Course Recommendation System

Team members
1) Khaled Hossain - khaled.hossain@colorado.edu
2) Jahoon Koo - jahoon.koo@colorado.edu
3) Jusung Kim - itmekjs@gmail.com
4) Carter Andrew - carter.andrew@colorado.edu

Our code can be accessed here: GitHub

Introduction

The purpose of our project proposal is to develop a recommendation system for the University of Colorado Boulder courses. Our main objective is to enhance the course selection process for students by providing personalized recommendations based on their academic interests, past performance, and feedback from previous students. This recommendation system will help students make better decisions about their course selection and have a more satisfactory experience. **Update:** Furthermore, our model can also be used for predicting the scores a student may get based on the attributes. This will be useful for the evaluation of our model and also allow teachers to give more time to students who may receive a low score.

Related Work & Literature Review:

Update: We have found a paper discussing a course recommender system built by Yinping Ma that includes a very large dataset of courses and associated student attendance and scores from Peking University. This dataset will be instrumental in our project and be the source for both our training and testing datasets.

We have done a comprehensive review of existing recommendation systems that are currently being used in the educational domain. Our review included numerous topics, such as collaborative filtering techniques, content-based filtering approaches, and hybrid models. By reading the papers we were able to identify how to proceed with our final project. Here are the papers and this is what we learned from them:

1) Comparison of User-Based and Item-Based Collaborative Filtering Recommendation Services.

In this paper, the authors present a comprehensive analysis and comparison of User-Based and Item-Based Collaborative Filtering Recommendation Services, which are considered the cornerstone of recommendation systems. The paper thoroughly explores and substantiates the superiority of user-based collaborative filtering over item-based approaches by conducting two dataset experiments using Adjusted Cosine Similarity and Root Mean Square Estimation (RMSE). Additionally, the authors shed light on five inherent challenges associated with collaborative filtering systems. **Update:** The goal of our project is to compare the performance of different collaborative filtering systems and this paper will allow us to compare our results.

2) Intelligent Recommendation System for Course Selection in Smart Education [4]

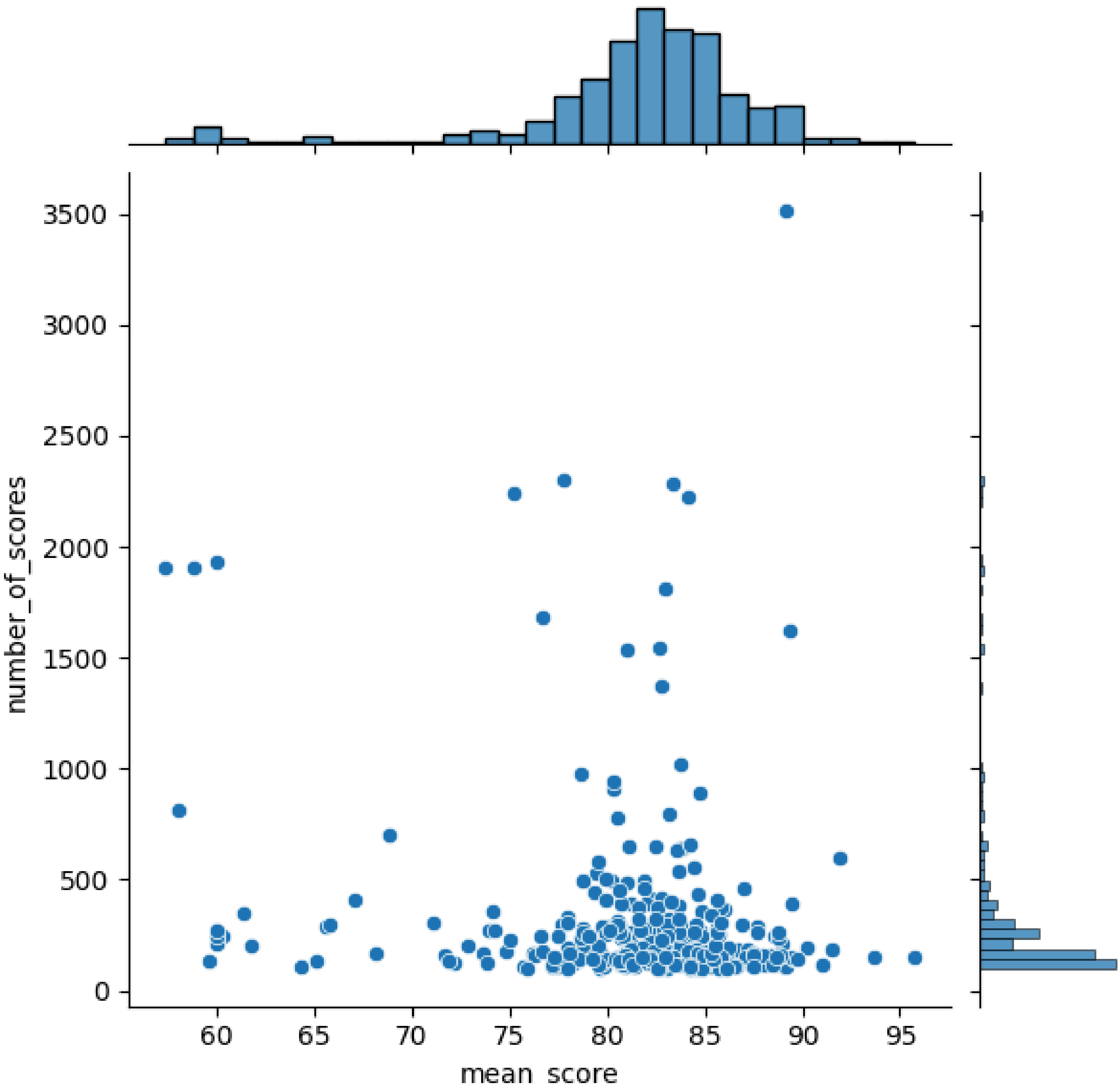
The authors of this paper proposed a novel recommendation system for course selection in the specialty of information management in Chinese Universities. They used the Sparse Linear method (SLIM) to generate top-N recommendations of courses that are appropriate for the students. From this paper, we were also able to learn about the three systems recommendation systems are generally categorized into collaborative, content-based, and knowledge-based. For evaluation, they used the hit rate and the average reciprocal Hit-Rank. After comparing it with existing state-of-the-art models [1] they were able to showcase better results using their strategy of using sparseness aggregation strategy.

3) A Book Recommender System Using Collaborative Filtering Method [3]

A book recommender system aims to provide personalized recommendations for Arab readers. The system utilizes user-based and item-based collaborative filtering and matrix factorization with a Singular Value Decomposition (SVD) algorithm. Accuracy is evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), precision, and recall. At the same time, training and testing times are compared to the K-Nearest Neighbor (KNN) based model. The KNN-based algorithm performs better in terms of speed, but the SVD algorithm achieves the highest accuracy.

4) A non-IID Framework for Collaborative Filtering with Restricted Boltzmann Machines [2]

This paper discusses the use of a hybrid user-user and item-item collaborative filtering-based recommendation system. To achieve this the user-item matrix is considered as a singular input



into a RBM training model with two hidden layers created for processing each dimension of the

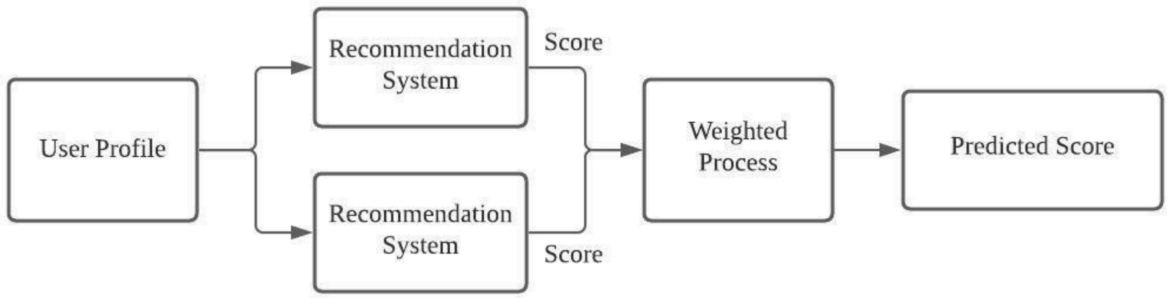
data. By expanding on Restricted Boltzmann Machines (RBMs) we can produce models that are as accurate as previously proposed models while using less data.
The reason the RBM can see this kind of accuracy is that it can be used to generate data that can then be combined to produce more generalizable results. It is in essence a very complex form of bootstrapping that could be useful for gaining insights into more readily available datasets for our project.

Proposed Work:

Update: Our proposed work involves the following steps:
a) Datasets: For the purpose of analyzing students' academic success, this research paper collected an anonymized dataset from Peking University between 2014 to 2021. The dataset comprises 4568 students, 5591 courses, and a total of 208949 course enrollments. Each student is accompanied by education (e.g. undergrad or postgrad) and major. A course enrollment indicates that a student completed the class and has a student score. The course data covers 53 departments within Peking University. Each course contains a course introduction that provides a concise description of the course.
b) Data Preprocessing: After collecting our dataset we had to do some data cleaning and preprocessing to cast it into the correct format. We began by dropping all the rows which didn't have any values, and also unnecessary columns that do not add any value to our project. The plot shows the sparsity of our dataset after filtering the course dataset to include data with a score count exceeding 100. We have the mean score on the x-axis and the number of scores on the y-axis.

c) Knowledge Extraction: We will use a hybrid collaborative filtering method (Talked about more extensively in part d) combining and weighting different outputs for predictive models and item-item similarity measures to get a final ranking that is both accessible to accuracy evaluation and produces interesting results. We plan to use the following metrics for knowledge extraction: user-user collaborative filtering for the grade of a student in a given course normalized by their

GPA, the cosine difference of a vectorized course introduction, and if we have time collaborative filtering for determining if challenging or easier courses should be recommended. Another benefit of this approach is we can use a regression model to change the weighing of our metrics to produce results that favor strong recommendation metrics for that specific query.



d) Recommendation Algorithm Development: Our initial plan was to utilize a collaborative filtering algorithm to implement our course recommendation system. We have decided to compare the performance of 3 kinds of collaborative filtering methods, which are: user-based filtering, item-based filtering, and lastly, a hybrid model. User-based collaborative filtering focuses on finding similarities between students based on their past interactions or preferences. Item-based collaborative filtering identifies similarities between courses based on the preferences of students who have taken the course. In a hybrid recommendation system, predictions from user-based and item-based filtering models are combined by taking the weighted average of their predictions. As part of our plan, we intend to build a regression model that will determine the optimal weights for each model.

$$Sim(a,b) = \frac{\sum_p (r_{ap}-\bar{r}_a)(r_{ab}-\bar{r}_b)}{\sqrt{\sum (r_{ap}-\bar{r}_a)^2}\sqrt{\sum (r_{bp}-\bar{r}_b)^2}}$$

$r_{up}: rating\ of\ user\ u\ against\ item\ p$

$p: items$

Milestones

	6/26/2023	7/3/2023	7/10/2023	7/17/2023	7/24/2023
Prior Investigations					
Data Collection					
Data Preprosessing and Feedback					
Model Training					
Model Testing and Evaluation					

Changes: In the light of the dataset we have acquired we can make confident user-user and item-item collaborative filtering systems. This does not change much from our original plan in terms of strategy. The most notable difference is our method for evaluating accuracy. Because we have continuous data for our label we can have a much better indicator if our algorithm is overfitting. Because our dataset is sufficiently large we can freely drop missing data points without losing a significant amount of our dataset. This means we can avoid sparsity issues and simplify our data preprocessing significantly.

Metrics & Evaluations:

To evaluate the effectiveness of the recommendation system, we will be comparing the outputs of a predictive model for the score (normalized by student GPA) a student will receive in their class, we will then feed partial student careers into our model and find the difference between real and predicted values. Because we are using continuous data for these scores, the difference is simple to calculate. We will compute similar metrics for other predictive values used for the final ranking. For example, when evaluating how relevant a class is to a given student's career we will compare it to a binary "probability" of 1 or 0 for if the course was in the test student's career. Lastly, we will compare our predicted scores with the test dataset and calculate the Mean Squared error which calculates the average difference between predicted and actual values and also other metrics suitable for continuous data.

Challenges

Initially, our intention was to utilize aggregated data about CU boulder classes or the Coursera Course dataset from Kaggle. Unfortunately, these datasets solely provide information about courses and lack user datasets, restricting us to constructing only a content-based filtering model.

However, content-based filtering is a form of unsupervised learning without a ground truth, making evaluation challenging. As a solution, we have discovered student and course datasets from Peking University that include scores. By using this data, we can now develop a collaborative filtering model capable of predicting scores, which enables us to conduct evaluations effectively.

Conclusion

Our recommendation system will be a huge benefit for students, teachers, and parents. Initially, we had planned to create a recommendation system only for the students of Peking University. But due to the challenges of finding a suitable dataset and decided to expand on our initial plan to make the recommendation system more general so that it can be used by a larger audience. Our model will give interesting recommendations quickly with minimal input. Once completed it will provide students with personalized recommendations based on their preferences. This will save time for both students and counselors by providing a head start on finding the proper classes for a student to take. Because of our large data set we do not have to worry about sparsity so by employing a hybrid ranking system based on multiple predictive and comparative metrics we will produce accurate and interesting results that can be tweaked.

Changes:

Most of the changes made have been highlighted with a **Changes:** or **Update:** at the beginning. Sections that do not have the bolded changes included have been completely overwritten. This includes: Challenges, Conclusion, Proposed Work, Introduction, and Metrics/evaluation.

Works Cited

BOSTRÖM, PETER, and MELKER FILIPSSON. "Comparison of User Based and Item Based Collaborative Filtering Recommendation Services." *EXAMENSARBETE INOM TEKNIK*, vol. 15, no. 2017, 2017, pp. 1 - 24. *Diva Portal*, https://www.diva-portal.org/smash/get/diva2:1111865/FULLTEXT01.pdf.

Chiang, Jeffery. "7 Types of Hybrid Recommendation System | by Jeffery chiang | Analytics Vidhya." *Medium*, 26 June 2021, https://medium.com/analytics-vidhya/7-types-of-hybrid-recommendation-system-3e4f78266ad8. Accessed 17 July 2023.

Georgiev, Kostadin, and Preslav Nakov. "A non-IID Framework for Collaborative Filtering with Restricted Boltzmann Machines." *PMLR*, vol. 28*3), no. 2013, 2013, pp. 1148-1156. *Proceedings of the 30th International Conference on Machine Learning*, http://proceedings.mlr.press/v28/georgiev13.html.

Khalifeh, Sewar, and Amjed Al-Mousa. "A Book Recommender System Using Collaborative Filtering Method." *ACM*, vol. 1, no. April 2021, 2021, pp. 131–135. *ACM Digital library*, https://dl-acm-org.colorado.idm.oclc.org/doi/10.1145/3460620.3460744.

Lin, Jinjiao, et al. "Intelligent Recommendation System for Course Selection in Smart Education." *Procedia Computer Science*, vol. 129, no. 2018, 2018, pp. 449-453. *Science Direct*, https://doi.org/10.1016/j.procs.2018.03.023.