Theory-driven Learning Analytics and Open Learner Modelling: The Teacher's Toolbox of Tomorrow?

Michael D. Kickmeier-Rust
Graz University of Technology
Knowledge Technologies Institute
8010 Graz, Austria
+43 316 873 30636
michael.kickmeier-rust@tugraz.at

Dietrich Albert
Graz University of Technology
Knowledge Technologies Institute
8010 Graz, Austria
+43 316 873 30640
dietrich.albert@tugraz.at

ABSTRACT

Big Data and data technologies increasingly find their way into school education. Learning Analytics and Educational Data Mining are focal research areas. However, technical solutions often fail to meet the practical requirements of teachers or to really mirror human learning processes. The LEA's BOX project aims at developing a practical web platform that hosts tools for a theory-based approach to Learning Analytics and that offers tools to open and negotiate learner models.

Keywords

Learning analytics, data visualization, open learner models.

1. INTRODUCTION

Using Learning analytics and educational data mining are more than recent buzz words in educational research: they signify one of the most promising developments in improving teaching and learning. While many attempts to enhance learning with mere technology failed in the past, making sense of a large amount of data collected over a long period of time and conveying it to teachers in a suitable form is indeed the area where computers and technology can add value for future classrooms. However, reasoning about data, and in particular learning-related data, is not trivial and requires a robust foundation of well-elaborated psycho-pedagogical theories.

The fundamental idea of learning analytics is not new. In essence, the aim is using as much information about learners as possible to understand the meaning of the data in terms of the learners' strengths, abilities, knowledge, weakness, learning progress, attitudes, and social networks with the final goal of providing the best and most appropriate personalized support. Thus, the concept of learning analytics is quite similar to the idea of formative assessment. "Good" teachers of all time have strived to achieve exactly this goal. However, collecting, storing, interpreting, and aggregating information about learners that originates from a school year, or even in a lifelong learning sense) requires smart technology. To analyse this vast amount of data, give it educational meaning, visualize the results, represent the learner in a holistic and fair manner, and provide appropriate feedback, teachers need to be equipped with the appropriate technology. With that regard, a substantial body of research work and tools already exist. This project aims to continue and enrich on-going developments and facilitate the broad use of learning analytics in the "real educational world.

2. LEA's BOX

LEA's BOX (www.leas-box.eu) is a project, funded under the EU's Seventh Framework Programme and stands for a practical LEarning Analytics tool Box, that provides

- a competence-centred, multi-source formative assessment methodology,
- based on sound psycho-pedagogical models (i.e., Competence-based Knowledge Space Theory and Formal Concept Analysis),
- intelligent model-based reasoning services,
- innovative visualization techniques,
- and features to open and negotiate learner models;

LEA's BOX is dedicated to develop a learning analytics toolbox that is intended to enable educators to perform competence-centered, multi-source learning analytics, considering their real practical needs. Thus, the project spends significant efforts on a close and intensive interaction with educators in form of design focus groups and piloting studies.

The tangible result of LEA's BOX manifest in form of a Web platform (Figure 1) for teachers and learners provide links to the existing components and interfaces to a broad range of educational data sources. Teachers will be able to link the various tools and methods that they are already using in their daily practice and that provide software APIs (e.g., Moodle courses, electronic tests, Google Docs, etc.) in one central location. More importantly, the platform hosts the newly developed LA/EDM services, empowering educators to conduct competence-based analysis of rich data sets. A key focus of the platform will enable teachers not only to combine existing bits of data but to allow

Figure 1. Central web platform.

them to "generate" and collect data in very simple forms, not requiring sophisticated hard- or software solutions. Finally, we want to open new ways to display the results of learning analytics - leaving the rather statistical dashboard approach, moving towards structural visualizations and towards opening the internal learner models.

3. Open Learner Models

Learner models contain and dynamically update information regarding users' learning: current knowledge, competencies, misconceptions, goals, affective states, etc. There is an increasing trend towards opening the learner model to the user (learner, teacher or other stakeholders) to support reflection, encourage greater learner responsibility for their learning, and help teachers to better understand their students [2]. The core requirement is that such visualizations must be understandable to the user. Although this may appear to be similar to the more recent work on LA, open learner models (OLM) concentrate more on the current state of learners, with less references to activities undertaken, scores obtained, materials used, contributions made, etc. OLMs typically focus on concepts, competencies, and guiding learners with regard to conceptual issues rather than specific activities and performance. Various OLM visualization examples have been described in the literature for university students (see [2] for a more detailed overview). The most common visualizations used in courses include skill meters, concept maps and hierarchical tree structures.

In addition to visualizing the learner model, various methods of interacting with the learner model exist, ranging from simple inspectable models, which allow some kind of additional evidence to be input directly by users, to negotiated learner models, in which the content of the learner model is discussed and potentially updated. We focus on the latter. Key features of negotiated learner models are not only that the presentation of the learner model must be understandable by the user, but also that the aim of the interactive learner modelling should be an agreed model. Most negotiated learner models are negotiated between the student and the teaching system. However, other stakeholders can also be involved, and the notion of "the system" can be broadened to include a range of technologies, such as the ones used in technology-enhanced learning. Here we consider (i) fully-negotiated learner models; (ii) partially-negotiated learner models; and (iii) other types of learner model discussion. They are all relevant to our notion of negotiating the learner model or its content, and they are adapted for LEA's BOX (Figure 2).

Figure 2. OLM in LEA's BOX

4. COMPETENCE SPACES

In the context of formative LA, a competence-oriented approach is necessary. Thus, a Hasse diagram can be used to identify and display the latent competencies of a learner in the form of so-called competence states. An elaborated theoretical approach to do so is Competence-based Knowledge Space Theory (CbKST). The approach originates from Jean-Paul Doignon and Jean-Claude Falmagne [6, 7] and is a mathematical psychological, set-theoretic framework for addressing the relations among problems (e.g., test items). It provides a basis for structuring a domain of knowledge and for representing the knowledge based on prerequisite relations. While the original Knowledge Space Theory focuses only on performance (the behavior; for example, solving a test item), its extension CbKST [1] introduces a separation of observable performance and latent, unobservable competencies, which determine the performance [1]. This is a psychological learning-theoretical approach, which highlights that competencies (e.g., the ability to add two integers) are unobservable latent constructs and which can only be observed or assessed indirectly.

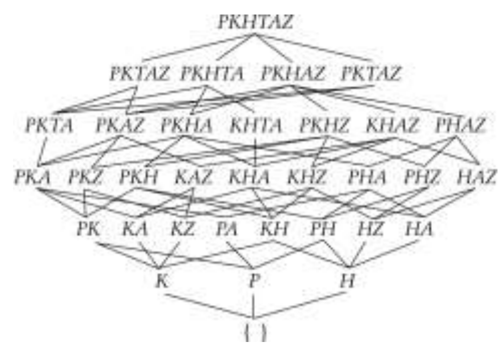

Figure 3. A simple Hasse diagram.

We interpret the performance of a learner (e.g., mastering an addition task) in terms of holding or not holding the respective competency. In addition, recent developments of the approach are based on a probabilistic view of having or lacking certain competencies. In our example, mastering one specific addition task allows the conclusion that the person is able to add two numbers (to hold this competency) only to a certain degree or probability. When thinking of a multiple-choice item with two alternatives, as another example, mastering this item allows only to 50 percent that the person has the required competencies/ knowledge.

On the basis of these fundamental views, CbKST is looking for the involved entities of aptitude (the competencies) and a natural structure, a natural course of learning in a given domain. For example, it is reasonable to start with the basics (e.g., the competency to add numbers) and increasingly advance in the learning domain (to subtraction, multiplication, division, etc.). As indicated above, this natural course is not necessary linear, which bears significant advantages over other learning and test theories.

As a result we have a set of competencies in a domain and potential relationships between them. In terms of learning, the relationships define the course of learning and thus which competencies are learned before others. In CbKST such relationships are called prerequisite relations or precedence relations. On the basis of competencies and relationships, in a next step, we can obtain a so-called competence space, the ordered set of all meaningful competence states a learner can be in. As an example, a learner might have none of the competencies, or might be able to add and subtract numbers; other states, in turn, are not included in this space, for example it is not reasonable to assume that a learner holds the competency to multiply numbers but not to add them. By the logic of CbKST, each learner is, with certain likelihood, in one of the competence states.

5. VISUALIZING COMPETENCE SPACES

Hasse diagrams are capable of holding a number of important information for an educator to evaluate the learning progress and also to make recommendations. In this paper we want to highlight such advantages.

5.1 Competence States and Levels

As outlined, a competency space is the collection of meaningful states a learner can be in. Depending on the domain, the amount of possible states might be huge. The big advantage, however, is that depending on the degree of structure in the domain, by far not all possible combinations of competencies are reasonable and thus part of the space. When zooming into the diagram, a teacher can exactly identify the set of competencies that is most likely for the learner, by zooming out color-coding can illustrate the most likely locations of a learner within the space. When looking at the entire space, it is obvious at first site at which completion level a learner is approximately (rather at the beginning or almost finished). These zoom levels are shown in Figure 4. Technically, there is a variety of options to achieve the coding, for example, bolding, greying, or color coding, whereas likely states are displayed more distinctly than such with low probability.

Equal to individual states, Hasse diagrams can represent group distributions. Defined by a certain confidence interval of probabilities those states and areas can be made more salient that hold the highest percentage of learners of a group. By this means, specific areas in the competency space become apparent within which the most learners are and, in contrast also positive or negative outliners pop out the diagram. A different method was suggested by [10], who altered the size of the nodes to represent the groups' sizes; the larger a node the more learners hold a particular state.

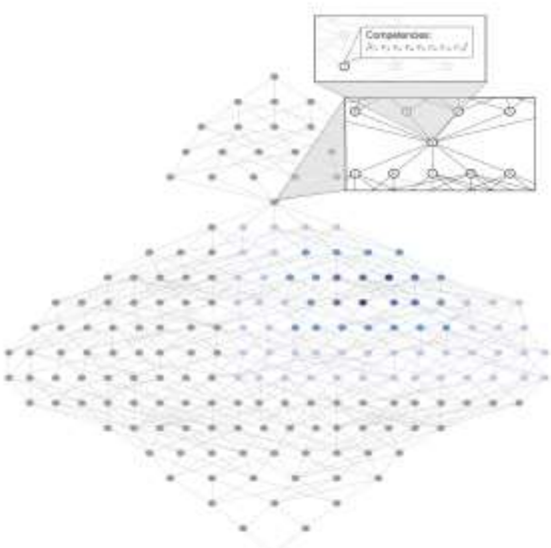

Figure 4. Hasse diagram illustrating the probability distribution over a competence space on three zoom levels.

5.2 Learning Paths

In addition to having insight into groups' and individuals' current states of learning, the learning history, the so-called learning paths, are of interested for educators; on the one hand for planning future activities, on the other hand, for negotiation and documenting the achievements of a learning episode (e.g., a semester). Learning paths can be simply displayed by highlighting the edges between the most likely state(s) over time. As for the states, various probable paths can be realized by making more likely paths more intensive (by color coding or line thickness). Figure 5 shows a simple example. A key strength of presenting learning paths, as indicated, is opening up the learner model to the learners (perhaps parents) themselves [10] – to explain where they started at the beginning of a course and how they proceeded during the course and which competencies they hold today. This perhaps can be complemented with comparisons to others or groups. Not least, learning paths can unveil information about the effectiveness and impact of certain learning activities, materials, or the teacher herself.

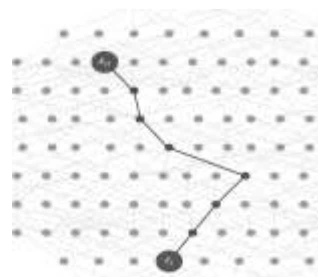

Figure 5. Learning Path. The cutout is part of the structure shown in Figure 4.

6. CLASSROOM DATA COLLECTION

The features of Hasse diagrams and the arising advantages for LA appear all well and good. However, the key question is, where do they data for computing the probabilities of competence states come from. And everything stands or falls with this question. As for all techniques of LA, it depends on a data rich approach to education, the more and the better data exist, the better is the quality of LA conclusions. CbKST and Hasse diagrams are no exception to that. However, the approach of separating latent competencies, which more or less develop and exist in the black box 'human brain', and the performance they determine, bears particular advantages. On the one hand, performance, e.g. test scores, classroom participation, homework, etc., is not only determined by competencies or aptitude; there is a variety of aspects contributing to a certain performance, e.g., motivation, daily constitution, tiredness, external distractors, nutrition, health status, etc. On the other hand, CbKST-ish competence spaces are rather stable, once set up and validated properly. The advantage lays in the fact that performance such as test results, behaviors, achievements, etc. is considered as probability-based indicators for certain competencies. Mathematically this relationship is established in form of interpretation and representation functions [1], which links an arbitrary set of performances/behaviors to one or more competencies, either in an increasing or in a decreasing sense. This, in the end, allows linking all available and perhaps changing data sources to one and the same competence space. It's not about a single test, it's about all available information we can gather, even it is considered being of little importance, all sorts of information may contribute to strengthen the model, the view of the learner. In case the amount or quality of data is weak, CbKST allows conservative interpretations, based on the arising probability distributions, in case there is a richer data basis, the probability distributions are more reliable, valid, and robust. For the educator, and this is important, the uncertainty is mirrored in the degree of likelihood. On a weak data basis, the probabilities of competence states differ substantially less than on the basis of richer data. Such information, however, can change the educator's view and evaluation of a student's achievements. In the end, this approach supports a fairer and more substantiated approach to grading or providing formatively inspired feedback.

7. CONCLUSIONS AND OUTLOOK

There is little doubt that frameworks, techniques, and tools for LA will increasingly be part of a teacher's professional life in the near future. The benefits are convincing – using the (partly massive) amount of available data from the students in a smart, automated, and effective way, supported by intelligent systems in order to have all the relevant information available just in time and at first sight. The ultimate goal is to formatively evaluate individual achievements and competencies and provide the learners with the best possible individual support and teaching. Great. The idea of formative assessment and educational data mining is not new but the hype over recent years resulted in scientific sound and robust approaches becoming available, and usable software products appeared. However, when surveying the educational landscape, at least that of the EU, the educational daily routines are different. We face technology-lean classrooms and schools, we face a lack of proper teacher education in using ICT in schools – not mentioning of using techniques of LA in schools. We face a certain aloofness to use breaking educational technologies and a well-founded pedagogical view that learning ideally is analogous and socially embedded and doesn't occur in front of some kind of electronic device. These are all experiences and results of a large scale European research project named Next-Tell (www.next-tell.eu) that was looking into educationally practices across Europe and that intended to support teachers where exactly they are today with suitable ICT as effective and as appropriately as possible.

The framework of CbKST offers a rigorously competence-based, probabilistic, and multi-source approach that accounts for the latent and holistic abilities of learners and therefore accounts for the recent conceptual change in Europe's educational systems towards a more competence-oriented education including multi-subject competencies and superordinate 21st century (soft) skills. No matter if data are rich or lean; a teacher is supported to the best possible degree and with a variety of important information about individual and group-based learning processes and performances and not least about the performance of learners and about the educator's own performance. The probabilistic dimension allows teachers to have a more cautious view of individual achievements – it might well be that a learner has a competency but fails in a test; vice versa, a student might luckily guess an answer.

From an application perspective, in the context of European projects we developed and evaluated tools that cover the techniques and approaches described in this paper. In the Next-Tell project, for example, we developed a software tool named ProNIFA, which allowed linking multiple sources of evidence of learning and building CbKST-based learner models. We piloted various school studies and gathered feedback from teachers. In the end, and this can be considered an outlook for future developments, we had to find out that the 'massive' Hasse diagrams are overburdening teachers' understanding and mental models about individual and class-based learning. Moreover, in order to understand the classical Hasse diagrams, it required (too) massive efforts in training teachers to fully utilize the potentials of those diagrams. Large scale surveys yielded that most educators still prefer simple but information-wise shallow visualizations such as traffic lights or bar charts significantly over more information-rich approaches such as Hasse diagrams or, just to mention another interesting approach, parallel coordinates. Therefore, recent efforts, e.g., in the LEA's BOX (www.leas-box.eu) project, seek to adjust and advance the classical Hasse diagrams to such visualizations that are intuitively understood by educators and, at the same time, hold the same density of information. In particular, focus of research is on an advancement of Hasse diagrams towards specific mental models teachers may hold, such as a starry night sky or organic, biological structures such as cells of a living being. Also, abstraction and simplification techniques are investigated, e.g., fisheye lenses or streamgraphs. In conclusion, the utility of CbKST-ish approaches to LA, involving a separation of latent competencies and observable behaviors/performance, as well as having a conservative, probabilistic, multi-source approach appears to be a striking classroom-oriented, next-level contribution to LA, learner modelling, and model negotiations.

8. ACKNOWLEDGMENTS

This work is based on the Next-Tell project and the LEA's BOX project, contracted under number 619762, of the 7th Framework Programme. This document does not represent the opinion of the EC and the EC is not responsible for any use that might be made of its content.

9. REFERENCES

[1] Albert, D., & Lukas, J. 1999. Knowledge Spaces: Theories, Empirical Research, and Applications. Mahwah, NJ: Lawrence Erlbaum Associates.

[2] Bull, S. & Kay, J. (2010). Open Learner Models, In R. Nkambou, J. Bordeau and R. Miziguchi (Eds.), *Advances in Intelligent Tutoring Systems* (pp. 318-338), Berlin-Heidelberg: Springer-Verlag.

[3] Doignon, J., & Falmagne, J. 1985. Spaces for the assessment of knowledge. International Journal of Man-Machine Studies, 23, 175–196.

[4] Doignon, J., & Falmagne, J. 1999. Knowledge Spaces. Berlin: Springer.

[5] Nakamura, Y., Tsuji, H., Seta, K., Hashimoto, K., and Albert, D. 2011. Visualization of Learner's State and Learning Paths with Knowledge Structures. In A. König et al. (Eds.), KES 2011, Part IV. Lecture Notes in Artifical Intelligence 6884, pp. 261-270. Berlin: Springer.