

Understanding Experts' and Novices' Expertise Judgment of Twitter Users

Q. Vera Liao

Department of Computer Science
University of Illinois at Urbana-Champaign
201 N. Goodwin Ave, Urbana, IL 61801
liao28@illinois.edu

Peter Pirolli

Palo Alto Research Center
3333 Coyote Hill Rd, Palo Alto, CA 94304
pirolli@parc.com

Claudia Wagner

DIGITAL- Institute for Information and
Communication Technologies
Steyrergasse 17, Graz, Austria
claudia.wagner@joanneum.at

Wai-Tat Fu

University of Illinois at Urbana-Champaign
201 N. Goodwin Ave, Urbana, IL 61801
wfu@illinois.edu

ABSTRACT

Judging topical expertise of micro-blogger is one of the key challenges for information seekers when deciding which information sources to follow. However, it is unclear how useful different types of information are for people to make expertise judgments and to what extent their background knowledge influences their judgments. This study explored differences between experts and novices in inferring expertise of Twitter users. In three conditions, participants rated the level of expertise of users after seeing (1) only the tweets, (2) only the contextual information including short biographical and user list information, and (3) both tweets and contextual information. Results indicated that, in general, contextual information provides more useful information for making expertise judgment of Twitter users than tweets. While the addition of tweets seems to make little difference, or even add nuances to novices' expertise judgment, experts' judgments were improved when both content and contextual information were presented.

Author Keywords

Twitter, recommendation system, expertise judgment

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI):
Miscellaneous.

General Terms

Human Factors, Experimentation

INTRODUCTION

The rising popularity of micro-blogging enables individuals to become broadcasting channels. While most micro-

blogging platforms give users a high level of autonomy to choose which channel to follow, it is difficult to discover real experts from the vast pool of micro-bloggers, especially for people who are new to a topic or community. External recommendation systems that aim at identifying users who are experts of a topic are often considered an ideal solution to this problem. For example, Twitter, the most popular micro-blogging service, has attracted the development of a number of such recommendation systems to allow users to selectively follow micro-bloggers who can deliver the most relevant information. Among them, hybrid recommender approaches which analyze the topical relevance of users' published contents, and combine it with social network graph (e.g., follower information) are proved to be powerful in finding topical experts (e.g., [1][3] [5]).

The plethora of information on a Twitter page makes it challenging to assess a person's expertise accurately. In addition to the *tweets* (i.e., micro-blog messages), the biographical section (*bio*), with fewer than 160 characters, often contains important information that indicates the expertise of the user, such as his/her self summarized interests, career information, and links to his/ her personal web page. Another recently introduced feature of Twitter that could potentially be useful for assessing user's expertise is the support of *user list*. User lists allow users to organize people they are following into labeled groups. If a user is added to a list, the list label and short description of the list will appear on his/her Twitter page. Unlike bio information, which may contain self-reported expertise indicator, lists reflect external, i.e., followers' judgment about one's expertise and provide straightforward cues about this judgment to other users. While both bio and list content are textual information that could be analyzed by machine learning and natural language processing techniques, to the best of our knowledge, most existing systems do not utilize them but focus only on tweet content.

Ideally, a recommendation system should always provide results that users are satisfied with. Disagreement may

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2012, May 5-10, 2012, Austin, TX, USA.

Copyright 2012 ACM xxx-x-xxxx-xxxx-x/xx/xx...\$10.00.

happen if the system and users differ systematically in what information they use to make the expertise judgment. Therefore, it is important to understand how users judge the level of expertise based on various bits of information on the interface. Another factor that could potentially influence the judgment is how much the user knows about the topic. Customizable recommendation system often differentiates between expert and novice users because they tend to have different information needs and use different information processing strategies. There exists substantial evidence that when evaluating information quality, experts tend to actively deliberate on the quality of message content, while novice tend to rely on peripheral cues that are easier to make sense, such as the authoritativeness of information source [4]. Therefore, with regard to Twitter, it is possible that novices may pay more attention to contextual cues (user profile features) than experts since they often provide more salient cues for the judgment of expertise.

In this paper, we are interested in understanding the ability of experts and novices of a given topic to correctly judge the expertise of Twitter users when the judgment is based on the contents they published (tweets), self-reported and externally-reported contextual information (bio and user lists), or both contents and contextual information. Results will provide important information that guides the design of micro-blogger recommendation system to provide users with favorable results, and especially inform the design of customizable user recommendation system.

METHOD

Participants

We chose “semantic web” to be the topic for the experiment. We recruited 8 expert participants by contacting the faculties and students of a semantic web summer school held at a university and 8 novices from a university town in the United States. Participants’ age ranged from 20 to 34. Participants were asked to complete a knowledge test to verify if they were experts or novices of the given topic. The knowledge test included a few knowledge questions regarding semantic web, and a fluency task that asked the participant to generate as many terms related to semantic web as possible in 1.5 minutes. Each of the knowledge questions scored 5 points and each term in fluency task scored 1 point. The knowledge test showed that the two groups differ significantly in their knowledge about the topic (novice mean score 7.0; experts 29.1).

Experimental Materials and Procedure

We used WeFollow (www.wefollow.com), one of the most popular Twitter user recommendation systems, to select candidate Twitter users to be judged. WeFollow is a user powered Twitter directory where users can sign up for at most 3 directories. Wefollow ranks all users based on a proprietary algorithm, i.e., users who are followed by more users signed up for a topic directory get a higher rank of the particular topic. Wefollow suggested 276 Twitter users

relevant to the topic of semantic web. For candidates to represent high level of expertise, we randomly selected six users from rank 1-20 and six users from rank 93-113. For candidates of low expertise, we randomly selected six users from rank 185-205 and six users from the public Twitter timeline who did not show any relation to the topic. We excluded users whose tweets contained non-English language and users who had fewer than 50 tweets.

According to Cambridge Handbook of Expertise and Expert Performance [2], canonical ranking and social judgments by peers are commonplace for identifying expertise. We assume the way that WeFollow functions, by ranking users according to the number of followers in the same field, is a reflection of such social judgment. To further validate our manipulation, we randomly paired one candidate from our high expertise level with one from the low level, and ask semantic Web researchers to answer which one has higher expertise by checking their Twitter pages. We asked 3 raters to evaluate 50 pairs. The results showed that all of them had 95% or higher agreement with our manipulation, and the inter-rater agreement was 0.94. This result proved that our expertise level manipulation was successful.

Our experiment tested three conditions: participants saw 1) the latest 30 tweets and contextual information (bio and latest 30 lists) ; 2) only the latest 30 tweets; 3) only the bio and the latest 30 lists (or all lists if fewer than 30 were available). Each of the 24 pages we selected in step one was randomly assigned to one condition. In other words, for each condition, we have four users of high expertise and four users of low expertise. To tease out the influence of Twitter interface, we presented all the textual information in a table. For condition 1 the table has two columns to present tweets and contextual information separately. For condition 2 and 3 the table has only one column. The user’s name, profiles pictures and all the list creators’ names were removed to avoid the influence of uncontrolled variables.

Before the task, participants were asked to answer demographical questions and complete a knowledge test. Then they were presented with 24 evaluation tasks (three conditions, eight pages for each condition) in sequence. They were told that the information in the table was derived from a real Twitter user, and asked to rate how much this person knew about the topic, semantic web, on a one (least) to five (most) scale. The tasks took about 30-40 minutes.

RESULTS

We analyzed experts and novices’ expertise rating by performing three-way ANOVA with participants’ knowledge level (high/low) as between subjects variable, and Twitter user expertise (high/low) and conditions as within subjects variables. Interestingly, there is a three-way interaction between condition, Twitter user expertise and participants’ knowledge level ($F(1,14)=5.86, p<0.01$). To illustrate the three-way interaction, we plot novices and experts’ average ratings in different conditions in Figure 1.

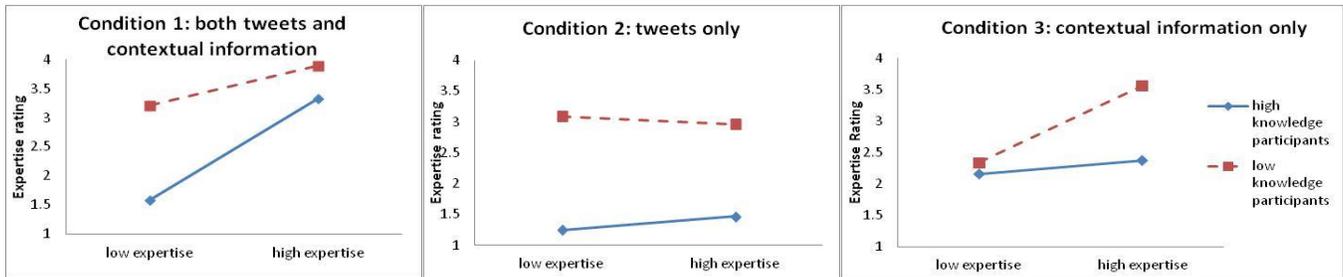


Figure 1. Average expertise rating given to users with high/low expertise by participants with high/low knowledge

It suggested the difference between experts and novices' judgment about Twitter user with high or low topical expertise varies for different conditions. For the analysis below, we focused on the comparisons of the slopes in Figure 1, which reflect the ability of differentiating between Twitter users of high and low expertise, to compare the ability of correctly judging the expertise between expert and novice participants, or between different conditions.

To understand the three-way interaction, for each of the three conditions, we performed two-way ANOVA on expertise ratings, with participants' knowledge and Twitter user expertise as independent variables. The interaction between Twitter user expertise and participants' knowledge level is significant in condition 1, where both the tweets and contextual information were shown ($F(1,14)=8.91, p=0.01$), and condition 3 where only contextual information was shown ($F(1,14)=9.36, p<0.01$), but not significant in condition 2 where only tweets were shown ($F(1,14)=2.47, p=0.14$). Referring to the differences of average ratings given to high and low expertise (the slopes) as shown in Figure 1, these results suggested that when both tweets and contextual information were shown, experts, as compared to novices, performed better in differentiating between Twitter users with high and low expertise. However, novices outperformed experts in condition where only contextual information was shown. When only tweets were shown, there is no significant difference between experts and novices' ability in correctly judging the expertise.

For experts, to compare their ability of making correct expertise judgments in different conditions, we performed two-way ANOVA with Twitter user expertise and conditions as independent variables for each pair of the three conditions. We focused on the significance of two-way interaction between Twitter user expertise and condition since it indicates differences of participants' ability of differentiating between high and low expertise between two conditions. The results showed this interaction was significant for condition 1 versus condition 2 ($F(1,7)=33.41, p<0.01$), and condition 1 versus condition 3 ($F(1,7)=17.67, p<0.01$), but not for condition 2 versus condition 3 ($F(1,7)=0, p=1$). It suggested that experts perform the expertise judgment best when seeing both content and contextual information. The same analysis performed for novices showed the same two-way interaction was significant for condition 3 versus condition

2 ($F(1,7)=7.00, p=0.01$), condition 3 versus condition 1 ($F(1,7)=5.40, p=0.05$), and marginally significant for condition 1 versus condition 2 ($F(1,7)=4.57, p=0.07$). The results indicated that novices performed best when only shown the contextual information.

Noting that the judgment becomes challenging for both experts and novices when only the tweets were presented, it seems to suggest that tweets are less informative for making expertise judgment than contextual information. On the other hand, experts' judgments were significantly improved when seeing both the content and contextual information than solely seeing either of them, but there was no such effect for novices. It implied that, as compared to novices, tweets, or the integration of tweets and contextual information, had a higher, positive impact on experts' judgments of Twitter user's expertise. Surprisingly, the addition of content cues seems to add nuances to novices' expertise judgment. While we do not have a clear answer for the causes from the current study, it is possible that novices could more efficiently rely on heuristics that utilize contextual cues to infer the Twitter user's expertise by avoiding the "information overload" from the tweets, for which they may not have adequate knowledge to correctly detect and comprehend every expertise related cues.

To further explore the difference between experts and novices, we investigated the difference of their judgments for each Twitter page. For each condition, we calculated the rank of the eight pages judged by experts and novices according to the average ratings given by each group. Then we calculate the absolute differences of the ranks between experts and novices for each page. The percentages of pages for which experts and novices' judgments has minor, medium, or high difference are shown in Table 1. It showed that the agreement between experts and novices was relatively high when seeing both content and contextual cues, and was lowest when only seeing the tweets.

We are particularly interested in the two pages where experts and novices' judgments differed the most. The first page in condition 2 was a highly ranked user. When only seeing the person's tweets, averagely experts ranked him second but novice ranked him seventh out of all the eight pages in condition 2. Interestingly, we found that this person tweeted about a specific event, "big data webinar" and frequently mentioned "cloud" and "cloud computing". Without explicitly seeing the term "semantic web", it is

possible that experts could easily relate these cues to the topic, while novices were less able to infer their association with the topic. The other page for which experts and novices disagreed on was another highly ranked user in condition 3. While novices ranked him as the second, experts ranked him as the sixth out of the eight users. In the experiment, we showed the participants his bio as below:

“CEO of XXX.com, a mobile, cloud-based, personal information management system. Founder & former CEO of Health xxx Network”

Also this person’s lists information frequently mentioned words like “business” and “entrepreneur”. It is possible that these cues had different meanings to experts and novices. While some authority related heuristics might lead novices to rate this person with a high expertise, the expert participants, who were mostly researchers in our study, may not directly relate the person’s administrative position to high level of expertise, but possibly expect more specifically knowledge related cues.

Difference of ranks	0-1	2-3	≥ 4
Condition 1	75%	25%	0
Condition 2	37.5%	50%	12.5%
Condition 3	62.5%	25%	12.5%

Table 1. The percentage of pages for which experts’ and novices’ average judgments differed for 0-1, 2-3 or more than 4 ranks

DISCUSSION

One interesting finding of this study is the low quality of expertise judgment based solely on tweets’ contents, even for users who had adequate knowledge about the topic. It highlighted the large variance of information people tweet about. Experts of a particular topic do not necessarily tweet about the topic all the time, if any. When we inspected some of the semantic web experts on the top ranks of WeFollow, they tweeted about almost everything, such as their daily activities, comments on news, promotion of their company, etc. It requires some efforts to distinguish expertise related cues from all the noises. In contrast, bio and user list information provide salient and straightforward cues for expertise judgments since they often provide descriptive information about the person himself, such as personal interests, professional experience, community the person belongs to. It is therefore questionable if algorithms for estimating and modeling users’ expertise should rely heavily on tweets. While previous research on finding topical authorities and experts on Twitter explored different ways of exploiting content features (e.g., [1] [3] [5]), most of them are solely based on tweets. According to our findings it is likely that these algorithms could be improved by also taking contextual information into account.

From the perspective of customizability, our study seems to suggest that there are differences between experts and

novices in terms of what information they use to make expertise judgment. Consistent with information processing theories, experts were better able to utilize content cues to infer their expertise judgment. Although the quality of expertise related cues in tweets alone may not allow experts to make satisfactory judgments, the integration of content and contextual information seems to achieve good results. Their worse performance when only seeing contextual information, as compared to seeing both contextual and content cues, further supports experts’ needs for evaluating cues from both sources when making expertise judgments. In contrast, novices’ judgments seem to be primarily informed by salient cues in contextual features, and are more efficient by focusing on this information alone. It suggests that recommender system should seek to fit the knowledge level of its users. For example, it is possible that while expert recommendations based on contextual information could achieve satisfactory results for novice information seekers, expert information seekers may have additional requirements for evaluation of contents.

Our results suggest that novices who are unfamiliar with a topic tend to rely on cues from contextual features. The current Twitter interface, however, does not make all contextual information directly available. Users have to click the “listed” number on the Twitter page to view the lists to which the user was added. It may be helpful to make the information in the “list” feature more salient to improve users’ topical expertise or relevance judgment. In addition, by highlighting the profile information (bio, following and followers, lists, etc) in recommendation results, users can more easily decide which user to further check on. This may improve users’ satisfaction with recommended results.

ACKNOWLEDGMENTS

We thank the members of Augmented Social Cognition Group of Palo Alto Research Center for support and feedback. We also thank ARL Network Science CTA for funding this research.

REFERENCES

1. Canini, K. R., Suh, B., & Pirolli, P. L. (2011). Finding credible sources in Twitter based on relevance and social structure. In Proc. SocialCom 2011 (In Press).
2. Ericsson, K. A., Charness, N., Hoffman, R. & Foltovich, P., *The Cambridge handbook of expertise and expert performance*. New York: Cambridge University Press
3. Pal, A., Counts, S. Identifying Topical Authorities in Microblogs. In Proc. WSDM 2011, ACM Press (2011)
4. Petty, R.E., Cacioppo, J.T, The Elaboration Likelihood Model. *Advances in Experimental Social Psychology* 19. 123-205 (1986).
5. Weng, J., Lim, E., Jiang, J., He, Q., TwitterRank: finding topic-sensitive influential twitterers, In Proc. WSDM 2010, ACM Press . 261-270 (2010).