Influencing Factors Analysis of People's Answering Behaviours on Social Network Based Questions

Zhiwei Sun[1], Wenge Rong[1,2], Yikang Shen[3], Yuanxin Ouyang[1,2], Chao Li[1,2], Zhang Xiong[1,2]

[1]School of Computer Science and Engineering, Beihang University, Beijing 100191, China
[2]Research Institute of Beihang University in Shenzhen, Shenzhen 518057, China
[3]Sino-French Engineer School, Beihang University, Beijing 100191, China
{sunzhiwei, w.rong, yikang.shen, oyyx, licc, xiongz}@buaa.edu.cn

Abstract—Recently the social question/answering systems have become a natural way for people to get information from their social connections. Despite its popularity in knowledge sharing, its development still faces several challenges, among which a notable one is that a majority of questions get no response. How to improve the response ratio and reduce the number of unanswered social questions has become a challenging problem. Though some research in this context has been well studied, intrinsic factors that reveal profiles of social answerers are still less understood. In this paper, we try to discover: what kind of person is likely to be an answerer? To answer this question, we study multiple hypothetical factors from the perspectives of social and personal analysis. A set of experiments are conducted to identify the relationship between these factors and the action of answering. We present a comprehensive description alongside our analysis of behavioural patterns of answerers in the scenario of social information seeking.

Keywords-Social Network; Question/Answering; Factor Analysis; Answering Behaviour

I. INTRODUCTION

The explosive development of social networking services (SNS), such as Twitter and Facebook, make social networks the new channel for information exchanges between individuals. Individuals can easily publish the latest news or describe their current status, emotion, and/or other information, while at the same time their friends can also join the discussion by responding freely. The unique structure of social networks makes it possible to build individual social relationships [1], thereby making social media an appropriate platform to assist people for information seeking. A survey by Morris et al. illustrates that 50.6% of participants had at some point asked questions through a social network [2]. The statistics report from Efron and Winge also reveals that 13% of tweets on Twitter are questions [3].

Social question/answering (SQA) presents a new style of information seeking [4], and has become complementary to traditional search platforms and community based question/answering sites. It is a nature way for people to get answers from social media when they have questions [5], because they believe that their friends are familiar with their personal background to the question and therefore can provide more precise answers to fulfil their information needs [6]. Even though SQA platform enhance social interactions, their development also meet several challenges and one in particular

is that a majority of questions get no responses [7]. As such how to improve the response ratio then becomes a challenging problem for investigation. Several studies have been carried out to explore the challenges facing SQA and typical examples include what kind of questions are more likely to get answers and what extrinsic factors [8]; how to utilise a historic question archive to help answer a new questions [9]?

An alternative method is to explore this problem from the answerer's perspective. Research suggests that a lot of questions are missed by potential answerers due to the flood of social interaction [7]. As a result, automatic recommendation of potential question answerers has been considered an efficient way to enhance the question response rate [10]. It can be seen that from an answerer's perspective, although starting from different points, there are always two fundamental factors: who can answer the question? Who is willing to answer? However, a comprehensive investigation to identify the factors and their influence on people's response to social question has yet to be undertaken. In this work, we limit the scope of our study to discovering what kind of people would likely to be answerers. A core element in SQA is the correct interpretation of people's motivation to answering.

In this paper, we try to comprehensively investigate the answerer behaviour in the context of social media. We explore what drives users to provide answers; what is the purpose of the answering behaviour, for sharing knowledge, to help others, or something else? To evaluate the relevance of the explored factors, we conduct a set of experiments to identify which factors influence the response received.

The rest of the paper is organised as following. Section 2 gives an overview of related work in social question answering. In Section 3, we introduce the explored factors and related hypotheses. Section 4 illustrates the method for features selection and present our experimental study on a Weibo dataset, lessons learned from the experimental study are also discussed. Finally, Section 5 gives a conclusion and outlines possible future works.

II. RELATED WORK

SQA is a special kind of social information seeking mechanism. It has received much attention for exploration. SQA behaviour consists of four important parts: social questions,

IEEE computer society

social answers, question askers and question answerers. Several research problems in this context have been conducted and well studied from these four different angles.

To better understand SQA behaviour, researchers have tried to get an understanding of what kind of questions will be asked and/or answered in a social network. Morris et al. carried out a survey involved 624 participants. By analyzing participants' question habits, Morris found that several factors could provide great insight into the questions asked in SQA, such as how the questions are phrased, what kind and type of question is most likely to be asked [2].

From an answer's perspective, to reduce the number of unanswered questions, one possible solution is to utilise past answers in the archives [9]. This inspiration has been widely used in community based question/answering systems. One kind of intuitive approach is to find previously published possible matching question/answer pairs regard to the new questions [11]. An alternate method is to automatically generate answers by referring to existing knowledge from a historic archive and/or other online resources [12].

Some researchers have studied SQA from the viewpoint of questioners and past investigation on this interesting topic has obtained some insights. In the study by Morris et al., participants are presented with several questions to explain their motivation for asking questions in a social network [2]. They found that the most common reason to search socially, rather than through a search engine, is that participants had more trust in the responses provided by their friends.

Besides the three above mentioned directions, another important perspective to increase the answer rate in SQA, is proactive identification of possible answerers. Several efforts have been applied to identify users who have the capability to answer the questions and a typical example is the social search engine Ardvark that can automatically find the right person to satisfy a questioner's information needs [13].

Research into SQA has attracted much attention, but certain topics in this domain are still emerging areas for exploration. In this paper, we try to understand the profile of social answerers and present an empirical evaluation to reveal the factors that may influence the response behaviour in SQA.

III. ATTRIBUTES

In this section we investigate the factors which can depict the profile of people who has a higher tendency to answer social questions. In Weibo, when one user repost a message from others, he can choose whether to add the repost-content to the message comment list. We define "answer" behaviour in our paper: reply to the question and the repost behaviour which agree the adding. To this end, we propose ten factors and related hypothesis on the impact of a potential answerer. The factors cover not only features of people's social connections, but also general characters of personal information.

1) Follower of questioner

Generally in social media, a follower is notified of and may view any posts made by accounts they follow [14]. Most social media posts are public by default, and can be found by viewing a public feed, or searching, but only those users who are following the poster will be notified of their posts [15]. Therefore, given a question posted by user A, we can assume that the users who are followers of A would be more likely to answer A's question and propose the first hypothesis.

Hypothesis 1: People who follow the questioner are more likely to respond to the social questions..

2) Friendship between different answerers

In the real society, there is a very common phenomenon that when our friends attend a topic discussion, we may also have interest in following friends and joining the event [16]. So we wonder that in the social network, would it also affects our motivation of answering behaviours in SQA that if our friends are active in answering social questions?

Hypothesis 2: The friend's behaviour of answering question will have positive affect users to also answer social questions.

3) Prior Interaction with questioner

The friendship between answerer and questioner is defined based on their bidirectional follow relationship. It just shows users' interest of knowing each other, can't reflect the closeness. But the social interaction such as @-mentions, post and reply behaviours [17] can. The more times users interact with questioner before, the more familiar they are. Compared to strangers, we assume users may be more likely to answer the questions posted by "acquaintance".

Hypothesis 3: Users who have frequent interactions with questioner are more likely to respond social question.

4) Number of followers

The number of followers can to some degree represent a person's influence on SNS [18]. People with a large number of followers on SNS pay great attention to their public image. Of all the possible ways to improve their influence, helping others is a good way to obtain acknowledgement and respect from the society, such as "weiwen" a Weibo account that often helps others and has more than 500,000 followers.

Hypothesis 4: People with more number of followers are more likely to answer social question posted by others.

5) Frequency of posts

The posting behaviour, which mainly just means updating the status in this paper, is of great importance in social network. The number of messages post on social media is considered as an significant indicator to measure the activity of a Weibo user [11]. People heavily involved in SNS show a greater sense of participation willingness when meeting questions. This simple and intuitive arguments lead to the following hypothesis.

Hypothesis 5: People with higher posting frequency are more likely to respond social questions.

6) At-mentions

The people being mentioned in social network could be notified of the post directly. At-mentions play an important role for information diffusion and social communication [19]. But in our study, at-mentions are considered another important indicator that can describe the interaction between friends. In a question, if one answer could not solve the problem in his ability, he may @ some friends for help.

Hypothesis 6: Users with frequent at-mentions behaviour are more likely to respond social question.

7) Account creation-time

People who actively use SNS as a communication tool for a long time would get used to it, and are more likely to spend more time on it. The age of an account may correlate with higher social rank for information [20] and we wonder if senior users would be more positive to answer questions because they are more experienced.

Hypothesis 7: People who are long term SNS users are more likely to respond to social questions.

8) Description

Personal description is a short section of text created by Weibo users freely to introduce themselves. Those users who complete a personal description are more likely to spend more time and effort in SNS communication. Furthermore, the hobbies and other key words listed in the description represent skills and topics in which they are proficient. As such they are likely to answer questions on topics that fall in these or similar domains [21].

Hypothesis 8: People with personal description are more likely to respond social questions.

9) Gender difference from questioner

One interesting that holds in both the real world and on SNS is that people are more interested in pictures or topics from people of different gender [22]. It can be imagined that under the context of social question and answering, a quick answer to questions posted by fiends of different gender is an efficient way to display a user's talent and further gain attention.

Hypothesis 9: People are more likely to respond social questions posted by fiends of a different gender.

10) Geographical differences

SNS provide a platform where people from different areas can easily publish information and join a discussion with others since the information is universally shared. The variety of culture and economic development in different areas will inevitably lead to the different levels of exposure to social media [23]. We assume that users in one province are same-area because they may have some common living habits.

Hypothesis 10: People who are located close to the questioner are more likely to respond social questions.

IV. EXPERIMENT

A. Date Set

In this study, we perform experiments on a data set from Sina Weibo[1], which is the largest social network in China [24]. Just like Facebook, Twitter, and other SNS applications, it also provides users various ways to interact with others [25]. For example, users can read recent posts, reply, and repost, and also use the symbol '@' to mention their friends.

Using Weibo search API, we collected 1,066,129 Weibo messages that were posted between May 2013 and Aug 2013. In this process, the way we identify the questions from the normal messages is similar to the method employed in [8],

[1] http://weibo.com/

which used several key interrogative words such as "where"[2], "who knows", "what", "Is there" and the question mark '?'. A simple test was carried out to check the questions we collected by this way and results showed that about 85% of the questions are real ones.Finally, we obtained 88,713 questions, of which 20.8% (18,475) received more than one answer. The median number of answers for all the questions is just 1.19, this increases to 5.72 if we removed the unanswered questions for the convenience of our experiment. In calculating median value celebrity or public questions contributed greatly because they generate many replies and most ordinary questions only get one or two answers.

We aim to discover the most important factor in inciting replies by analysing this data. As such, in this paper, the underlying association of different hypothesis are studied by using Principal Components Analysis (PCA). The exact degree of how much the answerer is influenced by Multiple Regression Model (MRM) is also further studied.

B. Principal Components Analysis

PCA is a widely used multivariate method that is able to analyse a data table in which observations are described by several inter-correlated quantitative dependent variables [26]. This technique is employed in this research attempting to reveal the internal structure of all of the ten features defined in previous section and then reduce them to a smaller number of principal components. In this research, the factors are extracted by PCA and rotated by varimax, afterwards we will further calculate factor loading and the score of common factors.

The KMO and Bartlett's Test show that the value of Kaiser-Meyer-Olin Measure of Sampling Adequacy of collected data is 0.859 (>0.5), which means that PCA could be able to find potential correlations between variables. To identify the right number of main components, two rules are often used in the literature. Kaiser recommends retaining factors with an eigenvalue greater than 1, while Cattell uses factorial scree to find the smoothing point in the graph and decides the keeping results [14]. Taking these two rules in consideration, both Table I and Fig. 1 suggest to reduce the factors to 3 principal components which account for 59.83 % of all the variances.
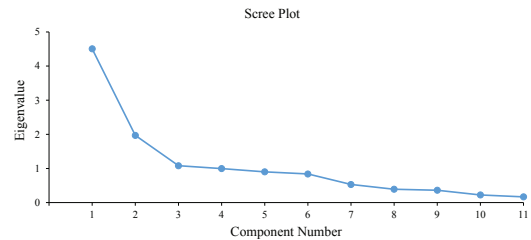

Fig. 1. Plot

TABLE I
TOTAL VARIANCE EXPLANATION

Component	Eigenvalues	% of Variance	Cumulative %
Factor 1	3.52	31.99	31.99
Factor 2	1.98	17.96	49.94
Factor 3	1.09	9.89	59.83

Table I describes the information of all of the three new factors extracted by PCA. The "Eigenvalues" column represents the weight of the factors and "% of variance" shows the contribution to the total variance of all variable. The greater the value is, the more significant the factor is.

TABLE II
ROTATED COMPONENT MATRIX

	Component		
	Factor 1	Factor 2	Factor 3
ReplyCount	0.01	0.69	0.31
H1:Followship	0.61	0.60	-0.12
H2:Answerer-friendship	0.78	0.10	-0.12
H3:Prior-interaction	0.79	0.25	0.01
H4:FollowNumber	-0.20	0.03	0.69
H5:Post-frequncey	-0.03	0.04	0.24
H6:At-mentions	0.83	0.11	0.01
H7:Account-creation	-0.43	0.27	-0.59
H8:Description	0.58	0.68	-0.05
H9:Same-sex	0.45	0.63	-0.16
H10:Same-area	0.68	0.36	-0.22

The coefficient in Table II represents the detail of the three factors that is the linear combination of the ten features. In order to explain the main reflection of each factor, a threshold value of 0.5 is set to eliminate non-significant features. As a result, Hypothesis 5, the affect of post-frequency should be rejected since its impact is too low with coefficients smaller than 0.5 in all of the three new factors.

Factor 1 is the most important principle component and describes the data from two aspects. The first is from the perspective of relationships between questioners and answerers and includes the follow, and friendship with questioners. Another aspect, which contains the prior interaction and the at-mention behaviours, reflects one's attitude towards social answering to some extent for both of them are subjective actions. Particularly, when users have prior interaction with each other, either intentionally or unintentionally, a positive attitude has some influence on the SQA behaviour.

Factors 2 and 3 mainly focus on the following features: gender difference, personal description, number of followers, and account age. These features are able to describe the influence of users' inherent attributes on their answering behaviour. Usually, the account age reflects one's qualifications on SNS and the average number of posts the questioner makes everyday reflects the user's liveness. But some special cases are not consistent with these rules. A typical example is British Prime Minister David Cameron's Weibo, which only posted 23 Weibo since account was created, but one of his questions has received nearly two thousand replies and answers.
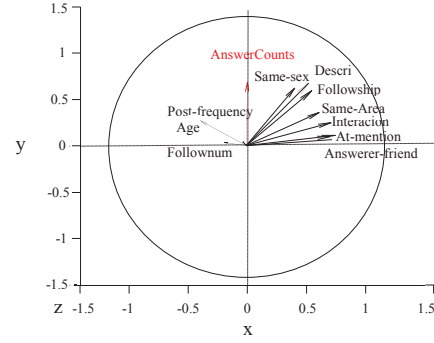

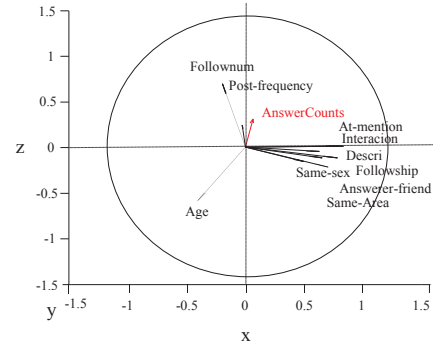Fig. 2. PCA factor map with Factor1(x-axis), Factor2(y-axis)

Fig. 3. PCA factor map with Factor1(x-axis), Factor3(y-axis)

In Figs. 2 and 3, each feature in Table I corresponds to a vector, and the vector's coordinates are the correlation between different factors. For example, the vector representing account creation-time in Fig. 2 is $(-0.43, 0.27)$, where x-axis value is the contribution to factor 1 and y-axis is factor 2. The directions of all the vectors reveal the various associations between the proposed hypotheses and answerer behaviours in SQA. Generally speaking, there should be three figures for the different combinations of the three principle component. But after studying Table II carefully, it is observed that all the contributions of each feature in factor 2 are positive numbers. As such the main difference is mainly in the coefficient of Factors 1 and 3. In Fig. 3, "Description", "Same sex", "Followship" are in the same quadrant as the vector "Reply Count", and the weight in the y-dimension is quite big, suggesting that users' answering behaviour is positively affected by these features. At the same time, Fig. 3 shows that the account creation-time has a negative influence to the reply counts, as it has the opposite direction of "Reply count".

C. Multiple Regression Model

The PCA Model is able to find potential correlation between different features. But sometimes the result of PCA is not exact, thereby making it necessary to have more detailed

informations. Multiple Regression Model offers further detailed information and is a widely used technique to reveal quantitative relationship. Before applying the Multiple Regression Model, we use bivariate correlations method to examine the multicollinearity of data. The checking is often adopted among three or more variables. The p-value, which inspects the probability of error of first kind in conclusion, could decide whether the result of Multiple Regression Model is reliable. The result of applying Multiple Regression Model is listed in Table III, which shows that the data is statistical significant and no obvious multicollinearity among independent variables, for the p-value is less than 0.05 [27].

TABLE III
MULTIPLE REGRESSION RESULT

Features	Coefficient	p-value
(Constant)	-2.349	0.019
H1:Followship	7.933	0.000
H2:Answerer-friendship	-6.479	0.000
H3:Prior-interaction	12.511	0.000
H4:FollowNumber	-0.799	0.424
H5:Post-frequncey	-0.101	0.919
H6:At-mentions	0.983	0.326
H7:Account-creation	-2.371	0.018
H8:Description	14.203	0.000
H9:Same-sex	4.019	0.000
H10:Same-area	-12.032	0.000

Normally the multiple features of social answering behaviour are not a simple linear model. They can be affected by various hidden factors. As such the fit of the Multiple Regression Model usually would not be very high. Our experimental study shows that all of the ten proposed predicted variables account for 27% the total variance which is much higher than 19% in similar research on social question attributes [8].

After studying the Multiple Regression Model result, we decide to remove those features that have no significant effect, i.e., their p-value is greater than 0.05. This means removing features 4 (Followers), 5 (Frequency), and 6 (At-mentions), indicating that for our Weibo dataset these had no significant impact on user enthusiasm to answer social questions. Among the three removed features, the insignificant affect of number of followers and posting-frequency have also been similarly revealed by previous studies [14], while they think larger number of follower and suitable post-frequency would do good to message spreading. "Number of followers" is also an important feature for questioner being able to spread the question [2]. But it has little effect on motivating user's answering willingness. What is more, the elimination of hypothesis 6 (At-mentions) is unexpected but reasonable, since we are trying to discover what attributes make a user more enthusiastic to answer. No matter we are willing to answer or not, we would judge whether we could answer the question by our knowledge subconsciously, when we see a question at first. Only when we felt the question is interesting enough or beyond our capability, we would consider using at-mention to inform our friends. Compared with the fact that at-mention function is usually used in Weibo content, the number of at-mention

when user reply other Weibo is relatively less. Therefore it is acceptable that at-mention behaviour is not significant with users' answering passion.

After combining the experimental results of PCA and MRM, we can come to a general conclusion. Hypothesis 1 (Followship with questioner), 3 (Prior interaction) and 8 (Personal description) are acceptable and do have positive influence to users' answering. Although Hypothesis 2 (Friendship between answerers), 7 (Account creation-time), 9 (Gender difference) and 10 (Geography difference) are also significantly correlated with the reply numbers, though negatively.

D. Features impact the answer behaviour

Though we have known several features which may lead a user to answer the question in social network. However, after looking into the dataset retrieved from Weibo (which have 88,713 Weibo and only 18,475 questions got answers with number of 105,748), it is also found that the data presentation varies in a large range. To better understanding the detailed influence of each extracted feature, this section will present in-depth analysis on the dataset.
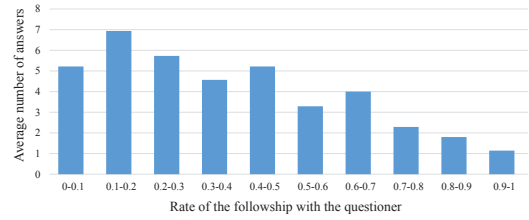

Fig. 4. Followship with questioner

Followship with questioner. Generally, questions posted by users with many followers get more replies. But the question could be replied by anyone, as such we do not know how many replies are contributed by the follower of the questioner. In order to find out whether the follow relationship between answerers and questioner impacts the user's answer behaviour, we carry out the following analysis. For each question, we count the number of answers who follows the questioner and then divided it by the total number of unique answerers. In this way, we get a ratio of followship of each question. By grouping all the questions with different ratio, as shown in Fig. 4, it is found that with the ratio of the followship increased, the number of answers decreased slowly. But we do not think it means that the follow relationship between questioner and answerers would do bad to the answering behaviour. Questions with small number of answers had relatively high rate of followship, while questions with large numbers of answers had low rate. It seems reasonable, since it is much easier to have a high "rate of followship" among a small group of answerers than a large group.

Prior interaction. In all collected data, 78% (82,861) answers are contributed by strangers, which means that there are no interaction between answerers and questioner before. It seems that this factor plays a weak role and has no influence

on reply counts. But within the answers who have prior interaction, Fig. 5 tells that the prior interaction would have obvious promotion to user's answering willingness. Although vast majority of answers on SNS are indeed posted by the random users, which implies that the probability of a question seen by massive strangers is much higher than friends, it does not conflict with our conclusions. For each question, we count the total number of prior interactions between the questioner and answerers and group all the questions by the number. Those no-prior interaction questions only get one or two answers, which is far less than average 7.6 in the prior-interaction aware ones. Despite of the possible existing friendship between questioner and answerer, the more times they interacted, no matter in this question or some other Weibo status as long as in the social network of Weibo community, the more answers the questioner may get. Users who already established relationship with the questioner would be willing to join the question discussion again.
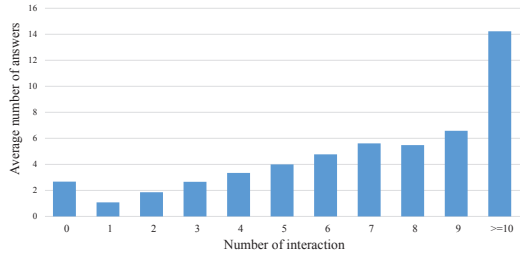

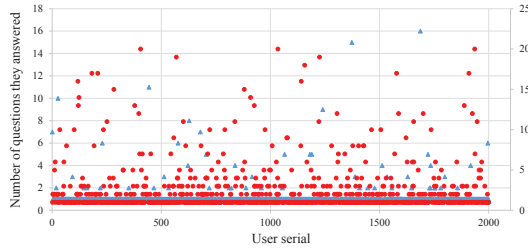Fig. 5.   Prior interaction with questioner

Fig. 6.   Prior interaction history

In Fig. 6, x-axis is the serial number of users, and the y-axis is the times of questions each user attended. We exact 2,000 users randomly from the dataset and group them by the times of prior interaction. The blue triangles represent the users who did not communicate with questioners, while the red pots indicate the users who have prior interactions. It is observed that the distribution of red ones shows that prior interaction behaviour does have positive influence on the enthusiasm of users answering.

Description. The personal description in some ways could reflects a user's attitude on his own images in social networks. In hypothesis 8, it is assumed that users who have no descriptions would not positively show their hand to the help

information seeking in SNS. Because there are high probability that users would not be positive or enthusiastic to the social network if they even do not care the image of themselves. For each question, we get the ratio of description by accounting the answerers who have description and divided by the total answerer number. As shown in Fig. 7, the result verifies the finding in the multiple regression analysis. The user who has description tends to be more positive and more likely to play a active role in social question/answering system. For those "responsible users", the probability of sharing their knowledge and experience is relatively higher than ordinaries.
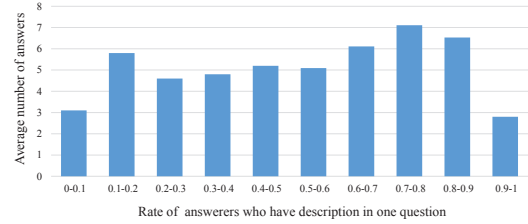

Fig. 7.   Description

Friendship between answerers. Friendship between users is a typical research hotspot and of great significance in helping the spreading of social question/answering system. Many studies focused on the friendship between questioner and answerer [10], but paid less attention on the relationship between answerers. Actually, social conformity may also play an important role in the social network. In this paper, we define that two user are friends if they are follower of each other and study how user's answering behaviour is affected by friends' behaviours on SNS. We count the number of friendship between different answerers and divided by the combination number of two answerers. For example, question Q receive three answers from user A, B, C and only A, B are friends. So, the number of friendship in question Q is 1. The combinations of two different answerers are AB, AC, BC and the combination number of two answerers is 3. Divided friendship number by the combination number, we could calculate the ratio of friendship in the question. By grouping all the questions with different ratio, the result is shown in Fig. 8. The number of answers in the 0-0.1 interval is much less than others. This is because most of the questions could just get a few answers, and be seen by anyone on the social network randomly. In most cases, there would be small probability existing friendship among answerers if the group is small. As we can see the trend of other intervals in Fig. 8, the average number of answers per question decreased with the rate of friendship increases. It seems that the friends' attitude on the social network would have some negative impact the user's enthusiasm to the answering behaviour.

Accounts creation-time. To better understand the influence of time user stayed on SNS, we conduct an experiment which counts the average account-age of the answerers for each question and group them by the average account-age. Grouped by year in x-axis and y-axis representing the average number
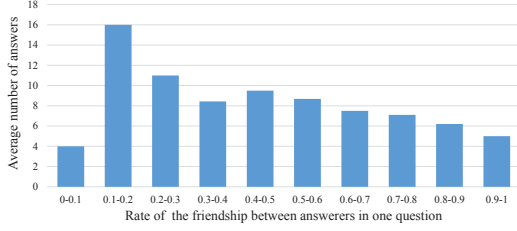
Fig. 8.   Friendship between answerers

of answers they contribute, we plotted a chart which reflects the relationship between the time users stayed on SNS and their answering behaviours. As shown in Fig. 9, although the longer users stayed on SNS, the more stable and experienced their social behaviour tend to be, the enthusiasm of join social question and discussion will degrade.
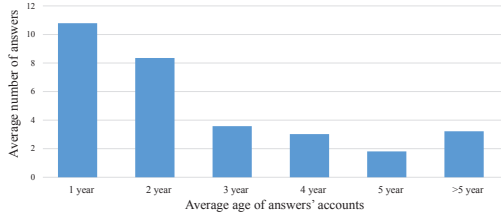

Fig. 9.   Average age of account

Same-sex answerers. Could the difference of genders impact his or her answer behaviours in SNS? This experimental study leads to a very interesting result, which did not verify the proposed assumption in the past. In the question/answering dataset, 30.4% (32,143) answers are contributed by opposite-sex answerers and 69.6% (73,605) are posted by same-sex answerers. Different from the previous study where opposite-sex users seem to get more special attention on SNS. As illustrated in the Fig. 10, the five intervals between 0 and 0.5 all get a greater average number of answers, this is because only a few questions belongs to these intervals and be easily influenced by some special cases. Most answers are contributed by same-sex users and we could see a growth with the rate of same-sex answers increase in the intervals between 0.5 and 1. The curve which represents the affection of gender on answer number reveals that opposite-sex relationship between questioner and answerers do not make users be more positive.
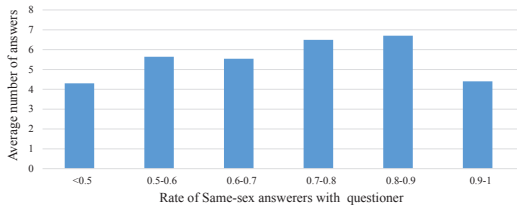

Fig. 10.   Same-sex with questioner

Geography difference. Users in Weibo are across all over the world and then the cultures vary between different areas. As shown in Fig. 11, when the rate of answerers who are in the same area with questioners increases, the less attention the question would get. On the one hand, this may due to the fact that on the Internet the distance in reality between social network users makes no sense. On the other hand, we may have more interest in some topics discussion held by users from different regions, since their culture might be new to us. Furthermore, we studied the economy status in affecting user's social answering behaviour. By using the information of geographic location in longitude, we group answerers into categories named Eastern China, Middle of China and the Western China (in terms of economy development). Fig. 12 is the contribution of different areas to all the answers and shows that the performance of users in relatively developed Eastern China are much more active in social answering.
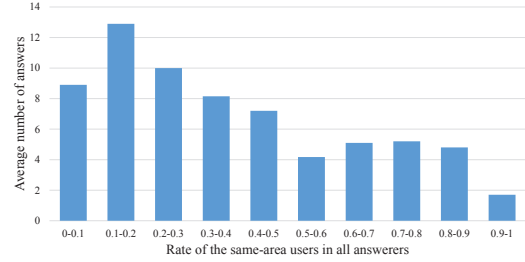

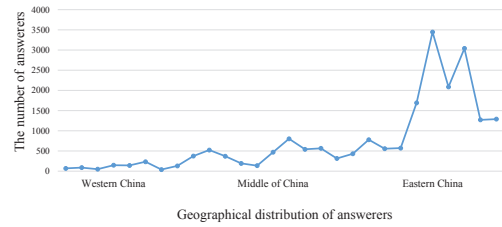Fig. 11.   Same-area with questioner

Fig. 12.   Answerer distribution in China

V. CONCLUSION

In this paper, we present a comprehensive study to answer such questions: what kind of people contribute mostly to answer the questions posted on social platform. Why is it compelling to understand the answerer profile in social context? It attributes to the big challenge that majority of social questions got no responses. Social information seeking represents the new search model in future. The solutions helping to improve the response ratio and reduce the amount of unanswered social questions is quite important. A lot of methods are therefore proposed from various perspectives, such as social question character exploration, social question recommendation, and etc. Different from these existing methods, our study is implemented with the focus on answerer. The target is to provide a complete image to characterise social answerer.

By analysing the data from Weibo, we try to find out the relationship between ten extrinsic features and answering willingness. These features cover not only the social relationship between SNS users, e.g., followship, friendship, social interaction and etc., but also user's personal characters, e.g., posting number, geographical information and etc. We further conduct a series of principal component analysis and multiple regression analysis. Some interesting and unexpected facts are revealed by the experiment. Some features such as number of followers, friendship between answerers, at-mention behaviour play significant roles in social question spreading, while we do not find strong evidence to support that they could significantly impact one's social answering willingness. The features such as followship with questioner, description, prior interaction, and same-sex all have positive impaction on the answering and will influence user's enthusiasm to answering behaviour positively. But different from the positive effect to the retweet [14], the same-area, the age user staying on the SNS have negative influence to the user's answering.

Although the study reveals some features that affect users' social answering behaviours, there are some limitations in current work. All the question data are extracted from Weibo by keywords, which is a relatively simple and convenient way to quickly distinguish questions from all the ordinary Weibo messages. However, keyword filter method would also possibly miss lots of questions which do not contain these limited interrogative words. Although the ten features extracted in this study covered most of user attributes, some other hidden features could not be easily described. For example, the background of knowledge, the mood they meet the question, all these may lead an entirely different answer behaviour. How to perfect the way of extracting question and hidden features could be possible directions of later work.

REFERENCES

[1] R. Kumar, J. Novak, and A. Tomkins, "Structure and evolution of online social networks," in *Link mining: models, algorithms, and applications*, 2010, pp. 337–357.

[2] M. R. Morris, J. Teevan, and K. Panovich, "What do people ask their social networks, and why?: a survey study of status message q&a behavior," in *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, 2010, pp. 1739–1748.

[3] M. Efron and M. Winget, "Questions are content: A taxonomy of questions in a microblogging environment," in *Proceedings of 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem - Volume 47*, 2010, pp. 27:1–27:10.

[4] X. Tu, D. Feng, X.-J. Wang, and L. Zhang, "Analogical reasoning for answer ranking in social question answering," *IEEE Intelligent Systems*, vol. 27, no. 5, pp. 28–35, 2012.

[5] S. A. Paul, L. Hong, and E. H. Chi, "Is twitter a good place for asking questions? a characterization study," in *Proceedings of 5th International AAAI Conference on Weblogs and Social Media*, 2011, pp. 578–581.

[6] M. R. Morris, J. Teevan, and K. Panovich, "A comparison of information seeking using search engines and social networks," in *Proceedings of 4th International AAAI Conference on Weblogs and Social Media*, 2010, pp. 23–26.

[7] Sysomos Inc., "Replies and Retweets on Twitter," http://www.sysomos.com/insidetwitter/engagement/, 2010, [Online; accessed Sept-2010].

[8] Z. Liu and B. J. Jansen, "Factors influencing the response rate in social question and answering behavior," in *Proceedings of ACM 2013 Conference on Computer Supported Cooperative work*, 2013, pp. 1263–1274.

[9] A. Shtok, G. Dror, Y. Maarek, and I. Szpektor, "Learning from the past: answering new questions with past answers," in *Proceedings of 21st International Conference on World Wide Web*, 2012, pp. 759–768.

[10] K. Panovich, R. Miller, and D. Karger, "Tie strength in question & answer on social network sites," in *Proceedings of ACM 2012 Conference on Computer Supported Cooperative Work*, 2012, pp. 1057–1066.

[11] C. Sibona and S. Walczak, "Unfriending on facebook: Friend request and online/offline behavior analysis," in *Proceedings of 44th Hawaii International Conference on Systems Science*, 2011, pp. 1–10.

[12] E. Mendes Rodrigues and N. Milic-Frayling, "Socializing or knowledge sharing?: characterizing social intent in community question answering," in *Proceedings of 18th ACM Conference on Information and Knowledge Management*, 2009, pp. 1127–1136.

[13] D. Horowitz and S. D. Kamvar, "The anatomy of a large-scale social search engine," in *Proceedings of 19th International Conference on World Wide Web*, 2010, pp. 431–440.

[14] B. Suh, L. Hong, P. Pirolli, and E. H. Chi, "Want to be retweeted? large scale analytics on factors impacting retweet in twitter network," in *Proceedings of 2nd IEEE International Conference on Social Computing*, 2010, pp. 177–184.

[15] H. Kwak, H. Chun, and S. Moon, "Fragile online relationship: a first look at unfollow dynamics in twitter," in *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, 2011, pp. 1091–1100.

[16] F. L. Pinheiro, M. D. Santos, F. C. Santos, and J. M. Pacheco, "Origin of peer influence in social networks," *Physical Review Letters*, 2014.

[17] D. Centola, "The spread of behavior in an online social network experiment," *Science*, vol. 329, no. 5996, pp. 1194–1197, 2010.

[18] M. Cha, H. Haddadiy, F. Benevenutoz, and K. P. Gummadi, "Measuring User Influence in Twitter: The Million Follower Fallacy," in *Proceedings of 4th International Conference on Weblogs and Social Media*, 2010, pp. 10–17.

[19] J. Yang and S. Counts, "Predicting the speed, scale, and range of information diffusion in twitter," in *Proceedings of 2010 AAAI Conference on Weblogs and Social Media*, 2010, pp. 355–358.

[20] J. Sallet, R. Mars, M. Noonan, J. Andersson, J. O'Reilly, S. Jbabdi, P. Croxson, M. Jenkinson, K. Miller, and M. Rushworth, "Social network size affects neural circuits in macaques," *Science*, vol. 334, no. 6056, pp. 697–700, 2011.

[21] M. J. Blooma, J. C. Kurian, A. Y.-K. Chua, D. H.-L. Goh, and N. H. Lien, "Social question answering: Analyzing knowledge, cognitive processes and social dimensions of micro-collaborations," *Computers & Education*, vol. 69, pp. 109–120, 2013.

[22] M. Thelwall, D. Wilkinson, and S. Uppal, "Data mining emotion in social network communication: Gender differences in myspace," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 1, pp. 190–199, 2010.

[23] J.-P. Onnela, S. Arbesman, M. C. González, A.-L. Barabási, and N. A. Christakis, "Geographic constraints on social network groups," *PLoS one*, vol. 6, no. 4, p. e16939, 2011.

[24] L. Yu, S. Asur, and B. A. Huberman, "Artificial inflation: The true story of trends in sina weibo," *arXiv preprint arXiv:1202.0327*, 2012.

[25] Q. Gao, F. Abel, G.-J. Houben, and Y. Yu, "A comparative study of users microblogging behavior on sina weibo and twitter," in *User Modeling, Adaptation, and Personalization*, 2012, pp. 88–101.

[26] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[27] B. Slinker and S. Glantz, "Multiple regression for physiological data analysis: the problem of multicollinearity," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 249, no. 1, pp. R1–R12, 1985.