Recommender systems to support socio-collaborative learning in educational discussion forums

by

Zhaorui Chen

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Computing Science

University of Alberta

Abstract

With the popularity of online education, many educational technologies have been introduced to support students' learning. Among them, asynchronous discussion forums are widely used to support students' socio-collaborative learning processes. However, the forum's complex thread structure and lengthy posts often lead to poor learning experiences: Students' limited time is spent searching and filtering posts that match their interests. Accordingly, their time for engaging in more meaningful learning activities (i.e., discussion) is reduced. To address this issue, personalized learning support is needed. Forum post recommender systems are one of the possible solutions. However, none have been created for small-scale socio-collaborative learning forums that do not rely on a-priori domain knowledge and none integrate learning theories into the recommendation algorithm design. In this thesis, I introduce two similarly-structured multirelational graph-based recommender systems, CSCLRec and CoPPR. The recommender designs account for several learning theories and incorporate learner modeling, social network analysis, and natural language processing techniques. They customize forum post recommendations for learners with different social learning behaviors in order to accommodate individual learner needs in socio-collaborative online learning contexts. In the experiments with small course discussion forums, both CSCLRec and CoPPR delivered significantly better results than their competitors in terms of recommendation precision while achieving acceptable diversity and novelty performance. The results demonstrate that CSCLRec and CoPPR can predict

students' behavior and recommend relevant forum posts. These also imply the recommenders' underlying ability to solve the information overload issue and increase student engagement in discussion-related learning activities.

Preface

The thesis is original work of Zhaorui Chen. Part of the results from this thesis were originally published at the 13th Educational Data Mining conference as:

Chen, Z., & Demmans Epp, C. (2020). CSCLRec: Personalized Recommendation of Forum Posts to Support Socio-collaborative Learning. In the *International Conference on Educational Data Mining (EDM).* (pp 364-373). https://educationaldatamining.org/files/conferences/EDM2020/papers/paper_64.pdf

The first author, Zhaorui Chen developed the algorithms and conducted the experiments. The second author, Dr. Carrie Demmans Epp advised on various aspects of the project.

This study received research ethics approval from the University of Alberta Research Ethics Board, Project Name "Understanding and Evaluating E-Learning System Use: Modelling Learners, Describing Learning, and Developing Features", No. Pro00078624, May 15, 2018. The approval letter is included in Appendix B.

Acknowledgements

There are many people that I would like to express my appreciation to. In fact, I want to thank everyone I have met in the past 25 years. Every help you provided, every suggestion or criticism you offered, every word you said, either kindly or maliciously, together makes who I am now. Thank you all, for letting me now walk proudly on the road I love.

Over the past two unforgettable years, I am extremely grateful to Carrie for her careful guidance and caring support. It is my greatest luck to meet such a brilliant supervisor here that matches my research interests. Thank you for the profound belief in my ability and my work. Thank you for acting as a role model, inspiring my determination to further study in the HCI field.

I also wish to sincerely appreciate everyone in the EdTeKLA group. I am very happy to be part of this great team and work with everyone towards the goal of bringing more technology to the ones who need it. Your passion for research, enthusiasm for people, and all kinds of help to me have been of great significance to me.

I should also acknowledge the consistent support and love of my parents, my girlfriend, and the rest of my family. Your encouragement and company has been invaluable to me, especially at some difficult moments. At last, thank you, grandpa and my pet Lulu, for your blessings all the time. You are my heroes, and I miss you very much.

Contents

List of Tables

List of Figures

Glossary

Socio-collaborative learning is a pedagogical method in which learners collaborate in a group, relying on each other, being responsible for each other, and accomplishing a common task together.

Computer-supported collaborative learning (CSCL) is a type of collaborative learning. It involves collaboration among group members that is carried out through technologies and network services.

Knowledge co-construction is a learning process where learners create new knowledge or advance their current understanding of concepts through collaborative learning activities. It takes place in socio-collaborative learning contexts.

Knowledge building is a specific type of knowledge co-construction process that focuses on the creation of ideas in online learning settings and emphasizes the community's knowledge advancement as a whole.

Personalized PageRank (PPR) is a graph-based link analysis algorithm. It is often used in the research field of information retrieval and recommender systems.

Social network analysis (SNA) is the study of social relations among a set of participants inside a specific social context.

Top-K recommendation refers to a type of recommendation task in which the recommender system suggests a list of K items to each target user. In my case, $K = 10$ so a recommendation list of 10 forum posts was given to each target user.

Precision at 10 (P@10) is a recommendation accuracy measure that is used when recommending 10 items. It is the proportion of recommended items that are positive. In my case, it reflects the percentage of the recommended 10 forum posts that were actually consumed by the target user in the test week.

Recall at 10 (R@10) is another recommendation accuracy measure that is used when recommending 10 items. It is the proportion of positive items being recommended. In my case, it reflects the percentage of all forum posts that were actually consumed by the target user in the test week and that were included in the recommendation list.

Intra-list diversity (ILD) is a recommendation diversity measure. It measures how semantically diverse the recommendation list is.

Mean inverse user frequency (MIUF) is a recommendation novelty measure. It measures how unpopular a recommended item is based on how many users have consumed it. The fewer people who have interacted with the post, the higher inverse user frequency (IUF) is.

Chapter 1

Introduction

In 2020, online education has unexpectedly become a necessity though it faces tremendous challenges. In the field of higher education, as more and more post-secondary courses are forced to move online, how to achieve comparable learning experiences as traditional teaching through online education scenarios and platforms is attracting the interest of many educators and educational technology researchers. As online assessment tools make remote exams possible, researchers are exploring how to make these tools support more problem types (for example, problems involving drawing pictures are difficult to complete by computer) and to prevent possible cheating problems. Video lectures with playback functions are widely used as a primary teaching approach to facilitate learning by students with different schedules (O'Callaghan et al., 2017). However, because pre-recorded videos do not support interaction, student confusion cannot receive timely resolution. Students may also feel isolated. Although virtual classrooms with video conferencing support can help alleviate this problem, they lack support for courses that require implementation and practice (such as chemistry or nursing). In the absence of appropriate tools and environments, just watching teachers' instruction on a topic cannot effectively enable students to fully learn skills. Addressing the lack of interpersonal interaction caused by spatial separation is the key to enabling interaction within the online learning environment.

To improve the interpersonal interaction that is most commonly used, this research focuses on the asynchronous discussion forum, which is a type of educational technology that uses forum posts to share knowledge among learners. Today, in many fully online non-STEM courses, forums are increasingly used to replace traditional face-to-face discussions. When teaching in person, these discussion-based courses usually follow a socio-collaborative learning approach. Teachers encourage students to learn from each other through discussion, to jointly create knowledge, and improve their understanding of concepts (Hmelo-Silver, 2013). In such a learning context, the process of students collaboratively creating knowledge together is called knowledge co-construction. As the course shifts to online delivery, the introduction of asynchronous forums is believed to enable many advances in helping the process of co-building knowledge. For example, asynchronous forums can provide flexibility so that students can participate at times convenient to them (Hull & Saxon, 2009; Johnson, 2006). The discussion content can also be retained in the form of a post to facilitate students' review (Andresen, 2009). However, the use of asynchronous discussion forums still has limitations, which may pose challenges to students and teachers in online courses. For example, due to lengthy discussions and complex thread structure, it is difficult to find the desired information, which may cause students to lose interest or affect the efficiency of teachers (Abel et al., 2010; Hew & Cheung, 2012). This may further negatively impact the students' participation in knowledge co-construction. At the same time, students who need instructor attention may not be noticed because monitoring student status through forum posts is more difficult than talking to or directly observing students, with many questions going unanswered in large online courses (Chandrasekaran et al., 2017).

This research intends to address these concerns using recommender systems. Throughout the previous studies of forum recommender systems, most were deployed in question-and-answer (Q & A) forums in large online learning environments, for example, massive open online courses (MOOC). They are designed to reduce the number of unanswered questions. Compared to MOOC and other Q & A forums, discussion forums for university-level non-STEM courses tend to have far fewer users. The purpose of these forums is to encourage close interaction and in-depth discussions among students. Due to these differences, recommender system design for MOOC forums cannot be directly applied in the socio-collaborative learning context where students are expected to co-construct knowledge. Given that these types of Q & A forums have been a focus of recommender system work, there are few recommender systems specifically designed for smaller socio-collaborative contexts. To fill this research gap, this study attempts to design a recommender system specifically for small asynchronous discussion forums to solve these previously mentioned challenges and support the knowledge co-construction process. Therefore, this study investigates the following question: Do recommender algorithms that integrate concepts from socio-collaborative learning better predict student behaviors in a small socio-collaborative online learning environment?

I introduce two recommender systems. Both systems are guided by educational principles and they incorporate social network analysis, modeled learner types, and natural language processing techniques. The contribution of this research mainly lies in the following two points:

1. Two recommender systems that account for often ignored pedagogical considerations are introduced

2. These recommenders outperform other approaches and may better support the knowledge co-construction process in socio-collaborative learning contexts

This thesis contains 6 chapters. Chapter 2 introduces the asynchronous discussion forum as my research context and the educational theories that support this research (e.g., collaborative learning and knowledge building). It also covers similar works and introduces the commonly-used recommendation algorithms that will serve as reference points, in later chapters, to evaluate my methods against. In Chapter 3, I introduce the two algorithms that I developed (CSCLRec and CoPPR) and detail the features of the learning environment, PeppeR, where these algorithms might be used. The PeppeR system is an educational platform that hosts asynchronous discussion forums and it is the source of the data used in later evaluations. In Chapter 4, I describe the evaluations of the proposed recommendation algorithms. Along with comparison against other common algorithms, I discuss the strengths and drawbacks of various types of recommender systems. In Chapter 5, I further discuss the potential for CSCLRec and CoPPR to support the knowledge co-construction process. At the same time, I list the limitations of our methods, potential improvements, and various implications.

Chapter 2

Literature Review

2.1 Introduction

The highlight of this research lies in the consideration of educational theories in the design of recommender systems. Therefore, this chapter begins with a discussion of collaborative learning theories and the process of knowledge co-construction, which support or relate to my methods. Asynchronous discussion forums are one of the technologies that can be used to facilitate collaborative learning. They are also the environment where I intend to use recommender systems to support this type of learning. Given that recommender systems could alleviate some of the challenges that students face in discussion forums, I discuss recommendation algorithms and their use in educational contexts. Moreover, since social interactions in the forum are an important component of socio-collaborative learning processes, my method also combines social network analysis. Therefore, this technique is discussed at the end of this chapter.

2.2 Socio-collaborative Learning

Socio-collaborative learning, also called collaborative learning, is a type of pedagogical method in which learners collaborate in a group, relying on each other, being responsible for each other, and accomplishing a common task together (Smith & MacGregor, 1992). The concept of collaborative learning rose from the theory of social constructivism, which was first proposed by Vygotsky (1978) who pointed out that in a society of learners, those more able can help others perform better. Piaget (1985) claimed that the cognitive conflicts generated during social interaction could help the learner reflect on their original point of view, thus enhancing their understanding. Subsequently, in the 1990s, collaborative learning gradually emerged as one of the most used pedagogical theories in education. In socio-collaborative learning contexts, students can improve themselves by interacting with others and drawing on the strengths of others, and the success of one individual is related to the success of others. This ubiquitous process can take place in face-to-face communication or through many computer technologies, such as online forums.

2.2.1 Computer-supported Collaborative Learning

Computer-supported collaborative learning (CSCL), as a type of collaborative learning, has the collaboration between group members carried out through educational technologies and network services. As indicated by Koschmann (1996), computer technology was first used as a strategy for educating individual learners in the 1970s. After the introduction of the CSCL concept in the 1980s, researchers began developing many software tools for CSCL. With the advent of big data in the past decade, technologies that incorporate intelligent decision-making have become the focus of CSCL research (Jeong & Hmelo-Silver, 2016).

CSCL focuses on using computers and the Internet as media. Any software or tools that are used to facilitate collaborative learning are realizations of CSCL theories. They can work solely online or blended with classrooms. The asynchronous discussion forum is a typical CSCL technology. Inside of it, information is transmitted asynchronously through posts to enable knowledge sharing among learners. Forum posts, in the form of threads and replies, are a medium that can carry information and knowledge exchanged between learners, resulting in a learning process called knowledge co-construction. This thesis focuses on this fundamental learning process within discussion forums.

2.2.2 Knowledge Co-construction

This learning process can take place in socio-collaborative learning contexts. Knowledge co-construction is often triggered by collaborative learning activities, for example, discussion, debate, or the negotiation of the meaning of concepts. As a result, learners are able to create new knowledge or advance their current understanding of concepts.

With the popularization of educational technology, there is an increasing research interest in studying the knowledge co-construction process in collaborative learning contexts (i.e., educational forums). For example, Tirado, Hernando, and Aguaded (2015) explored how network structures affect the quality of knowledge co-construction processes in asynchronous discussion forums. They employed content analysis to evaluate the knowledge co-construction process from the perspectives of social and cognitive presence. Through social network analysis and structural equation modeling, they discovered that the cohesion and centralization of the social network established in the forum can have a positive impact on the co-construction process. In another study, Fu, van Aalst, and Chan (2016) conducted extensive research of user discourse patterns in a discussion forum. With the help of qualitative coding, narrative analysis and teachers' opinions, they developed a set of coding schemes to distinguish 9 different discourse patterns. This result enabled them to further distinguish which learner discourse is in the mode of knowledge construction.

The study of knowledge building investigates a specific type of knowledge co-construction process that focuses on the creation of ideas through discussion in online learning settings (Scardamalia & Bereiter, 1994). Knowledge building treats a classroom as a learning community where knowledge sharing and construction can take place. It emphasizes the community's knowledge advancement as a whole and features a set of social principles and practices that could positively affect the learning process in the learning community (Scardamalia & Bereiter, 2006). Knowledge building emerged from the notion of intentional learning in 1988 (Scardamalia & Bereiter, 2010). Along with the proposal of knowledge building theory, the KnowledgeForum platform was created to support the collaborative construction of community knowledge (Scardamalia, 2004). A list of twelve principles was later proposed to act as guidelines in the design of many kinds of online educational platforms, including the asynchronous discussion forums that are a focus in this thesis (Scardamalia et al., 2002). Of the established principles, the following were considered in the recommender system design and further inspired the experimental design presented in Chapter 4.

Democratizing knowledge highlights that everyone's idea is equally important to knowledge co-construction within the community. Therefore, the forum design should encourage everyone to express their opinions and should not hinder any voices.

Idea diversity emphasizes the benefits of having diverse ideas in knowledge sharing communities. Besides people with similar views, those with conflicting views should also discuss with each other. Debates and negotiations triggered by this conflict allow for reflection and may lead to refined understandings (Piaget, 1985).

Symmetric knowledge advancement indicates that knowledge advancement is a bi-directional process. When those more capable learners offer help to other community members, they can benefit from the process, for example by consolidating and reviewing their understanding.

There are 9 other principles within knowledge building. Most of them involve teaching or administrative strategies that are used in the learning community. They emphasize the social learning practices of participants, such as encouraging rise-above comments (i.e., synthesizing previous ideas and raising the concepts to higher levels) or referencing authoritative sources. Such strategies are meant to be performed by humans and are not within the scope of recommender systems. In addition, some other principles have already been supported by socio-collaborative learning forums, for example, supporting improvable and revisable comments.

In the past decade, many suggestions have been made for how to support the knowledge co-construction process in order to promote learning in forums (Schellens & Valcke, 2006; Veldhuis-Diermanse, 2002). Motivated by suggestions that emphasize the help that adaptive systems give to members of a learning community (Hmelo-Silver, 2013; Rosé & Ferschke, 2016), this thesis sets out to explore whether adaptive systems (such as recommender systems) could support the process of knowledge co-construction.

2.3 Asynchronous Discussion Forums

The asynchronous discussion forum is a web-based communication space that enables users to discuss, share ideas, or solve problems together. In the past years, its support for online courses in higher education has attracted much research interest and various suggestions for improvement have been raised (Akcayir et al., 2020; Demmans Epp, Phirangee, & Hewitt, 2017; Dowell et al., 2017). The asynchronous nature of discussion forums keeps historical records and does not require both parties of the conversation to be online at the same time. Compared with traditional face-to-face discussions, this type of remote interaction is considered to have multiple advantages: course delivery can be flexibly adapted to every student's schedule; archived course materials can facilitate students' review and reduce miscues (Geer, 2005); and students' communication is not constrained by time zone differences (Hull & Saxon, 2009). Therefore, asynchronous discussion forums have become a widely used environment for distance learning and are playing different roles: they are sometimes used to assist in traditional face-to-face teaching (i.e., Q & A forum), and sometimes as a standalone tool to host all teaching activities of a fully online course (i.e., discussion board).

This thesis focuses on a discourse-based forum that supports small learning communities, such as discussions in graduate classes. This kind of collaborative learning community heavily relies on student discourse and has an emphasis on peer to peer interactions (Kear, 2004). Instructors act as moderators such that they post topic threads and guide students to reply under the post. As a result, students communicate with each other, debate, and exchange ideas, so as to co-construct their knowledge. Forum posts published in the form of replies are a medium that can carry the information and knowledge exchanged and transferred between learners.

While this type of information sharing and building is sometimes achieved, the design and use of asynchronous discussion forums could negatively impact the implementation of knowledge co-construction learning approaches. Among them, the most discussed challenge is low student engagement in the discussion forum (Mason, 2011), which may lead to a lack of shared resources (Hammond, 1999) and a high dropout rate in the entire course (Bonet & Walters, 2016; Xiong et al., 2015). Many approaches have been proposed and examined to alleviate these problems. For example, Comtella employed a reward-based approach, in which its system adaptively awarded participation points to students based on their contributions and the community needs in an attempt to stimulate student participation in learning activities (Vassileva, McCalla, & Greer, 2016).

A similar incentive mechanism was also applied in MOOC forums where some gamification strategies were trialed. Similar to Comtella, rewards and status were employed through the use of badges (Ding, Kim, & Orey, 2017; Ortega-Arranz et al., 2019) and leaderboards (Morales et al., 2016) that were given to students as they completed a milestone or an online course. Case studies showed that this kind of adaptive reward-based method helped increase student engagement and retention rates. On the other hand, some also realised that student engagement may be hindered as they are overwhelmed by the massive amount of data that is presented to them. As a solution, they developed a graphical interface (Marbouti & Wise, 2016) or customized recommendations (Yang et al., 2014) to alleviate the information overload problem brought by the complex thread structure and the continuously growing body of posts.

As the study of student engagement was brought to the individual level from the course or group level, researchers found that student participation and forum activity was also associated with different learner behaviors. For example, Hewitt (2005) identified a ”single pass” behavior that may cause low longevity of forum threads. The “single pass” behavior refers to the act of students not rereading posts they have previously viewed. As a result, due to the lack of continuous attention, most posts will be terminated early, resulting in reduced forum activity. Another notable study identified a "listener" type of student who only viewed posts without contributing their own voices to the discussion (Wise, Hausknecht, & Zhao, 2014). These students are also called lurkers (Brooks, Greer, & Gutwin, 2014). In a collaborative learning environment, students are creators of forum posts. If everyone performs the single-pass or listener behaviors, the entire learning community will lack resources for sharing and learning, which will reduce forum activities and knowledge construction (Guy, 2015).

To identify other sharing patterns, a study that employed social network analysis techniques aimed to provide customized prompts to encourage student activities. Users with different knowledge sharing patterns, including new users who just joined the community and peripheral users who lost interactive readers from the previous days, were analyzed and prompted to participate with encouraging notifications (Kleanthous & Dimitrova, 2013). These findings indicated that customized strategies are needed to cater to different types of learner behaviors such that student engagement can be enhanced in a way that improves forum activity and the knowledge construction process. As the goal of this thesis is to use personalized recommendations to improve the co-construction process, these findings about asynchronous forum activity have been taken into consideration in the recommender system design.

2.4 Recommender Systems in Education

The vast amount of information in today's webspace makes it difficult for users to find the content they may be interested in or that will be helpful to them. To alleviate this information overload problem, recommender systems have been introduced in various web platforms including educational discussion forums. These systems analyze and infer the user's preferences through algorithms to filter or prioritize content for users. The customized recommendations could save users the time to locate relevant materials, improve work efficiency, and enhance the user experience.

In e-learning research, many recommender systems have been developed to serve the purpose of enhancing learning experience or optimizing learning efficiency and outcomes. In a learning management system, researchers mined students' past behaviors and developed a recommender system that can suggest navigation shortcuts to them, thereby reducing the cognitive effort involved in using the software (Zaiane, 2002). Recommendations were also made available in a work-integrated learning environment to pair users with knowledgeable peers to facilitate their knowledge sharing and communication processes (Beham et al., 2010; Vassileva, McCalla, & Greer, 2016). Along with the growth of Massive Open Online Courses (MOOCs), online courses have been recommended to students to meet their career goals and current levels of knowledge (Bousbahi & Chorfi, 2015).

Prior research on recommender systems in educational forums has primarily focused on Q & A forums, where students are expected to answer questions from each other, thereby reducing the workload of instructors. In this way, instructors can have more energy to help students, so that each student can obtain a learning experience similar to traditional courses. In such forums, recommender systems usually recommend unanswered questions for students who are able to answer them, and recommend similar questions that have already been answered for users who are about to ask one, thereby reducing the number of unanswered questions. There is also research on recommending peers so that they can help each other (Vassileva, McCalla, & Greer, 2016). In recent years, many works in MOOC forums aimed to overcome the challenges of having large amounts of forum activity and a fast post updating rate (Lan et al., 2019; Yang et al., 2014). These studies have improved the accuracy and timeliness of recommendation results and have shown the potential of these techniques for enhancing student learning efficiency and reducing dropout rates.

However, the same recommender system design is not expected to provide the same results when used in smaller-scale socio-collaborative contexts as their application domains and pedagogical purposes are sufficiently different. Compared with MOOCs, small communities need to take precautions to handle the cold start problem. Unlike Q & A forums, socio-collaborative contexts ought to strengthen inter-user connections to meet the goal of fostering the knowledge sharing processes (Kleanthous & Dimitrova, 2013). Besides content that satisfies user tastes or preferences, it is necessary to include posts containing diverse and novel ideas from students who express different points of view so that students might learn from each other and those minority ideas are noticed (Scardamalia & Bereiter, 2006). In addition, as discussion forums usually have a complex thread structure and long discourse content, the majority of the students' limited time for learning are dedicated to filtering redundant information, rather than focusing on discussion and knowledge scaffolding (Abel et al., 2010; Hew & Cheung, 2012). When designing a recommender system in a small-scale socio-collaborative context, these aspects all require extra attention.

Few prior studies have addressed the aforementioned challenges in small socio-collaborative contexts. In one example, Abel and colleagues proposed a rule-based recommendation mechanism to select the optimal collaborative filtering algorithms in an effort to produce recommendations and alleviate the cold-start problem when working with small datasets (Abel et al., 2008). Another approach to solving this problem categorized forum messages according to a topic ontology and recommended messages that match student interests to support the broadcasting of collaborative knowledge (Chen & Persen, 2009). However, both approaches relied on prior knowledge of the domain-specific environment (i.e., an ontology of learning resources). This requirement demands prior access to domain expertise, is time-consuming, and has limited generalizability to unseen cases (Jones, Bench-capon, & Visser, 1998). A generalizable solution that requires little a priori domain knowledge of the recommended resources becomes necessary.

Going beyond trying to solve these problems, many studies have suggested that social relations (Jeong & Hmelo-Silver, 2016; Rosé & Ferschke, 2016) and needs from different types of learners (Bergmeir & Benítez, 2012; Manouselis et al., 2012; Santos, Boticario, & Pérez-Marín, 2014; Vassileva, 2008), should be considered in designing CSCL personalization technologies. Therefore, customized recommendation strategies for different learner knowledge sharing patterns are incorporated in the recommender system design.

2.5 Recommendation Algorithms

Recommender systems are driven by recommendation algorithms. Different algorithms can be adopted to meet different data patterns and recommendation goals. There exist two types of entities in every recommender system: users and items. "User" denotes people utilizing the recommender system and receiving recommendations, and "item" refers to the resources which the system recommends to users. At the stage of recommendation, each user is expected to receive a list of items. The list of items are chosen by the recommendation algorithm from a pool of "candidate items". The user who receives the recommendations is called the target user. In educational scenarios, users can be students, learners, or instructors; items can be courses, books, or any other types of learning resources.

To analyze user behavior so as to determine user preferences, recommender systems exploit "user-item interactions". These interactions are historical logs that record user feedback about those items or how the items were utilized by users. Feedback can be numerical, for example in the form of ratings or scores. Feedback can also be categorical such as the like or dislike indicator. In many scenarios, user feedback can also be less direct: for example, the system only records whether the user has purchased an item. This situation is called "implicit feedback" or "unary feedback", and it often causes uncertainty and complications to the analysis of user interests (Oard & Kim, 1998; Ricci, Rokach, & Shapira, 2015).

Implicit feedback is common in education recommendation scenarios, because many education systems do not force users to rate learning resources. As an example, many educational forums do not offer an option to rate posts so they only have user interaction records. Some researchers of recommendation algorithms are exploring how to use this type of feedback or mitigate the impact of ignoring it (Pan et al., 2008; Sindhwani et al., 2010). Although my research does not contribute to this aspect, some recommender systems considering implicit feedback are used as comparison points with my proposed systems.

Keeping this in mind, I introduce several commonly-used recommendation algorithms and instances of their use in the field of online education. Some of them were implemented in this thesis, so as to compare against the proposed approaches.

2.5.1 Collaborative Filtering

Collaborative filtering is one of the most applied approaches in recommender systems. In principle, collaborative filtering recommends resources to users according to what other users who share similar tastes liked in the past (Ricci, Rokach, & Shapira, 2015). Predictions are made based on other users' prior behaviors, activities, or preferences. All collaborative filtering algorithms follow a general idea that if a person, Abby, has a similar rating to another person, Bob, on one item, then Abby is more likely to have similar ratings to Bob on other items.

A possible drawback of collaborative filtering is the long tail problem. It refers to a circumstance where a small portion of the items were recommended because they have huge amounts of data, while the rest are left unnoticed (Park & Tuzhilin, 2008). It harms recommendation coverage and could also lower the diversity of recommendations. Collaborative filtering algorithms tend to be subject to the long tail problem because their underlying mechanism relies on user-item interactions which biases towards popular items: An unpopular or even never-used item is not treated as a common interest between any two users due to insufficient usage history, hence it will not be included in any collaborative filtering recommendations. This issue presents a challenge to social learning contexts in which everyone's opinion should be equally important because people learn from the debates or negotiations that they have over conflicting opinions. The long tail problem may only show recommendations of mainstream opinions and block the opportunity for learners to get to know what less popular views are, thus negatively affecting the co-construction process.

In addition to the long-tail problem, collaborative filtering methods are not robust to the cold-start problem. Cold-start refers to a situation where there are little to no resources for recommender systems to determine the appropriate recommendations. In the context of discussion forums, the cold-start problem can take two forms. The first type occurs when a new user joins. In this case, the recommender could have trouble figuring out what that user likes. The second type of cold-start problem is when a discussion forum just opens. As there are no forum posts, no recommendations can be generated by recommender systems. Because collaborative filtering only relies on interactions, when there are no interactions during some specific situations (i.e., new user or new forum), it will not be able to make any recommendations.

Despite their limitations, collaborative filtering approaches still receive considerable attention and a number of variations have been developed. Among them, memory-based collaborative filtering and model-based collaborative filtering are the mainstream methods for implementing this approach in CSCL (Drachsler et al., 2015).

Memory-based collaborative filtering

Memory-based methods are also known as neighbourhood-based methods. They are the most intuitive implementations of the general collaborative filtering principle. These methods rely on the historical interactions between items and between users. Such interaction history data are stored in the memory as a user-item matrix. An exemplar use is the user-user method: at the stage of recommendation, it first locates the most similar users of the target user based on the user-item matrix, and then uses the weighted average of these users' feedback as an estimate of the target user's feedback (Ning, Desrosiers, & Karypis, 2015).

Memory-based collaborative filtering algorithms are commonly used as baselines in educational recommender system research. For example, in a study that aims to sequentially recommend embedded assessments that adapt to student knowledge level, a memory-based collaborative filtering system was developed and compared with model based ones (Min et al., 2013). Their results showed that the memory-based approach, though outstanding in running time, was outperformed by its model-based counterparts in prediction accuracy, suggesting the potential for model-based approaches and a need to investigate them in my research context. In another case, memory-based collaborative filtering was used to provide recommendations in an educational resource repository (Karampiperis, Koukourikos, & Stoitsis, 2014). The highlight of this research lies in the application of sentiment analysis of user reviews to overcome the lack of explicit user ratings. This research suggests that it is also important to analyze the textual information contained in posts when analyzing forum interaction data. However, sentiment analysis does not apply to my specific context, because the discussion content may contain domain specific phrases that can be misclassified in sentiment lexicons. For example, the neutral argument "Students who concentrate tend to get good grades" may be classified as having a positive sentiment because of the word "good".

Model-based collaborative filtering

Model-based methods, or latent factor models, also rely on the user-item matrix despite the fact that the matrix usually does not come directly from historical data. In model-based collaborative filtering variations, users and items are represented by feature vectors in a lower dimensional space. The user-item matrix is an inner product of the feature matrix of all users and the feature matrix of all items. Therefore, the user's preference can also be estimated by the inner products of the user vector and the item vector (Koren & Bell, 2011).

Model-based collaborative filtering has become the most commonly-used recommendation algorithm to support learning in educational scenarios (Drachsler et al., 2015). For example, in a recommender system study for Massive Online Open Course (MOOC) forums, Yang et al. (2014) proposed a model-based collaborative filtering system that also incorporates content analysis and social network analysis. Their work achieved promising results when recommending forum posts, showing the potential to address the thread overload problem brought by the increasingly expanding class size.

2.5.2 Content-based Filtering

Content-based filtering is a type of recommender system that utilizes item features and characteristics (de Gemmis et al., 2015). By analyzing the item features that the user preferred in the past, the recommender will infer the features that the user may like and it will recommend unseen items with similar features. Genre-based music recommender systems are a typical example: they infer the user's favorite music genre from the user's listening history, and then recommend music belonging to this genre (Magno & Sable, 2008).

In addition to the type of content-based algorithm that utilizes item features, in some scenarios where items do not have known labels or features, automatic labeling is a common solution (de Gemmis et al., 2015). They often appear in scenarios that recommend textual resources, such as news (Capelle et al., 2013) or forum posts (Bach, Hai, & Phuong, 2016). Using various natural language processing (NLP) methods, such as keyword extraction or topic modeling, these text resources can be represented through automatically extracted features or represented as vectors. A common approach is to use the term frequency-inverse document frequency (TF-IDF) method to encode text, and then apply the latent semantic indexing (LSI) technique to reduce the embedding dimension, and finally obtain a vectorized representation of each textual item. Each user is also assigned with a user profile embedding, represented by the average of the item embeddings from all those they liked before. Finally, as both user profile embeddings and item embeddings are mapped to the same vectorized feature space, the algorithm recommends to the target user the nearest items in the feature space.

Because its recommendation only depends on item characteristics, content-based systems do not suffer from the long-tail problem. At the same time, these approaches tend to have good scalability since they do not rely on the user-item matrix for inference.

However, drawbacks of content-based approaches may be exposed when the characteristic labels are not available and the automatic labeling solutions (i.e., text embedding) are not applicable: a labor-intensive and time-consuming manual labeling procedure becomes necessary. This procedure has to employ a considerable number of domain experts or crowdsourcing resources. Moreover, the recommender system produced by this method is limited to the same domain topic. Application in other domains is not appropriate as it requires a redo of the entire manual labeling step. The recommender system design for university course discussion forums (such as this thesis's background) is a typical example. Forum posts do not usually have user-created tags. In addition, since discussions in each individual community are limited to domain-specific content, pre-trained text embeddings may have difficulty differentiating posts. For the above reasons, an automated recommender system that combines content-based methods with other recommendation mechanisms is necessary in order to reduce the system's dependence on excessive manual work.

Content-based methods are also considered to be prone to the problem of over-specialization (de Gemmis et al., 2015). Since the recommendation mechanism only depends on the similarity of item features, similar items can always obtain a higher recommendation score than others. As a consequence, when the number of these similar items is large enough, a filter bubble will be formed, and all recommendations will only be generated within the bubble (T. T. Nguyen et al., 2014). For example, when a student subscribes to a few machine learning courses, the course recommender system would prefer to only recommend machine learning courses in order to obtain high recommendation precision, while ignoring all other courses in similar domains. The over-specialization phenomenon also presents challenges to discussion forums, as it prevents users from jumping outside of their comfort zones. As a result, users are forced to receive repetitive and topically limited content to discuss. This may indirectly harm user experience and reduce their learning outcomes.

The cold-start problem can also happen in the case of a content-based recommender system. Carefully-designed content-based recommender systems can avoid the cold-start problem by asking for user preferences at the beginning, for example, prompting them to enter their topics of interest upon login or account creation (Rubens et al., 2015). However, this approach will not solve the cold-start problem brought on by a freshly created new forum.

Content-based algorithms are frequently used to personalize text-based environments such as digital libraries or social forums. In an exemplar case, Wang and Blei (2011) introduced a recommender system that suggests scientific articles using a topic modeling technique (i.e. Latent Dirichlet Allocation (Blei, Ng, & Jordan, 2003)) to extract features of user comments and build user profiles. In another study, content-based recommenders were deployed in online discussion forums to recommend posts that are helpful to users or that might be of interest to them (Albatayneh, Ghauth, & Chua, 2018). In this system, users were explicitly asked to provide ratings towards forum posts to express their interest. The authors applied content analysis techniques to analyze learners' negative reviews, so as to make the recommended items fit the user's background and improve the recommendation accuracy. In the subsequent evaluation of the user's learning outcomes, this addition is considered to have a positive effect on the learner's performance.

Content-based approaches' various benefits prompted me to integrate the content-based ideas in the proposed recommender system design. At the same time, some conventional content-based recommender system methods were also implemented for comparison with my proposed methods.

2.5.3 Graph-based Recommendation

Graph-based recommender systems have received research interest in recent years (Drachsler et al., 2015). They extend the idea of utilizing user-item interactions in collaborative filtering methods and represent the interactions in graph-like structures. Graph-based methods emerged as a solution to problems that traditional recommender system approaches, such as collaborative filtering could not solve (Ning, Desrosiers, & Karypis, 2015). Typical cases include the limited coverage and the cold start problem, especially when limited interactions are available.

A typical graph-based recommendation algorithm incorporates a bipartite graph that treats users and items as graph nodes. They are connected by edges that represent historical user-item interactions. At the stage of recommendation, they employ graph traversal algorithms, for example the personalized PageRank (PPR) algorithm (Haveliwala, 2003). A random walk agent is initialized from the target user node and set to go over the entire graph for a large number of interactions until the visiting probabilities of each node converge to a steady state. Such graph algorithms usually take actions to avoid falsely computed probabilities caused by self-loops and dead ends. One of these solutions includes using a teleportation mechanism that forces the random walk agent to occasionally restart from the target user node regardless of the transition probabilities. In the end, the most frequently visited item nodes are used as the recommendations for the target user. The proposed systems in Chapter 3 follow a graph-based recommendation mechanism based on the PPR algorithm.

Since these recommenders use graph search algorithms to find transitive relationships between data that are not connected (Ricci, Rokach, & Shapira, 2015), they not only preserve the original local relationships between data, but also can discover some potential relationships, thus alleviating the limited coverage problem (Ning, Desrosiers, & Karypis, 2015). For example, El Helou, Salzmann, and Gillet (2010) proposed a graph-based recommender system architecture and claimed that it can support personal learning in a CSCL environment. Their recommender system can provide heterogeneous recommendations, in which users and different types of resources can be suggested at the same time. Their method was tested on a non-educational dataset, and the results showed that it had a better recall rate than a user-based collaborative filtering system.

2.5.4 Hybrid Recommendation

Hybrid recommender systems have gained popularity in the past decade. They combine multiple recommendation approaches together to improve the recommendation output, since one approach's advantage can be used to overcome the shortcomings of another (Burke, 2002). For example, Bach, Hai, and Phuong (2016) presented a recommender system design which employed the learning-to-rank algorithm to connect both a memory-based collaborative filtering and a topic-modeling content-based algorithm. In comparison with pure collaborative filtering methods, the hybrid solution showed higher accuracy when broadcasting personalized user comments in social forums. In another research, a graph-based system that naturally hybridizes collaborative filtering and content-based algorithms outperformed the stand-alone systems when recommending books in a digital library setting (Huang et al., 2002). The algorithms that I propose in Chapter 3 are hybrid systems that combine graph-based algorithms with content-based ones, to overcome the limited coverage and over-specialization problems.

2.5.5 Other Algorithms

It should be noted that the types of recommendation algorithms are not limited to the above. However, not every type of recommender system can be widely adopted given the limits imposed by the characteristics of educational data. One example is deep learning recommendation algorithms. In recent years, with the boosting of computing power, these deep approaches are increasingly used in many non-educational scenarios such as music streaming (van den Oord, Dieleman, & Schrauwen, 2013) or commercial product reviews (Zheng, Noroozi, & Yu, 2017). However, these data-hungry methods can easily overfit when the data set is small (Zhang et al., 2019), as is common in educational settings. For example, in discussion forums for graduate-level classes, due to the limited number of enrollees and course length, there is usually not enough textual information (discussion posts, messages) to support the training and use of deep learning methods.

The use of a hybrid system in this research offers flexibility to adapt more features. To make my recommendations more in line with the needs of students in a social learning community, some educational theories and additional techniques have been incorporated into the design of the recommender system. One of the key approaches that I used involved network analysis techniques to derive the social network formed within each learning community.

2.6 Social Network Analysis

Social network analysis (SNA) is the study of the social relations among a set of participants inside a specific social context. The modern theory of SNA emerged in the 1930s when Moreno established a very basic principle in social network analysis: the position in the social structure has an impact on the people who occupy them (Moreno, 1934). Sociograms that visualize social connections were also introduced shortly after. As the mathematical modeling of social structure became popular in the 1950s, mathematical tools were developed to support social network theories (Cartwright & Harary, 1956). Nowadays, SNA is playing an important role in educational research. Researchers have applied it to different educational technology scenarios to complete various tasks, for example, exploring the formation of learning communities (Grunspan, Wiggins, & Goodreau, 2014) and discovering user characteristics (Hallinan & Williams, 1990).

SNA techniques are being applied in more and more CSCL studies. For example, in a study that modeled closely-knit communities as social networks, the authors analyzed user activities and categorized different behavior patterns (Kleanthous & Dimitrova, 2013). Thanks to the extracted information, the resulting adaptive system that broadcasts personalized prompts accomplished the system's goal of promoting the overall knowledge sharing process within the learning community. SNA techniques were also used in the proposed recommender system to provide more insight into the recommendation strategy design in order to adapt to different forum participation behaviors.

2.7 Conclusion

In this chapter, I introduced the background theory (collaborative learning) and application scenario (asynchronous discussion forum) which this thesis research is based on. After reviewing some related works, we noticed that there exists a demand for a recommender system that does not rely on a-priori knowledge of the course content when discussion forums that practice socio-collaborative learning approaches are used. Moreover, as few recommender system studies have drawn elements from educational theories, this thesis looks to fill this gap by integrating collaborative learning theories and principles into recommender system design. Therefore, two hybrid recommender systems that incorporate natural language analysis and social network analysis are proposed and evaluated in subsequent chapters.

Chapter 3

System Description

3.1 Introduction

This chapter's discussion starts with the introduction of PeppeR. It is a forum-like discussion workspace that has been used for many graduate-level classes in the humanities and social sciences. As PeppeR aims to facilitate collaborative learning processes, it was used to conduct the evaluation study and validate my recommender designs in later chapters.

To provide personalized support in PeppeR and similar social learning forums, two proposed recommender systems, CSCLRec and CoPPR, that account for socio-collaborative learning principles are presented in this chapter.

3.2 PeppeR

PeppeR is an educational platform deployed at the the Ontario Institute for Studies in Education (OISE) of the University of Toronto[1]. It provides socio-collaborative learning workspaces in the form of discussion forums for many graduate-level courses in education disciplines. PeppeR's asynchronous discussion feature enables students to communicate, exchange ideas, and reflect their insights without introducing time constraints.

For online courses that use discussion forums as the only teaching instrument, their forums on PeppeR are essentially learning communities that aim to implement collaborative learning and knowledge co-construction approaches

[1] https://pepperproject.ca/

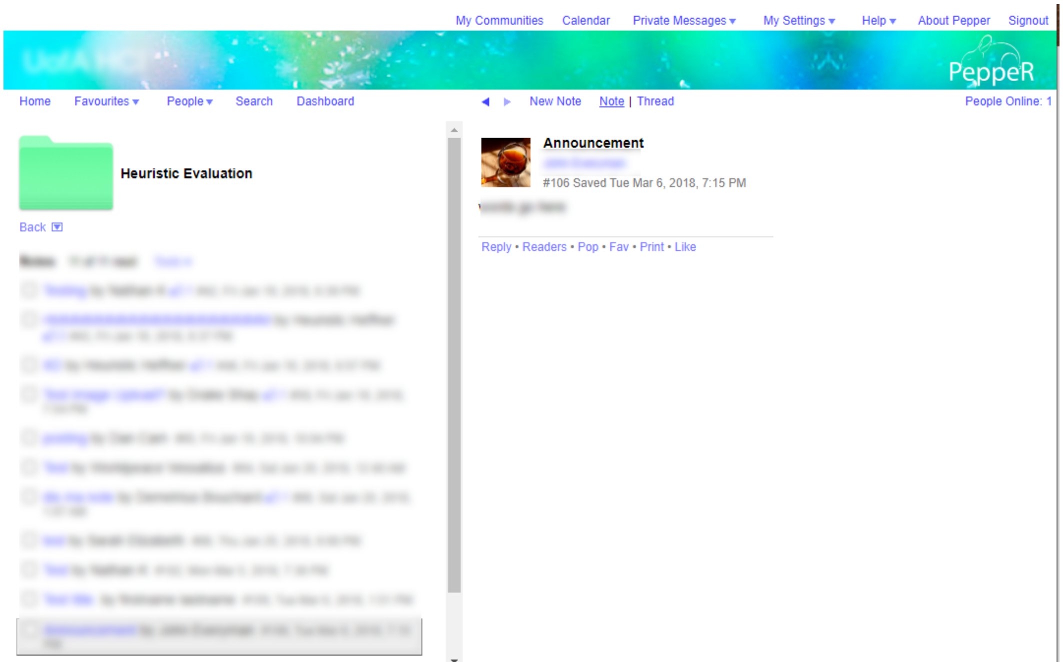

Figure 3.1: A screenshot of the PeppeR user interface. The left panel shows the thread structure overview, and the right panel shows the details of a forum post. The text has been blurred to protect student identities.

(Oztok et al., 2014). Students, instructors, and visitors are the three types of participants in the community. Among them, instructors are the administrators and coordinators. Their responsibilities include organizing the structure of the discussion topics, assigning tasks and guiding students to conduct learning activities through discussion in designated places. Students, on the other hand, are encouraged to collaboratively negotiate, discover, or create knowledge through a series of operations such as reading, posting, and replying to posts.

As shown in Figure 3.1, the PeppeR forum has a hierarchical structure where forum posts are nested together. The post title is displayed in the thread structure thumbnail on the left. Users click on the title to read the content. Below the content of the post, there are actions, such as reply, like, and link, that the user can choose from.

3.2.1 Users, Items, and Interactions

Since the recommender system is designed for students, I refer to students as users. The term learner and student will be used interchangeably. Similarly, I refer to forum posts as items to be recommended. They are also called learning resources.

Interactions between user and item (user to post interactions, U2P) are categorized into the following types:

- Creation: creation indicates that a user writes a post.
- Reply: reply indicates that a user replies to an existing post. As a reply itself is a post, the reply interaction always accompanies a creation interaction.
- Like: like is an interaction where the user clicks on the "Like" button to express their interest in a post. The post can be their own or can be created by others.
- Link: link refers to the action in which a user inserts hyperlinks to other posts in her own posts.
- Read: read data saves a certain user's post reading record. In PeppeR, posts that have never been read by the user will display the title in bold.
- Read anonymously: Some courses in PeppeR give users the privilege of viewing posts in the incognito mode. As this research does not explore the reasons why users hide their reading actions, those records are seen as equivalent to read records.
- Revisit: revisit data keeps a record of users repeatedly reading posts. Starting from reading a post for the second time, every read or anonymous read of the same post will be extracted as a revisit record.

Inter-user interactions, also known as user-to-user (U2U) interactions are categorized into 5 types: reply, like, link, read, and read anonymously. Unlike the aforementioned user-to-post interactions, this emphasizes connections of

Person	Post ID	Action	Week
Abby	60594	Read anonymously	2
Bob	74929	Like	7
Charles	85011	Create	13
David	93128	Read	4
Evan	84670	Read	13
Fred	105214	Link	9

Table 3.1: Excerpt of the forum activity data recorded in the PeppeR database.

one user with another. For example, if user Abby replied to user Bob's post, there will be a user-to-post interaction of reply type (Abby to Bob's post) and a user-to-user interaction of reply type (Abby to Bob). The presence of such interactions benefits the analysis of user behaviors and the construction of social relationship networks in the community.

As illustrated in Table 3.1, each interaction record has detailed information such as the action type, time, forum post identifier (ID), and the person who triggers this action. Historical data extracted from PeppeR was used for building and testing the recommender systems. The evaluation procedures will be provided in Chapter 4.

3.3 CSCLRec and CoPPR

This study presents two multi-relational graph-based recommender systems, CSCLRec and CoPPR[2]. Both of them incorporate social network analysis, learner categorization, and natural language processing techniques.

3.3.1 Structure Overview

Both CSCLRec and CoPPR are graph-based recommender systems as their core is a modified version of the personalized PageRank (PPR) graph (Haveliwala, 2003). The proposed CSCLRec and CoPPR algorithms use similar approaches. They have an identical design in the learner profiler and post filtering modules. The only difference lies in the content analyzer module where CoPPR

[2]Code available at: https://github.com/EdTeKLA/CSCLRec; My presentation video from EDM 2020 is available at: https://youtu.be/jeBRKsp5OHA.

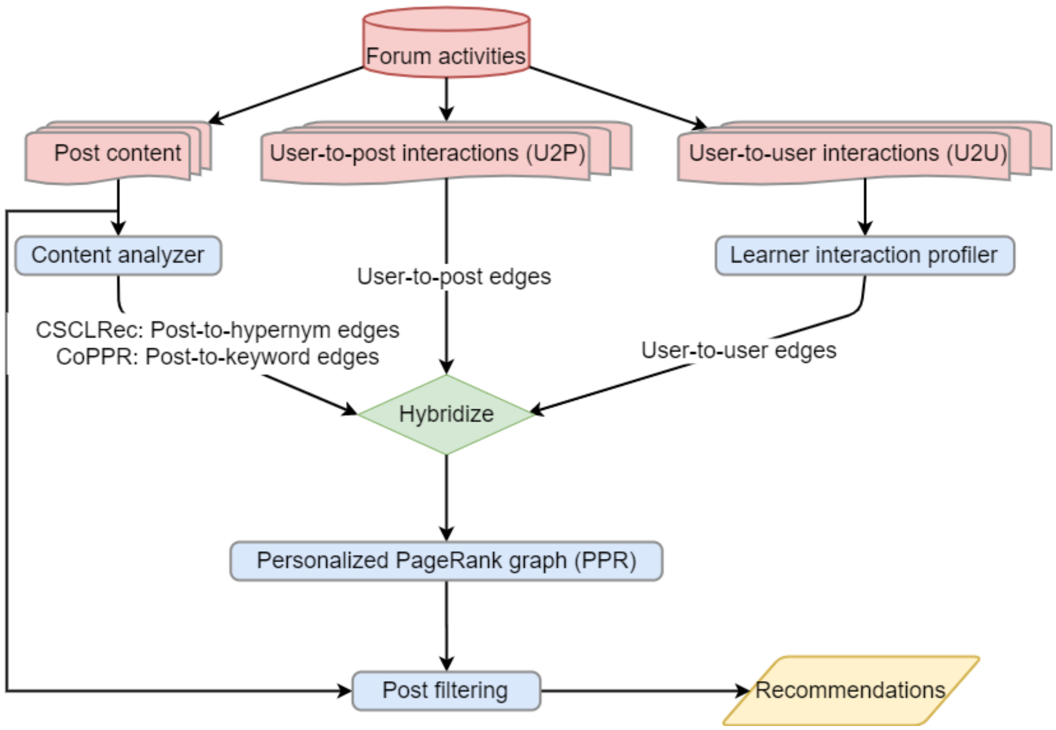

Figure 3.2: Overall workflows of CSCLRec and CoPPR

directly uses the extracted keywords as nodes and CSCLRec uses hypernyms.

Consider CSCLRec, as shown Figure 3.2. Its procedure starts with the extraction of three types of data from the logs of forum activities, namely the U2U interactions, the U2P interactions, and the post contents. As a hybrid recommender system, CSCLRec consists of four modules: a learner interaction profiler module that analyzes U2U interactions and generates the user-to-user edges in the social network graph; a content analyzer module that takes the post contents as input and outputs edges that connect posts; a personalized PageRank graph module that integrates the user-to-post edges extracted from U2P interactions with the other two types of edges; and a post filtering module that filters posts based on their content to remove less helpful ones before displaying the final recommendations. CoPPR follows a similar procedure. The only difference is that the content analyzer outputs the keyword nodes and post-to-keyword edges, as the input to the PPR graph.

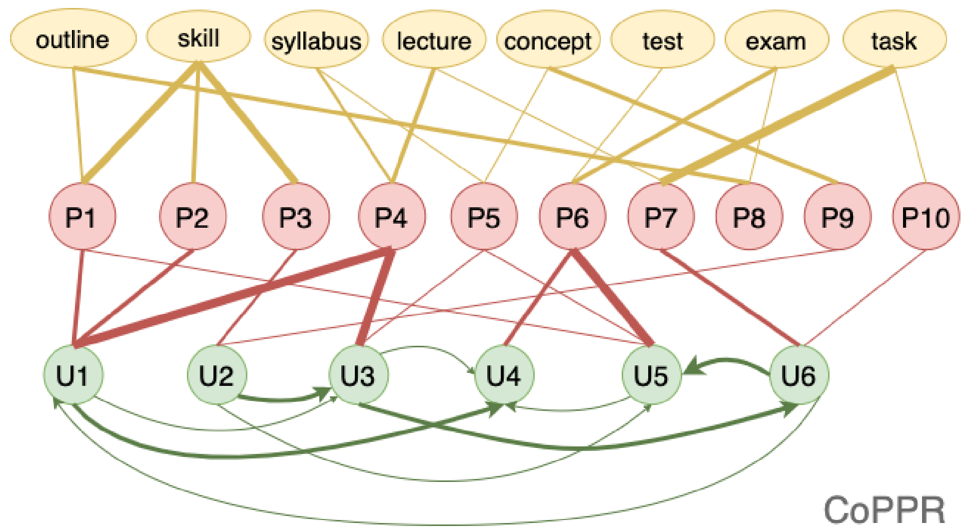

Figure 3.3: CoPPR's PPR module

3.3.2 PPR Module

The limited size of small socio-collaborative environments, such as PeppeR, motivates the use of PPR as the recommendation algorithm, as it handles sparse and small datasets well (Fazeli et al., 2014). As shown in Figure 3.4, in CSCLRec, the PPR graph has users, posts, and hypernyms as its nodes. A hypernym is a linguistic term that represents a type-of relationship, for example, the word "fish" is the hypernym of the words "tuna" and "salmon" because both of the latter are fishes.

Edges are generated either from the output of other modules or directly from the U2P interactions. In CoPPR, as shown in Figure 3.3, its PPR graph contains keywords and post-to-keyword edges. All edges are weighted. Their weights are determined by the strength of the relationship between two nodes. For example, the weight of a post-to-keyword edge will be higher if a keyword appears multiple times in a post. Since the graph contains different types of edges, this module takes precautions to ensure that they are comparable: the weight within each type of edge is unitized to have values between 0 and 1. More details regarding edge generation will be provided in the subsection describing each module.

To reflect the temporal effect of user preferences, the system imposes a

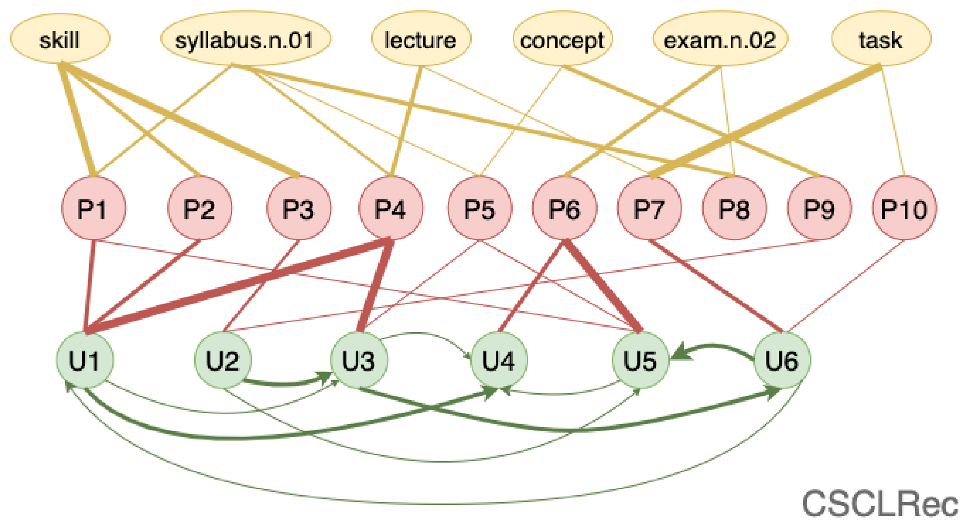

Figure 3.4: CSCLRec's PPR module

temporal decay rate to the weight of user-to-post edges in the graph. The earlier the interaction takes place, the lower the weight of the edge. This feature enables the recommendation algorithm to emphasize the user's recent interests.

To perform recommendation, the PPR algorithm runs a random walk procedure (Haveliwala, 2003). The student who receives recommendations is referred to as the "target student" for the current recommendation task. The algorithm begins with setting the target user node and its connecting post nodes as the "restart nodes". A random walk agent is initialized to start from the target user node, follow the edges, and traverse over the graph. During each iteration of node transitions, the agent's walking pattern is controlled by the damping factor α: the random walk agent has a fixed probability of α of transiting to its adjacent nodes, and another $(1 - \alpha)$ probability of teleporting back to the restart nodes. After sufficient iterations are completed, the probability of visiting each node will converge to a stable state. The most visited post nodes are the recommendations calculated by the system. In practice, I used power iteration to approximate the stationary probabilities, such that the computing-intensive random walk simulations are avoided (Hu, Koren, & Volinsky, 2008).

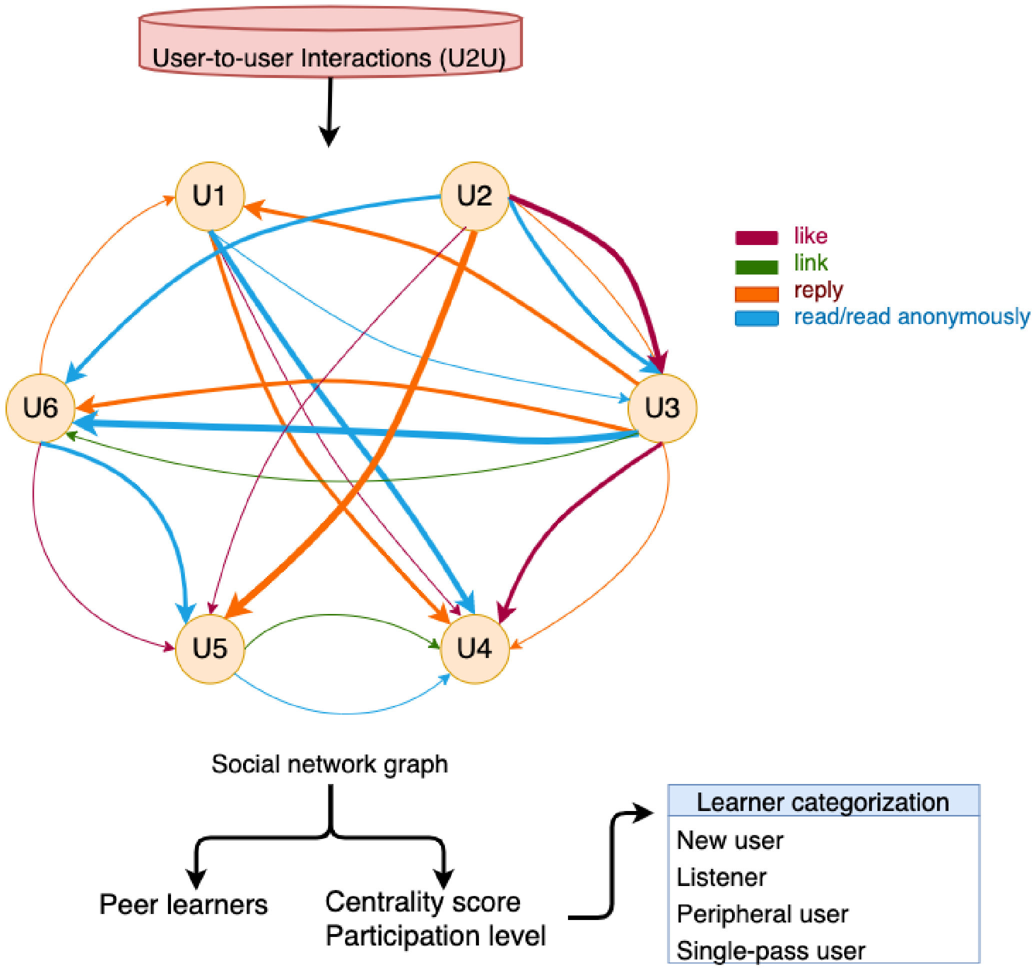

Figure 3.5: The workflow of the learner interaction profiler. Edge widths in the graph indicate the number of interactions.

3.3.3 Learner Interaction Profiler Module

The learner interaction profiler integrates pedagogical considerations. The module analyzes interactions among learners in a manner similar to social network analysis and brings together learners who share similar interests. Meanwhile, it also profiles user behaviors so that customized recommendation strategies can be delivered to accommodate students of each behavior type. The overall processing flow of this module is shown in Figure 3.5.

Internally, this module adopts a bidirectional graph (Wasserman & Faust, 1994) that consists of different types of U2U interactions (e.g., replies, likes, or reads). This research refers to the graph as the social network graph. This graph takes all users as nodes and represents the interactions as edges. For example, in Figure 3.5, the red edge from user U2 to user U3 indicates that U2 liked U3's post.

The first function is the identification of peer learners. The module generates edges that connect the target user to their peer learners, as shown by the green edges in Figure 3.3. These edges are then used to determine who the target user's peer learners are. For each target user, other students who have the most interactions with them are considered to be their peer learners. For example, U3 is U2's peer learner (Figure 3.3) because U3 is the one who had the most interactions with U2 in Figure 3.5. The determination of peer learners is uni-directional. Therefore, U2 is not necessarily U3's peer learner. In this study, the number of peer learners is a hyperparameter that is proportional to the total number of students in the forum. Because the cognitive conflicts triggered by negotiations and debates are essential for students to think and learn from each other's perspective (Fosnot & Perry, 1996), the set of peer learners are not limited to those who share similar perspectives. This set may also include those who have different opinions.

Peer learners are identified as the ones who have the most number of edges connecting to the target user in the social network graph. To determine this number, a threshold is used. To simplify the later evaluations, I fixed the number of peer learners for each target user. The value was determined through a tuning process from a list of potential values. Once deployed, this number can be determined through instructor configuration.

The learner interaction profiler module also measures two metrics for each student through analysis of the social network graph: the participation level and the degree of centrality. Both metrics provide information about the user's level of integration within the learning community. The participation level of a user is the number of outgoing edges of reply, like, and link types from their user node. It reflects how actively the student engages in the discussion. The degree of centrality of a user is the in-degree (the number of incoming edges in the social network graph) of their user node. Therefore, the target user's degree of centrality increases as other students frequently interact with the target user's posts. It reflects the importance of a user to others in the community.

Four types of learners that need differentiated recommendation strategies are identified through the two above measures.

New user

In the context of a socio-collaborative learning community, new users are learners who are new to the discussion. As they just joined the discussion, they do not have any logged interactions thus recommendations for these users will be affected by the cold-start problem. As a consequence, the recommender system will have trouble inferring the user's interest and provide suggestions because the lack of data prevents recommendation (Bobadilla et al., 2012). Our analyses have found that in the first two weeks of each semester, the cold start problem is the case for most users because students are free to enroll and withdraw from courses. Moreover, from the perspective of individual user needs, new users can suffer from greater information overload because they face many posts at once. They may have a greater need for tailored recommendations to help them filter information. This strategy could also help identify new user's interests early so that they can contribute their own voice.

In terms of recommendation strategies, rather than being linked to a fixed number of peer learners, new users are connected to every other user in the community in the PPR graph. The action builds up the new user's relationship to everyone else, thereby making it easy to access a variety of information.

Listener

The module categorizes users who read posts but who rarely post themselves as listeners (Wise, Hausknecht, & Zhao, 2014). As everyone is expected to be both a consumer and a producer of resources, if everyone is benefiting from the resources and not contributing any resources, then there are no resources to share and recommend (Guy, 2015). In addition, to foster forum activity and promote knowledge co-construction, it is critical to apply the knowledge building principle of collective responsibility and symmetric knowledge advancement, which suggests a need to encourage posting (Scardamalia et al., 2002).

To reduce the number of listeners, I would like to help these learners explore different topics, so that they are more likely to express opinions and the community's overall sustainability can be promoted (Kleanthous & Dimitrova, 2010). We, therefore, adopt the same recommendation strategy as that employed for new users.

Peripheral user

Peripheral users are those who have recently lost interactivity in their readership. In contrast to central users, the fewer other users who actively interact with the target user's posts, the more peripheral the target user is. The idea of detecting peripheral users comes from work by Kleanthous and Dimitrova (2008), Kleanthous and Dimitrova (2009). The authors believed that a growing number of such users can break the knowledge sharing chain and lead to reduced forum activity. As a solution, they suggested early detection and intervention.

Therefore, with the goal of promoting the knowledge sharing process, the learner profiler module automatically detects peripheral users and applies a special recommendation strategy to bring them to a more central position. I define U2U interactions of type reply, like, or link as active interactions as they actively create opportunities for further communication and express opinions explicitly, compared with other interaction types (i.e., read, read anonymously).

For each target user, the learner profiler monitors the number of interactive peers who had this type of active interaction with the target user. Prior to every round of recommendations, if the number has halved from the previous week, the target user will be marked as a peripheral user. To accommodate peripheral users, the module introduces a connection between the peripheral user and their lost interactive peers in the PPR graph to enhance these connections. In this way, they would be more likely to get to know each other's views. Some common interests may be discovered by both parties and stimulate more discussion.

Single-pass user

Single-pass refers to a behavior where someone only reads new posts once and ignores older posts. In a study conducted by Hewitt (2005), he argued that this type of behavior is one of the main causes of discussion threads terminating early. As socio-collaborative learning processes require topics to be progressively discussed and deepened (Scardamalia et al., 2002), it is suggested to prevent this behavior by encouraging students to revisit earlier posts (Hewitt, 2005). Inspired by this idea, the learner interaction profiler identifies students who have only read posts from the previous week. For example, those who have not read posts created before week 3 are marked as single-pass users in week 4. Once they start to follow the guidance to read old posts again, they are removed from the list of single-pass users.

Our measure decreases the temporal decay on earlier posts for single-pass users. As a result, earlier posts are down-weighted less and have larger probabilities to be recommended to these learners. The threshold number of old posts that categorize a user as a single-pass user can be determined by instructors given the course regulations, size, or other contextual factors. In the case of our evaluations in later chapters, the value is set at 0 such that any person who did not read old posts is considered a single-pass user.

3.3.4 Content Analyzer Module

The content analyzer offers a recommendation solution from a semantic perspective such that content of forum posts is taken into consideration. The design is motivated by the fact that recommender systems based on interaction records (i.e., collaborative filtering, personalized PageRank graph) could bias towards popularity (Steck, 2011). For example, popular posts get recommended more often; while others, especially those made by students who possess minority opinions, are left unnoticed. This contradicts the goal of a small socio-collaborative learning community where every voice needs to be heard and students need to be exposed to diverse perspectives (Scardamalia et al., 2002). It can also lead to reduced forum activity (Cremonesi, Koren, & Turrin, 2010).

Moreover, as forum thread structure is hierarchical, posts on the same topic tend to have similar interaction records. Therefore, pure graph-based or collaborative filtering may also cause a filter bubble problem: only posts with similar locations that users often interact with will be included in the recommendation list (T. T. Nguyen et al., 2014). This problem impairs students' learning experience because there may be posts that match their current interests that are located in another topic structure.

In the content analyzer module of both systems, forum post contents are processed and analyzed to identify the relationships between forum posts. As a result, the module generates post-to-hypernym or post-to-keyword edges to be used in the PPR graph. Which of these is used depends on the specific recommender algorithm.

CoPPR

As shown in Figure 3.6, CoPPR's content analyzer module involves two key steps: preprocessing of the extracted post content and keyword extraction for each post. Specifically, after the extraction of post content, a sequence of regular text preprocessing techniques is applied: First, texts are tokenized to identify individual words. The resulting tokens are converted to their root

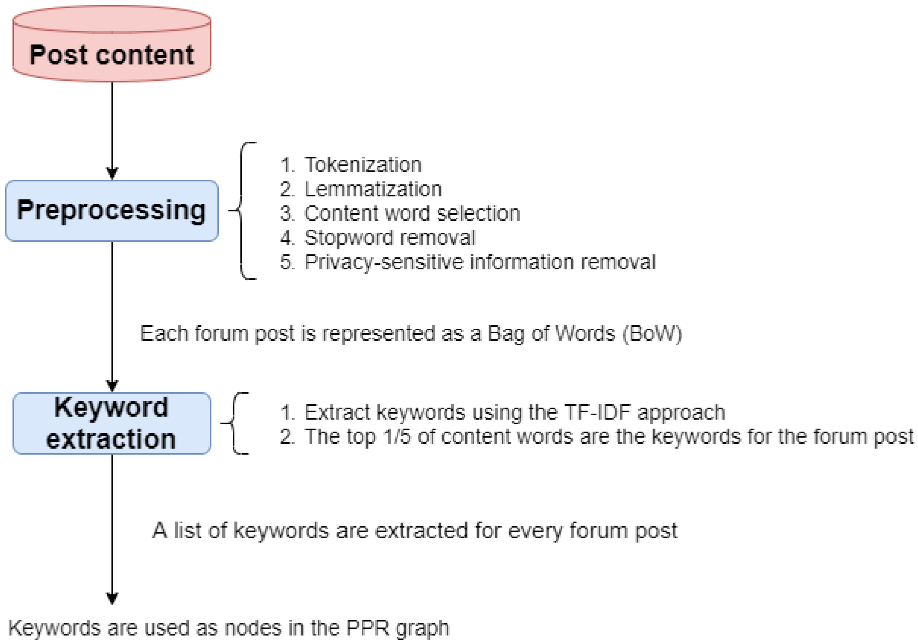

Figure 3.6: The workflow of CoPPR's content analyzer module.

forms in the lemmatization step. Then, according to a customized stopword list, frequently used words are deleted because they often do not affect the meaning of the text. The content analyzer only keeps the nouns and verbs. In the end, privacy-sensitive information is also removed (i.e., email address, names, web URLs).

After the preprocessing steps, each forum post is expressed as a bag-of-words (BoW). For example, a forum post that contains "Here is a description of our course. Please review the syllabus before coming to lectures" is represented as "description", "course", "review", "syllabus", "lecture".

In the keyword extraction step that follows, the module chooses keywords that can best differentiate the current post from others. It applies the term frequency-inverse document frequency (TF-IDF) technique to identify these words. After the TF-IDF scores have been computed for all words in a post's BoW, the ones with the top 1/5 scores are selected as the keywords to represent the post. Though the cutoff ratio is fixed at 1/5 in our setup, the value could be flexibly determined by instructors in real classroom uses.

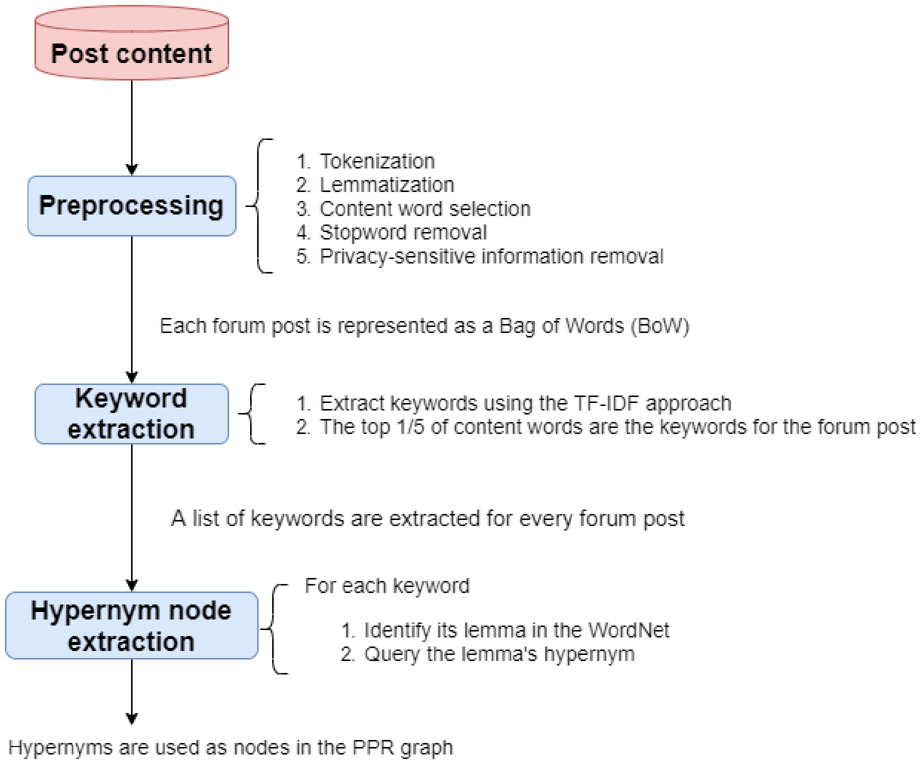

Figure 3.7: The workflow of CSCLRec's content analyzer module.

The extracted keywords are used as nodes in the PPR graph, as shown in Figure 3.3. The graph connects a post node to a keyword node by drawing a post-to-keyword edge if the post contains the keyword. The weight of the edge is proportional to the TF-IDF score of the keyword in the post's BoW.

CSCLRec

Although the content analyzer used by CSCLRec and CoPPR has many similarities, CSCLRec employs a different method to identify the key concepts in the post. CSCLRec's content analyzer goes through the same preprocessing and keyword extraction steps. Different from CoPPR, the output of CSCLRec's content analyzer module is a collection of hypernym nodes.

As shown in Figure 3.7, the module uses a semantic network, WordNet, for hypernym look-up and textual similarity measurement (Miller, 1995). The Lesk algorithm (Lesk, 1986) was applied to disambiguate the lemmas that correspond to the extracted keywords with help from the post content as a context. Then, through WordNet's collection of hypernyms, it fetches the highest level hypernyms of the extracted keywords. Similar to CoPPR, these extracted concepts are then used: they are the hypernym nodes in the PPR graph. When a post contains a keyword that can lead to this hypernym, the module draws an edge to connect the post node to the hypernym node. This hypernym-based design recommends posts that are thematically similar, even if the authors express opposite views, provided they mentioned similar concepts, which gives some posts more possibility to be connected. One benefit of the hypernym-based approach is that it can recommend posts that contain opposite views provided they mentioned similar concepts, which could potentially enable the creation of cognitive conflict and provide more opportunities for idea improvement and symmetric knowledge building.

3.3.5 Post Filtering Module

One limitation of the content analyzer module is that it does not take into consideration whether the recommended posts are sufficiently informative to advance the knowledge of the community. To compensate for this, I introduce the post filtering module as the guard to analyze, sort, filter, and rerank the recommendations produced by the PPR graph, before the generation of the final list. Without taking this step, posts such as "Thank you for the clarification, [name]" are likely to be suggested to students who have made similar posts. The post filtering module excludes these types of posts from the recommendation results using a combination of two filters:

- The first filter builds trigram models from all post contents to identify noun and verb phrases and extract them from each post. The module excludes posts with fewer than 3 key phrases.
- The second filter compares post content with the Academic Word List (AWL) (Coxhead, 2016). Posts with fewer than 3 AWL words are excluded.

Our decision of fixing the threshold at 3 key phrases and AWL words was determined through observing the sample results from a separate course that was not used for evaluation. Posts that have fewer than 3 keywords or AWL words usually did not contain enough information.

3.4 Conclusion

In this chapter, I described PeppeR, an educational platform that our recommender systems were originally designed to work with. Two recommender system algorithms, CSCLRec and CoPPR, that were designed to support learning in small asynchronous discussion forums were introduced. These algorithms mine learner interaction data using both modelled learner types and natural language processing techniques. Grounded in a modified graph-based recommendation algorithm, the design of these algorithms also accounts for learning theories. In the next chapter, using PeppeR data as the testbed, I describe the evaluations that examine how the proposed recommender systems perform in discussion forums.

Chapter 4

An Offline Experiment Using Historical Data

4.1 Introduction

In this chapter, I describe an experimental study that evaluates the proposed recommenders using historical data from PeppeR. The following research question is posed for the study: *How do recommender systems that incorporate principles from socio-collaborative learning (i.e., CSCLRec and CoPPR) perform in comparison with traditional recommenders when suggesting posts in small discussion forums?* To answer this question, I measure the predictive performance of CSCLRec and CoPPR, along with six other recommenders commonly used in educational contexts, through four quantitative indicators. In the following sections, I describe the design of this experiment and compare the recommender performance results.

4.2 Methods

In this experiment, the task is to provide forum post recommendations to students in PeppeR. At the end of each week, each student receives a customized list of recommendations that contains 10 posts. The student who receives recommendations is referred to as the "target student" for the current recommendation task. For simplicity, the weeks before and including the current week are called the "training weeks". Similarly, the first week after the current week is the "test week".

4.2.1 Data Preparation

Data from 6 courses from PeppeR were extracted for this study. Only fully online courses were chosen because they keep the full log of learning activities so that learner behaviours can be captured in detail. Of the six test courses, three were regular-length courses (13-weeks long; a regular term) and the others were short courses (6-weeks long; a summer term).

I extracted all historical data which includes forum posts, students and instructors, and all of their interactions with posts (i.e., posting, replying to others' posts, inserting hyperlinks to other people's posts, liking posts, and reading posts). The interactions were processed and categorized into 7 types: create, reply, like, link, revisit, read, and anonymously read. Post content was further processed into plain text. User activity statistics for each course are summarized in Table 4.1 and Table 4.2.

Each course consisted of fewer than 30 students, as shown in Table 4.1. The first three (LA, LB, LC) were 13-weeks long courses, and the last three (SA, SB, SC) were short courses (6-weeks long). The large standard deviation seen in the five columns indicates considerable variability in student interactions that is consistent with the different types of users identified and targeted: Some had a large number of forum activities, while the others seldom interacted with posts. This suggests the necessity of distinguishing different learner types and employing user-specific recommendation strategies. Figure 4.1 illustrates the number and distribution of different types of learners in the selected test courses. It can be noted that as long courses progress, the number of new users and listeners will decrease due to the gradual participation of students. In short courses, these numbers are 0 from the beginning. This may be because the short courses are intensive and fast-paced, so students often need to speak at an early stage (week 2).

Course	Weeks (#)	Students (#)	Posts (#)
LA	13	26	1751
LB	13	19	809
LC	13	30	2090
SA	6	23	627
SB	6	24	1142
SC	6	20	869

Table 4.1: Course length and participation information of the 6 selected sample courses from PeppeR.

Course	Interacted posts /student	Interactions /student	Reads /student	Likes student	Links /student
LA	1176 (550.2)	1628 (1010.3)	1314 (719.2)	76 (61.2)	1 (3.0)
LB	358 (246.0)	441 (298.8)	365 (247.0)	21 (25.0)	0.05 (0.2)
LC	1212 (686.4)	1373 (732.4)	1226 (698.9)	29 (23.9)	10 (15.2)
SA	362 (219.1)	505 (417.6)	405 (290.3)	15 (19.5)	0.26 (1.3)
SB	616 (269.9)	731 (281.9)	635 (270.1)	8 (10.0)	0.25 (1.0)
SC	507 (223.7)	631 (269.1)	521 (223.4)	44 (44.6)	0.55 (1.6)

Table 4.2: Student forum interactions as *M (SD)*.

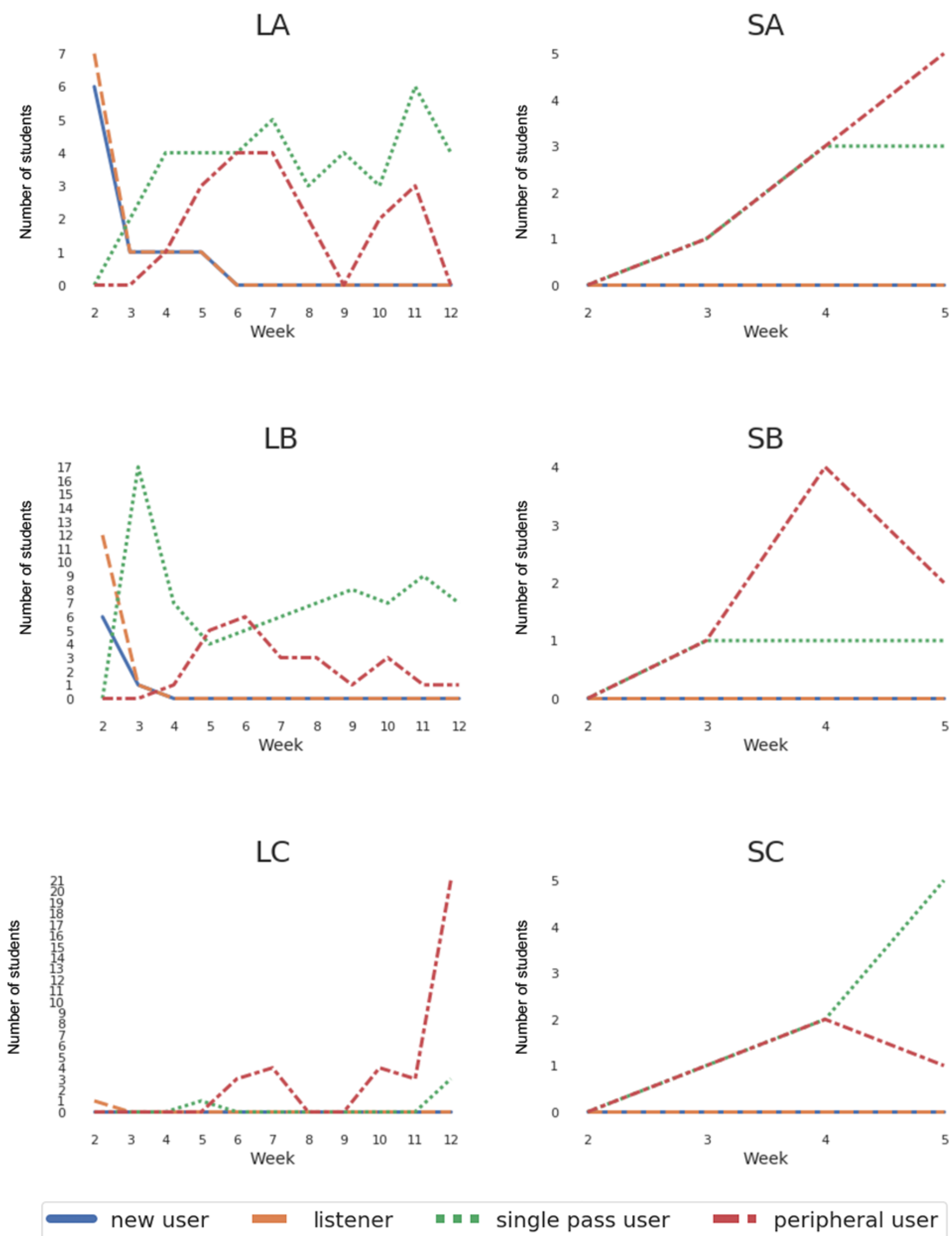

Figure 4.1: Distributions of the 4 learner behavior types within each course. Note that the first week (week 1) and last week (week 13 or week 6) are not included as the evaluations were not performed for these weeks.

4.2.2 Competing Recommenders

To examine the performance of CSCLRec and CoPPR, I compared them against 6 other recommenders that have been used in similar scenarios.

- Matrix factorization collaborative filtering (MCF) is the most popular model-based collaborative filtering algorithm in the educational recommender system domain (Drachsler et al., 2015). To deal with the unary positive-only feedback in our student interaction log, the specific version of the algorithm proposed by Hu, Koren, and Volinsky (2008) was applied. The algorithm exerts a confidence factor on implicit feedback such that unseen interactions can be considered as dislikes with lower confidence. An alternating least square approximation process is used to speed up the optimization. The confidence level that specifies the negative weight attributed to unseen interactions was tuned in the validation step.

- Keyword-based content-based recommender (KCB) uses TF-IDF to extract keywords from post content and encode posts into a vectorized form. After performing dimension reduction using the Latent Semantic Indexing (LSI) algorithm, a fixed-length vector is used to represent the post. Users are represented as the average of the post vectors they have interacted with before. At the end, KCB recommends candidate posts which are nearest to the target user in the vector space.

- Sentence embedding recommender (SCB) implements another type of content based system. It relies on the semantics of post content. It differs from KCB in that it directly encodes the entire post into vectors. The Universal Sentence Encoder was used to perform this step (Cer et al., 2018). The Universal Sentence Encoder is a deep learning-based text embedding model that works for sentences and paragraphs. It offers a few pre-trained models to fine-tune or directly apply to downstream tasks. Due to the limited training corpus in PeppeR, pretrained embeddings were used; thus, this system does not have hyperparameters to control.

- Personalized PageRank (PPR) incorporates a bipartite graph that only has user nodes and post nodes. There are edges connecting user nodes to post nodes that represent historical interactions. The edge weights are determined by the number of interactions in the past. Similar to CSCLRec and CoPPR, a random walk-based PageRank algorithm is applied when this system produces recommendations. Target user nodes were used as the restart nodes, and the node traversing procedure depends on the edge weights and a damping factor that controls the probability of teleporting to the restart node.

- The popularity-based recommender system (PPL) delivers recommendations based on the popularity of posts. Every user receives the same recommendations.

- The random recommender system (RND) randomly draws recommendations from the candidate list.

It should be admitted that the above list did not extensively cover all recommendation methods. There are still some well-known recommendation algorithms that have not been used on our data. The reason for this is that some structural and design particularities mean they do not inherently fit with the nature of our dataset. For example, the unary positive-only user feedback in our data presents challenges since topic modeling methods (i.e., LDA (Blei, Ng, & Jordan, 2003)) do not work well with short forum texts. Moreover, in a course forum that discusses class materials, topics are usually fixed to a limited domain. Therefore, as I found in earlier experiments, topic modeling methods tend to have difficulty discovering latent topics that can describe differences among posts. Deep learning based methods are also easily overfit due to the size of our dataset, and they lack transparency which could keep learners from trusting the provided recommendations. For both of these reasons, they were not used.

4.2.3 Recommender Training

I followed the evaluation protocol suggested by previous studies (Erdt, Fernandez, & Rensing, 2015; Shani & Gunawardana, 2011) in this experiment. As shown in the upper part of Figure 4.2, the recommender systems were built using all forum activities during the training weeks (blue and orange blocks) and the activities performed by everyone other than the target user in the test week (sketched yellow block). In other words, the target user's activities in the test week are hidden from training and are used as the ground truth for prediction.

As I note that all recommender systems come with hyperparameters, to control the consistency and reproducibility of the experiments, a validation step is conducted before the build-up of recommenders to determine the optimal value of hyperparameters. The selection of hyperparameters and their ranges is included in Table 4.3. However, due to the limited number of students, the diversity of student interaction behaviors, and the inter-dependence of time-series data adding the constraint of not being able to use future interactions to validate past interactions, random cross validation is not appropriate in this context. As an alternative, a last block validation procedure (Bergmeir & Benítez, 2012) was employed: For each weekly evaluation, the week prior to the test week is called the tuning week (week 7 in Figure 4.2). For each recommendation approach, I grid-searched a list of hyperparameters and trained the systems using the data from weeks before the tuning week (blue blocks in Figure 4.2). Then, I tested systems using data from the tuning week (orange block for the target user in Figure 4.2). The best performing hyperparameters in recommendation precision were then selected to build the recommenders for subsequent evaluations of the tuned recommenders.

Method	Hyperparameters	Values
CSCLRec	Temporal ratio applied to older posts (= 1-temporal decay)	[0.7, 0.75, 0.8]
	Number of peer learners for each target user to be connected in the PPR graph	[1/3, 1/5, 1/7, 1/10] of total student number
	Damping factor for the PPR	[0.8, 0.85, 0.9]
	Number of keywords extracted from each post	1/3 of the word count
	Cutoff threshold of AWL words to remove from recommendations	3
	Cutoff threshold of key content words to remove from recommendations	3
	Temporal decay for single-pass users	0.9
CoPPR	Temporal ratio applied to older posts (= 1-temporal decay)	[0.7, 0.75, 0.8]
	Number of peer learners for each learner to be connected in the PPR graph	[1/3, 1/5, 1/7, 1/10] of total student number
	Damping factor for the PPR	[0.8, 0.85, 0.9]
	Number of keywords extracted from each post	1/3 of the word count
	Cutoff threshold of AWL words to remove from recommendations	3
	Cutoff threshold of key content words to remove from recommendations	3
	Temporal decay for single-pass users	0.9
PPR	Damping factor for the PPR graph	[0.75, 0.8, 0.85]
MCF	A scaling term to represent the confidence exerted to implicit feedback	[15, 20, 25, 30]
KCB	Word vector dimension size	[10, 15, 20]
	Top 1/n content words being used as keywords to represent the post	[3, 5]

Table 4.3: List of hyperparameters chosen for the offline evaluations.

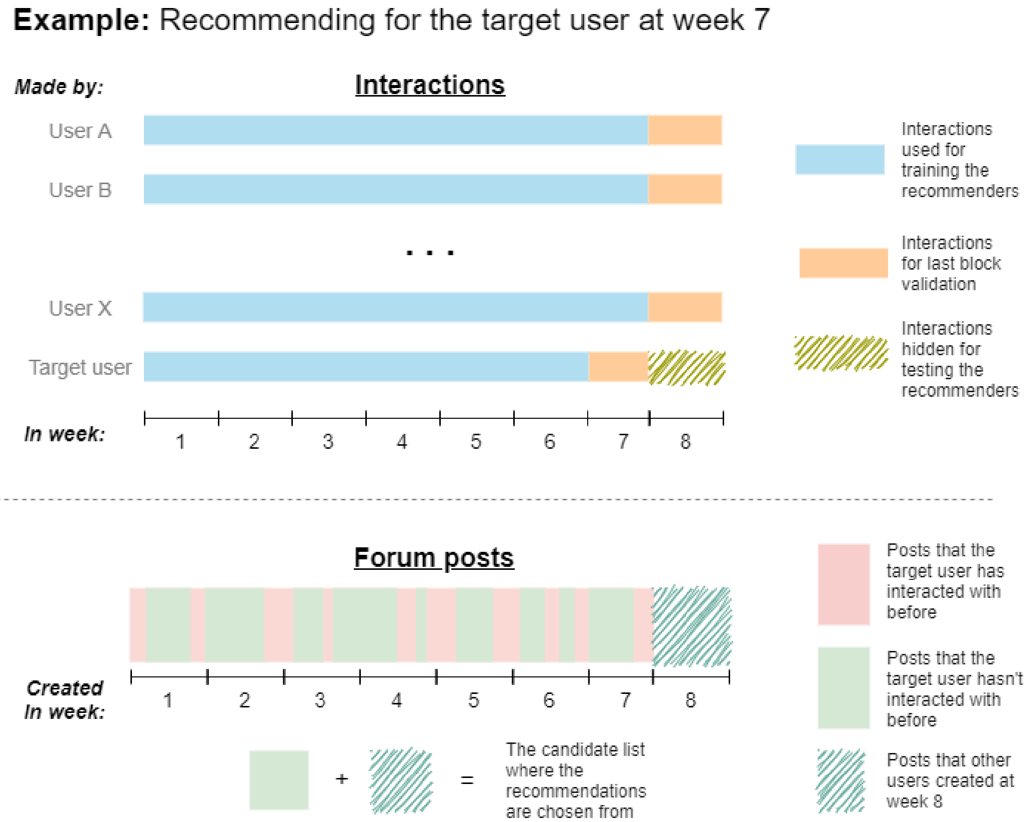

Figure 4.2: Data splitting procedure: illustrated through an example. The recommendation is performed at the end of week 7 (test week = 8).

4.2.4 Recommender Evaluations

The recommenders were tested against the target user's behaviours during the test week (sketched yellow block in the upper figure of Figure 4.2). Each target user is expected to receive 10 recommendations. As shown in the bottom of Figure 4.2, the recommendations were drawn from a candidate list that includes the posts that this target user has not interacted with during the training weeks (solid green blocks) and the posts created by other users in the test week (sketched green block).

The evaluation measures the recommendation performance based on the match of recommended posts with what the target user has actually interacted with in the test week. Since the forum had few resources that can be used for recommendation at the beginning of each course, the evaluation procedure starts from the end of week 2. In addition to the evaluation across course types, I also examined recommender performance across different user types.

As research indicated that the quality of educational recommender systems is not only determined by accuracy (Fazeli et al., 2017), I measured 3 dimensions of performance: accuracy (precision and recall), diversity, and novelty. Since I am going to present the formulae for calculating these metrics, the formula notations are displayed here.

1. $\mathcal{R}$ represents the set of post recommendations;
2. $|\mathcal{R}|$ represents the size of set $\mathcal{R}$;
3. $\mathcal{I}_A$ represents the set of posts that the user **A** actually interacted with in the test week;
4. p_a, p_b represents post **a** and post **b**;
5. $\overrightarrow{p_a}, \overrightarrow{p_b}$ denotes the vectorized embeddings of posts p_a and p_b;
6. $\text{cd}(\overrightarrow{u}, \overrightarrow{v})$ represents the cosine distance between vectors u and v, such that

$$\text{cd}(\overrightarrow{u}, \overrightarrow{v}) = 1 - \frac{\overrightarrow{u}\,\overrightarrow{v}}{\|\overrightarrow{u}\|\|\overrightarrow{v}\|};$$

7. $\mathcal{U}_p$ represents the set of users who have interacted with the post p before;
8. $\mathcal{U}$ represents all users.

To measure accuracy, this study reports both the Precision at K (P@K see Formula 4.1) and the Recall at K (R@K see Formula 4.2) where K is the number of recommendations. Precision is the most important measure in terms of our context, because it reflects how well the recommender systems predict user preference given their past interactions in offline evaluations. A higher precision value means that the system is more likely to reduce the time it takes for users to find materials of interest, making them more focused on learning activities. Additionally, the R@K measure is not always useful since the number of recommendations is fixed at 10 in our evaluations. For example, when the target user actually interacted with fewer than 10 posts, the R@10 will never reach 1 even if all of them have been successfully recommended. Therefore, I also listed the maximum R@10 as reference in the table. The maximum R@10 is also reported to reflect the relative performance of the recommenders.

$$\text{Precision} = \frac{|\mathcal{R} \cap \mathcal{I}_A|}{|\mathcal{R}|} \tag{4.1}$$

$$\text{Recall} = \frac{|\mathcal{R} \cap \mathcal{I}_A|}{|\mathcal{I}_A|} \tag{4.2}$$

For diversity, this study used intra-list diversity (ILD) which measures the average pairwise distance between recommended items (Castells, Hurley, & Vargas, 2015; Smyth & McClave, 2001), see Formula 4.3. As forum posts were encoded in pre-trained Universal Sentence Encoder embeddings, the intra-list distance is the average pairwise cosine distance between all recommended post embeddings in the vectorized embedding space. In general, a higher cosine distance is preferred because it means that the textual similarity of the recommended posts is lower, which indicates a more diverse recommendation list. Diverse recommendations help students get in touch with all the required knowledge of the course, instead of just focusing on one topic they talk about most often. However, since diversity is a trade-off to accuracy, an overly-diverse recommendation result could make the recommender no different from a random recommender, which does not deliver personalization.

$$\text{ILD} = \frac{\sum\limits_{p_a, p_b \in \mathcal{R}} \text{cd}(\overrightarrow{p_a}, \overrightarrow{p_b})}{(|\mathcal{R}|(|\mathcal{R}| - 1))/2} \tag{4.3}$$

Mean inverse user frequency (MIUF) is used to measure recommendation novelty (Breese, Heckerman, & Kadie, 1998), see Formula 4.4. The measure exhibits the same notion of inverse user frequency: The fewer people who have interacted with the post, the higher the IUF of the post, thus indicating higher novelty. Novel recommendations could bring new content to the target student, which has the potential to help them discover new interests or expand their world view. It can also help them discover less popular posts, and the student who created such posts could benefit from gaining greater visibility. Sharing these posts with different perspectives to more people can also benefit the entire community. Potential benefits include stimulating more discussions or inspiring more diverse perspectives of thinking. However, similar to diversity, novelty is not the main focus of the proposed recommender systems because novel recommendations may not be consistent with the learner's background knowledge and may not be able to help them build knowledge.

$$\text{MIUF} = -\frac{1}{|\mathcal{R}|}\sum_{p\in\mathcal{R}}\log_2\frac{|\mathcal{U}_p|}{|\mathcal{U}|} \tag{4.4}$$

4.3 Results

4.3.1 Recommender Performance for Short and Long Courses

Evaluation results across course types are shown in Table 4.4. As the recommendation to each individual student every week is treated as a sample point, in total 825 samples were observed for the long courses and 268 samples were observed for the short courses. For the simplicity of displaying the results, results for the same course type are combined. To reflect the consistency of algorithm performance comparison, the mean and standard deviation are both reported. The results were averaged over the results of each individual student every week. At the same time, I conducted a statistical test on each measure for each course type. For each case, a recommender's weekly result is treated as a sample.

	Long courses (LA, LB, LC)			
Algorithm	**P@10**	**R@10**	**ILD**	**MIUF**
CSCLRec	**0.73 (0.319)**	0.22 (0.305)	0.27 (0.125)	0.61 (0.360)
CoPPR	**0.72 (0.324)**	0.22 (0.304)	0.22 (0.110)	0.64 (0.380)
PPR	0.54 (0.408)	0.18 (0.310)	0.39 (0.162)	0.47 (0.251)
MCF	0.48 (0.391)	0.18 (0.313)	0.45 (0.192)	0.84 (0.532)
SCB	0.30 (0.355)	0.16 (0.313)	0.08 (0.047)	**1.22 (0.453)**
KCB	0.29 (0.359)	0.15 (0.315)	0.22 (0.105)	1.05 (0.490)
RND	0.31 (0.335)	0.16 (0.312)	0.41 (0.174)	**1.17 (0.404)**
PPL	0.41 (0.407)	0.18 (0.310)	0.42 (0.164)	0.42 (0.243)
	Short courses (SA, SB, SC)			
Algorithm	**P@10**	**R@10**	**ILD**	**MIUF**
CSCLRec	**0.75 (0.310)**	0.19 (0.254)	0.19 (0.059)	0.48 (0.140)
CoPPR	**0.73 (0.315)**	0.18 (0.243)	0.156 (0.048)	0.50 (0.137)
PPR	0.57 (0.383)	0.14 (0.248)	0.24 (0.106)	0.41 (0.144)
MCF	0.45 (0.406)	0.13 (0.265)	**0.36 (0.151)**	0.80 (0.485)
SCB	0.40 (0.378)	0.12 (0.247)	0.08 (0.019)	0.92 (0.251)
KCB	0.40 (0.369)	0.12 (0.247)	0.19 (0.072)	**1.04 (0.311)**
RND	0.35 (0.336)	0.11 (0.248)	**0.35 (0.130)**	**1.20 (0.385)**
PPL	0.48 (0.402)	0.14 (0.249)	**0.31 (0.136)**	0.35 (0.135)

Table 4.4: Summary of evaluation results - Short courses and Long courses. The measures are as *M (SD)*. The best performers in precision measure are bolded. Average max R@10 in long courses: 0.351 (0.343), in short courses: 0.303 (0.317). Sample size in long courses: 825, in short courses: 268.

Course Type	**Metric**	***F*-stat**	***p*-value**	η^2
Short course	P@10	$F(21,64) = 0.734$	0.782	0.035757
Short course	R@10	$F(21,64) = 0.237$	1.000	0.042683
Short course	ILD	$F(21,64) = 0.321$	0.997	0.022624
Short course	MIUF	$F(21,64) = 0.321$	0.997	0.015522
Long course	P@10	$F(21, 176) = 0.738$	0.927	0.056534
Long course	R@10	$F(21, 176) = 0.062$	1.000	0.020983
Long course	ILD	$F(21, 174) = 0.213$	1.000	0.029796
Long course	MIUF	$F(21, 176) = 0.225$	1.000	0.039168

Table 4.5: ANOVA Result: Week × Algorithm interaction effects.

Course Type	**Metric**	***F*-stat**	***p*-value**	η^2
Short course	P@10	$F(3,64) = 2.982$	0.038	0.020892
Short course	R@10	$F(3,64) = 1.923$	0.135	0.048780
Short course	ILD	$F(3,64) = 0.251$	0.860	0.002715
Short course	MIUF	$F(10,64) = 2.031$	0.118	0.014067
Long course	P@10	$F(10, 176) = 9.001$	0.000	0.098492
Long course	R@10	$F(10, 176) = 1.946$	0.042	0.093654
Long course	ILD	$F(10, 174) = 0.998$	0.447	0.019864
Long course	MIUF	$F(10, 176) = 1.677$	0.089	0.041618

Table 4.6: ANOVA Result: Main effect of Week.

A two-way ANOVA was first conducted to check whether there is an interaction effect between the week and the choice of recommender system. Results show that there was no significant interaction effect between the two factors in all cases (Table 4.5). They also indicate that week was not a significant factor that affected recommender performance (Table 4.6), which implies that there is no measurable difference in recommender performance from one week to the next.

Course Type	**Metric**	***F*-stat**	***p*-value**	η^2
Short course	P@10	$F(7,64) = 48.724$	0.000	0.794295
Short course	R@10	$F(7,64) = 7.089$	0.000	0.396341
Short course	ILD	$F(7,64) = 31.677$	0.000	0.756561
Short course	MIUF	$F(7,64) = 50.913$	0.000	0.822662
Long course	P@10	$F(7, 176) = 85.167$	0.000	0.652372
Long course	R@10	$F(7, 176) = 1.119$	0.353	0.037615
Long course	ILD	$F(7, 174) = 43.305$	0.000	0.603717
Long course	MIUF	$F(7, 176) = 27.761$	0.000	0.482354

Table 4.7: ANOVA Result: Main effect of Recommender Algorithm.

On the other hand, recommender algorithm performance varied significantly (Table 4.7). Therefore, Tukey's HSD post-hoc analysis was performed to compare each pair of recommenders. The resulting winner recommender system, in each case, is bolded in Table 4.4. Detailed significance test results are provided in Appendix A.

As shown in Table 4.4, CSCLRec and CoPPR outperform all competitors according to the P@10 by more than 17% in both types of courses; PPR places third.

Outstanding performance can also be observed in R@10. Considering the maximum possible recall is capped at 35.1% and 30.3%, CSCLRec and CoPPR's R@10 performance in both courses suggest their successful retrieval of most relevant items. However, the post-hoc analysis does not show significant differences among the performance of recommender systems because the variability among CSCLRec, CoPPR, and the other algorithms' R@10 performance was large.

Recommendation algorithms that emphasize interactions (MCF and PPR) had the best diversity among the personalized approaches. Apart from them, the two baseline recommenders (RND and PPL) perform very well in ILD. Content-based systems (KCB and SCB) have the lowest recommendation diversity scores (always placed last or second last in the pool of 8 recommenders). Though CSCLRec and CoPPR's performance in ILD is in the middle (3rd and 4th out of 6 personalized recommenders under both scenarios), their performance is still acceptable as diversity is a tradeoff to accuracy. This is illustrated through Table 4.8, where I displayed a recommendation list presented to a sample target student. The sample course mainly focused on the issue of anti-oppression education. Within this round of recommendation, the target user had records of recent interactions with Reading 1, 2, and 4. As can be seen through the example in the table, CSCLRec generated diverse recommendations around all 3 readings, instead of focusing on one of them. The recommendations covered both discussions of these reading materials and build-on comments within those discussions. Meanwhile, CSCLRec produced a novel recommendation (Item #6) that is from Reading 3 that the user had not previously focused on. This post had a relatively new concept that emphasized the importance of gay-straight alliance and the role of teachers in educating students about anti-homophobia or anti-bullying issues. This recommendation still fits the user's background because it was situated in the scope of anti-homophobia education.

According to MIUF, content-based algorithms (KCB and SCB) are the best performers apart from the random recommender system, followed by MCF. As another tradeoff to high precision, novelty is not best achieved with CoPPR or CSCLRec though they still beat the traditional graph-based system (PPR). PPL is ranked at the bottom due to the fact that it recommends based on post popularity which is inherently in contrast to recommendation novelty.

4.3.2 Recommender Performance Over Time

To illustrate differences in recommender performance over time, I show examples from the LA and SA courses in Figure 4.3 and Figure 4.4 respectively. In both figures, week 1 is not included because the evaluation starts from week 2. Final weeks (week 13 or week 6) are also not included because no resources were generated after the courses end.

Similar patterns to those seen in the LA and SA courses were present in other courses. In general, CSCLRec and CoPPR remained the best performing recommender systems for precision and recall throughout the semester.

Another pattern visible in Figure 4.3 is the performance of recommenders in the first two weeks of both courses. Our proposed algorithm, CSCLRec, beats its competitors in both precision and recall, and it continues this trend until near the end of the course. Another graph-based method, PPR, ranked third. In contrast, when few inputs from students were available, the performance of content-based approaches was worse than the baselines at the beginning of courses. These results suggest that the proposed methods are better at handling the cold-start problem caused by new users.

Item #	Thread location	Topics covered
1	Discussion on reading 4	Sexual identity issues; Early childhood education; Heteronormativity
2	Discussion on reading 4	Early childhood education; Need for and lack of male educators
3	Build-on comment on reading 1	Instruments to create a positive classroom environment
4	Discussion on reading 1	Safe school model; Positive school planning; Queer school models
5	Task brief for reading 4	Diversity and difference in early childhood education
6	Discussion on reading 3	Teacher's role in teaching anti-homophobia education; (Gay-Straight Alliance) GSA; Recommending additional resources;
7	Discussion on reading 2	Anti-homophobia; Anti-heterosexism; Cultural diversity
8	Build-on comment on reading 1	Gender binaries
9	Build-on comment on reading 2	Recommending additional resources to read
10	Build-on comment on reading 1	Suggesting more instruments to create a positive classroom; Homophobic name-calling; Anti-oppressive pedagogy

Table 4.8: Recommendations generated by CSCLRec for a sample user.

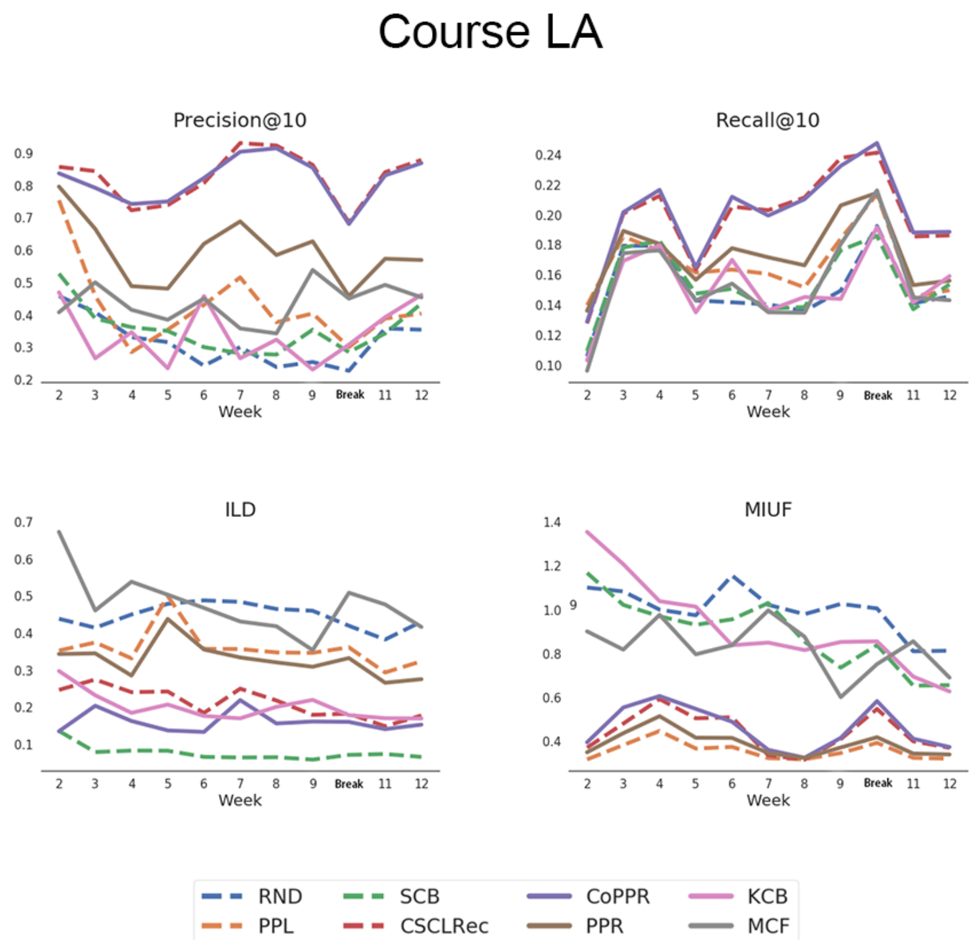

Figure 4.3: Weekly recommendation results for the LA course.

In addition, at the end of week 10, the line plot of the long course shows a drastic decrease in precision and a peak in recall. After investigation, I found that this was caused by an in-term break. During this week-long break, the forum activity was substantially reduced (a decrease of 54.7% compared to the previous week). As a result, the number of posts available for recommendation and samples that can be regarded as ground truth are reduced, leading to a decrease in precision. As the opposite of the precision measure, recall increases due to the fact that some students did not have interactions ($recall = 1.0$).

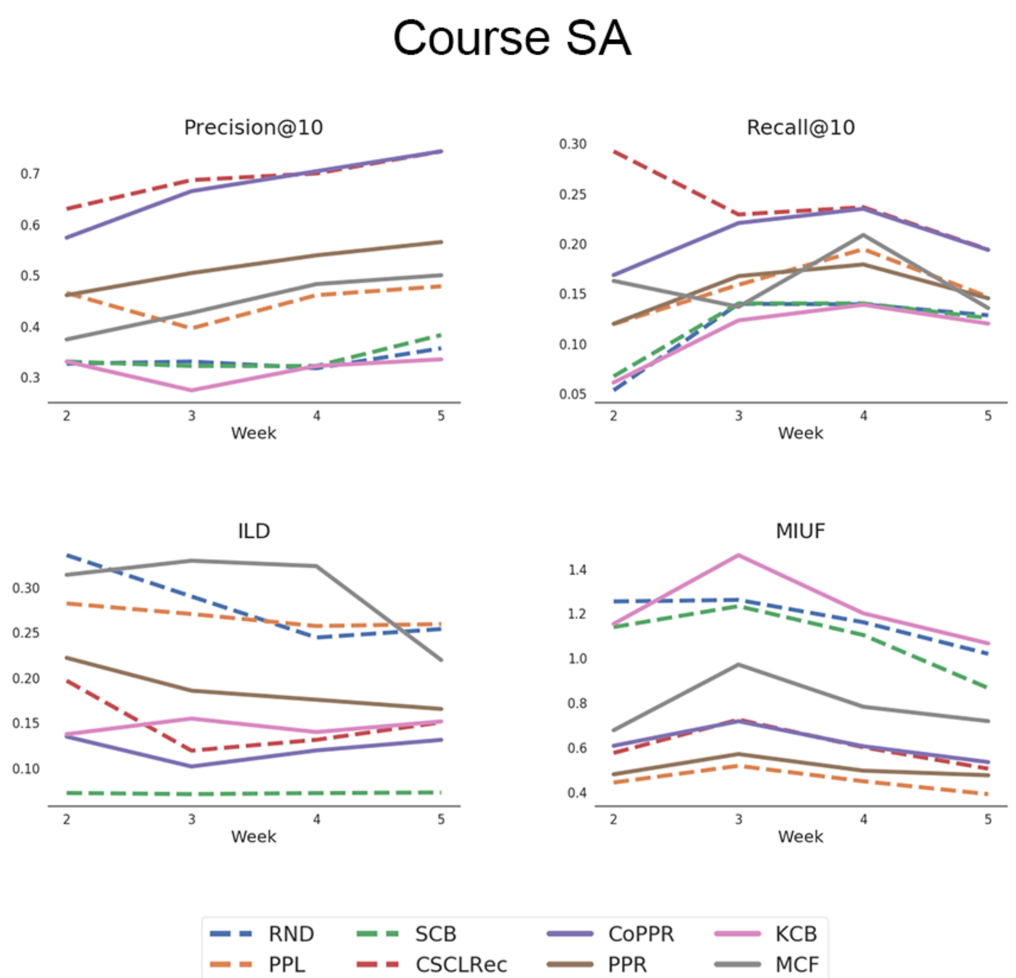

Figure 4.4: Weekly recommendation results for the SA course.

4.3.3 Recommender Performance Across User Types

To examine the recommenders' performance for different user types, I present Table 4.9. Similar result patterns to those seen for the model as a whole are observed from the evaluation performance for each learner subgroup. However, I did not perform statistical tests on these data due to the fact that the number of observations were small in most cases. For example, as shown in Figure 4.1, there was a limited number of listeners in all six test courses. Through the customized strategies, CSCLRec and CoPPR lift prediction performance for the four identified user types. In Table 4.9, as expected, the proposed approaches (CSCLRec and CoPPR), almost always had the highest precision and recall across all learner types.

In most cases, Table 4.9 shows the content-based recommenders working well with respect to novelty, while the collaborative filtering algorithms perform better than average at providing semantically diverse recommendations. The user type-specific results also reveal some inconsistencies with the overall results in Table 4.4. For example, the unpersonalized RND method outperformed others, especially the content-based systems, in recommendation novelty for peripheral and single-pass users. For the same two user types, PPL works better than other systems according to the diversity measure.

	P@10	**R@10**	**MIUF**	**ILD**
New User				
CSCLRec	0.868 (0.243)	0.384 (0.398)	0.245 (0.067)	0.411 (0.146)
CoPPR	0.860 (0.222)	0.394 (0.393)	0.257 (0.072)	0.387 (0.172)
PPR	0.821 (0.223)	0.390 (0.386)	0.215 (0.052)	0.483 (0.167)
MCF	0.326 (0.454)	0.293 (0.443)	0.147 (0.480)	0.517 (0.072)
SCB	0.200 (0.302)	0.299 (0.433)	1.924 (0.580)	0.325 (0.116)
KCB	0.068 (0.192)	0.268 (0.450)	1.421 (1.396)	0.382 (0.079)
RND	0.284 (0.299)	0.315 (0.424)	1.125 (0.359)	0.501 (0.133)
PPL	0.574 (0.398)	0.391 (0.385)	0.191 (0.049)	0.504 (0.158)
Listener				
CSCLRec	0.873 (0.216)	0.322 (0.341)	0.279 (0.091)	0.403 (0.129)
CoPPR	0.871 (0.201)	0.332 (0.339)	0.290 (0.094)	0.384 (0.158)
PPR	0.821 (0.230)	0.330 (0.333)	0.255 (0.087)	0.461 (0.159)
MCF	0.407 (0.402)	0.247 (0.373)	0.513 (0.742)	0.545 (0.134)
SCB	0.336 (0.359)	0.248 (0.364)	1.712 (0.611)	0.269 (0.127)
KCB	0.096 (0.240)	0.186 (0.387)	1.043 (1.286)	0.360 (0.087)
RND	0.361 (0.312)	0.249 (0.361)	1.160 (0.361)	0.471 (0.126)
PPL	0.661 (0.370)	0.334 (0.332)	0.229 (0.080)	0.482 (0.150)
Peripheral User				
CSCLRec	0.740 (0.331)	0.369 (0.422)	0.604 (0.292)	0.247 (0.124)
CoPPR	0.728 (0.338)	0.369 (0.421)	0.625 (0.309)	0.191 (0.086)
PPR	0.579 (0.417)	0.336 (0.438)	0.460 (0.227)	0.352 (0.164)
MCF	0.537 (0.425)	0.333 (0.441)	0.854 (0.590)	0.407 (0.175)
SCB	0.185 (0.325)	0.323 (0.447)	1.103 (0.396)	0.068 (0.019)
KCB	0.204 (0.331)	0.321 (0.447)	1.036 (0.372)	0.222 (0.110)
RND	0.208 (0.314)	0.321 (0.447)	1.121 (0.350)	0.371 (0.166)
PPL	0.266 (0.354)	0.338 (0.438)	0.409 (0.225)	0.430 (0.164)
Single-pass User				
CSCLRec	0.870 (0.253)	0.612 (0.445)	0.617 (0.312)	0.282 (0.118)
CoPPR	0.868 (0.255)	0.617 (0.441)	0.638 (0.317)	0.207 (0.096)
PPR	0.778 (0.347)	0.602 (0.456)	0.440 (0.228)	0.397 (0.150)
MCF	0.738 (0.363)	0.598 (0.459)	0.826 (0.520)	0.464 (0.152)
SCB	0.179 (0.316)	0.595 (0.463)	1.093 (0.441)	0.088 (0.054)
KCB	0.112 (0.261)	0.582 (0.481)	0.918 (0.542)	0.302 (0.148)
RND	0.163 (0.286)	0.588 (0.469)	1.124 (0.371)	0.451 (0.159)
PPL	0.193 (0.318)	0.598 (0.459)	0.407 (0.217)	0.470 (0.147)

Table 4.9: Recommender performance by user type as *M (SD)*.

4.4 Conclusion

Evaluations were conducted to compare CSCLRec and CoPPR with other commonly-used recommender systems. This chapter focuses on analyzing and explaining the patterns discovered from the offline evaluation results. These results show that CSCLRec and CoPPR perform far better than the other methods in accurately predicting the behaviors of students. In order to further illustrate their usability in a collaborative learning environment, their underlying help for student learning and potential extensions will be discussed in the next chapter.

Chapter 5

Discussion

5.1 Introduction

In Chapter 3, two recommender algorithms that support collaborative learning environments, CSCLRec and CoPPR, were introduced. In Chapter 4, evaluations using historical data from the PeppeR system demonstrated the performance of these recommender systems. Through the comparison with commonly-used recommender systems, results show that CSCLRec and CoPPR achieved higher accuracy in terms of forecasting user actions. They also showed acceptable recommendation diversity and novelty. These two findings suggest CSCLRec and CoPPR are appropriate choices for recommending posts in small discussion forums.

In this chapter, based on the comparison results from Chapter 4, I discuss the performance of each recommendation algorithm. Then I further analyze the strengths, implications, and limitations of the two proposed algorithms. In this way, I attempt to answer how they might help the knowledge co-construction process and the overall socio-collaborative learning environment.

5.2 Performance Comparisons and Analyses

As was observed in the results, the good performance of content-based (CB) recommender systems (KCB and SCB) in recommendation novelty demonstrates their ability to discover unpopular posts. This result is attributed to their recommendation mechanism: they only care about the similarity of the text content and not about user interaction. The results imply that, if the end goal is to help students specialize in a particular area of knowledge, some difficult-to-find but conceptually related discussions can be located faster. This expediency may help students consolidate and advance their understanding. This may also increase forum equity by increasing the visibility of the posts made by students with minority opinions that may otherwise go unnoticed in a popularity-based recommender scheme. More attention from fellow learners can be injected to help their integration into the discussion.

The results on recommendation diversity show that the CB algorithms suffer from the over-specialization problem. Their recommendation mechanism relies solely on content similarity, thus making such algorithms inherently unable to propose recommendations that are semantically diverse. Since discussions on the same thread usually have similar content, the suggestions provided by CB recommendation algorithms are likely to direct users to a few specific threads, which may prevent exposure to new ideas. This goes against the general teaching goals of many learning communities where students are expected to discuss different topics. Therefore, despite their good recommendation novelty, challenges are still presented for CB algorithms' integration into socio-collaborative learning contexts.

As the most commonly-used collaborative filtering (CF) recommender system, MCF shows performance above average in all cases, which demonstrates its good adaptability to educational datasets. However, since its mechanism focuses solely on user-item interactions, it tends to bias towards popular posts and result in low recommendation novelty. This can be illustrated by comparison with another class of pure recommenders, the CB methods. As shown in some other investigations, the lack of novelty may prevent the participation of students who possess minority opinions (Phirangee, Demmans Epp, & Hewitt, 2016). At the same time, another factor that presents challenges to CF systems is the lack of text analysis techniques. Because discourse is the only and most important medium in a social learning environment, the textual content of forum posts need to be considered (i.e., include content analysis methods).

Evaluation results across course types and user types suggest that graph-based systems have effectively discovered the potential interests of students, thanks to their feature of traversing nodes and edges to enrich recommendation possibilities. The proposed CSCLRec and CoPPR algorithms achieved the best precision and recall in all cases, followed by PPR. On the basis of PPR, CSCLRec and CoPPR hybridize content-based recommendation. At the same time, unlike PPR's one-size-fits-all recommendation strategy, they employ different strategies (i.e., adding more user-to-user edges) depending on different modelled learner types by taking advantage of socio-collaborative learning principles. The effectiveness of these add-on modules in predicting user preferences are demonstrated by the accuracy results.

When comparing CSCLRec and CoPPR, I found that their difference of performance with respect to the test dataset is subtle. CSCLRec is better than its counterpart according to accuracy and diversity, although the post-hoc test does not suggest a measurable difference. The potential slight difference in recommendation diversity is attributed to the design of CSCLRec's content analyzer module as it is the only difference between CSCLRec and CoPPR. Specifically, CoPPR connects post nodes through keyword matching. Consequently, some potential links between posts may be missed. For instance, it is likely to miss the connection between the post containing "assignment" and the post containing "homework". In CSCLRec, the design of concept-level content matching between posts can avoid this problem. However, I expect CoPPR to work better than CSCLRec in environments where the domain of discussion topics is more specific and narrower than what I saw in my dataset. For example, in STEM courses, terms used are very close to each other. The use of hypernyms in CSCLRec is prone to mis-classification due to the granularity of the WordNet ontology. For example, in Chemistry, both "Sodium chloride" and "Copper(II) sulfate" are a "chemical compound", but it makes little sense to link these two terms together as students might be talking about different things. The direct keyword matching mechanism of CoPPR makes it a better alternative to CSCLRec when this type of learning context, that requires more precise terminology, is being supported.

The RND and PPL methods' relatively good performance in ILD and

MIUF is largely due to the randomness introduced in them. Although they are the unpersonalized methods, they may sometimes still be an ideal choice in socio-collaborative learning contexts. This argument is supported by the results of the single pass and peripheral user in Table 4.9. Single pass users who only focus on newly generated posts often have a limited scope of activities. In this case, personalized recommendations (including MCF or SCB) may heavily rely on what they have been doing, so that the recommendations generated are also limited to the recently published posts. In contrast, unpersonalized systems can break this restriction by using strategies such as recommending random old posts even if the target user only focuses on recent posts. Similarly, peripheral users may lose readers as a result of not actively participating in recent discussions. Therefore, personalized systems may lack resources to predict the content that these users might prefer.

It should be acknowledged that the most suitable recommender system is not unique and fixed, as needed support varies for different learner types. For example, listeners not actively engaging in the discussions could be attributed to the recommendation lacking diversity. In this case, instead of CSCLRec, collaborative filtering approaches such as MCF might be a better choice. Moreover, new users may benefit from unpersonalized recommenders. For example, PPL could be used when the system lacks information about that learner because popular discussions may pique newcomer's interest and encourage them to participate. These findings provide a new idea for the practical application of recommender systems: in order to adapt to different types of learning behaviors, established teaching goals, and teachers' judgments of student needs, different recommender systems can be used or changed to meet a user's specific needs at that specific time.

5.3 Strengths and Implications

CSCLRec and CoPPR's outstanding accuracy demonstrates their ability to predict posts that students are looking to interact with. This implies that the systems are able to reduce the time spent on locating relevant information. As a result, students are left with more opportunities to participate in core learning activities such as discussion, review, and reflection. Such increased engagement in discussion-related activities enables more knowledge co-construction in discussion forums, as suggested in many previous studies (Hew & Cheung, 2012; Schellens & Valcke, 2006). At the same time, as more knowledge is shared in the community, students' critical thinking abilities are further developed (Carini, Kuh, & Klein, 2006) and better learning outcomes can be achieved (Cheng et al., 2011; Palmer, Holt, & Bray, 2008; Webb et al., 2004; Weber et al., 2008). In addition, for the forum as a whole, increased student participation means more discussions, resulting in more knowledge being created in this learning community. When this knowledge is shared and circulated in the community, it may attract the interest of more students and make them participate, thereby further increasing forum activity.

Although diversity and novelty are considered to be important measures, they were not best achieved through my recommender systems, as indicated in Chapter 4. The reason is that my design prioritizes accuracy (precision and recall) while keeping novelty and diversity within a reasonable range. In the specific context of educational recommender systems, there are reasons to make such considerations. First, recommendation diversity and novelty are attributed to the randomness of the recommendation algorithms. High diversity and novelty will lead to a decline in recommendation accuracy, and fewer recommendations would meet users' interests. Although they may discover new interests because of the broader recommendations, it is also likely that the recommendation list does not attract users' interest, thereby reducing their learning participation and overall forum activity. At the same time, recommendations that are too diverse or novel may not help students learn because students may not be prepared to interact with that novel content. In discussion forums with complex thread structures, diversified recommendations may consist of posts distributed on different topic threads with the novel ones usually holding non-mainstream views. Since the generation of such recommendations does not depend on the user preferences inferred from the user's past behavior, students may lack a basic understanding of the background of these recommended posts. This lack of background knowledge and relevant previous experience means that they may not be able to build on this post or improve knowledge, harming learning. Moreover, this knowledge mismatch is likely to cause users to lose their willingness to rely on the recommendation system.

As the analysis is moved to the individual student level, the outstanding performance in the results across user types highlights the strength of the learner categorization design. On one hand, it takes care of the unique needs of individuals and hopefully corrects their social learning behavior. On the other hand, this design for automatically identifying potentially problematic behaviours could reduce the burden of teachers, especially when course enrollments grow.

5.4 Potential Expansions

CSCLRec and CoPPR should have good generalizability. This is because the design of CSCLRec and CoPPR only relies on three types of data sources: user-to-user interaction, user-to-post interaction, and forum post content. These three sources are available in most forums, so that my recommender system design can be extended to similar discussion-based online communities, such as discussion boards for vocational training. Thanks to these systems' expected generalizability, several improvements could be made when applying the algorithms to support online learning communities.

5.4.1 Recommender Strategy Improvements

In addition to the above strengths, CSCLRec and CoPPR can be adjusted to allow for more functionality, for example, recommending peer learners. The current learner profile module finds peer learners who share the most edges with the target student. Due to the limitations of historical data and the requirement of consistency of the evaluations, my current design limits the number to a fixed percentage of the total number of students based on the number of interactions. As an example of how the design could be changed, the selection of peers can be decided based on more metrics (e.g. reply rate, interest overlap). As those more appropriate peers are identified, the recommender system can recommend both posts and users at the same time. This heterogeneous recommendation design would add flexibility. The recommended users might be more likely to stimulate richer discussion to support each other's learning (Vassileva, McCalla, & Greer, 2016) by encouraging the effective co-construction of knowledge. Should this happen, the change would make it more in line with the knowledge building principle of symmetric knowledge advancement (Scardamalia et al., 2002), where using information from others, contributing knowledge, and offering help to others are equally important.

5.4.2 Supporting Recommender Transparency and Increasing User Awareness

The practice of connecting users, posts, and keywords through a graph structure enables reasonable levels of transparency for recommendation results (de Gemmis et al., 2015; Tintarev & Masthoff, 2015). Extensions can be applied to the system to enhance user experience by making the decision process more visible. One approach would be to display the social network graph and the PPR graph to end users. Since the graph nodes and edges are intuitive to interpret, the users may understand why they receive such suggestions. At the same time, an open learner model (Bull & Kay, 2010), which uses visualizations to enable students to understand what the algorithms think they know or prefer, can be built. This way, users could understand their positions in the community. Instructors could also be given access to an open learner model so that they can focus on a particular student model or a model of all of the students in their course. This could help them monitor course progress and easily find students who need attention.

5.5 Limitation

5.5.1 Content Analysis Challenges

The content analyzer depends on WordNet which only supports nouns and verbs, meaning that some terms commonly used in educational scenarios, such as "office hour", will be disassembled into two words with completely different meanings, which creates some bias for my recommendation algorithm.

Without compromising the graph-based recommendation design, a potential solution could be to replace the WordNet analysis by using knowledge graphs. Relying on Linked Open Data (LOD), knowledge graphs are built by treating concepts or entities as nodes and then linking these nodes by the formal relationships defined in the LOD (Berners-Lee, 2009). Nowadays, incorporating knowledge graphs in graph-based recommender systems has shown promise in many recommender system studies (Musto et al., 2017; P. Nguyen et al., 2015; Pereira et al., 2018). They usually have a wider coverage of entities which allow them to overcome the problem of not having phrases as nodes (Shen, Wang, & Han, 2014). Moreover, the presence of various types of formal relations offers a greater opportunity to discover hidden connections among posts. Entity linking tools can be used to query post content so that key phrases would be detected and linked to entities in the knowledge graph. Instead of the hypernym nodes or keyword nodes in CSCLRec and CoPPR, sub-knowledge graphs containing these queried entities could be appended to the PPR graph to connect post nodes.

When I tried to incorporate some entity linking tools, (e.g., DBpedia spotlight (Daiber et al., 2013) and TagMe (Ferragina & Scaiella, 2010)), the performance seemed poor with regard to the PeppeR dataset. Many key phrases were not identified and failed to be linked to the correct knowledge graph entities. The main reason may be that poorly composed texts, such as forum posts in my case, present challenges for entity linking tools to disambiguate and match the correct knowledge graph entities (Heitmann & Hayes, 2010). Moreover, some open-source knowledge bases (i.e., DBpedia (Auer et al., 2007)) have a limited number of verb entities because most commonly-used verbs are treated as relations and only important verbs that are themselves resources are included. As a result, domain-specific technical terms were not included in the knowledge graph, for example "intervene". Because these domain-specific terms are missing, the establishment of a knowledge graph would depend on the input of course instructors, which would be a time-consuming process as it usually requires considerable effort.

5.5.2 Automatic Selection of Hyperparameters

Recommender performance depends on the configuration of hyper-parameters (e.g., the number of peer learners for each student, the temporal decay of older posts for different types of users) because they influence the selection and ordering of recommendations. For example, if the overall weight of post-to-hypernym edges is increased, the recommendation will include more thematically-similar items. The most suitable hyperparameters are often obtained by analyzing real user feedback. However, historical data does not allow for that. I could only apply an offline hyperparameter tuning process to test and select the most suitable hyperparameter value from a fixed list. A more dynamic approach could be used once the algorithms are integrated into a live system. Moreover, the present approach will have had minimal impact on the performance comparison results because recommenders with the same hyperparameters also have the same candidate values. When the user's feedback is available, two possible approaches could be attempted to optimize the hyperparameter tuning process.

First, the automatic adjustment of hyper-parameters can be informed through user feedback. In practice, real-time feedback collected from students can be used to adjust the strategies for the next round of recommendations, thereby sequentially improving recommender performance.

Second, the system can incorporate direct engagement from instructors and students, making itself into a user-controllable system. In this way, not only can the quality of suggestions be flexibly improved, but they can also be made more transparent to increase user satisfaction (Verbert et al., 2012). Therefore, in the future deployment study, course instructors can contribute to the adjustment of the hyper-parameters, because the teaching objectives, course settings, and student behavior vary from one course to the next. For example, they can be asked to decide how many different topics the recommendations should cover, so that a hyperparameter that controls the recommendation diversity can be changed accordingly. Moreover, both instructors' and students' feedback can help the recommendation engine to decide which of the algorithms to use in each specific scenario. Following such a dynamic approach, as more information is collected from end users, the recommendations can gradually better reflect their understanding of the student needs and fit established teaching plans, as suggested by McCalla (2004).

5.5.3 Need for Human Evaluations

Further evaluations that involve reviewers and actual users are needed to fully validate the performance of CSCLRec and CoPPR. In Chapter 4, the performed evaluations reflected the ability of the recommender system to predict student behavior and provide semantically diverse or novel recommendations. In terms of recommendation accuracy, because of the successful prediction of user preferences, students are assumed to save time for more important learning activities. However, this initial evaluation does not measure whether the recommendations satisfy learner interests or support their learning process. Therefore, student feedback, in the form of surveys or reviews, need to be acquired in future evaluations. Their reactions to the recommendations can also be logged by the system and be used to continually improve the recommendations. Similarly, as the recommendation diversity metric only depends on textual similarity, more evaluations need to be done to measure how diverse the discussion topics or the knowledge components contained in the recommended posts are.

To move towards this broader evaluation of the algorithms' impacts, we are working on another study in which expert reviewers were invited to evaluate the recommender systems from a socio-collaborative learning perspective. They have been asked to judge whether the suggestions given by the recommenders can effectively promote the knowledge co-construction of each student or the socio-collaborative learning process of the community, based on these experts' understanding of the course goals, the learning needs of students, and their understanding of collaborative learning processes. While not part of this thesis, the results should enable us to improve existing algorithms, adjust the configurations, and provide practical suggestions to help operators (e.g., course instructors) prepare for a future in-class deployment study.

5.6 Conclusion

In this chapter, I analyzed and explained the patterns discovered from the offline evaluation results. The strengths, limitations, and use cases of the recommendation algorithms were discussed. In addition, this chapter argued for the potential of CSCLRec and CoPPR to support learning in socio-collaborative learning workspaces. Specifically, these recommenders can provide students with more time for learning through accurate predictions that direct students to relevant posts rather than requiring students to go through everything to find those learning materials. As a possible extension, I propose that after deploying the system, the internal graph model can be presented to users and an additional feature of recommending peer learners can be provided. Some future work includes trying to modify the settings of the content analyzer and conducting follow-up experiments to address some limitations of my method and current system design.

Chapter 6

Conclusion

Today, with the popularity of online education, there is an increasing demand for personalization. Many online learning environments urgently need personalized technology to provide a more efficient and effective online learning experience. These include small-scale discussion forums deployed for online courses, which suffer from a reduced level of student participation due to the information overload presented by complex thread structures and lengthy discussions.

To address this problem, this thesis presents two graph-based hybrid recommender systems, CSCLRec and CoPPR, to recommend posts for students in asynchronous discussion forums. The proposed recommender systems employ multiple techniques including SNA and NLP to generate recommendations from the perspective of both user interaction and text semantics. Their design is grounded in educational theories and principles that focus on collaborative learning and knowledge building. Therefore, the systems can categorize students based on knowledge sharing behaviors such that customized recommendations can be delivered to meet different types of needs.

Compared with other approaches to supporting student learning in online discussion formats, this research focuses on the characteristics and needs of small-scale discussion forums. It fills in the research gap of lacking forum post recommenders in this context and contributes to the need for personalization techniques to support learning. As my proposed recommender algorithms (CoPPR and CSCLRec) innovatively detect four different types of problematic knowledge sharing behaviors, I explored the support of my methods for students who exhibit these behaviors. Because the recommender systems only depend on three types of data sources that are available in most forums, they are also expected to achieve satisfactory performance in other similar areas, for example, forums for skill development.

Through evaluations, the performance of CSCLRec and CoPPR were compared with 6 other commonly-used recommender systems. Using historical data from a forum-based educational platform, the recommenders were tasked with recommending posts and evaluated on four dimensions: precision, recall, diversity, and novelty. Results from the offline experiment showed that my proposed algorithms had an outstanding performance with respect to precision and recall. This demonstrates their ability to accurately predict student preferences from past behaviors. This result further suggests that the systems can save student time when searching for useful resources, thereby potentially increasing student engagement in learning activities. On top of this result, I argue that the recommenders may improve the knowledge co-construction process in a collaborative learning context, which could lead to better learning outcomes. Meanwhile, CSCLRec and CoPPR's acceptable performance on diversity and novelty measures implies they balance between providing recommendations that help develop interest and recommendations that are aligned with the student's background.

Nevertheless, I acknowledge that the current offline evaluation method cannot validate the help the recommender system provides to students. Therefore, more evaluations that involve actual users are expected to be conducted as future work. A review study is underway in which education experts are evaluating how the recommendations might facilitate the knowledge co-construction process. Following this, a deployment study can be conducted to evaluate how the use of the recommenders influences student experience and learning. A later deployment could also be used to study the on-demand adjustment of hyperparameters by instructors.

References

Abel, F., Bittencourt, I. I., Costa, E., Henze, N., Krause, D., & Vassileva, J. (2010). Recommendations in online discussion forums for e-learning systems. *IEEE Transactions on Learning Technologies*, *3*(2), 165–176.

Abel, F., Bittencourt, I. I., Henze, N., Krause, D., & Vassileva, J. (2008). A rule-based recommender system for online discussion forums (W. Nejdl, J. Kay, P. Pu, & E. Herder, Eds.). In W. Nejdl, J. Kay, P. Pu, & E. Herder (Eds.), *Adaptive hypermedia and adaptive web-based systems*, Hannover, Germany, Springer.

Akcayir, G., Chen, Z., Demmans Epp, C., Pandeliev, V., & Munteanu, C. (2020). Two case studies of online discussion use in computer science education: Deep vs. shallow integration and recommendations, In *Handbook of research on online discussion-based teaching methods*. IGI Global.

Albatayneh, N. A., Ghauth, K. I., & Chua, F.-F. (2018). Utilizing Learners' Negative Ratings in Semantic Content-based Recommender System for e-Learning Forum. *Journal of Educational Technology & Society*, *21*(1), 112–125. http://www.jstor.org/stable/26273873.

Andresen, M. A. (2009). Asynchronous discussion forums: Success factors, outcomes, assessments, and limitations. *Journal of Educational Technology Society*, *12*(1), 249–257. http://www.jstor.org/stable/jeductechsoci.12.1.249.

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). Dbpedia: A nucleus for a web of open data (K. Aberer, K.-S. Choi, N. Noy, D. Allemang, K.-I. Lee, L. Nixon, J. Golbeck, P. Mika, D. Maynard, R. Mizoguchi, G. Schreiber, & P. Cudré-Mauroux, Eds.). In K. Aberer, K.-S. Choi, N. Noy, D. Allemang, K.-I. Lee, L. Nixon, J. Golbeck, P. Mika, D. Maynard, R. Mizoguchi, G. Schreiber, & P. Cudré-Mauroux (Eds.), *The semantic web*, Berlin, Heidelberg, Springer.

Bach, N. X., Hai, N. D., & Phuong, T. M. (2016). Personalized recommendation of stories for commenting in forum-based social media. *Information Sciences*, *352-353*, 48–60. https://doi.org/https://doi.org/10.1016/j.ins.2016.03.006.

Beham, G., Kump, B., Ley, T., & Lindstaedt, S. (2010). Recommending knowledgeable people in a work-integrated learning system [Proceedings of the 1st Workshop on Recommender Systems for Technology Enhanced

Learning (RecSysTEL 2010)]. *Procedia Computer Science*, *1*(2), 2783–2792. https://doi.org/https://doi.org/10.1016/j.procs.2010.08.003.

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation [Data Mining for Software Trustworthiness]. *Information Sciences*, *191*, 192–213. https://doi.org/https://doi.org/10.1016/j.ins.2011.12.028.

Berners-Lee, T. (2009). Linked-data design issues. http://www.w3.org/DesignIssues/LinkedData.html.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993–1022.

Bobadilla, J., Ortega, F., Hernando, A., & Bernal, J. (2012). A collaborative filtering approach to mitigate the new user cold start problem. *Knowledge-Based Systems*, *26*, 225–238. https://doi.org/https://doi.org/10.1016/j.knosys.2011.07.021.

Bonet, G., & Walters, B. R. (2016). High impact practices: Student engagement and retention. *College Student Journal*, *50*(2), 224–235.

Bousbahi, F., & Chorfi, H. (2015). Mooc-rec: A case based recommender system for moocs [World Conference on Technology, Innovation and Entrepreneurship]. *Procedia - Social and Behavioral Sciences*, *195*, 1813–1822. https://doi.org/https://doi.org/10.1016/j.sbspro.2015.06.395.

Breese, J. S., Heckerman, D., & Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering, In *Proceedings of the fourteenth conference on uncertainty in artificial intelligence*, Madison, Wisconsin, Morgan Kaufmann Publishers Inc.

Brooks, C., Greer, J., & Gutwin, C. (2014). The data-assisted approach to building intelligent technology-enhanced learning environments. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 123–156). New York, NY, Springer. https://doi.org/10.1007/978-1-4614-3305-7_7.

Bull, S., & Kay, J. (2010). Open learner models. In R. Nkambou, J. Bourdeau, & R. Mizoguchi (Eds.), *Advances in intelligent tutoring systems* (pp. 301–322). Berlin, Heidelberg, Springer. https://doi.org/10.1007/978-3-642-14363-2_15.

Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, *12*(4), 331–370. https://doi.org/10.1023/A:1021240730564.

Capelle, M., Hogenboom, F., Hogenboom, A., & Frasincar, F. (2013). Semantic news recommendation using wordnet and bing similarities, In *Proceedings of the 28th annual acm symposium on applied computing*, Coimbra, Portugal, Association for Computing Machinery. https://doi.org/10.1145/2480362.2480426.

Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student Engagement and Student Learning: Testing the Linkages. *Research in Higher Education*, *47*(1), 1–32. https://doi.org/10.1007/s11162-005-8150-9.

Cartwright, D., & Harary, F. (1956). Structural balance: a generalization of Heider's theory. *US*, American Psychological Association. https://doi.org/10.1037/h0046049.

Castells, P., Hurley, N. J., & Vargas, S. (2015). Novelty and diversity in recommender systems, In *Recommender systems handbook*. Springer.

Cer, D., Yang, Y., Kong, S.-y., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Et al. (2018). Universal sentence encoder. *arXiv preprint arXiv:1803.11175*.

Chandrasekaran, M. K., Demmans Epp, C., Kan, M.-Y., & Litman, D. (2017). Using discourse signals for robust instructor intervention prediction, In *Proceedings of the thirty-first aaai conference on artificial intelligence*, San Francisco, California, USA, AAAI Press.

Chen, W., & Persen, R. (2009). A recommender system for collaborative knowledge, In *Proceedings of the 2009 conference on artificial intelligence in education: Building learning systems that care: From knowledge representation to affective modelling*, Brighton, UK, IOS Press.

Cheng, C. K., Paré, D. E., Collimore, L.-M., & Joordens, S. (2011). Assessing the effectiveness of a voluntary online discussion forum on improving students' course performance. *Computers & Education*, *56*(1), 253–261.

Coxhead, A. (2016). Reflecting on coxhead (2000), "a new academic word list". *TESOL Quarterly*, *50*(1), 181–185. http://www.jstor.org/stable/43893808.

Cremonesi, P., Koren, Y., & Turrin, R. (2010). Performance of recommender algorithms on top-n recommendation tasks, In *Proceedings of the fourth acm conference on recommender systems*, Barcelona, Spain, Association for Computing Machinery. https://doi.org/10.1145/1864708.1864721.

Daiber, J., Jakob, M., Hokamp, C., & Mendes, P. N. (2013). Improving efficiency and accuracy in multilingual entity extraction, In *Proceedings of the 9th international conference on semantic systems*, Graz, Austria, Association for Computing Machinery. https://doi.org/10.1145/2506182.2506198.

de Gemmis, M., Lops, P., Musto, C., Narducci, F., & Semeraro, G. (2015). Semantics-aware content-based recommender systems. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (pp. 119–159). Boston, MA, Springer. https://doi.org/10.1007/978-1-4899-7637-6_4.

Demmans Epp, C., Phirangee, K., & Hewitt, J. (2017). Student actions and community in online courses: The roles played by course length and facilitation method. *Online Learning*, *21*(4). https://doi.org/10.24059/olj.v21i4.1269.

Ding, L., Kim, C., & Orey, M. (2017). Studies of student engagement in gamified online discussions. *Computers Education*, *115*, 126–142. https://doi.org/https://doi.org/10.1016/j.compedu.2017.06.016.

Dowell, N. M., Brooks, C., Kovanović, V., Joksimović, S., & Gašević, D. (2017). The changing patterns of mooc discourse, In *Proceedings of the fourth*

(2017) acm conference on learning @ scale, Cambridge, Massachusetts, USA, Association for Computing Machinery. https://doi.org/10.1145/3051457.3054005.

Drachsler, H., Verbert, K., Santos, O. C., & Manouselis, N. (2015). Panorama of recommender systems to support learning, In *Recommender systems handbook*. Springer.

El Helou, S., Salzmann, C., & Gillet, D. (2010). The 3a personalized, contextual and relation-based recommender system. *J. UCS*, *16*(16), 2179–2195.

Erdt, M., Fernandez, A., & Rensing, C. (2015). Evaluating recommender systems for technology enhanced learning: A quantitative survey. *IEEE Transactions on Learning Technologies*, *8*(4), 326–344.

Fazeli, S., Drachsler, H., Bitter-Rijpkema, M., Brouns, F., Van der Vegt, W., & Sloep, P. B. (2017). User-centric evaluation of recommender systems in social learning platforms: Accuracy is just the tip of the iceberg. *IEEE Transactions on Learning Technologies*, *11*(3), 294–306.

Fazeli, S., Loni, B., Drachsler, H., & Sloep, P. (2014). Which recommender system can best fit social learning platforms? (C. Rensing, S. de Freitas, T. Ley, & P. J. Muñoz-Merino, Eds.). In C. Rensing, S. de Freitas, T. Ley, & P. J. Muñoz-Merino (Eds.), *Open learning and teaching in educational communities*, Graz, Austria, Springer International Publishing.

Ferragina, P., & Scaiella, U. (2010). Tagme: On-the-fly annotation of short text fragments (by wikipedia entities), In *Proceedings of the 19th acm international conference on information and knowledge management*, Toronto, ON, Canada, Association for Computing Machinery. https://doi.org/10.1145/1871437.1871689.

Fosnot, C. T., & Perry, R. S. (1996). Constructivism: A psychological theory of learning. *Constructivism: Theory, perspectives, and practice*, *2*, 8–33.

Fu, E. L. F., van Aalst, J., & Chan, C. K. K. (2016). Toward a classification of discourse patterns in asynchronous online discussions. *International Journal of Computer-Supported Collaborative Learning*, *11*(4), 441–478. https://doi.org/10.1007/s11412-016-9245-3.

Geer, R. (2005). Imprinting and its impact on online learning communities, In *Balance, fidelity, mobility: Proceedings of the 22nd ascilite conference*, Brisbane, Australia. Australian Society for Computers in Learning in Tertiary Education Brisbane.

Grunspan, D. Z., Wiggins, B. L., & Goodreau, S. M. (2014). Understanding Classrooms through Social Network Analysis: A Primer for Social Network Analysis in Education Research. *CBE life sciences education*, *13*(2), 167–179. https://doi.org/10.1187/cbe.13-08-0162.

Guy, I. (2015). Social recommender systems. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (pp. 511–543). Boston, MA, Springer. https://doi.org/10.1007/978-1-4899-7637-6_15.

Hallinan, M. T., & Williams, R. A. (1990). Students' characteristics and the peer-influence process. *Sociology of Education*, *63*(2), 122–132. https://doi.org/10.2307/2112858.

Hammond, M. (1999). Issues Associated with Participation in On line Forums—the Case of the Communicative Learner. *Education and Information Technologies*, *4*(4), 353–367. https://doi.org/10.1023/A:1009661512881.

Haveliwala, T. H. (2003). Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search. *IEEE Transactions on Knowledge and Data Engineering*, *15*(4), 784–796.

Heitmann, B., & Hayes, C. (2010). Using linked data to build open, collaborative recommender systems, In *2010 aaai spring symposium series*, Palo Alto, CA, USA.

Hew, K. F., & Cheung, W. S. (2012). *Student participation in online discussions: Challenges, solutions, and future research*. Springer Publishing Company.

Hewitt, J. (2005). Toward an understanding of how threads die in asynchronous computer conferences. *Journal of the Learning Sciences*, *14*(4), 567–589.

Hmelo-Silver, C. (2013). *The International Handbook of Collaborative Learning*. Routledge. https://doi.org/10.4324/9780203837290.

Hu, Y., Koren, Y., & Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets, In *2008 eighth ieee international conference on data mining*, Pisa, Italy.

Huang, Z., Chung, W., Ong, T.-H., & Chen, H. (2002). A graph-based recommender system for digital library, In *Proceedings of the 2nd acm/ieee-cs joint conference on digital libraries*, Portland, Oregon, USA, Association for Computing Machinery. https://doi.org/10.1145/544220.544231.

Hull, D. M., & Saxon, T. F. (2009). Negotiation of meaning and co-construction of knowledge: An experimental analysis of asynchronous online instruction. *Computers Education*, *52*(3), 624–639. https://doi.org/https://doi.org/10.1016/j.compedu.2008.11.005.

Jeong, H., & Hmelo-Silver, C. E. (2016). Seven Affordances of Computer-Supported Collaborative Learning: How to Support Collaborative Learning? How Can Technologies Help? *Educational Psychologist*, *51*(2), 247–265. https://doi.org/10.1080/00461520.2016.1158654.

Johnson, G. M. (2006). Synchronous and Asynchronous Text-Based CMC in Educational Contexts: A Review of Recent Research. *TechTrends*, *50*(4), 46. https://doi.org/10.1007/s11528-006-0046-9.

Jones, D., Bench-capon, T., & Visser, P. (1998). Methodologies for ontology development.

Karampiperis, P., Koukourikos, A., & Stoitsis, G. (2014). Collaborative filtering recommendation of educational content in social environments utilizing sentiment analysis techniques. In *Recommender systems for technology enhanced learning* (pp. 3–23). Springer.

Kear, K. (2004). Peer Learning Using Asynchronous Discussion Systems in Distance Education. *Open Learning: The Journal of Open, Distance and e-Learning*, *19*(2), 151–164. https://doi.org/10.1080/0268051042000224752.

Kleanthous, S., & Dimitrova, V. (2008). Modelling semantic relationships and centrality to facilitate community knowledge sharing (W. Nejdl, J. Kay, P. Pu, & E. Herder, Eds.). In W. Nejdl, J. Kay, P. Pu, & E. Herder (Eds.), *5th international conference on adaptive hypermedia and adaptive web-based systems*, Hannover, Germany, Springer.

Kleanthous, S., & Dimitrova, V. (2009). Detecting changes over time in a knowledge sharing community, In *2009 ieee/wic/acm international joint conference on web intelligence and intelligent agent technology*, Milan, Italy.

Kleanthous, S., & Dimitrova, V. (2010). Analyzing community knowledge sharing behavior (P. De Bra, A. Kobsa, & D. Chin, Eds.). In P. De Bra, A. Kobsa, & D. Chin (Eds.), *2010 international conference on user modeling, adaptation, and personalization*, Big Island, HI, USA, Springer.

Kleanthous, S., & Dimitrova, V. (2013). Adaptive Notifications to Support Knowledge Sharing in Close-knit Virtual Communities. *User Modeling and User-Adapted Interaction*, *23*(2), 287–343. https://doi.org/10.1007/s11257-012-9127-y.

Koren, Y., & Bell, R. (2011). Advances in collaborative filtering. In F. Ricci, L. Rokach, B. Shapira, & P. B. Kantor (Eds.), *Recommender systems handbook* (pp. 145–186). Boston, MA, Springer. https://doi.org/10.1007/978-0-387-85820-3_5.

Koschmann, T. (Ed.). (1996). *CSCL: Theory and practice of an emerging paradigm*. Hillsdale, NJ, US, Lawrence Erlbaum Associates, Inc.

Lan, A. S., Spencer, J. C., Chen, Z., Brinton, C. G., & Chiang, M. (2019). Personalized thread recommendation for mooc discussion forums (M. Berlingerio, F. Bonchi, T. Gärtner, N. Hurley, & G. Ifrim, Eds.). In M. Berlingerio, F. Bonchi, T. Gärtner, N. Hurley, & G. Ifrim (Eds.), *Machine learning and knowledge discovery in databases*, Würzburg, Germany, Springer International Publishing.

Lesk, M. (1986). Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone, In *Proceedings of the 5th annual international conference on systems documentation*, Toronto, Ontario, Canada, Association for Computing Machinery. https://doi.org/10.1145/318723.318728.

Magno, T., & Sable, C. (2008). A comparison of signal-based music recommendation to genre labels, collaborative filtering, musicological analysis, human recommendation, and random baseline. *ISMIR 2008 - 9th International Conference on Music Information Retrieval*, 161–166.

Manouselis, N., Drachsler, H., Verbert, K., & Duval, E. (2012). *Recommender systems for learning*. Springer Science & Business Media.

Marbouti, F., & Wise, A. F. (2016). Starburst: a New Graphical Interface to Support Purposeful Attention to Others' Posts in Online Discussions. *Educational Technology Research and Development*, *64*(1), 87–113. https://doi.org/10.1007/s11423-015-9400-y.

Mason, R. B. (2011). Student engagement with, and participation in, an e-forum. *J. Educ. Technol. Soc.*, *14*, 258–268.

McCalla, G. I. (2004). The ecological approach to the design of e-learning environments: Purpose-based capture and use of information about learners. *Journal of Interactive Media in Education*, *2004*(1). https://doi.org/http://doi.org/10.5334/2004-7-mccalla.

Miller, G. A. (1995). Wordnet: A lexical database for english. *Commun. ACM*, *38*(11), 39–41. https://doi.org/10.1145/219717.219748.

Min, W., Rowe, J. P., Mott, B. W., & Lester, J. C. (2013). Personalizing embedded assessment sequences in narrative-centered learning environments: A collaborative filtering approach (H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik, Eds.). In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *16th international conference on artificial intelligence in education*, Memphis, TN, USA, Springer.

Morales, M., Amado-Salvatierra, H. R., Hernández, R., Pirker, J., & Gütl, C. (2016). A practical experience on the use of gamification in mooc courses as a strategy to increase motivation (L. Uden, D. Liberona, & B. Feldmann, Eds.). In L. Uden, D. Liberona, & B. Feldmann (Eds.), *Learning technology for education in cloud – the changing face of education*, Hagen, Germany, Springer International Publishing.

Moreno, J. L. (1934). Who shall survive?: A new approach to the problem of human interrelations. *Washington, DC, US*, Nervous; Mental Disease Publishing Co. https://doi.org/10.1037/10648-000.

Musto, C., Basile, P., Lops, P., de Gemmis, M., & Semeraro, G. (2017). Introducing linked open data in graph-based recommender systems. *Information Processing & Management*, *53*(2), 405–435.

Nguyen, P., Tomeo, P., Di Noia, T., & Di Sciascio, E. (2015). An evaluation of simrank and personalized pagerank to build a recommender system for the web of data, In *Proceedings of the 24th international conference on world wide web*, Florence, Italy, Association for Computing Machinery. https://doi.org/10.1145/2740908.2742141.

Nguyen, T. T., Hui, P.-M., Harper, F. M., Terveen, L., & Konstan, J. A. (2014). Exploring the filter bubble: The effect of using recommender systems on content diversity, In *Proceedings of the 23rd international conference on world wide web*, Seoul, Korea, Association for Computing Machinery. https://doi.org/10.1145/2566486.2568012.

Ning, X., Desrosiers, C., & Karypis, G. (2015). A comprehensive survey of neighborhood-based recommendation methods. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (pp. 37–76). Boston, MA, Springer. https://doi.org/10.1007/978-1-4899-7637-6_2.

Oard, D., & Kim, J. (1998). Implicit feedback for recommender systems, In *Proceedings of the aaai workshop on recommender systems*.

O'Callaghan, F. V., Neumann, D. L., Jones, L., & Creed, P. A. (2017). The use of lecture recordings in higher education: A review of institutional, student, and lecturer issues. *Education and Information Technologies*, *22*(1), 399–415. https://doi.org/10.1007/s10639-015-9451-z.

Ortega-Arranz, A., Er, E., Martínez-Monés, A., Bote-Lorenzo, M. L., Asensio-Pérez, J. I., & Muñoz-Cristóbal, J. A. (2019). Understanding Student Behavior and Perceptions toward Earning Badges in a Gamified MOOC. *Universal Access in the Information Society*, *18*(3), 533–549. https://doi.org/10.1007/s10209-019-00677-8.

Oztok, M., Zingaro, D., Makos, M., Hewitt, J., & Brett, C. (2014). Towards understanding threads as social and cognitive artifacts for knowledge building in online learning, In *Annual meeting of american educational research association*, Philadelphia, Pennsylvania, American Educational Research Association.

Palmer, S., Holt, D., & Bray, S. (2008). Does the Discussion Help? the Impact of a Formally Assessed Online Discussion on Final Student Results. *British Journal of Educational Technology*, *39*(5), 847–858. https://doi.org/10.1111/j.1467-8535.2007.00780.x.

Pan, R., Zhou, Y., Cao, B., Liu, N. N., Lukose, R., Scholz, M., & Yang, Q. (2008). One-class collaborative filtering. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 502–511. https://doi.org/10.1109/ICDM.2008.16.

Park, Y.-J., & Tuzhilin, A. (2008). The long tail of recommender systems and how to leverage it, In *Proceedings of the 2008 acm conference on recommender systems*, Lausanne, Switzerland, Association for Computing Machinery. https://doi.org/10.1145/1454008.1454012.

Pereira, C. K., Campos, F., Ströele, V., David, J. M. N., & Braga, R. (2018). BROAD-RSI – educational recommender system using social networks interactions and linked data. *Journal of Internet Services and Applications*, *9*(1), 7. https://doi.org/10.1186/s13174-018-0076-5.

Phirangee, K., Demmans Epp, C., & Hewitt, J. (2016). Exploring the relationships between facilitation methods, students' sense of community, and their online behaviors. *Online Learning*, *20*(2), 134–154. https://doi.org/10.24059/olj.v20i2.775.

Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development*. University of Chicago Press.

Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender systems: Introduction and challenges. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (pp. 1–34). Boston, MA, Springer. https://doi.org/10.1007/978-1-4899-7637-6_1.

Rosé, C. P., & Ferschke, O. (2016). Technology Support for Discussion Based Learning: From Computer Supported Collaborative Learning to the Future of Massive Open Online Courses. *International Journal of Ar-*

tificial Intelligence in Education, *26*(2), 660–678. https://doi.org/10.1007/s40593-016-0107-y.

Rubens, N., Elahi, M., Sugiyama, M., & Kaplan, D. (2015). Active learning in recommender systems. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (pp. 809–846). Boston, MA, Springer. https://doi.org/10.1007/978-1-4899-7637-6_24.

Santos, O. C., Boticario, J. G., & Pérez-Marín, D. (2014). Extending web-based educational systems with personalised support through user centred designed recommendations along the e-learning life cycle [Software Development Concerns in the e-Learning Domain]. *Science of Computer Programming*, *88*, 92–109. https://doi.org/https://doi.org/10.1016/j.scico.2013.12.004.

Scardamalia, M. (2004). Csile/knowledge forum®. *Education and technology: An encyclopedia*, *183*, 192.

Scardamalia, M. Et al. (2002). Collective cognitive responsibility for the advancement of knowledge. *Liberal education in a knowledge society*, *97*, 67–98.

Scardamalia, M., & Bereiter, C. (1994). Computer Support for Knowledge-Building Communities. *Journal of the Learning Sciences*, *3*(3), 265–283. https://doi.org/10.1207/s15327809jls0303_3.

Scardamalia, M., & Bereiter, C. (2006). Knowledge Building: Theory, Pedagogy, and Technology., In *The cambridge handbook of: The learning sciences.* Scardamalia, Marlene: Ontario Institute for Studies in Education, 252 Bloor Street West, Toronto, ON, Canada, M5S1V6, mscardamalia@oise.utoronto.ca, Cambridge University Press.

Scardamalia, M., & Bereiter, C. (2010). A brief history of knowledge building. *Canadian Journal of Learning and Technology/La revue canadienne de l'apprentissage et de la technologie*, *36*(1).

Schellens, T., & Valcke, M. (2006). Fostering knowledge construction in university students through asynchronous discussion groups. *Computers Education*, *46*(4), 349–370. https://doi.org/https://doi.org/10.1016/j.compedu.2004.07.010.

Shani, G., & Gunawardana, A. (2011). Evaluating recommendation systems, In *Recommender systems handbook*. Springer.

Shen, W., Wang, J., & Han, J. (2014). Entity linking with a knowledge base: Issues, techniques, and solutions. *IEEE Transactions on Knowledge and Data Engineering*, *27*(2), 443–460.

Sindhwani, V., Bucak, S. S., Hu, J., & Mojsilovic, A. (2010). One-class matrix completion with low-density factorizations. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 1055–1060. https://doi.org/10.1109/ICDM.2010.164.

Smith, B. L., & MacGregor, J. T. (1992). What is collaborative learning. Washington.

Smyth, B., & McClave, P. (2001). Similarity vs. diversity, In *International conference on case-based reasoning*, Vancouver, BC, Canada. Springer.

Steck, H. (2011). Item popularity and recommendation accuracy, In *Proceedings of the fifth acm conference on recommender systems*, Chicago, Illinois, USA, Association for Computing Machinery. https://doi.org/10.1145/2043932.2043957.

Tintarev, N., & Masthoff, J. (2015). Explaining recommendations: Design and evaluation. In *Recommender systems handbook* (pp. 353–382). Boston, MA, Springer. https://doi.org/10.1007/978-1-4899-7637-6_10.

Tirado, R., Hernando, Á., & Aguaded, J. I. (2015). The Effect of Centralization and Cohesion on the Social Construction of Knowledge in Discussion Forums. *Interactive Learning Environments*, *23*(3), 293–316. https://doi.org/10.1080/10494820.2012.745437.

van den Oord, A., Dieleman, S., & Schrauwen, B. (2013). Deep content-based music recommendation (C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger, Eds.). In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems 26*. Curran Associates, Inc. http://papers.nips.cc/paper/5004-deep-content-based-music-recommendation.pdf.

Vassileva, J., McCalla, G. I., & Greer, J. E. (2016). From Small Seeds Grow Fruitful Trees: How the PHelpS Peer Help System Stimulated a Diverse and Innovative Research Agenda over 15 Years. *International Journal of Artificial Intelligence in Education*, *26*(1), 431–447. https://doi.org/10.1007/s40593-015-0073-9.

Vassileva, J. (2008). Toward social learning environments. *IEEE transactions on learning technologies*, *1*(4), 199–214.

Veldhuis-Diermanse, A. (2002). *Csclearning?: Participation, learning activities and knowledge construction in computer-supported collaborative learning in higher education* (Doctoral dissertation). Wageningen University Research. Netherlands, Wageningen University.

Verbert, K., Manouselis, N., Ochoa, X., Wolpers, M., Drachsler, H., Bosnic, I., & Duval, E. (2012). Context-aware recommender systems for learning: A survey and future challenges. *IEEE Transactions on Learning Technologies*, *5*(4), 318–335.

Vygotsky, L. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Wang, C., & Blei, D. M. (2011). Collaborative topic modeling for recommending scientific articles, In *Proceedings of the 17th acm sigkdd international conference on knowledge discovery and data mining*, San Diego, California, USA, Association for Computing Machinery. https://doi.org/10.1145/2020408.2020480.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press. https://doi.org/10.1017/CBO9780511815478.

Webb, E., Jones, A., Barker, P., & van Schaik, P. (2004). Using E-learning Dialogues in Higher Education. *Innovations in Education and Teaching International*, *41*(1), 93–103. https://doi.org/10.1080/1470329032000172748.

Weber, K., Maher, C., Powell, A., & Lee, H. S. (2008). Learning Opportunities from Group Discussions: Warrants Become the Objects of Debate. *Educational Studies in Mathematics*, *68*(3), 247–261. https://doi.org/10.1007/s10649-008-9114-8.

Wise, A. F., Hausknecht, S. N., & Zhao, Y. (2014). Attending to others' posts in asynchronous discussions: Learners' online "listening" and its relationship to speaking. *International Journal of Computer-Supported Collaborative Learning*, *9*(2), 185–209. https://doi.org/10.1007/s11412-014-9192-9.

Xiong, Y., Li, H., Kornhaber, M. L., Suen, H. K., Pursel, B., & Goins, D. D. (2015). Examining the relations among student motivation, engagement, and retention in a mooc: A structural equation modeling approach. *Global Education Review*, *2*(3), 23–33.

Yang, D., Piergallini, M., Howley, I., & Rose, C. (2014). Forum thread recommendation for massive open online courses, In *Educational data mining 2014*, London, United Kingdom.

Zaiane, O. R. (2002). Building a recommender agent for e-learning systems, In *Proceedings of the international conference on computers in education*, Auckland, New Zealand.

Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM Comput. Surv.*, *52*(1), 1–38. https://doi.org/10.1145/3285029.

Zheng, L., Noroozi, V., & Yu, P. S. (2017). Joint deep modeling of users and items using reviews for recommendation, In *Proceedings of the tenth acm international conference on web search and data mining*, Cambridge, United Kingdom, Association for Computing Machinery. https://doi.org/10.1145/3018661.3018665.

Appendix A

Supplementary Statistical Tests

In Chapter 4.3.1, recommendation method was shown to be a significant factor in recommendation performance in the ANOVA tests. Therefore, I performed post-hoc tests to further explore their differences. Tukey's HSD tests were employed to make pairwise comparisons between the recommendation methods. As 7 out of the 8 ANOVA test results showed significant results in Table 4.7, the post-hoc analyses were run 7 times. The results are presented in the following tables.

Rec 1	Rec 2	meandiff	lower	upper	p-adj
CSCLRec	**CoPPR**	-0.0192	-0.1159	0.0775	0.9000
CSCLRec	**KCB**	-0.3523	-0.4490	-0.2556	0.0010
CSCLRec	**MCF**	-0.3018	-0.3986	-0.2051	0.0010
CSCLRec	**PPL**	-0.2727	-0.3694	-0.1760	0.0010
CSCLRec	**PPR**	-0.1849	-0.2817	-0.0882	0.0010
CSCLRec	**RND**	-0.4013	-0.4980	-0.3046	0.0010
CSCLRec	**SCB**	-0.3497	-0.4464	-0.2530	0.0010
CoPPR	**KCB**	-0.3332	-0.4299	-0.2364	0.0010
CoPPR	**MCF**	-0.2827	-0.3794	-0.1859	0.0010
CoPPR	**PPL**	-0.2535	-0.3502	-0.1568	0.0010
CoPPR	**PPR**	-0.1658	-0.2625	-0.0690	0.0010
CoPPR	**RND**	-0.3821	-0.4789	-0.2854	0.0010
CoPPR	**SCB**	-0.3305	-0.4272	-0.2338	0.0010
KCB	**MCF**	0.0505	-0.0462	0.1472	0.7102
KCB	**PPL**	0.0796	-0.0171	0.1764	0.1859
KCB	**PPR**	0.1674	0.0707	0.2641	0.0010
KCB	**RND**	-0.0490	-0.1457	0.0477	0.7383
KCB	**SCB**	0.0026	-0.0941	0.0993	0.9000
MCF	**PPL**	0.0292	-0.0676	0.1259	0.9000
MCF	**PPR**	0.1169	0.0202	0.2136	0.0073
MCF	**RND**	-0.0995	-0.1962	-0.0028	0.0393
MCF	**SCB**	-0.0479	-0.1446	0.0488	0.7595
PPL	**PPR**	0.0877	-0.0090	0.1845	0.1041
PPL	**RND**	-0.1286	-0.2254	-0.0319	0.0020
PPL	**SCB**	-0.0770	-0.1737	0.0197	0.2210
PPR	**RND**	-0.2164	-0.3131	-0.1197	0.0010
PPR	**SCB**	-0.1648	-0.2615	-0.0680	0.0010
RND	**SCB**	0.0516	-0.0451	0.1483	0.6891

Table A.1: Pairwise Tukey's HSD test among recommender methods for P@10 performance in short courses.

Rec 1	Rec 2	meandiff	lower	upper	p-adj
CSCLRec	**CoPPR**	-0.0113	-0.0537	0.0311	0.9000
CSCLRec	**KCB**	-0.0717	-0.1141	-0.0294	0.0010
CSCLRec	**MCF**	-0.0578	-0.1001	-0.0154	0.0014
CSCLRec	**PPL**	-0.0476	-0.0900	-0.0053	0.0165
CSCLRec	**PPR**	-0.0460	-0.0883	-0.0036	0.0239
CSCLRec	**RND**	-0.0737	-0.1161	-0.0313	0.0010
CSCLRec	**SCB**	-0.0697	-0.1121	-0.0273	0.0010
CoPPR	**KCB**	-0.0604	-0.1028	-0.0181	0.0010
CoPPR	**MCF**	-0.0465	-0.0888	-0.0041	0.0214
CoPPR	**PPL**	-0.0364	-0.0787	0.0060	0.1478
CoPPR	**PPR**	-0.0347	-0.0770	0.0077	0.1927
CoPPR	**RND**	-0.0624	-0.1048	-0.0200	0.0010
CoPPR	**SCB**	-0.0584	-0.1008	-0.0160	0.0012
KCB	**MCF**	0.0140	-0.0284	0.0563	0.9000
KCB	**PPL**	0.0241	-0.0183	0.0665	0.6256
KCB	**PPR**	0.0258	-0.0166	0.0682	0.5533
KCB	**RND**	-0.0019	-0.0443	0.0404	0.9000
KCB	**SCB**	0.0020	-0.0403	0.0444	0.9000
MCF	**PPL**	0.0101	-0.0322	0.0525	0.9000
MCF	**PPR**	0.0118	-0.0306	0.0542	0.9000
MCF	**RND**	-0.0159	-0.0583	0.0265	0.9000
MCF	**SCB**	-0.0119	-0.0543	0.0304	0.9000
PPL	**PPR**	0.0017	-0.0407	0.0441	0.9000
PPL	**RND**	-0.0260	-0.0684	0.0163	0.5425
PPL	**SCB**	-0.0221	-0.0644	0.0203	0.7129
PPR	**RND**	-0.0277	-0.0701	0.0146	0.4687
PPR	**SCB**	-0.0237	-0.0661	0.0186	0.6406
RND	**SCB**	0.0040	-0.0384	0.0464	0.9000

Table A.2: Pairwise Tukey's HSD test among recommender methods for R@10 performance in short courses.

Rec 1	Rec 2	meandiff	lower	upper	p-adj
CSCLRec	**CoPPR**	-0.0350	-0.1051	0.0351	0.7522
CSCLRec	**KCB**	-0.0030	-0.0732	0.0671	0.9000
CSCLRec	**MCF**	0.1686	0.0985	0.2388	0.0010
CSCLRec	**PPL**	0.1192	0.0490	0.1893	0.0010
CSCLRec	**PPR**	0.0514	-0.0188	0.1215	0.3193
CSCLRec	**RND**	0.1600	0.0898	0.2301	0.0010
CSCLRec	**SCB**	-0.1128	-0.1829	-0.0426	0.0010
CoPPR	**KCB**	0.0320	-0.0382	0.1021	0.8302
CoPPR	**MCF**	0.2036	0.1335	0.2738	0.0010
CoPPR	**PPL**	0.1542	0.0840	0.2243	0.0010
CoPPR	**PPR**	0.0864	0.0162	0.1565	0.0058
CoPPR	**RND**	0.1950	0.1248	0.2651	0.0010
CoPPR	**SCB**	-0.0778	-0.1479	-0.0076	0.0191
KCB	**MCF**	0.1716	0.1015	0.2418	0.0010
KCB	**PPL**	0.1222	0.0520	0.1923	0.0010
KCB	**PPR**	0.0544	-0.0157	0.1246	0.2505
KCB	**RND**	0.1630	0.0928	0.2331	0.0010
KCB	**SCB**	-0.1098	-0.1799	-0.0396	0.0010
MCF	**PPL**	-0.0495	-0.1196	0.0207	0.3693
MCF	**PPR**	-0.1172	-0.1874	-0.0471	0.0010
MCF	**RND**	-0.0087	-0.0788	0.0615	0.9000
MCF	**SCB**	-0.2814	-0.3516	-0.2113	0.0010
PPL	**PPR**	-0.0678	-0.1379	0.0024	0.0661
PPL	**RND**	0.0408	-0.0294	0.1109	0.6024
PPL	**SCB**	-0.2319	-0.3021	-0.1618	0.0010
PPR	**RND**	0.1086	0.0384	0.1787	0.0010
PPR	**SCB**	-0.1642	-0.2343	-0.0940	0.0010
RND	**SCB**	-0.2727	-0.3429	-0.2026	0.0010

Table A.3: Pairwise Tukey's HSD test among recommender methods for ILD performance in short courses.

Rec 1	**Rec 2**	**meandiff**	**lower**	**upper**	**p-adj**
CSCLRec	**CoPPR**	0.0191	-0.1636	0.2018	0.9000
CSCLRec	**KCB**	0.5480	0.3653	0.7307	0.0010
CSCLRec	**MCF**	0.3207	0.1380	0.5034	0.0010
CSCLRec	**PPL**	-0.1321	-0.3148	0.0505	0.3355
CSCLRec	**PPR**	-0.0763	-0.2590	0.1064	0.8999
CSCLRec	**RND**	0.7109	0.5282	0.8936	0.0010
CSCLRec	**SCB**	0.4390	0.2563	0.6217	0.0010
CoPPR	**KCB**	0.5289	0.3462	0.7116	0.0010
CoPPR	**MCF**	0.3016	0.1189	0.4843	0.0010
CoPPR	**PPL**	-0.1513	-0.3340	0.0314	0.1806
CoPPR	**PPR**	-0.0954	-0.2781	0.0873	0.7100
CoPPR	**RND**	0.6918	0.5091	0.8745	0.0010
CoPPR	**SCB**	0.4199	0.2372	0.6026	0.0010
KCB	**MCF**	-0.2273	-0.4099	-0.0446	0.0051
KCB	**PPL**	-0.6801	-0.8628	-0.4974	0.0010
KCB	**PPR**	-0.6243	-0.8070	-0.4416	0.0010
KCB	**RND**	0.1629	-0.0198	0.3456	0.1167
KCB	**SCB**	-0.1090	-0.2917	0.0737	0.5754
MCF	**PPL**	-0.4529	-0.6356	-0.2702	0.0010
MCF	**PPR**	-0.3970	-0.5797	-0.2143	0.0010
MCF	**RND**	0.3902	0.2075	0.5729	0.0010
MCF	**SCB**	0.1183	-0.0644	0.3010	0.4819
PPL	**PPR**	0.0559	-0.1268	0.2386	0.9000
PPL	**RND**	0.8431	0.6604	1.0258	0.0010
PPL	**SCB**	0.5712	0.3885	0.7539	0.0010
PPR	**RND**	0.7872	0.6045	0.9699	0.0010
PPR	**SCB**	0.5153	0.3326	0.6980	0.0010
RND	**SCB**	-0.2719	-0.4546	-0.0892	0.0010

Table A.4: Pairwise Tukey's HSD test among recommender methods for MIUF performance in short courses.

Rec 1	Rec 2	meandiff	lower	upper	p-adj
CSCLRec	**CoPPR**	-0.0125	-0.1080	0.0830	0.9000
CSCLRec	**KCB**	-0.4498	-0.5453	-0.3543	0.0010
CSCLRec	**MCF**	-0.2422	-0.3378	-0.1467	0.0010
CSCLRec	**PPL**	-0.3324	-0.4279	-0.2368	0.0010
CSCLRec	**PPR**	-0.1841	-0.2796	-0.0885	0.0010
CSCLRec	**RND**	-0.4264	-0.5219	-0.3309	0.0010
CSCLRec	**SCB**	-0.4354	-0.5309	-0.3399	0.0010
CoPPR	**KCB**	-0.4373	-0.5328	-0.3418	0.0010
CoPPR	**MCF**	-0.2297	-0.3252	-0.1342	0.0010
CoPPR	**PPL**	-0.3198	-0.4153	-0.2243	0.0010
CoPPR	**PPR**	-0.1715	-0.2670	-0.0760	0.0010
CoPPR	**RND**	-0.4139	-0.5094	-0.3183	0.0010
CoPPR	**SCB**	-0.4228	-0.5184	-0.3273	0.0010
KCB	**MCF**	0.2076	0.1121	0.3031	0.0010
KCB	**PPL**	0.1175	0.0220	0.2130	0.0051
KCB	**PPR**	0.2658	0.1703	0.3613	0.0010
KCB	**RND**	0.0234	-0.0721	0.1190	0.9000
KCB	**SCB**	0.0145	-0.0811	0.1100	0.9000
MCF	**PPL**	-0.0901	-0.1856	0.0054	0.0806
MCF	**PPR**	0.0582	-0.0373	0.1537	0.5673
MCF	**RND**	-0.1841	-0.2797	-0.0886	0.0010
MCF	**SCB**	-0.1931	-0.2886	-0.0976	0.0010
PPL	**PPR**	0.1483	0.0528	0.2438	0.0010
PPL	**RND**	-0.0940	-0.1895	0.0015	0.0573
PPL	**SCB**	-0.1030	-0.1985	-0.0075	0.0245
PPR	**RND**	-0.2423	-0.3378	-0.1468	0.0010
PPR	**SCB**	-0.2513	-0.3468	-0.1558	0.0010
RND	**SCB**	-0.0090	-0.1045	0.0865	0.9000

Table A.5: Pairwise Tukey's HSD test among recommender methods for P@10 performance in long courses.

Rec 1	Rec 2	meandiff	lower	upper	p-adj
CSCLRec	**CoPPR**	-0.0590	-0.1379	0.0199	0.3061
CSCLRec	**KCB**	-0.0569	-0.1371	0.0232	0.3746
CSCLRec	**MCF**	0.1789	0.1000	0.2578	0.0010
CSCLRec	**PPL**	0.1470	0.0682	0.2259	0.0010
CSCLRec	**PPR**	0.1198	0.0409	0.1987	0.0010
CSCLRec	**RND**	0.1389	0.0600	0.2178	0.0010
CSCLRec	**SCB**	-0.2096	-0.2885	-0.1307	0.0010
CoPPR	**KCB**	0.0021	-0.0781	0.0822	0.9000
CoPPR	**MCF**	0.2378	0.1590	0.3167	0.0010
CoPPR	**PPL**	0.2060	0.1271	0.2849	0.0010
CoPPR	**PPR**	0.1788	0.0999	0.2577	0.0010
CoPPR	**RND**	0.1978	0.1189	0.2767	0.0010
CoPPR	**SCB**	-0.1506	-0.2295	-0.0717	0.0010
KCB	**MCF**	0.2358	0.1556	0.3159	0.0010
KCB	**PPL**	0.2040	0.1238	0.2841	0.0010
KCB	**PPR**	0.1767	0.0966	0.2569	0.0010
KCB	**RND**	0.1958	0.1156	0.2759	0.0010
KCB	**SCB**	-0.1527	-0.2328	-0.0725	0.0010
MCF	**PPL**	-0.0318	-0.1107	0.0471	0.9000
MCF	**PPR**	-0.0591	-0.1380	0.0198	0.3040
MCF	**RND**	-0.0400	-0.1189	0.0389	0.7523
MCF	**SCB**	-0.3885	-0.4673	-0.3096	0.0010
PPL	**PPR**	-0.0273	-0.1062	0.0516	0.9000
PPL	**RND**	-0.0082	-0.0871	0.0707	0.9000
PPL	**SCB**	-0.3566	-0.4355	-0.2777	0.0010
PPR	**RND**	0.0191	-0.0598	0.0980	0.9000
PPR	**SCB**	-0.3294	-0.4083	-0.2505	0.0010
RND	**SCB**	-0.3484	-0.4273	-0.2696	0.0010

Table A.6: Pairwise Tukey's HSD test among recommender methods for ILD performance in long courses.

Rec 1	Rec 2	meandiff	lower	upper	p-adj
CSCLRec	**CoPPR**	0.0305	-0.2013	0.2624	0.9000
CSCLRec	**KCB**	0.4166	0.1848	0.6485	0.0010
CSCLRec	**MCF**	0.2247	-0.0071	0.4565	0.0652
CSCLRec	**PPL**	-0.2106	-0.4424	0.0212	0.1056
CSCLRec	**PPR**	-0.1668	-0.3986	0.0651	0.3569
CSCLRec	**RND**	0.5543	0.3225	0.7861	0.0010
CSCLRec	**SCB**	0.5808	0.3489	0.8126	0.0010
CoPPR	**KCB**	0.3861	0.1543	0.6179	0.0010
CoPPR	**MCF**	0.1941	-0.0377	0.4260	0.1757
CoPPR	**PPL**	-0.2412	-0.4730	-0.0093	0.0350
CoPPR	**PPR**	-0.1973	-0.4291	0.0345	0.1604
CoPPR	**RND**	0.5238	0.2920	0.7556	0.0010
CoPPR	**SCB**	0.5502	0.3184	0.7821	0.0010
KCB	**MCF**	-0.1920	-0.4238	0.0399	0.1872
KCB	**PPL**	-0.6273	-0.8591	-0.3954	0.0010
KCB	**PPR**	-0.5834	-0.8152	-0.3516	0.0010
KCB	**RND**	0.1377	-0.0941	0.3695	0.5951
KCB	**SCB**	0.1641	-0.0677	0.3960	0.3789
MCF	**PPL**	-0.4353	-0.6671	-0.2035	0.0010
MCF	**PPR**	-0.3914	-0.6233	-0.1596	0.0010
MCF	**SCB**	0.3561	0.1243	0.5879	0.0010
PPL	**PPR**	0.0439	-0.1880	0.2757	0.9000
PPL	**RND**	0.7649	0.5331	0.9967	0.0010
PPL	**SCB**	0.7914	0.5596	1.0232	0.0010
PPR	**RND**	0.7211	0.4893	0.9529	0.0010
PPR	**SCB**	0.7475	0.5157	0.9794	0.0010
RND	**SCB**	0.0265	-0.2054	0.2583	0.9000

Table A.7: Pairwise Tukey's HSD test among recommender methods for MIUF performance in long courses.

Appendix B

Research Ethics Board Approval Letter

The study went through the review of the research ethics board. The approval letter is shown in B.1.

Notification of Approval

Date:	May 15, 2018
Study ID:	Pro00078624
Principal Investigator:	Carrie Demmans Epp
Study Title:	Understanding and Evaluating E-Learning System Use: Modelling Learners, Describing Learning, and Developing Features
Approval Expiry Date:	Tuesday, May 14, 2019
Sponsor/Funding Agency:	NSERC - Natural Sciences And Engineering Research Council NSERC
Sponsor/Funding Agency:	University of Alberta, Faculty of Science

RSO-Managed Funding:

Project ID	Project Title	Speed Code	Other Information
RES0038658	Evidence-based adaptation of Learning: Supporting informed decision making		[redacted]
RES0039792	University of Alberta Startup Funds	[redacted]	

Thank you for submitting the above study to the Research Ethics Board 2. Your application has been reviewed and approved on behalf of the committee.

A renewal report must be submitted next year prior to the expiry of this approval if your study still requires ethics approval. If you do not renew on or before the renewal expiry date, you will have to re-submit an ethics application.

Approval by the Research Ethics Board does not encompass authorization to access the staff, students, facilities or resources of local institutions for the purposes of the research.

Sincerely,

Stanley Varnhagen, PhD
Chair, Research Ethics Board 2

Note: This correspondence includes an electronic signature (validation and approval via an online system).

Figure B.1: Research ethics board approval letter (speed codes redacted).