Knowledge Discovery and Data Mining Models

MODELING EDUCATIONAL PROCESSES IN MODERN SOCIETY BY NAVIGATING MULTIDIMENSIONAL NETWORKS

Sergey Maruev, Eugene Levner, Dmitry Stefanovskyi, Alexander Troussov

***Abstract**: To graduate from a university and receive a diploma the student must follow curricula, have good command of certain topics, pass certain tests and exams. All the above mentioned artifacts of educational processes could be viewed as nodes in a large network where nodes of various kinds are connected by typed arcs, indicating, for instance, that the knowledge of a particular book or a research paper is required in a particular item of a particular curricula, or that before enrolling for a particular examination one needs to pass through particular tests. In this paradigm the process of education becomes the navigation from the initial nodes corresponding to the student knowledge and qualifications to the nodes which represent her goals. For some students the goal could be just one node representing diploma, for other students, especially for self-motivated life-long learners, the goal is a set of nodes.*

In this paper we present the initial results in modeling educational process as the navigation in multidimensional networks and the pertaining algorithms of optimization of that navigation. The results of our research could be useful for the building of educational resources (for instance, by finding structural weakness in existing networks), as well as for the personalization of the education.

The practical importance of our research stems from the processes of globalization, personalization of education, and from the explosive growth of the availability of good quality on-line training courses. Multidimensional networks of educational processes in modern are huge, naïve (common-sense) navigation tools are not sufficient for their analysis, and new computer-based navigation tools are to be designed.

***Keywords**: big data, graph-based methods, education.*

***ACM Classification Keywords**: Algorithms, Economics, Education, Experimentation, Theory.*

Introduction

We live in the interconnected space of socio-technical systems, where layers of technological infrastructures interact with the social context, which drives their everyday use and development. Most of the content is generated in public systems like LinkedIn, Facebook, Delicious, Twitter, blogs and microblogging systems, as well as in the social software used in the enterprises. These socio-technological systems already transformed computer information systems in all domains of human activities: document collections became a highly interconnected socio-semantic space, where documents are shared, discussed and edited collaboratively, and are filtered following the interests of individual users and social groups.

This process is now penetrating the modern educational systems by the explosion of corresponding internet resources and on-line platform for the distribution. Massive Open Online Courses (MOOCs), developed in universities, corporation and analytical centers are capable to compete with traditional methods of education. Such decentralized big data approaches influence methods of education; they allow individualization of education, building different trajectories of education. The problem of navigating and choosing the educational trajectory becomes actual. The trajectory must depend on the background, interests and expectations of students, on the availability of resources and various features of students and educational resources.

Resources, concepts, documents, individuals and organizations form multidimensional network, where nodes represent actors, concepts and other artifacts; links are also typed and weighted [Troussov et al. 2011]. Trajectories of education form paths in this network. This representation of modern types of education allows formalization and use of powerful abstraction provided by graph-based methods to optimize the trajectories according to given criteria.

The title of this paper has two notions characterizing the object of our investigation: Educational Processes in Modern Society. Both components are important and interdependent. In modern society many connections and relations function on the layers of technological structures and software. Interactions between agents become different, and these new types of interactions in their turn drive the future development of the technology. Educational processes evolve in the technological media, the trajectories becomes individual, though might be influenced by decisions taken by other students.

Wide spread of internet technologies and educational tools make actual the task of choosing the on-line objects relevant to educational curricula. Traditional methods for the solution of this problem, based on the titles and metainformation of the resources, do not allow discriminate, for instance, between textbooks using different approach, different conceptualization and granularity of topics.

Modern age information technologies drastically changed forms of communications, information retrieval and management, as well as social interactions between people. During last decade, these innovation affected the education. Explosive emergence of massive open on-line courses (MOOCs) demonstrates one of the trend inherent in modern education. More than 4,5 millions of students across the globe use on-line courses EdX, Udacity and Coursera [Carr, 2012]

Availability of huge amount of ready to use educational data leads to a more disruptive and far-reaching changes related to the notion of “big data”. According to [Guthrie, 2013], universities can customize courses and learning modules for student’s needs. At the same time, phenomenon of big data poses the question of how to use these opportunities in the big data environment.

Individualization of education becomes the key success factor in modern education [Robinson,2010]. The student is given the opportunity to choose between so many courses, learning materials and tutors according to their own interests. “An embarrassment of riches” - overabundance of new good opportunities - generates new troubles for students in navigating and building educational trajectories.

Enterprises and recruiting agencies now also are interested to detect how the qualification obtained by the individualized education corresponds to requirements. They need tools to evaluate and compare various courses and modules.

Learning trajectory is a complicated concept using in many domains. Educators and cognitive psychologists define learning trajectory as a sequence of knowledge units (paradigm, concepts, methods etc.) internalized by students.

Speaking on a different scale, one can formalize the learning trajectory, for instance, simply as the sequence of courses finished by a student. Learning trajectory naturally lends itself to the formalized trajectory in a corresponding network of educational artifacts. Nodes represent units of knowledge, courses and resources needed to get the knowledge. Graph-based methods allow addressing the problem of learning trajectory optimization, for instance, by minimizing the duration of education. By attaching additional information to network nodes, one can compute optimization using additional criteria. Network models allow to compute the traffic in the network, and to solve the capacitated transportation problem.

Greedy strategy in building the trajectory suggests the usage of the iterative process. On each step of iteration, the agent chooses the most suitable object. In doing so we must consider two levels, scales of analysis, as it is depicted on the Fig. 1. The trajectory is build on the higher level of objects, the optimal choice of the element must be done by analyzing objects on the micro level (which because of their complexity usually also requires application of graph-based methods).

Information technology has changed the usual forms of communication, information work and social interaction between people. In the past 10 years, these technological innovations are changing education. Phenomenon MOOC (massive open on-line courses) demonstrates one of the trends characteristic of modern education. More than 4.5 million students in the world use online courses EdX, Udacity and Coursera [Carr, 2012]. A huge amount of data available belongs to a more disruptive and far-reaching change - "big data". According to [Guthrie, 2013] universities can use big data to customize courses and learning modules for student's needs. Individualization of education is becoming a key factor in the success of modern education [Robinson, 2010]. Individualization of education is an opportunity for students to choose courses of study, teaching materials, teachers in accordance with their interests.

New possibilities give rise to new problems. Opportunity to take courses or other learning resources by different authors from different sources creates a problem of resource or courses selection on the condition of the student interests and therefore the problem of constructing a trajectory in the space of learning resources and navigation between educational resources.

Recruitment agencies also became interested in what the applicant studied courses meet the requirements of a career position. They need to know the content of student learning modules or courses in different universities or training centers and need a tool to solve this problem.

Learning trajectory is a complicated concept using in many meanings. Educators and psychologists call the learning trajectory sequence of elements of knowledge (ideas, concepts, methods, etc.) that develops student. Changing the scale we come to understand the learning trajectory as a sequence of courses that the student is studying. Learning trajectory is naturally represented as a graph. Nodes of the graph represent the knowledge, training or resources necessary for their development. Methods of the graph theory can solve some problems by optimizing learning trajectory, such as optimization of training time. Adding a description of the node of her prerequisites, you can automate the construction of possible learning trajectories. Graph models allow to analyze the capacity of the resultant structure [Maruev and Shilin, 2012] and evaluate the necessary resources [Maruev and Gorbunova, 2012] for her work.

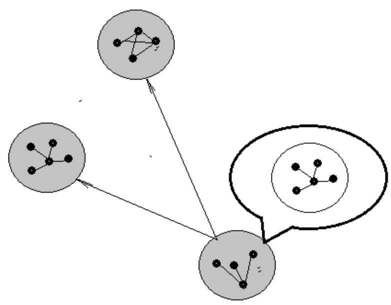

Fig. 1. Levels of modelling. Gray nodes – mezo-level, black nodes – micro-level.

Agent builds its learning trajectory as an iterative process. He selects the next element at each step of the process. Choice is optimal if the selected item is more resemble than other to the expectation of other agent. We are working on two levels of scale in Fig.1. Building a path of generalized elements on mezo-level and doing the optimal choice using their internal structure on micro-level.

Application of Graph-mining for the Selection of the Most Suitable Educational Resource

In this section we demonstrate the application of network modeling and graph-based methods for the selection of the most relevant educational module to cover a particular topic. To show the proposed methods in sufficient detail and to validate the results we used the following use case.

We took real life Russian language curriculum on the topic of macroeconomics, and two Russian textbooks on this topics; this curriculum and textbooks we will refer to as *C*, *T1*, and *T2*. We use *C* to automatically extract vocabulary for the macroeconomics and to build a network or semantic relations between concepts as it is seen from the text and the structure of the curricula. This vocabulary is then used to analyze *T1* and *T2* and to build networks corresponding to the relations between terms in these resources. We present a novel generic method for comparing graphs which allows us to quantify how well an educational resource covers the program.

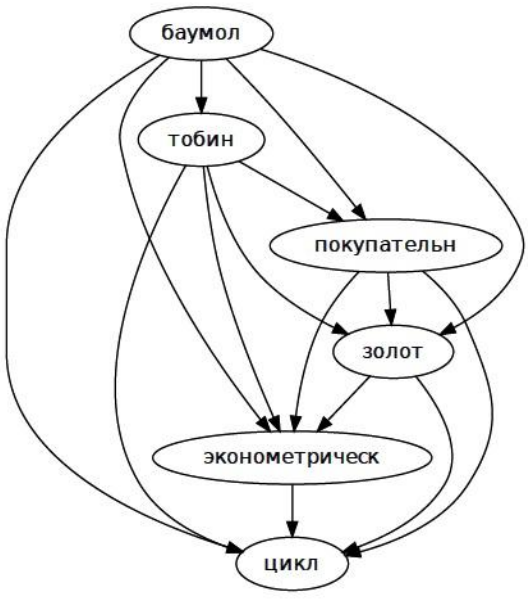

Fig. 2. A fragment of the network of concepts extracted from the curriculum on microeconomics. The nodes represent stems for several Russian words, terms and names of prominent economists (William Baumol, James Tobin, gold, econometrics, cycle). The links represent collocations of the words in the textbook 1. The whole configuration suggests, that the curriculum includes the Baumol–Tobin economic model of the transactions demand for money. Although our modeling doesn't use ontologies and recognition of multiword terms, our methods of modeling and mining are capable to indirectly capture and detect graph configuration of terms used in this model.

CONSTRUCTION OF THE NETWORK MODELS OF CARRICULUM AND EDUCATIONAL RESOURCES

We preprocess the text of the curriculum to find orthographic words, to filter out so called “stop” words (using Google's list of stop words for Russian), and to stem words to index forms using Porter stemmer for Russian (see, for instance, [Jurafsky and Martin 2009]). Words which are met within the window of three sentences are connected with an arc; the weight of the arc is calculated based on the number of the connections. An example of such collocation graph is shown on the Fig. 2.

Collocations per se do not represent semantic relations (since collocations captures various relations, including, for instance, sintagmatic relations), but the resulting network in the context of our study might be considered as semantic network.

MINING – FINDING STRUCTURAL SIMILARITIES IN NETWORKS WITH THE SAME SET OF NODES

Apparently, the global topology of these three network models – C, T1, and T2 - could be quite different. However, there must be some similarity at the level of the local topology, especially at the level of micro- and mezzo-clustering. To measure similarities which might be related to the coverage of topics, we propose the following methods.

We assume that the local clusterization in the graph *C* must be high. For instance, all multiword expressions like *real balance effect*, form complete graphs, where each word used in the multiword term is connected to another words in the expression. We also might put forward a hypothesis that if we select at random a set of concepts from the vocabulary of *C*, and compute the number of nodes which are within the distance two from the initial set in all three graphs *C*, *T1*, and *T2*, these numbers should be approximately the same. It is also clear that, from the other hand, if instead of a real text *T1* we'll take a random list of word forms from the vocabulary in *C*, most of the multiword terms will not be seen in the topology of such meaningless text. Therefore, If, for instance, this number is high in the network *C* and *T1*, but small in the network *T2*, we might suppose that a certain topic (or topics) of the curriculum is not covered in the textbook *T2* in sufficient details.

To find the neighborhood of a set of nodes, and to ensure the extensibility of the method to work with fuzzy sets of nodes, we use method of spreading activation (see, for instance, [Crestani 1997], [Troussov et al. 2009]) with two iterations. From the considerations above, we suppose that the high number of activated nodes in our experiments is a good predictor that the textbook covers the curriculum in sufficient details.

EVALUATION

The number of nodes in networks *C, T1*, and *T2* is the same and equals to 260. We split this number into 52 sets with 5 elements. Correspondingly, we generated 66 sets of nodes and used spreading activation to propagate the activation to neighbor nodes in networks *T1* and *T2*. The cumulated number of activated nodes in the network *T1* is 617, for the network *T2* this number equals to 1152. Therefore, according to our metrics, we conclude that the textbook *T2* is better learning object for the curriculum in the question. This result computed by our graph-based method metrics is consistent with the manual evaluation provided by experts in the field. However, the amount of the experimentation is too small to conclude that the feasibility of our method is proven.

Conclusions and Future Work

The practical importance of our research stems from the processes of globalization, personalization of education, and from the explosive growth of the availability of on-line training courses. Multidimensional networks of modern educational processes are huge; therefore naïve (common-sense) navigation tools are not sufficient for their analysis, and new computer-based navigation tools are to be designed.

The central element of the navigation tools is a block decision making choosing one of the many educational elements based on a comparison of their contents. If the elements exist in the form of free texts, someone must read, understand all the texts and make a decision based on comparison of them. Such endeavor is impossible when we move into big data. It is therefore necessary to develop methods of automated processing big volumes of training resources for the decision to select an element of a learning trajectory.

In this paper we present the initial results in modeling educational process as the navigation in multidimensional networks and the pertaining algorithms of optimization of that navigation. The results of our research could be useful for the building of educational resources (for instance, by finding structural weaknesses in existing networks), as well as for the personalization of the education.

The presented model of elements of a learning trajectory is a multidimensional network. We have developed a method of constructing such a multidimensional network and the algorithm for selecting the most similar element on each step of the iterative algorithm of a learning trajectory constructing. Our algorithm used the spread activation method, which has been successfully used for mining multidimensional networks. The efficiency of the algorithm is shown in the example of choosing a textbook for a particular program (curriculum). Further research in this direction involves experiments with big data, construction of models with more dimensions and improvement of the spreading activation algorithm (an example of a new strand of spreading activation algorithms is provided in [Troussov et al. 2011a]).

Bibliography

[Carr, N., 2012]. The Crisis in Higher Education MIT Technology Review magazine November/December 2012/ http:// technologyrReview.com/featuredstory/429376/ the-crisis- in-higher-education

[Crestani, F., 1997]. Application of Spreading Activation Techniques in Information Retrieval. Artificial Intelligence Review, 11(6), 453-482.

[Guthrie, D., 2013]. The coming Big Data Education Revolution. US News. http://www.usnews.com/opinion/articles/2013/08/15/why-big-data-not-mooc-will-revolutionize-education

[Jurafsky, D. and Martin, J.H., 2009]. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition (2ed.), Prentice Hall.

[Maruev S. and Gorbunova E., 2012b]. Analysis of the Throughput of the Process of Distance Learning. Artificial Intelligence Driven Solutions to Business and Engineering Problems. ITHEA Rzeszow-Sofia, 2012. P.7-11.

[Maruev S. and Shilin K., 2012a]. Model of Resource Potential Estimation for Quality of Education Ensuring. Economic Policy. #1, 2012, P.78-86. (In Russian)

[Robinson, K., 2010]. Changing Education Paradigms http://www.ted.com/talks/ken_robinson_changing_education_paradigms

[Troussov, A., Darena, F., Zizka, J., Parra, D., and Brusilovsky, P., 2011a]. "Vectorised Spreading Activation Algorithm for Centrality Measurement". Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis. sv. LIX, no. 7, s. 469--476. ISSN 1211-8516.

[Troussov, A., Jundge, J., Alexandrov, M., and Levner, E., 2011]. Social Context as Machine-Processable Knowledge. Proceedings of the International Conference on Intelligent Information and Engineering Systems INFOS 2011, Rzeszów - Polańczyk, Poland, pp. 104-114, ISBN: 978-954-16-0053-5.

[Troussov, A., Levner, E., Bogdan, C., Judge, J., and Botvich, D., 2009]. Spread of Activation Methods. In Dynamic and Advanced Data Mining for Progressing Technological Development, Y. Xiang and S. Ali (eds) IGI Global.

Authors' Information

Sergey Maruev – Professor, The Russian Presidential Academy of National Economy and Public Administration (RANEPA), Prosp. Vernadskogo 82, bld. 5, Moscow, 119571, Russia; e-mail: maruev@rane.ru

Major Fields of Scientific Research: mathematical modeling, economics of education, business process modeling, social network analysis, operations research

Eugene Levner – Ashkelon Academic College, 52, Golomb St, Holon 68102 Israel; e-mail: levner@Hit.ac.il

Major Fields of Scientific Research: combinatorial optimization, operations research, design and analysis of computer algorithms, algorithm complexity and computability, scheduling theory, grid optimization, network analysis, and risk analysis

Dmitry Stefanovskyi – Assoc. Prof., Ph.D, The Russian Presidential Academy of national economy and public administration (RANEPA), Prosp. Vernadskogo 82, bld. 5, Moscow, 119571, Russian Federation; e-mail: dstefanovskiy@gmail.com

Major Fields of Scientific Research: mathematical modeling, world economy

Alexander Troussov – Director of the International Research Laboratory for Mathematical Methods for Social Network Mining, The Russian Presidential Academy of National Economy and Public Administration (RANEPA), Prosp. Vernadskogo 82, bld. 5, Moscow, 119571, Russian Federation; e-mail: troussov@gmail.com

Major Fields of Scientific Research: Natural language processing, Information Retrieval, Social and semantic web, social network analysis, graph-based methods