

Construction of Adaptive Educational Forums Based on Intellectual Analysis of Structural and Semantics Features of Messages

Alexander Kozko

Chelyabinsk State University,
Chelyabinsk, Russia
alkozko@gmail.com

Abstract. The paper deals with organization of interaction and communication between subjects of the distance learning process through forums and comment system. Considered existing software tools, their structure and disadvantages. Propose a model of adaptive educational forums, as well as the structural and semantic similarity metrics for extracting dialogues and thematic discussions from arrays of individual comments, as a basis for the construction of adaptive educational forums.

Keywords: distance learning · online discussion forums · semantic similarity

1 Introduction

Development of information technologies makes distance education more available for everyone, and technology of management and support of education process becomes more and more important. Many scientific works dedicated to LMS (Learning Management System), for example works [1, 2, 3], give a description of LMS, and works [4, 5] devoted to review of existing LMS products. However, these works consider of learning management tools mainly from the following perspectives: development, management, courses and distribution of education materials, but not paying enough attention to the issues of communication and information exchange between the different subjects of education process.

At present, modern learning theories such as the theory of connectivism proposed by George Siemens and Stephen Downes, indicate as the basic conditions for successful learning activities not only communication with the teacher, but also the interaction between students and the exchange between them so-called “sensemaking artifacts” [6]. Sensemaking artifacts means blog posts, notes, podcasts, and other educational materials that created by a student for discussion with other students. In the paper [7] analyzed this learning approach and claimed it more effective in comparison with traditional.

Thus, the theme of the organization of communication and interaction between subjects of distance education process is extremely relevant today, but it isn't considered in the existing works.

2 Instruments of communication and informational exchange for education process

Consider the instruments of communication that are used in distance education.

On the one hand LMS, like Moodle, ILIAS, and Sakai and the other hand, of the platform for organization MOOCs (Massive open online course), such as Stepic, Udemy, Coursera, EdX, Udacity propose to use for interaction within the course online forums, blogs, chat rooms, questions and answers websites, wiki-pages and in addition allowed commenting learning materials. For some courses, authors suggest to use third party sites such as Reddit, StackExchange or social networks instead of internal platform tools.

Irrespective of the specifics and type of a course the main tools that are used for communication within the course are standard forums, blogs, questions and answers system, as well as various materials commenting system. However, such instruments do not take into account the specificity of education process and do not provide any applicable analytical functions for organization of education.

3 Overview of informational exchange tools

3.1 Interaction tools

Forum, in generally, - is an online tool for website's visitors communication. The essence of any forum is to create topics with its subsequent discussion. Users can comment on created topics, ask questions and receive answers, and answer other forum user's questions of the forum and give them advice. Thus, each topic is an initial entry with a set of comments.

Blogs are a set of copyright entries, sorted by creation time, usually from the newest to the oldest. Blogs characterized by the ability to publish reviews (comments) by other users, and that makes blogs an instrument for information exchange. Technically, each blog entry as well as forum topic is the initial post with a set of comments to it.

Comment systems additionally used in distance education platforms to allow discuss of any course materials, such as videos or articles.

3.2 Comments structure

Structuring of post's comments may also vary depending on the implementation. There are three main types of structuring:

- Tree comments - comments list is presented as a hierarchical tree view. New message is placed right after the previous one (quoting it isn't necessary). New comment can also start its own discussion branch.
- Linear (flat) comments - comments within the same topic are published under each other, as they become available; new message is placed last (usually at the bottom); interactive relationship between comments is based on specially decorated citations of references to the author and other means.

- Hybrid comments - represent a cross between a tree and a linear structure, now, it is the most popular form of comments submission.

Comments are usually ordered by date, popularity and the number of votes. Each comment has text but it also has the following attributes: Author Name and Timestamp.

4 Adaptive learning forums

As noted earlier, the above instruments of online communication do not take into account the specifics of education process and do not provide any specific functions for it. Furthermore, they are not designed for a large number of participants, which is usual for MOOC. Problems such as a large number of overlapping topics, unanswered questions, incorrectly exposed statuses and tags, are an evidence that the forum is not an effective communication instrument, we lack of effectiveness of the learning process.

To increase the effectiveness of forum usage, as an instrument of communication and information exchange within education process we propose develop the technology of adaptive educational forums, based on data mining.

Adaptive educational forum is an online forum, whose structure is rebuilt depending on student educational trajectory, his information needs, the features of the course and users activity. For teachers, the analysis of information from educational forum can be a source of data for implicit course quality feedback, student's problems with the understanding of educational materials, student's activity evaluation, etc. Thus, the technology of adaptive educational forums could be a way to increase the quality of distance learning.

5 Analysis of comments on online discussion forum

Based on the overview, it can be argued that in such systems, the main content is often not the whole topic and not individual comments, but thematic discussions and dialogues, consisting of comments and united by one common theme. Therefore, the primary tasks are extracting individual discussions of the whole comments list, and the problem of determining the semantic similarity of discussions among themselves. For dialogues extracting of the whole comments list we offer use at both the semantics of messages and their location in the structure of comments.

5.1 Extracting tree commenting relations from linear comments

Because linear structure's main distinction is implicit comments relations, we propose an approach to the reduction of the linear structure to explicit comments tree structure.

For comments linear structure conversion to the tree and find comments relations, proposed focus on the comment text and message attributes, for example, author nickname, post timestamp, user's nickname which are responsible, text of citation, position in comments list.

Using these attributes, and the semantics of messages for analyze, for linear comments conversion into a tree we propose determine the pairwise comments relations. In general, the numerical metric defines an association between two messages comments represented by equation 1.

$$rel(c_1, c_2) = k_{sem} \cdot sim_{sem}(t_1, t_2) + k_{attr} \cdot sim_{attr}(a_1, a_2) \quad (1)$$

$rel(c_1, c_2)$ – relations between the two commentaries, $sim_{sem}(t_1, t_2)$ – semantic similarity of the two comments texts, $sim_{attr}(a_1, a_2)$ – attribute similarity of the two comments, defined by their attributes, k_{sem} и k_{attr} – coefficients.

To extract tree structure, it is necessary to calculate each comment's c1 degree of relation with all previous time comments c2, then c2 comment that has a maximum relation score becomes the parent of the current comment c1 in a tree structure.

5.2 Extraction of thematic discussions

In a discussion with a large number of participants some of the participants might drift away from the main theme and start to discuss unrelated topics. Such thematic discussions (subtopics) within the main topic can also provide useful information for the user, but their detection difficult. Therefore we propose to allocate such subtopics in separate entities and use them in the construction of adaptive educational forums.

In tree-like comment systems, a message has the following attributes: author nickname, post timestamp, comment depth in tree. Based on this, we proposed similarity metric for two comments, represented by equation 2.

$$sim(c_1, c_2) = k_{sem} \cdot sim_{sem}(t_1, t_2) * \left(\frac{1}{dist(c_1, c_2)} + k_{attr} \cdot sim_{attr}(a_1, a_2) \right) \quad (2)$$

$sim(c_1, c_2)$ – similarity between the two commentaries, $sim_{sem}(t_1, t_2)$ – semantic similarity of the two comments texts, $sim_{attr}(a_1, a_2)$ – attribute similarity of the two comments, $dist(c_1, c_2)$ – the distance between the two comments in the tree, k_{sem}, k_{attr} – coefficients.

To unite a group of comments in a thematic discussion we proposed to use an approach based on clustering algorithms, the result of which (clusters of comments) should be used as boundaries for thematic subtopics.

5.3 Semantic similarity metrics for comments

We consider separately the question of calculating the semantic similarity between the two comments texts. First it is worth noting the specifics of messages - basically all posts extremely short - from a few words to 2-3 sentences, so the use of methods for calculating the semantic similarity of documents can be difficult. It is proposed to determine the similarity of the documents, as similarity of containing concepts.

Existing semantic similarity metrics can be divided into several classes. In work [8] it is proposed to the following classification:

1) Measures based on a corpus of texts: for example, LSA, Web-based - NGD and PMIIR.

2) Measures based on ontologies: measures by Wu and Palmer, Leacock and Chodorow, Resnik, Lin and others.

3) Measures based on definitions: ExtendedLesk, GlossVectors.

Another group of metrics that can be allocated -

4) The metrics based on Wikipedia: Wikipedia Link-based Measure (WLM), Explicit Semantic Analysis (ESA), WikiWalk, WikiRelate! and others.

Choosing a semantic similarity metric for concepts it's necessary to pay attention to the specific learning courses. Themes of learning courses may be beyond ontology but in domain that is disclosed in course may contain specific concepts which well-known ontologies as WordNet don't include and specific ontologies for course does not exist. Therefore, a similarity measure metrics based on ontologies can be used to analyze educational forums, only if a teacher will build the ontology of course by himself.

For this reason assumed to use the metrics based on the online encyclopedia Wikipedia. The advantages of using Wikipedia as a source of data are a volume and wide range of different themes, relevance and partial structure.

6 Related Work

Forums analysis is not widely discussed in scientific papers, however it is possible to highlight the following investigation.

Work [9] devoted to the study groups Usenet, it proposes a method for measuring the similarity of the different groups on the activity of participants in them, as well as introduce a measure for evaluation post belonging specific group, which allows exclude cross-posts from the analysis.

In study [10] proposes a model to estimate the probability of involvement or non-involvement of a user in a specific online discussion based on the activity of his friends and his interests, and the list of friends and user's interests based on the previous activity in other topics.

In the work [11] is proposed an approach for extracting context information, as well as questions and answers from the topic using the method of SVM, in [12] considered other methods used for this purpose.

Finally, in [13], the authors propose an approach that combines both structural and semantic analysis for search discussions and find of key messages in the threads.

7 Conclusion

Thanks to the Internet, distance education has been made available for millions of people around the world, and now it's getting increasingly important role of supporting communication within education process. Number of students enrolled to courses is increased, and growth of the distance education percentage in the education process realm sets new requirements that existing communication tools, do not respond, it reduces the efficiency of education process as a whole.

We propose the concept and the model of adaptive educational forums, the use of which, in our opinion, will increase the efficiency of interaction between students, as well as enhance the role of communication environment in distance education process and give teachers the ability for automatically collect information about students as well as the quality of the course.

For creation of adaptive educational forums, we propose to use an approach that includes allocation of individual thematic subtopics that in turn bases on messages structural and semantic features.

References

1. Dalsgaard, C.: Social software: E-learning beyond learning management systems. *European Journal of Open, Distance and E-Learning*, 2 (2006)
2. Sclater, N.: Web 2.0, personal learning environments, and the future of learning management systems. *Research Bulletin*, 13 (2008)
3. Coates, H., James, R., & Baldwin, G.: A critical examination of the effects of learning management systems on university teaching and learning. *Tertiary education and management*, 11, 19-36 (2005)
4. Ketcham G., Landa K., Brown K., Charuk K., DeFranco T., Heise M., McCabe R., Youngs-Maher P.: *Learning Management Systems Review*. (2011) http://openscholar.purchase.edu/sites/default/files/keith_landa/files/doodle_lmsreport_final.pdf
5. Siemens G.: *Learning or Management Systems? A Review of Learning Management System Reviews*. Learning Technologies Centre, University of Manitoba (2006)
6. Siemens G.: *Connectivism: Design and Delivery of Social Networked Learning*. *The International Review of Research in Open and Distance Learning* 2011; 12:3-11. (2011)
7. Mott J., Wiley D.: *Open for Learning: The CMS and the Open Learning Network*. In *education*, 15(2). (2009)
8. Panchenko A.: *Similarity Measures for Semantic Relation Extraction* (Ph.D. Thesis). Université catholique de Louvain & Bauman Moscow State Technical University, (2012-2013)
9. McGlohon M., Hurst M.: *Community Structure and Information Flow in Usenet: Improving Analysis with a Thread Ownership Model*. ICWSM. (2009).
10. Wu, H., Bu, J., Chen, C., Wang, C., Qiu, G., Zhang, L., Shen, J.: *Modeling Dynamic Multi-Topic Discussions in Online Forums*. In *AAAI*. (2009)
11. Cao, Y., Yang, W. Y., Lin, C. Y., Yu, Y.: *A structural support vector method for extracting contexts and answers of questions from online forums*. *Information Processing & Management*, 47(6), 886-898. (2011)
12. Cong, G., Wang, L., Lin, C. Y., Song, Y. I., Sun, Y.: *Finding question-answer pairs from online forums*. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 467-474). ACM. (2008)
13. Lin, C., Yang, J. M., Cai, R., Wang, X. J., Wang, W.: *Simultaneously modeling semantics and structure of threaded discussions: a sparse coding approach and its applications*. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval* (pp. 131-138). ACM. (2009)