University Degree Recommendation System

Islam Ehab Anwar

Faculty of Computer Science

University of Prince Edward Island, Cairo Campus

June 2023

A thesis submitted to McGill University in partial fulfillment of the requirements

of the degree of Bachelor of Science

Acknowledgements

Words cannot express my appreciation and gratitude towards my professor, project adviser and Dean of the faculty of computer science Dr. Ahmed Elsheikh for his unwavering support and guidance. Additionally, this would not have been possible without my research assistants Heba Abdelkader and Ola Galal whom without their guidance and experience I would not have been able to achieve this. Furthermore, I would like to thank the university for providing the necessary data for this project.

I am also grateful to Dr. Amal Mohamed, professor at the faculty of computer science for her unconditional moral support and significant impact in inspiring me to pursue this field of study.

Lastly, I would be remiss in not mentioning my family, especially my parents and brother. Their belief and encouragement has kept my spirits and motivation high during this trying process. I would never have reached this stage without them.

Abstract

Abrégé

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

With the accelerating advancement of internet technology, there has been a recent overload of information available to online users such as news content, e-commerce products and services, digital entertainment, and so on (1) (2). A common issue arising from this predicament is the inability to find the information that is desired by a user, whether that be for education, entertainment, or regarding a business application. Recommender Systems (RS) are a potential solution to this problem that aims to offer a subset of items or a collection of data that a corresponding user might be interested in (1).

In the previous 20 years, most RS has been used with a focus on business applications with strong efforts catering to a user-focused recommendation system (3). One of those applications that have sky-rocketed in success is Netflix recommends movies and TV shows based on your history of watching shows or films. Furthermore, these recommendations can be enhanced further by integrating ratings as a part of the recommendation algorithm to improve sophistication (4).

A multitude of RS exists today that provide meaningful recommendations. However, there are various methods in which RS are created with unique architectures and designs that favor the application they serve. One of the most popular recommendation techniques is collaborative filtering (CF), which works by representing unique users and items with unique representation vectors. These representation vectors' interactions can be analyzed using various machine learning approaches such as Neural Networks (NN) (1). However, a disadvantage of CF recommendation is suffering from data sparsity and the cold start issue, hence other techniques have been developed to tackle these problems (2).

Another common recommendation technique is Knowledge Graphs (KG). KG are heterogeneous graphs, whose nodes represent the unique users, and the arcs are the relations between the user and the corresponding items. KG recommendations have been widely used recently such as Freebase and Google's Knowledge Graph. The added value of KG is that user-item links have different latent relations, which can improve recommendation proficiency. KG can be designed in a variety of ways, including embedding-based methods, path-based methods, as well as unified methods which combine the former techniques to build the latter (2). There is a large number of other recommendation techniques including content-based, demographic-based, utility-based recommendations, etc. (5).

A student degree recommendation system is a problem that has not been tackled by many recent publishers. With the vast diversity of university programs available to prospective students, many are left wondering what their real passion is or if a particular degree is the one for them. A recommendation algorithm is suitable to mitigate this issue, by analyzing student academic performance during the latter end of high school as well as other factors such as familial status, personal status, and other daily lifestyle choices that could shape the overall performance of a student and impact the recommended degrees.

Many recent papers have struggled to use recommendation systems to address issues such as the student degree recommendation due to various reasons including the cold start problem. This problem significantly impacts the algorithm performance when there is an insufficient amount of data in the early stages for the recommendation algorithm to work as intended. Another issue at hand is data sparsity. Users tend to only interact with certain items, leaving other items completely blank and it can affect the recommendation process in the long term as items with no relations at all will not be recommended at all (5).

This paper aims to provide a student degree recommendation system that can accurately suggest a collection of degrees that correspond to a set of data features including the type of subjects taken in high school as well as final grades of those respective subjects, familial status, number of hours spent going out and more. Machine learning algorithms will be deployed as well as optimized to identify student performances and then enhanced further to provide predictions on the degree(s) of choice.

The rest of the paper is as follows: Chapter 2 describes recent existing work regarding recommendation systems. Chapter 3 provides a background on the methodology taken to reach a viable recommendation system. Chapter 4 discusses the environment setup, diving into steps taken including feature importance, hyper-parameter optimization, and so on. Chapter 5 provides a discussion of results obtained. Future works and conclusions are highlighted in Chapter 6.

Chapter 2

Literature Review

There has been extensive recent research performed in the domain of recommender systems. H. Wang et al. (1) proposed a knowledge graph convolutional network. The purpose of this architecture was to capture relationships between several items of interest to a user through data mining techniques. Some of the featured techniques involved association rules and others in order to identify relations between attributes on a graph. This was performed by sampling data from respective neighbors per entity in a particular graph and fusing the information already gained with the bias in order to calculate an accurate representation of each entity's graph relations. Three datasets were used for testing including movie, book and music recommendation datasets. Results indicated that the proposed netwrok outperformed the baseline recommendation techniques with an Area Under the Curve (AUC) of 0.9 or higher with two of the three datasets (1).

Content-based recommender systems are amongst another popular technique used to gather information about user preferences. They are based on past preferences of users and recommendations are suggested with similar items of similar characteristics (6). Chen et al. (7) used correlation analysis in an experiment regarding the education field in order to group certain courses. This was performed by segmenting the data into three categories based on a rule-space model using a content-based approach to optimize the learning path for each individual. Similarly, Shu et al. (8) utilised the historical data of students to create predictions regarding the provided learning materials using a content-based algorithm. The most common learning algorithms used in this domain are fuzzy-based as well as rule-based clustering that relies on a probabilistic methodology, and similarity between neighbors. However, a disadvantage of content-based recommendation is the overreliance on past data to create predictions. It cannot overcome the cold start problem of having no data in the initial stages as well as data sparsity (6).

Another favored technique in recommendation is collaborative filtering, which has been used frequently in recent systems as it mitigates the drawbacks of content-based filtering as we discussed earlier (6). Liu (9) recommended a collaborative filtering approach that focusedon the influence of e-learning group behavior to improve the accuracy of predictions even in presence of data sparsity. Moreover, collorative filtering has been used with unsupervised learning. El-Bishouty et al. (10) utilised a k-means algorithm to extract the learning path and objects of interest for each individual learner. Collaborative filtering does improve on content-based systems, but it still comes with its own set of difficulties. It can be difficult to create relations between attributes to their respective items, which can affect recommendation accuracy. It also suffers from cold-start and scalability issues (6).

Certain hybrid methods have been experimented with including the combination of content-based and collaborative filtering to counter the cold start problem (6). Hussain et al. (11) opted to use a collection of models including an artificial neural network, decision tree, logistic regression, and support vector machine to predict the troubles students face during an online learning course. Similarly, Karga and Satratzemi (12) used a similarity matrix to create the relations between learners and their respective learning paths according to their needs and preferences. This methodology allows both content-based and collaborative filtering methods to complement each others weaknesses while improving prediction accuracy (6).

Finally, knowledge-based approaches provide recommendations based on how certain item features meet user needs. Wan and Niu (13) used a knowledge-based approach with an underlying self-organization method to propose learning objects. This can improve accuracy but suffers from increase time for computations and stacking of multiple algorithms.

Many of these recommendation techniques have specific advantages and disadvantages. However, hybrid techniques seem to outperform them due its combination of the individual techniques to overcome common adversities in recommendation systems (6).

Chapter 3

Necessary Background

Chapter 4

Experimental Setup

Chapter 5

Results and Discussion

Chapter 6

Conclusion and Future Work

Bibliography

[1] H. Wang, M. Zhao, X. Xie, W. Li, and M. Guo, "Knowledge graph convolutional networks for recommender systems," in *The world wide web conference*, pp. 3307–3313, 2019.

[2] Q. Guo, F. Zhuang, C. Qin, H. Zhu, X. Xie, H. Xiong, and Q. He, "A survey on knowledge graph-based recommender systems," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 8, pp. 3549–3568, 2020.

[3] S. Milano, M. Taddeo, and L. Floridi, "Recommender systems and their ethical challenges," *Ai & Society*, vol. 35, pp. 957–967, 2020.

[4] N. Hug, "Surprise: A python library for recommender systems," *Journal of Open Source Software*, vol. 5, no. 52, p. 2174, 2020.

[5] Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim, and R. Kashef, "Recommendation systems: Algorithms, challenges, metrics, and business opportunities," *applied sciences*, vol. 10, no. 21, p. 7748, 2020.

[6] S. S. Khanal, P. Prasad, A. Alsadoon, and A. Maag, "A systematic review: machine learning based recommendation systems for e-learning," *Education and Information Technologies*, vol. 25, pp. 2635–2664, 2020.

[7] Y.-H. Chen, C.-H. Tseng, C.-L. Huang, L. Y. Deng, and W.-C. Lee, "Recommendation system based on rule-space model of two-phase blue-red tree and optimized learning path with multimedia learning and cognitive assessment evaluation," *Multimedia Tools and Applications*, vol. 76, pp. 18237–18264, 2017.

[8] J. Shu, X. Shen, H. Liu, B. Yi, and Z. Zhang, "A content-based recommendation algorithm for learning resources," *Multimedia Systems*, vol. 24, no. 2, pp. 163–173, 2018.

[9] X. Liu, "A collaborative filtering recommendation algorithm based on the influence sets of e-learning group's behavior," *Cluster Computing*, vol. 22, no. Suppl 2, pp. 2823–2833, 2019.

[10] M. M. El-Bishouty, A. Aldraiweesh, U. Alturki, R. Tortorella, J. Yang, T.-W. Chang, S. Graf, *et al.*, "Use of felder and silverman learning style model for online course design," *Educational Technology Research and Development*, vol. 67, no. 1, pp. 161–177, 2019.

[11] M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, and S. Ali, "Using machine learning to predict student difficulties from learning session data," *Artificial Intelligence Review*, vol. 52, pp. 381–407, 2019.

[12] S. Karga and M. Satratzemi, "A hybrid recommender system integrated into lams for learning designers," *Education and Information Technologies*, vol. 23, pp. 1297–1329, 2018.

[13] S. Wan and Z. Niu, "An e-learning recommendation approach based on the self-organization of learning resource," *Knowledge-Based Systems*, vol. 160, pp. 71–87, 2018.