Discovering learning antecedents in learning analytics literature

Kobayashi, V.; Mol, S.; Kismihók, G.

Publication date
2015

Document Version
Final published version

Published in
Proceedings of the First International Workshop on Visual Aspects of Learning Analytics

Discovering Learning Antecedents in Learning Analytics Literature

Vladimer Kobayashi
CJKR, HRM-OB, ABS
University of Amsterdam
Netherlands
V.Kobayashi@uva.nl

Stefan Mol
CJKR, HRM-OB, ABS
University of Amsterdam
Netherlands
S.T.Mol@uva.nl

Gábor Kismihók
CJKR, HRM-OB, ABS
University of Amsterdam
Netherlands
G.Kismihok@uva.nl

ABSTRACT

We investigated various learning antecedents that have been the research subjects of Learning Analytics (LA) studies and explored the content and quantity of the LA literature with respect to each antecedent through text mining the LAK dataset. Our goal was to simultaneously reveal to what extent do LA researchers address learning antecedents and how they incorporated these in the implementation of LA solutions (e.g. models and software technologies) to facilitate and augment student learning. Instead of taking a pure text mining approach, we undertook a slightly different strategy by (i) identifying antecedents of student learning by examining extant literature on learning and educational theories and (ii) identifying which among the theoretically relevant antecedents are currently reported in LA studies. The analytical techniques we employed were a mix of domain-based analysis and corpus analytics which included association analysis and key-phrase extraction. The results showed that most LA studies are geared toward capturing and measuring student awareness and promoting social learning and less on goal-setting and self-efficacy. Through this work we hope to encourage the LA community to dedicate research efforts to also investigate other relatively neglected yet promising learning antecedents.

Keywords

student learning, corpus analytics, learning analytics

1. MOTIVATION AND OBJECTIVE

The Learning Analytics (LA) field uses analytics to understand and facilitate student learning. Since learning is influenced by various antecedents and circumstances, some LA researchers focus on capturing, measuring, and enhancing these antecedents in an effort to impact student learning. This is especially relevant nowadays with the proliferation of nontraditional venues for learning such as in online learning. Examples of these antecedents include awareness, social learning, and self-regulated learning to name but a few.

As LA studies flourish a need arises to address the question of how LA as a field has contributed so far to our understanding and to the enhancement of student learning. This can be answered in part by characterizing LA studies according to which learning antecedents they tackle. This could help researchers from various education-related disciplines to keep track, compare, and share knowledge and to identify opportunities for further research. It could also provide a basis for the adaption of LA projects and explicating how LA models and software technologies influence learning. How each element of an LA project imparts information or generates and uses data that valuate the determinants for student learning success is a major concern.

Our primary objective was to explore the content and quantity of LA literature that report each learning antecedent. In a parallel manner, we shifted the focus towards the antecedents by finding which antecedents are often addressed and which not. This approach would facilitate a more objective assessment and comparison of whether LA studies have achieved their intended outcomes.

For this study we used the dataset provided by the LAK dataset challenge [9] and other literature on student learning theories to accomplish our objective.

2. METHODOLOGY

As an overview, we used a text mining approach to discover learning antecedents. Although text mining is naturally an inductive approach we supplemented our investigation with domain information. The diagrammatic description of the steps we undertook is illustrated in Figure 1.

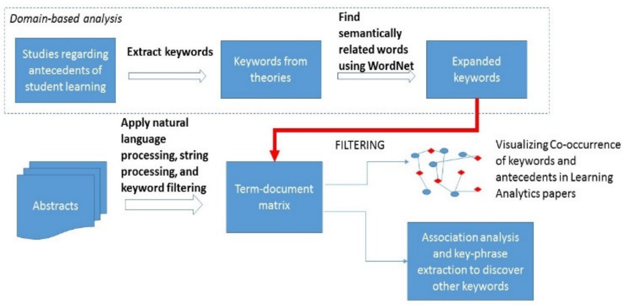


Figure 1: Diagrammatic view of the methodological steps followed in this study.

2.1 Domain-based Analysis

We first performed an inquiry regarding the antecedents that influence student learning and learning outcomes. From this inquiry we identified keywords that are usually strongly associated to each antecedent. The keywords represent the vocabulary used to refer to the antecedents that were extracted from existing literature on education and student learning theories. The antecedents are discussed in Section 3.

The list of keywords were further expanded by using a lexical database called WordNet[1] to find semantically similar words. This is a vital step because authors use varying terms to convey the same

[1] http://wordnet.princeton.edu/

concept. An example would be to use “participate” rather than “engage”. The expanded keyword list was used in the succeeding steps.

2.2 Corpus Analytics on LAK dataset

Corpus analytics was performed in the following manner.

First, we initially kept matters simple yet meaningful by choosing to perform corpus analytics only on the abstracts of each publication. There might be a downside to this such as missing otherwise important information but in exchange this has kept the analysis manageable. Moreover, this decision is sufficient for our purpose since the abstract contains the gist of the whole article and provides a summary about the paper’s objectives, methodology, and conclusion.

Second, we created a corpus containing abstracts of all papers in the LAK dataset. Each document was pre-processed by removing punctuation, removing numbers, transforming upper case letters to lower case, removing stopwords, and selectively stemming specific words. An example of the selective stemming was to treat the words “engaging” and “engagement” as just derivatives of the word “engage”. The method of stemming that we applied here is the look-up table method where the look-up table is the expanded keyword list from domain analysis.

Third, a further filtering was implemented to reduce the number of terms. The filtering process was done using the expanded keyword list in conjunction with association analysis so that potentially important words not present in the list could be identified and added.

Fourth and finally, the pre-processing stage culminated in the creation of the document-by-term matrix weighted by raw term frequencies. We were interested in determining which among the theory inspired antecedents (see Section 3) are discussed in each LA study. The document-by-term matrix acted as a springboard from which we explored the construction of other matrices (e.g. co-occurrence matrices) and application of other analytical techniques such as key-phrase extraction.

All analyses were done using the **R** software[2] and the packages **tm**[3], **wordnet**[4], and **igraph**[5].

2.3 Two assumptions

We assumed that the mention of keywords associated to a learning antecedent in the abstract of a paper would indicate that the paper is dealing with that learning antecedent. We anticipate a number of caveats with this assumption. One possible scenario is that the keyword is used in a different sense. An example is the keyword “goal”, in some papers the presence of this word does not mean that they are automatically dealing with Goal-setting but it could be the case that the word “goal” here refers to the goal of the study. Thus it is also important to consider the context in which the word is being used. We addressed this by examining other words in the abstract. Using association analysis we noticed that when the word “goal” is used in the sense of Goal-setting words such as performance, achievement, or learning are also encountered.

Another assumption is that the mention of keywords belonging to different learning antecedent in one abstract means that these two learning antecedents are simultaneously addressed and with the same emphasis in that paper. We can see a problem with this since some papers just use the concept but do not develop that concept further. This problem can be addressed by using the information on the raw frequencies of the term. The higher the raw frequency the more importance we can attach to it with respect to a particular paper.

3. NINE ANTECEDENTS OF STUDENT LEARNING

The keywords represent 9 common antecedents that have been reported by educational experts as antecedents for success in learning. The antecedents are: (1) Engagement, (2) Motivation, (3) Self-reflection (including self-assessment and self-regulation), (4) Social Learning (among students and between students and teachers), (5) Assessment (e.g. formatting testing and evaluation), (6) Recommendation (and feedback), (7) Goal-setting, (8) Awareness (social awareness, context awareness), and (9) Self-confidence. These were selected based on our previous content analysis of publications in the area of education and student learning.

Student engagement refers to the quality of effort and level of involvement that students invest in their learning. It has been shown to be positively linked to gains in general abilities, critical thinking, and grades [1]. Therefore it has worthwhile effects on student learning and success in education.

Motivation is a drive, a stimuli, an incentive or desire that causes someone to act or to expend effort to accomplish something [8]. Often, it is manifested when students are attentive, participative and active in class.

Self-reflection occurs when learners evaluate the breadth and scope of their knowledge. It is important in learning because it helps students to identify what they need to learn leading to effective self-regulation [5].

Some researchers view learning as a collaborative process where learners interact and share knowledge. The roles, activities, and behavior that students assume in a social learning context ultimately impact their learning [2].

Testing and assessment in general has long been used to assess whether students have achieved specific learning outcomes. Furthermore, during testing information is stored in the brain for long term retrieval, which in turn is essential for learning transfer (i.e. using information in different contexts) and meaning generation.

Recommendation is seen as a potential antecedent of learning since it helps students track their learning achievement and improve their learning at the same time [3].

Goals direct attention, energize effort and promote persistence. Studies have shown the valuable effect of goal-setting to academic achievement, self-regulation, and deep learning strategies [6].

Awareness provides context for learning since it discloses information about other person’s activities and the environment where learning takes place. It has been shown to be crucial to learning and contributes to the quality of active participation [7].

Last is self-efficacy (colloquially termed as self-confidence) which is usually defined as belief in one’s own capability to accomplish tasks and achieve goals [4]. It is important in learning since students

[2] http://www.r-project.org/

[3] http://cran.r-project.org/web/packages/tm/index.html

[4] http://cran.r-project.org/web/packages/wordnet/index.html

[5] http://cran.r-project.org/web/packages/igraph/index.html

must believe in their own capacity to learn even if the material is difficult.

We added the Analytics to see which LA projects have incorporated advanced analytical tools on top of the basic summarization and visualization features.

4. MAIN FINDINGs AND DISCUSSION

Combining the keywords obtained from the domain analysis, association analysis, and corpus analytics we obtained the keyword list in Table 1 that are grouped according to the antecedents that are most likely associated to them.

Table 1: Keywords associated to each learning antecedent.

Learning Antecedents	**Keywords**
Engagement	engage, participate, active, access, resource
Motivation	motivate, encourage
Self-reflection	negotiate, self-regulate, self-reflect, self-aware, self-discipline, self-test, reflect, self-report, self-knowledfe
Social Learning	collaborate, network, interact, social, community, graph, connect
Assessment	test, assess
Recommendation	recommend, feedback, intervene
Goal-setting	goal, sub-goal
Awareness	aware, content-aware, track, monitor, compare
Self-Confidence	confidence, self-efficacy
Analytics	model, student model, user model, analytics, analytic, predict, valid, visual, classify

From the document-by-term matrix we identified which among the documents have used analytics and which learning antecedents are addressed in each document. We also constructed 4 co-occurrence matrices (see Figure 2) that reveal which learning antecedents are often treated simultaneously, and which keywords are often mentioned together. A sampling of output is presented in Figure 3.

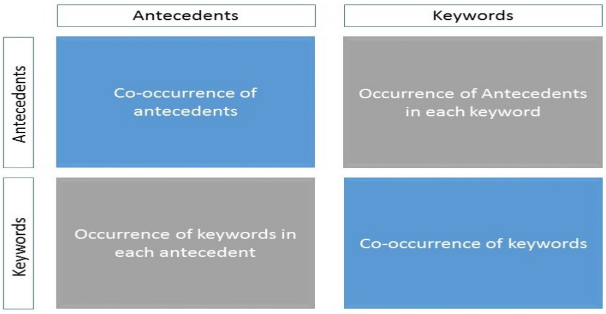


Figure 2: Four co-occurrence matrices constructed from the term-by-document matrix.

The first subfigure (Figure 3a) shows a bar plot that depicts the number of papers in the LAK dataset that have dealt with each learning antecedent. It can be vividly seen that the focus of many studies are the learning antecedents awareness, social learning, engagement, and assessment. This can be explained by the considerable interest of LA researchers in online learning settings where the capture, measurement, and monitoring of these antecedents are both challenging and crucial. On the other the less often discussed antecedents are goal-setting, motivation, and self-discipline. Although, goal-setting has a slightly higher bar than self-reflection this is because some studies that mention the word "goal" actually referred to the aim or objective of the studies.

Figure 3b depicts both the magnitude of studies that deal with each antecedent and the relationship (in the sense of co-occurrence) among the antecedents. The red circles are the antecedents and the green ones are the keywords. An edge connects a keyword to its associated antecedent and edges between antecedents represent relationship. We include "Analytics" to see which among the antecedents make heavy use of analytics and what type of analytics is commonly employed. It is not difficult to observe that social learning and awareness are the most related in terms of the number of publications that tackled them. It is followed by awareness and assessment, although there is a strong indication that assessment here may imply the students' assessment of their knowledge, context, peers, and environment and not about test or evaluation.

The last subgraph (Figure 3c) visually represent the relationship among words as well as the quantity of studies that mention each word (as expressed by the size of the circle). It is not surprising to observe that the word "model" is the leading keyword this is because most LA researchers are concerned with creating models to describe some learning-related phenomena, as to be expected from an LA research. Another observation that is worth mentioning is the conspicuousness of the three vertices that represent visual, network, and interact and the interconnections between them. These three are indicative of the social learning antecedent since interactions among students are usually visualized by means of a network structure.

In Table 2, we see the list of words that are highly associated to the keywords of each antecedent. We discovered these with the use of association analysis and key-phrase extraction. The list is incomplete since we just present the ones that were interesting in our opinion. These words could be used to further enrich our original keyword list. Moreover, we unearthed interesting relationships such as the association between "affect" and "engagement", "assessment" and "scores", "recommendation" and "similarity". Some of these associations reveal the kind of techniques used to analyze particular antecedents (e.g. the use of the idea of similarity in recommendation) and the underlying concepts that might govern an antecedent (e.g. the affective state of a student might indicate or influence engagement).

Table 2: Other terms associated to each antecedent.

Engagement	affect, peripheral, discussion, home
motivation	learnograms
self-reflection	cope, personal, health, feelings
Social learning	blackboard, intergroup, intranetwork, cyberlearner
Assessment	scores
Recommendation	similarity
Goal setting	orientation, temporal
Awareness	clues, cope
Self-confidence	Egocentric, high achieving

5. CONCLUSION AND FUTURE WORK

In this study we show how an analysis that combines domain-based information and corpus analytics could be used to uncover and analyse interesting concepts in LA literature. These concepts directly deal with the question of how LA has been used to improve our understanding and control of a number of learning antecedents. We believe that to fully answer that question a more detailed analysis should be undertaken such as investigating the measures and validity of the constructed models as described in the publications. Nevertheless, our approach clears the cloud to expedite such detailed analysis. Our study also highlights the need to study other antecedents that might be critical to student learning but do not yet receive due research attention. From an educator's perspective it is now becoming clearer how LA solutions impact learning and to which aspect the contribution is focused. It is now time that we move LA from a technique-laden endeavor to a more theory driven approach.

If ever, this work will be selected we also show our effort on the temporal analysis of these antecedents such as visualizing the evolution of focus of LA studies on each concept. Moreover, we aim to analyze how publications in educational data mining, learning analytics and technology-enhanced learning differ in this aspect.

6. ACKNOWLEDGEMENT

We gratefully acknowledge the publishers who have contributed to the LAK Dataset: ACM, International Educational Data Mining Society and Journal on Education Technology &Society. We are grateful for the financial support of the Eduworks Marie Curie Initial Training Network Project (PITN-GA-2013-608311) of the European Commissions's 7th Framework Programme.

7. REFERENCES

[1] Aguiar, E. et al. 2014. Engagement vs performance: using electronic portfolios to predict first semester engineering student retention. (2014), 103–112.

[2] Barr, J. and Gunawardena, A. 2012. Classroom Salon: A Tool for Social Collaboration. *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education* (New York, NY, USA, 2012), 197–202.

[3] Bramucci, R. and Gaston, J. 2012. Sherpa: Increasing Student Success with a Recommendation Engine. *Proceedings of the 2Nd International Conference on Learning Analytics and Knowledge* (New York, NY, USA, 2012), 82–83.

[4] Diseth, Å. 2011. Self-efficacy, goal orientations and learning strategies as mediators between preceding and subsequent academic achievement. *Learning and Individual Differences*. 21, 2 (Apr. 2011), 191–195.

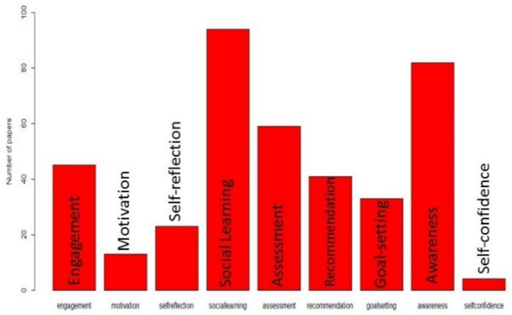
[5] Govaerts, S. et al. 2012. The Student Activity Meter for Awareness and Self-reflection. *CHI '12 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2012), 869–884.

[6] Latham, G.P. and Locke, E.A. 2007. New Developments in and Directions for Goal-Setting Research. *European Psychologist*. 12, 4 (Jan. 2007), 290–300.

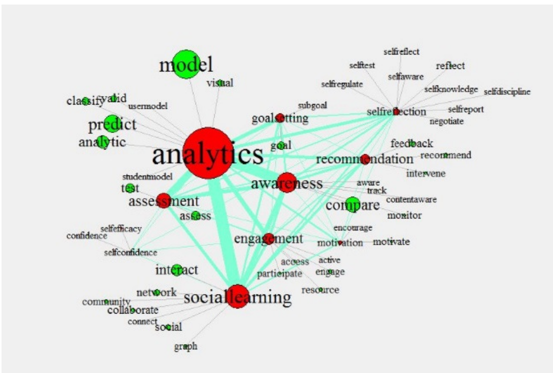
[7] Pohl, A. et al. 2012. Sensing the classroom: Improving awareness and self-awareness of students in Backstage. *2012 15th International Conference on Interactive Collaborative Learning (ICL)* (Sep. 2012), 1–8.

[8] Schiefele, U. Interest, Learning, and Motivation.

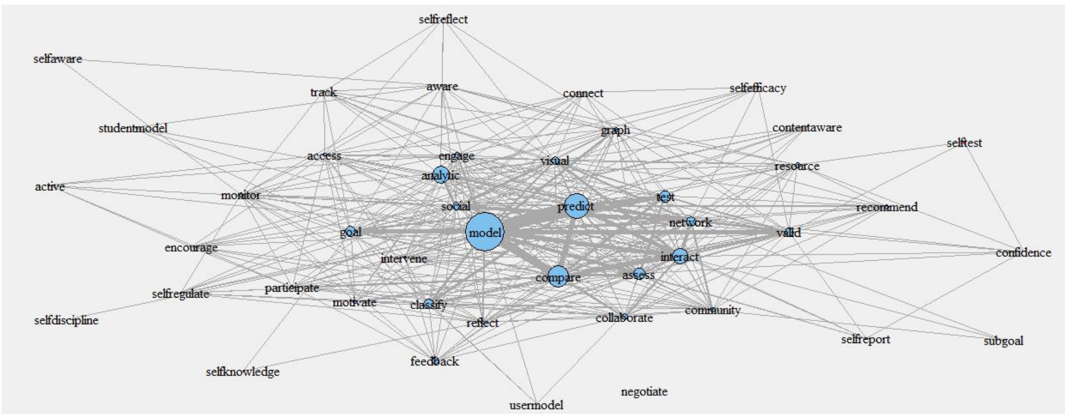
[9] Taibi, D. and Dietze, S. 2013. Fostering analytics on learning analytics research: the LAK dataset. In: *CEUR WS Proceedings Vol 974, Proceedings of the LAK Data Challenge*, held at LAK2013 - 3rd International Conference on Learning Analytics and Knowledge(Leuven, BE, Apr. 2013).



(a)



(b)



(c)

Figure 3: Sampling of the output from the analysis