Educational Question Routing in Online Student Communities

Jakub Macina
Slovak University of Technology
Ilkovicova 2
Bratislava 842 16, Slovakia
jakub.macina@gmail.com

Ivan Srba
Slovak University of Technology
Ilkovicova 2
Bratislava 842 16, Slovakia
ivan.srba@stuba.sk

Joseph Jay Williams
National University of Singapore
13 Computing Drive
Singapore 117417, Republic of Singapore
Harvard University
Cambridge, Massachusetts 02138, USA
williams@comp.nus.edu.sg

Maria Bielikova
Slovak University of Technology
Ilkovicova 2
Bratislava 842 16, Slovakia
maria.bielikova@stuba.sk

ABSTRACT

Students' performance in Massive Open Online Courses (MOOCs) is enhanced by high quality discussion forums or recently emerging educational Community Question Answering (CQA) systems. Nevertheless, only a small number of students answer questions asked by their peers. This results in instructor overload, and many unanswered questions. To increase students' participation, we present an approach for recommendation of new questions to students who are likely to provide answers. Existing approaches to such question routing proposed for non-educational CQA systems tend to rely on a few experts, what is not applicable in educational domain where it is important to involve all kinds of students. In tackling this novel *educational question routing* problem, our method (1) goes beyond previous question-answering data as it incorporates additional non-QA data from the course (to improve prediction accuracy and to involve more of the student community) and (2) applies constraints on users' workload (to prevent user overloading). We use an ensemble classifier for predicting students' willingness to answer a question, as well as students' expertise for answering. We conducted an online evaluation of the proposed method using an A/B experiment in our CQA system deployed in edX MOOC. The proposed method outperformed a baseline method (non-educational question routing enhanced with workload restriction) by improving recommendation accuracy, keeping more community members active, and increasing an average number of their contributions.

CCS CONCEPTS

•**Information systems** →**Recommender systems;** *Personalization; Information extraction; Clustering and classification;* •**Applied computing** →*E-learning;*

KEYWORDS

Question routing, Massive Open Online Courses, Community Question Answering, User Models

1 INTRODUCTION

The expansion of Massive Open Online Courses (MOOCs) could provide the potential for everybody with an internet connection to access educational resources, bringing thousands of students from around the world into a diverse online learning community.

However, a key issue in MOOCs is that many students who enroll do not finish the course. Dropout rates for courses can be as high as 94% [14]. While student completion of MOOCs is a complex issue, previous studies [1, 3] have shown that reduced dropout is associated with: (1) involvement in discussion tools (participating students are more likely to successfully finish the course), and (2) question answering success rate (students are less likely to dropout when a high proportion of discussion forum questions are answered). Discussion tools can reduce dropout by enabling students and course instructors to discuss problems with learning materials, gain deeper insight into a topic, and socialize to build a sense of community [16].

MOOC platforms such as edX[1] and Coursera[2] provide basic built-in discussion forums. Some MOOCs use also additional external discussion tools that incorporate elements of social networking sites, online chats or Community Question Answering (CQA) systems (e.g. [1]). In this work, we focus particularly on CQA systems, such as Stack Overflow[3] and Yahoo! Answers[4], which have been successfully used on the open web, in enterprise domains [13] and they are recently starting to be employed in MOOCs[5] as well. CQA systems are an alternative to standard discussion forums that provide more structured content around questions and their answers. In addition, CQA systems offer more possibilities for collaboration (e.g. best answer selection) and are more community-driven (e.g.

---

[1] https://www.edx.org/
[2] https://www.coursera.org/
[3] https://stackoverflow.com/
[4] https://answers.yahoo.com/
[5] E.g. https://cs50.stackexchange.com/

they provide community profiles and possibility to follow activity of other users) [17].

The challenge for MOOC discussion and CQA is that many questions go unanswered, and this proportion of unanswered questions can be as high as 50% [24]. As thousands of students are enrolled, course instructors are overloaded with too many students to serve. Ideally, students themselves could be a resource for answering each other's questions. However, only a small fraction of the student community actually answer questions [1]. Another part of the community (so called lurkers) consumes the content but does not actively contribute in discussion. This reflects the long-tail distribution commonly found in online communities, where a majority of content is created by a minority of users. Two reasons for low participation are (1) many users are not willing to participate, and (2) users who are willing to participate are overloaded with too many questions, and are not able to find the questions of greatest interest for them [4].

We tackle this problem of engaging more students in answering questions in online courses. Our approach is to recommend new questions to those students who are suitable candidates to answer them (so called question routing). To the best of our knowledge, this is the first work investigating question routing (as a task well-known from CQA systems on the open web) in the context of MOOCs. Our main contributions are as follows:

(1) A novel approach to question routing, which addresses the restrictions and novel opportunities of the educational context of MOOCs. While existing methods primarily focus on askers' needs, we take an answerer-oriented approach by considering not only students' expertise, but also willingness to answer a question.

(2) A model of student activity (beyond data about questions and answers) that enables us to involve more students in question answering and reduce the burden on individual users.

(3) An experiment aimed at online evaluation of our question routing method with more than 4600 MOOC students.

This paper is organized as follows. Section 2 reviews state-of-the-art approaches to question routing, and their limitations in educational domains. Section 3 introduces our formulation of the problem of educational question routing. Section 4 describes our method for question routing. Section 5 presents on experimental evaluation of our method in a MOOC.

2 RELATED WORK

CQA systems on the open web raise a wide range of problems in adaptively supporting collaboration among users, for which many recommendation-based solutions have been developed [17]. In this work, we focus on personalized question recommendation. The problem of question recommendation can be seen from two perspectives [23]: (1) providing users with questions which can be beneficial for them, or (2) routing unanswered questions to users who are likely to answer them.

In the first problem of *question recommendation* [21, 24], the task is given a user u to find ordered list of most relevant questions $q_1, ..., q_n$ to the user u. It can be characterized by analogy to product recommendation, where items are questions. Content-based,

collaborative filtering or hybrid approaches are applicable here [8]. Question recommendation recommends any type of questions, mostly resolved ones, to all kinds of users. It can be generated on demand or regularly, for example, as a part of a newsletter.

In the second problem of *question routing* [6, 27], given a newly posted question q, the task is to find an ordered list of users $u_1, ..., u_n$ who are most suitable to answer question q. Question routing uses mainly content-based recommendation, since a question only needs to be recommended to users until it is successfully answered. In contrast to question recommendation, question routing recommends *new* questions, that have not yet been answered, and must be achieved with the limited data available after question posting. Question retrieval, i.e. reusing content by automatic identification of semantically similar questions (e.g. [26]), can be used as a supplement to question routing to filter out repeating questions directly at question creation time.

Our work focuses on question routing, because of the potential to support collaboration among students, to involve more of the community in question answering and to increase the amount of students' contributions. The process of question routing in CQA systems can be understood in three phases [7]: (1) construction of a question profile, (2) construction of a user profile, and (3) matching the profiles of questions and users.

The purpose of the question profile is to represent the topic of the question. Models for representing a question include bag-of-words models based on tf-idf, topic models [15, 22] such as Latent Dirichlet Allocation (LDA), and combinations of both [21]. User profiles that capture a user's expertise in a topic can be inferred from the question-answer data from users' previous contributions in CQA system [15, 21]. Previous work has also modeled users' activity [12, 22], when a user is likely to respond [5], and a user's readiness to answer [13]. To match the profiles of questions and users, three main approaches have been used. The first is classification models that predict a match using methods like logistic regression [13] or Support Vector Machines (SVM) [27]. The second is collaborative filtering. For example, Dror [6] proposed a multi-channel hybrid recommender system that combined collaborative filtering and classification. The third is ranking models that draw on work from document retrieval [12, 15, 22].

Current state-of-the-art question routing methods applied in CQA systems on the open web can be characterized as asker-oriented (i.e. primarily focusing on askers' goals). They route new questions only to the best possible answerers (top experts) to satisfy the askers' needs best [6, 12, 15, 22]. In other words, they may overload a small set of experts while the full potential of the community is not utilized (e.g. one third of all answers on Yahoo! Answers are created by junior users not top experts [21]). This asker-oriented attitude may have several significant drawbacks [19]. From sustainability point of view, it is necessary to maintain healthy community ecosystem and utilize the potential of the online community more effectively by involving and satisfying needs of all potential answerers, including novices, lurkers and junior users. This is also especially important in educational domain.

Research work by Luo et al. [13] addressed this sustainability problem by question routing in enterprise CQA system. The uniqueness of their work is in systematically involving inactive users by routing new questions to them. A classification model is used to

find best potential answerers. Afterwards, inactive users most similar to the best potential answerers are matched by distance learning utilizing non-QA data from company's internal systems, e.g. current work state. Similarly in our previous work [20], we employed publicly available non-QA data (e.g. users' homepages or profiles on social networking sites) to obtain expertise estimation also for less active users.

To the best of our knowledge, the task of question routing has not been addressed in educational CQA systems yet, although there has been related work on educational CQA more generally. The first question recommendation method designed specifically for MOOCs was proposed by Yang et al. [24]. The main contribution is optimization of predictions given two constraints present in MOOC domain which are limited students' time for participation and adequate knowledge of an answerer to answer a question. It used a context-aware matrix factorization model that exploited three kinds of contextual information – student features, question features and implicit feedback. This approach regularly (e.g. once a week) recommended a list of answered and unanswered questions to all users. This kind of question recommendation may improve question answering success rate, but it does not route questions to individual users. A significant delay is introduced, as newly created questions cannot be immediately routed to suitable answerers. Some progress has been made also in automatically identifying struggling students based on their previous activity [10], gamification in educational CQA system [2] and in identifying low-quality answers that can mislead students [9].

3 INTRODUCTION OF EDUCATIONAL QUESTION ROUTING

Standard question routing solutions are not appropriate for educational domain. At first, educational question answering has a slightly different aim. While the traditional CQA systems stress the importance of asking challenging, unique questions and providing top quality fast answers, in MOOCs, we primarily aim at students' learning. And thus we may accept also more trivial questions and not so perfect and fast answers while students benefit from question answering. Secondly, we need to consider some constraints, students are usually learning about the topic for the first time and in addition they might be in various weeks of the course, thus having different levels of knowledge. At the same time, educational domain brings many additional opportunities in comparison with standard CQA systems on the open web, such as additional information about students (e.g. students' results in the course) and about questions (e.g. interconnection with learning materials).

3.1 Problem characterization

In this paper we introduce a new task of *educational question routing*, which refers to routing of new questions without any answer in educational CQA system to suitable answerers. The goal of educational question routing is to: (1) *decrease information load of users* by accurate recommendations, (2) *engage a greater part of the community in question answering*, and (3) *increase an average number of contributions*. This way educational question routing will also contribute to lower instructor workload and lower dropout rate.

In educational question routing, we should take three factors (constraints) into account: (1) answerers should have appropriate knowledge for answering a question, (2) answerers should have willingness to answer a question, and (3) answerers should not be overloaded with routed questions. The rationale for considering students' adequate expertise is to involve the majority of students (not only top experts) and at the same time recommend them questions with a reasonable difficulty (i.e. not too challenging questions so the necessary quality of the provided answers will be achieved). Some students might have a suitable expertise to answer a question but not all of them are also motivated to actually provide a contribution. Therefore, willingness to answer is also likely to improve the recommendation. And finally, it is necessary to balance recommended questions according to current students' workload, as answering others questions is not students' primary goal.

In educational question routing, we need to tackle two cold-start problems related to questions and users. Firstly, we want to involve students who do not have previous contributions in the CQA system. Thanks to additional possibilities of MOOCs, we are not limited in user modelling (i.e. estimation of expertise and willingness) only at users' previous activities directly in the CQA system. We can exploit several valuable sources of data in the MOOC course. For example, it is possible to estimate student's expertise more accurately by considering knowledge prerequisites for a routed question, such as a previous view of related lecture videos or a sufficient grade for related assignments.

Secondly, only limited data are available about questions at their creation time and we cannot rely on community feedback, such as voting, favorites or comments. However, utilizing information about the asker can help to estimate quantities like question difficulty. We can utilize the so called knowledge gap between an asker and a potential answerer. The knowledge gap denotes a pattern, which was observed by Lin et al. [11], in which the question asked by an expert has a high probability of being difficult and thus can be answered only by other expert users. The same pattern was also observed for users with low expertise and easy questions.

Finally, the method for educational question routing should meet requirements of fast-paced learning environments. It should be designed to route new questions in real-time after question creation. In addition, it should reflect recent activity in the CQA system and MOOC course throughout the period of a course (e.g. when students drop out of the course, the method should not consider them as suitable answerers any more).

3.2 Definitions

Definition 3.1. Categories in MOOCs are organized into a hierarchical structure, where at the first level a course consists of categories representing weeks and at the second level each week consists of categories dedicated to particular topics. This hierarchical structure is usually followed also in discussion tools. Each question is assigned to a particular topic category. If we define Q as the set of all questions in a course, then $Q_{topic} \subseteq Q_{week} \subseteq Q$.

Definition 3.2. Knowledge $K_{u,c}$ is the topical expertise of a user u for answering questions in a category c. In this paper, user knowledge $K_{u,c}$ is estimated as the sum of the total number of contributions in a category c, i.e. answers, comments and votes.
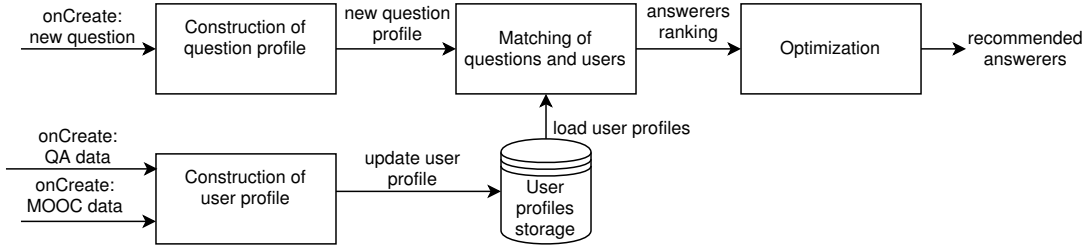
Figure 1: Educational question routing framework.

Definition 3.3. Knowledge gap is the difference in the knowledge of an answerer $K_{answerer,c}$ and knowledge of an asker $K_{asker,c}$ in a category c of a routed question.

Definition 3.4. Workload L_u is the number of questions routed to user u within the recent time period (e.g. during the last 7 days).

4 EDUCATIONAL QUESTION ROUTING FRAMEWORK

An overview of our educational question routing framework is depicted in Figure 1. The inputs for the framework are a new question and users' activities in the CQA system and MOOC. The output for a new question is a list of recommended answerers, sorted by the ranking $r_{u,q}$ of how likely user u is to answer a new question q. The framework is divided into three phases, which are typical for question routing methods, followed by the fourth optimization phase, which is added specifically for educational question routing.

4.1 Construction of question profile

The question profile is created immediately after a new question is posted into the CQA system. It captures the question's content (title and question body), as well as metadata, such as the hierarchy of assigned categories, and information about the asker, who is represented by a user profile (see Section 4.2). Question title and body are concatenated and preprocessed by tokenization, removal of stop words, and stemming. The question text profile θ_q is created as a bag-of-words model with tf-idf weights. The answer profile $\theta_{a,q}$ is created similarly, but since the answer does not have a title the concatenation step is unnecessary. Information about the asker and the categories for week and topic categories are also stored for use in the third question-user matching step.

4.2 Construction of user profile

The user profile is updated after each user activity in the CQA or the MOOC. It consists of two main parts that capture information about the (1) topic of questions a user previously provided answers on, (2) quantity, quality and time distribution of previous user activities carried out in the CQA system or in the MOOC.

Firstly, the user profile models user's interests in particular question topics. A user text profile θ_u is created as a sum of the bag-of-words vectors from question and answer text profiles:

$$\theta_u = \sum_{q \in Q_u} (\theta_q + \theta_{a,q}) \tag{1}$$

where Q_u is the set of questions to which the user provided an answer, θ_q is a question text profile and $\theta_{a,q}$ is an answer text profile of user u for question q.

Secondly, the user profile captures the current level of user activity in the CQA system itself (i.e. number of posted questions, answers, and comments) as well as activity in the MOOC (i.e. number of seen lectures). In addition to level of activity, the user profile captures quality of activity, through earned votes in the CQA system and assignment grades. Quantity and quality of all activities is stored separately for each week and each topic category. As user's activity in the system can vary significantly over time, the user profile also contains time related metrics, such as course registration date and time of last answer.

4.3 Matching of questions and users

The third phase, matching of the question and the user profile, aims to calculate ranking of each user given a new question. Using the profile of a new question and each potential answerer's user profile, we address this as an ensemble of two explicit classification tasks:

(1) Predicting whether a user has sufficient expertise to answer a new question.
(2) Predicting user willingness to answer a new question.

The rationale for splitting the classification into two subtasks is to provide more powerful and flexible global classifier. Moreover, it ensures that both sources of information will be appropriately weighted and used in a global classifier. For example, our approach ensures that expertise features are not the only consideration in question routing at the expense of features that provide a cue to willingness.

The final ensemble ranking combines the predictions of individual classifiers, as the probability of both events occurring – a user having the expertise as well as willingness to answer a new question:

$$P(y = 1) = P(exp = 1) \cdot P(will = 1) \tag{2}$$

where $P(exp = 1)$ is the probability of the positive class for expertise classifier prediction, and $P(will = 1)$ is the probability of the positive class for the willingness classifier. This final probability of each prediction is assigned as the value of $r_{u,q}$ for each user and new question, and used to rank predicted answerers for question routing.

Input features are divided into two groups for each classifier separately: expertise and willingness features (see Table 1). The features dedicated to expertise classification include:

Table 1: Expertise and willingness features, divided into subgroups according to their origin in the CQA system or the MOOC; and according to their type as non-educational (used also by CQA systems on the open web) and educational.

		Non-educational	Educational
Expertise	**CQA**	question-user text profiles similarity answers count within a week category answers count within a topic category earned votes count within a week category earned votes count within a topic category	overall knowledge gap knowledge gap within a week category knowledge gap within a topic category
	MOOC		average assignments grade portion of seen lectures within a week category portion of seen lectures within a topic category
Willingness	**CQA**	overall answers/comments/questions count overall earned votes count answers count in the recent period last answer time proportion of days with any activity in CQA	portion of seen questions within a week category portion of seen questions within a topic category
	MOOC	course registration date	portion of days with any activity in MOOC lecture freshness portion of seen lectures within a week category portion of seen lectures within a topic category

(1) simple expertise measures calculated for the question's week/topic category, such as *answers counts* and *earned votes counts*,

(2) *question-user text profiles similarity*, which is calculated as a cosine similarity between the question text profile θ_q and the user text profile θ_u,

(3) *knowledge gap* for the question's week/topic category, which models the level of difficulty for an answerer to answer a new question,

(4) knowledge prerequisites, such as *portion of seen lectures* and student's *average assignments grade* (the rationale is that students who have already seen lectures or have good grades for a question's topic are more likely to have the suitable expertise).

The features for classifying willingness to answer a question include:

(1) simple activity measures, such as *overall answers, questions or comments count*,

(2) *answers count in recent period* which determine the amount of latest activity,

(3) time related metrics, *last answer time t_u*, *lecture freshness t_l* (i.e. time of watching the lecture related to a new question) and *course registration date t_r* (which influences a student commitment as shown by [25]), are processed in relation to time of posting of the new question t_q:

$$t_q - t, \quad t \in \{t_u, t_l, t_r\} \tag{3}$$

(4) *portion of days with any activity in CQA/MOOC* is used to estimate a response time on recommendation,

(5) *portion of seen questions* within the question's week/topic category, which represents student interest in browsing questions in the CQA system,

(6) *portion of seen lectures* within the question's week/topic category, which estimates student activity in the course.

4.4 Optimization

In the last phase of our question routing framework, the constraints are applied similarly as in [13, 24]. Routed questions are balanced by applying a threshold on user's current workload L_u. If L_u is above a threshold, question is not routed to a user u. The goal of the optimization is to not overload users and to involve a greater part of the community in question answering.

5 EXPERIMENTS

This section presents the results of an offline and an online experiment. The goal of both experiments was to evaluate the performance of our educational question routing method. The online experiment enabled us to examine the real-world impact of our question routing method on the student community.

Both experiments were conducted in our open-source CQA system Askalot ported to the edX MOOC platform [18]. We used data from the first half of the course to fine tune parameters of our question routing framework. Then, we used the trained classification models during the online experiment, which was conducted on the second half of the course.

5.1 CQA system and course description

The MOOC course used for our experiments was *QuCryptox Quantum Cryptography*[6] offered by the California Institute of Technology and Delft University of Technology[7]. The course taught basics of quantum cryptography and required advanced knowledge of linear algebra and probability. It lasted 10 weeks from 10th October to 20th December 2016. As the course and the CQA system was available for students also before and after the official course duration, the presented results cover a longer period including two weeks before (26th September 2016) and after (2nd January 2017) course.

---

[6]https://courses.edx.org/courses/course-v1:CaltechDelftX+QuCryptox+3T2016/
[7]Permission to conduct this research was provided by the course instructors.

Table 2: Statistics on QuCryptox Quantum Cryptography course.

Metric	Quantity
Students enrolled in the course	8115
Students started the course	4618
Users participating in CQA (contributors + lurkers)	1098 (24%)
Users contributing in CQA	377 (8%)
Questions	361
Questions with answer	299 (83%)
Questions with best answer selected	53 (15%)
Answers	386
Comments	476

The course had estimated workload from 6 to 8 hours per week. Each week contained video lectures about several topics followed by quizzes, pen and paper assignments, and coding assignments. Summary statistics about the course are presented in Table 2. The proportion of answered questions was very high during the whole course (84.28% before and 80.30% during the online experiment).

5.2 Baseline

As there is not any other educational question routing method, which we can directly compare with, we used a non-educational variant of our question routing framework as a baseline. It does not consider the educational features, which are listed in Table 1. As we could not afford to overload students in the online experiment by many requests to answer new questions, we applied a workload restriction also on this baseline recommendation. Despite this workload restriction, which is not typical for existing state-of-the-art question routing methods, our selected baseline can be characterized as asker-oriented which is very common in the CQA systems on the open web.

5.3 Offline experiment

Research on methods for question routing are evaluated mostly by offline experiments, while online experiments are conducted very rarely (see [17] for a review). However, offline experiments are limited in several ways. They can only consider users who choose to answer a question as a positive example. However, a user who is a suitable answerer (a positive example) may simply not choose to answer a question if a high-quality answer was already provided by another user. Also, as most available data sets do not provide question views, it is not possible to identify users who see an unanswered question and do not provide any answer (a negative example). Past research therefore considers all users as negative examples, except those who choose to answer a question. Our work addresses these limitations by supplementing the offline evaluation with an ecologically valid online evaluation of our method.

5.3.1 Experimental setup. For training expertise and willingness classifiers, we generated positive ($y = 1$) and negative ($y = 0$) samples from data as defined in Table 3. The ensemble classifier was used to compute whether a question would be routed to a user, and was evaluated by comparision to which users actually

Table 3: Generation of positive and negative samples for expertise classifier and willingness classifier.

	Positive class (class 1)	Negative class (class 0)
Expertise classifier	- answer with positive votes difference - answer which is marked as best answer - answer with positive evaluation from teaching assistant	- answer with negative votes difference - answer with zero votes and another answer was added later - answer with negative evaluation from teaching assistant
Willingness classifier	- answer - comment	- question view without interaction

answered the question. The optimization step for load balancing was not applied in the offline experiment.

We performed feature selection by χ^2 selection and filtering out highly correlated features (these features are already removed from Table 1). When selecting appropriate classification algorithms, we required online training, or fast re-training on all previous activities. Particularly, we selected and trained three classifiers - SVM, Random Forest and Logistic Regression. The probability threshold for predicting the positive/negative class of each classifier is found dynamically by maximizing the AUC metric (Area Under the Receiver Operating Curve). We used grid search to find the best combination of parameters and tune the model hyper parameters. We used 10-fold stratified cross validation for evaluating both classification algorithms. Our experiment found that the Random Forest method performed the best on both classification tasks, and so we selected it for both classifiers.

5.3.2 Metrics. We used the following metrics for evaluating our question routing methods for ranking answerers: *Success (S@N)*, *Mean Average Precision (MAP@N)*, *Normalized Discounted Cumulated Gain (nDCG@N)* and *Mean Reciprocal Rank (MRR)*. S@N measures whether any real answerer is among the top *N* users, while the other metrics are standard in information retrieval.

5.3.3 Results. The results are computed for the whole duration of the course. As shown in Table 4, our educational question routing framework outperformed the baseline approach across all of metrics. If we would recommend new questions to the top 10 most suitable answerers, we would hit at least one actual answerer in 60.1% of all cases, in comparison to the 54.8% achieved by the baseline (see Figure 2).

We explored the feature importance values from the Random Forest classifier. For prediction of user expertise in answering a question, the most important features were similarity of the question and user text profiles, overall knowledge gap, and count of answers. For prediction of user willingness to answer a question, the most important features were the portion of seen questions within a topic category, and the recent period's answers count. Individual classifier AUC performance is $0.67(+/-0.06)$ for the expertise classifier and $0.73(+/-0.06)$ for the willingness classifier.

Table 4: Educational question routing and baseline question routing results for the offline experiment.

Metric	Educational			Baseline		
	N=5	N=10	N=100	N=5	N=10	N=100
S@N	0.418	0.601	0.924	0.369	0.548	0.894
MAP@N	0.221	0.242	0.260	0.193	0.216	0.234
nDCG@N	0.288	0.345	0.413	0.254	0.312	0.384
MRR	0.284			0.259		
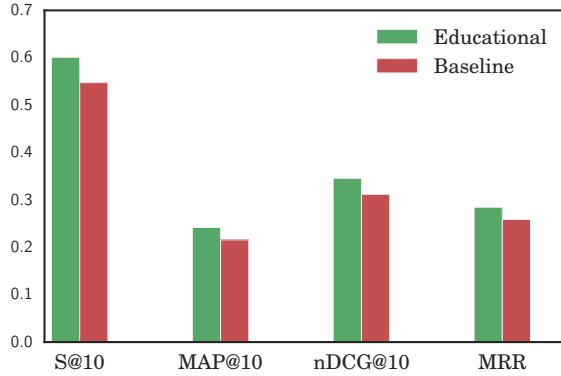
Figure 2: Question routing offline performance results compared to baseline for selected metrics with N=10.

5.4 Online experiment

5.4.1 Experimental setup. We conducted the online experiment as an A/B test in the second half of the QuCryptox Quantum cryptography course, starting from 14th November 2016 (week 6). At the beginning of week 6, the users registered in the CQA system were randomized into three groups of *n* users, as follows:

(1) *Educational group* (n = 1306). Potential answerers had questions recommended by our educational question routing framework.
(2) *Baseline group* (n = 1306). Users had questions routed by the baseline method.
(3) *Control group* (n = 1306). Users in the control group did not receive any question recommendations.

Randomized assignment was stratified by users' answer counts, to reduce variability. Students who signed up for the course during the online experiment were not included.

Each newly asked question was routed or recommended to the top 10 most suitable users in the educational group and to the top 10 users in the baseline group, using the group's respective question routing method. The constraint on workload L_u for user u was derived from the amount of natural activity of the most active student contributors before the online experiment, and it was set to a maximum of 4 recommendations per 7 days. The recommendation was sent to users as a notification in the CQA system, as well as appearing in the user's dashboard as a personalized list of new recommended questions.

Table 5: Accuracy of question recommendation.

	Educational	Baseline
CTR	23.25%	18.29%
Question routing success rate	15.91%	10.61%

User profiles were updated in real-time and the data loaded for matching questions and users (this ensured users' recent activity was immediately reflected in the question routing, reducing the cold-start problem). Both expertise and willingness classifiers were re-trained each day with the latest data.

5.4.2 Metrics. We evaluated the different conditions (question routing methods) on the following metrics:

- *Click-through rate (CTR)* - The proportion of clicked question recommendations (out of the total number of question recommendations).
- *Question routing success rate* - The proportion of routed questions for which one or more proposed answerer contributed.
- *Users' coverage rate* - The proportion of unique users involved in question routing, because they were selected as a suitable answerer and received a question recommendation (out of the total number of users in the group).
- *Time to answer* - Time in hours for a newly posted question to receive its first answer from a recommended answerer.
- *Answer quality* - This was computed as the difference between the positive and negatives votes by users.

5.4.3 Results. During the online experiment, 132 new questions were routed to potential answerers, resulting into 2640 recommendations. We present results for evaluating the educational question routing goals outlined in Section 3.1.

First, educational question routing needs to decrease the burden on answerers by generating relevant recommendations. One metric for relevant recommendation is a high CTR (see Table 5). The CTR was 23.25% for users in the educational group, but only 18.29% for the baseline group, which was significantly different, $\chi^2(1, N = 2640) = 10.03, p < 0.01$. This difference in CTR of 4.96% means that our method for educational question routing increased the chances of users clicking through by 27%.

A second metric for recommendation quality is the proportion of recommended questions that were answered, or the question routing success rate. The question routing success rate was 15.91% for the educational group, but only 10.61% for the baseline group. While this difference did not reach statistical significance, $\chi^2(1, N = 264) = 1.61, p = 0.20$, this is likely a reflection of the small sample size providing insufficient statistical power to detect differences, as the difference of 5.30% represents an increase of 50% beyond the baseline group. The improvements in recommendation accuracy were achieved by incorporating MOOC data that improved predictions of expertise and willingness to answer questions.

With respect to involving more of the community in educational question routing, we did not observe substantive differences in users' coverage rate, or the proportion of users who answered a recommended question. This rate was 10.72% for the educational group and 10.03% for the baseline group.
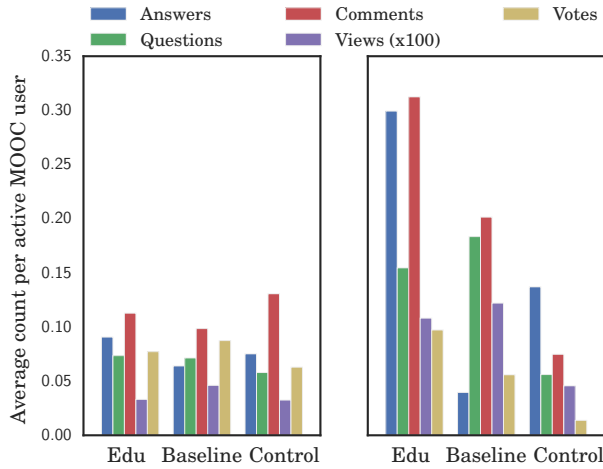
Figure 3: Average amount of activity per active MOOC user before (left) and during (right) the online experiment.

Table 6: Number of contributing CQA users / number of active MOOC users, before and during the online experiment.

	Educational	Baseline	Control
Before	62 / 816 (7.60%)	73 / 812 (8.99%)	74 / 811 (9.12%)
During	40 / 304 (13.16%)	26 / 278 (9.35%)	28 / 321 (8.72%)

In addition to the direct impact of question routing, introducing these methods can also increase overall activity in the CQA system. We report data for each of the three groups, both before and during the online experiment, although the small sample sizes precluded statistical tests. Figure 3 shows the usage behaviors in the CQA system, in terms of posting questions, answers, comments, and viewing and voting on questions. The average count of each behavior was obtained using a similar methodology to [21], where we normalized the absolute numbers of each behavior by the number of active MOOC users in each group (a user is considered as active if he/she had at least one course interaction during the first or second half of the course). The introduction of question routing appears to have led to greater activity in the use of the CQA system, in both the educational and baseline group, relative to the control.

In the Quantum Cryptography course, instructors spent a significant amount of time by collaboration and discussions with students. This high instructors involvement is reflected in the proportion of contributions posted by instructors or teaching assistants, which was 37.28% before the online experiment and only 31.25% during the online experiment. This significant drop confirms that instructors' workload decreased during the online experiment.

To further analyze the impact of question routing and lower instructor load, we verified that it did not have a negative influence on quality of answers and time to answer them. As we stated in Section 3.1, these metrics are not the main subject of optimization (as in the standard question routing methods). However, we do not want these metrics to be reduced at a cost of involving more of the student community or decreasing instructors' workload. The results

suggest that the question routing preserves the time to answer and answer quality.

Finally, we evaluated the dropout rate in each user group. We found a minor difference in overall dropout rate from the MOOC (76.72% for educational group vs. 78.71% for baseline). At the same time, we identified a difference in dropout of contributors from the CQA system itself. While in the control group, 21.86% of users stopped contributing to CQA system during the online experiment, it was only 15.61% in the baseline group and even 12.50% in the educational group. This result confirms a positive influence of question routing on keeping users motivated and devoted to question answering. It is also reflected in the proportion of contributing users to CQA system out of active MOOC users (see Table 6), which increased for educational group while it remained rather stable in baseline and control group.

6 CONCLUSIONS

In this paper, we described a new question routing approach for MOOCs. We proposed two innovations which makes our question routing method suitable for an educational domain. Firstly, our approach explicitly models user's willingness to answer the question and combines it with expertise of a user. Secondly, we utilized non-QA data from online course for question routing such as students' grades, activity in the MOOC course and knowledge prerequisites to successfully answer a new question.

There are many additional specifics of MOOCs that provide a possibility for further improvements. Currently, we tackle all types of questions equally. However, in MOOCs some kinds of questions (e.g. those about organizational matters) can be answered only by instructors. If we will be able to automatically and reliable identify them, we can limit the potential answerers to instructors only. Another promising future direction is a better approximation of students knowledge for computing knowledge gap. Finally, applying question routing in MOOC courses with thousands of students brings scalability issues which need to be addressed.

The proposed question routing approach was evaluated by the offline experiment (to fine tune the models and evaluate accuracy of recommendation) and the online experiment (to measure a total impact on the online student community). Based on the comparison with the baseline, it can be concluded that the proposed framework achieved a higher interest of users in the routed questions and engaged more students, who in addition provided more contributions. Moreover, we found out that the recommendations also contributed to the lower dropout rate of active contributors in CQA and to the lower instructors workload.

REFERENCES

[1] Carlos Alario-Hoyos, Mar Pérez-Sanagustín, Carlos Delgado-Kloos, Hugo A. Parada G, and Mario Munoz-Organero. 2014. Delving into Participants' Profiles and Use of Social Tools in MOOCs. *IEEE Transactions on Learning Technologies* 7, 3 (2014), 260–266.

[2] Chulakorn Aritajati and N. Hari Narayanan. 2013. Facilitating Students' Collaboration and Learning in a Question and Answer System. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion (CSCW '13)*. ACM, New York, NY, USA, 101–106. DOI:http://dx.doi.org/10.1145/2441955.2441983

[3] Lori Breslow, David E. Pritchard, Jennifer DeBoer, Glenda S. Stump, Andrew D. Ho, and Daniel T. Seaton. 2013. Studying Learning in the Worldwide Classroom: Research into edX's First MOOC. *Research & Practice in Assessment* 8 (2013), 13–25.

[4] C. G. Brinton, M. Chiang, S. Jain, H. Lam, Z. Liu, and F. M. F. Wong. 2014. Learning about Social Learning in MOOCs: From Statistical Analysis to Generative Model. *IEEE Transactions on Learning Technologies* 7, 4 (Oct 2014), 346–359. DOI:http://dx.doi.org/10.1109/TLT.2014.2337900

[5] Tianjian Chen, Jing Cai, Hao Wang, and Yu Dong. 2014. Instant Expert Hunting: Building an Answerer Recommender System for a Large Scale Q&A Website. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing (SAC '14)*. ACM, New York, NY, USA, 260–265. DOI:http://dx.doi.org/10.1145/2554850.2554915

[6] Gideon Dror, Yehuda Koren, Yoelle Maarek, and Idan Szpektor. 2011. I Want to Answer; Who Has a Question?: Yahoo! Answers Recommender System. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '11)*. ACM, New York, NY, USA, 1109–1117. DOI:http://dx.doi.org/10.1145/2020408.2020582

[7] Jinwen Guo, Shengliang Xu, Shenghua Bao, and Yong Yu. 2008. Tapping on the Potential of Q&A Community by Recommending Answer Providers. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM '08)*. ACM, New York, NY, USA, 921–930. DOI:http://dx.doi.org/10.1145/1458082.1458204

[8] Yutaka Kabutoya, Tomoharu Iwata, Hisako Shiohara, and Ko Fujimura. 2010. Effective Question Recommendation Based on Multiple Features for Question Answering Communities. In *Proceedings of Fourth International AAAI Conference on Weblogs and Social Media*. AAAI, 259–262.

[9] Long T. Le, Chirag Shah, and Erik Choi. 2016. Evaluating the Quality of Educational Answers in Community Question-Answering. In *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries (JCDL '16)*. ACM, New York, NY, USA, 129–138. DOI:http://dx.doi.org/10.1145/2910896.2910900

[10] Long T. Le, Chirag Shah, and Erik Choi. 2017. Bad Users or Bad Content?: Breaking the Vicious Cycle by Finding Struggling Students in Community Question-Answering. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval (CHIIR '17)*. ACM, New York, NY, USA, 165–174. DOI:http://dx.doi.org/10.1145/3020165.3020181

[11] Chih-Lu Lin, Ying-Liang Chen, and Hung-Yu Kao. 2014. Question difficulty evaluation by knowledge gap analysis in Question Answer communities. In *Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '14)*. IEEE, 336–339.

[12] Mingrong Liu, Yicen Liu, and Qing Yang. 2010. Predicting Best Answerers for New Questions in Community Question Answering. In *Proceedings of the 11th International Conference on Web-age Information Management (WAIM'10)*. Springer, Berlin, Heidelberg, 127–138.

[13] Lin Luo, Fei Wang, Michelle X. Zhou, Yingxin Pan, and Hang Chen. 2014. Who Have Got Answers?: Growing the Pool of Answerers in a Smart Enterprise Social QA System. In *Proceedings of the 19th International Conference on Intelligent User Interfaces (IUI '14)*. ACM, New York, NY, USA, 7–16. DOI:http://dx.doi.org/10.1145/2557500.2557531

[14] Daniel Onah, Jane Sinclair, and Russell Boyatt. 2014. Dropout Rates of Massive Open Online Courses: Behavioural Patterns. *Proceedings of the 6th International Conference on Education and New Learning Technologies (EDULEARN '14)* (2014), 5825–5834.

[15] Fatemeh Riahi, Zainab Zolaktaf, Mahdi Shafiei, and Evangelos Milios. 2012. Finding Expert Users in Community Question Answering. In *Proceedings of the 21st International Conference on World Wide Web (WWW '12 Companion)*. ACM, New York, NY, USA, 791–798. DOI:http://dx.doi.org/10.1145/2187980.2188202

[16] Carolyn Penstein Rosé, Ryan Carlson, Diyi Yang, Miaomiao Wen, Lauren Resnick, Pam Goldman, and Jennifer Sherer. 2014. Social Factors That Contribute to Attrition in MOOCs. In *Proceedings of the First ACM Conference on Learning @ Scale Conference (L@S '14)*. ACM, New York, NY, USA, 197–198. DOI:http://dx.doi.org/10.1145/2556325.2567879

[17] Ivan Srba and Maria Bielikova. 2016. A Comprehensive Survey and Classification of Approaches for Community Question Answering. *ACM Trans. Web* 10, 3, Article 18 (Aug. 2016), 63 pages. DOI:http://dx.doi.org/10.1145/2934687

[18] Ivan Srba and Maria Bielikova. 2016. Design of CQA Systems for Flexible and Scalable Deployment and Evaluation. In *Proceedings of 16th International Conference on Web Engineering (ICWE '16), Lugano, Switzerland, June 6-9, 2016*. Springer, Berlin, Heidelberg, 439–447. DOI:http://dx.doi.org/10.1007/978-3-319-38791-8_30

[19] Ivan Srba and Maria Bielikova. 2016. Why is Stack Overflow Failing? Preserving Sustainability in Community Question Answering. *IEEE Software* 33, 4 (July 2016), 80–89. DOI:http://dx.doi.org/10.1109/MS.2016.34

[20] Ivan Srba, Marek Grznar, and Maria Bielikova. 2015. Utilizing Non-QA Data to Improve Questions Routing for Users with Low QA Activity in CQA. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 (ASONAM '15)*. ACM, New York, NY, USA, 129–136. DOI:http://dx.doi.org/10.1145/2808797.2809331

[21] Idan Szpektor, Yoelle Maarek, and Dan Pelleg. 2013. When Relevance is Not Enough: Promoting Diversity and Freshness in Personalized Question Recommendation. In *Proceedings of the 22nd International Conference on World Wide Web (WWW '13)*. ACM, New York, NY, USA, 1249–1260. DOI:http://dx.doi.org/10.1145/2488388.2488497

[22] Yuan Tian, Pavneet Singh Kochhar, Ee-Peng Lim, Feida Zhu, and David Lo. 2014. Predicting Best Answerers for New Questions: An Approach Leveraging Topic Modeling and Collaborative Voting. In *Social Informatics: SocInfo 2013 International Workshops, QMC and HISTOINFORMATICS, Kyoto, Japan, November 25, 2013, Revised Selected Papers*. Springer, Berlin, Heidelberg, 55–68. DOI:http://dx.doi.org/10.1007/978-3-642-55285-4_5

[23] Fei Xu, Zongcheng Ji, and Bin Wang. 2012. Dual Role Model for Question Recommendation in Community Question Answering. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '12)*. ACM, New York, NY, USA, 771–780. DOI:http://dx.doi.org/10.1145/2348283.2348387

[24] Diyi Yang, David Adamson, and Carolyn Penstein Rosé. 2014. Question Recommendation with Constraints for Massive Open Online Courses. In *Proceedings of the 8th ACM Conference on Recommender Systems (RecSys '14)*. ACM, New York, NY, USA, 49–56. DOI:http://dx.doi.org/10.1145/2645710.2645748

[25] Diyi Yang, Tanmay Sinha, David Adamson, and Carolyn Penstein Rosé. 2013. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the 2013 NIPS Data-driven education workshop*, Vol. 11. 14.

[26] Kai Zhang, Wei Wu, Haocheng Wu, Zhoujun Li, and Ming Zhou. 2014. 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)*. ACM, New York, NY, USA, 371–380. DOI:http://dx.doi.org/10.1145/2661829.2661908

[27] Tom Chao Zhou, Michael R. Lyu, and Irwin King. 2012. A Classification-based Approach to Question Routing in Community Question Answering. In *Proceedings of the 21st International Conference on World Wide Web (WWW '12 Companion)*. ACM, New York, NY, USA, 783–790. DOI:http://dx.doi.org/10.1145/2187980.2188201