

Knowledge Sharing via Social Login: Exploiting Microblogging Service for Warming up Social Question Answering Websites

Yang Xiao¹, Wayne Xin Zhao², Kun Wang¹ and Zhen Xiao¹

¹School of Electronics Engineering and Computer Science, Peking University, China

²School of Information, Renmin University of China, China

{xiaoyangpku, batmanfly}@gmail.com

{wangkun, xiaozhen}@net.pku.edu.cn

Abstract

Community Question Answering (CQA) websites such as Quora are widely used for users to get high quality answers. Users are the most important resource for CQA services, and the awareness of user expertise at early stage is critical to improve user experience and reduce churn rate. However, due to the lack of engagement, it is difficult to infer the expertise levels of newcomers. Despite that newcomers expose little expertise evidence in CQA services, they might have left footprints on external social media websites. Social login is a technical mechanism to unify multiple social identities on different sites corresponding to a single person entity. We utilize the social login as a bridge and leverage social media knowledge for improving user performance prediction in CQA services. In this paper, we construct a dataset of 20,742 users who have been linked across Zhihu (similar to Quora) and Sina Weibo. We perform extensive experiments including hypothesis test and real task evaluation. The results of hypothesis test indicate that both prestige and relevance knowledge on Weibo are correlated with user performance in Zhihu. The evaluation results suggest that the social media knowledge largely improves the performance when the available training data is not sufficient.

1 Introduction

One of the main challenges for social startup websites is how to gain a considerable number of users quickly. A growing number of social startups outsource sign-up process to existing social networking services. They allow users to log in to the services using their existing social media accounts. For example, Quora allows users to log in with their Google, Twitter or Facebook accounts based on the OpenID technology. Lots of startup web services benefit from the huge number of users and rich relationships accumulated by social network sites. Social login helps the newborn web services to collect crowds of users in a short time. Moreover, startup web services can gain reliable profiles through social login. It also offers a convenient mechanism for users to surf the web using a unified social identity (e.g., Twitter account). For example, by the end of 2013, there are about 600,000 web services including mobile applications using social login offered by Sina Weibo.

When we go beyond simple import of profiles and consider the general problem of leveraging knowledge from social media, many subtasks arise. One of them is how to incorporate data from social media and startup web service to better predict user performance. In this paper, we take the largest social based question answering service Zhihu in China, which closely resembles Quora, as the testbed. Different from traditional CQA sites such as Baidu Zhidao, Zhihu have more prominent social features, which supports login with Sina Weibo accounts. Although Zhihu grows quickly and attracts more and more users, about 85% of the users answer fewer than 10 questions and 60% of the users answer fewer than 4 questions in our dataset, which is a large sample of Zhihu.

Previously, many studies have been proposed to improve expertise ranking on CQA services. Link analysis based approaches (Jurczyk and Agichtein, 2007; Zhang et al., 2007) exploit the question-answering relationships to construct a graph and run PageRank or HITS on the graph. Jeon et al. (2006)

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Page numbers and proceedings footer are added by the organisers. Licence details: <http://creativecommons.org/licenses/by/4.0/>

propose a method based on the non-textual behaviors. Moreover, co-training model (Bian et al., 2009) jointly infers answer quality and user expertise. Liu et al. (2011) formalize expertise ranking as a competition game with the insight that the best answerer beats other answerers in the same question thread. However, the above studies highly rely on the history data, which might not work well for newcomers or users with few answering records. For startup services, many users may not accumulate sufficient data to support the reliable estimation for their expertise levels. Indeed, the importance of newcomers has been noted in related studies, and it has been shown that the effective evaluation of users' performance at an early stage significantly affects the overall development of QA services (Nam et al., 2009; Sung et al., 2013).

In this paper, we propose a method that incorporates social media and social startup data to predict newcomers' performance. This problem is technically challenging due to the heterogeneous characteristics across websites. Given a user, we hypothesize that her capability of contributing high quality answers is dependent on her prestige and relevance. The more contents a user publishes on an area and the higher prestige a user has on social media sites, the higher likelihood that user can offer high quality answers. Thus, the first goal is to precisely measure the relevance between question and a user's tweets. Owing to the short question length and noisy tweet content, this problem brings technical challenges. We make use of user-annotated tags and adopt a translation based model to improve relevance estimation. For prestige, a straightforward way is to use the standard graph based ranking algorithm, however, Zhihu users have very sparse links on Weibo and the standard PageRank algorithm does not work well on sparse graphs. To address it, we add virtual links to alleviate the sparsity problem by finding available paths on a large Weibo graph. Furthermore, we propose a performance biased random walk algorithm and naturally incorporates Zhihu performance history as the supervised information.

We carefully construct a dataset of 20,742 users who have been linked across Zhihu and Weibo, which represent the social startup and the social media site respectively. We first conduct Spearman correlation test for these two hypotheses. Our results have shown that prestige in Weibo has a strong correlation with overall performance in Zhihu. For the performance in question level, we have found that the relevance of Weibo contents is also significantly correlated with answer quality in Zhihu. Based on these findings, we further incorporate the extracted prestige and relevance knowledge into the existing framework for user performance prediction. To simulate the process of history data accumulation, we also conduct experiments with the varying observed number of answers. The experiment results suggest that the borrowed social media knowledge, i.e., prestige and relevance information in Weibo, largely improves the performance when the available training data is not sufficient. Interestingly, we have found that even individual prestige feature can achieve very competitive results.

Although our approach is tested on a joint combination of Weibo and Zhihu, it is equally applicable to other knowledge sharing startup web services. The flexibility of our approach lies in that we identify two important and general types of knowledge that are easy to leverage from external social media sites.

The rest of this paper is organized as follows. The construction of the dataset collection and the problem formulation are given in Section 2 and 3 respectively. Section 4 presents the detailed feature engineering and is followed by the experiment part in Section 5. Finally, the related work and conclusions are given in Section 6 and 7 respectively.

2 Construction of the Dataset Collection

We focus on a popular social question answering website, Zhihu as the studied service. We select Sina Weibo, the largest Chinese microblogging service as the external website to help improve user expertise estimation task in Zhihu. We exploit the social login mechanism to identify the same user across these two platforms: if a user logs in Zhihu with her Weibo account, her Zhihu profile will contain the corresponding Weibo account link. This approach accurately links users across websites.

Zhihu dataset. Zhihu¹ is a social based question answering site in China, which is similar to Quora in terms of overall design and service. Zhihu has three major components: users, questions, and topics. Users on Zhihu can ask and answer questions, furthermore, they can comment on or vote for answers.

¹<http://www.zhihu.com>

Each question is usually assigned with a small set of topic tags by the asker and opens a discussion thread consisting of candidate answers. Topics are represented as tags and organized in a directed acyclic graph where a child topic can have multiple parent topics.

Zhihu was founded in January 2011, and we obtain the data between January 2011 and November 2013 via a Web crawler. The dataset contains 266,672 users, 819,125 questions and 2,730,013 answers. These questions are associated with 44,333 topic tags. Since the aim is to examine whether knowledge extracted from Weibo is helpful to improve tasks in Zhihu, we only keep the users who explicitly use social login and get 136,002 cross-site users, which roughly covers 50% of the users in our dataset. For a robust evaluation, we further remove users who have answered fewer than ten questions. Finally, we obtain a total of 20,742 users and summarize the data statistics in Table 1.

#users	#topics	#questions	#answers
20,742	44,333	335,145	883,373

Table 1: Basic statistics of Zhihu dataset for linked users.

Weibo Dataset. Sina Weibo is the largest Chinese microblogging service which has about 500 million registered users by the end of 2012. We have crawled all the detailed information of these 20,742 linked users, including tweets, followers, and following links. These users are indeed active on Weibo and have posted 21,121,955 tweets in total. In later sections, we will adopt the PageRank algorithm to estimate the prestige scores of these linked users, thus we need a dense following graph for reliable estimation. By using these linked users as seeds, we further crawl their followings and followers as well as the following links between all the crawled users. Finally, we obtain 253,361,449 edges between 1,322,425 users. Note that we only use these 20,742 linked users for further study, and the rest are only used to help compute more accurate PageRank scores.

In what follows, we refer to a user who has both a Weibo account and a Zhihu account in our dataset as a *linked user*.

3 Problem Formulation

Users are the most valuable resource in community question answering (CQA) services. Discovering users' expertise at an early stage is important to improve the service quality. A typical task on CQA services is to predict users' performance or expertise: given a question, it aims to estimate the user expertise level and identify experts who can provide good answers to this question.

Borrowing the ideas from information retrieval, we solve the performance prediction task via the learning to rank framework (Liu, 2009). Formally, we assume that there are a set of m questions (i.e., queries) $\mathcal{Q} = \{q^{(1)}, q^{(2)}, q^{(3)}, \dots, q^{(m)}\}$. A question is associated with a set of $n^{(i)}$ answers $\{a_1^{(i)}, \dots, a_{n^{(i)}}^{(i)}\}$ provided by $n^{(i)}$ users $\{u_1^{(i)}, \dots, u_{n^{(i)}}^{(i)}\}$ respectively. For each user, let $y_j^{(i)}$ denote the performance score of user $u_j^{(i)}$ with respect to query $q^{(i)}$. A higher value of $y_j^{(i)}$ indicates better performance for query $q^{(i)}$. In our work, we instantiate the performance score by *the number of votes* that a user receives on a question. A feature vector $\mathbf{x}_j^{(i)}$ is constructed based on a pair of question and user $(q^{(i)}, u_j^{(i)})$. The aim of the learning task is to derive a ranking function f such that, for each feature vector $\mathbf{x}_j^{(i)}$, it outputs a prediction score $f(\mathbf{x}_j^{(i)})$ for the performance of user $u_j^{(i)}$ on the question $q^{(i)}$. With this function, when a new question comes, we can predict who will be competent at it.

For prediction tasks, the answer information $\{a_1^{(i)}, \dots, a_{n^{(i)}}^{(i)}\}$ is not available during training. Besides users' accumulated history data on Zhihu, external knowledge from Weibo is available to help construct the query-user feature vector. We assume that the studied Zhihu users have already been linked to the corresponding Weibo accounts, and we can obtain their Weibo information, including tweets and followings/followers. The key of the learning to rank framework is how to derive effective features. In our task, we consider two types of features, i.e., Zhihu features and Weibo features. Our focus in this paper is how to leverage microblogging information for improving CQA service, i.e., how to incorporate knowledge from Weibo as features into the learning to rank framework.

4 Feature Engineering

In this section, we discuss how to derive effective features from both Zhihu and Weibo. In particular, we mainly study how to leverage Weibo knowledge for the current task.

4.1 Weibo features

In our work, we focus on two types of Weibo features: prestige and relevance. For prestige, it aims to capture the social status of a user. In our setting, it refers to the status or authority level of a user on online social networks (Anderson et al., 2012). We hypothesize that a user is likely to have similar status levels across multiple online communities, thus the prestige scores of Zhihu users can be roughly estimated based on the rich link information of Weibo. The second type of knowledge we consider is relevance. A user is more likely to be an expert on an area that she is interested in, and Weibo provides a good platform to identify users' interests. Since Weibo and Zhihu are text based websites, we hypothesize that a user will show similar interests on these two medias.

Prestige. Prestige features aim to capture the status of one user. Status characteristic theory posits that one with higher status characteristic is expected to perform better in the group task (Oldmeadow et al., 2003). Prestige estimation has been a classical problem in both web graph analysis and social networking analysis (Easley and Kleinberg, 2012). We are motivated by previous study on authority ranking in Twitter (Kwak et al., 2010), which utilizes the following relations as the evidence of authority. A straightforward way is to run standard PageRank algorithm on the Weibo subgraph consisting of these 20,742 linked users. However, the subgraph of these linked users is very sparse, each linked user has only about 5 out-links to other linked users on average. Such a sparse graph will not produce meaningful ranking results.

Our solution is to add virtual links between linked users. Let N ($N = 20,742$) denote the number of linked users and $\mathbf{M}_{N \times N}$ denote the transition matrix based on the graph of these linked users. Given two users u_i and u_j , we check whether there is a directed path between them on our large Weibo graph. Recall that we have 253,361,449 edges between 1,322,425 users in Weibo dataset. We run the breadth-first search algorithm to find the shortest path between two linked users. If there exists a directed path between two linked users, we add a virtual link between them and set the weight to *the reciprocal of the shortest path length*, i.e., $I(i, j) = \frac{1}{\text{len}(u_i \rightarrow u_j)}$, where $\text{len}(u_i \rightarrow u_j)$ denotes the length of the shortest path between u_i and u_j . In this way, we have $M_{ij} = \frac{I(i, j)}{\sum_k I(i, k)}$. By adding virtual links, we obtain a more dense graph of these linked users. Formally, the standard PageRank algorithm (Brin and Page, 1998) can be formulated as:

$$\mathbf{r}^{(n+1)} = \mu \cdot \mathbf{M}^T \cdot \mathbf{r}^{(n)} + (1 - \mu) \cdot \mathbf{y} \quad (1)$$

where μ is the damping factor usually set to 0.85 and \mathbf{y} is the restart probability vector usually set to be uniform (Yan et al., 2012). When the algorithm converges, we can obtain the stationary distribution of users (i.e., \mathbf{r}) as the prestige scores.

The above method assumes that users have same restart probability, which may not be true in reality. Since we are considering improving Zhihu service quality, we incorporate users' history data from Zhihu as supervised information. The main idea is that instead of using a uniform restart distribution \mathbf{y} , we use a performance biased restart distribution in Eq. 1. We set the restart probability of a user to her average vote ratio based on the questions she has answered. Formally, we set $y_u = \text{Average}(\sum_q \frac{\#vote(q, u)}{\sum_v \#vote(q, v)})$, where $\#vote(q, u)$ denotes the number of votes user u receives on question q and $\sum_v \#vote(q, v)$ denotes the total number of votes that all users receive on question q . We do not use other measures such as best answer ratio because we assume that the history window is very limited and our proposed method provides more robust estimation. Let us further explain the idea. At the beginning of each iteration, each user is assigned to her performance score estimated based on Zhihu data: the more competent she is, the larger score she has. During the iteration, each user begins to collect authority evidence from her incoming neighbors on the Weibo graph. The final score is indeed a trade-off between her own performance on Zhihu and her authority on Weibo.

There are also other measures to consider, e.g., the follower number and the times of being retweeted. In our experiments, we have tried these variants and found that no one is more effective than the above method.

Relevance. Intuitively, a user is more likely to be an expert on an area that she is interested in. In the setting of Zhihu, a user tends to perform better on the topics that are more relevant to her interests. Status characteristic theory also conveys that task relevance is an important factor which affects one’s performance (Oldmeadow et al., 2003). Weibo provides a good platform to infer users’ interests, which is helpful to derive relevance scores.

We formulate relevance estimation as an information retrieval task. Let \mathcal{V} denote a term vocabulary and w denote a word in \mathcal{V} . Note that we take the union of the Weibo vocabulary and Zhihu vocabulary. The interest of a user u is modeled as a multinomial distribution over the terms in \mathcal{V} , i.e., $\theta^u = \{\theta_w^u\}_{w \in \mathcal{V}}$. Given a question q , we also model it as a multinomial distribution over the terms in \mathcal{V} , i.e., $\theta^q = \{\theta_w^q\}_{w \in \mathcal{V}}$. Following (Zhai, 2008), the relevance score between question q and user u can be estimated by the negative Kullback-Leibler divergence between θ^q and θ^u :

$$\text{Rel}(q, u) = -KL(\theta^q, \theta^u) = - \sum_{w \in \mathcal{V}} p(w|\theta^q) \log \frac{p(w|\theta^q)}{p(w|\theta^u)} \quad (2)$$

We first estimate θ^q . The straightforward way is to estimate θ^q based on the question text. However, the question text is usually short and noisy, which does not yield good results in our experiments. Recall a question is associated with a small set of user-annotated topic tags, and tags are good semantic indicators of the question. A topic tag usually indexes a considerable amount of questions, and we can use tags to leverage semantics from the indexed questions. Formally, we adopt the translation based model (Zhai, 2008) to estimate the question model:

$$\theta_w^q \propto \sum_{t \in q} p(w, t|q) = \sum_{t \in q} p(w|t)p(t|q) \quad (3)$$

where $p(w|t)$ is the translation probability from a tag to a term, and $p(t|q)$ is the empirical distribution of tag t in question q . Here we make an independent assumption: given a tag, the question is independent of a word, i.e., $p(w|t, q) = p(w|t)$. The procedure can be interpreted as follows: sample a tag from the question and then compute the probability of translating the tag into a specific word. We estimate the tag-term translation probability as $p(w|t) = \frac{\#(w, t) + 1}{\sum_{w' \in \mathcal{V}} \#(w', t) + |\mathcal{V}|}$, where $\#(w, t)$ denotes the term frequency of w in the question text that tag t indexes. We use the additive-one smoothing.

We also try to incorporate the question text into the above estimation formula. However, it does not result in any improvement. The main reason is that the question words may be too specific, as a comparison, tags provide a general level of semantics, which is more effective to identify expertise areas.

Next, we estimate user interest model θ^u . We consider aggregating all the tweets of a user as a “document”, and then estimate the document-term probability as $\theta_w^u = \frac{\#(w, u) + 1}{\sum_{w' \in \mathcal{V}} \#(w', u) + |\mathcal{V}|}$, where $\#(w, u)$ denotes the term frequency of w in the aggregated document of user u .

4.2 Zhihu features

Now we describe the features extracted from Zhihu, and we refer to them as *baseline features* since we take the performance of them as a base reference. We summarize these features in Table 2.

These features have been extensively tested to be very effective by previous related studies (Song et al., 2010; Liu et al., 2011), which represent the state-of-art of the current task.

Summary. We have considered two general types of knowledge in social media which are potential to improve user expertise estimation in Zhihu. It is easy to see that our approach can be equally applicable to other third-party websites which is text based and contain manually annotated tags.

Features	Abbr	Formulas
Number of Best Answers	NBA	—
Number of Answers	NA	—
Number of Received Votes	NV	—
Average Number of Votes	AVA	—
Smoothed Average number of Votes	SAVA	$SAVA(u) = \frac{\sum_q \sigma(v(q,u))}{NA(u)}, \sigma(x) = \frac{1}{1+e^{-x}}$
Best Answer Ratio	BAR	$BAR(u) = \frac{NBA(u)}{NA(u)}$
Smoothed Best Answer Ratio	SBAR	$SBAR(u) = \frac{BAR(u)*NA(u)+BAR_{avg}*NA_{avg}}{NA_{avg}+NA(u)}$
Average Answer Length	AAL	—

Table 2: List of baseline features with corresponding abbreviations and formulas. Here u denotes a Zhihu user.

5 Experiment

Questions with fewer than five answerers do not receive much attention, and we only keep questions which involve at least six users. In this way, we have obtained a total of 25,262 questions. The number of votes is used as the measure of answer quality. The question threads are sorted by the post time, and we can simulate the cold-start phenomenon to examine the performance of different methods. We split the dataset into a training set and a test set by question threads with the ratio of 3:1. The “history” data of a user is put into the training set and the rest is treated as test data. We further vary the size of “history data” that can be used for performance prediction in three levels, i.e., at most 3, 5, and 10 “historical” question threads have been observed for a given user.

5.1 Hypothesis Testing

In this part, we first examine the fundamental hypotheses of our work: whether Weibo knowledge is potentially effective to improve the performance of tasks in Zhihu. We conduct significance test to examine the correlation between user features extracted from Weibo and user performance in Zhihu. We adopt the Spearman’s rank correlation coefficient as the test measure. For a sample of size n , the n raw scores X_i, Y_i are converted to ranks x_i, y_i , and the Spearman correlation coefficient ρ is computed as $\rho = 1 - \frac{6 \sum_i d_i^2}{n(n^2-1)}$, where $d_i = x_i - y_i$. The Spearman’s coefficient ρ lies in the interval $[-1, 1]$, and a value of “+1” or “-1” indicates a perfect, positive or negative Spearman correlation.

Test of prestige. In our test, the overall performance of a user is estimated by the average vote counts she receives per answer, and the prestige level of a user is estimated by her PageRank score on the original Weibo following graph with a uniform restart probability. With these two measures, it is straightforward to generate two rankings of users, either by user prestige level or by user performance. However, it is noted that ρ is usually very sensitive when the sample size is too large, and it is difficult to obtain robust correlation values in this case. To better capture the overall correlation patterns, we group users according to their prestige levels and examine the correlation degree in the group level. We sort users according to their PageRank scores in a descending order, and split users equally into 100 buckets. The correlation value between performance and PageRank is $\rho = 0.5617$ at the significance level of $9.879e^{-10}$, which indicates there is a strong correlation between performance and prestige.

Test of relevance. Different from prestige, relevance is defined to be question specific, so we cannot perform global correlation analysis. We perform the correlation analysis in the question level. For each question, we have two rankings of involved users: the relevance ranking and the question-specific performance ranking. Let ρ denote the correlation coefficient between the relevance ranking and the performance ranking for a given question. Formally, given a question, we have the null hypothesis H_0 being “ ρ is zero”, whereas H_1 being “ ρ is not zero”. If H_0 is rejected, we can conclude that prestige in Weibo is correlated with users’ performance on Zhihu for the given question. Our experiments have shown that 14.48% of the questions rejected the H_0 hypothesis at the confidence level of 0.9.

5.2 Evaluation metrics

In the above, we have shown that prestige and relevance knowledge extracted from Weibo are correlated with user performance in Zhihu. Next we are going a step further to examine the feasibility of using these external features to improve user performance ranking in CQA service. In this paper, we consider studying this problem in two aspects: in the first case, we only focus on the user who provides the *best answer*; while in the second case, we focus on the *overall ranking* of all engaged answerers in a given question thread. By following previous studies (Song et al., 2010; Deng et al., 2012), we adopt traditional evaluation metrics in information retrieval for evaluating user performance prediction in CQA services.

Best answer prediction. Our first task is to predict which user will provide the best answer given a question. The user who has received the maximum vote counts in a question thread will be labeled as relevant and the rest will be treated as non-relevant. Then we can adopt the widely used relevance metrics Precision at rank n ($P@n$) and Mean Reciprocal Rank (MRR).

Top expert recommendation. Unlike best answer prediction, top expert recommendation aims to provide a short list of candidate experts given a question. By following the study (Liu et al., 2011), we use nDCG (normalized Discounted Cumulative Gain) as the evaluation metrics. Let $vote(i)$ denote the vote counts of the answer ranked at i in a system output. To reduce the effects of large outliers, we set the gain value for an answer with the vote counts v to be $\log(v + 1)$. The metrics are formally defined as follows:

$$DCG@n = \sum_{i=1}^n \frac{\log(vote(i) + 1)}{\log(i + 1)} \quad (4)$$

$$maxDCG@n = \sum_{i=1}^n \frac{\log(vote^*(i) + 1)}{\log(i + 1)} \quad (5)$$

$$nDCG@n = \frac{DCG@n}{maxDCG@n} \quad (6)$$

where $vote^*$ denotes the vote counts list of the ideal ranking system, i.e., the answer list is sorted by vote counts in the descending order.

Similar to query-specific information retrieval tasks, all our experiments are question specific. For a system, we evaluate its performance of each question and then average all the results as the final performance.

5.3 Results

As studied in Section 3, the above two tasks can be formulated as the learning to rank problem. Following previous work (Song et al., 2010), we adopt SVMRank as the ranking model and implement SVMRank using the tool package SVMLight². We use the linear kernel for SVMRank, and report the results in Table 3 and Table 4.

We refer to the system with all Zhihu features as *Baseline*. We use two ways to compute prestige features: $P+UniformG$ denotes the system which implements the standard PageRank algorithm with uniform restart probability, while $P+HisG$ denotes the system which implements the biased PageRank algorithm with users' history performance on Zhihu as the restart probability. *Rel* denotes the system with only relevance features and *Baseline+Weibo* denotes the system with all the features.

Analysis of baseline results. The baseline system is built with all Zhihu features, which are estimated using history data, and it is natural to see that the performance of the baseline system improves with the increasing of the history data. Recall that all the question threads in our dataset contain more than six answers, indeed, 36.3% of them contain more than ten answers. A random algorithm to guess the best answer can only achieve a poor $P@1$ value of 11.07%. Results in Table 3 and Table 4 show that our baseline is competitive even on long question threads. In our experiments, the system performance begins to stay stable when the history window is set to ten question threads since quite a few users have engaged in fewer than ten question threads.

²<http://svmlight.joachims.org>

History Window Size	Systems	NDCG@1	NDCG@3	NDCG@5
NULL	P+UniformG	0.510	0.555	0.621
	Rel.	0.360	0.434	0.519
≤ 3 question threads	(B)aseline	0.508	0.582	0.656
	P+HisG	0.550	0.596	0.658
	B.+Weibo vs. B.	0.580 +14.17%**	0.617 +6.01%***	0.676 +3.05%***
≤ 5 question threads	(B)aseline	0.509	0.578	0.658
	P+HisG	0.556	0.603	0.668
	B.+Weibo vs. B.	0.589 +15.72%***	0.625 +8.13%***	0.687 +4.41%***
≤ 10 question threads	(B)aseline	0.534	0.602	0.671
	P+HisG	0.568	0.616	0.679
	B.+Weibo vs. B.	0.595 +11.42%	0.637 +5.81%	0.696 +3.73%*

Table 3: Overall ranking performance with varying history window sizes. “*”, “***”, “****” indicate the improvement is significant at the level of 0.1, 0.05 and 0.01 respectively.

Analysis of the effect of Weibo features. We now incorporate Weibo features and check whether they can help improve the system performance. In Table 3 and Table 4, we present the improvement ratios over baselines with the incorporation of Weibo features. We can see that Weibo features yield a large improvement over the baseline system, especially when the size of history window is small, i.e., ≤ 3 question threads. This indicates the effectiveness of Weibo features on alleviating the cold-start problem in Zhihu. When we have more history data, i.e., ≤ 10 question threads, the improvement becomes smaller.

It is noteworthy that the single prestige feature (i.e., *P+UniformG* and *P+HisG*) achieves good performance. Especially, *P+HisG* obtains very competitive results compared with the baseline system. *P+HisG* naturally combines history data on Zhihu and prestige information on Weibo, which largely improves the standard prestige estimation method *P+UniformG*. As a comparison, the relevance feature is not that effective but still improves the overall performance a bit. These findings indicate that the incorporation of social media data can be a very promising way to improve the tasks of startup services.

History Window Size	Systems	MRR	P@1	P@3
NULL	P+UniformG	0.457	0.261	0.544
	Rel.	0.353	0.157	0.404
≤ 3 question threads	(B)aseline	0.474	0.263	0.589
	P+HisG	0.498	0.303	0.604
	B.+Weibo vs. B.	0.516 +8.86%***	0.323 +22.81%**	0.624 +5.94%**
≤ 5 question threads	(B)aseline	0.478	0.271	0.590
	P+HisG	0.501	0.303	0.613
	B.+Weibo vs. B.	0.521 +9.00%***	0.327 +20.66%***	0.627 +6.27%***
≤ 10 question threads	(B)aseline	0.494	0.286	0.612
	P+HisG	0.514	0.316	0.627
	B.+Weibo vs. B.	0.530 +7.29%	0.332 +16.08%	0.643 +5.07%

Table 4: Best answer prediction performance with varying history window sizes. “*”, “***”, “****” indicate the improvement is significant at the level of 0.1, 0.05 and 0.01 respectively.

6 Related Work

Our task is built on community question and answering site and researchers have studied CQA from many perspectives. One perspective focuses on user expertise estimation. Generally, there are two principle methods for expertise ranking, interaction graph analysis and interest modeling. Interaction graph based methods (Jurczyk and Agichtein, 2007; Zhang et al., 2007) construct a graph using interaction (e.g., asking and answering) behavior, and rank users using some generalization of PageRank (Brin and Page, 1998) or HITS (Kleinberg, 1999). Interest modeling methods characterize users' interests using question category (Guo et al., 2008) or latent topic modeling (Liu et al., 2005). There are also methods that combine both interest modeling and graph structure (Zhou et al., 2012; Yang et al., 2013) to rank users. Another research perspective on question answering service is quality prediction including answer quality prediction (Harper et al., 2008; Shah and Pomerantz, 2010; Severyn and Moschitti, 2012; Severyn et al., 2013) and question quality prediction (Anderson et al., 2012). However, since the methods mentioned above are based on the history data, the system will experience the cold start problem. Our work explore to what extent can external features help relieve the problem.

This work is also concerned with mining across heterogeneous social networks. Recently, many researches focus on mapping accounts from different sites to one single identity (Zafarani and Liu, 2013; Liu et al., 2013; Kong et al., 2013). By utilizing these recent studies on linking users across communities, our work can be extended to larger scale datasets. From another perspective, cross-domain recommendation has also been widely studied. Zhang et al. (Zhang and Pennacchiotti, 2013a; Zhang and Pennacchiotti, 2013b) explore how Facebook profiles can help boost product recommendation on e-commerce site. Previous work (Zhang et al., 2014) analyze user novelty seeking traits on social network and e-commerce site, which can be used to personalized recommendation and targeted advertisement. Different from simply borrowing user's profiles or psychological traits, our work integrates user footprints from heterogenous social networks and captures performance related characteristics more precisely.

7 Conclusion

In this paper, we take the initiative attempt to leverage social media knowledge for improving the social startup service. We carefully construct a dataset of 20,742 users who have been linked across Zhihu and Weibo, which are social startup and external social media websites respectively. We hypothesize that a user with higher prestige and more relevant Weibo contents to a question is more likely to have better performance.

We first carefully construct testing experiments for these two hypotheses. Our results indicate that prestige in Weibo has strong correlation with overall performance in Zhihu. For question specific performance, we have found that relevance between questions and a user's tweets also correlates with user performance on Zhihu. Based on these findings, we further add prestige and relevance knowledge into existing user performance prediction framework. The experiment results show that prestige and relevance information in Weibo largely improve the performance when the available training data is not sufficient. Moreover, individual prestige feature achieves very competitive results. Our approach is equally applicable to other knowledge sharing web services with appropriate external social media information.

Acknowledgements

The authors would like to thank the anonymous reviewers for their comments. This work is supported by the National Grand Fundamental Research 973 Program of China under Grant No.2014CB340405 and the National Natural Science Foundation of China (Grant No.61170056). The contact author is Zhen Xiao.

References

Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2012. Discovering value from community activity on focused question answering sites: a case study of stack overflow. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 850–858. ACM.

- Jiang Bian, Yandong Liu, Ding Zhou, Eugene Agichtein, and Hongyuan Zha. 2009. Learning to recognize reliable users and content in social media with coupled mutual reinforcement. In *Proceedings of the 18th international conference on World wide web*, pages 51–60. ACM.
- Sergey Brin and Lawrence Page. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1):107–117.
- Hongbo Deng, Jiawei Han, Michael R Lyu, and Irwin King. 2012. Modeling and exploiting heterogeneous bibliographic networks for expertise ranking. In *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries*, pages 71–80. ACM.
- David Easley and Jon Kleinberg. 2012. Networks, crowds, and markets: Reasoning about a highly connected world.
- 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*, pages 921–930. ACM.
- F Maxwell Harper, Daphne Raban, Sheizaf Rafaeli, and Joseph A Konstan. 2008. Predictors of answer quality in online q&a sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 865–874. ACM.
- Jiwoon Jeon, W Bruce Croft, Joon Ho Lee, and Soyeon Park. 2006. A framework to predict the quality of answers with non-textual features. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 228–235. ACM.
- Pawel Jurczyk and Eugene Agichtein. 2007. Discovering authorities in question answer communities by using link analysis. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 919–922. ACM.
- Jon M Kleinberg. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632.
- Xiangnan Kong, Jiawei Zhang, and Philip S Yu. 2013. Inferring anchor links across multiple heterogeneous social networks. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 179–188. ACM.
- Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. 2010. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600. ACM.
- Xiaoyong Liu, W Bruce Croft, and Matthew Koll. 2005. Finding experts in community-based question-answering services. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 315–316. ACM.
- Jing Liu, Young-In Song, and Chin-Yew Lin. 2011. Competition-based user expertise score estimation. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 425–434. ACM.
- Jing Liu, Fan Zhang, Xinying Song, Young-In Song, Chin-Yew Lin, and Hsiao-Wuen Hon. 2013. What’s in a name?: An unsupervised approach to link users across communities. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, WSDM ’13*.
- Tie-Yan Liu. 2009. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval*, 3(3):225–331.
- Kevin Kyung Nam, Mark S Ackerman, and Lada A Adamic. 2009. Questions in, knowledge in?: a study of naver’s question answering community. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 779–788. ACM.
- Julian Oldmeadow, Michael Platow, Margaret Foddy, and Donna Anderson. 2003. Self-categorization, status, and social influence. *Social Psychology Quarterly*, 66(2):138–152.
- Aliaksei Severyn and Alessandro Moschitti. 2012. Structural relationships for large-scale learning of answer re-ranking. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 741–750. ACM.
- Aliaksei Severyn, Massimo Nicosia, and Alessandro Moschitti. 2013. Learning adaptable patterns for passage reranking. *CoNLL-2013*, page 75.

- Chirag Shah and Jefferey Pomerantz. 2010. Evaluating and predicting answer quality in community qa. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 411–418. ACM.
- Young-In Song, Jing Liu, Tetsuya Sakai, Xin-Jing Wang, Guwen Feng, Yunbo Cao, Hisami Suzuki, and Chin-Yew Lin. 2010. Microsoft research asia with redmond at the ntcir-8 community qa pilot task. In *NTCIR-8*.
- Juyup Sung, Jae-Gil Lee, and Uichin Lee. 2013. Booming up the long tails: Discovering potentially contributive users in community-based question answering services.
- Rui Yan, Mirella Lapata, and Xiaoming Li. 2012. Tweet recommendation with graph co-ranking. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, pages 516–525. Association for Computational Linguistics.
- Liu Yang, Minghui Qiu, Swapna Gottipati, Feida Zhu, Jing Jiang, Huiping Sun, and Zhong Chen. 2013. Cqarank: jointly model topics and expertise in community question answering. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 99–108. ACM.
- Reza Zafarani and Huan Liu. 2013. Connecting users across social media sites: a behavioral-modeling approach. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 41–49. ACM.
- ChengXiang Zhai. 2008. Statistical language models for information retrieval. *Synthesis Lectures on Human Language Technologies*, pages 1–141.
- Yongzheng Zhang and Marco Pennacchiotti. 2013a. Predicting purchase behaviors from social media. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1521–1532. International World Wide Web Conferences Steering Committee.
- Yongzheng Zhang and Marco Pennacchiotti. 2013b. Recommending branded products from social media. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 77–84. ACM.
- Jun Zhang, Mark S Ackerman, and Lada Adamic. 2007. Expertise networks in online communities: structure and algorithms. In *Proceedings of the 16th international conference on World Wide Web*, pages 221–230. ACM.
- Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, and Xing Xie. 2014. Mining novelty-seeking trait across heterogeneous domains. In *Proceedings of the 23rd international conference on World wide web*, pages 373–384. International World Wide Web Conferences Steering Committee.
- Guangyou Zhou, Siwei Lai, Kang Liu, and Jun Zhao. 2012. Topic-sensitive probabilistic model for expert finding in question answer communities. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 1662–1666. ACM.