

Approche narrative des processus d'analyse de traces d'apprentissage : un framework ontologique pour la capitalisation

Alexis Lebis^{1,2}, Marie Lefevre², Vanda Luengo¹ et 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

Résumé. Les processus d'analyse de traces d'apprentissage permettent de construire des connaissances issues des traces. La capitalisation de ces processus d'analyse est restreinte par deux facteurs importants : les contraintes techniques auxquelles ils sont soumis et leur dépendance au contexte. Cela empêche notamment de pouvoir les réutiliser aisément en dehors de leur cadre applicatif initial. Dans cet article, nous proposons une approche narrative pour émanciper les processus d'analyse de ces contraintes. Il s'agit de les représenter *via* un framework ontologique pour renforcer la compréhension, humaine et machine, des concepts impliqués, et permettre leur capitalisation au sein de la communauté.

Mots-Clefs : Learning Analytics, processus d'analyse de traces d'apprentissage, capitalisation, ontologie, contexte.

Abstract. Capitalization of analyses extracting knowledge from educational data is hindered by two important factors : technical constraints implied by runnable necessities and context dependencies. These factors prevent to easily reuse analysis processes outside of their original scopes. In this paper we propose a narrative approach, structured by an ontological framework, in order to improve human and machine understanding of such processes. Our aim is to enable the capitalization of analysis processes inside the TEL community.

Keywords : Learning Analytics, analysis process of learning traces, capitalization, ontology, context.

1 Introduction

L'analyse des données issues d'Environnements Informatiques pour l'Apprentissage Humain (EIAH) permet d'en extraire des connaissances pour identifier et étudier les méthodes et les phénomènes d'apprentissage, ainsi que pour assister les différents utilisateurs dans leurs tâches. Dès lors, ces connaissances sont importantes, et, dans le domaine des Learning Analytics par exemple, une attention particulière existe sur comment les expliquer et les partager [11]. En

revanche, les processus d'analyse, qui sont une succession d'opérateurs identifiables et réutilisables que l'on applique sur les éléments contenus dans les traces d'apprentissage [13], ne peuvent pas être traités de la même manière. En effet, ces processus d'analyse sont fortement dépendants de leur contexte d'implémentation, des spécificités des données utilisées (e.g. formalisme) et aussi des spécificités techniques liées aux outils utilisés : leur compréhension peut devenir complexe et leur partage peu pertinent [8]. L'objectif est alors de parvenir à une représentation unifiée des processus d'analyse. Cela permettrait d'impliquer la communauté, de la conception des processus à leur consultation, favorisant ainsi l'interopérabilité, la réutilisation et l'adaptation de ces processus à d'autres situations et besoins pédagogiques [9]. Nous définissons la capitalisation comme une action qui englobe à la fois l'intention de partager et de réutiliser l'existant, dans des contextes proches de ceux initiaux, et qui se destine à toute une communauté et non à un seul individu.

Dans cet article, nous expliquons l'intérêt de se focaliser sur les aspects descriptifs, conceptuels et relationnels des analyses et de leurs éléments associés, plutôt que sur leur aspect exécutable. Nous proposons un framework ontologique qui a vocation à être utilisé par les acteurs de l'analyse (i.e. statisticiens, analystes, chercheurs) pour créer, partager et permettre la capitalisation des processus d'analyse de traces d'apprentissage, à travers leur narration. Les différentes propriétés constituant cette capitalisation, définies par ce framework et décrites en section 4, permettent aussi de rendre ces processus narrés accessibles à la communauté (e.g. enseignants, analystes) et d'impliquer ces derniers *via* la consultation et l'annotation des processus rendus disponibles.

2 État de l'art et positionnement théorique

Processus d'analyse. De nombreux travaux de recherche ont été proposés et développés sur les analyses de traces d'apprentissage pour améliorer le cadre conceptuel et théorique de ces dernières. Il est ainsi possible de considérer les processus d'analyse comme systémiques. Ils prennent en entrée des données, sont décrits comme une succession cohérente d'opérations modifiant l'état des données manipulées pour produire en sortie des informations pertinentes [13]. Il est possible d'étudier les sous-parties de ces processus et de les réutiliser [2].

Ces travaux ont permis de faire émerger le cycle de vie d'un processus d'analyse, avec notamment les phases de pré-traitement, d'analyse et de post-traitement [2]. De plus, certains de ces travaux essaient de matérialiser les liens qui peuvent exister entre les analyses, les données utilisées, les objectifs visés et les acteurs des analyses, pour améliorer la pertinence des processus d'analyse.

Capitalisation, partage, réutilisation. Beaucoup de travaux se sont concentrés sur la question des données utilisées dans les processus d'analyse. Ils proposent divers formalismes pour les rendre partageables et exploitables par la machine [15]. Grâce à ces formalismes, il devient possible de faire correspondre lesdites données à des informations sémantiques prédéfinies, pouvant prendre en

considération des spécificités pédagogiques, comme xAPI¹, ce qui permet d'envisager la création de processus d'analyse reproductibles et réutilisables [7].

Des efforts sur le partage, le plus souvent en ligne, des processus d'analyse, et de leurs opérateurs, sont observables malgré les difficultés techniques [16]. En résulte alors un inventaire des processus d'analyse disponibles pour un outil particulier, ce qui est utile lors d'une démarche exploratoire [13]. Mais cela ne résout pas les limites de réutilisation et d'adaptabilité des processus. En effet, les contextes techniques ne sont pas les seuls à contraindre la capitalisation des processus d'analyse [6] : il y a aussi le contexte de l'analyse (e.g. le dispositif d'apprentissage). Pourtant, d'une manière générale, cette prise en considération des contextes des processus ne semble pas encore réellement établie dans les travaux de la communauté EIAH.

Cependant, la recherche autour des Workflows - représentant des processus de toutes sortes (e.g. biologie) - dévoile des pistes intéressantes pour capitaliser les processus. Ces travaux s'intéressent à la science reproductible, ce qui nécessite de fournir les informations contextuelles nécessaires pour prouver la validité des processus d'analyse. Cela est réalisé par exemple en utilisant des protocoles expérimentaux ou des ressources annexes permettant d'enrichir le contexte de l'analyse [10]. Mais, comme le font remarquer Belhajjame et al. [3], les Workflows présentent des problèmes intrinsèques de flexibilité et sont très sensibles aux contraintes techniques, limitant réutilisation et adaptabilité. Cela est imputable au prérequis d'être exécutable, ainsi qu'aux manques de description des différents éléments impliqués (e.g. opérateurs). Il devient alors intéressant de considérer comme pistes pour la capitalisation des travaux de description sémantique, comme wf4ever [14] ou [5], qui permettent de renforcer les possibilités de réutilisation et d'échange. Ainsi, étudier ces approches dans le cadre des EIAH se révèle intéressant pour exploiter les spécificités du domaine et permettre la capitalisation et le partage des processus d'analyses de traces d'apprentissage.

3 Motivation

Dans cette section, nous présentons un cas concret de processus d'analyse pour illustrer les différentes problématiques liées à la capitalisation et sa nécessité. Il s'agit des deux premières parties d'un tutoriel présenté lors d'un Workshop de la conférence EC-TEL 2016 [1]. Ce tutoriel est accessible en ligne, avec les données nécessaires pour le reproduire et la méthodologie pour chacune des étapes². Bien que ce processus d'analyse soit relativement simple, de nombreuses contraintes techniques et contextuelles existent.

3.1 Exemple d'un cas d'usage

L'analyse présentée dans ce tutoriel nous servira de cas d'usage. Elle a pour objectif la création de modèles permettant de prédire si un apprenant de MOOC

1. <https://experienceapi.com/>

2. https://github.com/Lewkow/EC-TEL_2016_Workshop

sera certifié et les différents facteurs déterminants pour sa réussite. L'analyse est effectuée avec le langage de programmation Python. Le processus présenté est décomposable en trois parties.

La première partie est un prétraitement des données. Elle permet de vérifier la qualité des données utilisées pour obtenir un ensemble exploitable et de les mettre en forme pour favoriser leur utilisation. Dans le tutoriel, des variables additionnelles sont créées pour faciliter la partie suivante, comme la transformation de la variable *Date of birth* vers *Age*.

La deuxième partie représente une analyse exploratoire qui permet d'obtenir des informations supplémentaires sur les données traitées. Ici, notamment, la corrélation - linéaire - pouvant exister entre les propriétés du cours et le succès des apprenants est étudiée avec le coefficient de Pearson.

Enfin, la troisième étape est la création de modèles prédictifs. Pour ce faire, trois ensembles de variables sont candidats à la prédiction de la certification et utilisés pour entraîner les modèles. Les modèles sont ensuite validés en évaluant la valeur de "l'AUC" (Area Under Curve).

3.2 Identification des limites et problématiques de la capitalisation

Ce cas d'usage décrit précisément son analyse et ses opérations - ce qui n'est pas toujours le cas [4], et permet de s'interroger sur comment capitaliser : *"Est-ce que fournir une documentation et présenter comment a été réalisé le processus d'analyse dans un outil est suffisant pour le capitaliser et le partager?"*. Pour répondre à cette question, différents aspects de l'analyse doivent être pris en compte.

Technique & Contextes. Tout d'abord, le formalisme des données exploitées dépend de facteurs contextuels (e.g. dispositifs d'apprentissage). Cela crée une contrainte technique et génère des limites computationnelles, solvables avec des étapes de prétraitement. Par conséquent, partager et réutiliser ces étapes, qui sont extrêmement dépendantes des données sources et du contexte, devient non pertinent si elles ne peuvent pas être adaptables par manque de compréhension. Par exemple, si une étape de prétraitement est réalisée pour normaliser les notes d'étudiants, initialement comprises dans un intervalle $[0; 100]$, dans un intervalle $[0; 20]$, alors elle ne pourra être correctement réutilisée avec un système de notation Nord Américain de type $[A+; F]$ et devra donc être adaptée.

De même, la dépendance des opérations entre elles constitue à la fois une limitation technique et contextuelle. Pour la reproductibilité de l'analyse, identifier leur ordre d'exécution est important. Cela peut être fait par une numérotation simple des étapes par exemple, comme c'est le cas dans le tutoriel présenté. Cependant, pour envisager la réutilisation, il est important de comprendre cet ordre et d'expliquer les relations qui lient ces opérations, afin de mieux les adapter aux spécificités d'un nouveau contexte.

Par conséquent, les processus d'analyses sont sujets à ces contraintes techniques et contextuelles, puisqu'étant eux-mêmes constitués d'opérations, qui varient selon les outils d'analyses. Les réutiliser - ainsi que les connaissances qu'ils

produisent - sans prendre en considération ces contraintes peut entraîner leur mauvaise utilisation et aboutir à de la désinformation. Par exemple, il peut être compliqué d'identifier si les modèles prédictifs du cas d'usage dépendent de spécificités liées aux données des MOOCs analysées ; et donc de savoir si ces modèles peuvent être réutilisés dans le cadre d'autres MOOCs.

Sémantique. Une autre limite à la réutilisation est liée aux données utilisées, qui représentent le contexte de l'analyse. Elles doivent être non ambiguës sur le plan sémantique pour permettre d'aligner les différents concepts nécessaires à l'analyse sur d'autres jeux de données. Par exemple, dans le cas d'usage, la variable "Number of chapters" n'est pas clairement définie, ni ses relations : une personne ayant un jeu de données quelconque possédant lui aussi une variable du même nom n'a pas l'assurance de la validité du processus dans son contexte.

De plus, puisque les opérateurs ne sont pas représentés ni implémentés de la même manière dans les outils d'analyse, leur réutilisation et leur partage sont restreints. Ainsi, l'opérateur de corrélation du tutoriel, et ses configurations, sont implémentés en Python et ne peuvent être directement réutilisés dans d'autres outils d'analyses, comme Weka ³. Dès lors, partager ces opérateurs - et donc les processus - en dehors des outils, sans informations descriptives, requiert un effort de mise en correspondance vers les concepts qu'ils représentent.

En conclusion, nous pensons que ces approches démonstratives (i.e. tutoriels, workflows) tirent leur force dans leur caractère reproductible et leur faculté à donner des solutions "clef-en-main" à des problèmes bien précis. Mais, au vu des limitations, elles ne permettent pas une capitalisation des analyses dans l'optique d'être réutilisées sur d'autres jeux de données et dans d'autres contextes.

4 Framework ontologique

Dans cette section, nous présentons un framework ontologique conçu pour permettre aux acteurs de l'analyse de capitaliser les processus d'analyse de traces d'apprentissage. Nous définissons la capitalisation comme un ensemble hiérarchique en six niveaux, où chaque niveau repose sur ceux antérieurs, comme le montre la Figure 1. Nous considérons l'adaptabilité comme une propriété complexe à obtenir. En effet, pour adapter les processus à différents contextes, il faut être en mesure de les partager entre les outils d'analyse. Cependant, leur partage n'a du sens que si leur réutilisation est réelle et qu'ils sont suffisamment explicites et compréhensibles (que ce soit par l'humain ou la machine). Cependant, pour permettre la réutilisation des processus, il faut préalablement qu'ils soient reproductibles en l'état, ce qui passe par une représentation correcte. Ces propriétés fondent notre framework et sont détaillées en section 4.2.

La particularité de notre approche est qu'au lieu de proposer un formalisme permettant l'exécution des processus d'analyse, nous avons choisi de nous intéresser à leur narration et à comment leurs concepts sont agencés. En effet,

3. <http://www.cs.waikato.ac.nz/ml/weka/>

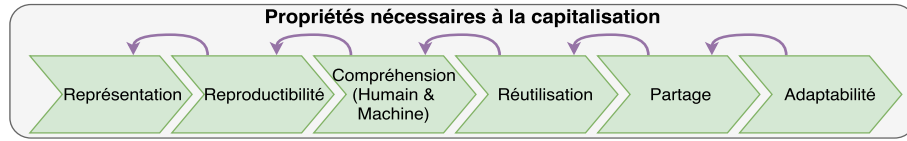


Figure 1. Illustration des propriétés identifiées comme nécessaires pour capitaliser, hiérarchiquement organisées en fonction de leurs dépendances (flèches violettes).

à notre connaissance, il n'existe pas de travaux s'émancipant réellement de cet aspect exécutable, qui pourtant génère d'importantes contraintes techniques, au détriment des concepts existants dans les processus [4].

4.1 Fondement de l'approche

Avant de présenter notre framework, il est important de présenter les fondements sur lesquels il repose. Comme vu en section 2, les processus d'analyse sont composés d'opérateurs utilisant des données en entrée pour calculer des données en sortie : ils sont tous soumis à des contraintes techniques. Dans nos précédents travaux, nous avons montré que ces contraintes techniques pouvaient être surmontées en utilisant un formalisme de plus haut niveau pour décrire les processus d'analyse et les opérateurs [12].

Premièrement, ce formalisme repose sur la notion d'opérateurs indépendants. Un opérateur indépendant est le concept associé à une opération au sein d'un processus d'analyse. Il est privé de toute contrainte technique et faculté d'exécution. Il est donc possible de mettre en correspondance des opérateurs implémentés dans différents outils d'analyse avec un même opérateur indépendant.

Deuxièmement, les données manipulées par les opérateurs indépendants sont aussi de plus haut niveau. Pour ce faire, les valeurs des différentes variables ne sont pas prises en considération, pour éviter les contraintes techniques qu'elles apportent (i.e. format de la donnée). À la place, seuls les concepts associés à ces données sont utilisés et fournis en entrée aux opérateurs indépendants pour représenter et expliquer l'évolution des données au cours de l'analyse.

Enfin, par transitivité, émerge la notion de processus d'analyse indépendants, qui ne sont donc pas destinés à calculer directement des résultats. En revanche, par leur nature plus abstraite, ils représentent une méthodologie de construction de connaissances, à partir de traces d'apprentissage, non altérée par des contraintes techniques.

4.2 Présentation du framework

Afin de rendre accessibles aux humains les processus d'analyse indépendants, et ainsi faciliter leur partage et leur réutilisation, mais également de les rendre intelligibles par la machine pour que celle-ci puisse assister les humains lors de la réutilisation d'un processus d'analyse, nous avons complété notre approche pour permettre la narration des processus d'analyse.

Cette narration porte sur le processus d'analyse, mais aussi sur chacun de ses éléments constitutifs, sur les relations qui existent entre eux, ainsi que sur des informations annexes décrivant le contexte de l'analyse. Pour ce faire, nous définissons un framework ontologique formalisant la représentation des processus d'analyse par la narration.

Notre framework ontologique a pour objectif de répondre aux 6 propriétés que nous avons identifiées comme nécessaires à la capitalisation des processus (cf. Figure 1). Il permet de contrôler la narration aux travers d'éléments sémantiques pertinents. L'ontologie est accessible en ligne⁴. Elle réutilise des terminologies définies dans des travaux existants, comme xAPI ou wf4ever, et qui ont apporté des éléments de réponses à la réutilisation et au partage. L'intérêt est de favoriser l'interopérabilité.

Pour illustrer les propriétés de notre framework, nous reprenons le tutoriel présenté en section 3, et plus particulièrement l'étape de corrélation entre les propriétés du cours et le succès des apprenants qui servira de cas d'usage. Ainsi, la Figure 3 représente cette opération de manière narrée et la Figure 2 montre comment notre ontologie est utilisée pour la décrire puis la partager.

Représentation. Représenter un processus d'analyse signifie être capable d'identifier les opérations qui doivent être appliquées, et dans quel ordre. Dans ce but, nous définissons les notions d'opérateur narré et de processus d'analyse narré.

Un opérateur narré représente un concept d'opération à appliquer. En exemple, considérons l'opération de corrélation réalisée dans le cas d'usage en section 3. Son équivalent est un opérateur narré ayant pour objectif de trouver une corrélation entre différentes variables (trapèze bleu (b) dans la Figure 3). Un processus d'analyse narré décrit une analyse réalisée de manière conceptuelle : il est constitué d'opérations narrées (opérateurs et processus) auxquelles sont rajoutés les éléments narratifs introduits dans les sections suivantes.

Ces opérations narrées sont formalisées dans notre framework, illustré en Figure 2 dans la partie bleue (α), de la manière suivante : chaque opération (trapèze bleu) possède une entrée (*Inputs*), une sortie (*Outputs*) et éventuellement des configurations (*Settings*). Les *Inputs* décrivent les données à traiter et leur contexte. Dans l'exemple de l'opération de corrélation (trapèze vert), les *Inputs* sont, entre autres, l'*Age de l'apprenant* et l'*Événement dans le cours*. La sortie représente l'effet qu'aura l'opération sur les données fournies en entrée. Dans l'exemple, l'opération va produire un nouveau concept en sortie (*Coefficient de Corrélation*). Cela signifie que les données fournies en entrée vont être enrichies avec la corrélation qui existe entre elles.

Notre framework permet d'associer à cet opérateur la manière de le configurer et les données qui ont été utilisées (respectivement *via* les relations *isConfiguredBy* et *areUsedWith*).

Reproductibilité. Cependant, la représentation ne permet d'obtenir que le squelette de l'analyse, *via* la succession des opérations narrées. Dans l'objectif

4. <http://liris.cnrs.fr/~alebis/CAPTEN/ontology.html>

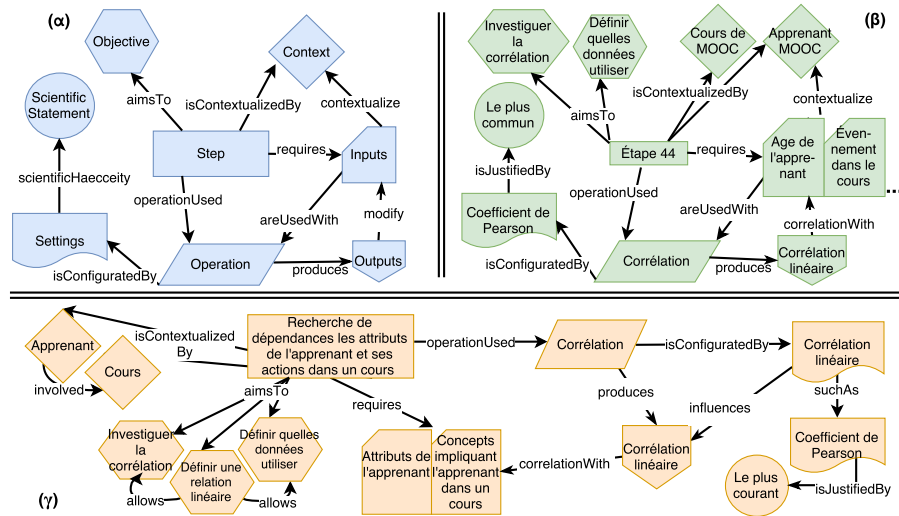


Figure 2. Illustration de notre framework en utilisant notre ontologie (α) pour réifier l'étape d'une analyse (β) et la narrer afin de permettre son adaptabilité (γ).

de rendre les processus d'analyse narrés reproductibles - c'est à dire pouvoir les réutiliser dans leur contexte initial, sur les mêmes données - les configurations, les données et les connaissances produites doivent pouvoir être représentées. Cela revient à réifier les concepts *Inputs* et *Settings* de la Figure 2.

En s'appuyant sur les résultats des expérimentations faites sur notre premier formalisme [12], nous proposons de représenter ces données et ces connaissances comme un graphe de concepts relationnel. En exemple, considérons le graphe (a) de la Figure 3 qui représente les concepts de données qui ont été utilisés lors de l'opération de corrélation. Nous nous intéressons au concept d'Age et à l'information qu'il est possible de déduire à partir de ce graphe. En effet, cela fait ressortir une information contextuelle intrinsèque à l'analyse réalisée, à savoir qu'il s'agit de l'âge d'un apprenant de MOOC : le contexte de l'étape s'en retrouve enrichi.

Ces graphes de concepts représentant les données utiles contenues dans les traces sont utilisés comme données d'entrée des opérations narrées. Les patrons d'entrée des opérateurs - qui sont eux aussi exprimés sous forme de graphes de concept - sont mis en relation avec les concepts contenus dans les graphes des données d'entrée (nœuds verts dans (a)). Les patrons de sortie - eux aussi sous forme de graphes - servent à représenter l'évolution de ces concepts d'entrée après les opérations. Le graphe (c) dans l'exemple montre comment les concepts verts dans (a) évoluent une fois que l'opérateur de corrélation est appliqué.

Compréhension. Comprendre les processus d'analyse permet de prévenir leurs mauvaises réutilisations ainsi que les problèmes liés à la qualité scientifique, comme une interprétation erronée des connaissances attendues. Cependant, les

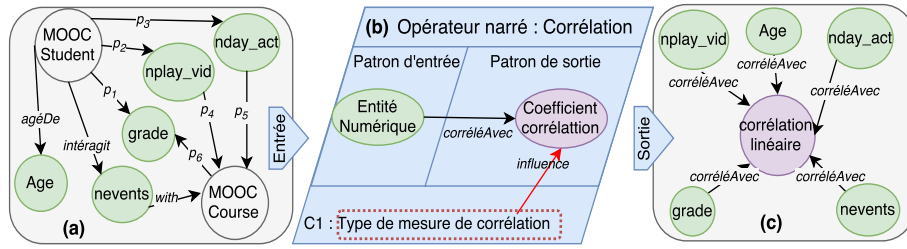


Figure 3. Vue détaillée d'une étape d'un processus d'analyse narré.

limitations descriptives induites par les prérequis d'exécutions des outils d'analyses ne permettent pas de formaliser ce besoin de compréhension. Dans notre framework, nous proposons de le formaliser en définissant des éléments narratifs dédiés, supports à cette compréhension. L'intérêt de cette formalisation est qu'elle est à la fois exploitable par l'humain et par la machine. Par exemple, visible sur la partie bleue de la Figure 2, notre ontologie définit des concepts comme l'objectif (*Objective*) ou bien encore des affirmations scientifiques (*Scientific Statement*) qui, une fois réifiés (cf. (β), partie verte), apportent des éléments de compréhension supplémentaires formalisés.

De plus, nous définissons un ensemble de relations qui sont sémantisées dans la propriété de partage du framework. Ces relations permettent de représenter les rapports que peuvent entretenir plusieurs éléments narratifs entre eux ou avec le processus narré. Pour illustrer ces relations, prenons l'exemple des auteurs du cas d'usage qui justifie l'utilisation du coefficient de Pearson par le fait que ce soit le plus commun. Le fait qu'ils justifient ce choix est une information importante, qu'il convient d'indiquer explicitement pour renforcer la description de cette étape - et qui serait perdu ou peu visible dans des outils d'analyse plus traditionnels. La partie verte de la Figure 2 montre comment réifier cette information dans notre framework : utiliser une relation de justification (*isJustifiedBy*) partant du *coefficient de Pearson* vers *Le plus commun*, qui est la réification du concept de *Scientific Statement*.

Réutilisation. Réutiliser consiste à utiliser des processus d'analyses déjà existants dans des situations similaires (i.e. même contexte d'analyse mais jeux de données différents). Il en émane un besoin computationnel évident ; or nous rappelons que nous avons délibérément choisi de ne pas doter les opérations narrées de cet aspect computationnel pour pouvoir changer le paradigme de représentation des analyses. Pour répondre ce besoin, notre framework adopte une approche systémique du processus d'analyse lors de la narration. Il définit le concept d'étape d'un processus d'analyse comme la transition d'un état de l'analyse vers un autre. Concrètement, une étape est constituée d'une opération narrée - pouvant être configurée, du graphe de concepts sur lequel cette opération est appliquée et du graphe résultant. Ainsi, l'ensemble de la Figure 3 est une étape représentant l'étape de corrélation effectuée dans le cas d'usage.

Par définition, l'étape est un élément structurant important : elle représente une intention lors de l'analyse. Notre framework permet une formalisation de cette intention en utilisant les éléments narratifs et les relations précédemment introduits. Chaque étape peut ainsi être décrite de manière structurée, et mise en relations avec les éléments de l'analyse - y compris avec d'autres étapes. La Figure 2 offre un aperçu de cette formalisation. L'ontologie définit une étape (*Step*, en bleu dans la partie (α)) comme possédant, en plus des concepts d'entrée et de l'opération, un objectif ou encore un contexte. En réifiant le cas d'usage, on peut ainsi décrire les motivations de cette étape (e.g. investiguer la corrélation) ou encore indiquer dans quel cadre elle a été initialement réalisée (cf. propriété *Adaptabilité*).

Enfin, notre framework se base sur l'encapsulation qu'offre ce concept d'étape pour permettre de décrire des pointeurs d'instanciation vers des outils d'analyses spécifiques. Il s'agit de décrire comment réaliser une étape d'un processus d'analyse narré dans un outil d'analyse particulier. L'objectif est de consolider la réponse au besoin calculatoire par l'interopérabilité des différents outils.

Partage. Partager et réutiliser les processus d'analyse dans d'autres outils d'analyse est non trivial, comme peuvent le montrer des travaux comme PMML⁵, qui se concentrent depuis presque vingt ans sur l'échange de modèles prédictifs entre différents outils. Cette complexité réside dans les possibilités computationnelles disparates des outils et de la divergence sémantique de leurs différents éléments [5].

Pour traiter ce problème de la divergence sémantique, notre framework utilise un vocabulaire contrôlé. L'objectif est de rendre intelligibles (1) les termes qui constituent l'ontologie et (2) les termes qui sont utilisés pour décrire les différents éléments du processus d'analyse narré. La finalité est de pouvoir représenter les mêmes choses, de la même manière, afin de favoriser l'échange et l'interopérabilité. Ainsi, dans la Figure 2, certaines terminologies de notre ontologie, comme *Input* et *Operation*, proviennent de travaux comme wf4ever. Il en va de même pour les concepts utilisés dans les graphes, comme ceux de la Figure 3, qui seront majoritairement issus de xAPI. Ces vocabulaires peuvent d'ailleurs être utilisés dans les éléments narratifs pour identifier de manière intelligible certains concepts, comme on peut le voir avec le losange vert *Apprenant MOOC* sur la Figure 2.

Adaptabilité. Utiliser des processus d'analyses déjà existant dans des situations différentes de leur réalisation revient à les adapter : cela peut entraîner une modification desdits processus. Nous avons constaté que les contextes de l'analyse ne sont que peu représentable dans les outils d'analyse traditionnels, alors qu'ils jouent un rôle majeur dans l'expression des spécificités de l'analyse.

Notre framework propose de formaliser les contextes de l'analyse par les quatre éléments narratifs suivants : (1) le contexte d'analyse (i.e. les contraintes impliquées par les éléments initiaux utilisés lors de l'analyse), (2) le contexte

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

d'utilisation de l'analyse (i.e. les situations pédagogiques dans lesquels un processus est utilisable), (3) le contexte d'une étape (i.e. les contraintes impliquées par les éléments utilisés lors d'une étape) et enfin (4) le contexte de viabilité des connaissances produites (i.e. le cadre d'usage prévu dans lequel les connaissances sont correctement exploitables). Le cadre narratif permet d'obtenir directement certaines informations pour enrichir ces contextes. La Figure 2 illustre comment, dans notre ontologie, le contexte est mis en relation avec une étape et comment les entrées influencent le contexte. En réifiant l'étape de corrélation dans notre framework, il est alors possible d'exprimer que la corrélation effectuée n'est valable que dans le contexte d'un *Cours de MOOC*, avec des *Apprenants de MOOC* : ces informations sont déduites des concepts d'entrée utilisés.

Enfin, le cadre narratif apporté par notre framework rend possible l'adaptation des processus d'analyse narrés par l'ajout, la suppression et la modification des différents éléments narratifs. La Figure 2 illustre cela. La zone verte est la réification de l'étape de corrélation du cas d'usage, telle qu'elle a été réalisée. Or, il est intéressant d'adapter cette étape pour permettre de la réutiliser dans des situations semblables : c'est ce que représente (γ), la partie orange de cette figure. Dans cette partie, les concepts d'entrée sont généralisés à des *Apprenants* - cela est rendu possible par le vocabulaire contrôlé, puisqu'un apprenant de MOOC est avant tout un apprenant. De ce fait, le contexte de l'étape peut être modifié et même précisé *via* la relation *involves*, précisant que cela concerne des *Apprenant* impliqués dans des *Cours*. Cette partie montre également comment les éléments narratifs peuvent être utilisés pour rajouter de l'information, comme c'est le cas avec la configuration de l'opération : le coefficient de Pearson est décrit comme étant une corrélation linéaire.

5 Conclusion et perspectives

Cet article présente une solution pour capitaliser les processus d'analyse de traces d'apprentissage par l'utilisation d'un framework ontologique conçu pour les narrer. Notre ontologie rend ces processus auto-descriptifs, puisque les concepts utilisés au sein de ces processus sont explicables grâce à du vocabulaire contrôlé et des relations et évite les altérations techniques. Grâce à l'abstraction générée par l'ontologie, les différents contextes de l'analyse peuvent aussi être décrits.

Nous envisageons d'évaluer et de renforcer notre framework par l'instanciation d'analyses existantes, notamment dans la littérature des *learning analytics*. Nous utiliserons ensuite ce framework pour découvrir et élaborer des règles d'inférence afin d'assister les différentes personnes impliquées dans les étapes complexes de l'analyse. De plus, cela pourra favoriser l'émergence et la compréhension d'habitudes d'analyse liées aux EIAH, tout en nous inscrivant dans une démarche de science reproductible.

Remerciement

Ce travail a été financé par le projet ANR HUBBLE (ANR-14-CE24-0015).

Références

1. Aghababayan, A., Agnihotric, L., Essa, A., Lewkow, N., Mojarad, S. : Educational data mining with python and apache spark : A hands-on tutorial. In : Workshop WAPLA at EC-TEL'16. Lyon, France (2016)
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. Belhajjame, K., et al. : Why workflows break — understanding and combating decay in taverna workflows. In : Proceedings of the 8th International Conference on E-Science. pp. 1–9. IEEE Computer Society, Washington, DC, USA (2012)
4. Belhajjame, K., et al. : Using a suite of ontologies for preserving workflow-centric research objects. *Web Semantics : Science, Services and Agents on the World Wide Web* 32, 16–42 (2015)
5. Bowers, S., Ludäscher, B. : An ontology-driven framework for data transformation in scientific workflows. In : International Workshop on Data Integration in the Life Sciences. pp. 1–16. Springer (2004)
6. Chatti, M.A., Dyckhoff, A.L., Schroeder, U., Thüs, H. : A reference model for learning analytics. *International Journal of Technology Enhanced Learning* 4(5-6), 318–331 (2012)
7. 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)
8. Clow, D. : An overview of learning analytics. *Teaching in Higher Education* 18(6), 683–695 (2013)
9. Cooper, A. : Learning analytics interoperability-the big picture in brief. *Learning Analytics Community Exchange* (2014)
10. De Roure, D., et al. : Towards open science : the myexperiment approach. *Concurrency and Computation : Practice and Experience* 22(17), 2335–2353 (2010)
11. Elias, T. : Learning Analytics : Definitions , Processes and Potential (2016), <http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf>
12. Lebis, A., Lefevre, M., Luengo, V., Guin, N. : Towards a capitalization of processes analyzing learning interaction traces. In : Proceedings of the EC-TEL'16. pp. 397–403. Springer (2016)
13. Mandran, N., Ortega, M., Luengo, V., Bouhineau, D. : Dop8 : merging both data and analysis operators life cycles for technology enhanced learning. In : Proceedings of LAK'15. pp. 213–217. ACM (2015)
14. Page, K., et al. : From workflows to research objects : an architecture for preserving the semantics of science. In : Proceedings of the 2nd International Workshop on Linked Science. Citeseer (2012)
15. Reffay, C., Betbeder, M.L., Chanier, T. : Multimodal learning and teaching corpora exchange : lessons learned in five years by the mulce project. *International Journal of Technology Enhanced Learning* 4(1-2), 11–30 (2012)
16. Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S.B., Ferguson, R., Duval, E., Verbert, K., Baker, R. : Open learning analytics : an integrated & modularized platform. Tech. rep., Society for Learning Analytics Research (2011)