Support Teachers' Predictions of Learning Success by Structural Competence Modelling

Michael D. Kickmeier-Rust
Graz University of Technology
Plüddemanngasse 104
8010 Graz, Austria
+43 316-873-30636
michael.kickmeier-rust@tugraz.at

Dietrich Albert
Graz University of Technology
Plüddemanngasse 104
8010 Graz, Austria
+43 316-873-30630
dietrich.albert@tugraz.at

ABSTRACT

Learning Analytics is one of the most promising major trends in educational technology. However, Learning Analytics is very often a rather statistical approach to the understanding of educationally relevant data. Theory-driven approaches are much sparser. In the context of the European Lea's Box project (www.leas-box.eu), we aim at developing methods for analysing data coming from multiple sources on the basis of psychological theories from the area of Intelligent Tutorial Systems, namely Competence-based Knowledge Space Theory (CbKST) and Formal Concept Analysis (FCA). These well-elaborated approaches allow us to identify competencies on an atomic level, to establish structural, multi-dimensional knowledge spaces, and to identify individual learning paths and knowledge gaps. In this paper we introduce an approach to utilize the mentioned theories to predict learning paths, the Learning Performance Vector, and individual limits, the so-called individual learning Horizon.

Keywords

Learning Analytics, Competence-based Knowledge Space Theory, Formal Concept Analysis, Learning Performance Vector, Learning Horizon

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.

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 for teachers and learners that provides 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 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.

2. THE LEARNING HORIZON

In the centre of conceptual research in the field of CbKST and FCA was the so called Learning Performance Vector (LPV) and the Learning Horizon. The principle idea of this constructs is to use CbKST and FCA as means of predictive analytics. The fundamental idea, thereby, is to consider the past learning performance in terms of CbKST-like learning paths , the current progress of an individual learner as well as a summary of peer performance (if available) and to match learning time and remaining time with the learning goals. In such a way we aim at deriving estimations of an individual's learning success and the degree to which a desired learning goal can be achieved. The foundations of this approach are not only competence structures and formal concepts (e.g., competencies over learners) but also temporal information, weighting information of activities and achievements, and difficulty aspects of future learning tasks. In the end, we try to establish an algorithm that is capable of melding those information into robust predictions of learning success – in other terms of the likelihood that a particular student can reach the learning goals in a given amount of time – the Learning Horizon. Of course, the predictions are unstable and blurred in the beginning and certainly the predications are more valid, the more time has passed and the more information the system has. Still, the approach is capable, so we hope, to give early indications of performance problems, so that it is still possible for educators to intervene appropriately. In addition, a particular strength is that the CbKST/FAC approach allows for finding concrete directions where a learner needs support and guidance.

3. ELEMENTS OF THE LEARNING HORIZON AND THE LPV

3.1 Competence Structures and Performance

The first element we consider is clearly a competence structure (Figure 1). Very briefly, we decompose a learning domain (e.g., 2nd grade maths) into atomic chunks of knowledge or aptitude. In a second step we try to find a natural course of learning or, in other terms, we try to find the prerequisite structure: which elements need to be learned before another piece can be acquired. This gives us a combinatorics model of a learning domain and a certain understanding of how learning and development occurs. Now, it must be highlighted that competencies and learning, abilities and aptitudes are latent constructs. One cannot directly observe the real "knowledge" of another person. It takes indicators and evidences, in its simplest form a school test. We know, very well, that tests are not necessarily objective. Students can be inattentive and fail although they have the knowledge or competence, some may guess the right answer incidentally. So in the end, there is a good portion of uncertainty in assessment. When talking about the underlying competencies, we need to account for this fact. And we need to account for that in a careful and conservative way. The CbKST approach does that by establishing stochastic relationships. Each indicator, each piece of evidence, each test result is only one indicator that contributes to the whole picture, but it contributes only with a certain probability. The more evidence we can aggregate, mirroring the same competencies and competence structures, the clearer and more robust our picture (our model of the learner) gets. Of course, we have to consider that different evidences have different weights, a different impact, on the learner model. A simple multiple choice test weighs less than an oral exam within which a teacher can explore the real knowledge of a student, exhibiting abilities in real live weighs more than filling in the right answers.

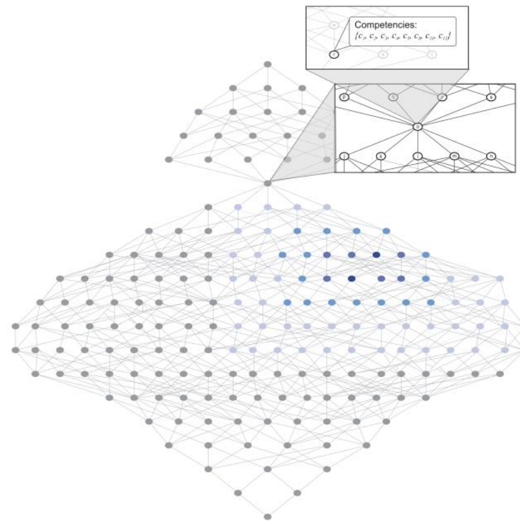

Figure 1. Illustration of a Competence structure; to bottom most node indicate the empty set of competencies, the lines are the possible learning paths, and the top most node indicates the possession of all the competencies of a domain.

3.2 Formal Contexts

FCA, the analysis of formal context, is a related formal psychological approach. The idea is to identify patterns in a universe of two dimensions. Imagine there is a set of competencies and a set of students. There is a multitude of clusters, some students hold the one some the other competencies. FCA allows to quickly analyse the patterns and identify relevant clusters, even more, hierarchies. If FCA is applied on the competency models of CbKST, we have the opportunity to meld pedagogically inspired domain models with pattern identification mechanisms. By this means we can identify clusters of good and not so good learners, we can establish a hierarchy of performance, and, at each step, we can determine which competencies are lacking, and therefore which educational measures would be necessary. In general, there is a broad variety of educationally relevant questions that can be addressed using the paired CbKST / FCA approach (cf. Bedek, Kickmeier-Rust & Albert, 2015).

3.3 Likelihoods, Weights, and their Extensions

In recent works we demonstrated that the traditional approaches of using Hasse diagrams for visualizing competence structures and lattice graphs for displaying formal contexts can be extended in meaningful ways. One idea suggested by (Kickmeier-Rust & Albert, 2015) was to extend Hasse diagram visualizations by adding a difficulty (a weight) dimension to the diagram by illustrating the length of edges in correspondence to their weight (difficulty). There are two important aspects to this idea. On the one hand, it introduces weights as levels of difficulties and the necessary efforts to make the step from one to another competence state, on the other hand, it provides valuable information to inspire the LPV and the estimation of a Learning Horizon. In addition to that, a simple yet important fact is that subject matter is increasing in difficulty over time. This definitely must be another variable in our model of learning.

3.4 What Peers are Doing

Now, when it's about to estimate a student's potential progress and chances to accomplish a course on time, e central element is a comparison to other learners. [It shall be highlighted that this is optional, since the LPV can be computed without peer information!] If a particular student appears being clearly ahead of the majority or, in a worse case, behind the majority, a teacher can receive corresponding and actionable information from analytics.

Here also a meta-perspective comes into play, namely the degree to which a teacher is capable of setting the right learning goals for a particular group of students and the ability to reach the goals. This is a non-trivial aspect to Learning Analytics tools. Oftentimes, a teacher is seen as the ultimate key luminary in a certain domain. This, however, is not necessarily true. Teacher may completely misjudge the abilities and potentials of a group of students (and there is a variety of reasons why this may happen). So, a dimension of a group comparison can add substantial information about individual progress as well as a teacher's plans. In the end, this analysis offers a fountain of deeper insights.

Finally, it's worth mentioning that a theoretically sound peer comparison offers the option for a motivation boost of individual efforts, almost like the principle of badging or gamification. Position and achievements in peer groups have tremendous motivational powers, however, the must be utilized very carefully and thoughtfully!

4. PREDICATION ALGORITHM

So what do we have: A competence structure (or competence space). This structure gives us a model of the learning domain, starting from point 0 (in this particular domain) leading to the complete mastery. In other terms, a competence structures is the manifestation of all possible and reasonable states a person can be in. This allows us to identify the progress of a particular learner given the timeline of a course. Mathematically speaking we have the sum of all possible learning paths. This indicates the average learning efforts, given that transitions have specific difficulties or weights (cf. Figure 2).

We have a set of competencies Q = {a, b, c, …. } with a relationship c □ c' among the competencies, which establishes the competence structure. The sum of the resulting competence states is □ (|Q|r). Given that the transitions from one competence state to another has a difficulty parameter, which in turn is the average of the difficulty parameters of the competencies being a part of the state, we have a set of tuples of the start competence state, the end state, and the difficulty □ = [s1, s2, w]. This results in a set of such tuples for the entire competence structure □ = □(□|Q). Also, we have a set of indicators providing evidences for competencies: I = { ei, {c} * w}, with a given weight w. Based on the evidences we can estimate the likelihood of each competency. The probability of a competence state is the average of its competencies □(s) = □(□)/n.

To identify the learning path of a person, we identify the state with the highest probability in certain time steps. Depending on the nature of the concrete use case this may rely on the events when evidences are put into the system or, alternatively on a timely basis (e.g., weekly or monthly).

Now for each step we compute the difficulty (as a value from 0 to 1). The sum of the values gives us an indicator for how many efforts a student has to spend on her learning history (the individual learning path). In a next step, given the concrete competence state of the learner, we have to identify the possible paths towards to defined learning goal, which is a (rather small) subset of all possible paths. Equally to the computation of the difficulty to reach the current state, we can compute the potential difficulty of all possible paths to the goal, whereas we have to compute the average difficulty of all possible paths. This now is an indicator for the efforts that are necessary for an individual learner to reach the learning goal.

When link the progress of a student within a given span of time, we can make a prediction about how far a student can come within the remaining time (of a course, for example). So, as a final step, we can identify exactly those states (and therefore the competencies) a particular will be able to reach within the time limits. The set of those states is, now finally, the student's Learning Horizon.

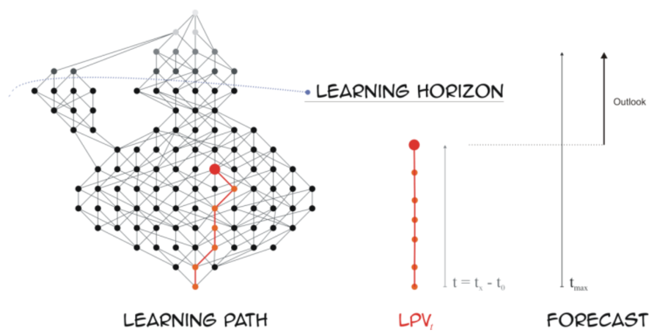

Figure 2. Conceptual illustration of the approach.

5. CURRENT STATUS AND OUTLOOK

The LPV and LH approach appear being an interesting method for educationally relevant predictions that compliments the existing rather statistical methods. While these methods usually make predictions on the basis of a comparison of an individual learner with a possibly large set of other students and their achievements, the introduced approach is primarily based on information about the learning domain, the competencies, their characteristics, and their relationships. The advantages are, on the one hand, that the noise of statistical comparisons is reduced; on the other hand, analyses and predictions can be made without referring to a large basis of existing student data. The latter point is of particular interest when focussing on school education: usually schooling is a diverse analogues setting where not much data is generated where data that is available are not aggregated and where the nature of data is extremely diverse (Kickmeier-Rust, Bull, & Albert, 2016).

The introduced approach is implemented in the Lea's Box Learning Analytics Toolbox (www.leas-box.eu). As emphasized in the introductory section, this online platform is tailored to the concrete demands of teachers and provides a set of internal Learning Analytics tools and, which is a key focus of the project, APIs to link a large number of external tools (such as learning apps, e-learning systems, cloud tools) to the system. The vision is to allow an easy aggregation of all the data that are available, even if there is not much data and not coherent data, and make the most of it in terms of a formative evaluation and feedback and a more evidence based individualisation of teaching.

Presently we are evaluating the validity of the approach on the basis of large data sets from professional learning solutions in Turkey and the US. In Turkey we have access to the Vitamin learning platform (https://www.vitaminegitim.com/vittrin/) which offers a broad offer of courses and tests. In addition we include data from the US product Adaptive Curriculum (https://www.adaptivecurriculum.com/us/), which offers courses for middle and high school levels. The first experiences are quite promising; the predications of the system yield a substantial fit to the patterns we find in the large data sets. We will investigate the predictive power further and will specifically address the question whether the analyses and predictions are also valid for the data lean school scenarios in comparison to the data rich evolution scenarios. The recent developments as well as the continuous study results are frequently posted on the Lea's Box website (www.leas-box.eu) as well on Lea's Facebook account (www.facebook.com/LeasLearning).

6. Acknowledgements

This work is based on the LEA's BOX project, contracted under number 619762, which is supported by the European Commission (EC) under the Information Society Technology priority of the 7th Framework Programme for research and development. 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.

7. References

[1] *Bedek, M. A., Kickmeier-Rust, M. D., & Albert, D. (2015). Formal concept analysis for modelling students in a technology-enhanced learning setting. In M. Kravcik, A. Mikroyannidis, V. Pammer, M. Prilla, T. D. Ullman (Eds.), Proceedings of the 5th Workshop on Awareness and Reflection in Technology Enhanced Learning at EC-TEL 2015, Toledo, Spain, September 15, 2015.*

[2] *Falmagne, J.-C., Albert D., Doble, D., Eppstein D., Hu X. (2013). Knowledge Spaces: Applications in Education. Berlin: Springer.*

[3] *Kickmeier-Rust, M. D., & Albert, D. (Eds.) (2012). An Alien's guide to multi-adaptive educational games. Santa Rosa, CA: Informing Science Press.*

[4] *Kickmeier-Rust, M. D., Bull, S., & Albert, D. (2016). LEA's BOX: Practical Competence-oriented Learning Analytics and Open Learner Modeling. In Proceedings of the workshop Learning Analytics for Learners at LAK'16 conference, 26-29 April, 2016, Edinburgh, UK.*

[5] *Kickmeier-Rust, M. D., Steiner, C. M., & Albert, A. (2015). Uncovering Learning Processes Using Competence-based Knowledge Structuring and Hasse Diagrams. In Proceedings of LAK15, Workshop Visual Approaches to Learning Analytics. March 16-20, 2015, Poughkeepsie, NY.*

[6] *Reimann, P., Bull, S., Kickmeier-Rust, M. D., Vatrapu, R., Wasson, B. (Eds.) (2015). Measuring and Visualizing Learning in the Information-Rich Classroom. New York, NY: Routledge.*