Adaptive Learning and the Use of Knowledge Graphs As Personal Data Repositories

by

David Sampson

INTE 6720

University of Colorado Denver

Introduction

Research suggests that adaptive learning (AL) will be an important part of educational environments in the near future, but implementing adaptive learning systems as a learning tool can be a difficult process due to the complexity of the AL environments (Bagheri, 2015; Bienkowski, Feng, & Means, 2012; Siemans, 2013). As Bagheri (2015) explains, adaptive learning systems use technology, learning theories, and psychological insights to create instructional systems which adapt their appearance, content, content delivery, and assessment to the knowledge level and personal traits of the learners. Siemans (2013) describes how data about learner performance, preferences, and abilities is an important part of adaptive learning systems (ALS), but due to the complex and proprietary nature of the data, it is difficult to share the data between adaptive learning systems. According to Bagheri (2015), most adaptive learning systems are either large, costly systems developed by private companies, or in research environments where the data is difficult to access.

In 2012 the U.S. Department of Education (DOE) published an issue brief through their Office of Educational Technology that focused on learning analytics and educational data mining (Bienkowski et al., 2012). In the issue brief they identified a number of issues that need to be addressed with adaptive learning systems, including data sharing between different systems, development of data analytic tools that can be used and understood by inexpert users, and learning analytics that are available on small scale systems (Bienkowski et al., 2012, pp. 47 - 49). The authors also identified the importance of empowering the individual learner in the management and use of their own knowledge data.

This review will look at the potential for establishing a standardized knowledge graph that can be transported between different adaptive learning systems and is owned by the individual learner. As Kerr (2015) explains, knowledge graphs (KG) are a method of representing complex data structures that reflect the knowledge and skills of individuals. Knowledge graphs can be used as part of the learning analytics that facilitate the adaptive processes in an ALS. Wilson and Nichols (2015) describe the use of knowledge graphs by Knewton, one of the companies currently attempting to implement large scale adaptive learning systems, as “a way to diagnose student understanding (and misunderstanding). It can also power intelligent tutoring strategies” (p. 6). As described by Kerr (2015), knowledge graphs can be used by teachers, researchers, or institutions to map concepts, link learning objects (LO), or create information repositories about knowledge and knowledge gaps.

As proposed in this review, a personal standardized knowledge graph could be an important component for addressing the issues identified in the DOE brief by creating a mechanism for sharing learner data between adaptive learning systems large and small, providing an easy to understand visualization of the learner’s knowledge, and giving the individual learner control of when/where they share their data. This review will also include an investigation into some of the important issues facing the adoption of this type of standard, including data security, KG development, tracking of life-long learning information, and knowledge verification.

Adaptive Learning and Data

The Complexity of Shared Data in Adaptive Learning Systems

Bienkowski et al. (2012) recognized that there are a number of issues that will significantly affect future development of adaptive learning systems. Currently, there is very little interoperability between existing adaptive learning systems which means that sharing data between systems is a difficult and expensive activity that requires specialized skills. In addition, they found that student data is often found in a hierarchy of systems ranging from individual classroom systems with task specific information up to large institutional systems that may also include demographic information, comparative information, or other aggregate data. Each of these data sets may have unique and complex characteristics that are not compatible with other systems.

Siemans (2013) observes that a single source of data is insufficient if learning is to be considered a holistic and social process. In addition, he found that a learner's interactions with distributed systems create different "identities" which make it difficult to determine how each of these identities map to individuals in other systems. Dispersed systems may not be available to analyze and create a coherent picture of the learner. According to Greller and Drachsler (2012) there currently exists a proliferation of systems creating data that could be relevant to adaptive learning processes, including learning management systems (LMS), e-portfolio systems, personal learning environments (PLE), and other systems. The data in these systems are not readily accessible for use in adaptive learning environments. Greller and Drachsler (2012) explain how the linking of these datasets would facilitate the development of more learner oriented systems that could lead to improved personalization of learning environments. Although there are movements toward creating open access of educational data sets, the authors describe how many educational institutions are reluctant to share the data due to perceived legal liabilities.

Hendler and Berners-Lee (2010) contend that current systems are limited because they are developed in isolation and that in order to solve the problems that our society faces technology needs to be developed that allows user communities to construct, share, and adapt "social machines" that can evolve through trial, use, and refinement. Through the efforts of entire communities of users these social machines can evolve at a much faster rate than typical systems developed by a single entity. They also contend that models need to be developed that allow people to explore information at a higher level of abstraction, or "at the level of a global graph of interconnected people and ideas" (p. 157). Hendler and Berners-Lee (2010) also suggest that there needs to be the development of mechanisms that allow the users of these social machines to freely share data without worrying if it's used in inappropriate ways.

Siemans (2013) also asserts that the development of adaptive learning systems that accurately reflect the complexity of the learning process requires the development of systems that can generate learner models or profiles. However, these profiles are typically static snapshots that don't necessarily reflect the current state of a learner's knowledge. In order to keep the data relevant, the systems need to include the capacity to provide information from observation and human intervention. According to Bienkowski et al. (2012), in addition to the highly skilled people needed to implement and manage these systems, there also need to be people trained to interact with the data and make sense of the data. "Smart data consumers" can be teachers, administrators, or the students themselves who are able to interpret data, determine the learning priorities, recognize problem areas, and decide on reporting needs. Bienkowski et al. (2012) also point out that it is important that the data in these systems is secured and not shared with anyone who does not have permission to access it.

Slade and Prinsloo (2013) assert that, "Providing an adaptive and interactive environment tailored to the learner's needs is one of the most important areas of research in e-learning environments" (p. 461). But with the creation and implementation of these adaptive learning systems, there are also the ethical challenges of interpreting the data correctly, getting informed consent, addressing privacy concerns, deidentification of data, and determination of the correct learning path. Slade and Prinsloo (2013) also suggest that providing a learner with an adaptive learning path that is optimal to their learning preferences, learning styles, and existing knowledge is fundamental to the implementation of a successful adaptive learning system.

Recommendations for Research and Development in Adaptive Learning Systems

In their 2012 report titled *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief,* the U.S. Department of Education made numerous recommendations for research and development that are relevant to learning analytics, data mining, and adaptive learning systems (Bienkowski et al., 2012, pp. 47 – 49). More specifically, a number of their recommendations are related to creating a data model that can be shared by different adaptive learning systems, provide a visualization mechanism for non-technical users, be implemented in small-scale systems in order to have wide spread impact, and ensure that proper ethical standards are adhered to in the systems. The report included the following recommendations:

- Develop models for how learning analytics and recommendation systems developed in one context can be adapted and repurposed efficiently for other contexts…

- Continue to perfect the anonymization of data and tools for data aggregation and disaggregation that protect individual privacy yet ensure advancements in the use of educational data…

- Develop decision supports and recommendation engines that minimize the extent to which instructors need to actively analyze data…

- Conduct research on the usability and impact of alternative ways of presenting fine-grained learning data to instructors, students, and parents. Data visualizations provide an important bridge between technology systems and data analytics, and determining how to design visualizations that practitioners can easily interpret is an active area of research…

- Data mining and analytics can be done on a small scale. In fact, starting with a small-scale application can be a strategy for building a receptive culture for data use and continuous

> improvement that can prepare a district to make the best use of more powerful, economical systems as they become available… (Bienkowski et al., pp. 47 – 49).

The Emergence of Knowledge Graphs in Large Scale Adaptive Learning Systems

Czopek and Pietrzak (2016), in their review of early efforts at adaptive learning systems, found that early research systems were not scalable to the massive amount of content across all educational domains. Because the systems were limited to the research domains, the number of students that could benefit from the systems was very limited. As they describe it, one of the main reasons for the limited scope was the problem of modelling student knowledge and understanding. According to Czopek and Pietrzak (2016), it was the emergence of the Knewton system in 2013 that provided adaptive learning solutions on a large scale, and it was the use of knowledge graphs that allowed for student modelling across content.

Wilson and Nichols (2015), two data scientists who work at Knewton, describe how their ALS uses knowledge graphs, "to address the problems of student modeling that focuses directly on scalability across content…" (p. 6). As they describe it, the Knewton knowledge graph uses a flexible and expressive ontology which allows a diverse set of content to be easily represented and connected with each other. Or as they state, "…the Knewton knowledge graph [is] a core instrument in enabling Knewton's scalability across content domains" (p.7). As described, these Knewton KGs can be automatically generated and content data can be combined with student interaction data to provide a comprehensive representation of the learner's knowledge within a content domain.

In the next section I will introduce personal knowledge graphs which have the potential to provide adaptable data models, protection of privacy, ease of use, data visualization, and the ability to integrate into small scale systems. In addition, knowledge graphs can be implemented with more advanced features such as automatic question generation, learning object testing, and user authorization controls which will make their use more accessible to teachers, administrators, and students. And significantly, knowledge graphs have the potential, given the development of appropriate standards, to be transportable from one knowledge domain to another knowledge domain, and from one adaptive learning system to another adaptive learning system.

Knowledge Graphs as Personal Data Repositories

Knowledge Graphs

Kerr (2015) explains that adaptive learning assumes that what is learnt can be analyzed and defined at a granular level (or small learning objects), and these objects can be organized into a knowledge graph. Sophisticated adaptive learning systems can be dynamic and allow for some modification of the knowledge graphs as a learner interacts with the material. By combining an individual's knowledge graph with an existing knowledge domain, a more complete picture of the learner's knowledge can be established.

Yan, Wang, Cheng, Gao, and Zhou (2016) describe knowledge graphs as a semantic graph consisting of nodes and edges. The nodes (also known as learning objects) represent concepts or entities. A concept refers to general categories such as house, doctor, etc. An entity is a physical object such as a specific person or location. The edges represent the semantic relationships between the objects. Entities and concepts can be connected to form a complete and structured knowledge repository. The structure of the knowledge graph facilitates the management, retrieval and understanding of the information it contains.

According to Yan et al. (2016), knowledge graphs are able to infer the conceptual meanings of data which can lead to better query results. In addition, KGs can serve as repositories of structured knowledge that support applications related to big data analytics. Some examples of applications that use KGs to support related concepts are search engines such as Google and Bing, e-commerce companies such as Walmart's product search and recommendations, and Facebook's Graph Search. Yan et al. (2016) also describe efforts underway to create large knowledge graph databases such as Freebase that consist of repositories built by crowd sourcing and contain millions of concepts and tens of millions of relationships between those concepts.

In their report on building knowledge bases (the underlying structure of knowledge graphs), Deshpande et al. (2013) describe how knowledge bases (KB) can be divided into two types of KBs. Domain specific KBs are specific to a well-defined topic or domain of interest, such as algebra or Shakespeare. Global knowledge bases attempt to cover the entire world, such as Google or Facebook reference features. The authors also describe the difference between an "ontology-like KB [that] attempts to capture all important concepts, instances, and relationships in the target domain" vs a "source-specific KB … [that] are built from a given set of data sources…and cover these sources only" (Deshpande et al., 2013, pp. 1210 – 1211). If an ontology-like KB already exists, it is relatively easy to build a source-specific KB in the same domain because the ontology-like KB can be used as a domain dictionary to help locate important information from the given data sources.

Personal Knowledge Graphs

Based on the model described by Deshpande et al. (2013), it is conceivable that different entities (students, teachers, institutions, etc.) could own a knowledge graph that is unique to their needs. Institutions could own and manage a domain specific, ontology-like knowledge base that reflected the entirety of the institution's educational program. Teachers could own and manage a domain specific, ontology-like knowledge base that reflects the topics of interest to them. And students could own and manage a source-specific knowledge base that reflects their knowledge and skills, and that when merged with domain specific, ontology-specific KBs (teacher's or institution's), would provide an actionable picture of the learner's existing knowledge and skills as compared to the domain specific KBs. The resulting knowledge graph could be used in an adaptive learning system to provide learning processes that are specific to each learner's needs.

Personal knowledge graphs that are transportable between adaptive learning systems would be a significant step towards addressing the issue of single data sources that was identified in this paper (Bienkowski et al, 2012; Siemans, 2013; Greller and Drachsler, 2012). As the knowledge graph is updated by each adaptive learning system, the newly updated data can be transported to another system if that system is using a compatible knowledge graph structure. Using data knowledge base maintenance and update techniques mentioned by Deshpande et al. (2013), the learner's information can also be removed from an institution's knowledge graph at the request of the learner.

With a personal knowledge graph, learners could be in control of when, where, and with whom they share the data about their knowledge and skills. As Slade and Prinsloo (2013) state, "Not only should students provide informed consent regarding the collection, use, and storage of data, but they should also voluntarily collaborate in providing data and access to data to allow learning analytics to serve their learning and development, and not just the efficiency of institutional profiling and interventions" (p. 1519). If they change learning institutions, they would be able to bring their knowledge graph with them for integration into other adaptive learning systems. Personal knowledge graphs can also be used to address a number of other issues that have been identified with adaptive learning systems.

A Closer Look at Personal Knowledge Graphs

Bienkowski et al. (2012) also recommended that decision support and recommendation engines be developed that minimize the extent to which instructors need to actively analyze data. With personal knowledge graphs, instructors would be responsible for maintaining their domain specific, ontology-specific knowledge base, which could be updated and modified independently of the learner's KG. In addition, Seyler, Yahya, and Berberich (2015) propose the use of knowledge graphs to automatically generate natural language questions from the KGs. Using a knowledge graph, and comparing it to a topic in Wikipedia, they were able to generate questions and related answers of varying difficulty (easy or hard) that can be specific to the learner. Automatic question generation could further reduce the need for instructors to intervene in individual learner skills and free them up to concentrate on their domain of knowledge.

Sette, Tao, Gai, and Jiang (2016) propose the development and use of knowledge graphs that include advanced Learning Objects (LO) which include assessment as part of each LO in order to ensure that the learner sufficiently understands the learning object before being considered competent in the LO topic. These proposed learning objects can also include indicators as to whether the learner is considered a beginner, intermediate or expert in the LO topic. These indicators can then be used to match the learner to the appropriate level of learning activities in diverse knowledge domains that may or may not be in the same adaptive learning system. These LOs can either be accessed directly by the learners for modifications or access can be restricted depending on the needs of the learning systems.

Siemans (2013) mentions some of the issues associated with tracking a learner's lifelong learning activities, such as how a learner's data is handled when a learner leaves an institution, how long an institution holds on to the data, should the data be shared with perspective employers or other institutions, and what are the effects of privacy laws are on the ability to share data. As Siemans indicates "These and numerous equally intractable problems will need to be addressed." (p. 1394) The use of personal knowledge graphs will allow users to address some of these issues by tracking a learner's lifelong learning activity by giving the learner a data set that they own and can keep with them for as long as needed.

A number of significant issues still need to be addressed in future research in order for personal knowledge graphs to be accepted in the education community, including implementation of standardized knowledge base structures that would allow for the merging of data. Currently, there are efforts by the National Center for Education Statistics to create and encourage the use of standards for student data, specifically the Common Education Data Standards (https://ceds.ed.gov/), which will help with this effort. Data verification practices that ensure that the data is accurate and learners haven't tampered with the data will become increasingly important with transportable data. And training users on the visualization tools that will allow them to understand the knowledge graph will be a key to successful implementation.

Conclusion

The development and implementation of adaptive learning systems is an important trend in educational technology because they provide a personalized learning experience that can adapt to the learners existing knowledge, preferences, and needs. These systems are mainly based on recent developments in learning analytics and educational data mining as describe by the U.S. Department of Education. One of the significant issues facing developers of adaptive learning systems is the exclusive nature of the data as it resides in discrete systems. To get a more accurate picture of the knowledge and skills of learners, it is important to be able to incorporate the data from a variety of systems into a single data structure.

This paper argues that knowledge graphs are the data structures that are the most compatible with personal ownership of learning objects. Knowledge graphs are already used in a number existing ALS, and can be modified to reflect the complex knowledge and skills of individual learners. In addition, knowledge graphs are structured such that personal learner data can be an independent entity that can be integrated into a variety of different knowledge domains, making the personal knowledge graph a potential means of transporting data between different adaptive learning systems.

A number of issues still need to be addressed before personal knowledge graphs can become a reality. But the potential benefits of adaptive learning systems that can adapt to a learner's lifelong learning quest, and is controlled by the learner themselves, is worth further research.

References

Bagheri, M. (2015). Intelligent and adaptive tutoring systems: How to integrate learners. *International Journal of Education, 7*(2), 1 – 16. doi: 10.5296/ije.v7i2.7079

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *US Department of Education, Office of Educational Technology*, 1-57.

Czopek, A., & Pietrzak, P. (2016). Unlocking the potential of technology in education. *e-mentor, 3(*65*),* 78-82.

Deshpande, O., Lamba, D., Tourn, M., Das, S., Subramaniam, S., Rajaraman, A., Harinarayan, V., & Doan, A. (2013). Building, maintaining, and using knowledge bases: A report from the trenches. *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data.* 1209 – 1220. doi: 10.1145/2463676.2465297

Greller, W. & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society, 15*(3), 42-57.

Hendler, J., & Berners-Lee, T. (2009). From the semantic web to social machines: A research challenge for AI on the World Wide Web. *Artificial Intelligence, 174*(2), 156 – 161. doi: 10.1016/j.artint.2009.11.010

Kerr, P. (2015). Adaptive learning. *ELT Journal, 70* (1), 88 – 93. doi: 10.1093/elt/ccv055

Sette, M., Tao, L., Gai, K., & Jiang, N. (2016). A semantic approach to intelligent and personal tutoring system. *3rd International Conference on Cyber Security and Cloud Computing, 3*, 261-266*.* doi: 10.1109/CSCloud.2016.50

Seyler, D., Yahya, M., & Berberich, K. (2015). Generation quiz questions from knowledge graphs. *Proceedings of the 24th International Conference on World Wide Web,* 113 -114. doi: 10.1145/2740908.274722

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, *57*(10), 1380 – 1400. doi: 10.1177/0002764213498851

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1510 – 1529. doi: 10.1177/0002764213479366

Wilson, K., & Nichols, Z. (2015). *The Knewton platform: A general-purpose adaptive learning platform*. Retrieved from https://www.knewton.com/wp-content/uploads/knewton-technical-white-paper-201501.pdf

Yan, J., Wang, C., Cheng, W., Gao, M., & Zhou, A. (2016). A retrospective of knowledge graphs. *Frontiers of Computer Science,* 1 – 20. doi: 10.1007/s11704-016-5228-9