SUMMER 2019 MIP PROPOSAL
A DETAILED STUDY ON RECOMMENDER SYSTEMS

FAHMIDA HAMID

Department of Computer Science,
Grinnell College,
Grinnell, IA 50112

Abstract. I plan to conduct two related research projects with a set of first and/or second-year students over this summer in order to introduce them to a relatively new research field. Over the ten-week period, students will investigate the full particulars of recommender systems and their evaluation mechanisms on different datasets.

1. Overview

I propose two related projects for several students to participate in a 299/MIP for summer 2019. The projects require a basic foundation in computer science and/or mathematics.

My projects will be on designing automatic recommender systems and exploring their evaluation mechanisms for different application domains. Designing an automatic recommender system is one of the most recent and popular research tracks in the field of Information retrieval. Students will explore the available datasets and implement the existing best-performing systems. Then they will study the performance of each of those systems in at least two different domains. In one project, students will discover new properties or clues that can improve the performance of the systems and make them less domain specific. In the second project, students will focus on the existing evaluation mechanisms for the recommender systems and suggest/report modifications to build a more robust evaluation metric. We plan to use text-based information.

I have done a fair amount of research in text-based information retrieval (summarization, keyphrase extraction, evaluation of automatic summarization and keyphrase extraction systems) during my Ph.D. studies. I have been studying the techniques to develop recommender systems for the last couple of months and have attended SIGIR 2018 (the major conference for IR researchers). I plan to use the techniques that I have learned from my past research in these projects.

I am interested in hiring up to four 299 students, particularly from underrepresented groups in CS, to introduce them to empirical computing research on important problems. Hence, the proposed budget is for four students at $3,400 each, for a total of $13,600.

2. Project Description

It is often necessary to make choices without sufficient personal experience of the alternatives. In everyday life, we rely on recommendations from other people either by word of mouth, recommendation letters, movie and book reviews printed in newspapers, or general surveys [1]. With the advent of world wide web, we are now living the age of information: we have numerous sites and personal blogs where people post their opinions on different items/products/venues/events and so on. Different organizations advertise their goods/events/resources through several sites. It is interesting to both of the parties (end user and producer) if some automated systems could integrate these pieces of information from several sources and present them

E-mail address: hamidfah@grinnell.edu.

in a useful way. Starting from Youtube, Netflix, Amazon, to big research institutions, almost all are now using different types of recommender systems to facilitate themselves and their clients.

The goal of recommender systems is to recommend a list of relevant information or tools or items based on the users' requirements. Designing such an automatic system is very challenging due to several reasons: the accuracy of a recommended list depends not only on the mood of the user or the quality and attributes of the items but it also depends on time. For example, collaborative recommender systems which depend on the similar users' past history to prepare the output struggle with a new user. On the other hand, content-based recommender systems suffer if the attributes are not adequately described [2]. Sometimes completely random selection might show similar performance as the some of the best performing algorithms. We plan to study these challenges in detail.

As the students get familiar with the systems we plan on discovering new techniques to enhance the performance of the systems. We plan to explore both supervised and unsupervised techniques to get the maximum exposure to the datasets and to the systems. We also plan to focus on automatically detecting the significant attributes of the items and the users. For example, the attributes of a collection of books (topic, price, author, time of publishing, language, context, etc.) are quite different from the attributes of clothes (price, color, brand, style, length, age of the client, etc.). Hence, it is an extremely challenging job to automatically detect/adjust the dominant features of the items on different domains. If we can define a mathematical model to automatically adjust (or discover) the dominant features of a collection for a set of users, we may have a successful recommender model that will work across different domains.

3. Student Preparation

Students who have completed our department's introductory and/or advanced course(s) (like CSC151, CSC161, CSC 207) will have a basic preparation for the projects. As this is a special research track, students will have to study a couple of articles to establish their preliminary understanding.

4. Outcomes & Timeline

Students will begin laying the groundwork for these projects with me by the end of April. Students will be working in pairs to facilitate mutual cooperation and support.
The following is an approximate timeline for the summer:

Week 1: begin writing introductions and background for MIP report; background and exploration on models for their respective tasks
Week 2: explore algorithm behavior
Week 3: begin investigating alternative methods
Week 4: develop software for additional behavioral exploration
Week 5: fine-tune model and parameters
Weeks 6-7: Set up experiments (generate software and/or data), and begin running experiments
Weeks 8-9: Preliminary analysis, continuing and/or modifying experiments
Week 10: Conclude experiments: integrate final report with results and conclusions

In addition to the 10-week schedule, students will be expected to present and distribute their work in several venues.

- We will aim for a short paper or a poster for the upcoming RecSys/SIGIR/ECIR (2020). Deadlines to these conferences are usually by the mid of January or April. Based on the outcomes, we will decide what can be the best possible event to submit the students work:
- We will also present the work at the Department of Computer Science Thursday Extras weekly seminar series.

The following page contains (will be updated as time progresses) details of the proposal: `https://www.cs.grinnell.edu/~hamidfah/summer_research/`

References

[1] Paul Resnick and Hal R. Varian. Recommender systems. *Commun. ACM*, 40(3):56–58, March 1997.
[2] Francesco Ricci, Lior Rokach, and Bracha Shapira. Recommender systems: introduction and challenges. In *Recommender systems handbook*, pages 1–34. Springer, 2015.