Uncovering Learning Processes Using Competence-based Knowledge Structuring and Hasse Diagrams

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

Christina M. Steiner
Graz University of Technology
Knowledge Technologies Institute
8010 Graz, Austria
+43 316 873 30640
christina.steiner@tugraz.at

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

ABSTRACT

Learning analytics means gathering a broad range of data, bringing the various sources together, and analyzing them. However, to draw educational insights from the results of the analyses, these results must be visualized and presented to the educators and learners. This task is often accomplished by using dashboards equipped with conventional and often simple visualizations such as bar charts or traffic lights. In this paper we want to introduce a method for utilizing the strengths of directed graphs, namely Hasse diagrams, and a competence-oriented approach of structuring knowledge and learning domains. After a brief theoretical introduction, this paper highlights and discusses potential advantages and gives an outlook to recent challenges for research.

Keywords

Learning analytics, data visualization, Hasse diagram, Competence-based Knowledge Space Theory.

1. INTRODUCTION

Using methods and tools from Learning Analytics (LA) can be considered best practice and is a key factor for making education more personalized, adaptive, and effective. Analyzing a variety of available data to uncover learning processes, strengths and weaknesses, competence gaps undoubtedly is a prerequisite for a formatively-inspired guidance, for changing and adjusting educational measures and teaching, and not least for disclosing and negotiating learner models [4]. Usually, the benefits are seen in the potential to reduce attrition through early risk identification, improve learning performance and achievement levels, enable a more effective use of teaching time, and improve learning design and instructional design [10]. On the basis of available data, ideally large scale data sets, smart tools and systems are being developed to provide teachers with effective, intuitive, and easy to understand aggregations of data and the related visualizations. There is a substantial amount of work going on this particular field; visualization techniques and dashboards are broadly available (cf. [2,4,7]), ranging from simple meter/gauge-based techniques (e.g., in form of traffic lights, smiley, or bar charts) to more sophisticated activity and network illustrations (e.g., radar charts or hyperbolic network trees).

However, LA operates in a delicate and complex area. On the one hand, facing today's classroom realities, we often find technology-lean environments, which do not easily allow or support recording the necessary data. Also, from a socio-pedagogical perspective, learning must be seen as a process of social interaction that not always occurs in front of some electronic. Thus, LA must be based on fewer data. On the other hand, it is rather easy to visualize learning on a superficial level using perhaps the aforementioned traffic lights or bar charts. The added value to the teachers is likely of limited utility to them. To provide a deeper and more formative insight into the learning history and the current state of a learner (beyond the degree to which a teacher might know it intuitively) requires finding and presenting complex data aggregations. This, most often, bears the significant downside that it is hard to understand. Challenges for LA and its visualizations, for example, are to illustrate learning progress (including learning paths) and - beyond the retrospective view - to display the next meaningful learning steps/topics.

In this paper we introduce the method of directed graphs, the so-called Hasse diagrams, for structuring learning domains and for visualizing the progress of a learner through this domain.

2. HASSE DIAGRAMS AND COMPETENCE-BASED KNOWLEDGE SPACES

A Hasse diagram is a strict mathematical representation of a so-called semi-order in form of a directed graph that reads from bottom to top. A semi-order is a type of mathematical ordering of a set of items with numerical values by identifying two items as equal or comparable if the values are within a given interval of error or noise. Semi-orders were introduced in mathematical psychology by Duncan Luce in 1956 [8] in human decision research without the assumption that indifference is transitive. This approach is also crucial for handling human learning and the resulting performance that is prone to all sorts of errors and peripheral aspects (perhaps failing in a test although the learner holds the knowledge due to being tired). A Hasse diagram is one way of displaying such ordering – in our case competences or competency states (which is to be explained in the following section). The technique was invented in the 60s of the last century by Helmut Hasse. The diagram exists of entities (the nodes), which are connected by relationships (indicated by edges).

The mathematical properties of a semi-order and the Hasse diagrams are (i) reflexivity, (ii) anti-symmetry, and (iii) transitivity. Reflexivity refers to the view that an item, perhaps a competency, references itself in a cause/effect sense. Anti-symmetry demands that if one entity is a prerequisite of another, this relationship is not invertible; as an example, if competency x is a prerequisite to develop competency y, y cannot be the perquisite of competency x. Finally, transitivity means that whenever an element x is related to an element y, and y is in turn related to an element z, then x is also related to z. In principle, the direction of a graph is given by arrows of the edges; by convention however, the representation is simplified by avoiding the arrow heads, whereby the direction reads from bottom to top. In addition, the arrows from one element to itself (reflexivity property), as well as all arrows indicating transitivity are not shown in Hasse diagrams. The following image (Figure 1) illustrates such a diagram. Hasse diagrams enable a complete view to (often huge) structures. Insofar, they appear to be ideal for capturing the large competence or learning spaces occurring in the context of assessment and learning recommendations (for example, all the competencies involved in the math curriculum for a specific age).

In an educational context, a Hasse diagram can display the non-linear path through a learning domain starting from an origin at the beginning of an educational episode (which may be a single school lesson but could also be the entire semester). Moreover, the elements in the diagram may refer to (latent) competencies, to learning objects or test items. Figure 1 illustrates the simple example of typical learning objects in a certain domain. The beginning of a learning episode is usually shown as { } (the empty set) at the bottom of the diagram. Now a learner might attend three learning objects (K, P, H), which is indicated by the edges; this, in essence, establishes three possible learning paths. After H, as an example, this learner might attend K, or H but not T yet, which in turn opens further three branches for the learning path until reaching the final state, within which all learning objects have been attended.

As claimed initially, 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 [5, 6] 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.

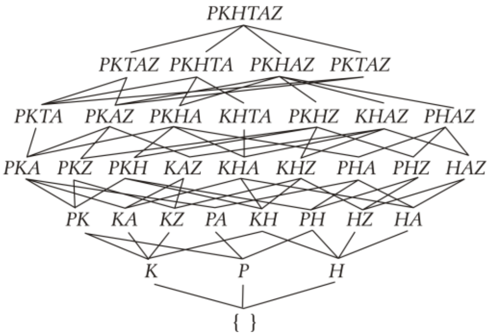

Figure 1. A simple Hasse diagram.

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.

3. VISUALIZING COMPETENCE SPACES

As claimed, 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.

3.1 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 2. 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,

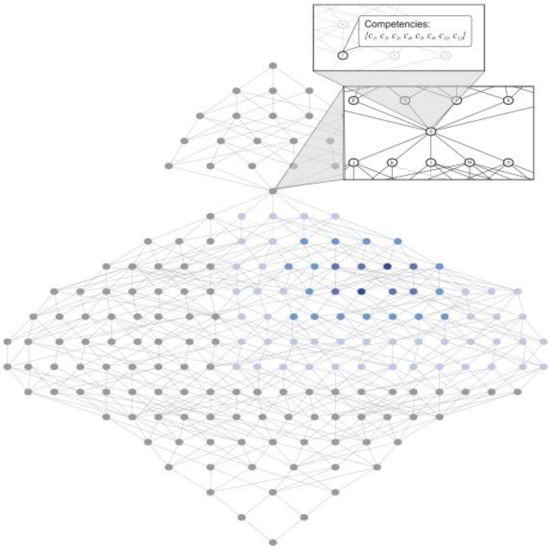

Figure 2. Hasse diagram illustrating the probability distribution over a competence space on three zoom levels.

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 [9], who altered the size of the nodes to represent the groups' sizes; the larger a node the more learners hold a particular state.

3.2 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 3 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 [9] – 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.

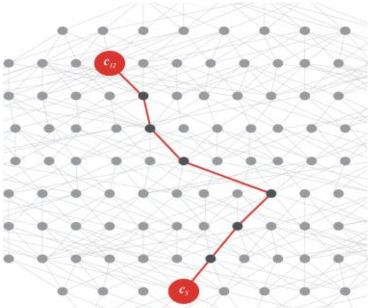

Figure 3. Learning Path. The cutout is part of the structure shown in Figure 2.

3.3 Tests and Recommendations

Hasse diagram offers information about two very distinct concepts, the inner and outer fringes. The inner fringe indicates what a learner can do / knows at the moment. Mathematically it refers to all sets of competencies, which hold all competencies of the current state but one. This inner fringe is a clear hypothesis of which test/assessment items this learner can master within the margins of a certain probability. Such information may be used to generate effective and individualized tests. The test generation can be complemented with group information. If an educator has very clear information in which competency areas of the space most of the learners are, she can generate or select test item covering exactly those competencies. The big advantage of such approach is the effectiveness of a test for identifying competency states or for ranking the learners can be maximized while the efforts for this evaluation (e.g., the number of test items) can be minimized. And of course the test can be optimized to differentiate different learners and the individual capabilities.

On the other hand, the outer fringes determine which competencies should be addressed in a next educational step. Mathematically is refers to all states which include all the competencies of the current state plus one. These fringes provide a clear set of recommendations about the most effective learning activities for a specific individual or a specific group of learners. Moreover, outer fringes, together with learning paths, allow specifically planning the most effective ways of reaching a specific learning goal (which not necessarily is the final stage of the competence space, the full set, and which is not necessarily the same goal for all individual learners).

3.4 Costs and Pace

When supporting teachers with information about learning processes, the concept of costs or learning pace (sometimes referred to as learning trajectories) is of distinct importance. Cost and pace can be considered as the time or any other measure of effort it takes to proceed from one competence state to another. In a Hasse diagram this information can be displayed by varying the length of the edges accordingly. If an educational leap requires a lot of efforts or time the edges are displayed proportionally longer than such that happens rather quickly. This method was introduced initially by [9]; an example is shown in Figure 4. Such information unveils criteria for the effectiveness of certain learning materials or acts of teaching. Particular outliers obviously pop out of the diagram and call educators to action to adapt teaching or teaching materials for a specific individual or a group.

3.5 Subordinate Concepts and General Notions of Achievement, Bottlenecks

A further important aspect in the context of LA is aligning the rather fine grained and low level approach to view competencies on a deeper level of granularity to more general concepts or rather superordinate notions of achievement. A general concept can be considered a higher level cluster of competencies; for example, sub-dividing mathematics into clusters like linear equations, non-linear equations, and vector arithmetic. Lower level competencies can be linked to one or more of those 'chapters'. Equally, one might view learning processes in a domain in terms of maturity. For example, writing skills can be on a low level of maturity, involving certain competencies and abilities, and on a higher one. Such approach is given, for example, in the CEFR language skills (cf. http://en.wikipedia.org/wiki/Common_European_Framework_of_Reference_for_Languages). Finally, teaching might involve the achievement of certain milestones, which should be reached step by step. Hasse diagrams allow identifying such milestones even if they were unclear or unknown initially. Considering that milestones as bottlenecks, i.e. unique competence states, each learning must pass, such bottlenecks immediately pop out in of the diagram. In a formative sense, it is easy for an educator to located their learners in their approach to or exceeding of such milestones (cf. Figure 2). A slightly different variant was introduced by [9] who used additional graphical elements (e.g., intersecting lines) to separate certain levels of maturity (whereas these authors used the CMMI[1] method; cf. Figure 5).

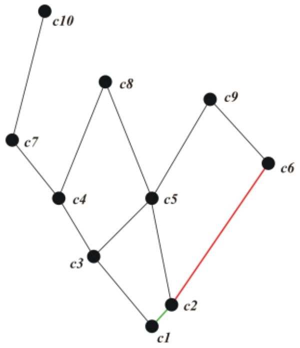

Figure 4. Illustrating learning efforts (as costs or pace). The longer the more efforts/time it took to acquire a further competency.

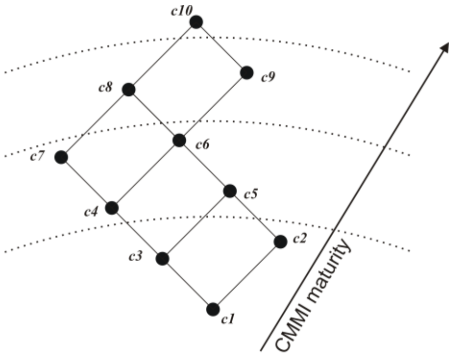

Figure 5. Illustrating maturity levels.

4. WHERE DO DATA COME FROM?

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.

5. 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

---

[1] CMMI refers to the so-called *Capability Maturity Model Integration* approach which models development processes (e.g., in production) on different predefined levels [3].

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.

6. ACKNOWLEDGMENTS

This work is based on the finalized project Next-Tell, which was supported by the European Commission (EC) under the Information Society Technology priority of the 7th Framework Programme for research and development as well as the running 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.

7. REFERENCES

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

[2] Ferguson, R., and Buckingham Shum, S. 2012. Social Learning Analytics: Five Approaches. In Proceedings of the 2nd International Conference on Learning Analytics & Knowledge, 29 Apr - 02 May 2012, Vancouver, British Columbia, Canada.

[3] Forrester, E. C., Buteau, B. L., and Shrum, S. 2009: CMMI for Services. Guidelines for Superior Service. Addison-Wesley.

[4] Dimitrova, V., McCalla, G. and Bull, S. 2007. Open Learner Models: Future Research Directions (Special Issue of IJAIED Part 2), International Journal of Artificial Intelligence in Education 17(3), 217-226.

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

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

[7] Duval, E., 2011. Attention Please! Learning Analytics for Visualization and Re-commendation. In Proceedings of the 1st International Conference on Learning Analytics & Knowledge, 27 Feb – 1 March 2011, Banff, Alberta, Canada.

[8] Luce, R. D. 1956. Semiorders and a theory of utility discrimination. Econometric,a 24, 178–191.

[9] 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.

[10] Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Buckingham Shum, S:, Ferguson, R., Duval, E., Verbert, K., and Baker, R.S..J.D. 2011. Open Learning Analytics: an integrated & modularized platform: Proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques. Available online at http://solaresearch.org/OpenLearningAnalytics.pdf