Agents and Analytics
A Framework for Educational Data Mining with Games based Learning

Harri Ketamo
Satakunta University of Applied Sciences, Tiedepuisto 3, Pori, Finland
Eedu Ltd, Satakunnankatu 23, Pori, Finland

Keywords: Educational Data Mining, Learning Analytics, Games based Learning, Artificial Intelligence.

Abstract: This paper focuses on data mining and analysis framework behind Eedu elements mathematics game. The background of the game is in learning-by-doing, learning-by-teaching and to some extent learning-by-programming. The data modelling behind the game is based on semantic networks. When all the skills and knowledge is modelled as semantic network, all the data mining can be done in terms of network analysis. According to our studies, this approach enables very detailed and valid learning analytics. The novelty value of the study is in games based approach on learning and data mining.

1 INTRODUCTION

Experienced teachers are aware that when a pupil is asked to teach another pupil, both pupils learn. This fact has not been applied enough in educational games, mostly because of a lack of technology and game AI that enables players to teach conceptually challenging themes still remaining easy-to-use game play. Furthermore, we know that children are ready to do more work for their game characters that what they are ready to do for themselves. This goes also for learning.

In terms of constructive psychology of learning, people actively construct their own knowledge through interaction with the environment and through reorganization of their mental structures. The key elements in learning are accommodation and assimilation. Accommodation describes an event when a learner figures out something radically new, which leads to a change in his/her mental conceptual structure. Assimilation describes events when a learner strengthens his/her mental conceptual structure by means of new relations (Mayer, 2004).

In economical game theory (Shoham and Layton-Brown, 2009) an agent behavior is widely studied in terms of Nash equilibrium. In this the agents are assumed to know the strategies of the other agents, and no agent has anything to gain by changing only its own strategy. A theory about existence of finite number of agents and their arbitrary relations based on other agent (Dukovska and Percikova, 2011) describes a set of attributes or properties that are useful when evaluating the agent behavior: 1) every agent is an entity, 2) every agent exists even it does not have a physical characteristics, 3) every agent chose to be in a state of direct knowledge with other agent according to its free will and 4) every agent is different from others in what it is.

Behavior modeling has a long research background: Neural and semantic networks, as well as genetic algorithms, are utilized to model a user's characteristics, profiles and pat-terns of behavior in order to support or challenge the performance of individuals. Behavior recording have been studied and used in the game industry for a good time. In all recent studies the level of behavior is limited, more or less, to observed patterns (Brusilovsky, 2001); (Houllette, 2003). Furthermore, agent negotiation and it's scripted behavior (Kumar and Mastorakis, 2010) as well as agent based information retrieval (Popirlan, 2010) in web-based information systems has been studied for a long time.

In this study, user behavior, competence and learning were seen as Semantic (neural) network that produces self-organizing and adaptive behavior/interaction. The behavior is evaluated in terms of the theory about existence of finite number of agents. The AI technology developed, emulates the human way to learn: According to cognitive psychology of learning, our thinking is based on conceptual representations of our experiences and

Ketamo H..
Agents and Analytics - A Framework for Educational Data Mining with Games based Learning.
DOI: 10.5220/0004331403770382
In *Proceedings of the 5th International Conference on Agents and Artificial Intelligence* (ICAART-2013), pages 377-382
ISBN: 978-989-8565-39-6

relations between these concepts. Phenomena when the mental structure change is called learning.

The data mining and analytics are based on this semantic modeling. When all the skills and knowledge is recorded as semantic network, all the mining can be done in terms of network analysis.

The novelty value of this study is in approach: to build games based technologies that enable easy construction of intelligent and human like behaviours and so enables detailed analysis of learning achievements.

2 EEDU ELEMENTS-GAME

The background of eedu elements is in learning-by-doing, learning-by-teaching and to some extent learning-by-programming. The approach is learner centric: the game introduces mathematics in a way that learner can build his/her mental conceptual structures by adding new concepts into known ones.

Technically it is relatively easy to produce games, but designing games that are pedagogically valid, and still attracts pupils, is challenging. No matter what is the technological implementation of game, the story behind the game is the key element for motivational game play. That's why interactive exercises can't be converted into games by just adding background characters. Nevertheless, entertaining games can't be converted into education by only adding calculator instead of guns; that breaks the story.

Eedu elements connects learner into things they can experience on daily basis when teaching knowledge for their game characters. The game characters learn like humans do: inductively case-by-case by building relations between new and existing concepts. The AI consists of teachable agents: Each game character is a teachable agent that learns through interactions and evaluations during the gameplay. Computationally the AI is based on semantic neural networks. The advantage of the method is in extensibility and scalability of learning: the AI can learn knowledge, behavior and strategy even in undefined domains (Ketamo, 2011).

The background of eedu elements is in learning-by-doing, learning-by-teaching and to some extent learning-by-programming. The approach is learner centric: the game introduces mathematics in a way that learner can build his/her mental conceptual structures by adding new concepts into known ones. According to cognitive psychology of learning, people actively construct their own knowledge through interaction with the environment and through reorganization of their mental structures.

When the player is responsible for character's mental development, he/she records also his/her mental conceptual structure during the gameplay. Eventually, we can say that while teaching his/her virtual character, learner reproduces a conceptual network about his/her mental conceptual structures.

A teaching phase consists of a question creation and evaluation – pair. Each teaching phase adds new relations into the conceptual structure. Furthermore, if the concept is not taught before, the new concept is also added into the conceptual structure during the teaching phase. The following example briefly describes the development of conceptual structures in the agent's mind during teaching phases. The understanding of how an agent's conceptual structure develops during playing is important in order to be able to interpret the results of the study. Each teaching phase is recorded in a semantic (conceptual) network within the game AI with one or more 'is (not/option) related to', 'is (not) bigger', 'is (not) equal', etc. relations. The following example is based on is (not) bigger and is (not) equal relations.

At first, the player teaches the relation between 1 and 1/2. The question, created by the player is: "Is ½ smaller than 1?" The agent does not have previous knowledge, so it will guess. In case it guesses "true" and the player's evaluation is "Correct." The relation "½ is smaller than 1." is formed in the conceptual structure (Figure 1a). The same would occur in a case where the agent guesses "False" and the player evaluates "Wrong".

In the second teaching phase, the player teaches a relation between 0.3 and ½, with the question "Is 0.3 bigger than ½?" The player knows that the question is false, but the agent answers (guesses) "True". So the player evaluates "wrong" and the agent determines that the correct answer is either "0.3 is equal to ½" or "0.3 is smaller than ½". The conceptual network in the agent's mind grows by both of these relations (Figure 1b).

In the third teaching phase a player forms a question in another way and asks "is 0.3 equal to ½?". Again, we know the statement is false. The agent can guess that statement is either "true" according to an "is_equal_to" relation or "false" according to a "is_smaller_than" relation. The agent guesses "false". When the player evaluates the answer as "correct", the agent determines that correct answer must be either "0.3 is smaller than ½" or "0.3 is greater than ½". After adding relations into conceptual structure, the agent knows that the correct answer is “0.3 is smaller than ½” because it is the mode (average) relation (Figure 1c).

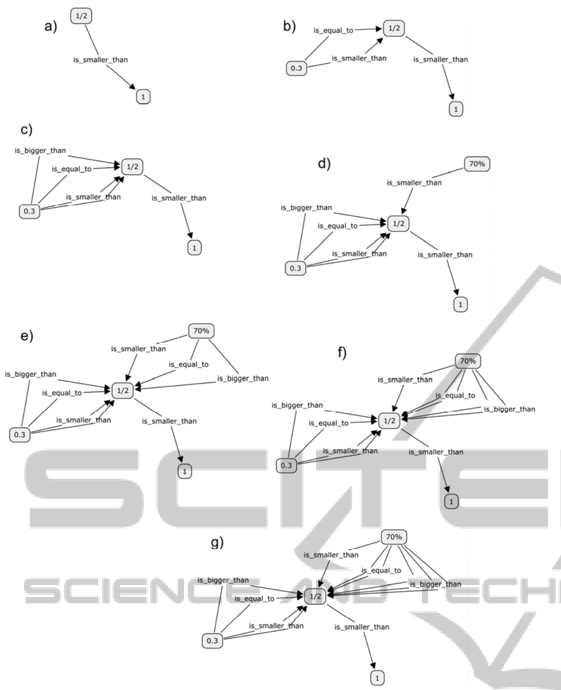

Figure 1: Semantic network and its development during the teaching phases.

In the fourth teaching phase the player asks, “Is 70% smaller than ½?” and on purpose, s/he teaches it the wrong way. The agent guesses that the statement is “true” and the player evaluates the answer as “Correct”, which forms an “is_smaller_than” relation in the conceptual structure (Figure 1d).

In the fifth teaching phase the player starts to correct the conceptual structure. S/He asks again,“ Is 70% smaller than ½?”. According to previous teaching, the agent knows that the answer is “true”. Because the player now knows that it is incorrect answer, the player evaluates it as “incorrect”. In this case the agent determines, that 70% must be equal to ½ or 70% must be greater than ½. After adding relations, the conceptual structure has all the possible comparing statements (Figure 1e) and basically behaves like an empty structure.

In the sixth teaching phase, the player asks for the third time, “Is 70% smaller than ½?”. Because there is no strongest relation, the agent guesses “true”. The player evaluates it again as “incorrect”. Again, the agent determines, that 70% must be equal to ½ or 70% must be greater than ½ and adds those relations to the conceptual structure (Figure 1f).

In the seventh teaching phase, the player decides to change the question to, “Is 70% more than ½?”. The agent guesses “True”, because ‘is_equal’ and ‘is_greater_than’ do contain the same probability. The player confirms that the answer was correct and one more “is_greater_than” relation was added into the conceptual structure (Figure 1g). After that the agent knows that the correct answer is “70% is greater than ½”, because such a set of relations are the strongest.

Figure 2: Eedu elements UI.

Technically eedu elements is a client-server solution where the client operates in a presentation layer (graphics, sounds and user interface) and the server operates with game mechanics and artificial intelligence (AI). This kind of architecture enables different devices and user interfaces (UI) connect to the game. In eedu elements, the UI is build with HTML5 and optimized for iPad, so it is compatible with browsers that implements full HTML5. Unfortunately at this point, only Chrome and Safari works perfectly and Firefox do have some minor challenges. Most important advantage is that it is possible to produce native applications from HTML5 to iOS, Android, Windows Mobile and MeeGo (figure 2).

One of the special focuses has been scientific proof of concept: The educational outcomes as well as motivation towards teaching virtual pets has been studied under laboratory experiment settings. In general, more than 60% of players increases their skills remarkably during the two hours gameplay (Kiili et al., 2011). The outcome in natural learning environment with possibility to longer gameplay is even greater: In fact, the best outcome is achieved when there are enough breaks and informal discussions between game play (Ketamo and Kiili 2010).

The most important finding is that assessment done according to learning data collected during the game play correlates with assessment done with traditional paper tests (Ketamo, 2011). Because of this, we can produce detailed diagnostic information about learning. This assessment information is meant for parents and teachers, not for the children.

3 EDUCATIONAL DATA MINING

Games and other virtual environments can provide relevant and meaningful information for individual learner, his/her parents, teachers and finally for educational system in an national level. In following we focus on 1) in-game analytics for player, parents and teachers and 2) analytics tool for national curriculum development.

In-game analytics tool (figure 3) is meant for parents or teachers to quickly observe what learner has taught for his/her pet. The visualization shows correctly taught concepts in the upper part of the skills -area and wrongly taught concepts in the lower part of the area. The quantity of the teaching is visualized in a way that concepts that are taught a lot appears in the right side of the area and little taught concepts on the left side. Quantity of teaching also mens that what more relations a concept do have, that more right it is located. Concepts that has not been taught do not appear in the skills -area.

When focusing on dependencies between the taught of conceptual structure and pupils achievements measured with traditional paper tests, we can find out that the taught conceptual structure is strongly related to paper tests score received after game play ($0.4<r<0.7$) with all tested content on mathematics and natural sciences. This is an important result in terms of reliability of the game as assessment/evaluation instrument.

In the game, the content in one level represents approximately one school week in Finnish school. Player can get one to three stars when completing the level. One star represents satisfactory skills, three stars represent good skills.

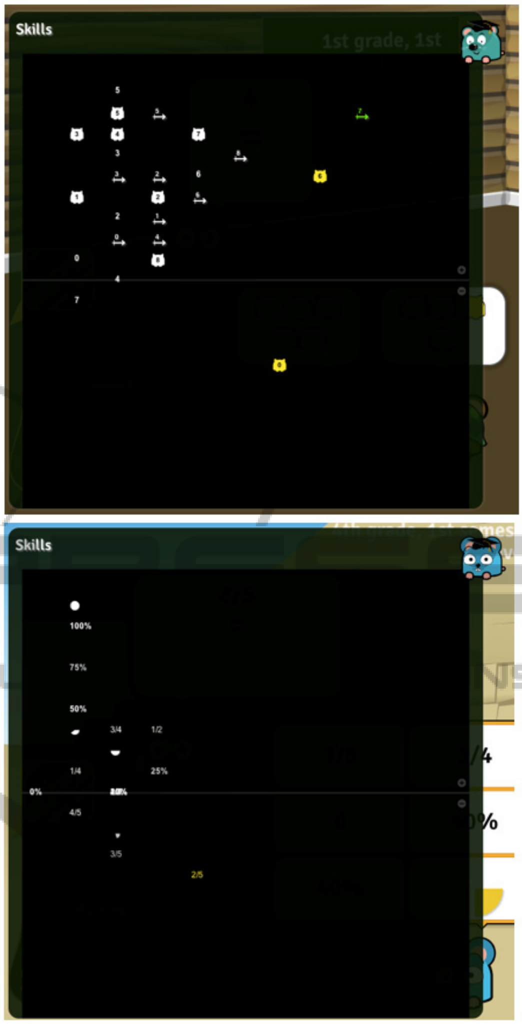

Figure 3: In-game analytics tool. On the upper screenshot a relatively good progress in 1st grade number concepts. On the lower screenshot a parent or teacher can observe difficulties with odd nominated fraction numbers.

However, the results of the gameplay are always a bit fuzzy: player can have just good luck and receive three stars with two stars performance. Furthermore, once and a while a nearly perfectly taught game character can have non-optimal performance because of one difficult task. So the evaluation/assessment with eedu elements in a single level is only indicative, but completing a whole grade requires skills that would be required to pass the same grade in a Finnish school.

In figure 4 a summary on progress in 1st grade according to Finnish curriculum is visualized. The idea is to show the pupil, parents and teacher the current position in game play and progress in terms of curriculum. This progress in terms of curriculum also shows diagnostic assessment: green characters represents good skills, yellow characters represents average or satisfactory skills, while red characters shows themes that are not opened or not completed yet.

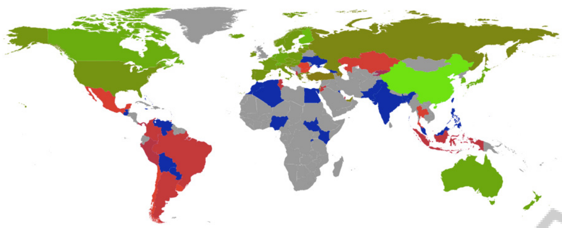

Figure 4: World scale analytics.

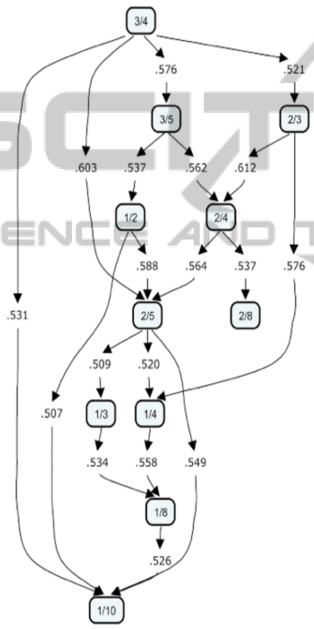

Figure 5: Misunderstood numbers and the strongest dependencies between misconceptions.

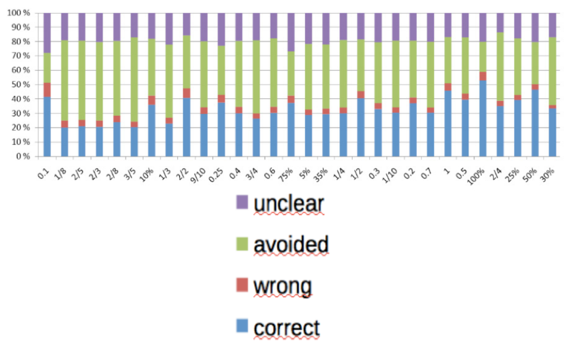

Figure 6: Frequencies on correct answers, wrong answers and avoiding the number. Unclear means that in some cases an individual player has understood such number correctly while in other cases he/she has not.

Analytics for national curriculum development. When summarizing the individual game achievements, schools and national level policy makers can receive analysis about competences and skills in general level. They can apply this in order to develop their teaching instructions or formal curriculum. Our goal is not to rank countries, we'll provide information for developing the practice. The full analytics shows all the countries we do have data to analyze (Figure 4).

We apply PISA data in general positioning, but when going inside our data, the analytics are that detailed that we can point out general bottlenecks of education. No matter how good some country is in PISA, there is always something to improve: e.g. in Finland there is an interesting bottleneck related to fraction numbers with odd nominator (figure 5). These numbers mediates or connects nearly all difficulties related to converting numbers between decimal numbers, fraction numbers and percent numbers. In other words, in Finland we should pay attention on how to teach odd nominated numbers.

When going deeper in details, wrong answers or misconceptions are not the only relevant factor explaining learning outcome. According to data received form gameplay, avoiding number (or concept) indicates directly poor performance in such concept. In figure 6 some of the numbers and frequencies avoiding the numbers during the gameplay are presented. In fact we can see that once again the most avoided numbers are the odd nominated fraction numbers.

4 CONCLUSIONS

Games and interactive virtual environments can offer much more than just entertainment, they can provide relevant and meaningful information for individual learner, his/her parents, teachers and even for whole educational system in an national level. This, however, requires careful planning and years of research on game design and game and learning analytics.

In Eedu elements, the game itself and data modeling is designed to support educational data mining. The analytic tools are embedded into the game and they provide real time analysis on learning process, difficulties in learning and challenges in curriculum. Future research consist of (big) data collection and experimental studies in order to validate the framework in real life context.

One major challenge within educational games is fragmentation: as long as there are one game for multiplying, another for subtracting and third for geometry, we can be sure that games will not produce any added value on learning analytics or curriculum design.

REFERENCES

Brusilovsky, P. (2001). Adaptive Hypermedia. User Modeling and User-Adapted Interaction, vol 11, p. 87-110.

Dukovska, S. C. & Percinkova, B. (2011). A model that presents the states of consciousness of Self and Others. *International Journal of Mathematical Models and Methods in Applied Sciences*. Volume 5(3), pp. 602-609.

Houlette, R. (2003) Player Modeling for Adaptive Games. In Rabin, S. (ed.) AI Game Programming Wisdom II. Massachusetts: Charles River Media, Inc.

Ketamo, H. & Kiili, K. (2010). Conceptual change takes time: Game based learning cannot be only supplementary amusement. *Journal of Educational Multimedia and Hypermedia*, vol. 19(4), pp. 399-419.

Ketamo, H. (2011). Sharing Behaviors in Games and Social Media. International *Journal of Applied Mathematics and Informatics*, vol. 5(1), pp. 224-232.

Kiili, K., Ketamo, H. & Lainema, T. (2011). Reflective Thinking in Games: Triggers and Constraints. In Connolly, T. (ed.) Leading Issues in Games-Based Learning Research. Ridgeway Press, UK, pp. 178-192.

Kumar, S. & Mastorakis, N.E. (2010). Novel Models for Multi-Agent Negotiation based Semantic Web Service Composition. WSEAS Transactions on Computers. Volume 9(4), pp.339-350.

Mayer, R. (2004) Should there be a three-strikes rule against pure discovery learning? American Psychologist, 59,14-19.

Popirlan, C.I. (2010). Knowledge Processing in Contact Centers using a Multi-Agent Architecture. WSEAS Transactions on Computers. Volume 9(11), pp. 1318-1327.

Shoham, Y. & Leyton-Brown, K. (2009). Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations. New York: Cambridge University Press.