

An Automated Approach for suggesting Tag based Questions and their Solutions

Gulshan Vaish, Anirudh Negi, B. Jothi

Students, Students, Assistant Professor
SRM University, Chennai

Abstract—Keywords or ‘tags’ are useful for providing more efficient and quicker approach to search for required content. Question and answer (Q&A) sites also use tags to help ensure that their users are aware of questions related to their areas of expertise or interest. In some cases, people are unsure of the exact questions for which they need solutions, use of tag based suggestions may help them define their questions. In this paper, we categorize the questions that are asked by people based on the keywords or tags they provide, we automatically suggest question tags to help people choose appropriate tags for eliciting a response. To help questioners to get their preferable answers, a set of possible expert users should be recommended.

Index Terms—tags, suggestions, experts

I. INTRODUCTION

A Q&A site helps online community to share knowledge. People with questions get the answers they need. In information systems, tagging is a popular way to categorize information and to search content. Therefore, almost all online newspapers, blogs, question-answer communities, and other similar sites make use of tags to categorize articles, posts, questions, answers, and so on. Similarly, our system uses tags to categorize computer science related questions so that users can find similar questions on the same topic or find questions that they may be able to answer. A user can add up to three tags to categorize a question using existing tags or using new tags they create. However, a user must have a certain level of reputation to create a new tag, which is determined by the site using a reputation score. Appropriate tagging of questions may be useful to get a quick answer because potential responders are able to be notified when a question is posted with a tag related to their interests.

The great success of CQA services comes with the research studies in different directions in community question answering. The major research contains three aspects: expert learning, best answer detection and similar question retrieval [4]. In this paper, we mainly focus on expert user recommendation, which has received increasing attention in recent years. Given a question, the problem is how to identify highly ranked authoritative answerers for askers. An early study on the expert finding is to build user profile based on the questions they have previously answered, and then rank those user profiles according to the query questions. The latest work considered that the expertise of users should be learnt based on topics and voting scores of their previous Q&A posts. However, the tags in each question were ignored in most work, which tend to be more informative than user profiles or user interested topics. In each question, the tags provided by askers can be viewed as the most representative words of the focus and intent of the asker. Generally, the voting score of an answer can reflect the expertise of its provider towards the corresponding question, which also can be viewed as the reflection of expertise on the tags of this question. Figure 1 is a sample of Q&A posts from the popular Stack Overflow site. Three tags for the question “What do the getUTC* methods on the date object do?” are ‘JavaScript’, ‘date’ and ‘utc’, which are more representative than the generic topic for a question.

Due to the importance of tags, in this paper, we proposed to learn the user expertise only based on tags and the voting scores of answers in CQA.

Unlike existing methods, we won’t use the content of Q&A posts. Specifically, we build user-tag expertise matrix to depict the tag-level expertise of users. We assume that a user with the highest expertise score towards the tags of a given question tends to be the best candidate to be recommended.

TABLE I. A SAMPLE OF USER-TAG SCORE PAIRS FOR A GIVEN QUESTION.

User Id \ Tag Id	T_1	T_2	T_3	T_4
U_1	0.25	0.05	0.70	0.50
U_2	0.35	0.45	0.25	0.05
U_3	0.05	0.25	0.15	0.80

An example for ranking 3 candidate users based on 4 tags of a new question is shown in Table I. It can be easily computed that the average score for user U_1 towards all the 4 tags is $(0.25+0.05+0.70+0.50)/4=0.375$. And such for U_2 and U_3 are 0.275 and 0.3125 respectively. Then we could get the ranking of users as U_1 , U_3 , and U_2 in descending order.

II. RELATED WORK

In recent years, there have been increasing research studies in expert learning in community question answering. In traditional methods, the user profiles or user topics are learnt based on the previously answered questions. To find experts, the user profiles were created from the previously answered questions, then these profiles were ranked by language models, where each given question was seen as a query.

Lately a topic expertise model (TEM) was proposed as a probabilistic graphical model by integrating user topics and topical expertise for CQA learning, in which the user topical expertise are learnt based on the Q&A posts and the user expertise is related to the voting scores of answers. Moreover, the social link information was considered together with the results of TEM to further improve the system performance, where the user expertise is still studied under the general topics. The above mentioned approaches considered the overall profiles or topics of each user. However, the tags of user previously answered questions are actually more informative and representative in depicting user's specialty, which can be seen as the subtopics. To this end, we model the user expertise under tags rather than the general topics.

III. FRAMEWORK OF EXPERT RECOMMENDATION

In this section, we will introduce our framework for expert recommendation in community question answering. We first build user-tag expertise matrix based on the training dataset, then the latent user feature matrix and tag feature matrix will be obtained via user-tag matrix factorization. Finally we conduct expert recommendation based on the predicted expertise scores.

• User-Tag Matrix Factorization

In some CQA services, like Stack Overflow, each question has been attached several tags by askers, these tags are more representative than the content of questions. We assume that a user with high expertise regarding a question is likely to be skilled at its tags. Therefore, we learn the user expertise under tags, and we create the user-tag expertise matrix based on the user previously answered questions. In the user-tag expertise matrix, suppose there are M users and N tags, s_{ij} represents the averaged expertise score of user i for tag j . $U \in \mathbb{R}^{L \times M}$ and $Y \in \mathbb{R}^{L \times N}$ are latent user and tag L -dimensional feature matrices.

• Expert Recommendation

Once we get the user matrix U and tag matrix Y , the missing expertise value s_{ij} in S can be predicted by using $U^T i Y_j$, then we can conduct expert user recommendation with the help of the completed user-tag expertise matrix.

IV. PROPOSED SYSTEM

Uses MySQL as the main database storage. Question and Answers can also be posted to Facebook, Twitter, Tumblr and Blogger all at the same time, using post-processing feature that does not slow down the page load at all - all API calls are made after the page loaded and browser connection is closed. Users are rewarded with reputation points for posting content to external sites (Twitter, Tumblr, Facebook, and Blogger). Members can follow questions they are interested in monitoring. User is rewarded with extra points for accepting answers. User can vote questions and answers up/down, can also un-vote (cancel vote). Votes increase/decrease reputations on question/answer posters. There is an optional support for code editor and code highlighter that supports all of the popular programming languages as well as CSS and HTML code highlighting (this feature can be turned on and off from config file).

The system has the following modules:

- **Login/Register Module:** Users who want to ask new questions or answer to a question need to login into the system if already registered. This will provide confidentiality and non-repudiation. Registrations to the system include registrations through social networking sites like Facebook, google+, etc.
- **Search module:** Users can search for the questions which have already been asked by other registered users. Suggestions are provided by the system based on the tags/technical terms they mention.
- **Add Module:** This module enable users to add their questions to the database so that other users can provide solutions to them.

- **Chat Module:** This is to provide one-to-one conversation between two users for discussion on a common topic.
- **Ranking/Rating Module:** Based on the number of valid solutions provided and number of commendations a user is rated by the system. This rating is compared with other users and a global ranking is generated.
- **Code Editor Module:** This module supports most of the popular programming languages to distinguish between code snippet and description.

V. Conclusion and Future Work

In community question answering sites, askers post their questions and expect to get reasonable answers from experts. However, the number of expert users is limited and their expertise on each topic can be different. It is very useful and necessary to recommend authoritative users to answer different questions according to their expertise [1]. Existing methods consider that the user expertise is defined on certain topics. Nevertheless, each topic can be divided into many subtopics, two questions about the same topic can belong to two different subtopics [2]. Hence it is better to find out an expert specialized on this subtopic. In this paper, we view the tags of each question as the representative words of the subtopic of each question. In other words, we study the user expertise under tags rather than topics [6]. The tags of each question are usually not independent, some of them can appear simultaneously in many questions. It might be useful to explore the relationship between tags and incorporate it in the study of expert recommendation in CQA.

Acknowledgment

The paper was completed under the gracious guidance of Mrs B. Jothi, as a guide for the final year project. I am also obliged to the Department of Software Engineering for conducting a “NATIONAL CONFERENCE ON IMPACT OF SOFTWARE ENGINEERING ON COMPUTER TECHNOLOGIES” for giving us an opportunity to be a part of this event and submit our ideas and research work.

REFERENCES

- [1]. J. Guo, S. Xu, S. Bao, and Y. Yu, “Tapping on the potential of q&a community by recommending answer providers,” in *Proceedings of the 17th ACM conference on Information and knowledge management*. ACM, 2008, pp. 921–930.
- [2]. J. Sung, J.-G. Lee, and U. Lee, “Booming up the long tails: Discovering potentially contributive users in community-based question answering services,” in *Seventh International AAAI Conference on Weblogs and Social Media*, 2013, pp. 602–610.
- [3]. T. Hao and E. Agichtein, “Finding similar questions in collaborative question answering archives: toward bootstrapping-based equivalent pattern learning,” *Information retrieval*, vol. 15, no. 3-4, pp. 332–353, 2012.
- [4]. Q. Tian, P. Zhang, and B. Li, “Towards predicting the best answers in community-based question-answering services,” in *Seventh International AAAI Conference on Weblogs and Social Media*, 2013, pp. 725–728.
- [5]. J. Liu, Y.-I. Song, and C.-Y. Lin, “Competition-based user expertise score estimation,” in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 2011, pp. 425–434.
- [6]. Baoguo Yang, Suresh Manandhar, “Tag-Based Expert Recommendation in Community Question Answering” in “2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)”