

Supporting Online Student Communities by Utilization of Questions and Answers Archives

Adrián HUŇA*

*Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
xhunaa@stuba.sk*

Recently Massive Open Online Courses (MOOCs) attracted a lot of attention as they represent a new educational trend with potential to transform education as we know it nowadays. The leading MOOC providers have millions of users nowadays. The main success behind MOOCs is their availability – anyone can join for free, and online nature – all one needs for participating in a course is a device with Internet connection. Many MOOCs are provided by world's top universities and delivered to thousands of users worldwide. These aspects bring new challenges into educational process, especially in managing the massive amount of learners.

Communication between students and course instructors is usually realised in discussion forums that are part of courses in MOOC platforms. However, due to the large number of users instructors play a secondary role, since they cannot address all students' questions [1]. Other students of the course often assist their partners and keep discussions alive. This model of knowledge exchange among community is on the open World Wide Web best represented in Community Question Answering (CQA) systems. Some well-known CQA systems are Yahoo! Answers, StackOverflow, and Quora.

CQA systems emerged as a way to get personalized answers for questions that are too specific or complex to be understood by search engines. Both discussion forums in MOOCs and CQA systems contain a large repository of knowledge stored in questions and answers archives. In our work we focus on utilization of the archives for automatic question answering. Occurrence of duplicated questions in CQA systems is common. Authors in [2] identified that in certain categories in Yahoo! Answers CQA system, 25% of questions are recurrent. This problem is even more visible in MOOCs as similar questions are naturally repeating during each iteration of a course.

Our main goal is improve question matching model for finding similar questions by exploiting specifics of online student communities. The most important specifics identified are: recurrence of questions in similar time from beginning of courses; natural

* Supervisor: Ivan Srba, Institute of Informatics, Information Systems and Software Engineering

authorities in form of course instructors; and connection between courses' content and content of related questions.

The most important part of our model is, however, finding similar questions on lexical level. We identified multiple models for question retrieval based on lexical similarity in related work: vector space model, language model, translation model, translation-based language model, syntactical tree structures, and LDA model. In our work we plan to employ word2vec model and compare its performance with LDA.

Because our ultimate goal is to automatically answer new question with answers from archive of answers, we must also take answers quality into account. Answer quality can be measured by simple means such as community feedback, but in our work will also exploit user role (e.g. course instructor) to rank their answers as more suitable. The last important component of our matching model is information about posting time since the course started. The complete matching model is shown in Figure 1.

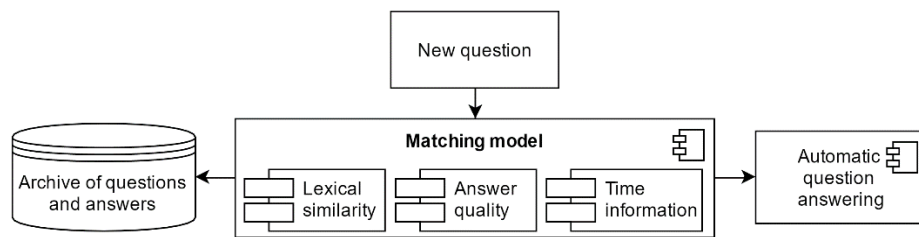


Figure 1. Components of our model for automatic question answering.

There are multiple sources from which latent topics can be learned. It has been argued that text of answers should not be used for learning translation probabilities in translation based models. However, if we take advantage of only quality answers, performance of topic modelling methods is improved [3].

We plan to evaluate our method both offline and online. The online experiment would be performed in MOOCs provided on educational platform edX – one of the leading MOOCs providers.

References

- [1] Alario-Hoyos, C. et al.: Analysing the Impact of Built-In and External Social Tools in a MOOC on Educational Technologies. In: Hernández-Leo, D. et al. (eds.) Proceedings of 8th European Conference, on Technology Enhanced Learning - EC-TEL '13. pp. 5–18 Berlin, Heidelberg, Paphos, Cyprus (2013).
- [2] Shtok, A. et al.: Learning from the past: Answering New Questions with Past Answers. In: Proceedings of the 21st international conference on World Wide Web - WWW '12. pp. 759–768 ACM Press, New York, New York, USA (2012).
- [3] Zhang, K. et al.: Question Retrieval with High Quality Answers in Community Question Answering. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management - CIKM '14. pp. 371–380 ACM Press, New York, New York, USA (2014).