analysis processes, through a capitalization step, can occur at two moments: S2 or S3, when they are designed (e.g. drew on paper), and S4 once they are implemented inside analysis tools. In both cases, this capitalization should be made by describing analysis processes with the formalism presented in the section 4.3, which is not constrained by technical tools specificities. The description can be realized by the analyst due to his/her analysis expertise, and also by the e-learning tool expert since designed analysis processes are pertinent to be capitalized. As we can note here, one of our proposition's strengths is its integration to the analysis procedure without modifying it: we enrich it and provide potential supports to actors involved.

4.2 Independence Using Operators

According to the non-linear assumption (A3), in order to describe a designed or implemented analysis process independently from analysis tools, all its inner operators must also be described independently. Thus, we represent an independent operator as the concept conveyed by semantic equivalent operators, implemented in different analysis tools. For example, let us consider a temporal filter. The way it is implemented, as well as the way of using it, differ between analysis tools. However, the underlying concept is to apply a filtering over a time: this is what is represented by such an independent operator *Temporal Filter*.

The cognition assumption (A1) implies that independent operators, and then *independent analysis process* (IAP), do not process data directly: data are designed with concepts instead of specific instances of concepts. Hence they only manage data concepts, for keeping track of what is available at each step of the analysis (see Fig.1). Data are conceptualized under a notion of *type of traced elements* (TTE) representing the concept they convey. For instance, if a student made an event at 11:02 am, then 11:02 am is a datum's value and the TTE is *time*. Besides, a *temporal filter operator* will use as input *time*, not 11:02 am.

Hence, IAP are adaptive concerning initial data requirements because they are not based upon values themselves but upon TTEs, offering capitalization abilities. Also, since the produced knowledge is also expressed with TTEs, the description of a process through independent operators potentially grants a greater semantic and insights of it. Accordingly, this ensures that, for any IAP, if given data match the prerequisites TTEs, then expected knowledge can be obtained.

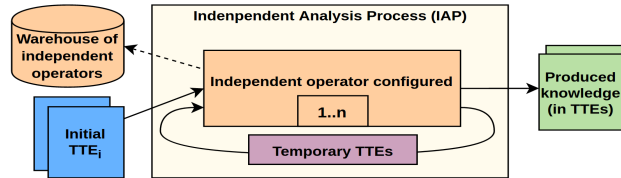


Fig. 1. Representation of the description of an independent analysis process (IAP).

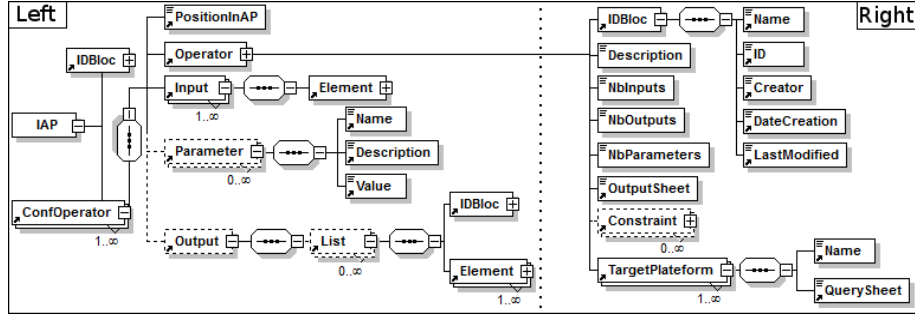


Fig. 2. Meta-model of IAP at the left and independent operator at the right.

4.3 Meta-models of our Approach

We present in this section how both IAP and independent operators have to be represented in order to allow the capitalization of analysis processes.

Firstly, the independent operator meta-model (see Fig.2, right part) has been obtained by an iterative process of identification and comparison between operators in different analysis tools such as UnderTracks, RapidMiner or Weka (<http://www.cs.waikato.ac.nz/ml/weka/>). It describes how an operator has to be constructed in order to be independent of technical specificities. Moreover, it describes how input TTEs will evolve when applying operator on them, according to processing behavior rules *OutputSheet*. For instance, the rule for a clustering operator can be to create a new TTE, representing new groups.

Furthermore, independent operators require few properties in order to exist per se, such as the number of input and output TTEs (*NbInputs* and *NbOutputs*) and the number of parameters (*NbParameters*). They are not directly specified because otherwise, these operators would be constrained before use. Independent operators also require information on which analysis tools are able to implement them, given by *TargetPlatform*. Consequently, it is possible to produce indications for the implementation of an IAP in a specific tool.

Secondly, the IAP meta-model shown in Fig.2, left part, describes an analysis process. It respects the A3 assumption stating that an analysis is a non linear combination of operators with *ConfiguredOperator* which is an ordered step in the IAP: a triplet (*Inputs*, *Operation*, *Outputs*). The inputs are TTEs that will be processed by an independent operator, producing eventually some output TTEs according to the rules of its *OutputSheet*. Then outputs of such operators can be used as the inputs of other ones. This chaining is guaranteed by the partial order property *PositionAP*. Consequently, a *ConfiguredOperator* is reflexive, antisymmetric and transitive, enabling to reliably capitalize an analysis. Hence, a relationship between the expected knowledge and initial TTEs is set up. Moreover, an IAP can be entirely used in another one if the knowledge produced by the first one fits initial TTEs requirements of the second one. This combination offers great perspectives about conception of new capitalized analysis processes.

5 Discussion & Future Works

We have implemented our approach in a web-based prototype⁷ to test its viability through experimentations with 6 subjects used to work in the TEL domain. Results strongly encourage that analysis processes can be described as suggested in this paper. However, experimentation also shows that there is a lack of semantic power concerning TTEs and a lack of feedback available during the description.

Our future works will focus about how supporting actors of the analysis procedure, using capitalized analysis processes. Firstly our efforts will be focused on driving a meaningful description of analysis processes, from a TEL point of view, such as which elements are able to discriminate analysis processes and enrich them. This will help to establish an effective and informative warehouse of IAPs. Secondly, we aim to enable the reuse of these independent analysis processes according to analysis tools and traced data. We assume that analyses will then be more accessible with more supports and insights (e.g. producing relevant instructions). Furthermore this reuse of IAPs can lead to interesting interoperability perspectives between the analysis tools available in the community.

Acknowledgement. This work has been supported by the HUBBLE project (ANR-14-CE24-0015).

References

1. Baker, B.M.: A conceptual framework for making knowledge actionable through capital formation. doctoral dissertation, University of Maryland (2007)
2. Baker, R.S., Yacef, K.: The state of educational data mining in 2009: A review and future visions. *JEDM* 1(1), 3–17 (2009)
3. Choquet, C., Iksal, S.: Usage tracking language: a meta language for modelling tracks in tel systems. In: *Proceedings of ICSOFT’06*. pp. 133–138. INSTICC (2006)
4. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. *AI magazine* 17(3), 37 (1996)
5. Jeong, H., Biswas, G.: Mining student behavior models in learning-by-teaching environments. In: *Proceedings of the 1st International Conference on Educational Data Mining*. pp. 127–136 (2008)
6. Mandran, N., Ortega, M., Luengo, V., Bouhineau, D.: Dop8: merging both data and analysis operators life cycles for technology enhanced learning. In: *Proceedings of the LAK 2015*. pp. 213–217. ACM (2015)
7. Rosch, E.H.: Natural categories. *Cognitive psychology* 4(3), 328–350 (1973)
8. Stamper, J.C., al.: Managing the educational dataset lifecycle with datashop. In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) *Proceedings of the AIED 2011*. pp. 557–559. Springer Berlin Heidelberg, Berlin, Heidelberg (2011)
9. Tsantis, L., Castellani, J.: Enhancing learning environments through solution-based knowledge discovery tools: Forecasting for self-perpetuating systemic reform. *Journal of Special Education Technology* 16(4), 39 (2001)
10. Volle, M., Malinvaud, E.: *Le métier de statisticien*. Economica (1984)

⁷ Experimentation materials at : <http://liris.cnrs.fr/~alebis/iogap.html>