PERSONALIZED APPROACHES TO SUPPORTING THE LEARNING NEEDS OF LIFELONG PROFESSIONAL LEARNERS

A Thesis Submitted to the

College of Graduate and Postdoctoral Studies

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

In the Department of Computer Science

University of Saskatchewan

Saskatoon

By

Oluwabukola Mayowa Ishola

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other uses of materials in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science
176 Thorvaldson Building
110 Science Place University
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5C9
Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9
Canada

Abstract

Advanced learning technology research has begun to take on a complex challenge: supporting lifelong learning. Professional learning is an essential subset of lifelong learning that is more tractable than the full lifelong learning challenge. Professionals do not always have access to professional teachers to provide input to the problems they encounter, so they rely on their peers in an online learning community (OLC) to help meet their learning needs. Supporting professional learners within an OLC is a difficult problem as the learning needs of each learner continuously evolve, often in different ways from other learners. Hence, there is a need to provide personalized support to learners adapted to their individual learning needs.

This thesis explores personalized approaches for detecting the unperceived learning needs and meeting the expressed learning needs of learners in an OLC. The experimental test bed for this research is Stack Overflow (SO), an OLC used by software professionals. To date, seven experiments have been carried out mining SO peer-peer interaction data. Knowing that question-answerers play a huge role in meeting the learning needs of the question-askers, the first experiment aimed to detect the learning needs of the answerers. Results from experiment 1 show that reputable answerers themselves demonstrate unperceived learning needs as revealed by a decline in quality answers in SO. Of course, a decline in quality answers could impact the help-seeking experience of question-askers; hence experiment 2 sought to understand the effects of the help-seeking experience of question-askers on their enthusiasm to continuously participate within the OLC. As expected, negative help-seeking experiences of question-askers had a large impact on their propensity to seek further help within the OLC.

To improve the help-seeking experience of question-askers, it is important to proactively detect the learning needs of the question-answerers before they provide poor quality answers. Thus, in experiment 3 the goal was to predict whether a question-answerer would give a poor answer to a question based on their past peer-peer interactions. Under various assumptions, accuracies ranging from 84.57% to 94.54% were achieved. Next, experiment 4 attempted to detect the unperceived learning needs of question-askers even before they are aware of such needs. Using information about a learner's interactions over a 5-month period, a prediction was made as to what they would be asking about during the next month, achieving recall and precision values of 0.93 and 0.81. Knowing the learning needs of question-askers early creates an opportunity to predict prospective answerers who could provide timely and quality answers to their question. The goal of experiment 5 was thus to predict the actual answerers for questions based only on information known at the time the question was asked. The

success rate was at best 63.15%, which would only be marginally useful to inform a real-life peer recommender system. Thus, experiment 6 explored new measures in predicting the answerers, boosting the success rate to 89.64%. Of course, a peer recommender system would be deemed to be especially useful if it can provide prompt interventions, especially to get answers to questions that would otherwise not be answered quickly. To this end, experiment 7 attempted to predict the question-askers whose questions would be answered late or even remain unanswered, and a success rate of 68.4% was achieved.

Results from these experiments suggest that modelling the activities of learners in an OLC is key in providing support to them to meet their learning needs. Perhaps, the most important lesson learned in this research is that lightweight approaches can be developed to help meet the evolving learning needs of professionals, even as knowledge changes within a profession. Metrics based on the experiments above are exactly such lightweight methodologies and could be the basis for useful tools to support professional learners.

Acknowledgements

I have received tremendous help from Gord McCalla from the get-go to the completion of my thesis. During the challenging times of my doctoral research, the continuous encouragement and excellent guidance received are so valuable. I am incredibly grateful for the financial support and all opportunities you provided to me as your student. Gord is an outstanding supervisor who shows care and concern about every aspect of his student life. I am grateful for your kind words, encouragement and understanding upon the early arrival of Peace. Thanks for being supportive throughout the years.

I am grateful to my committee members—Rick Schwier, Julita Vassileva, and Michael Horsch, Debajyoti Mondal and late Jim Greer. Thank you all for your time, feedback and suggestions you provided to me during my doctoral study. I have gained valuable insights from your questions and suggestions on many occasions. Special thanks to Gwen Lancaster for her compassion and support during my program.

Thanks to my fellow graduate students in the ARIES Lab. Thanks to Jennifer Seaton, Terry Peckham, Edgar Lelei and Stephanie Frost for your support, contribution and time spent listening to my talk. Special thanks to my fellow African students in the department of computer science, Rita Orji, Johnson Iyilade, Bunmi Olakanmi, Richard Lomortey, and Kiemute Oyibo for the time spent together.

Special thanks to my father, late Abel Kolawole for your sacrificial love and the time you invested in me. I am also sincerely grateful to my mother, for the time she spent taking care of Peace which availed me the opportunity to remain focussed on my studies. To all my siblings and relatives who have overwhelmed me with so much love from far and near, you are indeed appreciated. I am fortunate to have you all as my siblings.

Special appreciation to my dearest husband, Ademola Idowu for your tremendous help and sacrifice during my study. Thanks a lot for the valuable insights you provided to my research often. Thanks for your love and for standing by me to ensure I complete my doctoral studies. Special thanks to my son, Peace Idowu whose arrival has brought joy and strength to press forward to the end of my Ph.D. studies. Finally, I thank the almighty God, who has been a good father and anchor to me from the beginning of my life.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviations

NWTU	Normalized Weighted Tag Usage
OLC	Online Learning Community
PA	Prediction Accuracy
PQA	Percentage of Answer Quality
PIFQ	Percentage Increase in the Frequency of Questions.
RTFA	Response Time for First Answer
SO	Stack Overflow
TBE	True Bayesian Estimate

List of Equations

CHAPTER 1

INTRODUCTION

Lifelong learning (LLL) is a burgeoning area of advanced learning technology research, whose goal is to support learners with the knowledge and skills needed to succeed in a rapidly changing world (Dolog, Kay, and Kummerfeld, 2009). LLL is crucial for professionals who have an obligation to stay up to date throughout their professional practice. Technological advances are driving rapid ongoing changes in professions (Simons and Ruijters, 2004), work and society. It is easy for professionals to be so overwhelmed with work responsibilities that they are ignorant of valuable new knowledge that exists. Moreover, professionals can be unaware of their ignorance (Dunning, 2011).

Traditionally, professional learning has been accomplished through job training, short courses, and self-directed learning (Bruce, 1999), which does not scale beyond the workplace. Even with the job training received by professionals, millions of professionals still depend on online learning communities to help them overcome problems they may encounter daily (Ishola and McCalla, 2016a). Most users[1] in these support communities are learners helping each other to resolve their learning needs. My research is focused on supporting users in such online learning communities (OLCs).

Specifically, this thesis investigates issues involved in providing personalized support to software programmers in detecting and meeting their learning needs. As the experimental base of this research, the interaction data of software programmers in Stack Overflow (SO) was mined to detect patterns which would aid in inferring and meeting the individual learning needs of users. SO is a "question and answer site for professional and enthusiast programmers" [http://stackoverflow.com/]. SO contain the questions and answers, profiles, badges, reputation scores, and other data of over 8.4 million users, although only a proportion

[1] I will use the term "user" in this thesis rather than "learner" when specifically discussing users in online learning communities (for example Stack Overflow) 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.

of the users are active users. SO is a large-scale repository of information about programmers and their learning needs.

Applying advanced learning technology techniques such as social media mining, learner knowledge diagnosis, and personalization to the needs of professional learners in OLCs has not been a focus of research (Ley et al., 2016). Appropriate support that scales up to millions of learners and aligns with the evolving learning needs of the learners becomes necessary. In an OLC where the learning needs of most users are driven by technological advancement in their profession rather than by a well-defined curriculum (as in traditional learning environments), the challenge of detecting the evolving learning needs of users is acute. To support the diversified learning needs of users, there is a need to diagnose their individual learning needs and provide help to overcome them.

1.1 Scope of Research

Learning needs can be defined as the gap between what the learner knows and what the learner ought to know to do the job (Robinson, 1998). These learning needs are generated from various sources as shown in Figure 1.1, whether or not the professionals are aware of them. A brief explanation of the various sources of learning needs is described below (Manninen and Hobrough, 2000).

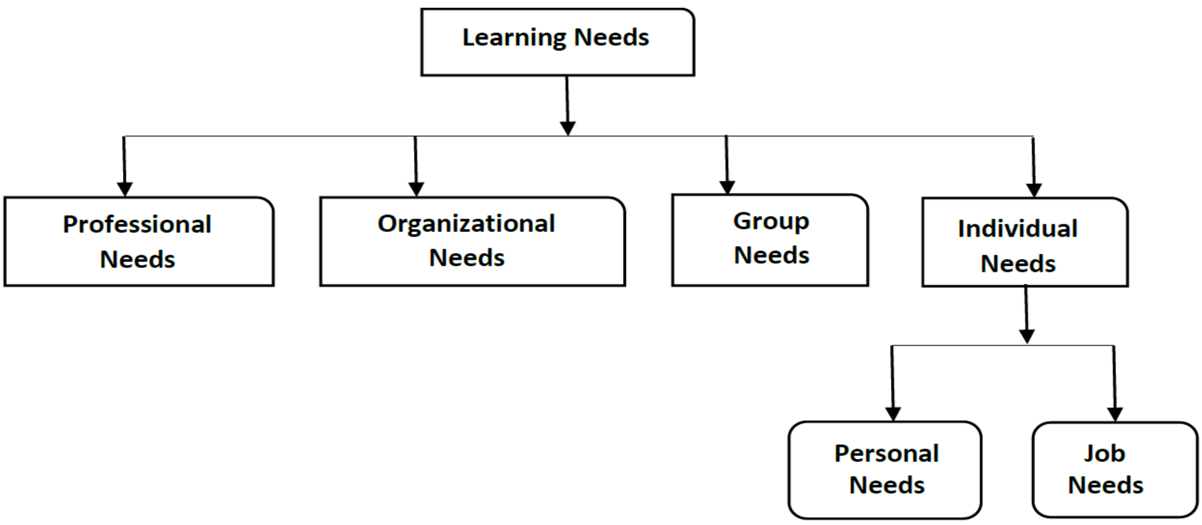

Figure 1.1 Sources of Learning Needs

- *Professional Needs:* The onus is on the profession to define the licensing requirements, standards and maintain the body of knowledge for the profession. The profession must ensure the safety of society. The profession ensures organizations and individuals are competent, ethical and professional in their day-to-day practices [https://www.apega.ca/enforcement/].
- *Organizational Needs:* The organization must ensure safe working conditions for its employees. Also, the organization must meet their corporate objectives, client satisfaction, financial and administrative requirements, and their employee needs. Further, the organization identifies any shortcomings in skill in the professional and recommends necessary training to help meet them.
- *Group Needs:* A group can define the norms and values of the community, its social needs, and even the learning needs of its members. Group needs could originate from a group comparing itself with other groups.
- *Individual Needs:* Each professional has their own needs. The needs of an individual could be personal needs or job needs. The personal needs are the needs that are identified by the professional based on the skills or expertise or communication ability he/she lacks, or social needs. Job needs are the needs required for the professional to perform efficiently on the job, based on tasks, skills, and standards needed for the job.

According to Dunning-Kruger (2011), learners themselves are limited in recognizing their own shortcomings. Regardless of where the learning needs come from, every professional will have needs of various forms. Learning needs are also distinguished along other dimensions as described by Ratnapalan and Hilliard (2002). Postgraduate medical education identifies the following types of learning needs:

- *Normative needs:* The gaps between the standards set by the profession and the individual's knowledge. These are the expectations of the profession for each professional to continue to be competent and up to date in their skills.
- *Prescribed needs:* The learning needs that require educational intervention as identified by the organization, educators or program planners by looking at the

normative needs defined by the profession. These needs are generated by the organization.

- *Perceived needs:* The needs identified by the individual learner about what they want to learn. Perceived needs are determined by the learner’s self-assessment as opposed to the *normative* and *prescribed needs* that are identified by the profession or educators.
- *Expressed needs:* The needs explicitly expressed by an individual. Not all *perceived needs* are *expressed needs*, as learners might be limited in communicating some of their needs due to barriers to expression. Examples of barriers to expression are lack of opportunity to show their needs, or lack of motivation or assertiveness to express their needs.
- *Comparative needs:* The needs identified by the learner by comparing their knowledge against the knowledge of their peers rather than against the standards defined by the profession. Comparative needs could also represent the needs of a group identified by comparing their group with another.
- *Unperceived needs:* The gaps in knowledge of the learner which they do not know they have. The *unperceived needs* could be generated from the profession, organization, group or even job needs.

For the *normative needs* and *prescribed needs* of the learner to be diagnosed normally the profession must define a body of knowledge. However, maintaining a body of knowledge could be difficult in a constantly evolving world, and determining gaps manually would be time-consuming. To help the profession automate these tasks is also difficult and would at least require an ontology and a diagnostic engine that could determine what a professional knows and doesn't know.

This research will propose an alternative to such knowledge and computationally intense methods that involves lightweight computation with no or minimal need for ontologies. To provide personalized support to users within an OLC, the data collected about users were mined to detect and meet their *unperceived* and *expressed needs*.

1.2 Research Questions

The main objective of this research is to *examine whether personalized support can be provided to professionals in detecting and meeting their learning needs as they interact with peers in online learning communities.* A support system for professionals must be able to adapt to the specific learning needs, styles and preferences, and learning goals of each professional even as such needs evolve. The research reported upon in this thesis provides insight into how to provide adaptive feedback to learners that would help meet their learning needs within an OLC. In addressing how to provide such feedback, the online activities of learners were tracked within an OLC to determine (using lightweight approaches) how to:

a. help learners detect their unperceived learning needs;
b. provide personalized support in meeting the individual learning needs of users;
c. track the evolving learning needs of learners.

The approaches adopted in this research are preventive rather than corrective. In employing a preventive approach, the ability to predict the quality of answer the user can provide will reduce the chances that his/her answers will be voted as weak. Likewise, the user who asked the question will have a higher chance of receiving answers of better quality. Also, by diagnosing the unperceived learning needs of users, learning resources could be recommended that would help in meeting those needs before they become evident to them. Finally, predicting a peer helper who can help answer a question on time can help reduce the answer response time. The better the learning needs of a learner can be detected early, the more the learner can be helped to meet these needs.

1.3 Overview of the Experiments Undertaken

This research extends previous studies (Ley et al., 2016) in supporting professionals by analysing the peer-to-peer interaction of learners in diagnosing their learning needs. Stack Overflow (SO) data were the experimental base for my research. SO is an example of an OLC used by millions[2] of programmers who ask questions and receive answers from peers. Mining such OLCs creates the opportunity to draw on some advanced learning technology

[2] Although SO is a large community, there are various sub-communities within it whose learning needs are narrower. Moreover, even though there are millions of users in SO, I mostly used active users for this study.

techniques, with an emphasis on learner modelling and personalization techniques. Techniques such as educational diagnosis, recommender systems, educational data mining, and intelligent help and tutoring were employed. Seven experiments have been carried out mining SO peer-peer interaction data with these goals:

- In the first experiment (as described in Section 4.1) the goal was to detect the learning needs of reputable question-answerers by mining their answer quality (Ishola and McCalla, 2016a). Unexpectedly, a decline in the answer quality by the reputable answerers was observed. A decline in answer quality could affect the help seeking experience of the question-askers.
- In experiment 2 (as described in Section 4.2) the goal was to study the effects of the help-seeking experience of question-askers on their ability to express their learning needs (Ishola and McCalla, 2016c; Ishola and McCalla, 2018c). Results of experiment 2 show that users who had negative help-seeking experiences had a large decrease in their propensity to request further help, hence the need to develop measures to mitigate against such a decrease.
- As a step in improving the help-seeking experience of question-askers, in experiment 3 (as described in Chapter 5), I predicted the ability of a question-answerer to provide a quality answer to a question (Ishola and McCalla, 2017a). In predicting the answer quality of question-answerers their past answer quality was employed, using a Naïve Bayes classifier. Prediction accuracies of 85% and higher were achieved.
- In experiment 4 (as described in Chapter 6), in supporting the learning needs of question-askers, I predicted the future learning needs of question-askers (Ishola and McCalla, 2016b). If a system can know what knowledge area a user would be asking questions about, then relevant questions and answers could be recommended to the user. Recall and precision values of 0.93 and 0.81 were achieved using a 5-month baseline of SO data in predicting the future learning needs of users.
- Aside from poor quality answers, question-askers in SO are experiencing longer response times to receive the first answer to their questions. Hence in experiment 5 (as reported in Section 7.1), I created measures to predict the question-answerers who will provide both timely and quality answers to questions in SO (Ishola and McCalla,

2017b). A success rate of 63% was achieved in predicting the actual first answerer among the top 20 predicted users (S@20). The success rates obtained in experiment 5 likely require further improvement if they are to be useful in a real-life peer recommender system.

- Hence, in experiment 6 my goal was to improve upon the success rate obtained in experiment 5 by refining my previous measures and by creating new measures (Ishola and McCalla, 2018a). In experiment 6 (as described in Section 7.2) I introduced a workload balancing approach by exempting answerers who recently provided help. With the workload balancing approach the goal was to ensure that answerers are well rested. In this experiment, I achieved a much higher success rate of nearly 90% while predicting well rested answerers even for questions that received their first answer after 8 days. The results obtained in experiment 6 show the prospect of building a peer recommender system that could with reasonable accuracy identify well-rested helpers.
- In the practical use of a peer recommender system inside an OLC, it is important to ensure active involvement of users by only recommending helpers for questions that would otherwise receive a late answer, or no answer. Therefore, in experiment 7 (as described in Chapter 8) my goal was to predict the questions that would be answered late or remain unanswered in SO (Ishola and McCalla, 2018b). Prediction accuracies of 60% and 68% respectively, were obtained in predicting questions that will be unanswered or answered late. These results likely still require further improvement to inform a recommender system usefully.

1.4 Contributions

This thesis contributes to research in advanced learning technology by exploring ways of supporting professional lifelong learners interacting in an online learning community. Specifically, this thesis contributes to research areas such as lifelong and professional learning, personalization, educational diagnosis, recommender systems, and intelligent help in these ways:

- First, this research extends the potential to support professionals beyond on the job training into an online learning community. Hence, this research opens the opportunity to support the millions of users relying on OLCs to meet their learning needs. Also, the research reported in this thesis creates the opportunity to judge the competency of a professional in the context of other professionals, mostly outside their own work place.
- Also, this research creates and tests lightweight approaches to detect the evolving learning needs of professionals within an OLC without using an ontology. The measures developed in this thesis show promise to predict the current and the future learning needs of professionals. Proactive approaches can be taken to provide personalized help tailored towards meeting the learning needs of each user.
- Further, this thesis creates methods that are scalable and can address, for example the ongoing increasing rise in the response time for the first answer to questions that can happen in OLCs as the user base grows (as happened in SO). The research reported in this thesis creates measures to reduce the response time to questions by recommending prospective answerers who can provide on time and quality answers just after the question is asked. A success rate of nearly 90% was achieved which greatly exceeds the success rate of about 23% achieved by Tian, Kochhar, Lim, Zhu, and Lo (2013) while predicting prospective helpers. This thesis advances previous research in intelligent help and peer recommender systems by providing methods that will help to ensure the helpers are well rested. Ensuring that helpers are well rested could create opportunities for inactive users to participate within the community.

1.5 Dissertation Outline

The dissertation is organized into ten chapters:

- Chapter 1 is the introduction to the dissertation and the background to the research. I present the concept of learning needs and the forms of learning needs. Further, I specify the research questions I am asking in this research, and the approaches and contributions of this thesis.

- In Chapter 2, I review work related to the research. I start with the literature on lifelong learning and the technologies used in supporting lifelong learners. I extend the discussion to a subset of lifelong learning: workplace and professional learning. I highlight the current approaches used in supporting professionals, and the current shortcomings of these approaches. I also discuss the need to extend the support provided to professionals to include professional learning networks. I focus the discussion on Stack Overflow (SO) as an example of such online learning community that support professional learning networks. I then narrow the discussion down to current research efforts employed in supporting users on Stack Overflow.
- In Chapter 3, I summarize Stack Overflow, the learning community I have used as a testbed for my research. I describe the reward systems in Stack Overflow and the current challenges experienced by users within the community. These challenges include an increase in the number of questions answered late or remain unanswered, and a decline in the quality of the answers to questions.
- In Chapter 4, I investigate the current issues experienced by users in SO that are barriers to peer-peer interactions within the community. Results from Chapter 4 show the need to track and detect the unperceived learning needs of question-answerers, as a decline in quality answers by reputable users were observed.
- In Chapter 5, I address the decline in answer quality as identified in Chapter 4 by predicting the answer quality of question-answerers using Naïve Bayes. The results obtained in this chapter show the promise to diagnose the knowledge of question-answerers within an OLC with no well-defined curriculum. The opportunity to diagnose the knowledge of question-answerers could help improve the help-seeking experience of question-askers by improving the quality of answers provided to their questions.
- In Chapter 6, I employ a proactive approach in predicting the future learning needs of users using the Bayesian estimation approach. To predict the future learning needs of users, I experimented using previous data about each user using short-term (5-month) and long-term (3-year) baselines. Lower precision and recall values were obtained in predicting the future learning needs of users using the long-term baseline compared to

the short-term baseline. The results from this chapter indicate that, mining the short-term information about the users is sufficient in meeting their learning needs, as with longer baselines their learning needs could have evidently evolved.

- In Chapter 7, I discuss the measures developed in recommending prospective answerers to questions, to reduce the response time for the first answer. I incorporated a workload balancing approach, which helps to ensure that the helpers recommended to provide answers to a question are well-rested. The results achieved in this chapter show the prospect of building a peer recommender system that can identify well rested helpers within an OLC.
- In Chapter 8, using the information available when a question is asked, I predict whether a question will be answered or not. If yes, I then predict if it will receive an answer early or late. This information could inform a recommender system that could suggest a peer who could intervene quickly to help answer a question that would otherwise be answered late or remain unanswered.
- In Chapter 9, I summarize how the results could inform the development of appropriate technology to support the evolving learning needs of professional learners.
- Finally, Chapter 10 presents a summary of this thesis, the research contributions, the limitations of this research and potential directions for future work.

CHAPTER 2
RELATED WORK

This chapter describes explorations of lifelong learning, lifelong user modelling and workplace and professional learning.

2.1 Lifelong Learning

Lifelong learning (LLL) is concerned with providing learners with the knowledge and skills needed to succeed in a changing world. According to Sharples (2000), this form of learning neither embraces the traditional educational system nor challenges it; rather it complements it. Drachsler, Hummel, and Koper (2008) present the major concepts of LLL, which are:

- Learning is a lifelong process which transcends beyond the childhood and youth stages of life.
- Learning is not limited to traditional school systems.
- Lifelong learners are responsible for their learning.

2.1.1 Forms of Lifelong Learning

Buntat et al. (2013), Tang and Kay (2013) and Laal and Salamati (2012) describe the three categories of lifelong learning:

- **Formal Learning**: Formal learning is the learning that takes place in a formal educational environment to obtain a qualification. Learning in this context is structured and organized.
- **Non-formal Learning:** Non-formal learning takes place outside the educational environment which may not necessarily result in any specific qualification. Non-formal learning includes individuals engaging in professional development, job-related training, technical courses, personal growth, and community organized programs.

- **Informal Learning:** Informal learning is unstructured learning which helps individuals learn day-to-day things. Informal learning occurs daily at work, home, leisure and elsewhere. Informal learning includes planned and unplanned learning. It is also called experiential learning or accidental learning.

Non-formal learning is a loosely structured form of learning that differs from formal learning because learners do not have a course syllabus, a learning curriculum and the certification associated with formal learning. Non-formal learning is more structured than informal learning which takes place as part of daily life activities.

2.1.2 Motives for Lifelong Learning

According to Fischer (2000), lifelong learning includes the application of knowledge and must consider the context of learning, rewards, opportunity for collaboration and support for learning. Carre (2000) describes two major motives for lifelong learning:

- An **intrinsic motive** is the satisfaction and pleasure expected to be derived from the learning. Intrinsic motives could be *epistemic*, *socio-emotional*, or *hedonic*. Epistemic means learners expect the learning activities to equip them with knowledge, skills and attitudes. A socio-emotional motive means that learners envisage the formation of new relationships. Learners with a hedonic motive are motivated to participate in learning because of the pleasure they will derive from it.
- An **extrinsic motive** is related to external rewards that learners believe they will gain from lifelong learning. These could be *economic* gains such as promotion or increases in wages. Extrinsic motives could also be *prescriptive* motives which means learners are pressured by peers or employers to participate in learning. Learners could also have *aspirational* goals to satisfy a professional need to remain competent or to gain social status.

In Carre's (2000) opinion these motives are temporary, changeable and contingent on the context of learning. Regardless of what learners' motives are for participating, lifelong learning should afford learners the opportunity of engaging in meaningful activities. Such activities should have provision for discussion and collaboration with other learners. Therefore, lifelong learning platforms should transcend just dumping course contents onto learners but should be open systems that allow learners to be co-developers (Fischer, 2000). In addition, lifelong learning platforms should encourage collaboration among learners to enhance teamwork. Also, lifelong learning platforms should support domain-oriented design environments (Sharples, 2000), where learners can set learning goals and modify learning contexts to suit their interests and problems.

2.2 Lifelong User Modelling

With the wide spread of mobile devices, the quantity of data generated by people has expanded enormously. Lifelong user modelling focuses on supporting lifelong learning by keeping information about the user over a long period. Hence, lifelong user modelling has the potential of aiding the user to achieve their long-term goals, and in detecting changes in the learning behavior of the user (Tang and Kay, 2013). Kay and Kummerfeld (2009) shared some goals of lifelong user modelling (LUM) which are described next.

2.2.1 Goals of Lifelong User Modelling

While the goals of LUM are specified in terms of users and general adaptive systems, they would also apply to learners and learning systems specifically. The following are the goals of lifelong user modelling as described by Tang and Kay (2013) and Kay and Kummerfeld (2009):

- *Supporting Long Term Personalization Goals*: Using the user's long-term goals, preferences, interests and prior knowledge, information and support could be personalized for them. In providing personalized support, the short-term goals, such as the user's current tasks, could be the focus.
- *Reusability*: Reusability allows parts of the model to be shared between different applications as exemplified in a personal user model cloud.
- *Self-Monitoring and Reflection*: A main goal of open user modelling is to create opportunities for users to reflect, plan and monitor learning goals. Opening the user model also allows the user to view the user model of other users, which gives room for competition among peers. Since the model is open, the user can also ensure their user model is correctly represented. In achieving these goals, a suitable interface is needed.

- *User Control*: While the user model might be open to the user and peers, the user should be able to control what part of the model can be shared with others. When to access the model and where it can be accessed is also controlled by the user.
- *Life Logging*: Another goal of LUM is life logging, that is gathering fine-grained information about user activities and logging such information over a long time. With life logging, the user model could be constructed from tracked data rather than relying on explicit information given by users.

2.2.2 Issues and Challenges of Lifelong User Modelling

The corresponding issues and challenges in realizing the goals of lifelong user modelling are discussed below, drawing on previous studies by Tang and Kay (2013), Kay and Kummerfeld (2009), and Dolog et al. (2009):

- *Issues in Supporting Long-Term Personalization Goals*: With the low cost of storage, information could be retained but issues arise such as privacy, control, users' understanding of their old model, and the possibility of forgetting past information. While recent information might be the most useful, some older artifacts might also be necessary in reminding users of past knowledge. In such circumstances, users might desire to explore their old user model.
- *Issues with Reusability*: Sharing the user model among several applications creates ontological challenges, standardization issues across the applications, and accessibility issues. Interpreting and using older contents of the model might not always be accurate. Another challenge is establishing how long the contents in the model can be reused. For instance, competency gained years back might not be relevant in the present context.
- *Issues with Self-Monitoring and Reflection*: Opening the user model requires creating an effective user interface, which the user can easily understand. Ensuring the user properly interprets the user model is an issue here. What the user thinks they know might not actually be known. Likewise, the system could also have difficulties ascertaining what the user knows because it's either outdated or what is to be learned has changed. Another issue is the accuracy of models in predicting real patterns of behavior rather than the many bogus patterns that often emerge from data mining.

- *Issues with User Control and Sharing*: Giving the user the right to determine what part of the model can be opened creates an issue of designing an effective interface which can support this goal.
- *Issues with Life Logging:* Deciding on the granularity of information to be logged is an issue. While fine grained data might provide detailed information, storing too much data could also result in the difficult problem of sorting out the relevant and useful data from noise.

2.3 Technologies for Support of Lifelong Learning

Information technology and learning has provided the possibility of developing personal technologies to support the lifelong needs of learners. Learners with mobile devices can access learning from any location, communicate with peers, and retrieve information to enhance learning (Sharples, 2000). This section identifies some of the existing learning systems used in supporting lifelong learning.

2.3.1 Personal Information Management Systems

In the digital world, learners store a lot of information, resulting in too much information for the user to manage. Indratmo and Vassileva (2008) in their survey paper stated that personal information management systems (PIMS) help lifelong learners in acquiring, organizing, and retrieving their personal information. Examples of personal information are e-books, emails, messages, contact lists, pictures, videos, appointments, calendars, bookmarks, and any other documents. Information in PIMSs could be arranged using hierarchical, flat, linear or network approaches, each of which has their merits and demerits. Information can be arranged in a tree structure, reducing the search space for information, although this might require more cognitive effort in remembering information. With a flat structure, learners can assign tags or keywords to items; this might be useful for lightweight retrieval of information and search for content. The problem with a flat structure is that it could lead to inconsistency when multiple tags are assigned to an item and further users might regard this as a tedious effort. Linear structure allows information to be arranged in a specific order whether alphabetical or chronological. Also, linear structure allows users to easily summarize information according to the order defined, although searching for information might be cumbersome where the information is large. In a network structure, learners can easily share information with peers although when the URL links are removed, broken link errors could be encountered.

2.3.2 Personal Recommender Systems

Recommending suitable learning activities that align with the individual learning goals of learners is the goal of developing personal recommender systems for lifelong learning (Fischer, 2000). In developing recommender systems to support lifelong learning, Drachsler, Hummel, and Koper (2008) identified the need to consider the context of learning, the learning strategy, and the evolving changes in learning needs of learners. One crucial aspect of learning context for a lifelong learner is learning at a distance and the learning strategy which would be the best fit for the learning process. The phases in cognitive development, preferred media of learning, and characteristics of the content to be learned should be considered when designing such systems. In effectively designing a recommender system, knowing the learning goal, prior knowledge, learning preferences, historically successful learning paths, and learning strategies are important.

Existing recommendation techniques can be classified as either model-based or memory-based (Adomavicius and Tuzhilin, 2005). Model-based techniques employ data mining algorithms to make predictions. Examples of such techniques are Bayesian models and neural networks (Hamalainen, Suhonen, Sutinen, and Toivonen, 2004). Memory-based techniques, on the other hand, continuously employ information about the users or items to offer recommendations. Such techniques employ user ratings to determine item and user similarities. Examples are collaborative filtering techniques, content-based techniques, and hybrid techniques. Collaborative filtering techniques recommend items used by similar users while content-based techniques recommend items similar to items used previously. The hybrid techniques combine both collaborative filtering and content-based techniques.

Recommender systems use historical information about the learner or learning resources to determine which recommendation technique would provide the most relevant recommendations. This leads to a “cold start” problem where an initial data set is needed for recommendation (Drachsler, Hummel and Koper, 2008). MovieLens, a recommender system for movies (Rashid et al., 2002) has addressed this problem by asking new users to rate their preferences for movies before the system can provide recommendations. Learners usually cannot rate learning activities in advance because they probably do not have adequate prior knowledge about the learning activity. More often, the learners' profiles are utilized for this cold start problem.

2.3.3 Personal Lifelong User Model Clouds

Dolog et al. (2009) provide an architecture for personal user model clouds which means all applications of a user are granted access to their user model when the user desires. The goal of having a personal user model cloud is to allow users to reuse their user model with several applications. The personal clouds could contain only a part of the user's complete model distributed across several applications. The user has control over what part of the model can be accessed by which application, and when and where such access can be granted. Accessing the user model with multiple applications creates a challenge of defining a representation for the user model in ways it can be reused by all applications. With cloud computing, a common approach in providing a common representation is the use of Service Oriented Architecture (SOA). Using the SOA architecture, the information about the user model can be represented as a service so different applications can access different part of the user model if desired (Kay and Kummerfeld, 2009). To access user model information across several applications, a semantic web approach has been suggested which means a common conceptual model exists as a standard for such integration (Klamma et al., 2007).

2.3.4 Life Annotation and Life-logging

Life annotation is accomplished by adding descriptive meta-data to resources. For instance, adding comments to a program or adding geographical information to an image file. Life annotations can help learners remember basic information about a past event such as date, time or location the event took place. Annotations could be structured in "top-down" or "bottom-up" schemes (Klamma et al., 2007). Top-down annotation schemes are based on automatic data generation functions such as indexing, storage, search and retrieval of information while bottom-up schemes allow learners themselves to tag and annotate learning resources.

An example of a top-down annotation system is the MACE[3] Project (Stefaner and Muller, 2007). The MACE project enriches learning repositories with metadata such as usage, competence, attention data, and metadata provided by annotators and authors. With this rich metadata, social recommendations can be provided, and relevant feedback can be provided to authors in improving their learning resources (Najjar, Meire and Duval, 2005). Social software approaches use automated metadata generation in providing social recommendations to lifelong learners (Klamma et al., 2007). Examples of bottom-up annotation are social tagging, bookmarking, folksonomies and personal mark-up annotation. Another example is Web based Intelligent Design System (WINDS) (Kravcik, Specht, and Oppermann, 2004). WINDS allows learners to annotate learning objects and discuss them. Using adaptive annotation techniques, recommendations are provided to the learners and communications among learners are enhanced within the learning system. Ultimately with WINDS feedback is provided to both instructors and tutors as well as learners. Beyond the classroom, a bottom-up approach has been employed in RAFT mobile applications (Kravcik, Specht, and Oppermann, 2004) that annotate the communication between learners at various locations. For instance, communication between learners in the field and those in the classroom could be annotated, and these annotations serve as aids to learners in reporting their experience in the field. Blogs and wikis are examples of collaborative tools used in supporting learners by allowing them to create, edit, comment and review learning objects that could be logged over a long time.

2.4 Workplace and Professional Learning

Rapid technological advances are leading to a massive on-going change in society and work, driving the need for lifelong learning (Simons and Ruijters, 2004; Sharples, 2000). Workplace learning is a form of professional learning associated with the daily involvement of professionals at work, usually in collaboration with peers, that often requires sharing of

[3] http://www.mace-project.eu

skills and knowledge (De Laat and Schreurs, 2013, Ley et al., 2014). This section describes some of the research efforts aimed at supporting professionals in the workplace.

2.4.1 Competency Modelling in the Workplace

Traditionally, the competency of professionals is modelled in their workplace and linked to organizational strategies or goals (Green, 1999). Competency modelling is usually performed through a controlled experiment whereby the behaviour of high performers is compared to that of average performers. The frequency of behavioural differences between these two groups is tested for statistically significant differences (Barrett and Depinet, 1991). Ultimately, a model is developed consisting of the competencies uniquely demonstrated by the high performers. Although this approach to competency modelling adequately identifies the specific behaviours of high performance users, it is very time and cost intensive (Athey and Orth, 1999).

The Knowledge Space Theory (KST) has been proposed in diagnosing the knowledge of professionals. With KST the domain knowledge is represented as a set of knowledge states which depict the tasks in the domain. With qualitative assessment, the knowledge of a professional is validated against expected answers which are empirically determined. According to Falmagne, Doignon, Koppen, Villano, and Johannesen (1990), the knowledge state of the user in solving the problem is represented by the subset of problems the user can solve correctly. The major advantage of KST is that it provides an efficient and economic assessment of knowledge as the knowledge states are validated by comparing the user's answer to empirically observed answer patterns. Korossy (1997) argues that a major shortcoming of KST is that it does not capture the underlying cognitive processes of the task, therefore it is impossible to transfer diagnosis to another task. Hence, Korossy (1999) suggests the need to incorporate the set of skills needed to solve the problems in KST as this would be useful in suggesting learning measures to the user.

Ley et al. (2008) argue that the limitation of the KST is the disconnection between the knowledge states and the actual tasks performed by professionals in real-life. Hence, KST leads to unclear assessments of competency. Ley et al. (2008) in their study model the competency of professionals using the Competence-based Knowledge Space Theory

(CbKST). CbKST extends the KST by mapping the set of skills to the problem to be solved in the work place environment. CbKST tracks the tasks performed by users as they engage in their job duties and the knowledge and skills required in successfully performing those tasks is assumed to be acquired by the professional. The CbKST approach poses a limitation in the workplace context as it is difficult to implicitly determine successful task performance. Also, the daily activities of professionals transcend beyond performing tasks alone to include searching for knowledge, contacting others or helping others. Hence, capturing task performance alone does not capture all information about the knowledge of the professional.

Ley and Kump (2013) in their study employed Knowledge Indicating Events (KIEs) in modelling the competency of professionals. The KIE approach is a non-invasive approach to diagnosing the knowledge of professionals in an adaptive work-related learning environment. With KIE, the behaviour of professionals within the system is tracked to provide evidence of their skills and knowledge. To model the competency of the professional, KIE uses certain behaviours as indicators, such as clicks to request help, keystroke and scrolling behaviours. Also, learning artefacts viewed, annotations made, successful task performance, help provided to peers, peers contacted, and hints requested are all tracked. KIE has limitations. First, emotional states of professionals could interfere with their performance (Ley et al., 2008), even down to their keystroke and scrolling behaviours. Besides, using KIE leads to collecting too much detailed information about a user, even though only a small fraction of the information collected would eventually be useful for adapting and personalizing the system (Drachsler, Hummel and Koper, 2008).

Athey and Orth (1999) have identified the key components of competency models to include the following:

- knowledge required for current job performance;
- emerging knowledge required for future success;
- knowledge of best practices within the profession;
- knowledge that enhances organization performance; and
- knowledge that provides competitive advantage to organizations.

To meet changing organizational needs, Athey and Orth (1999) suggest the need for information systems to be adapted to the emerging individual needs of professionals and the profession. Tracking the knowledge of professionals can aid in modelling their competency and in diagnosing gaps in their knowledge.

2.4.2 Future Work in Supporting Lifelong Professional Learners

Traditionally, professional learning has been accomplished through a combination of on the job training, short courses, and self-directed learning. Ley et al. (2014) have suggested that applying technology in supporting workplace learning would create opportunities for reflection and support the emergence of knowledge within the discipline. Applying techniques from social media mining, learners' knowledge diagnosis, and personalization in workplace learning has not been a focus of research (Ley et al., 2016). Even though such techniques would be beneficial for knowledge management among professionals, and for support of human resource management and organizational learning.

In enhancing workplace learning, adaptive technologies could be applied in creating workplace learning solutions, and in facilitating self-regulated learning. Grazyna Bartkowiak (2014) performed a user study to determine ways of enhancing workplace learning. Results obtained from this study show that employees identified ineffective communication, lack of practical experience, and poor business management as major barriers to learning. Employers identified inefficiency of staff, inability to combine theory and practice, and lack of experience in the organization as barriers. Bartkowiak (2014) concludes that high self-awareness and ability to apply theoretical knowledge to practical problems are important in enhancing workplace learning.

Previous studies on modelling the competency of professionals have been focussed on work-integrated learning. Modelling competency based on work-integrated learning is not sufficient to judge professionals' overall competence in their profession (Athey and Orth, 1999). Ley and Kump argue that competency will at most be judged compared to the staff of the organization rather than with professionals in society at large (Ley and Kump, 2013). Another shortcoming of modelling the competence of professionals within the workplace is that it will likely require only a subset of the skills they need to be fully capable. Learning in the workplace context rarely scales beyond the immediate context of the workplace (Ley et al., 2014).

In scaling up workplace learning, Ley et al. (2014) have identified some of the challenges professionals face in participating in informal learning:

- applying workplace norms to daily practice,
- extending good work practices among professionals and organizations,
- detecting learning needs as they emerge within the profession, and
- documenting learning experiences that could allow for reflection.

De Laat and Schreurs (2013) identified the implicit and spontaneous nature of informal learning as posing challenges to capture and analyse traces of learning in the everyday life of professionals. The opportunity to capture information about professionals would allow the competence of professionals to be examined beyond their workplace. With the emergence of learning networks among professionals facilitated by information and communications technology, this possibility is now a reality. However, within the advanced learning technology research community modelling the competency of professionals as they seek and receive help in an online learning network has not been looked at. Modelling the competency of professionals in such learning networks opens up the possibility to judge the competency of a professional in the context of other professionals. More importantly, this creates the opportunity to model the competency of professionals even as knowledge emerges within their profession. Further, this could be done on a large scale.

2.5 Learning Networks

Learning networks (LNs) are distributed sets of learners who collaborate to create and share knowledge to develop their competence in a discipline (Koper et al., 2005). LNs allow lifelong learners to update their skills and competence, share their knowledge and support other learners. LNs are key in supporting lifelong learners because they connect distributed learners together providing a virtual community of learners in a discipline. The designs of LNs are usually flexible and learner-centric with learning activities provided by different users in the network (Drachsler, Hummel and Koper, 2008). Users can create, modify, delete and rate learning activities within the LNs. LNs differ from virtual learning environments in that learning activities are controlled by contributions of their members instead of having education institutions design them. Also, LNs provide avenues for learning resources from different sources such as schools, companies, libraries and learners themselves to be linked and made available. Getting learning resources from different sources implies the need to organize learning resources. Koper et al. (2005) employed software agents and open learning technology standards to create an interoperable network for the different sources of the learning resources. LNs include self-directed learning, learning communities, and agent technologies. Learning networks can be formed within social networks which consist of social entities such as groups of people, organizations, and communities. Learners within the learning networks share social relationships as friendships, acquaintances, or linkages based on being co-workers or peers (Garton, Haythornthwaite, and Wellman, 1997).

2.5.1. Professional Learning Networks

Professional learning networks support professionals in developing and maintaining social relationships to help their learning and professional development. For effective professional development, learning must be a normal part of day-to-day work that is not isolated from work but instead requires daily participation among professionals (Boud and Hager, 2012). Lohman (2006) has argued that the benefits of such interactions between peers in informal social learning environments could be a key in carrying out work related tasks and in acquiring new knowledge.

The goal of networked learning communities is to explore how these social relationships aid in professional development and how they can be maintained. Relationships among peers range from weak relationships, acquaintances, to long lasting relationships with peers. Analysis of social networks has shown that weak ties are useful for acquiring new knowledge (Jones, Ferreday, and Hodgson, 2008). Strong ties are useful to embed knowledge related to daily-shared practices and to enhance commitment among peers. Communication could be in interactions between colleagues via email, telephone or Skype conversations with peers; or through social media (De Laat and Schreurs, 2013).

Analysing posts from discussion forums and online peer-peer interactions is important in detecting learning behaviours and in creating awareness about learning activities within the community (Siemens, 2010). De Laat and Schreurs (2013) carried out a study among teachers from 70 schools with an aim to understand the learning processes among the professionals in a naturalistic setting using:

- social network analysis (SNA) to determine 'who is talking to whom',
- content analysis (CA) to determine 'what they are talking about', and
- contextual analysis (CxA) to find out 'why they are talking as they do' in the context of their workplace.

The study was carried out using a network awareness tool that allows for visualizations of social networks based on learning activities professionals engage in. With SNA it is possible to help professionals to form learning partnerships on issues that currently matter. CA reveals the learning needs of professionals, while CxA provides insight as to how professionals learn together based on work related issues.

2.5.2 Current Approaches in Supporting Users in Learning Networks

With the proliferation of social media, several knowledge resources and computer mediated technologies exist that support lifelong learners in their day-to-day activities. Examples of such forums include:

- Twitter (https://twitter.com),
- Facebook (https://www.facebook.com/),
- Myspace (https://myspace.com/),
- Stack Overflow (http://stackoverflow.com/),
- Quora (https://www.quora.com/),
- Flickr (https://www.flickr.com/),
- Wikipedia (https://en.wikipedia.org/wiki/Main_Page),
- LinkedIn (https://www.linkedin.com/) amongst others.

The success of social media mining can be attributed to the vast data available about the users from online forums. Research efforts emerging from mining such social media forums include sentiment analysis which analyses learners' attitudes, sentiments, and emotions from discussion forums and blogs using natural language processing (Liu, 2012). Also network analysis is performed to understand the connections among users (Haythornthwaite and De

Laat, 2010). In addition, social influence analysis (Tang et al., 2009) and engagement of users in online forums (Attfield, Kazai, Lalmas, and Piwowarski, 2011) have been studied. Studies on reciprocity in social forums (Kwak, Lee, Park, and Moon, 2010); tag prediction (Heymann, Ramage, and Garcia-Molina, 2008); social recommender systems (Song, Zhang and Giles, 2011) have been used to provide personalized support to users.

Specifically, in mining the Stack Overflow (SO) online learning community for programmers, Treude, Barzilay, and Storey (2011) classified the various questions asked in SO as:

- questions asking for instructions on how to perform a task,
- questions related to environment issues after deployment,
- questions about error messages,
- questions aiming to know more about a concept,
- questions asking for code review from peers, and
- questions related to performance issues.

Investigation by Treude et al. (2011) shows that questions related to code reviews and questions asking for how to do a task are more likely to be well answered. This is because these types of questions can have more than one correct answer. Questions in other categories get poorer answers or are even unanswered. Nasehi, Sillito, Maurer, and Burns (2012) highlighted the characteristics of good answers in SO as answers with concise, clear explanations and containing code examples that follow best practices. Asaduzzaman, Mashiyat, Roy, and Schneider (2013) attempted to provide answers to the reasons questions are unanswered, largely by extracting features from the question. The content of the question asked, the length of the question asked, the title length, the readability of the question, and the presence of code in the question are indicators of how soon a question will receive its first answer. Low precision and recall values of 0.35 and 0.403 were achieved by Asaduzzaman et al. (2013) when they tried to use these factors to predict the response time to receive the first answer. The low precision and recall values achieved by Asaduzzaman et al. (2013) would require further improvement for real-life application.

In addressing the increasing number of unanswered questions, Bhat, Gokhale, Jadhav, Pudipeddi, and Akoglu (2014) stated the importance of assigning appropriate tags to questions. Stanley and Byrne (2013) applied a Bayesian probabilistic model to predict tags for questions in SO and achieved an accuracy of 65%. Similarly, Tian et al. (2013) predicted the best answerers to questions using a topic modelling approach. The study achieved a success rate of 21.5% in predicting the actual best answerer among the top 100 predicted answerers who could answer the question.

Previous studies have also analysed the SO dataset to identify learning trends among the programmers using SO. Barua, Thomas, and Hassan (2014) used a Latent Dirichlet allocation approach to model the main topics discussed by users in SO. There is an increasing trend to topics related to web development corresponding to recent interest in mobile application development. In supporting the programmers' learning needs, Ponzanelli, Bavota, Di Penta, Oliveto, and Lanza (2014) propose the use of an IDE. With the IDE, programmers can automatically retrieve answers from SO which relate to the context of their work. With this approach, programmers will spend less time searching for information on their own.

Vasilescu, Filkov, and Serebrenik (2013) investigated the association between a user's participation in SO and their rate of software development, as indicated by the number of code changes committed to GitHub. The study shows that users who actively commit code to GitHub ask fewer questions in SO but answer more questions, while active SO question-askers have less code committed to GitHub. Also, the association of gender with the online participation of users in SO was studied (Vasilescu, Capiluppi, and Serebrenik, 2014). Their study reveals that men are more active and therefore earn more reputation points than women earn. Further, Bazelli, Hindle, and Stroulia (2013) analysed the personality traits of users in SO using the Linguistic Inquiry and Word Count (LIWC) instrument. They found that users with higher reputation score are more extroverted when compared to users with lower reputation score.

Some of the results from the previous studies by Asaduzzaman et al. (2013) and Bhat et al. (2014) and Tian et al. (2013) require further experimentation before practical use of their approaches to be considered. Hence, my thesis research develops lightweight measures that aim to improve the results from the studies mentioned above. Also, I aim to adopt a lightweight approach that would allow tracking of the learning needs of users even as they change. In addition, my research aims for a preventive approach rather than a corrective approach whereby the learning needs of users can be diagnosed early, perhaps even before the users themselves become aware of them. To achieve these goals, I need to identify the challenges experienced by users using SO. The next chapter presents a more detailed overview of SO, an example online learning community and some of the challenges experienced by professional learners in this community.

CHAPTER 3
ONLINE LEARNING COMMUNITY: STACK OVERFLOW

Many software professionals are part of online communities that help them to stay up to date and to overcome problems they may encounter in their professional lives. Most users in these support communities are learners helping each other to resolve their learning needs. The overall goal of my research is to provide personalized support to such learners as they interact with peers in such "learning communities." My research seeks to provide adaptive feedback to learners that would help support their learning needs within an online learning community (OLC). As the experimental base, I turned to the Stack Overflow (SO) question and answer forum, which has been used for years by programmers. As of December 31, 2017, SO had 8.4 million users which is a significant growth from the 4.9 million users as of December 31, 2016. Overall, only about 43% of SO users have provided at least one question or one answer post.

3.1 Overview of Stack Overflow

Stack Overflow is a "question and answer site for professional and enthusiast programmers" [http://stackoverflow.com]. This section summarizes the SO community.

3.1.1 Posts

The two forms of posts in SO are question and answer posts. Users in SO can vote up or down questions and answers. All questions are tagged in SO to indicate the subject area the question falls under as shown in Figure 3.1. Tags are employed in SO as a word or phrase to describe the question being asked which also helps answerers to determine the questions they can answer. While creating a question, a maximum of five tags and a minimum of one tag can be employed. For instance, the question depicted in Figure 3.4 has three tags "*ios*", "*osx*" and "*swift*".

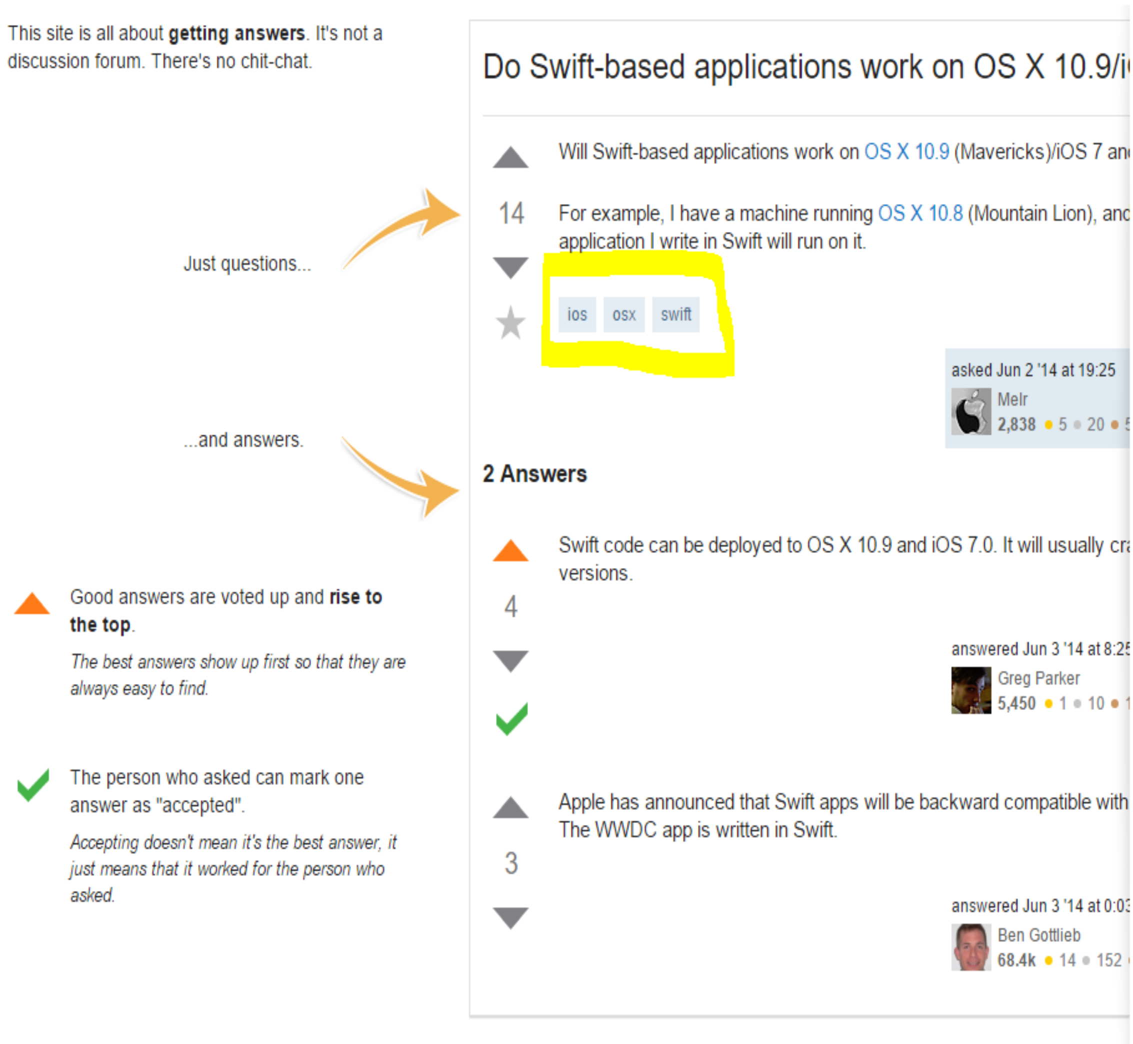

Figure 3.1. Illustration of a Question and Answer in Stack Overflow (Adapted from http://stackoverflow.com/tour)

Further, the person who asks the question can mark one of the answers given as accepted; the right sign in Figure 3.1 signifies this. Posts in SO differ also in popularity based on the number of views received by the post. Both the number of views by Stack Overflow users and anonymous users who are not Stack Overflow users adds up to the total number of views of a post. To ensure the quality of posts in SO, posts can be flagged as *spam*, *abusive*, *duplicate*, *off-topic*, *too-broad*, or a *very low-quality* post. Figure 3.2 below explains the reasons why posts can be flagged.

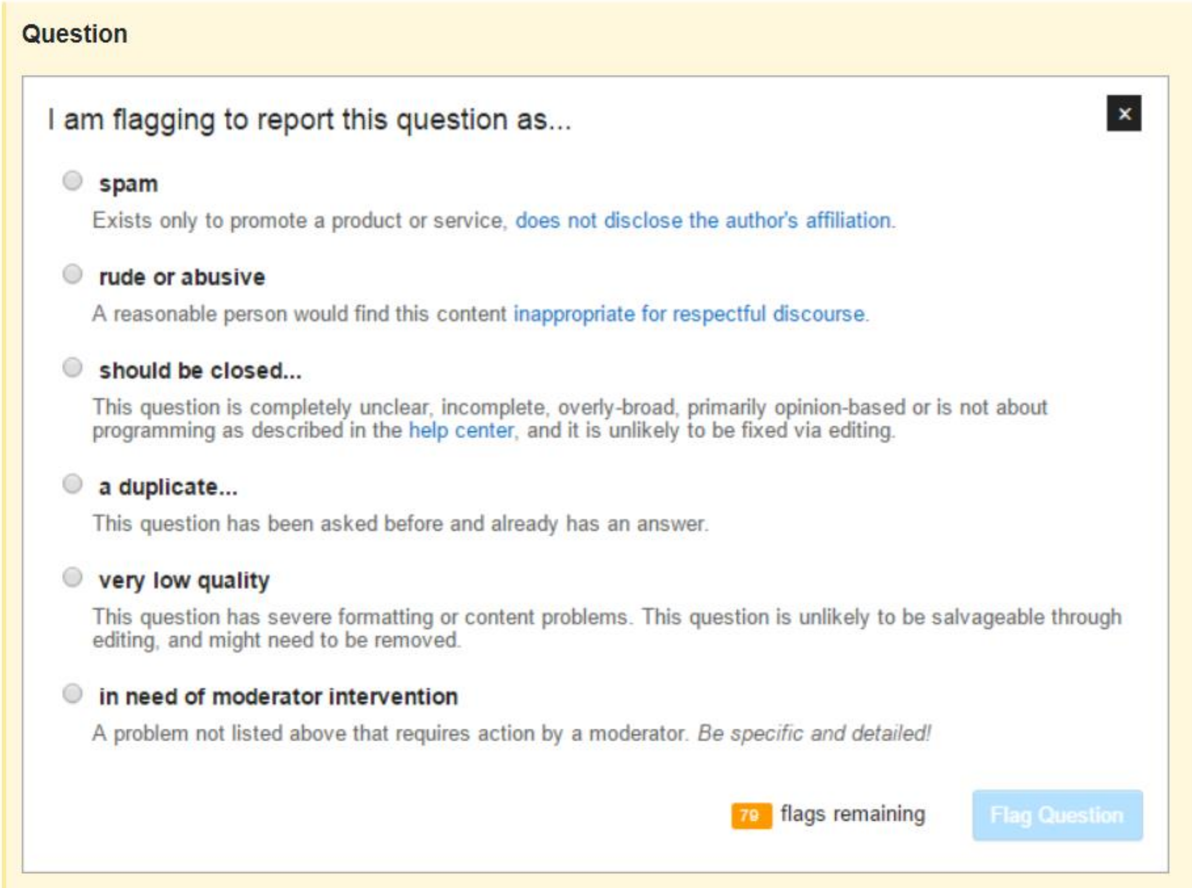

Figure 3.2. Post Flagging in Stack Overflow

(Adapted from https://stackoverflow.com/help/privileges/flag-posts)

Posts flagged by any of the reasons described in Figure 3.2 are brought to the attention of the moderator (as described in Section 3.1.3). A post flagged by at least 3 users would be banished from the front page. Likewise, a post flagged by at least 6 users would be deleted and the owner of the post would lose 100 from their reputation score. The next section summarizes how users gain and lose reputation score.

3.1.2 Reputation Score

Users in SO can earn reputation score for their participation in question asking, for providing answers and for editing question or answer posts. In SO reputation score can be earned or lost from participating in any of the online activities as shown in Table 3.1 below.

Table 3.1. Reputation Score in Stack Overflow

Online Activity	Score Earned
Question asked is voted up	+5
Answer provided is voted up	+10
Answer provided is marked "accepted"	+15
Question-asker accepted an answer	+2
Edit provided to a post that is accepted	+2
Question asked is voted down	-2
Answer provided is voted down	-2
User who votes down an answer	-1
User posts receive 6 spam or offensive flags	-100

Besides the online activities depicted in Table 3.1 reputation score can also be earned in SO when a *bounty* is awarded to a question. A *Bounty* is a special reputation score awarded to a question to help draw the attention of question-answerers to answer the question. To award a bounty to a question a reputation score between 50 and 500 must be awarded. Likewise, the user who awards a bounty to a question would lose the bounty amount they awarded from their reputation score. The aggregate of all reputation scores earned by a user for his/her online activities in SO is presented in their user profile.

3.1.3 Privileges

As users earn reputation score in SO, they gain more privileges in SO. The privileges in SO can be categorized as described below:

- *Creation*: Users with creation privileges in SO can create questions, answers, and tags in SO. This is the most basic privilege and it is available to all users in SO regardless of their reputation score.
- *Communication*: Users with communication privileges can dialogue with other users in SO through the chat rooms, or even provide comments to posts. Users need to have earned at least a 20-reputation score to participate in a chat discussion. Further,

comments in SO serve as notes to question or answer posts. A user must have at least a 50-reputation score to provide comments to posts in SO.

- *Moderation:* Moderation privileges allow users to peer review posts and to suggest new features to be incorporated in the community. Various moderation roles require different reputation scores; for instance, basic activities like to vote up posts or to flag posts (15 reputation score), to vote down posts (125 reputation score), to edit posts (2000 reputation score). These basic activities could be reviewed by moderators in SO. Moderators in SO are experienced users who can delete and undelete posts, edit posts without going through peer review, and review suggested edits. To gain moderator privileges in SO requires the user to have at least a 10,000-reputation score.
- *Milestone*: A milestone badge is a special privilege provided to users for their great service to the community, i.e. users with a reputation score of at least 25,000. With the milestone privilege such users have access to internal and Google site analytics for SO.

3.1.4 Badges

Badges can also be awarded to users for their activities within the community. There are seven main categories of badges in SO as shown in Table 3.2 below.

Table 3.2. Badge Categories in Stack Overflow

Badge Type	Description
Question Badge	Earned by participating in question asking activities.
Answer Badge	Earned by providing useful answers to questions.
Moderation Badge	Earned by reviewing posts of other users.
Participation Badge	Earned by providing comments to posts.
Tag-based Badge	Earned by showing high participation in answering questions related to a tag.
Documentation Badge	Earned by reading the documentation tour of the Stack Overflow welcome page.
Other Badge	Earned by completing the Stack Overflow survey.

Examples of question badges, answer badges and tag-based badges as shown in Figure 3.3 – 3.5. The first column is the badge name, the second column is the description of the badge and the last column shows the number of badges awarded. For instance, as shown in Figure 3.2, the *Scholar* badge is awarded to users who ask a question and accept one of the answers provided as useful.

Altruist	First bounty you manually award on another person's question	8.4k awarded
Benefactor	First bounty you manually award on your own question	40.3k awarded
Curious	Ask a well-received question on 5 separate days, and maintain a positive question record	291.8k awarded
Inquisitive	Ask a well-received question on 30 separate days, and maintain a positive question record	29.1k awarded
Socratic	Ask a well-received question on 100 separate days, and maintain a positive question record	3.5k awarded
Favorite Question	Question favorited by 25 users	43.9k awarded
Stellar Question	Question favorited by 100 users	6.4k awarded
Investor	First bounty you offer on another person's question	16.8k awarded
Nice Question	Question score of 10 or more	509.5k awarded
Good Question	Question score of 25 or more	165.8k awarded
Great Question	Question score of 100 or more	28.1k awarded
Popular Question	Question with 1,000 views	3.9m awarded
Notable Question	Question with 2,500 views	1.9m awarded
Famous Question	Question with 10,000 views	553.5k awarded
Promoter	First bounty you offer on your own question	70.9k awarded
Scholar	Ask a question and accept an answer	1.6m awarded
Student	First question with score of 1 or more	2.0m awarded
Tumbleweed	Asked a question with zero score, no answers, no comments, and low views for a week	1.0m awarded

Figure 3.3. Question Badges in Stack Overflow

(Adapted from https://stackoverflow.com/help/badges)

Likewise, as shown in Figure 3.4, the *Great Answer* badge is awarded to question-answerers whose answers earn a score of 100 or more.

Enlightened	First to answer and accepted with score of 10 or more	338.2k awarded
Explainer	Edit and answer 1 question (both actions within 12 hours, answer score > 0)	57.7k awarded
Refiner	Edit and answer 50 questions (both actions within 12 hours, answer score > 0)	1.8k awarded
Illuminator	Edit and answer 500 questions (both actions within 12 hours, answer score > 0)	98 awarded
Generalist	Provide non-wiki answers of 15 total score in 20 of top 40 tags	1k awarded
Guru	Accepted answer and score of 40 or more	112.2k awarded
Nice Answer	Answer score of 10 or more	1.1m awarded
Good Answer	Answer score of 25 or more	353.6k awarded
Great Answer	Answer score of 100 or more	61.6k awarded
Populist	Highest scoring answer that outscored an accepted answer with score of more than 10 by more than 2x	17.5k awarded
Reversal	Provide an answer of +20 score to a question of -5 score	294 awarded
Revival	Answer more than 30 days after a question was asked as first answer scoring 2 or more	337k awarded
Necromancer	Answer a question more than 60 days later with score of 5 or more	473.8k awarded
Self-Learner	Answer your own question with score of 3 or more	126.6k awarded
Teacher	Answer a question with score of 1 or more	1.3m awarded
Tenacious	Zero score accepted answers: more than 5 and 20% of total	49.5k awarded
Unsung Hero	Zero score accepted answers: more than 10 and 25% of total	19.4k awarded

Figure 3.4. Answer Badges in Stack Overflow

(Adapted from https://stackoverflow.com/help/badges)

Users who have actively participated in providing answers to questions related to a given tag are awarded a tag-based badge as shown in Figure 3.5. As shown in Figure 3.5, badges in SO are further categorized at three levels: “*Bronze Badge”*, “*Silver Badge*” and “*Gold Badge”*. All badges in the seven categories described in Table 3.2 could also be classified into the three badge levels.

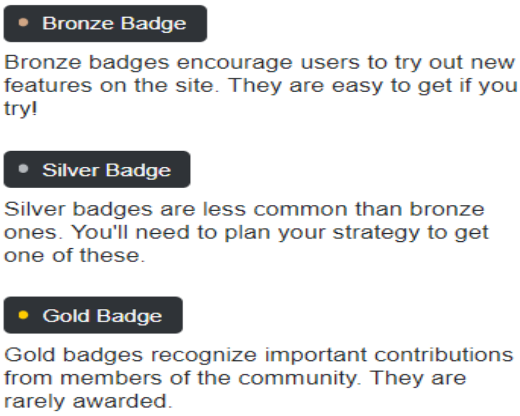

Figure 3.5. Tag-based Badges in Stack Overflow

(Adapted from https://stackoverflow.com/help/badges)

A "*Bronze Badge*" is awarded to users to encourage other users to try new features in the forum while the "*Silver Badge*" requires strategic planning by users who earn them. The "*Gold Badge*" which represents the highest level of a badge is rarely awarded and is only awarded to recognize important contributions by the user. Typically, a bronze tag-based badge is awarded after providing at least 20 answers with a total score of 100 or more. A silver tag-based badge is awarded after answering at least 80 questions with a total score of 400 or more. A gold tag-based badge is awarded after answering at least 200 questions with a total score of 1000 or more. Whichever category of badge is awarded to a user, the aim of the badge is to encourage the user to continue participating in useful and important activities within the community.

3.2 Challenges in Supporting the Learning Needs of Users in Stack Overflow

Despite these various reward mechanisms (either earning reputation score or privileges or badges), users in Stack Overflow still experience challenges meeting their learning needs. I have tracked the activities of SO users by looking deeply at the forum. I mined the question and answer activities of users in SO, the tags employed by users, and the badges earned as proxies in studying their learning needs within SO. Some of the challenges I have identified which are experienced by users in achieving their learning needs are discussed below along with my analysis of SO data that illustrates the significance of these challenges.

3.2.1 Evolving Learning Needs

With millions of users in SO having different learning needs, in providing personalized support to users, it is vital to detect their current learning needs. Since tags attached to a question likely indicate the knowledge required to answer the question, therefore tags can be employed as a proxy for the knowledge that the user needs to know. In inferring the learning needs of an individual user, it is crucial to classify the tags employed by a user into suitable computing related classes. Using the approach exemplified by Wang, Lo, and Jiang (2013), I defined 19 computing related classes. Some examples of the computing-related classes with corresponding tags mapped to them are shown in Table 3.3.

Table 3.3. Tag Classification

Computing Class	No. of Tags	Example of Tags
General Computing	31,409	rounding, detection, vocabulary
Software Coding	1,567	java, c#, python
Web Application	1,293	php, html, asp.net
Mobile Application Development	951	android, ios, blackberry
Database Systems	731	mysql, sql, oracle
Framework and Library	440	.net, playframework, zend-framework2
Operating Systems	330	windows, powershell, active directory

I was interested in tracking changes in the learning needs of SO users over the years. A computing class displayed in Table 3.3 above was assigned (by me) to each question based on the tags that constitute the question post. Where more than one computing class could be inferred based on the disparity of tags that forms the question, the computing class with the highest number of tags is assigned to the question. A baseline of SO questions asked from 2009 to 2017 (4.14 million questions) was used to study how the knowledge that users need to learn has changed over the years. The three top computing classes: *software coding*, *web* *application*, and *mobile application development* were considered for this study. For this study, the five most used tags for the top three computing classes were considered to be indicative of possible trends in the evolving learning needs of users. For each computing class, the percentage of questions asked about the five most used tags shown in Figure 3.6 were computed. Results from the analysis are shown in Figure 3.6.

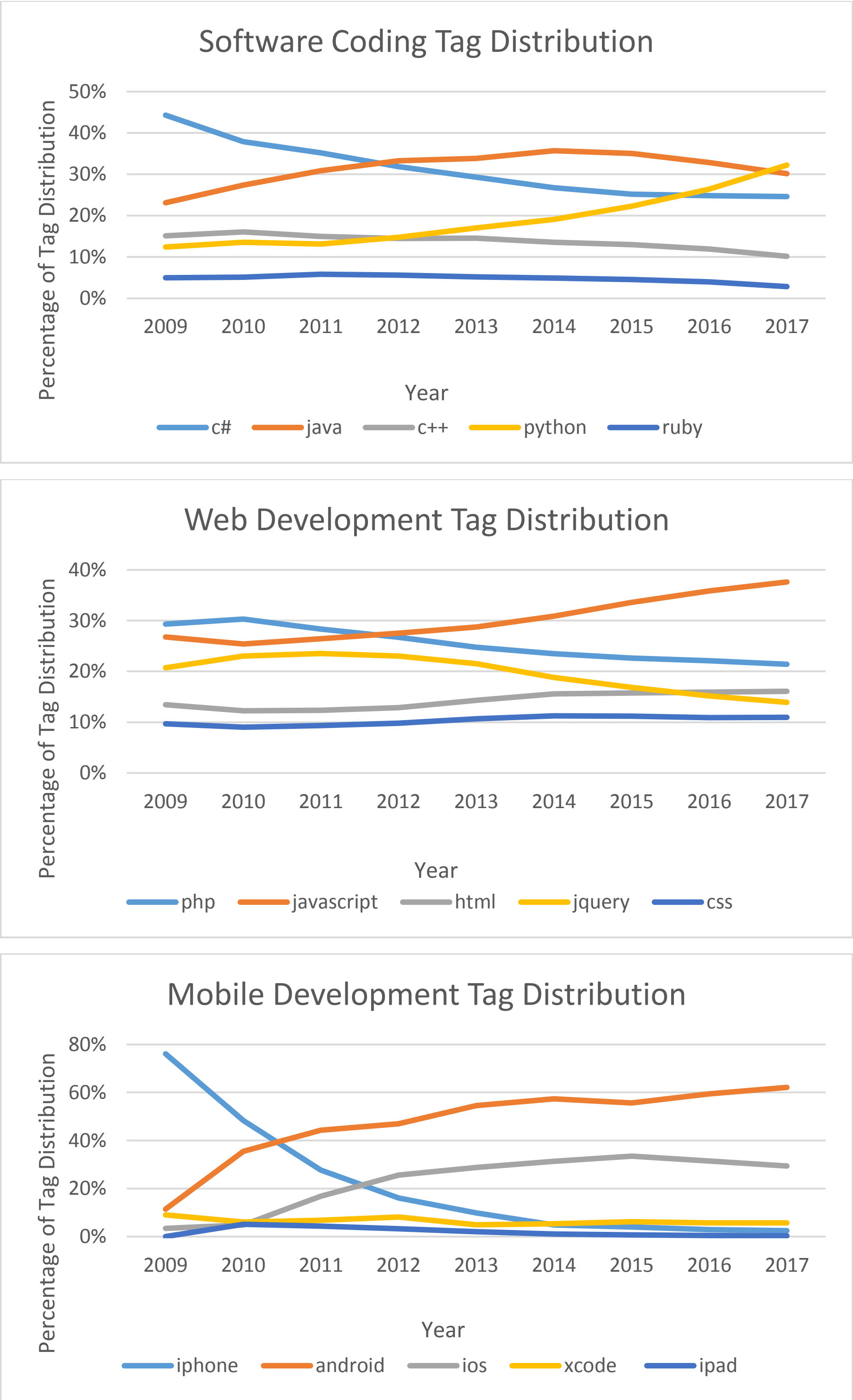

Figure 3.6. Evolving Learning Needs within the Computing Classes

Changes in the learning needs of the mobile application developers are revealed by their increasing interest in learning android as shown in Figure 3.6. The widespread adoption of Android phones from 2010 could be linked with the continued popularity of the java, javascript, android tags among users within the OLC. As technology will continue to drive changes in every profession, there is a need to ensure that the support provided to professionals adapts according to the evolving learning needs of the users and the profession (Ishola and McCalla, 2016b).

3.2.2 Increase in the Question Response Time

As the learning needs of users evolve, questions asked become more diversified. In cases where a user does not receive answers from their peers, the user could provide answers to their questions or sometimes, the questions remain unanswered. Key to the success of an OLC is the ability of users to receive prompt answers to their questions (Bhat et al., 2014). Recently, the answer response time of questions in SO has significantly increased. In addition, the distribution of questions answered by question-askers themselves and the proportion of unanswered questions have increased. The median response time[4] for first answer (RTFA) has risen from about 15 minutes in 2009 to 38 minutes in 2017. As of 2017 the average RTFA is about 4 days. The wide gap between the median and the average RTFA indicates that while some questions get answered early, other questions receive their first answer very late. Figure 3.7 shows the answer response time of questions for each of six defined RTFA time intervals for each year under consideration. In studying the changes in the response time to questions, about 13.99 million questions asked from 2009 to 2017, that received at least one answer were considered for this study.

[4] It should be noted that in computing the response time, I only considered the questions that have received at least one answer. The response time for questions which are yet to receive any answer is undefined.

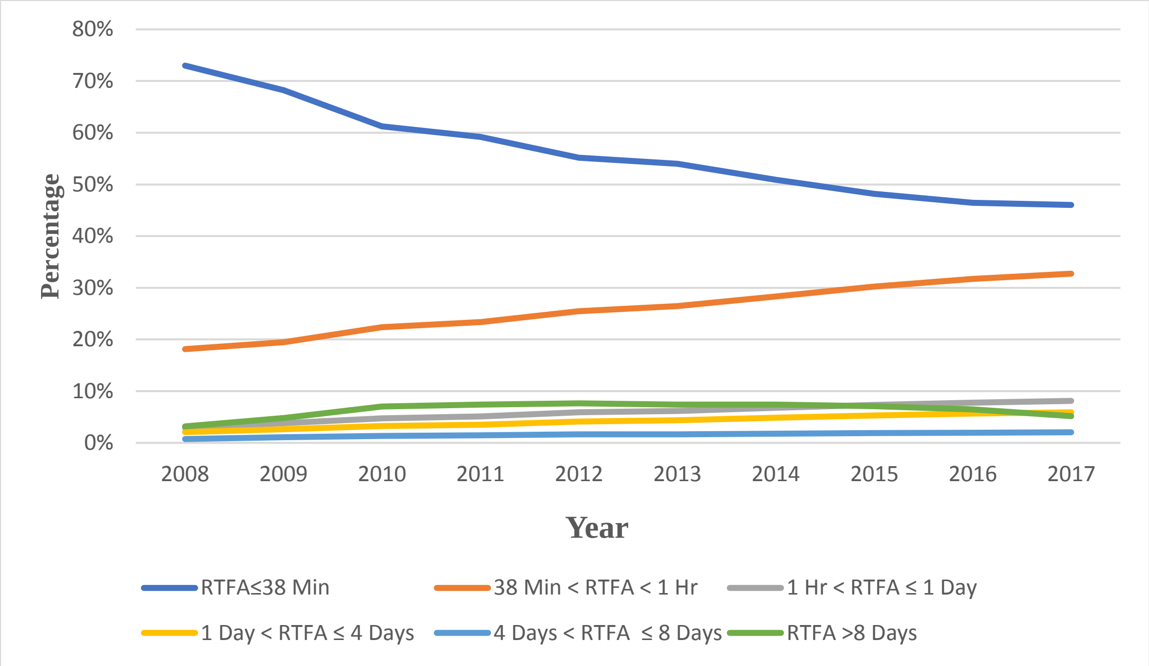

Figure 3.7. Response Time for First Answer

Figure 3.7 shows that the highest proportion of questions in SO are answered within 38 minutes (this is the median RTFA in SO), although a continuous decrease over time is also observed. Also, questions with response times above 38 minutes have continually increased. Even if the decline in the RTFA could be because of the increase in the number of questions asked over time, there is still the problem of helping users meet their learning needs on time.

3.2.3 Increase in the Number of Self-answered or Unanswered Questions

Some questions that received late answers were answered by the question-askers themselves. The year by year count of the questions answered by the question-askers themselves is shown in Table 3.4 (Ishola and McCalla, 2017a).

Table 3.4. Questions Answered by the Question-asker

Year	Count
2009	35,469
2010	71,051
2011	104,038
2012	162,954
2013	204,839
2014	225,256
2015	251,678
2016	252,075
2017	239,155

Specifically, the total number of questions self-answered by the question-askers has increased from 35,469 in 2009 to 239,155 in 2017 as shown in Table 3.4. Another challenge in supporting users in SO is the increase in the number of unanswered questions. Specifically, the number of unanswered questions has increased from 1,219 in 2009 to 685,965 in 2017. This implies a rise in the percentage of unanswered questions from 0.4% to 28% between 2009 and 2017 as shown in Figure 3.8. For this study, all the questions (about 15 million questions) asked in SO from 2009 to 2017 were considered.

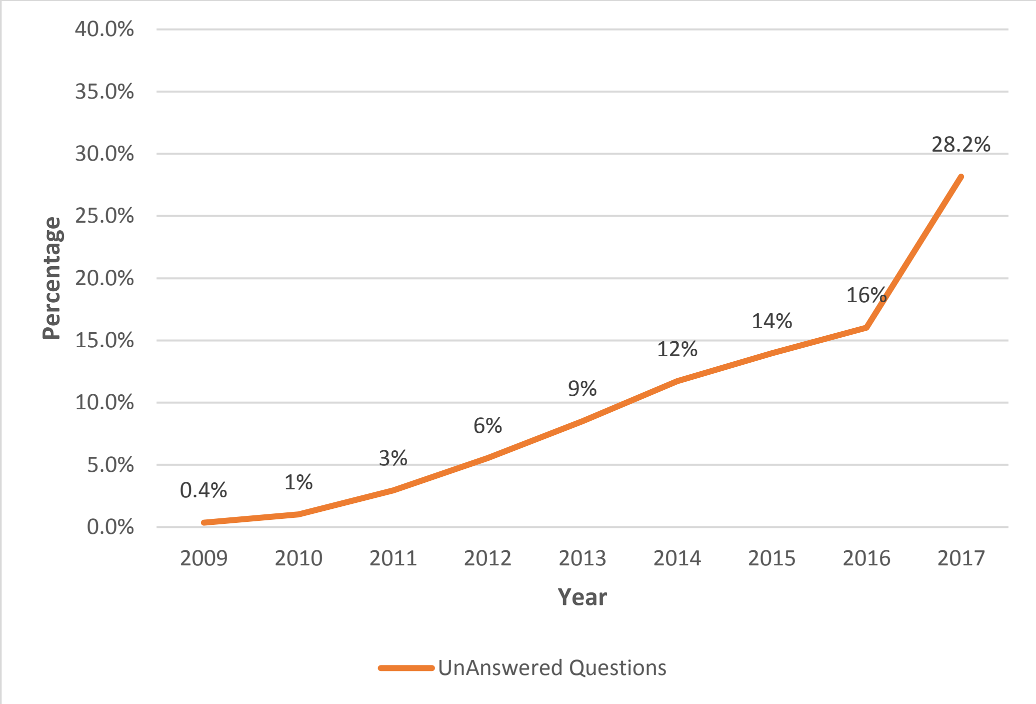

Figure 3.8. Percentage of Unanswered Questions

This growth in the proportion of unanswered questions is worrying, as having 685,965 unanswered questions is a challenge that could affect the users concerned in achieving their learning needs. 49.3% of the unanswered questions are asked by new users (users who have been in SO for a year or less) compared to the 17.5% of unanswered questions by older users (users who have been in SO for five years or more). Of course, not receiving answers to questions or having to answer your question yourself could deter a user from subsequently using the forum.

3.2.4 Decrease in the Proportion of "Quality Answers"

Another issue users are faced with in SO is the decrease in the proportion of quality answers. In this study, *quality answers* are defined as answers with scores of greater than 1. For some of the questions that get answered, the answers received are not good answers. The distribution of *quality answers* in SO from 2009 to 2017 as shown in Figure 3.9 shows a decrease in the proportion of quality answers (Ishola and McCalla, 2017a). 23.35 million answers provided to questions from 2009 to 2017 were considered for this study.

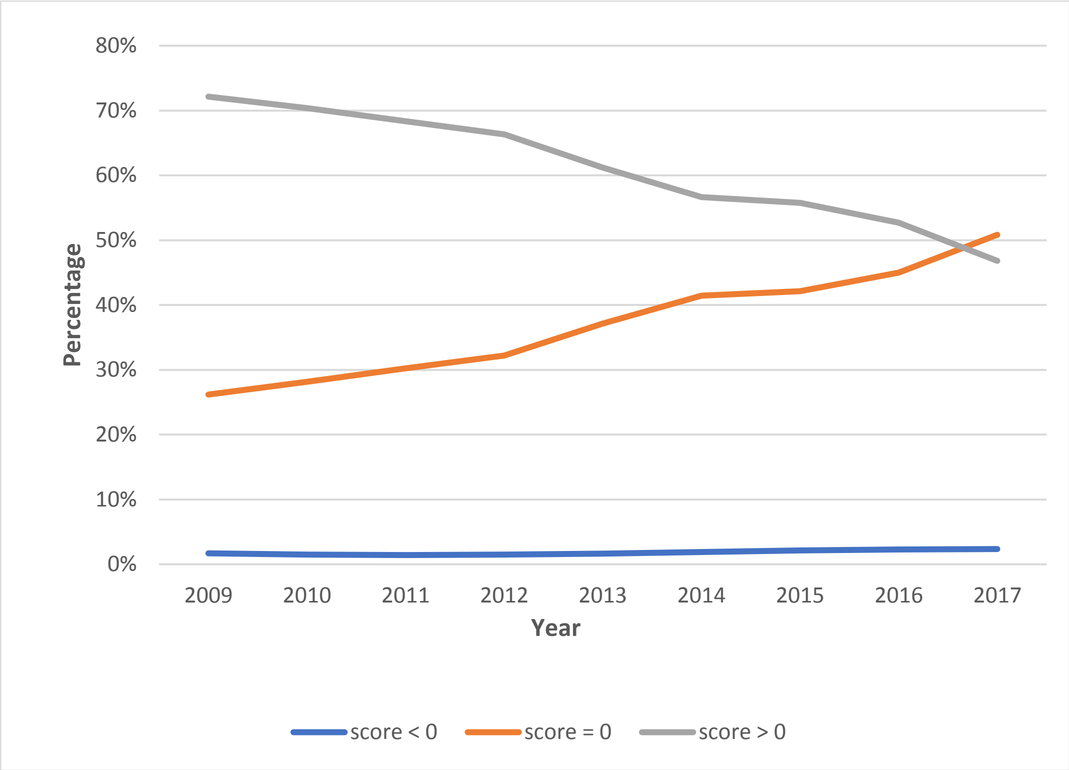

Figure 3.9. Answer Score Distribution in SO

The proportion of *poor answers* (with score < 0) in SO has remained at about 2%, although the total number of poor answers has also grown from 15,903 in 2009 to 69,574 in 2017, because of the overall growth in the SO user base. Looking closely into the reasons for poor answers in SO, the comments provided by community members to these low-quality answers were studied. In SO the comments provided serve as feedback to the user who provided such an answer. A qualitative analysis of comments (353) provided to poor answers with a score of -10 and below (which I will call *very poor answers*) was carried out. A score of -10 means at least ten SO users have ranked the answer as *poor*. The reasons provided for the *very poor answers* were manually extracted from the comments. Phrases such as "*misleading answer*", "*providing off track answers*", "*solution does not follow best practice*", "*answer provided does not work in all instances*", "*misinterpreted question*", "*recommending a new language different from users' interest*" were the most common feedback provided about such answers. The counts of these common kinds of feedback are shown in Table 3.5.

Importantly, many of these reasons for poor answers result from users providing answers when they lack relevant knowledge. The observed trend in Figure 3.6 points out the need to diagnose the learning needs of users early and to provide feedback to users before they create poor answers.

Table 3.5. Reasons for Very Poor Answers

Reason for Down Voted Answer	Count	Percentage
providing off track answer	77	21.81%
misleading answer	70	19.83%
solution does not follow best practice	38	10.76%
recommending a new language different from user's interest	21	5.95%
describing a solution without an example	20	5.67%
answer provided does not work in all instances	19	5.38%
answer wrongfully criticizes a product	19	5.38%
controversial opinion	16	4.53%
outdated solution	16	4.53%
code provided contains bug or is incorrect	13	3.68%
wrong application of knowledge	13	3.68%
misinterpreted question	11	3.12%
inefficient answer that does not optimize performance	10	2.83%
answer provides link to other external sources which are not useful	6	1.70%
unpopular answer	2	0.57%
attacking new product or not been open to new discoveries	1	0.28%
promoting one product over another without sufficient knowledge	1	0.28%

3.2.5 Increase in the Proportion of "Unaccepted Answers"

Another issue is the usefulness of the answer: does the answer, late or not, meet the user's learning needs? In SO, question-askers can mark as "accepted" one of the answers they received as the most useful in answering their question. Answers marked as accepted can be deemed to have met the user's learning needs. Analyzing the proportion of accepted answers compared to the total number of questions answered from 2009 to 2017, an increase in the proportion of *unaccepted answers* was observed as shown in Figure 3.10 below. For this study, all the questions (about 14 million questions) asked in SO from 2009 to 2017 were considered.

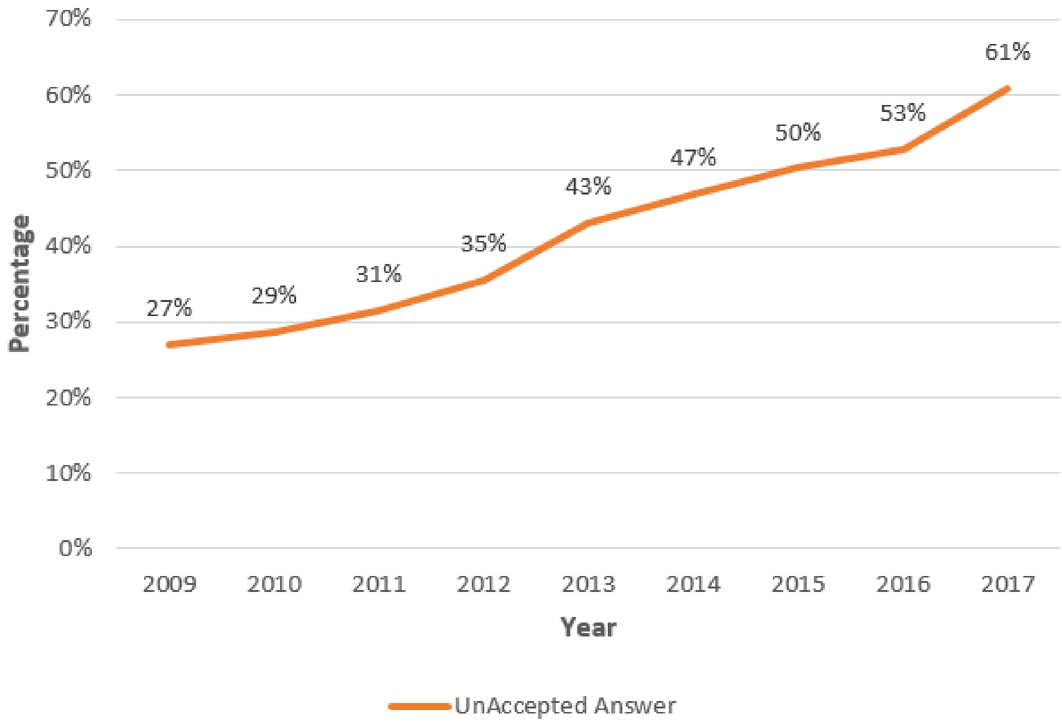

Figure 3.10. Percentage of Unaccepted Answers

Further study was undertaken into the perceived usefulness (user's acceptance) of the answers received by question-askers and how this related to how long it took to receive the first answer. For this study, all answered questions asked in 2017 (about 1.75 million questions) were considered. Figure 3.11 shows the time frame for the first answers and the proportion of these first answers not accepted by the user who asked the question.

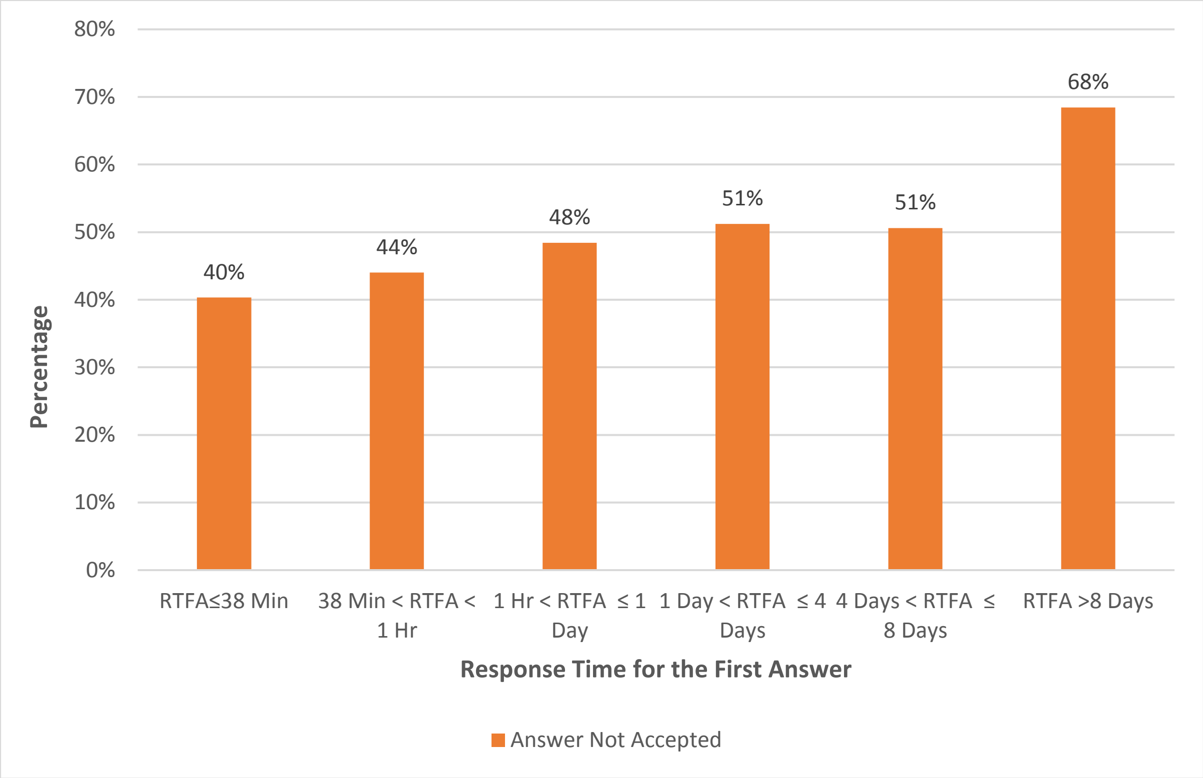

Figure 3.11. Proportion of Unaccepted Answers based on RTFA

The results obtained in Figure 3.11 show that as the RTFA increases, the perceived usefulness of the answer reduces (in the eyes of the question-asker), especially for question-askers who received their first answer to their question after 8 days. Some of these answers may prove useful to people other than the question-asker. In fact, some of the comments provided to the questions answered after 8 days before the questions received their first answer indicate that other users face similar issues. Examples of such comments are: "*facing the same issue, please help us*", "*have you resolved the problem?" What did you do eventually?"* Therefore, the diminished perceived usefulness of answers provided to questions with RTFA after 8 days at the very least signals a decrease in user satisfaction.

Having identified some of the reasons for *very poor answers* and identified some of the current issues experienced by users in SO, the conclusion is that it is important to support both the question-askers and answerers to ensure their learning needs are met, and they are up to date with emerging advances in their profession. The next five chapters present details of seven experiments carried out that attempted to resolve some of the issues described in Sections 3.2.1. to 3.2.5.

CHAPTER 4

TOWARDS FOSTERING PEER-PEER INTERACTIONS BETWEEN USERS

The goal of my research is to provide personalized support to professional learners that would be useful in addressing some of the issues described in Sections 3.2.1 to 3.2.5. To investigate the possible ways to foster peer-peer interactions in SO, I sought to understand the learning needs of question-answerers and question-askers. Hence, my research addresses issues involving support for peer help, which is an old area of research in the advanced learning technology research community (Greer et al., 1998a; Greer et al., 1998b; Vassileva et al., 1999). Peer helping is a process whereby peers with equal status support each other in solving a task (Ladyshewsky, 2006). Ladyshewsky (2006) argued that peer help should only be complementary to professional teaching and not a substitute for it. For software professionals who do not always have access to professional teachers, help from their peers in an OLC is one of the best options. Knowing the potential of peer help at improving the performance of users (Tsuei Mengping, 2012; Song, Loewenstein, and Shi, 2018), the main goal of my research is to detect the learning needs of professional learners. Afterwards, measures would be developed to provide personalized support to meet these learning needs. This chapter is divided into two sections aimed at analysing the learning needs of question-answerers and question-askers in SO.

4.1 Analysing the Answer Quality of Reputable Question-Answerers

Diagnosis of learning needs have been determined in the past mainly by tracking learner's job performance (Ley et al., 2008). Judging the knowledge of a professional based on their job performance has limitation since the knowledge required for the job is only a subset of the knowledge they ought to have to be proficient professionals. Ley and Kump (2013) have argued that an assessment of tasks performed alone is a weak measure in assessing the knowledge of a professional, as such assessment will at most be judged compared to fellow workers rather than with professionals in society at large.

Working in the professional programming domain, my study goes beyond these limitations in several ways. First, I aim to detect the learning needs of professionals from an OLC with no explicitly predefined required body of knowledge. In this experiment, the assessment provided by peers on the answer posts of a user were used as an indicator to their learning needs. Specifically, the answers provided by users voted up are regarded as what the users know and vice versa. Hence, my approach extends the previous studies by Ley et al. (2008) by diagnosing the learning needs of users based on feedback from peers beyond their workplace. Particularly, in this experiment my aim is to understand the learning needs of question-answerers in SO.

4.1.1 Methodology

For this study, I selected users who created an answer post in 2017 and have a reputation score above 100. Hence, only active and reputable users in SO were selected for this study. The selected users were classified into four activity levels based on how active they have engaged in help giving in SO, as determined by the tag-based badges in Figure 3.5. The four activity levels defined are shown in Table 4.1.

Table 4.1. Activity Levels Definition

Activity Levels	Definition	Number of Users	Badge Type
Level 1	Users with a minimum reputation score of 100 but who have not earned any tag-based badge.	187,027	No Tag-Based Badge
Level 2	Users who have earned a 100-reputation score and have earned a bronze tag-based badge for providing at least 20 answers to questions relating to a specific tag.	10,418	Bronze Tag-Based Badge
Level 3	Users who have earned a 400-reputation score and have earned a silver tag-based badge for providing at least 80 answers to questions relating to a specific tag.	2,458	Silver Tag-Based Badge
Level 4	Users who have earned a 1000 reputation score and have earned a gold tag-based badge for providing at least 200 answers to questions relating to a specific tag.	989	Gold Tag-Based Badge

As defined in Table 4.1, the set of question-answerers categorized as level 4 are the most reputable question-answerers who have earned the gold badge recognition for their participation. Question-answerers at all 4 levels have shown willingness to help provide answers to the questions of other users. To infer the learning needs of each set of users, I assume that users that provide a higher percentage of quality answers have lower learning needs. For this study, quality answers are defined as answers which are voted up by their peers and have at least a score of 1. So, for each set of users defined in Table 4.1, I wanted to study how answer quality provided by the active question-answerers in SO changed from 2009 to 2017. The percentage quality answers (PQA) for each set of users *u* for a given year was computed as in Equation (4.1) below:

$$PQA = \frac{\Sigma_u(\, Q_A^{up})_u}{\Sigma_u(Q_A)_u} * 100\% \qquad \text{Equation (4.1)}$$

where Q_A^{up} represents number of questions answered and upvoted, and

Q_A represents number of questions answered.

4.1.2 Results

The answer quality distribution for each set of users was tracked from 2009 to 2017 as shown in Figure 4.1 below.

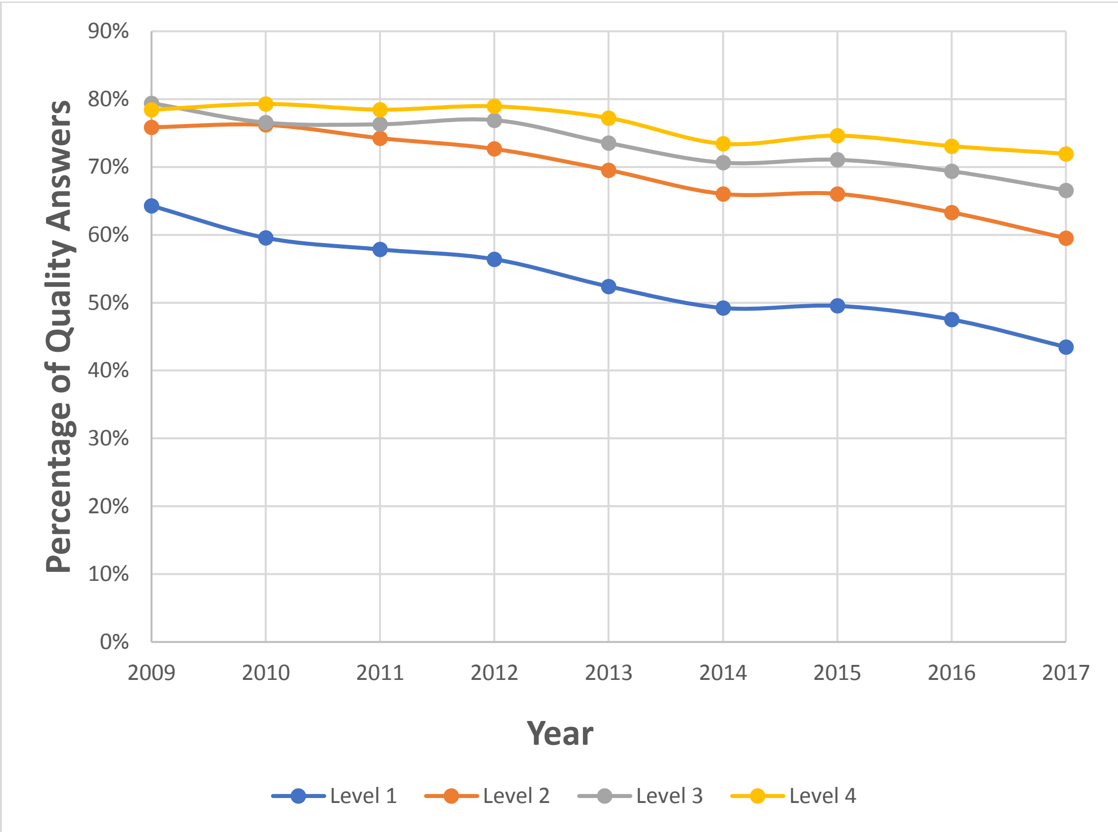

Figure 4.1. Distribution of Answer Quality by Active Answerers in SO

The observed trend in SO shows a decline in the quality of answers in SO, even by the most reputable users (level 4). The observed trend could be because of increase in poor quality questions or perhaps more difficult questions are being asked in SO. Another possible reason for the trend in Figure 4.1 might be that some of the answers were provided late and were thus no longer perceived as useful. Also, as technology advances within the profession, new knowledge emerges, widening the horizon of the knowledge required to provide quality answers to questions. The results from the comments analyzed as shown in Table 3.5 indicate that poor quality answers could be an indication of a deficiency in the knowledge of the question-answerers. Further studies would have to reach a definite conclusion about the observed trend. The observed trend in Figure 4.1 is at least concerning as the reputable question-answerers are considered subject-matter expert in SO relatively to novice users. Moreover, the growth in poor-quality answers could also affect the help-seeking experience of question-askers, of which a large proportion of question-askers are new users.

4.1.3 Discussion

In this experiment, attempts were made to detect the learning needs of users without depending upon an ontology. Instead, the proportion of up voted answers provided by the user to questions was used as a proxy to infer the learning needs of users. Results from this experiment show an overall trend of a decrease in quality of answers in SO as demonstrated in Figure 4.1. Regardless of the reason for the observed trend in Figure 4.1, a decrease in the quality of answers by the most reputable users (level 4) is concerning. As the most reputable users also seemed to have gaps demonstrated in their knowledge which they might be ignorant of at the point they provided the poor-quality answers. Hence, the need to help users diagnose their unperceived learning needs before they become apparent.

Despite the novelty of this experiment, there are limitations in this approach. Of course, not all the users selected for this study would have participated in peer help for the entire period from 2009 to 2017, which might have affected the validity of the results obtained. However, as the results in Figure 4.1 are based on the proportional changes in the quality of answers, the figure accounts for given individuals moving in and out of the active user category in that period. Moreover, accessing the learning needs of the question-answerers beyond SO would be worthwhile to validate that poor answer quality represents gaps in the knowledge of the users. Even if poor answers do not represent learning needs of the users, filtering out such answers could be useful since they are less likely to help meet the learning needs of the question-askers. This experiment was partially reported in the UMAP 2016 Extended Proceedings (Ishola and McCalla, 2016a). Unfortunately for this experiment, I did not have independent access to these users outside of their recorded behaviour in SO. Therefore, in subsequent experiments, as described in subsequent chapters I proceeded with a different approach that allows for the validation of the results of the diagnosis within SO.

4.2 Analyzing the Help-seeking Experience and Enthusiasm of Question-Askers

The decline in the answer quality provided by even the most reputable question-answerers in Figure 4.1 is concerning as this could have corresponding effects on the help-seeking experience of the question-askers. As shown by Howley (2015), the previous help-seeking experience of a question-asker plays a role in determining if they would rely on the OLC to meet their subsequent learning needs. Therefore, to enhance peer-peer interactions among users within the OLC, it is important to have help-seekers express their learning needs confidently whenever they have questions to ask. The ability of question-askers to express their learning needs when necessary is the first step in ensuring the learning needs of users can be met.

Hence, the goal of this experiment is to track the effects of the help-seeking experience of users on their ability to express their learning needs and to meet the learning needs of others. Also, this study seeks to establish the effects of the help-seeking enthusiasm of question-askers on their evolving answer quality in SO. In this study, help-seeking experience of a user is defined as whether the user received help or not when they posted a question in SO. The two forms of help-seeking experiences considered in this study are:

- positive help-seeking experience which represents a user who sought help and received help, and
- negative help-seeking experience which represents a user who sought help but didn’t receive help.

Similarly, help-seeking enthusiasm of a user is defined as how often a user seeks help and two forms of help-seeking enthusiasm were considered in this study:

- inactive help-seeking enthusiasm which represents a user who did not seek help frequently, and
- active help-seeking enthusiasm which represents a user who sought help frequently.

A previous study has shown that badges and scoring systems are forms of incentives that can encourage user participation in online forums (Grant and Betts, 2013). Gibson, Ostashewski, Flintoff, Grant, and Knight (2015) also studied the influence of badges on online educational activities to increase learner engagement. My experiment deepens and refines these studies by using badges as proxies to study the effects of the help-seeking experiences and help-seeking enthusiasm of question-askers. A detailed description of the approaches employed is presented in the next section.

4.2.1 Methodology

In SO, badges are awarded to question-askers to promote quality questions and continuous participation of the users within the forum. In a study by Grant and Betts (2013), the activities of the users before and after the badge was awarded were studied to establish the effects of badges in influencing the participation of users in SO. Specifically, the frequency of the online activities of the users in SO 2 months before and after a badge was earned were compared. For example, in their study they checked the frequency of posts edited by the users 2 months before and after earning the badge to see the effects of the edit badges shown in Figure 3.4. In this experiment, badges were used as a proxy to study the help-seeking experience and help-seeking enthusiasm of question-askers. Similarly, I studied the effects of help-seeking experiences of users on their ability to express their learning needs and to meet the learning needs of other users by mining their activities in SO. I examined the frequency of activities of users in SO a month before and a month after the time they earned the badge.

The four badges described in Table 4.2 were used as proxies to study the help-seeking experience and help-seeking enthusiasm of users in SO. For this study, users who have earned any of the four badges presented in Table 4.2 in 2017 were considered while users who earned more than one of the badges in Table 4.2 were exempted.

Table 4.2. Badges Representing Help-seeking Experiences and Enthusiasm of Users

Badge Name	Description	No. of Users
Scholar	Users that successfully sought and received help	227,997
Tumbleweed	Users that sought help but didn't receive help	158,576
Curious	Users who asked good questions on 5 separate days	36,037
Inquisitive	Users who asked good questions on 30 separate days	3,664

The distinct number of such users for each badge is shown in Table 4.2. The Scholar and Tumbleweed badges were used to determine the help-seeking experiences of the users. The Curious and Inquisitive badges were used to determine the help-seeking enthusiasm of the users. The description of the four badges as defined in SO is shown in Figure 3.3 while the description of these badges in the context of this study are stated below:

- *Scholar Badge***:** This badge is awarded when a user accepts one of the answers given to his or her question as being helpful. Users who earned the scholar badge had a positive help-seeking experience and their expressed learning needs on the question have been met.
- *Tumbleweed Badge***:** This badge is awarded to users who asked a question that earned a zero score, received no answers and no comments and low views for a week. Users who earned this badge had a negative help-seeking experience as they did not have their expressed learning needs met.
- *Curious Badge:* This badge is awarded to users who have asked at least 5 questions voted up in SO. Users who earned this badge would be the less unenthusiastic users (compared to those who earned the Inquisitive badge).
- *Inquisitive Badge:* This badge is awarded to users who asked at least 30 questions well received in SO. Users who earned this badge would be regarded as the more enthusiastic users.

Each category of users represented by the badges above has shown some effort to have their learning needs met. In analyzing the help-seeking experiences and help-seeking enthusiasm of question-askers in SO, three research questions were asked as stated below:

- ***Q1**: What are the effects of the help-seeking experiences of users on their propensity to express their learning needs?*

To answer question Q1, the number of questions asked by users who received the Scholar and Tumbleweed badges during the month before and during the month after they received the badge were compared. The percentage increase in the frequency of questions (PIFQ) asked by a user *u* was computed shown in Equation (4.2).

$$\text{PIFQ} = \frac{(N^{Q}_{after} - N^{Q}_{before})_B}{(N^{Q}_{before})_B} * 100\% \qquad \text{Equation (4.2)}$$

where N^{Q}_{before} = number of questions Q asked during the month before earning the badge *B*, and

N^{Q}_{after} = number of questions Q asked during the month after earning the badge *B*.

Hence, in computing PIFQ, the question asking activities of all users a month before and a month after earning the badge *B* were considered.

- ***Q2**: What are the effects of the help-seeking experiences of users on their propensity to meet the learning needs of other users needing help?*

To answer question Q2, the number of answers provided by users who received the Scholar and Tumbleweed badges during the month before and during the month after they received the badge were compared. The percentage increase in the frequency of answers (PIFA) for user u was computed as shown in Equation (4.3).

$$\text{PIFA} = \frac{(N^{A}_{after} - N^{A}_{before})_B}{(N^{A}_{after} -)_B} * 100\% \qquad \text{Equation (4.3)}$$

where N^{A}_{before} = number of answers A provided during the month before earning the badge *B*, and

N^{A}_{after} = number of answers A provided during the month after earning the badge *B*.

Like PIFA the answer giving activities of all users a month before and a month after earning the badge *B* were considered.

- ***Q3**: What are the effects of the help-seeking enthusiasm of users on their evolving answer quality distribution?*

In answering Q3, I compared the percentage of quality answers provided by users who do not seek help frequently (the unenthusiastic users) and the users who frequently ask questions (the enthusiastic users). For the unenthusiastic and enthusiastic users, the percentage of quality answers (with score > 0) was computed as shown in Equation (4.1). For each of the first 6 months after the user earned the badge.

The answers to the research questions are presented below.

4.2.2 Results

In analyzing the help-seeking experiences and help-seeking enthusiasm of question-askers in SO, three research questions were asked.

- ***Q1**: What are the effects of the help-seeking experiences of users on their propensity to express their learning needs?*

The results obtained in answering question Q1 are shown in Table 4.3.

Table 4.3. Percentage Increase in the Frequency of Questions Asked

Badge Name	Total Questions Asked BEFORE Earning the Badge	Total Questions Asked AFTER Earning the Badge	Percentage Increase
Scholar (positive experience)	263,720	178,701	-32.2%
Tumbleweed (negative experience)	226,605	66,125	-70.8%

Results in Table 4.3 show that regardless of the help-seeking experience of users, there is a decrease in the number of questions asked. A decrease for users who had negative help-seeking experiences is not a surprise, but a decrease of nearly three quarter (71%) in the number of questions asked was unexpected. Of course, the users with negative help-seeking experience might have received help to answer their question outside of SO. Even if help was gotten from outside SO, for the sustainability of the community and to avoid user churn it is important to investigate measures that would mitigate a decrease like this. Investigating ways to avoid such a large decrease is necessary as most questions in SO are asked by new users who have been in SO for less than a year.

Although I had expected that users who had positive help-seeking experiences would be motivated to ask more questions, the results in Table 4.3 show otherwise. Perhaps, these users have learned from the answers, and need little help - they are in fact "scholars". It is also possible the questions these users intend to ask have been asked by other users in SO. Another explanation might be that although the questions of these users were answered, they were answered late and were no longer relevant to the users as shown in Figure 3.8. Another possibility is that some of these question-askers received poor quality answers to their questions as shown in Figure 4.1. A better understanding of the reasons for the results in Table 4.3 is necessary in developing measures that would help improve the help-seeking experience of the users. My next goal was to study the effects of the help-seeking experience of users on their propensity to provide help subsequently to other users who require help.

- ***Q2**: What are the effects of the help-seeking experiences of users on their propensity to meet the learning needs of other users needing help?*

The results obtained in answering question Q2 are shown in Table 4.4.

Table 4.4. Percentage Increase in the Frequency of Answers Provided

Badge Name	**Total Answers Provided BEFORE Earning the Badge**	**Total Answers Provided AFTER Earning the Badge**	**Percentage Increase**
Scholar (positive experience)	96,970	121,476	25.3%
Tumbleweed (negative experience)	58,269	46,432	-20.3%

Results in Table 4.4 show that the users with positive help-seeking experiences become more likely to support other users in meeting their learning needs. The ones with negative help-seeking experiences are less eager to provide help. The results in Table 4.4 again point out the need for users to have their learning needs met if the supply of potential helpers is to be sustained in an OLC. Pro-active recommendation of potential helpers may be a partial solution to the problem of negative help-seeking experiences and to the problem of poor answer quality. While it is desirable to have users reciprocate the help received by answering the questions of other users in SO, it is also important these users provide quality answers. Hence, with question Q3 the goal was to study the evolving answer quality provided by question-askers in SO.

- ***Q3**: What are the effects of the help-seeking enthusiasm of users on their evolving answer quality distribution?*

Figure 4.2 below shows the percentage increase in quality answers as computed in Equation (4.3) above.

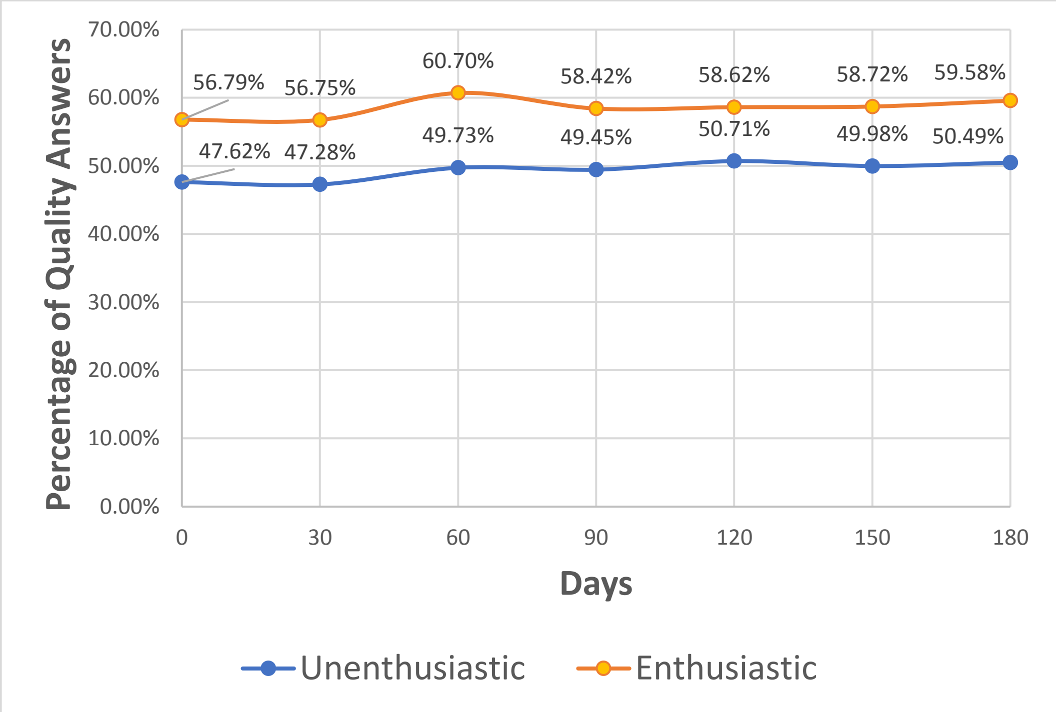

Figure 4.2. Evolving Answer quality of Unenthusiastic Versus Enthusiastic Users

The results in Figure 4.2 show that the more enthusiastic users demonstrated a higher percentage of quality answers than the more unenthusiastic users. Perhaps, the answers provided by the unenthusiastic users were not answered as promptly as the enthusiastic users, hence reducing the chances that their answers would be voted up. Should the answers provided by the unenthusiastic users be answered late, then perhaps informing users about newly asked relevant questions would be necessary. Another possible explanation for the lower percentage of quality answers demonstrated by the unenthusiastic users might be that they are not taking active steps to meet their learning needs. In such a case, informing these users about their individual learning needs might be helpful. Completely understanding the differences in the answer quality of unenthusiastic versus enthusiastic users would be helpful in identifying the best measures to be adopted to mitigate such a trend. While further investigations would be required to explain the results in Figure 4.2, the results at least suggest it would be useful to encourage more users to seek help more often.

4.2.3 Discussion

The end goal of peer help in an OLC is to have the expertise of help-seekers improve after seeking help. The first step in reducing the learning needs of users is to ensure users express their learning needs when necessary. Hence, this experiment attempted to establish the effects of the help-seeking experience of users on their propensity to express their learning needs and to meet the learning needs of other users. The effects of help-seeking enthusiasm of users on their evolving answer quality were also studied.

In carrying out this experiment, three research questions were considered. Answers to these questions Q1 and Q2 reveal the importance of supporting help-seekers in meeting their expressed learning needs. Users with negative help-seeking experiences were less likely to rely on the learning community in meeting their future learning needs compared to users with positive help-seeking experiences. Even for the users who had positive help-seeking experiences, they themselves still need to be further supported in ensuring they receive quality help.

Unlike the reputable users as shown in Figure 4.1 who had a decline in the percentage of their answer quality, for the enthusiastic users a marginal rise in their answer quality was observed as shown in Figure 4.2. Therefore, the observed trend in Figure 4.1 and Figure 4.2 indicate that both question-answerers and question-askers require support in ensuring their knowledge is kept up to date. One opportunity is to take a proactive approach to detecting the unperceived needs in the knowledge of users even before such needs are evident to the users. If it were possible to diagnose the unperceived needs of users before they are evident to the users, then it would be possible to provide prompt feedback to the users about such learning needs. Moreover, by letting the question-answerers have adequate information about what they know and what they don’t know, they could be more informed about questions they can answer well. For the results of such diagnoses to achieve their intended effects, it would be important to detect the individual learning needs of users. Therefore, in my subsequent experiments as discussed in Chapter 5, my goal will be to detect the individual unperceived learning needs of users.

A limitation to this study is that the effects of help-seeking experiences and enthusiasm of users were examined within 1 month and 6 months timeframes. A threat to the validity of the results might be posed if the time frame is extended to a longer duration. However, this experiment was carried out in line with other experiments on badges where researchers have also used a similarly short time frame of 2 months (Grant and Betts, 2013). Also, further studies and more concrete evidence would be required to completely explain the results obtained from these analyses. Even in the absence of these further studies, the variation in the results of users with positive versus negative help-seeking experiences, creates the need to improve the help seeking experience of users. Final, this study was carried out on the Stack Overflow forum. Results from this study, therefore might not exactly model the behavior of professionals in other online forums. However, this experiment was carried out on about 426,274 users, a sample size that is large enough to provide real insight into the behavior of users, if not yet definitive conclusions. Portions of this experiment were reported in the PALE 2016 workshop (Ishola and McCalla, 2016c) and the IMS 2018 workshop (Ishola and McCalla, 2018c).

CHAPTER 5

PREDICTING ANSWER QUALITY

The experiments discussed in Chapter 4 suggest the need to enhance the support provided to users within an OLC. With the decline in the answer quality of the active users (as shown in Figure 4.1), it means even the users believed to have a higher reputation in SO might not always know what their learning needs are. To deal with the decline in the answer quality of the active users, it would be good to promptly detect the learning needs of users. Knowing in advance the quality of the answer a user may give, advice might be provided to that user about whether to try to answer a question and even to inform them of their own unperceived needs.

The goal of the experiment reported in this chapter is to use the previous answers provided by a user to questions to predict the quality of the user's possible answer to a question. The accuracy of the diagnosis can be validated by comparing the actual quality of the answer the user provided with the predicted answer quality. More specifically, the quality of an individual user's answer performance (whether a user will give good or poor answers to a question) was predicted using a naïve Bayes model. Using a probabilistic classifier like naïve Bayes creates the opportunity to assess the confidence of the prediction. Hence, where there is uncertainty in the prediction, human expert intervention could be employed. Moreover, given the few data instances available about the previous activities of each user in SO, using a naïve Bayes classifier is ideal, as it performs well even with small data sets. A detailed description of the approach employed is discussed next.

5.1 Methodology

This experiment models the knowledge of users within SO, an open-ended online learning community (where there is no well-defined curriculum). In previous work by Conati (2010) for instance the knowledge of users is modelled in a well-defined domain, where a curriculum can guide expectations about changes in the user's knowledge. In other studies of well-defined domains, modelling the individual knowledge elements for each user has been shown to improve the overall performance of the model (Lee and Brunskill, 2012; Pardos and Heffernan, 2010; Yudelson, Koedinger and Gordon, 2012). In an open ended longer-term learning domain like SO, the user's learning needs do not align with a curriculum and the disciplinary knowledge changes over time. Hence, modelling the knowledge of users using a lightweight diagnosis technique that does not require extensive knowledge engineering would be useful. In the SO domain, I used the tags attached to questions as proxies for the knowledge elements to indicate the ability of the user to answer a question. In carrying out this experiment 5 answer classes were defined as described in Section 5.1.1. Thereafter, a naïve Bayes model was used to model the knowledge of SO users based on their past performance in answering questions as described in Section 5.1.2. The quality of their future answer performance was then predicted.

5.1.1 Dataset Description

In SO, each question is annotated with tags indicative of the knowledge elements required to answer it correctly. In this experiment, I mined the tags employed in questions answered by active SO users. Active users in this study are defined as those users who have provided at least 200 answers in the forum from 2009 - 2014. There were 834 such active users in this period with 4,038,969 answer posts containing 20,158 distinct tags. Each answer was categorized into five answer classes. These answer classes were defined based on the aggregate score received by the answer as evaluated by other users in SO. First, the answer classes were discretized, as shown in Table 5.1, based on the score requirements for the four answer badges as defined in Figure 3.2.

Table 5.1. Answer Badge Classification in SO

Badge Name	Score Range
Great Answer	score $\geq$ 100
Good Answer	25 $\leq$ score < 100
Nice Answer	10 $\leq$ score < 25
Teacher	1 $\leq$ score < 10

Rather than the awkward SO terminology, the "*Teacher*" answer badge class will be called "*Satisfactory Answer*" in the rest of this discussion. Next, in addition to the four badge categories, an extra answer class referred to as "*Poor Answer*" was created for answers with scores below 0. In this study, answers with score = 0 were not considered. Next the goal is to gain insight into the appropriateness of the five answer classes defined by studying the proportion of accepted answer distribution per answer class. The proportion of accepted answers and non-accepted answers to the total number of answers for each answer class defined above was computed as shown in Figure 5.1 below.

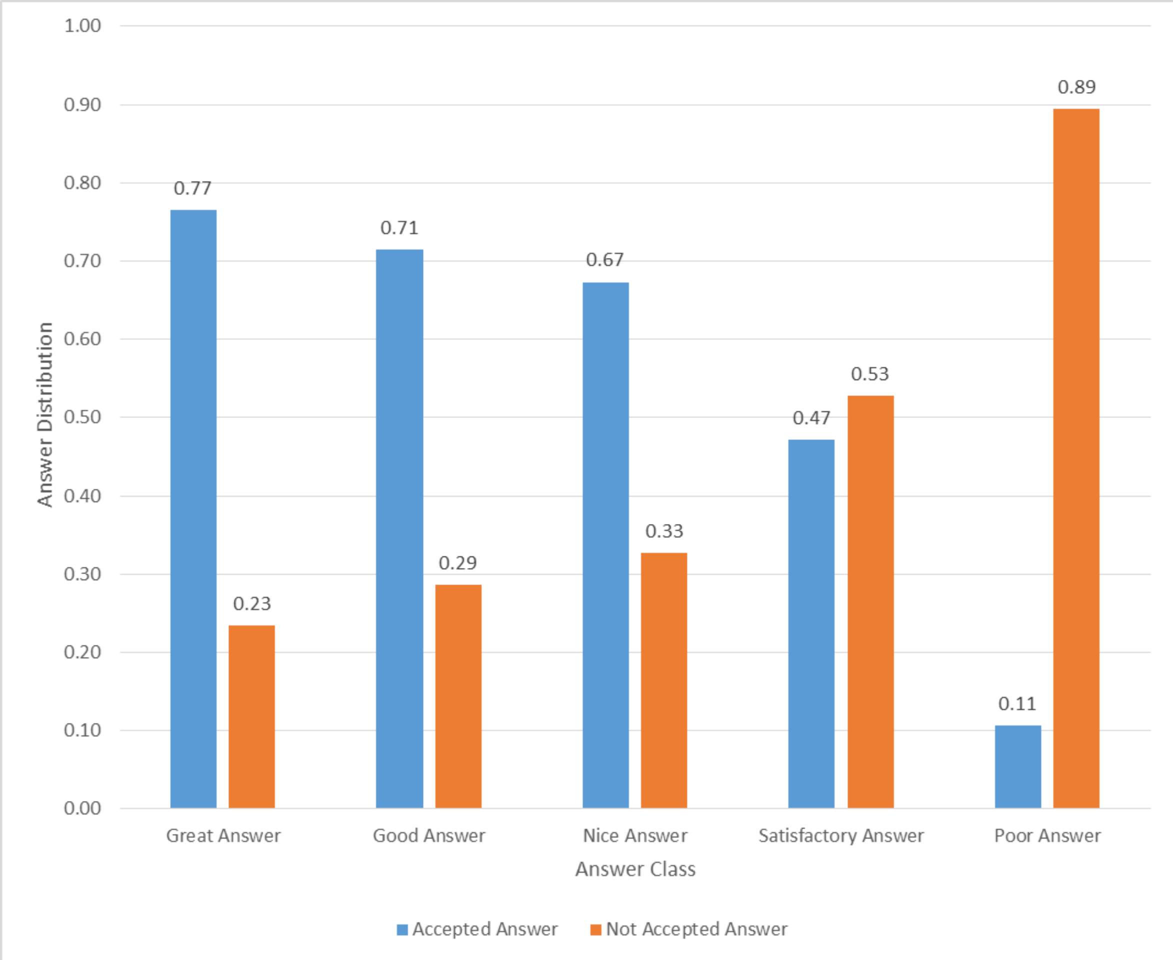

Figure 5.1. Proportion of Accepted Answer per Answer Class

Figure 5.1 shows that answers in the answer class "*Great Answer*" are more frequently accepted while answers in the answer class "*Poor Answer*" are rarely accepted. And there is a nice progression in between, with correspondingly higher proportions of "worse" answers not being accepted. This is the expected pattern if the answer classes reflect the quality of the answer.

5.1.2 Modelling Framework

This section describes the model employed in this experiment. The naive Bayes classifier was used to compute the probability of each of the 5 answer classes for each question and for each user. First a brief description of the naïve Bayes classifier is provided. This is followed by a description of how naïve Bayes was applied in predicting the answer classes.

The generative model of naïve Bayes is defined as shown in Equation (5.1) below (Murphy, 2012):

$$p(y|x) = \frac{p(y)\, p(x|y)}{p(x)} \qquad \text{Equation (5.1)}$$

where *x* represents the input feature vectors and *y* represents the output class labels. In the SO context, input feature vectors are represented by the tags ($tag_i,\ldots,tag_k$) contained in the question while the output class labels are represented by the answer classes.

Given the naïve Bayes assumption, that the features are conditionally independent given the class label, the likelihood $p(x|y)$ can be decomposed as shown in Equation (5.2):

$$p(x_i, \ldots, x_k|y) = \prod_{i=1}^{k} p(x_i|y) \qquad \text{Equation (5.2)}$$

By substituting Equation (5.2) into Equation (5.1), the posterior probability $p(y|x)$ can be rewritten as shown below (Mitchell, 2005):

$$p(y = c|x_i, \ldots, x_k) = \frac{p(y=c)\, p(x_i,\ldots,x_k|y=c)}{p(x_i,\ldots,x_k)} \qquad \text{Equation (5.3)}$$

The resulting model represented by Equation (5.3) is called a naïve Bayes classifier (NBC). Although the model is naïve because it assumes that the features are conditionally independent of the class label, even when the assumption does not hold, the model still works well (Al-Aidaroos, Bakar, and Othman, 2010). Using the generative model of naïve Bayes, the probability distribution of the feature vectors for each answer class *c* can be learnt separately and combined as shown in Equation (5.4) (Murphy, 2012).

$$p^u(A = c \mid Tag_i, \dots, Tag_k) = \frac{p^u(A=c)\ \prod_i^k (p^u(Tag_i \mid A=c))}{p(Tag_i, \dots, Tag_k)} \qquad \text{Equation (5.4)}$$

where A is the set of possible output answer class labels,

c represents any of the specific answer class labels,

k is the number of tags used in the question,

Tag_i represents each of the tags used in the question, and

$Tag_i, \dots, Tag_k$ represents the input features which depict the tags assigned to the question by the question askers.

For any given question of any given user u, the probability that an answer will belong to an answer class is computed as shown in Equation (5.4). The goal is to assign the answer class label with the maximum probability value for user u, as shown in Equation (5.5):

$$C^u = argmax_c \left[\frac{p^u(A=c)\ \prod_i^k (p^u(Tag_i \mid A=c))}{p(Tag_i, \dots, Tag_k)}\right] \qquad \text{Equation (5.5)}$$

As the denominator is constant for all c, the denominator need not be included in the equation as shown in Equation (5.6):

$$C^u = argmax_c \left[p^u(A = c)\ \prod_i^k (p^u(Tag_i \mid A = c))\ \right] \qquad \text{Equation (5.6)}$$

where C^u represents the predicted most likely answer class for the user considered. Using the questions previously answered by user u, the prior probability $p^u(A = c)$ and the likelihood probability $p^u(Tag_i \mid A = c)$ are computed as shown below:

$$p^u(A = c) = \frac{\sum_u A=c}{\sum_u A} \qquad \text{Equation (5.7)}$$

$$p^u(Tag_i \mid A = c) = \frac{\sum_u A_i=c}{\sum_u A=c} \qquad \text{Equation (5.8)}$$

where $\sum_u A = c$ is the total number of answers with A= c previously provided by user u to questions asked,

$\sum_u A$ is the total number of answers previously provided by user u to questions asked, and

$\sum_u A_i = c$ is the total number of answers previously provided to questions about tag_i and with A= c provided by user u to questions asked.

The posterior probability as shown in Equation (5.6) can then be computed using the values obtained from Equation (5.7) and Equation (5.8). As prior knowledge about the previous answer quality of the user u is required to compute the posterior probability as shown in Equation (5.6), I only considered active question-answerers to avoid the cold start problem. Although only active users who have answered at least 200 questions were considered for this study, the same procedure would apply to users who have answered less than 200 questions. For new users who have not answered any question in SO, the posterior probability will be zero resulting in the cold start problem. As evidence of the answer quality of the less active question-answerers also becomes available, their ability to provide quality answers to a question can be estimated.

5.2 Results

The model was evaluated using 3 data sets, based on the number of answer posts of users:

- users with answer posts greater than or equal to 1000,
- users with answer posts between [500, 999], and
- users with answer posts between [200, 499].

For each of the 3 data sets, the NBC model is evaluated by computing the prediction accuracy (PA) as shown in Equation (5.9):

$$PA = \frac{\sum_{i=1}^{s} C_a = \bar{C}_a}{p} * 100\% \qquad \text{Equation (5.9)}$$

where C_a is the predicted answer class,

$\bar{C}_a$ is the actual answer class of the answer post,

s is the total number of accurate predictions, and

p is the total number of answer posts used for validation.

I used two approaches in computing this prediction accuracy: incremental and non-incremental.

5.2.1 Incremental Approach

In the incremental approach, the most likely answer class is predicted anew at each time *t* that a question is asked, using the NBC model shown in Equation (5.6). Specifically, to use NBC to predict the answer class of user *u's* answer to a question asked at time *t*, I employ the previous answers provided by the user before time *t* to construct the model. For the next question asked after time *t*, the model of the user (as shown in Equation (5.6)) is recomputed to include the observed quality of answer demonstrated by the user to the question answered at time *t*. By recomputing the posterior probability (Equation (5.6)) as evidence become available about the answer performance of a user to new questions, the evolving expertise levels of the user are captured. For all users, the prediction accuracy is computed by comparing the predicted answer class with the actual answer class of the answer post as shown in Equation (5.9). The results obtained are shown in Table 5.2:

Table 5.2. Prediction Accuracy on Different User Categories

Type of Measure	posts ≥ 1000	500 ≤ posts < 1000	200 ≤ posts < 500
Prediction Accuracy	94.5%	89.4%	87.4%

The results obtained in Table 5.2 show that the accuracy decreases as the number of posts decreases, although even for users with "only" 200 answer posts, an average of about 87% accuracy was achieved. Table 5.2 only reflects how the NBC model performs with varying numbers of data points for each user but does not show how the model performs for question posts with varying number of tags.

5.2.2 Non-Incremental Approach

With the non-incremental approach, the time sequence of when the answers were provided was not considered. Specifically, with the non-incremental approach I employed 10-fold cross validation to check the stability of the model with a varying number of tags. In validating the NBC model for each user, I divided the answer posts of each user into 10 randomized subsets without replacements. Thereafter, for each user, the answer posts from the 9 subsets were used to construct the NBC model and the remaining one subset was used to validate the respective model of each user. This process was repeated 10 times using different 9 subsets to construct the NBC model for each user as shown in Equation (5.6). The tenth subset is used to validate the model and it is excluded from the 9 subsets used in building the model of each user. Thereafter, for all users, the answer class predicted for each answer post is compared with the actual answer class of the answer post by computing the average PA as shown in Equation (5.9).

Using the 3 data sets defined above, I computed the PA based on the number of tags which varies between [1, 5] used in the question. For example, for all users with answer posts between [200, 499], I computed the PA separately for all answers provided to questions containing 1 tag, 2 tags, and so on up to 5 tags. Table 5.3 below shows the accuracy of the model based on the number of tags present in the question being answered.

Table 5.3. Prediction Accuracy with Varying Number of Tag(s)

Type of Measure	**posts ≥ 1000**	**500 ≤ posts < 1000**	**200 ≤ posts < 500**
1 Tag	91.7%	87.7%	85.7%
2 Tags	91.0%	86.8%	84.6%
3 Tags	90.8%	86.3%	84.8%
4 Tags	90.5%	86.4%	84.9%
5 Tags	90.8%	86.8%	85.4%

Results from Table 5.3 demonstrate that the model does not show a wide variation in its PA even with a varying number of tags. These are very good outcomes and show the promise of the tag-based model in making good predictions about the future from the analysis of past behaviour.

5.3 Discussion

A naïve Bayes model was employed to predict the quality of an answer a user will give to a given question in advance before that user provides an answer. This experiment assumes that the quality of answer provided by a user indicates how much they know about the question asked. Five answer classes were defined to represent the various levels of quality of answers. In meeting the learning needs of users, a user predicted to have the ability to provide the highest quality answer to the question (i.e. a "Great Answer") could be selected to answer the question. As Figure 5.1 shows a user predicted to provide a '*Great Answer*' has a better chance to provide an accepted answer to the question-asker. Hence, the question-asker has a higher chance to receive quality answers to their question.

Another contribution of this experiment to the advanced learning community is that it opens the opportunity to diagnose the knowledge of users even in an online learning community without a well-defined curriculum. Importantly, this experiment employed a naïve Bayes model which is a lightweight technique for diagnosing the knowledge of users rather than reasoning with an ontology. Using lightweight techniques avoids the need for creating, maintaining and keeping the ontology up to date. With the lightweight approach, as new tags are created, and data becomes available on what the user knows about the new tag, the approach employed in this experiment will still work. The model employed will evolve naturally by using both the new and the existing information about the user to diagnose his/her ability. Consequently, there would be no need of re-engineering the model employed as the user and the discipline knowledge changes over time. In addition, this approach scales well in that the performance of the model is good regardless of the number of data points, as shown in Tables 5.2 – 5.4.

The obvious limitation to the approach employed in this experiment is a cold start problem for users who have never used a tag before. For instance, for a new user in SO who has never provided answers before, as no prior knowledge about the previous answer quality about the user exists, the posterior probability would be 0. To avoid the cold-start problem, only users who have provided at least 200 answers to questions in the past were considered in this study. Also, in this diagnostic approach, only the answer activities of a user within SO were employed as opposed to possibly incorporating other sources of information about the knowledge of the user, for example, their job performance, training, resume, e-portfolio, certifications and so on. In future, it would be interesting to explore how external information about the knowledge of the user could be incorporated. This experiment was reported in AIED 2017 (Ishola and McCalla, 2017a). In the next experiment, attempts were made to study the influence of peers in detecting the unperceived needs of users.

CHAPTER 6
PREDICTING THE FUTURE LEARNING NEEDS OF QUESTION-ASKERS

In Chapter 5 my goal was to predict the current unperceived needs of question-answerers, while in this experiment the goal is to predict the future unperceived needs of question-askers. Detecting the unperceived needs early enough could help the user in keeping abreast of the latest discoveries in their profession. According to Grant (2002), the learning needs of a learner can be detected using these techniques:

- a learner's self-assessment to determine what they did well and what they didn't based on their past actions,
- a peer review whereby peers assess the knowledge of each other,
- a supervisor's observation of the learner while performing a task,
- gap analysis which compares what the learner knows to what the learner ought to know, and
- tracking and recording the learning needs of the learner over time.

Drawing upon Grant, I tracked the past question posts made by users within SO to predict their learning needs. The True Bayesian Estimate approach is used to predict the future learning needs of users by employing the past interaction data of each user in SO and the current trends within the SO community. Specifically, in this experiment, I predict the tags that a user will be asking questions about in the future. In evaluating the results obtained, I compare the predicted tags to the actual tags employed by each user in SO.

6.1 Methodology

In conducting this experiment, the questions asked by users in SO from August 2008 to September 2014 were examined. For this study, 5063 users who asked at least 100 questions in this period were considered. A total of 1,015,961 questions were mined. In personalizing support for lifelong learners, there is a trade-off between the usefulness of retaining old user model information and pruning the learner model to retain just current information about the learner (Kay and Kummerfield, 2009). Hence, this experiment compares the effect of using a short-term (in this case 5 months) versus a long-term (in this case 3 years) baseline of data on the accuracy of the predictions. The unperceived needs of the user were inferred over a short-term 5-month baseline (between March 2014 and July 2014) and a long-term 3-year baseline (between January 2009 and December 2011). The goal was then to compare these predictions to the actual learning needs of the user as exhibited in the questions they asked immediately after the baseline.

As described in Section 3.2.1 learning needs were computed based on the mapping of the tags on each user's question post to a computing class. In instances where more than one computing class is associated with a question, then the computing class with the highest number of tags was assigned to the question. Where ties exist, the order of tag usage in the question was considered. For instance, a question where two tags (tag_i, tag_k) belong to different computing class with tag_i being the first tag specified when the question was created, then the question was assigned to the computing class of tag_i. For all questions asked by a user during the baseline period, the computing class distribution is computed to infer the computing class where a larger number of the learning needs of each user lies.

6.1.1 Inferring the Current Unperceived Needs of a User

The next goal is to determine the computing class that would align with the future learning needs of the user. To determine the computing class where the future unperceived needs of an individual user lies, first I determined the computing class that reflects the current learning needs of the user. The current learning needs of a user are determined by computing the computing class distribution $D(u,t)$ as shown in Equation (6.1).

$$D(u,t) = \left(\frac{N_i}{N_q}, \ldots, \frac{N_p}{N_q}\right) \quad, where\ N_q = \sum_c N_c \qquad \text{Equation (6.1)}$$

where N_c represents the count of questions asked by user u for the computing class c before time t (when the computing class distribution is to be computed),

N_q shows the total number of questions asked by user *u* before time *t*, and

N_i … N_p represents each of the computing classes of the questions asked by user u.

The computing class c with the highest value from D(u,t) is inferred as the current learning need of the user. Of course, saying the learning needs of a user can be represented by a single computing class (e.g software coding) is too broad and would not necessary capture the specific learning needs of the user. Hence, I drilled further to detect the specific tags within the computing class that could represent the learning needs of the user.

For the computing class c with the highest value in D(u,t), the 100 most popular questions asked in the period right after the short and long term defined baseline are selected. The popularity of a question was determined by the number of views the question had (information available in SO). For instance, using the short-term baseline, the 100 most popular questions asked in August 2014 that relates to the computing class c with the highest distribution are selected for each user. Similarly, using the long-term baseline, the 100 most popular questions asked in 2012 that relate to the computing class c with the highest value in D(u,t) are selected. For each user, the questions asked by the user are excluded from the 100 questions selected. The 100 questions selected for each user are assumed to have the tags that might represent the future unperceived needs of the user. A detailed description of how I predicted the tags a user would be asking questions about in the future is presented in the next section.

6.1.2 Personalized Detection of the Future Unperceived Needs of a User

Using the questions selected for each user, the next goal is to determine the specific future learning needs of the user. In detecting the specific future learning needs of a user, the 100 questions selected for each user were ranked using a True Bayesian Estimate approach (TBE) (described below). The TBE approach employs the historical information about the user and the feedback from the community members in computing the ranking score for each question ranging between [0, 1]. Based on the ranking score computed for each question, the 100 questions selected are ranked and the questions with ranking score above a threshold value are selected. For each user, the tags employed in the questions with ranking score above the threshold value are predicted to represent the future unperceived needs of a user. By employing information about the user and those of other users within the community, the prediction made for each user can be tailored to the learning needs of the user and the learning trends within the community. Hence, the user can be better informed about what they want to learn and what they ought to learn i.e. things which other users are learning.

In diagnosing a learner's knowledge in intelligent tutoring systems (ITS), the Bayesian network has been successfully employed (Arroyo and Woolf, 2005; Conati, 2010; Zapata-Rivera and Greer, 2004). With a Bayesian approach, as the knowledge of the user changes over time, the effect of the changes can be propagated through the network (Desmarais and Baker, 2012). Working within the SO learning community, with millions of questions asked, using the Bayesian network could be computationally expensive; hence, I employed a lighter weight approach. Using the TBE approach all the question posts selected for each user are assigned a weighted score, which is computed by collectively using historical information about the user and those of other users in SO. This allows the questions to be ranked in order of importance.

The TBE approach has been applied in movie recommender systems to rank movies for a given user by employing information about the votes and ratings given to each movie to be ranked (Das and Chakrabarti, 2016; Kunaver, Pozrl, Pogacnik, and Tasic, 2007). Using the TBE approach each movie is assigned a weighted rating which is thereafter used to rank the items. The TBE is computed as shown in Equation (6.2) (Hatta, Wee, Cheah, and Wee , 2015):

$$w = \frac{v}{v+m}R + \frac{m}{v+m}C \qquad \text{Equation (6.2)}$$

where w = weighted rating,

R = average *rating* of the observed data, about the item to be ranked (in the movie domain this would be average rating of the movie to be ranked),

v = number of *votes* for the observed data, about the item to be ranked,

m = the weight given to the prior estimation for the observed data, (in the movie domain, the number of votes IMDB deemed necessary to be listed in the top 25,000 movies selected), and

C = the mean *vote* across the whole pool, where the whole pool represents all the items to be ranked.

With the TBE, as the value of v approaches zero, the weighted rating w will reflect the value of the mean vote C value. Likewise, as the value of v increases, the weighted rating w will reflect the average rating R value. Hence, as the number of votes v increases, the value of C becomes irrelevant and the weighted rating w is approximately the value of the sample average R. The Internet Movie Database (IMDB) employed the TBE approach to rank its top rated 250 movies from a list of 25,000 selected movies (Das and Chakrabarti, 2016; Kunaver et al., 2007). The parameters defined in Equation (6.2) in computing the weighted rating w for a given movie were computed for this movie database as shown below:

w = weighted rating computed for the movie to be ranked,

R = average rating[5] for the movie to be ranked,

v = number of votes[6] for the movie to be ranked,

m = minimum votes required to be listed in the 25,000 top movies, and

C = mean vote across the top 25,000 top movies (currently 7.0).

As in the movie domain, given the thousands of questions asked daily in SO, it is necessary to also rank the question posts, so the most relevant questions are selected for each user. The tags contained in the most relevant questions are assumed to be most indicative of the learning needs of the user. In this experiment, the parameters used in computing the value of w for each question posts are defined similarly to how IMDB defined their parameters. Since SO does not use a rating system, so, I employed the average tag count (shown in Equation (6.3)) used in the previous question posts by a user as a proxy to the rating value of IMDB. Just as the rating value represents the preference of a user for a movie compared to others, the average tag count, represent the preference of the user in learning more about a tag

[5] Rating in the IMDB context represents the preference of a user for a giving movie using a rating scale between [1, 10].

[6] Vote in IMDB are used by community members to indicate how helpful the reviews provided to a movie are.

compared to other tags. By employing the average tag count, the assumption is that a user who keeps asking questions about specific tags has learning needs related to those tags. Therefore, the tags with higher average tag count are more likely an important indicator of learning needs of the user. Since information about the previous tag usage by a user alone does not capture what other professionals are asking questions about now, the information about the past tag usage of a user alone is not sufficient. Therefore, I also incorporated the votes earned by a question post as indicated by other users in SO. Of course, a poorly asked question, even if it contains all the tags a user has used in the past, is not likely to interest the question asker.

In the SO scenario, the parameters used in computing *w* were applied to rank the 100 question posts selected for each user as stated below:

w = weighted tag count,

R = average tag count used in the previous question posts by the user over the baseline period,

v = number of votes earned by the respective question post being ranked, that is the summation of upvotes and downvotes awarded to the question,

m = minimum number of votes among all the 100 selected question posts for each user. The value of *m* would vary according to the 100 most popular question posts selected for each user,

C = the mean votes across all the selected questions posts for all the top 100 questions selected for each user.

In computing *R*, if a question has *java* and *android* tags, I compute the average number of times *java* or *android* tags have appeared in the previous question posts of the user. Specifically, *R* is computed as shown in Equation (6.3) below:

$$R = \frac{\sum_{i=1}^{k} f_i}{k} \qquad \text{Equation (6.3)}$$

where f_i is the count of the number of questions among the questions asked by a user in the past in which tag_i, occurs, and

k is the number of tags in the current question post being ranked.

Since, R is computed based on the average tag count used by a user for a set of questions asked, the preference of the user is put into consideration. Also, as C reflects the mean votes across all the selected question posts for all the questions, the quality of the question is considered as indicated by other users in the community. Hence, the weighted tag count w computed considers both the preference of the user and the preference of the professional community in ranking the question posts selected. Moreover, by using the TBE approach the weighted tag usage value can be estimated for each question post for each user even though its exact value is uncertain. Also, with the TBE approach even if a question contains tags a user has never used, the value of C would still act as a default to estimate the frequency of usage of a tag by the user.

For each user, the weighted tag count computed as shown in Equation (6.2) is used to rank the 100 posts selected for that user. The next goal is to select the top ranked question posts that contain tags that represent the future unperceived learning needs of a user. To select the top ranked questions for each user, I needed to choose a threshold value of w that would apply to all users. As the average tag count R computed for each user and for each question post has no fixed range, hence the ranges of the value of w computed across the question posts selected for each user varies. Therefore, I had to normalize the weighted average count w computed for each answer post for each user so it ranges from 0 – 1 as defined in Equation (6.4) below:

$$\text{NWTU} = \frac{(w_{Actual} - w_{Min})}{(w_{Max} - w_{Min})} \qquad \text{Equation (6.4)}$$

where w_{Actual} = actual weighted tag usage computed for a question post for a user,

w_{Min} = minimum weighted tag usage across all the 100 selected posts for a user, and

w_{Max} = maximum weighted tag usage across all the 100 selected posts for a user.

The normalized weighted tag usage (NWTU) value is computed as shown in Equation (6.4) for each question post, so that the NWTU value ranges from [0, 1]. I eventually selected questions which would represent the future learning needs of a user by experimenting with various NWTU threshold scores between [0, 1]. For instance, using a NWTU threshold value of 0.7, for each user, the tags from all question posts with normalized score above 0.7 would be predicted as the future unperceived learning needs of the user. For each user u, the predicted tags were compared with the actual tags used by the user in the period right after the baseline by computing the precision, recall and f-measure scores. For all the users considered in this study, the average of the precision, recall and f-measure scores computed for each user were thereafter computed as shown below:

$$Precision = \frac{1}{n}\sum\left(\frac{f}{g}\right)^{u} \quad \text{Equation (6.5)}$$

$$Recall = \frac{1}{n}\sum\left(\frac{f}{h}\right)^{u} \quad \text{Equation (6.6)}$$

$$F\text{-}measure = \frac{1}{n}\sum\left(\frac{2 \times Precision \times Recall}{Precision + Recall}\right)^{u} \quad \text{Equation (6.7)}$$

where, f = total number of relevant items retrieved, i.e. the total number of tags correctly predicted,

g = total number of items retrieved, i.e. the total number of tags correctly predicted (true positive) plus the number of tags wrongly predicted (false positive),

h = total number of relevant items, i.e. the total number of tags correctly predicted (true positive) plus the number of tags falsely predicted as not used (false negative), and

n = total number of users in the subset of SO data being considered.

Since F-measure is computed using both precision and recall, it allows the overall effectiveness of the tag prediction to be determined. The results obtained are presented in the next section.

6.2 Results

The 100 posts selected for each user were ranked using the TBE approach as computed in Equation (6.2). To predict the unperceived learning needs of each user, the 100 selected posts were further filtered so that only the relevant tags would be predicted. In filtering the posts, I experimented with varying NWTU values between [0.1, 1]. For each NWTU value employed in this experiment, the tags used in the questions with NWTU value above the threshold score were used infer the future unperceived needs of the user. The results were evaluated by comparing the actual tags used by a user to the tags predicted to represent the future learning needs of the user. For instance, with the short-term prediction in which data were gathered over March-July 2014, the tags predicted that a user would ask questions about were compared with actual tags the user asked questions about in August 2014. Similarly, for the long-term prediction in which data were gathered over the long-term 3-year baseline Jan 2009-December 2011, all the tags used by an individual user in the year 2012 were compared with the predicted tags. The average recall, precision and f-measure scores obtained by experimenting with varying NWTU score thresholds using the long-term time frame are shown in Table 6.1.

Table 6.1 Results Obtained using Various Normalized Weighted Scores for the Long-Term Baseline

Normalized Weighted Tag Usage Score Threshold	Recall	Precision	F-Measure
1	0.6	0.90	0.67
0.9	0.64	0.82	0.66
0.8	0.67	0.73	0.62
0.7	0.70	0.61	0.57
0.6	0.73	0.49	0.50
0.5	0.76	0.36	0.41
0.4	0.77	0.27	0.35
0.3	0.77	0.23	0.32
0.2	0.78	0.23	0.32
0.1	0.79	0.22	0.32

Results obtained in Table 6.1 show there is a tradeoff in precision versus recall values. Using a threshold score that is very low (e.g 0.1) reduces the precision as many tags are selected that do not represent the future learning needs of the user. Likewise, using a very high threshold (e.g. 1) there are also a good chance that important tags representing the unperceived learning needs of the user could be missed.

In order not to forfeit either recall or precision, a threshold score of 0.7 seems to be appropriate, as its recall value is approximately the mean among recall values and the precision value is a big jump up from a .6 threshold. Thus, I chose 0.7 as the threshold when predicting the tags, a user will be asking questions about in the future using the long-term and short-term baseline. To avoid redundancy in the results, the same threshold value was employed with both the long-term and short-term baseline. The average recall, precision and F-measure for the long-term and short-term baseline predictions with a NWTU threshold of 0.7 are shown in Table 6.2.

Table 6.2. Evaluation of Results

Time Duration	Recall	Precision	F-measure
Long-Term	0.70	0.61	0.57
Short-Term	0.93	0.81	0.83

Higher precision and recall values were observed with the predictions made using the short-term learning data as compared with the long-term data. Hence, the results obtained in Table 6.2 reveal that employing the short-term information about a user is likely not only sufficient but even better than using long term information to provide adapatable support to the user. This isn't too surprising since learning needs evolve and change as the user gains more knowledge or forgets things.

6.3 Discussion

As knowledge in a profession changes, the unperceived learning needs of each user evolve. Any technology that must support professionals should be adaptable to such changes. Therefore, this study takes a preventive approach to detect the unperceived needs of users even before they become evident to them. If these users could act on their unperceived needs earlier, it would help minimize the growth in their learning needs. Also, supporting users to diagnose their future unperceived needs could be useful in recommending learning resources that would help in meeting those needs earlier.

This study raises the possibility of creating a system that would inform a user about their unperceived learning needs. In the experiment, the future unperceived needs of each user were predicted using a short-term and a long-term baseline using information about the user and learning trends within the community. In Chapter 4, a marginal difference was obtained (as shown in Figure 4.2) in the quality of answers provided over a 6-month time frame. The results shown in Table 6.2 show a wider range of performance between using short-term (5 month) data about the user versus longer-term data (3 years). Results in Table 6.2 also show that over the passage of 3 years, the learning needs of the users have evidently evolved considerably given the much better results obtained with data gathered much more recently over the previous 5 months. In sum, in predicting what the user would want to learn in the future, employing shorter term information about the user's past behaviour proves to be more effective.

Of course, there are limitations in the approach employed in this study. First, in inferring the current unperceived needs of users, only the top most computing class was considered. However, each of the computing classes is broad with lots of related tags depicting the diverse relevant knowledge areas related to the computing class. Also, this experiment was carried out only within the SO community, so the results obtained from this experiment might not be generalizable to other online learning communities. This experiment was reported in EC-TEL 2016 (Ishola and McCalla 2016b).

CHAPTER 7

RECOMMENDING PEERS TO MEET THE LEARNING NEEDS OF USERS

A main step in supporting professional learners to improve their expertise is to detect their unperceived learning needs, which was exactly the goal of Chapters 5 and 6. The results obtained in Chapter 4 also establish the need to support users within an OLC in achieving positive help-seeking experience. One possible approach to achieving both goals is to use a peer help recommender system, a common tool in OLCs to suggest prospective answerers to questions. In this chapter, the main goal is to explore how to predict users who will provide timely answers to the questions asked by a user. As I did in the last chapter, I will show how to track the online learning activities of question-askers within an OLC to predict prospective answerers who will provide timely and high-quality answers to the questions.

Specifically, the goal of this chapter is to shed light on how to provide personalized support to help individual users meet their *expressed learning needs*. The expressed learning needs of a user in SO are evidenced by questions asked by the user. Just as it is important to have the learning needs of users met by providing quality answers to their questions, it is also important to ensure the answers are received promptly. This chapter addresses the problem of a user who posts questions in SO and either receives an answer too late to be of use or, worse, receives no answer. Questions which receive late answers or are unanswered, not only deprive the user himself or herself of useful feedback but also deprive the entire community of such feedback. This is an increasingly serious problem as shown in Figures 3.3 and 3.4. Two experiments were performed in supporting users within SO to receive timely answers to their questions, as described in Section 7.1 and Section 7.2.

7.1 Towards Recommending Prospective Peer Helpers to Provide Timely Answers

In earlier experiments reported in Chapters 5 and 6, attempts were made to detect the unperceived learning needs of users using lightweight approaches. Results obtained in Chapter 4 show that users who earned *Tumbleweed badges* (users who asked questions but didn't receive answers) do not have their learning needs met. Obtaining timely answers to questions is important (Bhat et al., 2014) in enhancing the sustainability of an online learning community. However, the answer response times to questions have increased as shown in Figure 3.7. Sometimes the question-askers answer their questions themselves, which can deter users from continuing to use SO. In this experiment, these issues are addressed by predicting prospective users who are likely to provide the timeliest answers to a question. Such users could then be immediately connected to the person asking the question to speed up the response time.

Research efforts have succeeded in the past in predicting potential peer helpers within a classroom-learning environment which encompasses just hundreds of students (Greer et al., 1998b; Merrill, Reiser, Trafton, and Ranney, 1992; Vassileva et al., 1999). A new challenge arises in an online learning community where learning is unstructured with thousands or millions of potential helpers with varied expertise and learning interests. There is a need for an appropriate recommendation technique that scales up to millions of users and aligns with the interests and skills of the helper. The availability, helpfulness, technical ability and social ability of the helper were employed as strategies considered in selecting an appropriate peer helper (Greer et al., 1998b; Greer et al. 1998a; Merrill et al., 1992; Vassileva et al., 1999). The experiment reported in this section will augment such research in providing peer helper seeking strategies that scale to very large numbers of users. The methodologies employed in this experiment are described below.

7.1.1 Methodology

Greer et al. (1998b) built a peer-help system to help computer science students find potential peer helpers among their classmates who are ready, willing and able to help in overcoming impasses. Vassileva et al. (1999) in their study with iHelp incorporated the social characteristics of the helper into determining an appropriate helper. The social characteristics of the helper were gleaned from the online activities of the helper such as votes received by the helper, questions asked, answers provided, and the marks received on assignments. This experiment seeks to predict potential just-in-time peer helpers using five measures for choosing such a helper, drawn from those employed by Greer et al. (1998b) and Vassileva et al. (1999).

Each of these five measures considers the relevance of the question to the past online activities of the prospective helpers. Based on the demonstrated knowledge of prospective helpers in answers they have given in the past, a score is computed for each measure. The score is computed by the co-occurrence of tags contained in the question with tags contained in the answers provided by the prospective helper in the past. Further, I incorporated a *timeliness* criterion which considers how quickly the prospective helpers have provided answers in the past. For each measure, personalized scores are assigned to each prospective helper based on their suitability to answer a question, as described below.

- *Regularity*: Regularity is defined by how often the prospective answerer provided an answer to a question on a given topic (defined in terms of the tags on the question). For instance, if a question contains *java* tags, I would count the number of questions answered containing *java* tags for the month before the question was asked. The higher the regularity of interaction with relevant questions in the past, the more likely the user would be to answer the question. The regularity score was computed by counting the number of answer posts A relevant to the question tag_i for user u as shown in Equation (7.1) below:

$$Regularity = \sum_u A_i \qquad \text{Equation (7.1)}$$

- *Knowledgeability*: Knowledgeability is defined as the know-how of the prospective answerer about a given topic. The knowledgeability measure shows how much a prospective helper knows about the question based on the total score *S* that user *u* has earned in answering past questions about tag_i as shown in Equation (7.2) below:

$$Knowledgeability = \sum_u S_i \qquad \text{Equation (7.2)}$$

Prospective helpers with a higher number of up votes would be ranked as better based on this measure.

- *Eagerness*: Eagerness is the keenness of the prospective answerer to deal with a topic compared to other topics. The eagerness measure depicts the probability that a user will answer a question related to the question tag, tag_i.

$$Eagerness = \frac{\sum_u A_i}{\sum_u A} \qquad \text{Equation (7.3)}$$

where $\sum_u A$ represents the total number of answers provided by the user to all questions. The eagerness measure seeks to assess the interest of the user in answering questions related to tag_i by considering the proportion of relevant questions answered. Prospective helpers with higher scores are ranked higher.

- *Willingness:* The willingness measure depicts the conditional probability that a user u will provide an answer to a question given that tag_i was used in the question. Bayes rule is applied in computing this peer matching measure. Bayes rule is defined as shown in Equation 7.4 for event X given event Y (Murphy, 2012).

$$P(X = x|Y = y) = \frac{p(X)\, p(Y|X)}{p(Y)} \qquad \text{Equation (7.4)}$$

In computing the willingness measure, the probability that a user u would provide an answer to a question given that the question is about tag_i is computed as shown in Equation (7.5).

$$p(U_a|tag_i) = \frac{p(U_a)\, p(tag_i|U_a)}{p(tag_i)} \qquad \text{Equation (7.5)}$$

where $p(U_a)$ is the probability of user u answering a question,

tag_i is a tag labelling the question,

$p(tag_i|U_a)$ is the likelihood that an answer to a question about *tag*$_i$ will be given by user u, and

$p(tag_i)$ is the probability that a question related to *tag*$_i$ will be asked in SO (this is the same for all prospective helpers).

$p(U_a)$ and $p(tag_i|U_a)$ can be computed as shown below:

$$p(U_a) = \frac{\sum_u A}{\sum_{j=1}^{n} A} \qquad \text{Equation (7.6)}$$

$$p(tag_i|U_a) = \frac{\sum_u A_i}{\sum_u A} \qquad \text{Equation (7.7)}$$

where $\sum_u A_i$ is the total number of *tag*$_{i\ answers}$ by user *u*,

$\sum_u A$ is the total number of answers by user *u*,

$\sum_{j=1}^{n} A$ is the total number of answers by all users *n* .

To compute the willingness score I substituted values from Equation (7.6) and Equation (7.7) into Equation (7.5). Prospective helpers with higher willingness scores are ranked higher.

- *Recency:* The recency measure corresponds to how recently the prospective helper has provided an answer to questions about *tag*$_i$. The recency measure is defined by the creation date and time of the last answer post provided by each prospective helper to questions about *tag*$_i$. To compute the recency score, the creation date and time of the last answer post about *tag*$_i$ as recorded in SO is extracted for each user as shown in Equation (7.8) below:

$$Recency = argmax_u[T_i^a] \qquad \text{Equation (7.8)}$$

where T_i^a represents the distribution of the dates and times that user *u* has provided answers to questions about *tag*$_i$. For each user, the date and time of the latest answer post is the recency value. With the recency measure, a user is expected to have at least answered one question about tag_i, which is exactly the case in this study as only active question-answerers were considered. Under this measure, prospective helpers who have answered related questions more recently would be ranked higher than those who answered such questions earlier. Greer et al. (1998b) argued that helpers who have recently provided help should be exempt, to avoid overworking a peer helper. In SO, there is a possibility that users who recently provided help might still be willing to provide help to earn some incentive (e.g. reputation score or various badges) from the forum.

7.1.2 Results

7.1.2.1 Evaluation of the Ranking Measures

In this section, the goal is to explore the effectiveness of the different measures described above in terms of their ability to predict a relevant peer-helper who will provide quick answers. The effectiveness of each measure was evaluated using the historical SO data about each prospective helper. Historical data going back 1 month, 3 months and 6 months from when the time a question was asked were employed. For this study, only *java*[7] questions (53,731) that received at least one answer within the first hour of creation were considered with 254,766 prospective helpers to choose from. These represent questions that would provide a good rationale in evaluating the effectiveness of the measures in predicting the timely answerers. Also, to keep the computation modest, although a question could be defined with more than one tag, only a *java* tag was used in the computation of the score for each of the five measures. Likewise, in a real-life situation only users who were available online within the first hour the question was posted could be regarded to be users from who prospective helpers could be chosen. They are the set of users more likely to view the questions earlier and provide a quicker response.

There also needs to be a success measure for the predictions. As in the study by Tian et al. (2013), I deem it a *success* if a user in the top N ranked users computed by a measure is also a user who answered the question under consideration in SO. The success rate S@N for

[7] Questions containing *java* tags were focussed on as this is the most used programming related tag in SO.

each measure can then be computed by dividing the total number of successes by the total number of questions as shown in Equation (7.9) below.

$$S@N = \frac{\sum_{i=1}^{s} U_a = \bar{U}_a}{p} * 100\% \qquad \text{Equation (7.9)}$$

where U_a are the users predicted by the measure to be able to answer the question,

$\bar{U}_a$ is the actual answerer for the question as recorded in the SO data,

s is the total number of accurate predictions where the actual answerer is in the top N ranked users, and

p is the total number of question posts.

Different values of *N* were used to get insight into how the prediction would perform as the number of prospective helpers predicted increases. The *N* values employed were *N* = 1, 5, 10, and 20. Finally, the effectiveness of the measures was compared using three evaluation criteria as stated below:

- predicting the answerer who responded first in SO,
- predicting the answerer who gave the first answer accepted by the user who asked the question, and
- predicting the best answerer, the user who gave the answer with the highest score.

***Predicting the first answerer*:** The ranked lists of prospective helpers predicted for each measure were evaluated with the aim to know their effectiveness at predicting the first answerer. The results obtained in predicting the first answerer using the historical information of 1 month, 3 months, and 6 months are presented in Tables 7.1 – 7.3 below.

Table 7.1. Success Rate at Predicting the First Answerer with 1 Month Data

Peer-Help Measure	1 Month			
	S@1	**S@5**	**S@10**	**S@20**
frequency	5.4%	18.9%	31.7%	49.1%
recency	2.4%	11.3%	20.3%	33.6%
eagerness	1.8%	9.9%	21.3%	43.6%
knowledgeability	5.6%	18.0%	28.1%	39.5%
willingness	5.7%	21.1%	35.9%	**54.2%**

Table 7.2. Success Rate at Predicting the First Answerer with 3 Months Data

Peer-Help Measure	3 Months			
	S@1	**S@5**	**S@10**	**S@20**
frequency	5.3%	18.9%	31.4%	48.2%
recency	2.6%	11.7%	20.7%	34.0%
eagerness	1.9%	10.1%	21.5%	43.8%
knowledgeability	5.5%	17.9%	28.1%	39.3%
willingness	5.6%	21.1%	35.4%	**52.9%**

Table 7.3. Success Rate at Predicting the First Answerer with 6 Months Data

Peer-Help Measure	6 Months			
	S@1	**S@5**	**S@10**	**S@20**
frequency	5.8%	20.0%	33.1%	50.8%
recency	2.8%	12.7%	21.8%	35.6%
eagerness	2.0%	10.3%	23.2%	47.0%
knowledgeability	6.0%	19.0%	29.8%	41.9%
willingness	6.1%	22.4%	37.4%	**55.9%**

The results in Tables 7.1 –7.3 show that the willingness of a prospective helper has the highest success rate of about 56% with S@20 using a time frame of 6 months. Also, the results obtained using the time frame of 6 months resulted in higher success rates for all measures. However, comparing the results obtained using the 1-month (as shown in Table 7.1) versus 6-month time frame (as shown in Table 7.3), overall only a marginal difference is observed in the success rate.

Predicting the accepted answerer**:** In SO, from the numerous answers provided to a question, the question-asker can mark only one of the answers as accepted. The goal of this evaluation criterion is to determine the success of the measures at identifying the accepted answerer from the ranked list of prospective helpers suggested. The results obtained are shown in Tables 7.4 –7.6 below.

Table 7.4. Success Rate at Predicting the Accepted Answerer with 1 Month Data

Peer-Help Measure	**1 Month**			
	S@1	**S@5**	**S@10**	**S@20**
frequency	5.3%	19.6%	31.8%	48.3%
recency	2.9%	12.2%	21.2%	33.9%
eagerness	1.8%	9.4%	20.0%	41.0%
knowledgeability	5.6%	19.2%	29.2%	40.7%
willingness	5.6%	21.4%	35.4%	**52.5%**

Table 7.5. Success Rate at Predicting the Accepted Answerer with 3 Months Data

Peer-Help Measure	**3 Months**			
	S@1	**S@5**	**S@10**	**S@20**
frequency	5.3%	19.6%	31.7%	47.3%
recency	3.2%	12.6%	21.5%	34.4%
eagerness	1.9%	9.76%	20.7%	41.4%
knowledgeability	5.6%	18.99%	29.3%	40.7%
willingness	5.6%	21.30%	35.1%	**51.5%**

Table 7.6. Success Rate at Predicting the Accepted Answerer with 6 Months Data

Peer-Help Measure	6 Months			
	S@1	S@5	S@10	S@20
frequency	5.8%	20.8%	33.3%	50.3%
recency	3.5%	13.7%	22.8%	36.1%
eagerness	2.0%	9.9%	22.1%	44.6%
knowledgeability	6.0%	20.3%	31.2%	43.5%
willingness	6.0%	22.8%	37.3%	**54.6%**

Again, the willingness measure has the highest success rate of about 55% with S@20 using the 6 months defined time line. The recency measure has the lowest success rate suggesting the need to have well-rested answerers. Again, only marginal differences are observed in the success rates obtained in Tables 7.4 and 7.5 compared to Table 7.6.

***Predicting the best answerer*:** Using this evaluation criterion, the effectiveness of the peer matching measures at predicting the user with the highest score was examined. Results from this evaluation are shown in Tables 7.7 – 7.9 below.

Table 7.7. Success Rate at Predicting the Best Answerer with 1 Month Data

Peer-Help Measure	1 Month			
	S@1	S@5	S@10	S@20
frequency	5.4%	20.0%	32.5%	49.4%
recency	2.9%	12.3%	21.6%	34.9%
eagerness	1.8%	9.3%	20.1%	42.1%
knowledgeability	5.8%	20.0%	30.3%	41.7%
willingness	5.7%	21.7%	36.2%	**53.6%**

Table 7.8. Success Rate at Predicting the Best Answerer with 3 Months Data

Peer-Help Measure	3 Months			
	S@1	**S@5**	**S@10**	**S@20**
frequency	5.5%	20.3%	32.5%	48.4%
recency	3.2%	12.9%	22.1%	35.3%
eagerness	1.9%	10.0%	21.2%	42.9%
knowledgeability	5.8%	19.9%	30.4%	41.8%
willingness	5.7%	21.9%	36.1%	**52.8%**

Table 7.9. Success Rate at Predicting the Best Answerer with 6 Months Data

Peer-Help Measure	6 Months			
	S@1	**S@5**	**S@10**	**S@20**
frequency	5.9%	21.5%	34.3%	51.4%
recency	3.7%	14.1%	23.4%	37.1%
eagerness	2.0%	10.2%	22.6%	46.1%
knowledgeability	6.1%	21.2%	32.3%	44.6%
willingness	6.1%	23.5%	38.5%	**56.0%**

The willingness of the prospective users has the highest success rate at predicting the best helper. The highest success, with a success rate of 56% using S@20, was obtained using the 6 months defined timeline. Again, there aren't too many differences in the success rates achieved for all the measures using 1-month data compared to using 6 months data.

Overall, over the three evaluation criteria, the highest success rate was with the willingness measure, and the lowest success rate was with the recency measure. In addition, as the number of months increases from 1 to 6 months, no significant difference in the success rate for the measures was seen. Nevertheless, Tables 7.1 – 7.9 show (unsurprisingly) that as N increases, the success rate of the prediction also generally increases. Comparing all three evaluation criteria, the highest success was achieved while predicting the best answerer, although the success rate obtained with the other criteria did not differ significantly. The poor success rates achieved using the recency measure with the three evaluation criteria, show the need to have well rested helpers while designing a peer recommender system. As the recency measure shows little promise in achieving reasonable success rate in predicting the prospective answerers, subsequently I will exempt it from the other measures considered in this study. As discussed in the next section, attempts were made to improve the success rates achieved by including an additional measure called *timeliness*.

7.1.2.2 Prediction of Timely Helpers

The main goal of this study is to predict timely helpers, i.e. helpers who would provide answers quickly. To predict the timely helpers, I excluded the recency measure as it is the weakest measure as shown in Tables 7.1 - 7.9 and I added a new *timeliness* measure. Thereafter, I explored the impact of the predictions of combining timeliness measure with each of the other measures described in 7.1.1. I included a *timeliness* criterion that considers how quickly a prospective helper would provide an answer to a question. A 15-minute time frame was used as it represents a "reasonable" duration in which to get a timely answer. Although, the percentage of questions answered within this time frame of 15-minute has been decreasing as shown in Section 3.2.2. For each prospective helper, the timeliness measure was computed as shown in Equation (7.10):

$$Timeliness = \frac{\sum_u A_{t\leq 15}}{\sum_u A} \qquad \text{Equation (7.10)}$$

where $\sum_u A_{t\leq 15}$ represents the number of questions the user answered within 15 minutes, and $\sum_u A$ represents the total number of answers provided by user u.

The timeliness score obtained with Equation (7.10) by each user was combined with their respective score on each of the other measures except for the recency measure. I excluded the recency measure in this prediction, as it is the weakest measure as shown in Tables 7.1 - 7.9. Moreover, the recency score computed as shown in Equation (7.8) is a date and time value that cannot be multiplied by the timeliness score, as can the numeric values obtain with other measures.

To keep the computation modest, the timeliness score and the score computed for each of the other measures were combined using the product-function. For example, product-function (regularity, timeliness) is computed by multiplying the regularity score and the timeliness score. Employing the product-function helps ensure that the prospective helpers selected have a score greater than 0 for each measure considered. Hence, with product-function (regularity, timeliness), the prospective helpers selected must have regularity and timeliness score above 0; they must be regular and timely answerers.

Finally, since no major differences were observed when the 1-month history data of the prospective helper were used as compared to the 6-month history data for the three evaluation criteria. So in predicting the timely helpers, I only employed the 1-month history data about the prospective answerers in computing the score for each measure. Using the 1-month time frame also saves a lot of computational time. The results obtained are shown in Tables 7.10 - 7.12 for each of the evaluation criteria.

Table 7.10. Timeliness Success at Predicting the First Answerer

Peer-Help Measure	**1 Month**			
	S@1	**S@5**	**S@10**	**S@20**
Product-function (regularity, timeliness)	6.5%	21.9%	36.2%	55.0%
Product-function (eagerness, timeliness)	5.5%	26.7%	43.3%	**63.2%**
Product-function (knowledgeability, timeliness)	6.1%	20.1%	30.5%	41.5%
Product-function (willingness, timeliness)	6.9%	24.9%	40.6%	60.3%

Table 7.11. Timeliness Success at Predicting the Accepted Answerer

Peer-Help Measure	**1 Month**			
	S@1	**S@5**	**S@10**	**S@20**
Product-function (regularity, timeliness)	6.1%	21.0%	34.1%	50.8%
Product-function (eagerness, timeliness)	3.8%	21.0%	35.3%	53.8%
Product-function (knowledgeability, timeliness)	5.9%	20.2%	30.4%	41.5%
Product-function (willingness, timeliness)	6.5%	23.6%	37.9%	**55.3%**

Table 7.12. Timeliness Success at Predicting the Answerer with the Best Answerer

Peer-Help Measure	1 Month			
	S@1	S@5	S@10	S@20
Product-function (regularity, timeliness)	6.2%	21.5%	35.0%	51.9%
Product-function (eagerness, timeliness)	3.9%	21.5%	36.5%	55.3%
Product-function (knowledgeability, timeliness)	6.0%	21.1%	31.4%	42.5%
Product-function (willingness, timeliness)	6.5%	24.3%	38.9%	**56.7%**

The results obtained from each measure described in Tables 7.1 - 7.9 were improved upon, by including an additional criterion called *timeliness*. A maximum success rate of about 63% was achieved in predicting the first answerer using S@20 as shown in Table 7.10. Using the willingness measure, success rates of 55% and 57% were achieved in predicting, respectively, the accepted answerer, and the best answerer. These values represent an improvement in the success rate. For instance, in predicting the first answerer using the 1-month time frame the success rate improved from 44% to 63% with the eagerness measure. Likewise, in predicting the accepted answerer and the best answerer respectively, the success rate improved from 53% to 55% and from 54% to 57% using the 1-month data with the willingness measure.

7.1.3 Discussion

The first step in supporting users to meet their expressed needs is to provide prompt answers to their questions. As Kay and Kummerfeld (2009) have identified, there is a trade-off between the usefulness of retaining older information about a lifelong learner and preserving only the recent data. The results show that employing 6 months of data about the learner was only marginally better when compared to the results achieved using 1-month information. In addition, the initial results from this study (from Tables 7.1 - 7.9) shows that *willingness* of peers to provide help to questions has the highest success rate. A success rate of 54% using S@20 was achieved using the 1-month time frame with the willingness measure in predicting the first answerer.

To improve this success rate, the timeliness criterion was introduced. The timeliness criterion considers the probability that a user will answer a question quickly. By incorporating the timeliness criterion, the success rate increased from 54% to about 63% in predicting the first answerer using S@20. While these results likely require further improvement if they are to lead to practical heuristics for predicting appropriate helpers, these values are an improvement over the previous work by Tian et al. (2013) who obtained a success rate of 13% and 23% using S@20 and S@100 while predicting the best answerer using the topic modelling approach. The results obtained in this study for every measure outperforms this previous work. The variation of the results obtained from those of Tian et al. (2013) is presumably because this study was restricted to questions answered within the first hour the question was created. These questions were focussed on because the goal of this study is to predict the just-in-time helpers who will provide quick answers to the questions, meaning that, questions answered late would not suffice.

Despite the improvement in the results obtained in this experiment over the previous work by Tian et al. (2013), there are obvious limitations in the approach employed. First, only questions that received answers within 1 hour were considered in this study. Second, in computing the score for each measure, only one tag was employed rather than using all the tags employed when the question is asked. In a follow-up study as described in Section 7.2, possible exploration of how all the tags included in the question can be used in computing the score will be considered. Employing all the criteria could provide an opportunity to improve upon the results obtained, something that is explored in the next section. This experiment was reported in EDM 2017 (Ishola and McCalla, 2017b).

7.2 Towards Reducing the Answer Response Time to Questions

As in Section 7.1, the goal of this section is to describe methods to predict answerers who can provide answers to user questions. Unlike Section 7.1, in this follow up study answerers will be predicted for questions that received late answers rather than considering only the questions answered within an hour. Predicting the prospective answerers for questions that received their first answer much later is a more difficult problem. This is because the actual question-answerer might have been offline when the question was asked. Having to consider the users who might have been offline when a question was asked implies the need to select the prospective answerers from a much larger pool of active question-answerers. Another goal of this study is to improve upon the success rate of 63% achieved in Section 6.1 for.

According to Bull, Greer, McCalla, Kettel, and Bowes (2001), it is important to ensure that the help provided is not only just-in-time but also targeted at achieving the learning goals of the help-seeker. In designing a peer recommender system, knowing the learning goals, prior knowledge of the helpers, their learning preferences, and historically successful learning paths are important (Drachsler, Hummel and Koper, 2008). Following up on this insight, in predicting the best and most timely answerers in this experiment, the past answer posts of each prospective answerer for a question were analyzed. To predict the prospective answerer, tag-based, response-based and hybrid approaches were employed. Specifically, each approach tracks historical information about the questions and the SO users using its own defined features and generates a ranked list of prospective answerers.

7.2.1 Methodology

In predicting the best and most timely answerers for questions, questions asked in SO in the period from January to May 2017 inclusive were used. Then the questions that received no answers were eliminated as there is no way to validate the answerers predicted would eventually be the actual answerers. Also, since I wanted to predict good answerers, I then focused only on questions answered by *active question-answerers*. As defined in Chapter 4.1, these are the reputable users between level 2 and level 4 activity levels as described in Table 4.1. So, the active question-answerers are users who have earned at least one of the tag badges for providing at least 20 quality answers to questions relating to a specific tag. For this study, I focused only on questions answered by active question-answerers with at least one *java* or *android* tag. By focusing on the questions answered by the active answerers, I hoped to avoid the cold-start problem in making the predictions (Drachsler, Hummel and Koper, 2008). The *java* and *android* tags cover broad areas of software and mobile development and are among the top 5 most used tags in SO. With these foci, I was left with 44,035 questions answered by 14,051 active question-answerers over the 5-month time frame from January to May 2017 inclusive.

Three general approaches were developed to predict the answerer to a question: *tag-based*, *response-based* and *hybrid* approaches. Using these three approaches, I then attempted to predict the prospective answerers to questions. For the experiment described in this section only the *best answerer*, the *timely answerer* and the *best* and *timely answerer* were predicted. The best answerer for a question is the answerer with the highest score. The timely answerer is the answerer who will provide an answer within 38 minutes from when the question was created. The best and timely answerer is the answerer who will provide the answer with the highest score within 38 minutes from when the question was created. The three approaches employed are discussed next. A threshold of 38 minutes was employed in defining the timely answerer as 38 minutes is the median response time for the first answer in SO as discussed in Section 3.2.2.

7.2.1.1 Tag–Based Approach

The tag-based approach focusses on the past usage of the tags attached to the question by all the active question-answerers for a given question (Ishola and McCalla, 2017b). For a given question, the *prospective answerers* were selected, who are those active question-answerers who had provided at least one answer to questions relating to any of the tags in the question up to a month before the question was asked. The one-month baseline was employed in tracking the historical information about the active question-answerers as the results in Section 6.1.3 show that tracking past activities of users beyond one month only provides marginal benefits. The selected prospective answerers are thereafter ranked using the score calculated by the four measures described below. Since a detailed description of each measure of the tag-based approach has been defined in Section 7.1.1, only a brief description is provided here. Although, the definition of the measures is the same as defined in Section 7.1.1, in computing the score for the measures in this section all the tags used in the question were considered.

- *Regularity:* Regularity is a count of the number of times a prospective answerer *u* has provided answers to questions containing each tag employed in the question as shown in Equation (7.11). For instance, if a question contains *java* and *android* tags, I would count the number of questions answered containing either *java* or *android* tags for the month before the question was asked.

$$Regularity = \sum_{ui=1}^{k} A_i \qquad \text{Equation (7.11)}$$

where A is the number of answers provided by user u to questions with a given tag, *tag*$_i$, and

k is the number of tags used in the question asked.

- *Knowledgeability:* Knowledgeability is the total score earned by a prospective answerer for providing answers to questions containing each tag employed in the question. Knowledgeability is computed as shown in Equation (7.12)

$$Knowledgeability = \sum_{ui=1}^{k} S_i \qquad \text{Equation (7.12)}$$

where S_i is the score obtained for providing answers to questions with a given tag, *tag*$_i$.

- *Eagerness:* Eagerness is computed by dividing the regularity score with the total number of answers provided by a prospective answerer as shown in Equation (7.13).

$$Eagerness = \frac{\sum_{ui=1}^{k} A_i}{\sum_{u} A} \qquad \text{Equation (7.13)}$$

where $\sum_{u} A$ represents the total number of answers provided by the user to all questions.

- *Willingness:* Willingness is defined by how active the prospective answerer is in providing answers about tags in the question. The willingness measure depicts the conditional probability that a user u will provide an answer to a question given all the tags used in the question. The willingness score is computed for all the tags in the question for each user by applying [Equation (7.5)].

$$p(U_a|tag_i) = \frac{p(U_a)\, p(tag_i|U_a)}{p(tag_i)} \qquad \text{[Equation (7.5)]}$$

where p(U_a) is the probability of user u answering a question,

tag_i represents all the set of tags used in the question,

$p(tag_i|U_a)$ is the likelihood that an answer to a question about tag_i will be given by user u, and

$p(tag_i)$ is the probability of a question related to tag_i will be asked in SO (this is the same for all prospective helpers).

$p(U_a)$ was computed as shown in Equation (7.6) and $p(tag_i|U_a)$ was computed as shown in Equation (7.7).

As the score computed for each measure described above has different ranges, to prevent a measure from dominating other measures when the scores are combined, the scores computed above for each measure are normalized to range between 0 and 1. The normalized score is computed as described in Equation (7.14) below:

$$Normalized\ Score = \frac{(S_{Actual} - S_{Min})}{(S_{Max} - S_{Min})} \qquad \text{Equation (7.14)}$$

where S_{Actual} represents a given prospective answerer's score,

S_{Min} represents the minimum score for all the prospective answerers selected for a given question, and

S_{Max} represents the maximum score for all the prospective answerers selected for a given question.

7.2.1.2 Response-Based Approach

With the response-based approach, the overall responsiveness of an active question-answerer to provide answers to questions regardless of the tags used in the question is computed. The overall responsiveness of an active question-answerer is judged based on all answers they had provided for a month prior to when the question was asked using the three measures described below:

- *Probability of First Answer (firstprob)*: Firstprob is the probability that a prospective answerer *u* provides the first answer to a question. Firstprob is computed as shown in Equation (7.15) below:

$$firstprob = \frac{\sum_{ui=1}^{k} A_i^{t=first}}{\sum_{ui=1}^{k} A_i} \qquad \text{Equation (7.15)}$$

where $\sum_{ui=1}^{k} A_i^{t=first}$ represents the number of questions for which the prospective answerer provided the first answer related to any of the tags ($tag_1 \ldots tag_k$) used in the question, and

$\sum_{ui=1}^{k} A_i$ represents the total number of answers provided by user *u* to any of the tags used in the question.

- *Probability of Fast Answer (fastprob)*: Fastprob is the probability that a prospective answerer *u* will provide an answer within 38 minutes (the median response time, as discussed in Section 3.2.2). Fastprob is computed as shown in Equation (7.16) below:

$$fastprob = \frac{\sum_{ui=1}^{k} A_i^{t\leq 38}}{\sum_{ui=1}^{k} A_i} \qquad \text{Equation (7.16)}$$

where $\sum_{ui=1}^{k} A_i^{t\leq 38}$ represents the number of questions that the prospective answerer answered within 38 minutes for any of the tags ($tag_1 \ldots tag_k$) used in the question.

- *Probability of Best Answer (bestprob)*: Bestprob is the probability that a prospective answerer will provide the best answer, the one with the highest score. Bestprob is computed as shown in Equation (7.17) below:

$$BestProb = \frac{\sum_{ui=1}^{k} A_i^{s_{max}}}{\sum_{ui=1}^{k} A_i} \qquad \text{Equation (7.17)}$$

where $\sum_{ui=1}^{k} A_i^{s_{max}}$ represents the number of questions for which the prospective answerer provided the answer with the highest score for any of the tags used in the question.

Since the scores computed for the three features above range from 0 to 1, there is no need to have the scores normalized. In predicting the actual prospective answerers for a question, the selected prospective answerers are ranked using the computed score from each of the three metrics described above.

7.2.1.3 Hybrid Approach

With the hybrid approach, the goal is to combine the scores from two measures[8]. Rather than forming hybrids for all combinations of measures, I focused on combining the measures with the top 2 success rates using the tag-based approach and the response-based approach. For example, for the tag-based approach, I selected the measures with the top 2 success rates and combined them using the functions described below. Likewise, the topmost measure from the tag-based approach was combined with the topmost measure from the response-based approach. The prospective answerers are then ranked according to their aggregate scores on the chosen hybrid measure.

The measures were combined using the 3 approaches described below:

- *Intersect*: with the intersect function only prospective answerers with non-zero scores on both measures are selected. The selected answerers are thereafter ranked by taking a sum of their score on both measures combined.
- *Product*: with the product function the product of the scores obtained from the two measures considered are computed to get an aggregate score. Users are ranked based on the aggregate score obtained from the two measures combined.
- *Sum*: with the sum function the sum of the scores obtained from the two measures considered is computed. The intersect ranking may differ from the sum ranking in that a user with a score of 0 using one measure, even if they ranked well in the sum list, would not appear on the intersect list.

[8] I also attempted to combine scores obtained from more than two measures, but it wasn't as effective as using just two measures. Hence, the results are not presented.

7.2.2 Results

The success of the prediction was determined by calculating the success rate (S@N) as defined in Equation (7.9). Like Tian et al. (Tian et al., 2013) *N* ranges from 10 to 100. Successfully predicting that the actual answerer is in the top 10 or even top 100 is not a trivial task given that there was a pool of about 14,000 active answerers to select from. Moreover, in a real-life recommender application, only a few of the predicted prospective answerers might be available and ready to answer the question.

7.2.2.1 Predicting Question-Answerers

The results obtained using the measures defined with the tag-based approach are shown in Tables 7.13 - 7.15 below.

Table 7.13. Predicting the Best Answerer Using the Tag-Based Measures

Tag-Based Measures	**Best Answerer**			
	S@10	**S@20**	**S@50**	**S@100**
Regularity	30.0%	42.7%	64.8%	80.3%
knowledgeability	24.3%	37.7%	63.4%	83.9%
Eagerness	19.5%	28.4%	45.3%	58.9%
willingness	45.9%	58.2%	75.7%	85.9%

Table 7.14. Predicting the Timely Answerer Using the Tag-Based Measures

Tag-Based Measures	**Timely Answerer**			
	S@10	**S@20**	**S@50**	**S@100**
Regularity	35.8%	49.5%	70.3%	84.0%
knowledgeability	29.7%	45.4%	70.3%	86.5%
Eagerness	12.9%	21.6%	40.8%	57.4%
willingness	37.3%	50.6%	71.0%	83.1%

Table 7.15. Predicting the Best and Timely Answerer Using the Tag-Based Measures

Tag-Based Measures	Best and Timely Answerer			
	S@10	S@20	S@50	S@100
Regularity	31.1%	44.0%	65.6%	80.9%
knowledgeability	24.9%	38.5%	64.0%	84.1%
Eagerness	19.4%	28.4%	45.7%	59.5%
willingness	37.3%	50.6%	71.0%	83.1%

As shown in Tables 7.13 - 7.15, higher success rates were achieved with increasing values of *N*. Amongst the four measures employed, the knowledgeability and the willingness measures had the top 2 highest success rates. Comparing the success rates of the measures over all values of *N*, the willingness feature seems to perform best, especially considering the results achieved at lower values of *N* also.

The results obtained using the measures defined with the response-based approach are shown in Tables 7.16 - 7.18 below.

Table 7.16. Predicting the Best Answerer Using the Response-Based Measures

Response-Based Measures	Best Answerer			
	S@10	S@20	S@50	S@100
fastprob	4.4%	7.8%	27.8%	62.5%
firstprob	7.5%	12.8%	31.6%	60.4%
bestprob	20.6%	24.0%	36.2%	61.8%

Table 7.17. Predicting the Timely Answerer Using the Response-Based Measures

Response-Based Measures	Timely Answerer			
	S@10	S@20	S@50	S@100
fastprob	7.1%	12.7%	41.2%	75.4%
firstprob	5.1%	9.7%	27.2%	55.0%
bestprob	11.1%	14.5%	26.0%	52.9%

Table 7.18. Predicting the Best and Timely Answerer Using the Response-Based Measures

Response-Based Measures	Best and Timely Answerer			
	S@10	**S@20**	**S@50**	**S@100**
fastprob	4.7%	8.5%	29.2%	63.6%
firstprob	7.6%	12.8%	31.5%	60.5%
bestprob	20.3%	23.7%	35.9%	61.5%

As shown in Tables 7.16 - 7.18, the fastprob and the bestprob measures had the top 2 success rates with the response-based measures.

I formed hybrids of the best measures from the tag-based and response-based approaches. The top 2 measures with the highest success rates with the tag-based and response-based approaches were combined using the functions defined for the hybrid approach as shown in Tables 7.19 - 7.21 below. Also, the topmost measure with the tag-based approach (willingness) and the topmost measure with the response-based approach (fastprob) were combined as shown in Tables 7.19 - 7.21 below.

Table 7.19. Predicting the Best Answerer Using the Hybrid-Based Measures

Hybrid-Based Measures	Best Answerer			
	S@10	**S@20**	**S@50**	**S@100**
Intersect-function (willingness, knowledgeability)	17.1%	29.0%	54.2%	75.7%
Product-function (willingness, knowledgeability)	45.5%	59.2%	78.5%	89.2%
Sum-function (willingness, knowledgeability)	41.5%	55.7%	76.6%	89.6%
Intersect-function (fastprob, bestprob)	1.0%	1.6%	10.0%	43.6%
Product-function (fastprob, bestprob)	3.2%	7.7%	30.7%	68.1%
Sum-function (fastprob, bestprob)	3.2%	7.8%	30.8%	69.0%
Intersect-function (willingness, bestprob)	0.9%	2.7%	19.2%	53.6%
Product-function (willingness, bestprob)	42.6%	55.0%	71.4%	83.8%
Sum-function (willingness, bestprob)	29.9%	37.0%	55.5%	76.8%

Table 7.20. Predicting the Timely Answerer Using the Hybrid-Based Measures

Hybrid-Based Measures	Timely Answerer			
	S@10	S@20	S@50	S@100
Intersect-function (willingness, knowledgeability)	20.4%	34.1%	58.6%	76.5%
Product-function (willingness, knowledgeability)	40.0%	54.5%	75.4%	87.7%
Sum-function (willingness, knowledgeability)	36.7%	52.3%	74.8%	88.7%
Intersect-function (fastprob, bestprob)	1.0%	1.6%	10.0%	43.6%
Product-function (fastprob, bestprob)	4.9%	12.0%	43.9%	77.8%
Sum-function (fastprob, bestprob)	4.9%	12.1%	43.7%	77.4%
Intersect-function (willingness, bestprob)	1.4%	4.3%	28.1%	63.2%
Product-function (willingness, bestprob)	43.0%	57.4%	76.1%	89.1%
Sum-function (willingness, bestprob)	31.9%	41.0%	63.2%	84.3%

Table 7.21. Predicting the Best and Timely Answerer Using the Hybrid-Based Measures

Hybrid-Based Measures	Best and Timely Answerer			
	S@10	S@20	S@50	S@100
Intersect-function (willingness, knowledgeability)	17.7%	29.8%	55.0%	76.1%
Product-function (willingness, knowledgeability)	45.8%	59.7%	78.9%	89.5%
Sum-function (willingness, knowledgeability)	41.8%	56.2%	77.0%	89.5%
Intersect-function (fastprob, bestprob)	0.8%	1.2%	7.9%	41.5%
Product-function (fastprob, bestprob)	3.4%	8.2%	31.7%	68.7%
Sum-function (fastprob, bestprob)	3.4%	8.2%	31.7%	69.6%
Intersect-function (willingness, bestprob)	1.3%	4.1%	27.1%	62.0%
Product-function (willingness, bestprob)	41.7%	56.1%	75.1%	88.5%
Sum-function (willingness, bestprob)	30.7%	30.7%	30.7%	83.7%

The success rates obtained for the response-based approach as shown in Tables 7.19 - 7.21 with S@10, S@20, and S@50 were poor for all the measures. With S@100 a decent success rate was obtained with fastprob, although not as high as the best tag-based measures. As shown in Tables 7.19 - 7.21, the intersect function had the lowest success rate, and isn't likely to be very useful. However, comparing the results obtained in Tables 7.19 - 7.21 with Tables 7.13 - 7.18, a marginally higher success rate was achieved with the best hybrid measure. Product (willingness, knowledgeability) generally outperforms the other hybrid-based measures although it is only marginally better than Sum (willingness, knowledgeability). In my subsequent analysis I will just use the Product (willingness, knowledgeability) hybrid at S@100.

7.2.2.2 Adopting Work Load Balancing in Predicting Answerers

Building upon the results obtained in Tables 7.1 – 7.9 wherein the recency measure performed the worst, I attempted to avoid overworking peers by exempting helpers who recently provided help. Reducing the number of prospective answerers creates a tradeoff between having rested helpers versus choosing helpers who can provide the best and most timely answers. I wanted to explore this tradeoff. So, my next goal was to determine in the SO data the effect on my predictions of "exempting" potential answerers who recently provided help from the potential list of prospective helpers. I experimented with various exemption intervals between the time the question was asked and the time the most recent answer was provided by a prospective answerer to any question. For instance, for an exemption interval of 15 minutes, I would exempt all prospective answerers who had provided an answer to any question within 15 minutes of when the question was asked. The remaining prospective answerers can then be ranked using whatever measure I am considering. The results obtained for exempting "overworked" helpers using Product (willingness, knowledgeability) with various exemption intervals are shown in Figure 7.1 below using S@100.[9]

[9] The data points in the figure between 0 minutes and 60 minutes are at 15 minutes and 38 minutes.

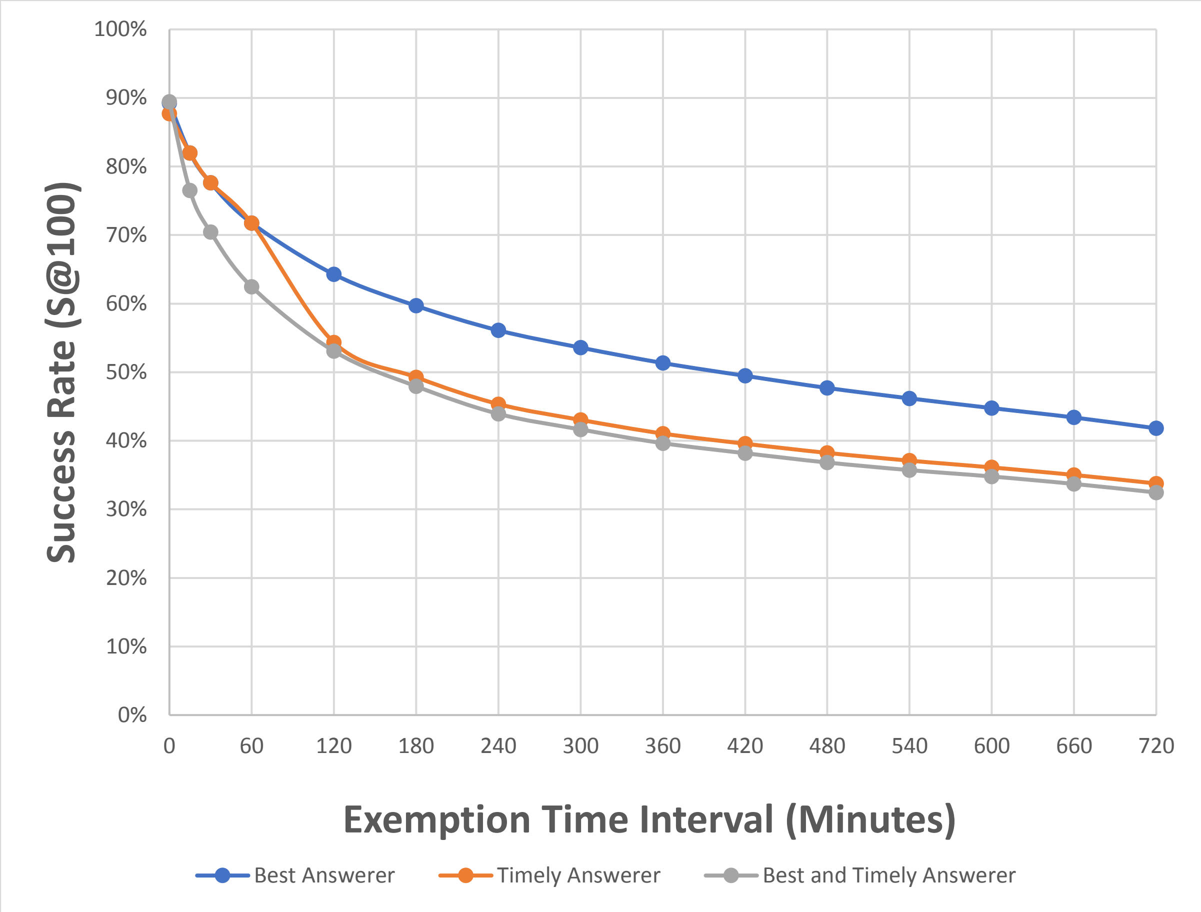

Figure 7.1. Success Rate Obtained with Varying Exemption Intervals for S@100

The exemption interval 0 as shown in Figure 7.1 represents the results obtained when no prospective answerers were exempted as shown in Table 7.21. As the exemption interval increases, the success rate steadily diminishes, reflecting the increasing likelihood that the actual answerer has been exempted and thus the prediction fails. However, the predictive success rate is still good up to a 60-minute exemption interval with over 60% for all the measures. Results in Figure 7.1 suggest there is the opportunity to exempt users for up to an hour after they have answered a question and still be able to recommend good helpers. Exempting frequent helpers from the selected prospective answerers would also encourage a more diverse set of question-answerers. The next goal is to study the performance of the best measure in predicting the answerers for questions answered late, assuming the prediction is to be made at the point the question is asked.

7.2.2.3 Providing Earlier Answers to Late Answered Questions

In this section I considered only the questions answered late. I wanted to know how much time can be saved if I succeeded at predicting the actual answerers to late answered questions at the point the question is asked. Once again, the best measure from the earlier analysis, Product (willingness, knowledgeability) with S@100, was used. Since there can be no timely answerers for questions that are answered late, for this analysis I only predicted the best answerer. I looked at various categories of late answers, ranging from just a little late with 38 Min < Response Time to First Answer (RTFA) <1 Hr all the way up to very late with RTFA >8 days. My goal was also to determine how the prediction accuracy at S@100 with the 60-minute exemption window compares to the prediction accuracy without exempting prospective answerers. The results are shown in Table 7.22 below.

Table 7.22. Success Rate Obtained with Varying Late RTFA Ranges

Product-function (Willing, Knowledgeability)	**38 Min < RTFA < 1 Hr**	**1 Hr < RTFA ≤ 1 Day**	**1 Hr < RFTA ≤ 3 Days**	**3 Days < RTFA ≤ 8 Days**	**RTFA > 8 Days**
S@100 Without the Exemption Interval	91.6%	93.6%	91.7%	86.8%	86.4%
S@100 With the Exemption Interval	78.7%	88.0%	88.3%	86.1%	79.6%

Overall, the success rates achieved with and without the 60-minute exemption interval are both relatively high through all the late answer categories. However, for answers that are just a little late (38 Min < RTFA <1 Hr), there is a bigger tradeoff with and without exempting the helpers, a drop from about 92% to 79%. Even for questions answered very late (RTFA > 8 days) a success rate of nearly 80% was achieved in predicting the best answerers who would also be rested. Hence, if the Product (willingness, knowledgeability) measure were incorporated in a peer recommender system, there is the possibility of saving the question-askers many days in waiting for an answer. If quicker answers can be provided to questions answered very late, not only will the question-asker benefit but other users within the community can gain insight from the answer.

7.2.3 Discussion

In building a peer recommender system, effective measures should be put in place to ensure that early response time and quality responses are provided to the help-seekers in meeting their learning needs. Such measures could be the basis for a peer recommender system to proactively find appropriate helpers. To ensure that the peer recommender system does not overload the same users with too many help requests, there is also a need to address the issue of "helper overload" in choosing prospective answerers. The issue of helper overload was addressed, by exempting helpers who had recently given help.

In this experiment, three general measures were explored: tag-based, response-based and hybrid-based. These measures allowed the prediction of the prospective answerers for questions asked in SO over a 5-month period. In exploring the efficacy of these measures, I achieved the best success rates using a hybrid approach that combines information about the user's willingness and knowledgeability to make the prediction. A success rate of about 90% was achieved using S@100 in predicting well-rested helpers who could provide both timely and quality answers to a question.

The results achieved in this experiment are an improvement over previous studies by Tian et al. (2013) who achieved a success rate 18% and 23% respectively using S@50 and S@100. Likewise, the results obtained in 7.22 are also an improvement over the previous results obtained in Tables 7.9 - 7.12 when a success rate of about 55% was achieved. The reason for the higher success rate achieved in this study compared to the results in Tables 7.9 - 7.12 is likely because I considered all the tags used in the question in computing the score for each measure. Also, the hybrid approach (where measures were combined) increased the success rate over the individual measures used alone. I believe with success rates nearing 90%, I already have the basis for effective helper recommendation, hence I did not explore more different hybrid techniques.

This study was carried out only on java and android related questions answered by active question-answerers, since I need some information about a question-answerer to make any reasonable predictions. I thus have no insight into those users who might have been good helpers but who were not active answerers. I hope that the same measures that allow predictions for active answerers also apply to other answerers, although such users must have some track record to be able to compute any of the measures. However, incorporating these measures into a recommender system that more proactively encourages question answering may stimulate a broader range of helpers. In a real-world context where developers had control over what capabilities to include in their model, it would also be possible to gather much more information about the answerers. The information gathered about the answerers could be used to form a richer long-term learner model to inform the components of the OLC environment, including the peer recommender system. Given the success rates achieved in this study using only the minimal information available, such rich models provide promise in achieving even more refined recommendations. This experiment was reported at AIED 2018 (Ishola and McCalla, 2018a).

CHAPTER 8

INFORMING A PEER RECOMMENDER SYSTEM

Most research in peer recommender systems in online learning communities (OLCs) is focused on the problem of identifying the answerers who can provide the best answers to a question. In Chapter 7 my goal was to predict the answerers who can provide answers soon after the question has been asked. In a real-world system, such peer recommenders would be especially useful if a learner would otherwise receive no answer or a late answer. In this chapter my aim to develop methodologies for predicting (when a question is asked) whether the question will receive an answer. If a question will be answered, I would also like to predict whether the answer will come early or late, i.e. before or after the median response time. In a large OLC with tens of thousands of questions asked daily, it is especially important to know the questions that would require the intervention of a peer recommender system. Recommending answerers for all questions in a large OLC might be expensive on the system resources. With the ability to identify (at the time the question is asked) questions that are likely to be answered late, then it would be possible to inform the recommender system. With the prompt intervention of the peer recommender system, the learner could be helped sooner.

In improving the help-seeking outcomes for learners, Aleven, Mclaren, Roll, and Koedinger (2006) employed an automatic help giving agent to provide feedback to students. With the I-Help system a currency system was introduced to foster peer-peer interactions between users (Bull et al., 2001; Greer et al., 2001). The experiment reported in this chapter extends previous studies by attempting to determine when the intervention of a peer recommender system would be required in supporting the users. Specifically, I hope to devise measures that would be useful in detecting early the questions that would take a longer time to receive an answer or, not be answered at all.

8.1 Methodology

In predicting whether a question will receive an answer and how soon such an answer will be provided (early or late), three general approaches were employed: using the features of the question asked (*question content-based* approach), using features of the active answerers (*answerer-based* approach) and using features that examine the popularity of the tags used in the question (*tag-popularity* approach). Among the three approaches 13 features were employed, based largely on those from past studies in SO (Ishola and McCalla, 2017b; Asaduzzaman et al., 2013; and Bhat et al., 2014).

8.1.1 Feature Extraction

8.1.1.1 Question Content-Based Approach

In the question content-based approach, to make the predictions aspects of the question and the number of questions a question asker has asked in the past were looked at. Five features are computed, drawn from features defined in (Ishola and McCalla, 2017b) as described below:

- *Title Length:* The title of a question provides a summarized description of the question. The title length is the number of characters in a question post title.
- *Body Length:* The body of a question contains the description of the question asked. The body length is the number of characters in the body of the question.
- *Code Present:* The code present feature is determined by whether or not the question contains code by searching for the *<code>* and *</code>* HTML tags in the body of the question. As code could be posted as an image, I also searched for the *<img* tag in the body of the question. The possible state of this feature is *yes* (1) or *no* (0).
- *Tag Count:* Tag(s) used in a question indicate the knowledge required to answer the question. A maximum of 5 tags can be used in a question. The tag count feature counts the number of tag(s) *k* included in the question *Q* asked as shown in Equation (8.1) below.

$$Tag\ Count = \sum_{i=1}^{k} tag_i^Q \qquad \text{Equation (8.1)}$$

- *Question Asked Count:* Question asked count is the total number of questions Q the question-asker u has asked in the past in SO as defined in Equation (8.2) below.

$$Question\ Asked\ Count = \sum_{uj=1}^{l} Q_j \qquad \text{Equation (8.2)}$$

where l is the total number of questions asked by user u.

8.1.1.2 Answerer-Based Approach

In the answerer-based approach, the features of potential answerers to questions were looked at. The tags used in a question serve as a guide to the knowledge required to answer the question. Therefore, for the answerer-based approach, only active answerers who have earned a tag-based badge for at least one of the tags in the question were considered. As in Section 7.2, active answerers who provided answers to questions containing any of the tags used in the question a month prior to when the question is asked were selected. The information about the answerers extracted is largely computed using the measures defined in Section 7.2.1.1.

- *Number of Active Answerers:* The number of active answerers was computed by counting the total number of distinct active answerers, U_{active}, who had provided answers to any of the tag(s) used in the question a month before the question was asked. This is computed as shown in Equation (8.3).

$$Number\ of\ Active\ Answerers = \sum_{i=1}^{k} U_{active}^{t \leq 30 days} \qquad \text{Equation (8.3)}$$

where U_{active} is the number of active question-answerers, users who have earned a tag-based badge for providing answer to questions about tag_i,
t is the time the question was asked as recorded in SO, and
k is the total number of tags used in the question.

- *Regularity:* Regularity represents how often the active answerers provide an answer to a topic, computed as in [Equation (7.11)].

$$Regularity = \sum_{ui=1}^{k} A_i \qquad \text{[Equation (7.11)]}$$

where A is the number of answers provided by user u to questions with a given tag, tag_i, and

k is the number of tags used in the question asked.

- *Knowledgeability:* Knowledgeability represents the know-how of the answerers about a given topic. Knowledgeability is the aggregate score earned by all active answerers as computed in [Equation (7.12)].

$$Knowledgeability = \sum_{ui=1}^{k} S_i \qquad \text{[Equation (7.12)]}$$

where S_i is the score obtained for providing answers to questions with a given tag, tag_i.

- *Eagerness:* Eagerness represents the keenness of the active answerers to deal with a topic when compared to other topics, computed as in [Equation (7.13)].

$$Eagerness = \frac{\sum_{ui=1}^{k} A_i}{\sum_u A} \qquad \text{[Equation (7.13)]}$$

where $\sum_u A$ represents the total number of answers provided by the user to all questions.

- *Willingness:* Willingness represents how vigorous the active answerers are in providing answers to questions on a topic as compared to other users. Willingness is computed as in [Equation (7.5)].

$$p(U_a|tag_i) = \frac{p(U_a)\, p(tag_i|U_a)}{p(tag_i)} \qquad \text{[Equation (7.5)]}$$

where p(U_a) is the probability of user u answering a question,

tag_i represents the set of tags used in the question,

$p(tag_i|U_a)$ is the likelihood that an answer to a question about tag_i will be given by user u, and

$p(tag_i)$ is the probability of a question related to tag_i will be asked in SO (this is the same for all prospective helpers).

$p(U_a)$ and $p(tag_i|U_a)$ can be computed as shown below:

$$p(U_a) = \frac{\sum_u A}{\sum_{j=1}^{n} A} \qquad \text{[Equation (7.6)]}$$

$$p(tag_i|U_a) = \frac{\sum_u A_i}{\sum_u A} \qquad \text{[Equation (7.7)]}$$

where $\sum_u A_i$ is the total number of tag_i answers by user *u*,

$\sum_u A$ is the total number of answers by user *u*, and

$\sum_{j=1}^{n} A$ is the total number of answers by all users *n*.

To compute the willingness score I substituted values from [Equation (7.6)] and [Equation (7.7)] into [Equation (7.5)]. Prospective helpers with higher willingness scores are ranked higher.

As the eagerness and willingness values computed above resulted in small numbers, I took the log of the respective values obtained. Also, as in the approach employed from Section 7.1.2.2, I multiplied the values computed for the last four features with the probability of an early answer. The probability of an early answer is computed as defined with [Equation (7.16)].

$$fastprob = \frac{\sum_{ui=1}^{k} A_i^{t \le 38}}{\sum_{ui=1}^{k} A_i} \qquad \text{[Equation (7.16)]}$$

where $\sum_{ui=1}^{k} A_i^{t \le 38}$ represents the number of questions that the prospective answerer *u* answered within 38 minutes for any of the tags *i* used in the question,

k is the number of tags used in the question, and

$\sum_{ui=1}^{k} A_i$ represents the total number of answers provided by user *u* to any of the tags used in the question.

8.1.1.3 Tag-Popularity Approach

In the tag-popularity approach, the features of the popularity of the tags used in the question were considered (Ishola and McCalla, 2007a; Asaduzzaman et al., 2013; Bhat et al., 2014). The three features extracted are described below:

- *Tag Popularity Count:* Tag popularity count is the count of the number of tags, used in the question q which are among the top 10% of all tags ($tag_{popular}$) used in SO.

$$Tag\ Popularity\ Count = \sum_{i=1}^{k} tag_{i=\ popular} \qquad \text{Equation (8.4)}$$

where $tag_{i=popular}$ represents the tag i used in question q which are among the top 10% mostly used tags in SO,

k is the total number of popular tags used in question q.

- *Minimum Tag Usage*: The minimum tag usage feature is the count of the number of times the least popular tag employed in the question has been used in SO. Minimum tag usage is computed as shown in Equation (8.5).

$$Minimum\ Tag\ Usage\ = \sum_{i=1}^{k} Q_{i=min} \qquad \text{Equation (8.5)}$$

where $Q_{i=min}$ represents the total number of questions asked in SO about the least popular tag i employed in the question, during the period the minimum tag usage measure is computed, and

k represents the total number of tags used in the question.

- *Average Tag Usage*: The average tag usage feature is the average number of times all the tags employed in a question have been used in SO. The average tag usage was computed as shown in Equation (8.6).

$$Average\ Tag\ Usage = \frac{\sum_{i=1}^{k} Q_i}{k} \qquad \text{Equation (8.6)}$$

where Q_i represents the total number of distinct questions asked in SO that contains any of the tag i used in the question, and

k is the total number of tags used in the question.

8.1.2 Predictive Modelling

The initial goals were to make two binary predictions. First, I wished to predict whether a question in SO would be answered or unanswered. Second, I wished to predict for an answered question whether it would be answered early or late. In carrying out these classification tasks, I employed some of the common classification techniques as described in Romero and Ventura (2010), which are:

- **Decision Tree**: Decision trees are a form of supervised learning which employs rules to generate a tree used in making predictions. The two decision tree models used in this experiment were *J48* and *Random Forest*, as defined in WEKA. The decision tree models were employed for this experiment because they are easy to understand, and they can be modelled with both numeric and categorical variables.
- **Bayesian Network**: A Bayesian network is a probabilistic graphical representation, which represents a set of variables and their conditional dependencies. The Bayesian network was chosen for this experiment as it works optimally with large data sets, as it is the case with the SO data. Since, the final prediction outputs from this experiment are categorical variables, the Bayes net approach was employed as defined in WEKA.
- **Logistic Regression**: Logistic Regression is used to find relationships between a dependent variable and one or more independent variables. The dependent variable is the variable to be predicted which takes on a binary value. Since the prediction tasks to be performed in this experiment are binary predictions, a logistic regression model was employed. The logistic regression as described in WEKA was employed for this experiment.

For the prediction tasks, the four predictive models employed were J48, Random Forest, Bayesian network and Logistic Regression. For each of the four predictive models, features from each the three approaches described in Section 8.1.1 were used as input into the model. For instance, using the tag-based approach, the tag popularity count, the minimum tag usage, and the average tag usage features would be used as input features into the respective predictive model. Using the definition for each of the predictive models in the WEKA tool kit, the input features from each approach are weighted equally and used to build the respective model for each predictive model. The model constructed by WEKA from each approach (e.g. tag-based approach) is then used to generate a binary output for the respective prediction task. For the prediction tasks, the output from the predictive model would depend on the classification task to be performed. For example, in predicting whether a question would be answered or not, the output from the respective predictive model employed would be *answered* or *unanswered.* Similarly, in predicting whether a question would be answered late or not, the output from the predictive model employed would be *early* or *late*.

The performance of each of the predictive models was evaluated using 10-fold cross-validation, again the computation is done by WEKA. WEKA fits the model to 90% of the data and the fitted model is used to predict the remaining 10% of the data (this process is repeated 10 times). The output from the WEKA tool kit is the overall precision, recall and prediction accuracy for each predictive task. The precision, recall, F-measure and the prediction accuracy of the prediction is defined as shown below.

$$Precision = \sum \left(\frac{f}{g}\right) \quad \text{Equation (8.7)}$$

$$Recall = \sum \left(\frac{f}{h}\right) \quad \text{Equation (8.8)}$$

where, *f* = total number of relevant items retrieved, i.e. the number of answer class correctly predicted,

g = total number of items retrieved. For example, in predicting whether a question would be *answered*, g is estimated as the number of answer posts correctly predicted as *answered* (true positive) plus the number of answer posts falsely predicted as *answered* (false positive).

h = total number of relevant items. For example, in predicting whether a question would be *answered*, *h* is estimated as the number answer posts correctly predicted as *answered* (true positive) plus the number of answer posts falsely predicted as *unanswered* (false negative).

$$F\text{-}measure = \sum \left(\frac{2 \times Precision \times Recall}{Precision + Recall}\right) \quad \text{Equation (8.9)}$$

$$PA = \frac{\sum_{i=1}^{s} C_p = C_a}{p} * 100\% \text{ Equation (8.10)}$$

where C_p is the predicted answer class,

C_a is the actual answer class of the answer post,

s is the total number of accurate predictions, and

p is the total number of prediction instances.

For my dataset, I used the 1,079,947 questions asked in SO from January to May 2017. Before making the first "answered or not" prediction. I noted that the dataset had 751,819 questions classified as answered and 328,128 classified as unanswered. These class sizes are skewed and would bias the results for the class with more data instances. To show the effects of the skewed class sizes I used precision, recall and F-measure to evaluate the model as computed in Equation (8.7) - (8.9) above. The results obtained in Table 8.1 – 8.3, shows the precision, recall and F-measure values obtained using logistic regression model.

Table 8.1. Logistic Regression using Raw Data with the Question Content-Based Approach

	Question Content-based		
Answer Class	**Precision**	**Recall**	**F-Measure**
Unanswered	0.38	0.00	0.01
Answered	0.70	1.00	0.82

Table 8.2. Logistic Regression using Raw Data with the Answerer-Based Approach

	Answerer Based Approach		
Answer Class	**Precision**	**Recall**	**F-Measure**
Unanswered	0.47	0.00	0.00
Answered	0.70	1.00	0.82

Table 8.3. Logistic Regression using Raw Data with the Tag-Based Approach

Answer Class	**Tag Based Approach**		
	Precision	**Recall**	**F-Measure**
Unanswered	0.00	0.00	0.00
Answered	0.70	1.00	0.82

The high recall values obtained in predicting the answered class indicate the model is highly biased towards the class with more data instances. Again, very low recall values were obtained in predicting the unanswered class. Chawla (2009) employed a random under-sampling technique to address the issue of a skewed class distribution. The random under-sampling was performed by randomly sampling the majority class samples, so that the majority class has the same number of instances as the minority class samples. Specifically, the random under-sampling was performed by randomly selecting 328,128 answers from the answered question class so that the two classes have the same amount of data. The results obtained with this under-sampling approach are shown in Table 8.4 - 8.6 using the question-based approach and logistic regression.

Table 8.4. After Under-Sampling with the Tag-Based Approach

Answer Class	**Question Content-based**		
	Precision	**Recall**	**F-Measure**
Unanswered	0.59	0.38	0.46
Answered	0.54	0.74	0.63

Table 8.5. After Under-Sampling with the Answerer-Based Approach

Answer Class	**Answerer Based Approach**		
	Precision	**Recall**	**F-Measure**
Unanswered	0.67	0.50	0.58
Answered	0.61	0.50	0.18

Table 8.6. After Under-Sampling with the Tag-Based Approach

Answer Class	Tag Based Approach		
	Precision	**Recall**	**F-Measure**
Unanswered	0.54	0.57	0.55
Answered	0.54	0.51	0.52

The results shown in Table 8.5 - 8.6 have higher precision and recall values compared to the results obtained in Table 8.1 - 8.3 in predicting the unanswered class. I have only shown the effects of skewed answer class sizes for logistic regression, but similar effects occur for all four predictive models. So, I used the under-sampling technique to eliminate such skewing throughout my experimentation. Further, as F-measure is computed using both precision and recall, so in evaluating the predictive models only the F-Measure and the prediction accuracy values obtained will be reported as defined in Equation (8.10).

8.2 Results

The results obtained for the four classification methods employed are shown in Table 8.7 - 8.8 below.

Table 8.7. Prediction Accuracy in Predicting Whether a Question Will be Answered or Not

Approach Used	Prediction Accuracy			
	J48	**Random Forest**	**Bayes Net**	**Logistic Regression**
Question Content-based	56.3%	56.4%	56.1%	55.8%
Answerer-based	59.4%	60.3%	58.9%	58.9%
Tag-Popularity	56.2%	56.7%	56.1%	53.8%

Table 8.8. F-Measure Values in Predicting Whether a Question Will be Answered or Not

Approach Used	F-Measure			
	J48	Random Forest	Bayes Net	Logistic Regression
Question Content-based	55.9%	56.1%	55.7%	54.2%
Answerer-based	59.2%	60.2%	58.8%	58.6%
Tag-Popularity	56.1%	56.7%	56.0%	53.7%

Comparing the results obtained with the three approaches the answerer-based approach has the highest prediction accuracy, although it is only marginally higher than the results achieved with the other approaches.

Next, I attempted to predict the "response time to first answer" (RTFA) class, to see if a question would be answered early or late. In defining the “RTFA classes” I used the median response time to the first answer of 38 minutes as the dividing line between the two classes. First, I eliminated all questions that never received an answer from the 1,079,947 questions considered in this study. Hence, I was left with class data sizes of 389,935 questions answered early (≤ 38 minutes) and 361,884 answered late (> 38 minutes), which creates relatively unbiased sample sizes, so I did not need to use under-sampling here. I present the prediction accuracy, and the F-measure obtained in predicting the RTFA class using the three approaches in Table 8.9 - 8.10 below.

Table 8.9. Prediction Accuracy in Predicting the RTFA Class

Approach Used	Prediction Accuracy			
	J48	Random Forest	Bayes Net	Logistic Regression
Question Content-based	58.2%	58.4%	58.0%	57.8%
Answerer-based	67.6%	68.4%	66.7%	65.1%
Tag-Popularity	63.9%	64.3%	63.6%	61.6%

Table 8.10. F-Measure Values in Predicting the RTFA Class

Approach Used	F-Measure			
	J48	Random Forest	Bayes Net	Logistic Regression
Question Content-based	57.9%	58.2%	57.7%	56.9%
Answerer-based	67.6%	68.4%	66.5%	65.1%
Tag-Popularity	63.8%	64.2%	63.6%	61.5%

Overall, as shown in Table 8.7 and Table 8.9, the question content-based approach had the least accuracy. The decision tree models (J48 and Random Forest) performed better than the other classifiers with Random Forest achieving the highest prediction accuracy and F-measure using the answerer-based approach. Overall, in predicting questions that will be unanswered or answered late, the answerer-based approach ranked higher.

8.3 Discussion

As described in Section 3.2.2 the opportunity to provide one-time support to users is especially important with the increasing growth in the proportion of unanswered questions or questions that are answered late. Hence, there is considerable importance in predicting whether a user's question might be unanswered or when it might be answered late. Knowing about such questions would be useful in advising a recommender system to intervene by explicitly recommending a peer helper. In this Chapter I explored three approaches to predicting unanswered or late answered questions: the question content-based approach, the answerer-based approach, and the tag-popularity approach.

The best prediction accuracy was achieved using the answerer-based approach, achieving about 60% and 68% accuracy in predicting the answer class and the RTFA class in Table 8.6 and Table 8.8. That the answerer-based approach works best is interesting and suggests that modelling users might be a key requirement when building tools to support the learners in an online learning community. The other approaches might also still be useful, even though they are less accurate than the answerer-based approach, perhaps allowing quicker predictions in some circumstances.

In future research I hope to improve upon the accuracy of these predictions, exploring whether other features might be useful. Based on the success of the answerer-based approach, the most promising place to look for such other features might be to look at other aspects of the answerer, i.e. to enhance the learner model that could be kept of learners in an OLC. I also would like to see if the features I have used can shed light on other facets of the learners. For example, can I discover their learning needs, especially their unperceived learning needs, using the answerer-based approach? Likely for these investigations, as with my future attempts to enhance the accuracy of my predictions of late or unanswered questions, new features will have to be explored. This experiment was reported in AIED 2018 (Ishola and McCalla 2018b).

CHAPTER 9

EXTENDING SUPPORT PROVIDED TO PROFESSIONAL LEARNERS

This thesis shows the possibility of supporting lifelong professional learners as they engage in their day to day learning activities as described in Chapters 4 to 8. With good support, the professionals themselves will benefit from having increased expertise. With increased expertise, professionals have a better chance of enhancing their their productivity at work, mitigating their susceptibility to layoffs, and increasing the prospect of getting better jobs. Another important aspect for professionals, especially licensed professionals such as doctors, lawyers, engineers, dentists, etc., is that they are legally required to keep their knowledge up to date. Further, the businesses that employ these professionals should benefit, reducing revenue loss due to incompetent or ignorant staff. This chapter describes the possible ways the results of this research could be applied to benefit the lifelong professional learner in meeting their learning needs and in improving upon their expertise levels.

9.1 Informing an Open Learner Model

One promise of my research is the ability of the methods I have developed to promote metacognitive activities by creating awareness in professionals about their learning needs. In achieving such awareness in professionals, my research tracked and maintained a record of what the professional currently knows and what they need to learn. This record is a called a "learner model", and could allow support systems to be appropriately reactive to the particular learning needs of the professional. The ability to inform an open learner model about the learning needs of individual professionals creates the opportunity for the professionals to take greater control and responsibility for their learning. In maintaining this model, the approaches discussed in Chapters 5 and 6 would evolve naturally with changes to each professional's learning interests and with changes in the disciplinary knowledge itself.

The opportunity to inform the learner creates the possibility of ensuring the learner model of each professional is updated as their learning needs evolve. If the learner model of professionals can be kept up to date, even the professional bodies could benefit by identifying common learning needs among the body of professionals. Of course, there are critical privacy issues in such third party use of open modelling (Anwar and Greer, 2008) which is one reason why the professional needs to have control over their learner model.

By opening up the learner model, users in an online learning community (OLC) would provide the opportunity to compare the learning needs of a particular professional with those of their peers. The ability to compare the learner models of peers would foster cooperative and competitive interactions amongst peers. Also, comparing the learner models of peers could create the opportunity to inform professionals about the learning trends within their discipline. Individual professionals could follow up such learning trends coupled with their learning needs to set their individual learning goals, create plans to achieve them, and reflect on their learning progress.

Many visualization issues must be addressed in the final design of an open learner model. There are also many HCI issues. The open learner model must determine user interaction protocols that will allow the user to have a richer ability to reflect on their learning needs. For example, by recommending posts that provide more context about the learning needs or by providing linkages to web content to help meet the learning needs.

9.2 Social Filtering of the Diagnosis

Having inferred the learning needs of a professional learner as described in Chapters 5 and 6, an opportunity exists to determine possible ways in meeting these needs using a social filtering approach (Klamma et al., 2007; Najjar, Meire and Duval, 2005). Through social filtering, learner *u* at time *t* could be compared to other learners with similar learner models at a given period *t1* in the past. These similar learners are useful in two ways. First, these similar learners could be a source of advice or help in meeting learner *u's* learning needs, assuming that these past similar learners have themselves subsequently resolved the learning needs now faced by u. But more interestingly with social filtering, it is possible to look at what happened to the similar learners to predict what will happen to learner *u* going forward. Such social filtering can allow inferences about what is important and not important to learn at time *t*. It can also allow inferences to be made on an appropriate order in which the learning needs could be met. Essentially through an instructional plan, what worked well for similar learners can be suggested to the learner (Frost and McCalla, 2015).

9.3 Informing Peer Recommender Systems

Peer help recommender systems are a common tool in OLCs to suggest prospective answerers to questions (Ruffo and Schifanella, 2009; Labarthe et al., 2016). In studies, the quality of the recommender system is often measured by its ability when a question is asked to predict answerers who will provide both high quality and on time answers. This thesis explored this issue in Chapter 7. Moreover, in a real-world system, such peer recommenders would be deemed to be especially useful if they can proactively determine users who would require intervention of the peer recommender system. If question-askers who require the intervention of the peer recommender system can be identified right after the question is asked, then it would be possible to help the learner soon after the question is asked. The ability to inform a recommender system of particularly serious learning needs is important in a large learning community with a lot of users to be helped, since recommending peers for all questions might be expensive in system resources. This thesis has identified some methodologies that would allow the prediction of users who would especially require the intervention of the peer recommender system. In instances where external information augments the online forum, the recommender system could even possibly provide learning resources to the user in meeting their learning needs.

9.4 Informing Educational Feedback Systems

Once the individual learning needs of professionals can be detected, adaptive feedback can be provided to the professionals about their learning needs. The opportunity to augment their learner model with other sources of information about the user and the user's activity within the OLC might make it possible to provide even more refined feedback to the professional. Other sources of information could include their resume or e-portfolio, their LinkedIn profile, the artifacts they produce (e.g. code), the tasks they have been assigned, job performance evaluations, etc. In my research augmenting the methodologies to include multiple sources of information about the professional has not been explored. This would be a promising future direction in improving the support provided to professionals. With the success of my research in detecting the learning needs of users, opportunities could exist to detect other aspects of the professional's knowledge such as forgetfulness.

9.5 Informing Professional Learning

As online learning communities gain more recognition among professionals, opportunities abound to extend professional learning beyond their workplace to open learning environments. Such a transition could lead to research in advanced learning technology shifting more from supporting traditional classroom learning to providing personalized learner-centered learning in distributed online environments. As more online learning environments emerge and gain more popularity, they have a huge potential to create avenues to gather more information about professional learners. Such an increasing awareness of the usefulness of online learning environments is evident by the millions of professionals already relying on OLCs such as Stack Overflow to meet their learning needs. As described in this chapter, supporting professionals in online learning environments holds promise to reshape professional learning. While learning environments that incorporate all the capabilities described in this chapter have not been implemented in this thesis research, they could be pursued as directions for future work.

CHAPTER 10
CONCLUSION

This chapter presents a summary of the thesis, the research contributions and directions for future work.

10.1 Summary

The main objective of this research has been to examine whether personalized support can be provided to professionals in detecting and meeting their learning needs within an online learning community. Supporting users within an OLC is important as it creates the opportunity to keep the professional up to date with current trends in the profession rather than tailoring the support only to their job roles. To keep professionals up to date with current trends in the profession, it is important to track the learning activities of the professional learner and the current trends within the profession as they evolve.

To achieve the goal of supporting professional learners within an OLC, I carried out seven experiments based on the Stack Overflow (SO) OLC. In the first experiment (as described in Section 4.1) the goal was to track the online learning activities of question-answerers to understand how the answer quality evolves based on how actively the answerers participate in help-giving. Results from this experiment (as shown in Figure 4.1) show a concerning trend of decline in answer quality of even the reputable users in SO. A decline in the quality of answers provided to questions could affect the help-seeking experience of question-askers. Hence, in experiment 2 (as described in Section 4.2) I saw the need to investigate the effects of help-seeking experiences of question-askers on their enthusiasm to express their learning needs. Results from experiment 2 reveal that negative help-seeking experiences could reduce the propensity of users to seek more help.

There is need to develop measures to help ensure users receive positive help seeking expreinces within the community. Afterward, I investigated the effects of the enthusiasm of question-askers on their evolving answer quality. As shown in Figure 4.2 a small rise was observed in the answer quality of even the enthusiastic question-askers. Results from Chapter 4 indicate the need to track the learning activities of users in detecting their unperceived learning needs and improve the help seeking experience of users.

Hence, the experiment in Chapter 5 tried to predict the answer quality of question-answerers in attempts to improve the quality of answers provided in SO. The tags of these questions were considered as indicators of the knowledge needed to answer the questions. So in experiment 3 (Section 5.1) a Naïve Bayes classifier was trained using the information about the tags contained in the question to predict the answer quality of a question-answerer to questions. Questions answered poorly were deemed suggestive of a user not having sufficient knowledge to answer such questions. Such questions were regarded as indicative of the unperceived learning needs of the user. Extracting information about tags proved effective in predicting the unperceived learning needs of question-answerers and under various conditions, accuracies of 85% and higher were achieved.

Chapter 6 (Section 6.1) explored how information about tags could be used in predicting the future unperceived learning needs of users. Unlike experiment 3, in experiment 4 the tags a question-asker would be asking questions about in the future were predicted. Again, extracting information about tags in assessing the learning needs of question-askers proved to be effective, as recall and precision values of 0.93 and 0.81 were achieved using a 5-month baseline. In sum, the results obtained in experiments 3 and 4 show the promise of detecting the learning needs in the knowledge of question-answerers and question-askers by extracting information about their tag usage.

Chapter 7 of this thesis takes on another serious issue in SO: the issue of increasing response times to questions and the rise in the proportion of unanswered questions in SO (as shown in Figure 3.10 and Figure 3.11). Therefore, there is a need to create measures that would allow the prediction of prospective answerers who will provide both timely and quality answers to questions in SO. In experiment 5 (Section 7.1), just as with the experiments in Chapter 5, I created measures that rely on information about question-answerers' past tag usage in predicting the prospective answerers to questions. Prediction accuracies of 63%, 55% and 57% were achieved in predicting (respectively) the first answerer, best answerer and the highest rated answerer to a question among the top 20 ranked users (S@20). Although, the results achieved in experiment 5 require further improvement if they are to be directly useful in a peer recommender system, the results obtained are a substantial improvement on the previous study by Tian et al. (Tian et al. 2013) who achieved a success rate of 13% at S@20 and a success rate of 23% at S@100.

The goal of experiment 6 (Section 7.2) was to improve on the prediction accuracy achieved in experiment 5. Using a hybrid-based approach which systematically combines the measures employed in experiment 5, prediction accuracies of nearly 90% were achieved. Furthermore, in experiment 6 workload balancing was also explored to ensure that answerers are well rested. Thus, helpers who recently provided help were exempted from the pool of prospective answerers. Despite the tradeoff of employing the work balancing approach which reduces the prediction accuracy, the results obtained in Figure 7.1 show that the prediction accuracy holds up fairly well if the exemption period isn't too long. By introducing the workload balancing approach the prospect of building a peer recommender system that could with reasonable accuracy identify well-rested helpers is enhanced.

In a real-world system, peer recommender systems would be deemed to be especially useful if a learner would otherwise receive no answer or a late answer. Therefore, in Chapter 8 attempts were made to predict the question-askers who will require the intervention of the peer recommender system to answer a question. . Specifically, in experiment 7 I attempted to predict the questions that would be answered late or remain unanswered, that is, the questions that would require the intervention of a peer help recommender system. As shown in Table 8.6 and Table 8.8, accuracies of 60% and 68% respectively show some promise in detecting questions that would require the early intervention of the peer recommender system. These results likely still require further improvement to inform a recommender system directly.

Despite the novelty of this research in seeking to provide personalized support to professionals in an OLC, there are obvious limitations to the approaches employed. First, all the experiments and results reported in this thesis were carried out using only the SO dataset. Therefore, some of the approaches and results reported might not be generalizable to other

OLCs. However, the SO dataset with over 27 million posts of users is large enough for my experiments to provide real insight if not yet definitive conclusions. In my most common approach users in SO were tracked and used to make predictions which were thereafter validated with the actual data of the user. Again, all the experiments performed were carried out offline, without any interaction(s) with the actual SO users, which means there is no way to independently confirm some of the predictions with information gathered outside of SO.

10.2 Contributions

This thesis research contributes to AIED areas such as personalization, educational diagnosis, educational recommender systems, educational data mining, and intelligent help, all in the context of lifelong learning. These general contributions have been made:

- First, this thesis research takes a novel approach in supporting the learning needs of professionals by mining the interaction of peers within an OLC. Using OLC data allows for the possibility of detecting the needs of the professionals beyond their job needs as they interact with peers. Further, this research has adopted a preventive approach by attempting to inform learners about their learning needs before such needs are obvious to them. These are all contributions providing insight into how to support professionals over the long-term with more advanced learning technology.
- Also, this thesis contributes to research in educational diagnosis and educational data mining by exploring lightweight approaches which can be easily adapted to OLCs rather than having to build an ontology which will continually have to be updated as new knowledge emerges.
- Further, this thesis contributes to intelligent help and recommender systems by developing measures that could be used in predicting prospective helpers who can help provide both quality and on time answers. Achieving a success rate of about 90%, the approaches employed in this study greatly exceed the success rate of about 23% achieved by Tian et al. while predicting prospective helpers. Further, my research advances previous research in intelligent help by exploring methods that could exempt the users who have recently provided help. This is especially helpful in

ensuring that helpers are well rested and to give opportunities to less active helpers to participate.

- Finally, this thesis contributes to learner personalization by employing approaches that would allow the tracking of changes in the learning needs of users naturally as they evolve. With the ability to track the learning needs of users as they evolve, the possibility exists to create tools to provide support adapted to their current learning needs. Even more promising, as the knowledge required for professionals to succeed in the profession itself evolves, the approaches employed in this thesis show promise to still be effective.

10.3 Future Work

This thesis has shown the possibility of extending support provided to users beyond their workplace to online learning environments. In future experiments, my goal would be to improve the prediction accuracy achieved in experiment 7. The results shown in Tables 8.6 - 8.8 achieve the highest levels of accuracy by extracting information about the answerers. In future experiments, I would like to explore other attributes about the question-answerers that could be useful in improving upon the prediction accuracy. Especially, I would like to gather information about the answerers beyond the OLC that might provide the opportunity for finer grained and more accurate diagnosis.

Although the approaches I have explored in this thesis research do not require an ontology, their reliance on tags alone might not always be optimal. Capturing the mappings of tags to actual knowledge elements in the profession using an ontology could be useful to determine an appropriate learning plan for each professional. Moreover, as technology and other aspects of the professional body of knowledge advance, the need to ensure that the profession maintains an up-to-date body of knowledge is important. To help the profession automate these tasks would at least require an ontology that their skills could be computed over and a diagnostic engine that could determine what a given professional knows and doesn't know. Fortunately, in the future professional bodies could take on the role of developing such an ontology and constantly updateing it as the body of knowledge changes.

Also, considering that software developers often work collaboratively as a team, in future research I would like to determine approaches that would help achieve the optimal success in such group learning. With the recent announcement of "*Stack Overflow for Teams*[10]" which would allow a group of developers to collaborate in a private space, the need to support group learning within the SO OLC seems to be emerging. As users in SO interact within closed groups, the need to develop tools that can perform discourse analysis to know what the users are talking about, the need to detect the group learning needs, and the need to support such learning needs would become necessary. In future experiments I could investigate how the measures explored in this thesis research could be extended to support group learning among the users within the OLC. Also, in future experiments, I would like to evaluate the benefits of group learning within the OLC. I would like to compare how users who collaborated within a private space compare to users who collaborated openly within the community.

To exploit the methodologies explored in this thesis, it would be necessary to incorporate technology that integrates these methodologies into the online learning environment. Some of the enhancements discussed in Chapter 9 could be incorporated into an OLC to help achieve this. With or without such enhancements the user will still have to interact with the OLC, and this would require an interface to be designed. Of course, any form of feedback mechanism that needs to be created would come with HCI issues. Some of the HCI issues are how to display any detected learning needs to the learner and how to explain the nature of the learning needs to the learner. Other HCI issues emerge when exploring what kinds of interactions and control the learner may have in updating the learning needs. These issues are the subject of future research and would need to be addressed to realize the desired end goals of this research to provide technology that resolves the learning needs and improves the expertise of professional learners.

[10] https://stackoverflow.com/teams

MY PEER-REVIEWED PUBLICATIONS WITH CONTENTS FROM THIS DISSERTATION

Eight papers have been written drawing from the experiments carried out in my doctoral research. All of the experiments described in each paper were carried out by me (Oluwabukola Ishola), including choosing and refining the research questions, designing the experimental methodology, carrying out the actual experiments, and analyzing the results. My supervisor (and co-author) provided feedback throughout the experimental process and provided strong editing support during the writing of each paper. The following are the peer-reviewed publications with contents from Chapter 4 – Chapter 8 of this thesis:

Idowu (Ishola), O. M. and McCalla, G. (2018a). Better late than never but never late is better: towards reducing the answer response time to questions in an online learning community. *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 184-197). London, United Kingdom. Springer, Cham.

Idowu (Ishola), O. M. and McCalla, G. (2018, June). On the value of answerers in early detection of response time to questions for peer recommender systems. *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 160-165). Springer, Cham.

Idowu (Ishola), O.M. and McCalla (2018c). Toward the enhancement of peer-peer mentoring systems in supporting lifelong professional learners. *Workshop on Intelligent Mentoring Systems*, London, United Kingdom, 10 pages.

Ishola (Idowu), O. M. and McCalla, G. (2017a). Personalized tag-based knowledge diagnosis to predict the quality of answers in a community of learners. *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 113–124). Wuhan, China. https://doi.org/10.1007/978-3-319-61425-0_10.

Ishola (Idowu), O. M. and McCalla, G.I. (2017b). Predicting prospective peer helpers to provide just-in-time help to users in question and answer forums. *Proceedings of the*

10th International Conference on Educational Data Mining (EDM), (pp. 238–243). Wuhan, China.

Ishola (Idowu), O. M. and McCalla, G. (2016a). Detecting and supporting the evolving knowledge interests of lifelong professionals. In *Proceedings of the European Conference on Technology Enhanced Learning* (pp. 595-599). Springer International Publishing.

Ishola (Idowu), O. M. and McCalla, G. (2016b). Diagnosis at scale: detecting the expertise level and knowledge states of lifelong professional learners. *24th User Modeling Adaptation, and Personalization (UMAP) Extended Proceedings.* Halifax, Canada, 4 pages.

Ishola (Idowu), O. M. and McCalla, G. (2016b). Detecting and supporting the evolving knowledge interests of lifelong professionals. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 595–599).

Ishola (Idowu), O. M. and McCalla, G. (2016c). Tracking and reacting to the evolving knowledge needs of lifelong professional learners. In *CEUR Proceedings of the Personalization Approaches in Learning Environments (PALE) Workshop*, (pp. 68-73). Halifax, Canada.

REFERENCES

Asaduzzaman, M., Mashiyat, A. S., Roy, C. K., and Schneider, K. A. (2013). Answering questions about unanswered questions of Stack Overflow. In *2013 10th Working Conference on Mining Software Repositories (MSR)*, (pp. 97–100). San Francisco, CA, USA. http://doi.org/10.1109/MSR.2013.6624015.

Adomavicius, G. and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, (Vol. 17(6), pp. 734-749). http://doi.org/10.1109/TKDE.2005.99.

Al-Aidaroos, K. M., Abu Bakar, A., and Othman, Z. (2010). Naïve Bayes variants in classification learning. In *Proceedings - 2010 International Conference on Information Retrieval and Knowledge Management: Exploring the Invisible World, CAMP'10* (pp. 276-281). http://doi.org/10.1109/INFRKM.2010.5466902.

Athey, T. R. and Orth, M. S. (1999). Emerging competency methods for the future. *Human Resource Management*, (Vol. 38(3), pp. 215-225). http://doi.org/10.1002/(SICI)1099050X(199923)38:3<215::AID-HRM4>3.0.CO;2-W.

Attfield, S., Kazai, G., Lalmas, M., and Piwowarski, B. (2011). Towards a science of user engagement (position paper). *Web Search and Data Mining (WSDM) Workshop on User Modelling for Web Applications* (pp. 9-12). Hong Kong, China.

Aleven, V., McLaren, B., Roll, I., and Koedinger, K. (2006). Toward meta-cognitive tutoring: a model of help seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education*, (Vol. *16*(2), pp. 101–128). http://doi.org/10.1.1.121.9138.

Anwar, M. and Greer, J. (2008). Role- and relationship-based identity management for private yet accountable e-learning. In J. C. D. Karabulut Y., Mitchell J., Herrmann P. (Ed.), *IFIP International Federation for Information Processing* (Vol. 263, pp. 343–358). Springer, Boston, MA. http://doi.org/10.1007/978-0-387-09428-1_22.

Arroyo, I. and Woolf, B. P. (2005). Inferring learning and attitudes from a bayesian network of log file data. *Proceedings of the 12th International Conference on Artificial Intelligence in Education* (pp. 33-40).

Barrett, G. V. and Depinet, R. L. (1991). A reconsideration of testing for competence rather than for intelligence. *American Psychologist*, (Vol. 46(10), pp. 1012–1024). http://doi.org/10.1037/0003-066X.46.10.1012.

Bartkowiak, G. (2014). The competence of professional practitioners as self-assessed and employer‑ assessed: Within and beyond boundaries of management. *In Warsaw School of Economics Press, Warsaw* (pp. 275-288). ISBN 978-83-7378-903-6.

Barua, A., Thomas, S. W., and Hassan, A. E. (2014). What are developers talking about? an analysis of topics and trends in Stack Overflow. *Empirical Software Engineering*, (Vol. 19(3), pp. 619–654). http://doi.org/10.1007/s10664-012-9231-y.

Bazelli, B., Hindle, A., and Stroulia, E. (2013). On the personality traits of StackOverflow users. In *IEEE International Conference on Software Maintenance, ICSM* (pp. 460–463). Washington, DC, USA. http://doi.org/10.1109/ICSM.2013.72.

Bhat, V., Gokhale, A., Jadhav, R., Pudipeddi, J., and Akoglu, L. (2014). Min(e)d your tags: analysis of question response time in StackOverflow . In *ASONAM 2014 - Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 328–335). http://doi.org/10.1109/ASONAM.2014.6921605.

Boud, D. and Hager, P. (2012). Re-thinking continuing professional development through changing metaphors and location in professional practices. *Studies in Continuing Education*, (Vol. *34*(1), pp. 17–30). http://doi.org/10.1080/0158037X.2011.608656.

Bruce, C. S. (1999). Workplace experiences of information literacy. *International Journal of Information Management*, (Vol. 19(1), pp. 33–47). http://doi.org/10.1016/S0268-4012(98)00045-0.

Buntat, Y., Puteh, N. A., Azeman, S. H., Nasir, A. N. M., Iahad, N., and Aziz, M. A. (2013). The need of lifelong learning towards learning community development in Malaysia. *Procedia - Social and Behavioural Sciences*, (Vol. 93, pp. 1541–1545). http://doi.org/10.1016/j.sbspro.2013.10.079.

Bull, S., Greer, J., Mccalla, G., Kettel, L., and Bowes, J. (2001). User modelling in i-help: what, why, when and how. In *User Modelling: Proceedings of The Eighth International Conference*, (pp. 117–126). Sonthofen, Germany: Springer, Berlin, Heidelberg. http://doi.org/10.1007/3-540-44566-8_12.

Carre, P. (2000). Motivation in adult education: from engagement to performance. *41st Adult Education Research Conference*, (pp. 66–67). Vancouver, Canada. Retrieved from http://www.adulterc.org/Proceedings/2000/carrep1-final.PDF.

Cha, S. H. (2007). Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, (Vol. 4(1), pp. 300-307).

Chawla, N. V. (2009). Data mining for imbalanced datasets: an overview. In *Data Mining and Knowledge Discovery Handbook* (pp. 875–886). Springer, Boston, MA. http://doi.org/10.1007/978-0-387-09823-4_45.

Conati, C. (2010). Bayesian student modelling. In Nkambou, R., Mizoguchi, R., Bourdeau, and J. (Ed.). In *Advances in Intelligent Tutoring Systems*, (Vol. 308(1), pp. 281-299). ISBN 978-3-642-14363-2. Springer, Berlin, Heidelberg.

Das, P. R. and Chakrabarti, T. (2016). Application of bayesian credibility theory in movie rankings to reduce financial risk of production houses. *KIIT Journal of Management*, (Vol. 12(2), pp. 95 – 106).

De Laat, M. and Schreurs, B. (2013). Visualizing informal professional development networks: building a case for learning analytics in the workplace. In Caroline H., Maartende L., and Shane D. (Ed.), *American Behavioural Scientist*, (Vol. 57(10), pp. 1421–1438). http://doi.org/10.1177/0002764213479364.

Desmarais, M. C. and Baker, R. S. J. D. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction*, (Vol. 22(1–2), pp. 9–38). https://doi.org/10.1007/s11257-011-9106-8

Dolog, P., Kay, J., and Kummerfeld, B. (2009). Personal lifelong user model clouds. In *Proceeding of the Lifelong User Modelling Workshop at UMAP*, *9*, (pp. 1–8). Trento, Italy.

Dunning, D. (2011). The Dunning-Kruger effect. On being ignorant of one's own ignorance. *Advances in Experimental Social Psychology*, (Vol. 44, pp. 247-296). http://doi.org/10.1016/B978-0-12-385522-0.00005-6.

Falmagne, J. C., Doignon, J. P., Koppen, M., Villano, M., and Johannesen, L. (1990). Introduction to knowledge spaces: how to build, test, and search them. *Psychological Review*, (Vol. 97(2), pp. 201–224). http://doi.org/10.1037/0033-295X.97.2.201.

Fischer, G. (2000). Lifelong learning more than training. *Journal of Interactive Learning Research*, (Vol. 11(3/4), pp. 265-294). Association for the Advancement of Computing in Education (AACE).

Frost, S. and McCalla, G. (2015). An approach to developing instructional planners for dynamic open-ended learning environments. In *CEUR Workshop Proceedings* (Vol. 1432, pp. 1–10). Madrid, Spain.

Garton, L., Haythornthwaite, C., and Wellman, B. (2006). Studying online social networks. *Journal of Computer-Mediated Communication*, (Vol. 3(1), pp. 0–0). http://doi.org/10.1111/j.1083-6101.1997.tb00062.x.

Gibson, D., Ostashewski, N., Flintoff, K., Grant, S., and Knight, E. (2015). Digital badges in education. *Education and Information Technologies*, (Vol. 20(2), pp. 403–410). http://doi.org/10.1007/s10639-013-9291-7.

Grant, J. (2002). Learning needs assessment: assessing the need. In Goldbeck-Wood, S., and Peile, E. (Ed.), *British Medical Journal (BMJ)*, (Vol. *324*(7330), pp. 156–159). http://doi.org/10.1136/bmj.324.7330.156.

Grant, S. and Betts, B. (2013). Encouraging user behaviour with achievements: an empirical study. In *2013 10th Working Conference on Mining Software Repositories (MSR)*, (pp. 65-68). San Francisco, CA, USA.

Green, P. C. (1999). Building robust competencies: linking human resource systems to organizational strategies. *Personnel Psychology*, (pp. 199-204). Retrieved from http://proquest.umi.com/pqdweb?did=748305andFmt=7andclientId=25836andRQT=309andVName=PQD.

Greer, J., McCalla, G., Collins, J., Kumar, V., Meagher, P., and Vassileva, J. (1998a). Supporting peer help and collaboration in distributed workplace environments. *International Journal of Artificial Intelligence in Education*, (Vol. 9(2), pp. 159-177). Retrieved from https://hal.archives-ouvertes.fr/hal-00588744/document.

Greer, J., McCalla, G., Cooke, J., Collins, J., Kumar, V., Bishop, A., and Vassileva, J. (1998b). The intelligent helpdesk: supporting peer-help in a university course. In Goettl, B.P., Halff, H.M., Redfield, C.L., Shute, V.J. (Ed), Proceedings *of the 4th International Conference on Intelligent Tutoring Systems (ITS)*, (pp. 494–502). San Antonio, Texas, USA. http://doi.org/10.1007/3-540-68716-5_55.

Hamalainen, W., Suhonen, J., Sutinen, E., and Toivonen, H. (2004). Data mining in personalizing distance education courses. *21st International Council for Open and Distance Education (ICDE) World Conference on Open Learning and Distance Education*, (pp. 18-21). Hong Kong, China.

Hatta, K. A. B. K., Wee, K., Cheah, W. P., and Wee, Y. (2015). A true bayesian estimate concept in LTE downlink scheduling algorithm. *In Telecommunication Networks and Applications Conference (ITNAC), 2015 International* (pp. 71-76). IEEE.

Haythornthwaite, C. and De Laat, M. (2010). Social networks and learning networks: Using social network perspectives to understand social learning. In Dirckinck-Holmfeld, L., Hodgson, V., Jones, C., De Laat, M., McConnell, D., and Ryberg, T. (Ed.), *7th International Conference on Networked Learning,* (pp. 183–190). Aalborg, Denmark.

Heymann, P., Ramage, D., and Garcia-Molina, H. (2008). Social tag prediction. *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval SIGIR 08,* (Vol. 31(4), pp. 531-538). Singapore. http://doi.org/10.1145/1390334.1390425.

Indratmo, J. and Vassileva, J. (2008). A review of organizational structures of personal information management. *JODI: Journal of Digital Information*, (Vol. 9(1), pp. 1-19).

Ishola (Idowu), O. M. and McCalla, G. (2016a). Diagnosis at scale: detecting the expertise level and knowledge states of lifelong professional learners. *In 24th User Modeling Adaptation, and Personalization (UMAP) Extended Proceedings.* Halifax, Canada.

Ishola (Idowu), O. M. and McCalla, G. (2016b). Detecting and supporting the evolving knowledge interests of lifelong professionals. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 595–599). https://doi.org/10.1007/978-3-319-45153-4_71.

Ishola (Idowu), O. M. and McCalla, G. (2016c). Tracking and reacting to the evolving knowledge needs of lifelong professional learners. In *CEUR Workshop Proceedings* of *Personalization Approaches in Learning Environments (PALE),* (pp. 68-73). Halifax, Canada.

Ishola (Idowu), O. M. and McCalla, G. (2017a). Personalized tag-based knowledge diagnosis to predict the quality of answers in a community of learners. *In International Conference on Artificial Intelligence in Education* (pp. 113–124). Wuhan, China.

Ishola (Idowu), O. M. and McCalla, G.I. (2017b). Predicting prospective peer helpers to provide just-in-time help to users in question and answer forums. *Proceedings of the 10th International Conference on Educational Data Mining (EDM)*, (pp. 238 – 243). Wuhan, China.

Idowu (Ishola), O. M. and McCalla, G. (2018a). Better late than never but never late is better: towards reducing the answer response time to questions in an online learning community. *In International Conference on Artificial Intelligence in Education* (pp. 184-197). London, United Kingdom. Springer, Cham.

Idowu (Ishola), O. M. and McCalla, G. (2018, June). On the value of answerers in early detection of response time to questions for peer recommender systems. *In International Conference on Artificial Intelligence in Education* (pp. 160-165). Springer, Cham.

Idowu (Ishola), O. M. and McCalla (2018c). Toward the enhancement of peer-peer mentoring systems in supporting lifelong professional learners. *Workshop on Intelligent Mentoring Systems*, London, United Kingdom.

Jones, C. R., Ferreday, D., and Hodgson, V. (2008). Networked learning a relational approach: Weak and strong ties. *Journal of Computer Assisted Learning*, (Vol. 24(2), pp. 90–102). https://doi.org/10.1111/j.1365-2729.2007.00271.x

Kay, J. (2008). Lifelong learner modeling for lifelong personalized pervasive learning. In Durlach, P. J., and Lesgold, A. (Ed). *IEEE Transactions on Learning Technologies*, (Vol. 1(4), pp. 215–228). Cambridge University Press. https://doi.org/10.1109/TLT.2009.9.

Kay, J. and Kummerfeld, B. (2009). Lifelong user modelling goals, issues and challenges. In *Proceedings of the Lifelong User Modelling Workshop, at UMAP'09 User Modelling Adaptation, and Personalization* (pp. 27–34). https://doi.org/10.1.1.149.8233.

Klamma, R., Chatti, M. A., Duval, E., Hummel, H., Hvannberg, E. T., Kravcik, M., Scott, P. (2007). Social software for life-long learning. *Educational Technology and Society*, (Vol. 10(3), pp. 72–83).

Koper, R., Giesbers, B., Van Rosmalen, P., Sloep, P., Van Bruggen, J., Tattersall, C., Brouns, F. (2005). A design model for lifelong learning networks. *Interactive Learning Environments*, (Vol. 13(1–2), pp. 71–92). https://doi.org/10.1080/10494820500173656.

Korossy, K. (1997). Extending the theory of knowledge spaces: a competence-performance approach. Zeitschrift fur Psychologie, (Vol. 205(1), pp. 53-82).

Korossy, K. (1999). Modeling knowledge as competence and performance. In Albert, D., and Lukas, J., (Eds.) *Knowledge Spaces: Theories, Empirical Research, and Applications*, (pp. 103-132). Mahwah, New Jersey.

Kravcik, M., Kaibel, A., Specht, M., and Terrenghi, L. (2004). Mobile collector for field trips. *Educational Technology and Society*, (Vol. 7(2), pp. 25–33). https://doi.org/10.1080/01449290512331319049.

Kravcik, M., Specht, M., and Oppermann, R. (2004). Evaluation of WINDS authoring environment. *Adaptive Hypermedia and Adaptive Web-Based Systems*, (pp. 166–175). Eindhoven, Netherlands. Springer Berlin Heidelberg. Retrieved from http://www.springerlink.com/index/5151H7EVRREABEAJ.pdf,

Kunaver, M., Pozrl, T., Pogacnik, M., and Tasic, J. (2007). Optimisation of combined collaborative recommender systems. *AEU - International Journal of Electronics and Communications*, (Vol. 61(7), 433–443). https://doi.org/10.1016/j.aeue.2007.04.003.

Kwak, H., Lee, C., Park, H., and Moon, S. (2010). What is twitter, a social network or a news media? *Proceedings of the 19th International Conference on World Wide Web - WWW '10*, (pp. 591-600). Raleigh, NC, USA. https://doi.org/10.1145/1772690.1772751.

Laal, M. and Salamati, P. (2012). Lifelong learning; why do we need it? In *Procedia - Social and Behavioral Sciences* (Vol. 31, pp. 399–403). https://doi.org/10.1016/j.sbspro.2011.12.073.

Labarthe, H., Yacef, K., Bouchet, F., Bachelet, R., Gallagher, M. S., Hogue, R. J., and Aldana, S. I. (2015). Does a peer recommender foster students' engagement in MOOCs? *Computers and Education* (pp. 418-423). https://doi.org/10.1016/j.compedu.2014.08.005.

Ladyshewsky, R. K. (2006). Building cooperation in peer coaching relationships: understanding the relationships between reward structure, learner preparedness, coaching skill and learner engagement. *Physiotherapy*, (Vol. 92(1), pp. 4–10). https://doi.org/10.1016/j.physio.2005.11.005.

Lee, J. I. and Brunskill, E. (2012). The impact on individualizing student models on necessary practice opportunities. In Yacef, K., Zaïane, O., Hershkovitz, A., Yudelson, M., and Stamper, J. (Eds.) *Proceedings of the 5th International Conference on Educational Data Mining*, (pp. 118-125).

Ley, T., Ulbrich, A., Scheir, P., Lindstaedt, S. N., Kump, B., and Albert, D. (2008). Modeling competencies for supporting work-integrated learning in knowledge work. *Journal of Knowledge Management*, (Vol. 12(6), pp. 31–47). https://doi.org/10.1108/13673270810913603

Ley, T. and Kump, B. (2013). Which user interactions predict levels of expertise in work-integrated learning? *In Proceedings of 8th European Conference on Technology Enhanced Learning (EC-TEL*) (pp. 178–190). Paphos, Cyprus. https://doi.org/10.1007/978-3-642-40814-4_15.

Ley, T., Cook, J., Dennerlein, S., Kravcik, M., Kunzmann, C., Pata, K., and Trattner, C. (2014). Scaling informal learning at the workplace: a model and four designs from a large-scale design-based research effort. *British Journal of Educational Technology*, (Vol. 45(6), pp. 1036–1048). https://doi.org/10.1111/bjet.12197.

Ley, T., Klamma, R., Lindstaedt, S., and Wild, F. (2016). Learning analytics for workplace and professional learning. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge - LAK '16* (pp. 484–485). https://doi.org/10.1145/2883851.2883860.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, (Vol 5(1), pp. 1–167). https://doi.org/10.2200/S00416ED1V01Y201204HLT016.

Lohman, M. C. (2006). Factors influencing teachers' engagement in informal learning activities. *Journal of Workplace Learning*, (Vol. 18(3), pp. 141–156). https://doi.org/10.1108/13665620610654577.

Manninen, J. and Hobrough, J. (2000). Skills gaps and overflows?: a European perspective of graduate skills and employment in SMEs. *Industry and Higher Education*, (Vol. 14(1), pp. 51–57). https://doi.org/10.5367/000000000101294869.

Merrill, D. C., Reiser, B. J., Trafton, J. G., and Ranney, M. (1992). Effective tutoring techniques: a comparison of human tutors and intelligent tutoring systems. *Journal of the Learning Sciences*, (Vol. 2(3), pp. 277–305). https://doi.org/10.1207/s15327809jls0203_2.

Mitchell, T. M. (2005). Generative and discriminative classifiers: naive bayes and logistic regression. In *Machine Learning*. https://doi.org/10.1093/bioinformatics/btq112.

Murphy, K. P. (2012). Machine learning: a probabilistic perspective. *MIT Press*. https://doi.org/10.1007/978-3-642-21004-4_10.

Najjar, J., Meire, M., and Duval, E. (2005). Attention metadata management: tracking the use of learning objects through attention. In Kommers, P. and Richards, G. (Ed.) *Proceedings of World Conference on Educational Media and Technology*, (pp. 1157-1161). Montreal, Canada.

Nasehi, S. M., Sillito, J., Maurer, F., and Burns, C. (2012). What makes a good code example?: a study of programming QandA in StackOverflow. In *IEEE International Conference on Software Maintenance, ICSM* (pp. 25–34). https://doi.org/10.1109/ICSM.2012.6405249. Trento, Italy.

Pardos, Z. and Heffernan, N. (2010). Modelling individualization in a Bayesian networks implementation of knowledge tracing. In De Bra, P., Kobsa, A., Chin, D. (Eds.) *Proceedings of 18th International conference on User Modeling, Adaptation, and Personalization (UMAP)*, (pp. 255-266). Hawaii, USA. Springer, Heidelberg.

Ponzanelli, L., Bavota, G., Di Penta, M., Oliveto, R., and Lanza, M. (2014). Mining StackOverflow to turn the IDE into a self-confident programming prompter. In *Proceedings of the 11th Working Conference on Mining Software Repositories - MSR 2014* (pp. 102–111). https://doi.org/10.1145/2597073.2597077

Rashid, A. M., Albert, I., Cosley, D., Lam, S. K., McNee, S. M., Konstan, J. A., and Riedl, J. (2002). Getting to know you. *Proceedings of the 7th International Conference on Intelligent User Interfaces - IUI '02*, (pp. 127-134). San Francisco, CA, USA. https://doi.org/10.1145/502716.502737.

Robinson, B. (1998). A strategic perspective on staff development for open and distance learning. Latchem, C., and Lockwood, F., (Ed.) *Staff Development in Open and Flexible Learning*, (pp. 33-44). London, United Kingdom.

Romero, C. and Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*. https://doi.org/10.1109/TSMCC.2010.2053532

Ruffo, G. and Schifanella, R. (2009). A peer-to-peer recommender system based on spontaneous affinities. *ACM Transactions on Internet Technology*, (Vol 9(1), pp. 1–34). https://doi.org/10.1145/1462159.1462163.

Sharples, M. (2000). The design of personal mobile technologies for lifelong learning. *Computers and Education*, (Vol. 34(3), pp. 177-193). https://doi.org/10.1016/S0360-1315(99)00044-5.

Simons, P. R. J., and Ruijters, M. C. (2004). Professional learning: gaps and transitions on the way from novice to expert. In Henny P.A. Boshuizen, H., Bromme, R., and Gruber, H. (Eds.) *Innovation and Change in Professional Education*, (Vol. 2, pp. 207-229).

Siemens, G. (2010). What are learning analytics? Retrieved from http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/ , on May 12, 2018.

Song, Y., Zhang, L., and Giles, C. L. (2011). Automatic tag recommendation algorithms for social recommender systems. *ACM Transactions on the Web*, (Vol. 5(1), pp. 1–31). https://doi.org/10.1145/1921591.1921595.

Song, Y., Loewenstein, G., and Shi, Y. (2018). Heterogeneous effects of peer tutoring: evidence from rural Chinese middle schools. *Research in Economics*, (Vol. 72(1), pp. 33–48). https://doi.org/10.1016/j.rie.2017.05.002.

Stanley, C. and Byrne, M. (2011). Predicting tags for StackOverflow posts. *Proceedings of 12th International Conference on Cognitive Modelling (ICCM)*, (pp. 414–419). Ottawa, Canada. https://doi.org/10.1.1.399.7835.

Stefaner, M. and M̈uller, B. (2007). Elastic lists for facet browsers. In *Proceedings - International Workshop on Database and Expert Systems Applications, DEXA* (pp. 217–221). https://doi.org/10.1109/DEXA.2007.44.

Tang, L. M. and Kay, J. (2013). Lifelong user modeling and meta-cognitive scaffolding: support self-monitoring of long term goals. In *CEUR Workshop Proceedings* (pp. 11 - 19). Rome, Italy.

Tian, Y., Kochhar, P. S., Lim, E. P., Zhu, F., and Lo, D. (2013). Predicting best answerers for new questions: An approach leveraging topic modeling and collaborative voting. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 55–68). https://doi.org/10.1007/978-3-642-55285-4_5.

Tsuei, M. (2012). Using synchronous peer tutoring system to promote elementary students' learning in mathematics. *Computers and Education*, (Vol. 58(4), pp. 1171–1182). https://doi.org/10.1016/j.compedu.2011.11.025.

Treude, C., Barzilay, O., and Storey, M.A. (2011). How do programmers ask and answer questions on the web? In *Proceeding of the 33rd international conference on Software engineering - ICSE '11* (pp. 804-807). Honolulu, HI, USA. https://doi.org/10.1145/1985793.1985907.

Vasilescu, B., Filkov, V., and Serebrenik, A. (2013). StackOverflow and GitHub: associations between software development and crowdsourced knowledge. In *Proceedings - SocialCom/PASSAT/BigData/EconCom/BioMedCom 2013* (pp. 188–195). https://doi.org/10.1109/SocialCom.2013.35.

Vasilescu, B., Capiluppi, A., and Serebrenik, A. (2014). Gender, representation and online participation: a quantitative study. *Interacting with Computers*, (Vol *26*(5), pp. 488–511). https://doi.org/10.1093/iwc/iwt047.

Vassileva, J., Greer, J., McCalla, G., Deters, R., Zapata, D., Mudgal, C., and Grant, S. (1999). A multi-agent approach to the design of peer-help environments. *Proceedings of the International Conference on Artificial Intelligence in Education (AIED 1999),* (pp. 38-45). Le Mans, France.

Wang, S., Lo, D., and Jiang, L. (2013). An empirical study on developer interactions in Stack Overflow. *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, (pp. 1019-1024).

Yudelson, M. V., Koedinger, K. R., and Gordon, G. J. (2013). Individualized bayesian knowledge tracing models. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 171–180). https://doi.org/10.1007/978-3-642-39112-5-18.

Zapata-Rivera, J. D. and Greer, J. E. (2004). Interacting with inspectable Bayesian student models. *International Journal of Artificial Intelligence in Education*, (Vol. 14, pp. 127–163). Retrieved from http://portal.acm.org/citation.cfm?id=1434859.