

Recommendation of Solved Questions from Archives in CQA Systems

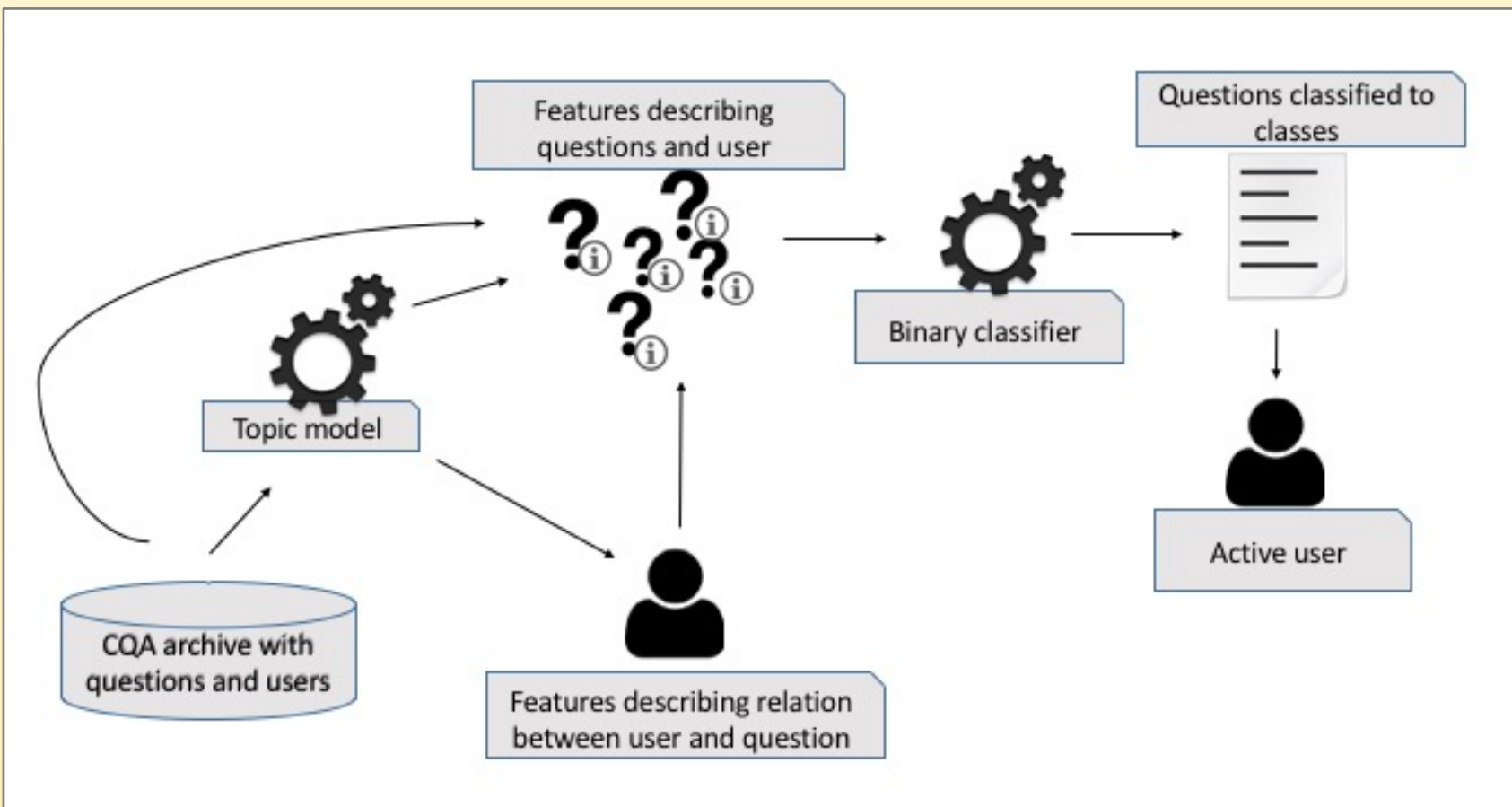
Viktória Lovasová

[viktoria.lovasova@gmail.com]

Supervised by: Ivan Srba

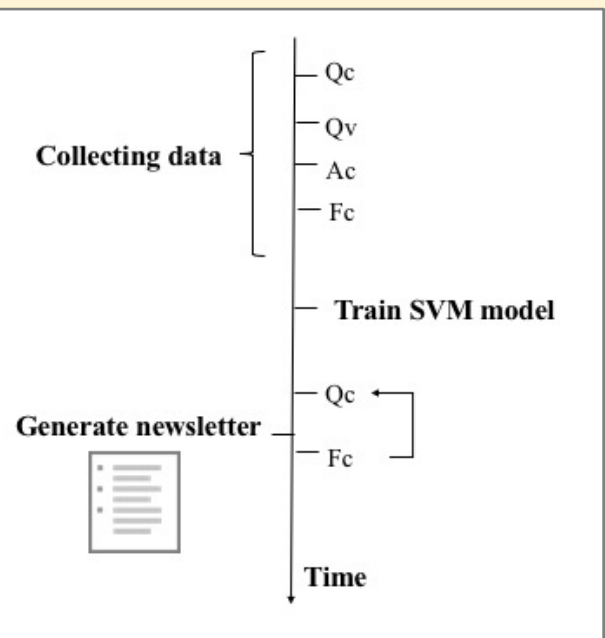


Conceptual design of the method



Evaluation of the proposed method in ranking

Qc – question create
Qv – question vote
Ac – answer create
Fc – favorite create



Features with weights entering the SVM model

Feature	Weight
1. Answers votes sum	2.97e-07
2. Question votes sum	1.81e-07
3. Views count	0.78e-07
4. Favorites count	0.64e-07
5. LDA similarity	0.5e-07
6. Best answer presence	0.1e-07
7. Score of best answerer	0
8. Highest score from answerers	0
9. Asker answers count	-0.11e-07
10. Close suggestions count	-0.66e-07
11. Answers count	-1.23e-07

Motivation

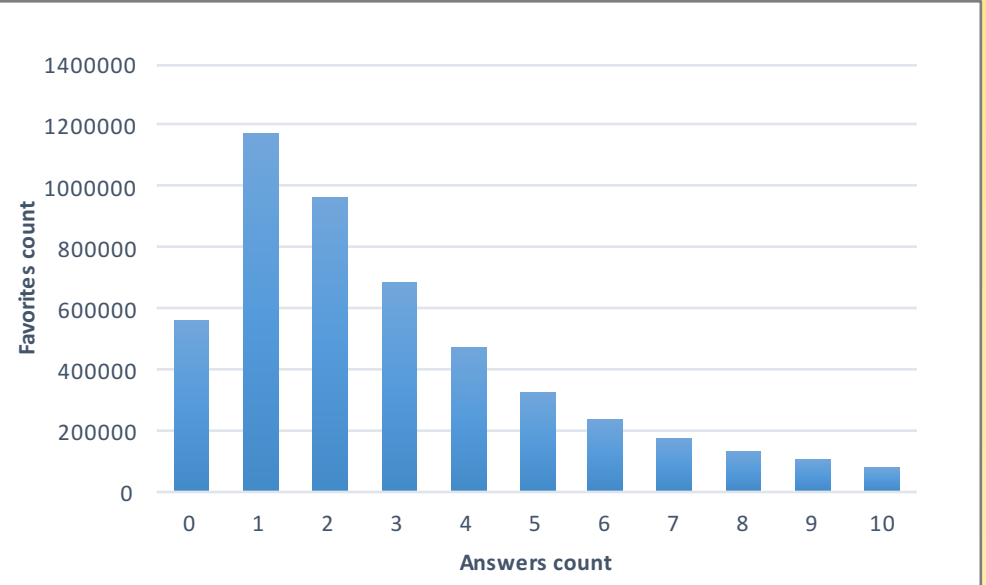
- Community Question Answering (CQA) sites are platforms to create, share and seek a massive volume of human knowledge.
- Stack Overflow and Quora have content that has long-term value (information value).
- > Recommend questions to users – existing work: non personalized recommendation
- We defined a new type of problem: propose a method for adaptive recommendation of questions which have information value for user what he/she usually expresses by favorites.
- Example: send newsletter with questions with information value to user once a week.

- 2 phase evaluation
 - Learning to classify - predict whether a user would favor a particular question or not.
 - Learning to rank – generate newsletter – 10 questions that have information value for user
- Experimental setup:
 - Offline
 - Dataset: Programmers Stack Exchange
 - Binary classifier: SVM – positive example favorite create, negative example – vote for closing the question.
 - Topic model: LDA
 - Data migration: Askalot

Evaluation results for phase 2

Algorithm	Features	Correct newsletters	MRR	Mean position
SVM	QF + Tags	0.608	0.414	16.092
SVM	QF + Tags + LDA	0.006 ?	0.014	95.869
SVM	QF + LDA	0.676	0.444	11.932
SGD	QF + Tags	0.472	0.364	21.347
SGD	QF + Tags + LDA	0.451	0.337	24.519
SGD	QF + LDA	0.52	0.381	18.945

Stack Overflow study: Favoring questions according the number of answers



Evaluation results for phase 1

Algorithm	Features	Precision	Recall	Accuracy
SVM	QF + Tags	0.784	0.833	0.801
SVM	QF+ Tags + LDA	0.794	0.817	0.803
SVM	QF + LDA	0.79	0.84	0.808
SGD	QF + Tags	0.827	0.805	0.818
SGD	QF + Tags + LDA	0.813	0.795	0.807
SGD	QF + LDA	0.771	0.753	0.765

- We achieved 67% success rate in generating newsletters with SVM algorithm -> successful given the type of problem.
- Future work – drawbacks of offline experiment -> live recommendation of questions.