

User Reputation in Community Question Answering

Adrián HUŇA*

*Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
huna.adrian@gmail.com*

The web is an enormous source of information which helps lots of people every day. Despite the amount of information available, there are still situations in which it is difficult to find specific information, or to answer a question that is too complex to be understood by a search engine. These types of situations helped Community Question Answering (CQA) systems gain popularity among users in the last years. Their principle is simple – everybody can ask a question, and anyone can answer it.

A traditional problem in systems that employ user-generated knowledge is recognition of expert users. It is important for example in order to extend their rights in managing the community, or to route hard questions. Expert identification is also an important part in the process of estimating user reputation. In general, reputation is an indicator of a value of a user to the community, and it should reflect a combination of user expertise and activity.

Reputation in CQA systems has been intensively studied in the previous years. We can broadly divide the methods into graph-based and feature-based approaches. Liu et al. [1] presented extensive performance comparison of graph-based as well as feature-based approaches. Even though their proposed graph-based approach performed as the best in some cases, very simple feature-based approach Best Answer Ratio achieved very similar or even better performance. Yang et al. [2] focused on the quality of users' contributions for expert identification. Users were labelled by a metric called Mean Expertise Contribution which takes question debatableness and answer utility into calculation. They, however, did not evaluate performance of their method.

Our main goal is to model users' reputation with accentuation on the quality of users' contributions. In our approach, reputation of a user consists of reputation gained for answering questions as well as for asking them. The gained reputation for such actions is added to previously earned reputation. We decided to calculate reputation for asking questions based on question difficulty in combination with question utility. We can expect that the longer it takes for the first answer to be added, the more difficult the

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

question is. Question utility is based on a question's score (number of positive votes minus number of negative votes). Reputation for answering questions has similar principle, as we work with question difficulty combined with answer utility.

For evaluation we have decided to use a comparison of two sorted lists of users for each question. The first list is sorted according to calculated reputation, while the second one is sorted according to the score of answers as accumulated in the CQA system. We chose four feature-based approaches as baselines to compare performance of the methods using standard information retrieval metrics. We evaluated performance of Precision@1 (P@1), Precision@2 (P@2), Mean Reciprocal Rank (MRR) and Normalized Discounted Cumulative Gain (nDCG). We tried two configurations of our method in order to completely eliminate activity factor – this variant is labelled as *average* in Table 1 which presents the performance of methods compared.

Table 1. Comparison of the performance of the methods.

| | P@1 (%) | P@2 (%) | MRR (%) | nDCG (%) | Questions |
|-------------------------------|---------------|---------------|---------------|---------------|-----------|
| Our method (<i>sum</i>) | 41.256 | 38.744 | 64.319 | 83.592 | 19861 |
| Our method (<i>average</i>) | 43.829 | 40.801 | 66.271 | 84.533 | 19861 |
| StackOverflow score | 40.360 | 38.218 | 63.707 | 83.239 | 19866 |
| Best Answer Ratio | 41.936 | 40.115 | 64.571 | 83.698 | 19651 |
| Z-score | 38.019 | 36.703 | 62.062 | 82.394 | 19866 |
| Number of answers | 38.176 | 36.998 | 62.204 | 82.500 | 19651 |

The results show that our method outperformed all the baseline methods, and thus we can confirm our assumption that consideration of content quality plays an important role in estimation of user reputation. The interesting observation is, however, the fact that the variant which completely eliminates user activity performed as the best. This suggests that we should focus only on the quality of user contributions, as even a small factor of activity can degrade precision of reputation estimation.

Extended version was published in Proc. of the 11th Student Research Conference in Informatics and Information Technologies (IIT.SRC 2015), STU Bratislava, 146-151.

Acknowledgement. This work was partially supported by the Cultural and Educational Grant Agency of the Slovak Republic, grant No. 009STU-4/2014.

References

- [1] Liu, J., Song, Y.I., Lin, C.Y.: Competition-based User Expertise Score Estimation. In: *Proc. of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '11, New York, NY, USA, ACM, 2011, pp. 425–434.
- [2] Yang, J., Tao, K., Bozzon, A., Houben, G.J.: Sparrows and Owls: Characterisation of Expert Behaviour in StackOverflow. In: *User Modeling, Adaptation, and Personalization*. Volume 8538 of *Lecture Notes in Computer Science*. Springer International Publishing, 2014, pp. 266–277.