

# Recommendation of New Questions in Online Student Communities

Jakub Mačina

Supervisor: Ivan Srba  
Slovak University of Technology

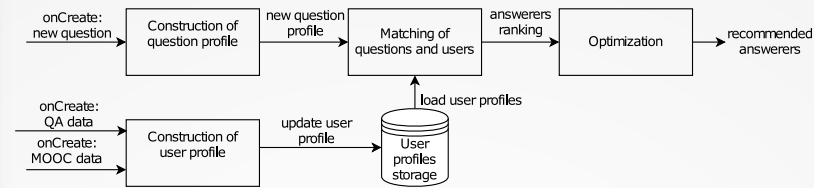
## Motivation

- Students' performance in Massive Open Online Courses (MOOCs) is enhanced by high quality discussion forums or recently emerging educational Community Question Answering (CQA) systems.
- Problems of discussion tools in MOOCs:
  - A small number of students answer questions asked by their peers.
  - An increasing proportion of unanswered questions (up to 50%).
  - Dropout rates for courses can be as high as 94%.
- Introducing a new task of educational question routing - routing of new questions without any answer in educational CQA system to suitable answerers

## Contributions

- 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.
- Incorporating additional non-QA data from the course to involve more students in question answering and reduce the burden on individual users.
- Online experiment as an A/B test of the proposed method with more than 4600 MOOC students.

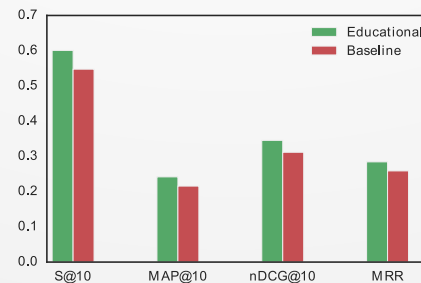
## Educational question routing framework



## Matching of questions and users

- Ensemble classifier
  - Predicting whether a user has sufficient expertise to answer a new question.
  - Predicting user willingness to answer a new question.
- Combination:
 
$$P(y = 1) = P(\text{expertise} = 1) * P(\text{willingness} = 1)$$

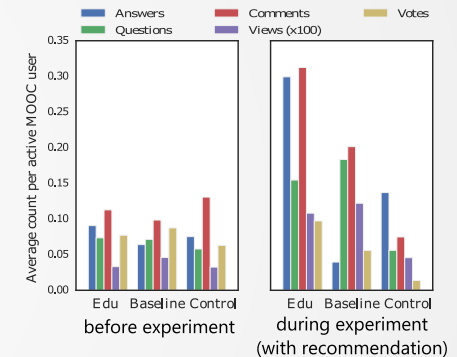
## Offline experiment



## Online experiment

- QuCryptox Quantum Cryptography offered by TU Delft and Caltech at MOOC system edX
- Open-source CQA system Askalot

Metric	Quantity
Students enrolled in the course	8115
Students started the course	4618
Users participating in CQA (with any question view)	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



## Accuracy of question recommendation

	Educational	Baseline	Statistical significance
CTR	23.25%	18.29%	$\chi^2(1, N=2640) = 10.03, p < 0.01$
Success rate	15.91%	10.61%	$\chi^2(1, N=264) = 1.61, p = 0.20$

## Conclusions

- Higher accuracy of answers prediction compared to baseline
- Higher interest of users in the routed questions and engaged more students, who in addition provided more contributions
- Lower dropout rate of active contributors in CQA and the lower instructors workload
- Cooperation with Harvard University
- Full paper accepted for ACM RecSys 2017 (acceptance rate 20.8%)