Predicting Prospective Peer Helpers to Provide Just-In-Time Help to Users in Question and Answer Forums

Oluwabukola Mayowa Ishola, Gordon McCalla
ARIES Laboratory, Department of Computer Science
University of Saskatchewan, Saskatoon Canada
bukola.ishola@usask.ca , mccalla@cs.usask.ca

ABSTRACT
Question and answer forums are becoming more popular as increasing numbers of lifelong learners rely on such forums to receive help about their learning needs. Stack Overflow (SO) is an example of such a forum used by millions of programmers. The ability of users to receive timely answers to questions is crucial to the sustainability of such forums and for successful lifelong learning. In SO we have observed that the number of questions answered within 15 minutes have diminished with more questions taking a longer time to get answered or remaining unanswered in some cases. This suggests the need for an effective approach in predicting prospective helpers who can provide timely answers to the questions. In this paper, we seek to explore strategies to match helpers and help seekers. In particular we wish to use these strategies to predict which SO users will provide timely answers to questions asked in SO, and then compare these predictions to the users who actually answered the questions. In making these predictions we looked at 3 time frames of user data: 1 month, 3 months and 6 months. We used 5 basic strategies: frequency, knowledgeability, eagerness, willingness, recency; and we compared the success rates of each strategy in making predictions on 3 different success criteria: predicting the first answerer, predicting the answerer most liked by the asker of the question, and predicting the answerer rated most highly by other SO users. We then incorporated a timeliness measure, which takes into consideration how quickly the user provides answers to questions in the past, which helped us to achieve a higher success rate. The results of our study are an improvement over a similar previous study of SO and we hope will form the basis of methods for recommending peers in online forums who can provide just-in-time help to lifelong learners as their knowledge needs evolve and change.

Keywords
peer help, lifelong learning, peer matching

1. INTRODUCTION
Professional lifelong learners depend on online learning forums to help to meet their learning needs [2]. Our research is focused on supporting lifelong learners as they interact in such open-ended learning environments. Stack Overflow (SO) is an example of an online question and answer (Q&A) forum which supports millions of programmers. Over time, the answer response times to questions have increased and the number of unanswered questions has also increased. According to Asaduzzaman et. al. [1], failure of the questions asked to attract expert users is the top reason for unanswered questions, accounting for about 21.75% of unanswered questions. Receiving prompt answers to questions is important to the sustainability of a Q&A forum [2] and for successful lifelong learning.

While research efforts have been employed in the past in predicting potential peer helpers within a classroom-learning environment which encompasses just hundreds of students [4, 8,10], a new challenge arises in an online learning environment that is open ended with thousands or millions of potential helpers with varied expertise and learning interests. The need for an appropriate recommendation technique that scales up to millions of available users[1], and also aligns with the knowledge, interests and competency of the helper could be necessary. Greer et al. [4] in their study (similar to other studies [3,8,10]) employed the availability, helpfulness, technical ability and social ability of the helper as strategies considered in selecting the appropriate peer helper from the available users.

In a previous study using SO users as surrogates for lifelong learners, we employed a tag-based Naïve Bayes model to predict the answer performance of users using their previous activity in the forum [6]. The possibility of this model to predict poor answers even before they are provided could be used to help to reduce the frequency of poor answers within SO. In this new study, our goal is to predict helpers who are likely to provide answers to users' questions quickly ("just-in-time"). We also aim to determine how much information about the user is sufficient to predict the helper (to deal with issues such as those raised by Kay and Kummerfeld [7] about how much information must be usefully retained about the user in lifelong learning contexts). Finally, we compare the results from this study with the topic modelling approach used by Tian et al. [9]. We hope this study will augment such studies as [3, 4, 8, 10] in providing peer helper seeking strategies that scale to very large numbers of users.

2. RELATED WORK
In supporting learners in computerized learning environments human helpers and intelligent agents have been employed. Merrill et. al. [8] compared the help provided by peer helpers with that provided by intelligent agents and conclusions from this study show that human helpers provide more flexible and subtle help. Similarly, Greer et al. [4], building on earlier work in finding peer helpers in workplace environments [3], built the iHelp system to help computer science students find potential peer helpers among their classmates who are ready, willing and able to help in overcoming impasses. In addition, Vassileva et al. [10] in their study with iHelp incorporated the social characteristics of the helper into determining an appropriate helper, gleaned from the

[1] We will use the term "user" in this paper rather than "learner" when specifically discussing SO users since they are likely not explicitly learners in their own minds. However, in the future most professionals will be using such forums to meet their lifelong learning goals. The term "learner" then will be highly appropriate. Since our research is aimed at helping develop tools for such professional lifelong learners, especially tools that support personalization to each such learner, it is, we believe, deeply and broadly relevant to advanced learning technology.

online activities of the helper such as votes received by the helper, questions asked, answers provided, and the marks received on assignments.

While these studies [3,4,10] have all successfully recommended just-in-time helpers for a relatively small number of students within classroom and workplace settings, in a typical question and answer forum, the number of users ranges from thousands to millions of users with more varied knowledge interests [5]. The sustainability of such a large-scale question and answer forum is dependent on providing quick responses to questions [2]. A study by Bhat et al. [2] reveals that in Stack Overflow, although most of the questions get answered in less than 1 hour, about 30% of the questions have a response time of 1 day with about 344,000 questions having a response time greater than 1 day. In addressing the increasing number of unanswered questions, Bhat et al. [2] revealed the importance of assigning appropriate tags to questions; Asaduzzaman et al. [1] predicted how long a question will remain unanswered; and Tian et al. [9] predicted the best answerers to questions using a topic modelling approach. Yang and Manandhar [11] identified the topic modelling approach as a less effective approach that is too general while the use of question tags was proposed as a more informative approach. The study by Tian et al. [9] in predicting best answerers achieved a success rate of 21.5% while recommending 100 users who could answer the question. This reveals the need to explore other methodologies in predicting best answerers to questions.

3. ANALYSIS OF QUESTION RESPONSE TIME AND UNANSWERED QUESTIONS IN STACK OVERFLOW

SO is a question and answer forum that provides a platform to support millions of programmers by providing opportunities for them to ask questions and obtain answers from peers [5]. In cases where users do not receive answers form their peers, the user could provide answers to their own questions or sometimes, the questions remain unanswered. Key to the success of such a forum is the ability of users to receive prompt answers to their questions [2]. We studied the answer response time of questions in SO from January 2009 to December 2015, the distribution of questions answered by question askers themselves, and the proportion of unanswered questions. We defined the answer response time as the time difference between the times when a question is asked to when it receives the first answer. Figure 1 shows the answer response time of questions for each of 6 defined time intervals (within 15 minutes, within 1 hour, within 1 day, within 1 week, within 1 month and over a month) for each year under consideration.

Figure 1 shows that the majority of questions in SO get answered within 15 minutes, although we also observe a continuous decrease over time in the percentage of questions answered within 15 minutes. In fact, in 2015 just 36% of the questions were answered within 15 minutes compared to 2009 when about 57% of the questions were answered within 15 minutes. Also, questions with response times above 15 minutes have continually increased. In fact, some of the questions which received late answers were actually answered by the question askers themselves. Specifically, the total number of questions in this category has increased from 1,946 in 2009 to 18,479 in 2015 as shown in Table 1. In fact, some of these questions never get answered. Figure 2 shows a rapid growth in the number of unanswered questions.

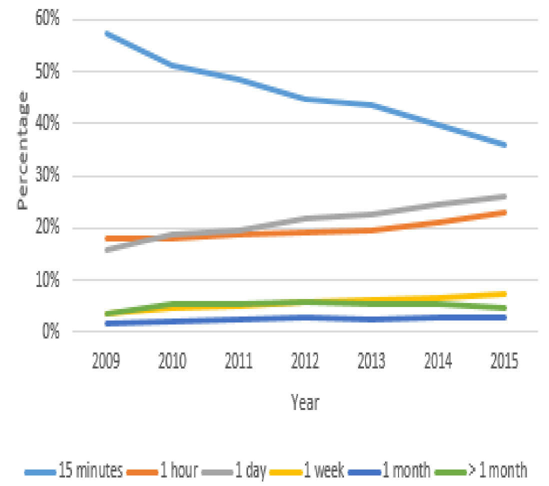

Figure 1: Response Time between Question Creation Date and First Answer Creation Date

Table 1: Questions Answered by the Question Asker

Year	Frequency
2009	1,946
2010	3,091
2011	6,701
2012	11,877
2013	16,936
2014	17,405
2015	18,479

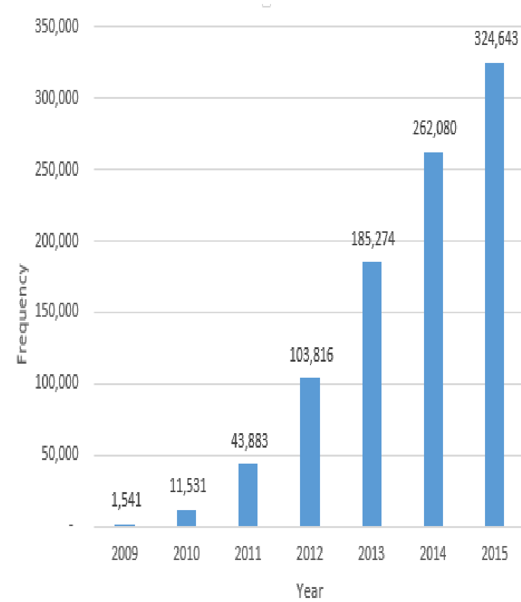

Figure 2. Number of Unanswered Questions

While this growth is partly a result of an increase in the number of questions asked in SO, we believe a growth from 1,541 in 2009 to 324,643 in 2015 is worth addressing. Moreso, Asaduzzaman et. al. [1] identified that the inability of questions to attract expert users is one of the main reasons they remain unanswered. Of course, not receiving answers to questions or having to answer your own question yourself could deter the user from subsequently using the forum. The goal of our research is to support users who depend on online forums to receive answers to their questions. We believe the ability to predict prospective answerers for questions is the first step at supporting users to achieve this goal.

4. RANKING STRATEGIES

Results in section 3 suggest the need to support users in question and answer forums with the aim of decreasing the answer response time to questions. Our study seeks to predict such potential just-in-time peer helpers using 5 strategies for choosing such a helper. Each of these strategies considers the relevance of the question to online activities and the demonstrated knowledge in answers of the potential helpers (other users) in the past (we defined this by the co-occurrence of tags contained in the question with tags contained in the answers provided by the potential helper in the past). For each proposed strategy, personalized scores are assigned to each prospective helper based on their suitability to answer a question, as described below.

4.1 Frequency

The frequency strategy measures how frequently the prospective helper has answered questions relevant to a particular question under consideration in the past. The higher the frequency of interaction with relevant questions in the past, the more likely the user would be to answer the question. The frequency score was computed by counting the number of answer posts A relevant to the question $tag(i)$ for user u as shown in equation 1 below:

$$Score_{ui}^{freq} = \sum A(i)_u \quad (1)$$

The prospective helpers with higher scores are ranked as better helpers based on this strategy.

4.2 Knowledgeability

Knowledgeability shows how much a prospective helper knows about the question based on the number of up votes the user has earned in answering past questions with the same tag (in SO questions and answers are voted upon to show how useful and appropriate they are). This is computed as shown in equation 2 below:

$$Score_{ui}^{know} = \sum Upvotes\ (A(i)_u) \quad (2)$$

that is the sum of all upvotes to answer posts A relevant to question $tag(i)$ for user u. Prospective helpers with a higher number of up votes would be ranked as better based on this strategy.

4.3 Eagerness

Eagerness is based on monitoring the online activity of a prospective helper as depicted by the proportion of answers they have provided in the past relevant to the question compared to the total number of answers provided by the user to all questions, as shown in equation 3 below. The eagerness measure depicts the probability that a user will answer a question related to $tag(i)$:

$$Score_{ui}^{eag} = \frac{Score_{ui}^{freq}}{N_u^a} \quad (3)$$

N_u^a represents the total number of answers provided by the user to all questions. This strategy seeks to measure the interest of the user in answering questions related to $tag(i)$ by considering the proportion of relevant questions answered. We assume that users will provide more answers to questions they are more interested in; therefore the higher the proportion of relevant questions answered, the higher the likelihood the helper would be interested in answering the particular question under consideration. Prospective helpers with higher scores are ranked higher.

4.4 Willingness

This measure is a combination of how active and eager the user has been in answering questions related to the question tag in the past. That is, a user who is eager to answer questions like the question under consideration and has answered such questions a lot should be more willing to answer the question under consideration. The Bayes theorem is applied in computing this peer matching measure as shown in equation (4) below:

$$P\left(U_u^a \middle| tag(i)\right) = \frac{P(tag(i)|U_u^a) * P(U_u^a)}{P\big(tag(i)\big)} \quad (4)$$

where $P(tag(i)|U_u^a)$ is the likelihood of an answer to a question related to $tag(i)$ will be given by a user u, which is computed as shown in equation (4a) below:

$$P(tag(i)|U_u^a) = \frac{Score_{ui}^{freq}}{N(i)_a} \quad (4a)$$

$N(i)_a$ represents the total number of answers provided to $tag(i)$ by all users. $P(U_u^a)$ is the prior probability of a user u answering a question related to $tag(i)$ which is equivalent to the eagerness of the user as computed in equation (3) above. $P\big(tag(i)\big)$ is the probability that a question related to $tag(i)$ will be asked (this is the same for all prospective helpers). To maximize the posterior probability as shown in equation (4), the numerator is maximized since the denominator is common to all the prospective helpers. The willingness score is therefore computed as shown in equation (4b) below (we substituted values from equation (4a) and (3) into equation (4)):

$$Score_{ui}^{will} = \frac{Score_{ui}^{freq}}{N(i)_a} * Score_{ui}^{eag} \quad (4b)$$

Prospective helpers with higher willingness score are ranked higher.

4.5 Recency

The recency strategy corresponds to how actively and recently the prospective helper has provided answers to relevant questions. The recency score is computed for each prospective helper based on the timestamp of the latest answer A provided relevant to the question $tag(i)$ as shown in equation 5 below:

$$Score_{ui}^{rec} = latest(Time\ A(i))_u \quad (5)$$

This simply means that the recency score for a user u who has provided answers A to questions with $tag(i)$ will be the timestamp of their latest answer (the maximum time). Under this measure prospective helpers who have answered related questions more recently would be ranked higher than those who answered such questions earlier. As the interests of potential helpers could evolve [5], providing answers to relevant questions in recent times could imply the prospective helper is still interested in answering questions related to the question tags. Although Greer et al. [4] argued that helpers who have recently provided help should be exempt, to avoid overworking a peer helper in SO, this might not be as true, as users might still be willing to provide help with the goal of earning some incentive from the forum (this could be the earning of a reputation score or of various badges).

5. EXPERIMENTAL EVALUATION AND RESULTS

The goal of our study is to explore the effectiveness of different peer-helper matching strategies in terms of their ability to predict a relevant peer-helper who will provide quick answers. For each of the strategies described in section 4, we evaluated their effectiveness using the historical SO data of each prospective helper going back 1 month, 3 months and 6 months from the time a question was asked. For this study we only focused on *java*[2] questions (53,731 of them) that received at least one answer within the first hour of creation with 254,766 prospective helpers to choose from. These represent questions that were answered fairly much in time which we feel would provide a good rationale in evaluating the effectiveness of the various strategies in predicting the just-in-time answerers. Likewise, we regarded only users who were available online within the first hour the question was created to be users who would be prospective helpers, as in a real life situation; they are the set of users who are more likely to view the questions earlier and provide quicker response. Also, we employed the one hour time frame in defining the online users as it aligns with the time frame of the questions considered in this study.

We also need a success measure for our predictions. Similar to the study by Tian et al. [9], we deem it a success if a user in the top N ranked users computed by a strategy is also a user who actually answered the question under consideration in SO. The success rate S@N for each strategy can then be computed by dividing the total number of successes by the total number of questions as shown in equation 6 below.

$$S@N = \frac{Total\ Number\ of\ Successes}{Total\ Number\ of\ Questions} * 100\% \quad (6)$$

We can use different values of N to get a glimpse into how our prediction would perform as the number of prospective helpers predicted increases. In our study we used N = 1, 5, 10, and 20. Finally, we wanted to compare the effectiveness of our strategies in three different prediction criteria: predicting the answerer who responded first in SO, predicting the answerer who gave the best answer according to the user who asked the question, and predicting the answerer whose answer other SO users ranked as having the best score.

***Predicting the first answerer*:** This criterion evaluates the ranked list of prospective helpers predicted for each of the strategies with the aim to know their effectiveness at predicting the user who will first provide an answer to the question. The results in table 2 show that considering the willingness of a prospective helper has the highest success rate of 55.86% with S@20 using a time frame of 6 months.

***Predicting the best answerer*:** In SO, from the numerous answers provided to a question, the question asker can mark only one of the answers as accepted which indicates the best answer according to the asker [9]. The goal of this evaluation criteria is to determine the success of the measures at identifying the best answerer from the ranked list of prospective helpers suggested. The results are shown in table 3 below. As in predicting the first answerer, we observed that the willingness peer matching strategy has the highest success rate of 54.62% with S@20 using the 6 months defined time line.

***Predicting the answerer with the highest score*:** Other community (SO) members also have the privilege to vote on the answers provided if they wish. In some cases the answer voted as best by the question asker might not necessarily be the answer with the highest score according to the community. With this evaluation criterion we want to examine the effectiveness of the peer matching strategies at predicting the user with the highest score. Results from this evaluation are shown in table 4 below. Amongst the 7 strategies considered, again we observed willingness of the prospective users has the highest success rate at predicting the user who obtained the highest success with a success rate of 56% with S@20 using the 6 months defined time line.

Overall, with the 3 evaluation criteria we achieved the highest success rate with the willingness measure and the least success with the recency strategy. Also, we observed that as the number of months increases from 1 to 6 months, we did not see any tremendous difference in the success rate for all the strategies. Tables 2 - 4 show (unsurprisingly) that as N increases, the success rate of the prediction also increases. Comparing all 3 evaluation criteria, we achieved the highest success while predicting the user with the highest score, although the success rate obtained with the other criteria (i.e. predicting the first answerer and best answerer) did not differ significantly using S@20. In the next section, we show how we attempted to improve the performance of these strategies by including an additional measure called ***timeliness.***

6. PREDICTION OF JUST-IN-TIME HELPERS

The main goal of this study is to predict helpers just-in-time, i.e. helpers who would provide answers as quickly as possible. Therefore we included a *timeliness* criterion that takes into consideration how quickly a prospective helper would provide an answer to a question. We used the 15 minutes time frame as it represents the average time in which most questions are answered (although, the percentage of questions answered within this time frame has decreased as shown in section 3). For each prospective helper, we computed the timeliness measure as shown in equation (7):

$$Score_u^{Tim} = \frac{N_u^{t\leq 15}}{N_u^a} \quad (7)$$

$N_u^{t\leq 15}$represents the number of questions the user answered within 15 minutes in the past while N_u^a represents the total number of answers provided by user u. To see how well our various strategies work in predicting such just-in-time helpers, we multiplied the timeliness score $Score_u^{Tim}$ obtained by each user by their respective score on each of the other strategies except for the recency strategy. We excluded the recency strategy in this prediction as it is the weakest measure as shown in tables 2-4. Moreover, the recency score computed as shown in equation 7 is a timestamp value which cannot be multiplied by the timeliness score as can the numeric values obtained with other strategies. Finally, since we did not observe any major differences when we used the 1 month history data of the prospective helper as compared to the 6 month history, in predicting the just-in-time helpers we only employed the history data of the prospective answerers over the 1 month time frame. This also saved a lot of computational time. The results obtained are shown in tables 5-7 for each of the evaluation criteria.

[2] We focused on questions containing *java* tags as this is the most used programming related tag in SO.

Table 2: Success Rate at Predicting the First Answerer

First Answerer	1 Month				3 Months				6 Months			
	S@1	S@5	S@10	S@20	S@1	S@5	S@10	S@20	S@1	S@5	S@10	S@20
frequency	5.40%	18.87%	31.65%	49.13%	5.27%	18.93%	31.37%	48.23%	5.81%	20.00%	33.13%	50.81%
recency	2.39%	11.31%	20.30%	33.60%	2.61%	11.67%	20.67%	33.96%	2.80%	12.66%	21.81%	35.59%
eagerness	1.81%	9.89%	21.29%	43.57%	1.88%	10.09%	21.53%	43.82%	2.01%	10.32%	23.15%	47.00%
knowledgeability	5.59%	17.97%	28.10%	39.52%	5.50%	17.85%	28.05%	39.32%	5.97%	19.03%	29.78%	41.94%
willingness	5.70%	21.06%	35.89%	**54.20%**	5.58%	21.11%	35.35%	**52.90%**	6.06%	22.43%	37.44%	**55.86%**

Table 3: Success Rate at Predicting the Best Answerer

Best Answerer	1 Month				3 Months				6 Months			
	S@1	S@5	S@10	S@20	S@1	S@5	S@10	S@20	S@1	S@5	S@10	S@20
frequency	5.27%	19.60%	31.84%	48.25%	5.27%	19.58%	31.72%	47.26%	5.77%	20.78%	33.34%	50.26%
recency	2.91%	12.20%	21.19%	33.91%	3.17%	12.55%	21.48%	34.35%	3.53%	13.70%	22.84%	36.12%
eagerness	1.75%	9.36%	19.98%	41.03%	1.89%	9.76%	20.69%	41.43%	1.97%	9.90%	22.06%	44.61%
knowledgeability	5.58%	19.18%	29.24%	40.66%	5.58%	18.99%	29.27%	40.66%	5.97%	20.33%	31.22%	43.54%
willingness	5.58%	21.40%	35.40%	**52.47%**	5.57%	21.30%	35.08%	**51.52%**	6.00%	22.80%	37.29%	**54.62%**

Table 4: Success Rate at Predicting the Answerer with the Highest Score

Highest Score	1 Month				3 Months				6 Months			
	S@1	S@5	S@10	S@20	S@1	S@5	S@10	S@20	S@1	S@5	S@10	S@20
frequency	5.43%	19.96%	32.48%	49.38%	5.47%	20.28%	32.46%	48.43%	5.90%	21.49%	34.34%	51.37%
recency	2.88%	12.26%	21.62%	34.90%	3.20%	12.92%	22.11%	35.33%	3.66%	14.11%	23.40%	37.13%
eagerness	1.82%	9.30%	20.09%	42.12%	1.94%	9.96%	21.18%	42.88%	2.02%	10.19%	22.61%	46.05%
knowledgeability	5.79%	19.99%	30.29%	41.73%	5.76%	19.92%	30.36%	41.80%	6.12%	21.16%	32.32%	44.63%
willingness	5.66%	21.71%	36.23%	**53.63%**	5.67%	21.89%	36.09%	**52.78%**	6.10%	23.45%	38.52%	**56.00%**

Table 5. Timeliness Success at Predicting the First Answerer

First Answerer Timeliness	1 Month			
	S@1	S@5	S@10	S@20
frequency	6.54%	21.86%	36.16%	55.01%
eagerness	5.46%	26.71%	43.31%	**63.15%**
knowledgeability	6.06%	20.10%	30.45%	41.54%
willingness	6.91%	24.89%	40.55%	60.34%

Table 6. Timeliness Success at Predicting the Best Answerer

Best Answerer Timeliness	1 Month			
	S@1	S@5	S@10	S@20
frequency	6.10%	20.95%	34.09%	50.84%
eagerness	3.80%	20.95%	35.27%	53.76%
knowledgeability	5.85%	20.19%	30.38%	41.45%
willingness	6.45%	23.64%	37.91%	**55.34%**

Table 7. Timeliness Success at Predicting the Answerer with the Highest Score

Highest Score Timeliness	1 Month			
	S@1	S@5	S@10	S@20
frequency	6.21%	21.46%	34.98%	51.94%
eagerness	3.90%	21.47%	36.47%	55.34%
knowledgeability	6.02%	21.06%	31.38%	42.54%
willingness	6.48%	24.30%	38.93%	**56.65%**

7. DISCUSSION

The aim of our research is to support lifelong learners as they interact with peers in open ended learning environments like SO. As lifelong learners are responsible for their own learning [7], millions of them depend on such learning forums to meet their learning needs on a daily basis. Obtaining timely answers to questions is important [2] in supporting lifelong learners and in enhancing the sustainability of such an online learning community. However, we observed (as shown in section 2) that the answer response times to questions have increased and in some cases the question askers have to answer their own questions themselves, which can deter the lifelong learner. In this study, we address this problem by predicting prospective users who are likely to provide the most timely answers to their question.

Previous studies by Greer et al. [3, 4] and Vassileva et al. [10] have identified the various strategies that could be used in predicting the prospective helpers within the classroom and workplace learning environments. In this study we explored the effectiveness of the various strategies at predicting prospective helpers in SO, an environment with vastly more learners seeking answers to their questions than in academic classes. We achieved the highest success rate S@20 of 54.20% using the 1 month time line with the willingness strategy. Also, with the recency measure, performing the poorest amongst all the measures defined, our study affirms the claim by Greer et al. [2] that helpers who have recently provided help would be less likely to provide answers and they should be exempted to avoid overworking a peer helper.

We improved upon the results obtained from each of the strategies described in section 4, by including an additional criterion called *timeliness*. This criterion takes into consideration the probability that a user would answer a question quickly. We achieved a maximum success rate S@20 of 63.15% (eagerness), 55.34% (willingness) and 56.65% (willingness) in predicting, respectively, the first answerer, the best answerer, and the answerer who will provide the highest score. These values represent an improvement in the success rate from 43.57% to 63.15% (eagerness), 52.47% to 55.34% (willingness), 53.63% to 56.65% (willingness) in predicting the first answerer, best answerer and the answerer who will provide the highest score respectively using the 1 month time frame (comparing our results from tables 2-4 with results obtained in tables 5-7). While these results likely require improvement, these values are an improvement over the previous work by Tian et al. [9] whom obtained a success rate S@20 of 12.57% and S@100 of 23.06% while predicting the best answerer using the topic modelling approach. We believe the results obtained in this study for all the strategies defined outperforms this previous work. The variation in our results from those of Tian et al. is presumably because our study was restricted to questions that were answered fairly much on time (i.e. questions with at least one answerer within the first hour the question was created). We focused on these sets of questions because the goal of our study is to predict the just-in-time helpers who will provide quick answers to the questions in which case, questions answered late would not suffice. Although Yang and Manandhar [11] argued for the use of the topic modelling approach in predicting the best answerer, our results suggest that this is a less informative approach.

For each of the peer matching strategies, we also studied their performance in predicting the relevant peer helpers using the history data for prospective peer helpers for the periods of 1 month, 3 months and 6 months. Our aim is to understand the tradeoff of using older data about the user vs newer data. As Kay and Kummerfield [7] already identified, there is a trade-off between the usefulness of retaining older information about the lifelong learner and preserving only the recent data. Our results show that employing older information (6 months) about the learner was at best only marginally better when compared to the results achieved with the newer information (1 month). This confirms an earlier study [5] we did in predicting (again in SO) what the user would want to learn in the future, where we showed that employing shorter term information about the user's past behavior proved more effective in predicting what the user would be learning in future

While we feel that we have achieved good prediction accuracy with our strategies (especially as compared to other studies), we would still like to enhance the accuracy to ensure the usefulness of our strategies in a real learning environment. So, in our next experiment, we aim to further improve on our results, pushing them well above our current success rates if we can. Our aim will be to develop new strategies that can identify users who would have been likely to help answer the question quickly. Overall, we feel this research is a promising first step for being able to show how we can find good peer helpers to help professional lifelong learners who are keeping themselves up-to-date through interactions with their peers in online forums.

8. ACKNOWLEDGMENTS
We would like to thank David Edgar Kiprop Lelei for his help with presenting the paper. Aso, we like to thank the Natural Sciences and Engineering Research Council of Canada and the University of Saskatchewan for providing funding to support this research.

9. REFERENCES
[1] Asaduzzaman, M., Mashiyat, A. S., Roy, C. K., & Schneider, K. A. (2013, May). Answering questions about unanswered questions of stack overflow. 10th IEEE Working Conference on Mining Software Repositories (MSR 2013) (pp. 97-100).

[2] Bhat, V., Gokhale, A., Jadhav, R., Pudipeddi, J., & Akoglu, L. (2014, August). Min (e) d your tags: Analysis of question response time in stackoverflow. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014) (pp. 328-335).

[3] Greer, J. E., McCalla, G., Collins, J. A., Kumar, V. S., Meagher, P., & Vassileva, J. (1998). Supporting peer help and collaboration in distributed workplace environments. International Journal of Artificial Intelligence in Education (IJAIED), 9, pp. 159-177.

[4] Greer, J., McCalla, G., Cooke, J., Collins, J., Kumar, V., Bishop, A., & Vassileva, J. (1998, August). The intelligent helpdesk: Supporting peer-help in a university course. International Conference on Intelligent Tutoring Systems (ITS 1998) (pp. 494-503).

[5] Ishola, O. M., & McCalla, G. (2016, September). Detecting and supporting the evolving knowledge interests of lifelong professionals. European Conference on Technology Enhanced Learning (EC-TEL 2016) (pp. 595-599).

[6] Ishola, O. M., & McCalla, G. (2017, June) Personalized tag-based knowledge diagnosis to predict the quality of answers in a community of learners. International Conference on Artificial Intelligence in Education (AIED 2017), (to appear).

[7] Kay, J., & Kummerfield, B. (2009, June). Lifelong user modelling goals, issues and challenges. In Lifelong User Modelling Workshop at the International Conference on User Modeling, Adaptation and Personalization (UMAP 2009) (pp. 27-34).

[8] Merrill, D. C., Reiser, B. J., Ranney, M., & Trafton, J. G. (1992). Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. The Journal of the Learning Sciences, 2(3), pp. 277-305.

[9] Tian, Y., Kochhar, P. S., Lim, E. P., Zhu, F., & Lo, D. (2013, November). Predicting best answerers for new questions: An approach leveraging topic modeling and collaborative voting. Workshops at the International Conference on Social Informatics (SocInfo 2013) (pp. 55-68).

[10] Vassileva, J., Greer, J., McCalla, G., Deters, R., Zapata, D., Mudgal, C., & Grant, S. (1999, July). A multi-agent approach to the design of peer-help environments. International Conference on Artificial Intelligence in Education (AIED 1999) (pp. 38-45).

[11] Yang, B., & Manandhar, S. (2014, August). Tag-based expert recommendation in community question answering. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014) (pp. 960-963).