Elective Recommendation System

Submitted by

Group: GA13

Aniket Bharti - University Roll No: 2100290140027

Sushmita Singh - University Roll No: 2100290140140

Session: 2022-23(4th Semester)

Under the supervision of

Mr. Prashant Agarwal

Associate Professor

KIET Group of Institutions, Delhi-NCR, Ghaziabad

DEPARTMENT OF COMPUTER APPLICATIONS

KIET GROUP OF INSTITUTIONS, DELHI-NCR, GHAZIABAD-201206

(April 2023)

ABSTRACT

This research paper presents a novel approach to developing a recommendation system for university elective courses that can provide valuable guidance to students on selecting courses based on their interests and academic backgrounds. The system uses a similarity-based approach that compares course templates to suggest suitable courses to students. This approach differs from traditional collaborative-based systems that rely on historical data and user feedback to generate recommendations. The proposed system's effectiveness was evaluated on a dataset of academic records of university students using popular algorithms such as Angular, Euclidean, Manhattan, and Hamming, and Dot, which were compared for their performance.

The results of the experiments showed that the proposed system using similarity-based algorithms outperformed collaborative-based approaches in terms of accuracy, achieving a rate of 86%. This demonstrates the potential of using similarity-based approaches to develop recommendation systems for university elective courses. The findings of this study can be valuable for universities seeking to enhance their elective course selection process and provide personalized guidance to students. Moreover, the system can benefit students who often struggle with selecting the most appropriate courses based on their academic records and interests. This research study could pave the way for future studies in exploring other similarity-based algorithms and evaluating their performance in the domain of university elective course recommendation systems.

TABLE OF CONTENTS

INTRODUCTION

Choosing elective subjects can be a daunting task for students in higher education. Some of the challenges that students face include limited capacity in some elective subjects, timetable clashes with compulsory subjects, and difficulty in finding the most suitable elective subject from the available ones. As a result, many students end up choosing elective subjects outside their majors because they perceive the subjects to be exciting. This often leads to students taking courses that do not align with their interests, goals, and aspirations.

To address these challenges, a recommendation system can be used to assist students in selecting the most suitable elective subjects. By analysing data on the subjects that students have completed, it is possible to categorize a student's interests and make recommendations based on those interests. The ability to predict student enrolment patterns for courses provides an opportunity to be effective in allocating resources and providing a high-quality learning experience. Additionally, predicting student grades in future courses before they take them is an essential tool that can be used to assist students in choosing elective subjects that align with their academic strengths and goals.

Recommendation systems must not only focus on broad outcomes such as courses but also on recommending learning resources and activities that will assist students in passing the recommended subjects. Such recommendations can take into consideration the student's needs, interests, preferences, and past activities. The purpose of recommendation systems is to recommend a product to a user that would possibly interest them based on the user profile. Therefore, recommendation systems in higher education should be designed to provide personalized guidance to students in selecting elective subjects that align with their interests, academic strengths, and goals. This research aims to explore the potential of using recommendation systems in higher education to enhance the elective course selection process and provide personalized guidance to students.

LITERATURE REVIEW

YouTube is a very large video platform that users use for an equally large variety of reasons. YouTube uses a recommendation system to help each user pick videos that suit their requirements. However, filtering these videos is not an easy task, and the recommendation system may make mistakes. This paper discusses what problems the recommendation system has and how dire these problems are through a survey in order to understand the advantages and disadvantages of the current recommendation system. Out of 59 participants, 46 of them find the recommendation system acceptable, good or excellent, but still point out several problems. Between the 24 creators participating in the survey, the recommendation system is still acceptable, but inclined much more towards being poor. We concluded that there is a lack of clarity on how the recommendation system works. This characteristic disliked mainly by creators because they are unable to direct their works to the prospective viewers.[1]

Matrix factorization (MF) is a prevailing technique in recommendation systems (RSs). Since MF needs to process a large amount of user data when generating recommendation results, privacy protection is increasingly being valued by users. Many existing privacy-preserving MF schemes only protect users' rating values, but ignore the privacy preservation of item sets rated by users. To make up for this shortcoming, a strategy based on piecewise mechanism (PM) is specially designed to simultaneously protect the privacy of rating values and item sets rated by users. To utilize data effectively, an improved MF based on PM (IMFPM) is proposed by dividing item profiles into global and personal information. Furthermore, in the IMFPM, random projection technology is used to reduce the influence of privacy noise on the estimation error. Theoretical analysis and experiment results show that the IMFPM not only provides strong differential privacy protection for rating values and item sets rated by users, but also has high prediction quality. Thus, the IMFPM is a good candidate scheme with privacy preservation for distributed recommendation systems.[2]

The recommendation system is fundamental technology of the internet industry intended to solve the information overload problem in the big data era. Top-k recommendation is an important task in this field. It generally functions through the comparison of positive pairs and negative pairs based on Bayesian personalized ranking (BPR) loss. We find that the contrastive loss (CL) function used in contrastive learning is well-suited for top-k recommendation.

However, there are two problems in the existing loss functions. First, all samples are treated the same, and hard samples are not considered. Second, all nonpositive samples are considered negative samples, which ignores the fact that they are unlabeled data containing items that users may like. Moreover, in our experiments, we find that when items are sorted by their similarities to the user, many negative items (or samples) appear before the positive items. We regard these negative items as hard samples and those at the top as potentially positive samples due to their high level of similarities with users. Therefore, we propose a ranking-based contrastive loss (RCL) function to exploit both hard samples and potentially positive samples. Experimental results demonstrate the effectiveness, broad applicability, and high training efficiency of the proposed RCL function.[3]

Universities have been focusing on increasing individualized training and providing appropriate education for students. The individual differences and learning needs of college students should be given enough attention. From the perspective of learning efficiency, we establish a clustering hierarchical progressive improvement model (CHPI), which is based on cluster analysis and context-dependent data envelopment analysis (DEA) methods. The CHPI clusters students' ontological features, employs the context-dependent DEA method to stratify students of different classes, and calculates measures, such as obstacles, to determine the reference path for individuals with inefficient learning processes. The learning strategies are determined according to the gap between the inefficient individual to be improved and the individuals on the reference path. By the study of college English courses as an example, it is found that the CHPI can accurately recommend targeted learning strategies to satisfy the individual needs of college students so that the learning of individuals with inefficient learning processes in a certain stage can be effectively improved. In addition, CHPI can provide specific, efficient suggestions to improve learning efficiency comparing to existing recommendation systems and has great potential in promoting the integration of education-related research and expert systems.[4]

Biodegradable starch films are promising as primary food packaging, and the k-Nearest Neighbour (KNN) algorithm enables selection and classification according to pre-established parameters. Here, the KNN algorithm and principal component analysis prove to be useful tools for sorting and selecting biodegradable starch packaging. Twelve biodegradable films were produced using starch from different botanical sources by the casting method. The KNN analysis evaluated data on thickness, water vapor permeability, tensile strength, elongation, water activity, transparency, and opacity, to obtain an information bank with 36 samples. Biodegradable films are visually homogeneous, transparent, without deformation, and easy to handle. The formulation (Cassava 5%) was classified as the best film, with WVP $1.21 \times 10-10$ (g. m−1. s−1. Pa−1), TS 2.34 (MPa), thickness 0.193 (mm), Aw 0.408, transparency 0.55 and opacity 0.63. The KNN algorithm and principal component analysis are advanced tools for classifying and selecting biodegradable starch films.[5]

Among the most popular collaborative filtering algorithms are methods based on the nearest neighbours (KNN). In their basic operation, KNN methods consider a fixed number of neighbours to make recommendations. However, it is not easy to choose an appropriate number of neighbours. Thus, it is generally fixed by calibration to avoid inappropriate values which would negatively affect the accuracy of the recommendations. In the literature, some authors have addressed the problem of dynamically finding an appropriate number of neighbours. But they use additional parameters which limit their proposals because these parameters also require calibration. In this paper, we propose a parameter-free KNN method for rating prediction. It is able to dynamically select an appropriate number of neighbours to use. The experiments that we did on four publicly available datasets demonstrate the efficiency of our proposal. It rivals those of the state of the art in their best configurations.[6] Business-to-business (B2B) social media efforts have largely focused on creating brand engagement through online content. We propose to analyse company social media texts (tweets) according to its two main dimensions, content and linguistic style, and to evaluate these in comparison to the overall content and style of the company's community of Twitter followers. We combine 15 million tweets originating from 254,884 followers of ten company profiles and link these to 10,589 B2B company tweets. Using advanced text analytics, we show that content similarity has positive effects on all engagement metrics, while linguistic style similarity mainly affects likes. Readability acts as a moderator for these effects. We also find a negative interaction effect between the similarity metrics, such that style similarity is most useful if content similarity is low. This research is the first to integrate content and linguistic style similarity and contributes to the brand engagement literature by providing practical message composition guidelines, informed by the social media community.[7]

Clarifying the mechanisms governing volumetric soil water content (VSWC) dynamics in soil profiles is essential, as it can help to elucidate soil water transport processes and improve the prediction accuracy of soil hydrological processes. Using Spearman's rank correlation and wavelet coherence analysis methods, similarity in soil profile VSWC dynamics and factors governing VSWC soil profile dynamics in upslopes and downslopes under three vegetation types (evergreen forest [EG], secondary deciduous forest mixed with shrubs [SDFS], and deforested pasture [DP]) at different time scales (hourly, daily, weekly, and monthly) and in different seasons were analyzed. The results revealed significant similarity in the VSWC of different soil depths ($P < 0.01$), with the similarity decreasing in accordance with the increment in soil depth. Greater VSWC similarity was found in EG than SDFS and DP sites and in upslope than downslope areas at both forest sites. The average significant coherence area (SCA) of VSWC similarity among surface and deep soil layers varied with the time scale, which was in the order of monthly (58.6%) > weekly (42.8%) > daily (21.8%). The effects of soil properties (e.g., texture, saturated hydraulic conductivity), rainfall, and potential evapotranspiration (ETp) on VSWC similarity were related to the time scale and season in which VSWC monitoring took place. Soil properties had apparent effects on VSWC similarity at longer time scales (i.e., monthly), with a high SCA. In contrast, the effects of rainfall and ETp on VSWC similarity were concentrated at weekly and daily scales, with a relatively low SCA. Rainfall and ETp dominated VSWC dynamics in the summer and fall, respectively. These results imply the use of measured VSWC at one soil depth to predict the VSWC at other soil depths was a reliable method. While the influence of time scale effects and seasonal variations on prediction accuracy of VSWC should be considered.[8]

PROJECT OBJECTIVE

1. **To develop an effective recommendation system for university students that suggests elective courses based on their academic performance in their past curriculum subjects:** This objective will focus on the development of a recommendation system that can accurately analyse a student's academic record and provide personalized recommendations for elective courses that match their academic strengths and interests.

2. **To evaluate and compare different recommendation algorithms in the context of elective course selection:** This objective will involve analysing the performance of different recommendation algorithms such as Angular, Euclidean, Manhattan, and Hamming distance, and comparing their effectiveness in generating accurate recommendations for elective courses.

3. **To provide an interactive user interface for students to input their academic performance data and receive personalized recommendations:** This objective will focus on the development of a user-friendly interface that will allow students to easily input their academic performance data and receive personalized recommendations for elective courses based on their unique profile. The interface should be intuitive and easy to navigate and should provide clear and concise recommendations for the student.

TECHNOLOGIES USED

The technologies used for developing this project are:

- **Python:** Python is a high-level, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation. Python is dynamically typed and garbage collected. It supports multiple programming paradigms, including structured (particularly procedural), object-oriented and functional programming.
- **NumPy:** NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.
- **Pandas:** pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license.
- **Annoy:** Annoy (Approximate Nearest Neighbours Oh Yeah) is a C++ library with Python bindings to search for points in space that are close to a given query point. It also creates large read-only file-based data structures that are mmapped into memory so that many processes may share the same data.

HARDWARE & SOFTWARE REQUIREMENTS

Hardware Requirements

S. N.	Description
1	PC with 10 GB or more Hard disk.
2	PC with 2 GB RAM.
3	PC with core i3 or above processor.

Software Requirements

S. N.	Description	Type
1	Operating System	Windows 10 or 11 or Ubuntu 18.04 or above
2	Language	Python 3
3	Front End	React 17
4	IDE	Google Colab, VS Code
5	Browser	Chrome, Firefox, Edge

MODULES IN PROJECT

- **User Interaction** is a critical component of the proposed recommendation system. A user interface will be developed to enable students to enter their scores of past subjects and receive recommendations based on their interests and academic strengths. The user interface will be designed to be user-friendly and intuitive, allowing students to input their information quickly and easily. Additionally, the user interface will provide feedback to users about their recommendations, giving them an opportunity to provide feedback and adjust their recommendations as needed.

- **Data Pre-processing** is an essential step in developing the recommendation system. It involves cleaning, transforming, and formatting the data in a way that the recommendation system can process it effectively. Data normalization and scaling are common data pre-processing techniques used to standardize the data and ensure that it is presented in the required format. Normalization ensures that data is consistent across different ranges, while scaling ensures that the data is transformed to a common scale for accurate analysis. These techniques will be used to ensure that the data is suitable for the recommendation system's analysis.

- **Recommendation Calculation** will be the backbone of the project, where all the marks of the student will be sent and embedding for that student will be calculated. The embeddings for the subjects will be calculated as well, and the similarity between the student's embedding and the embeddings for the subjects will be compared. This comparison will be used to generate personalized recommendations for each student based on their interests and academic strengths. The recommendation calculation process will be automated, and the system will be designed to provide recommendations quickly and efficiently. The accuracy of the recommendation system will be evaluated based on its ability to predict the students' interests and academic strengths and provide recommendations that align with their goals and aspirations.

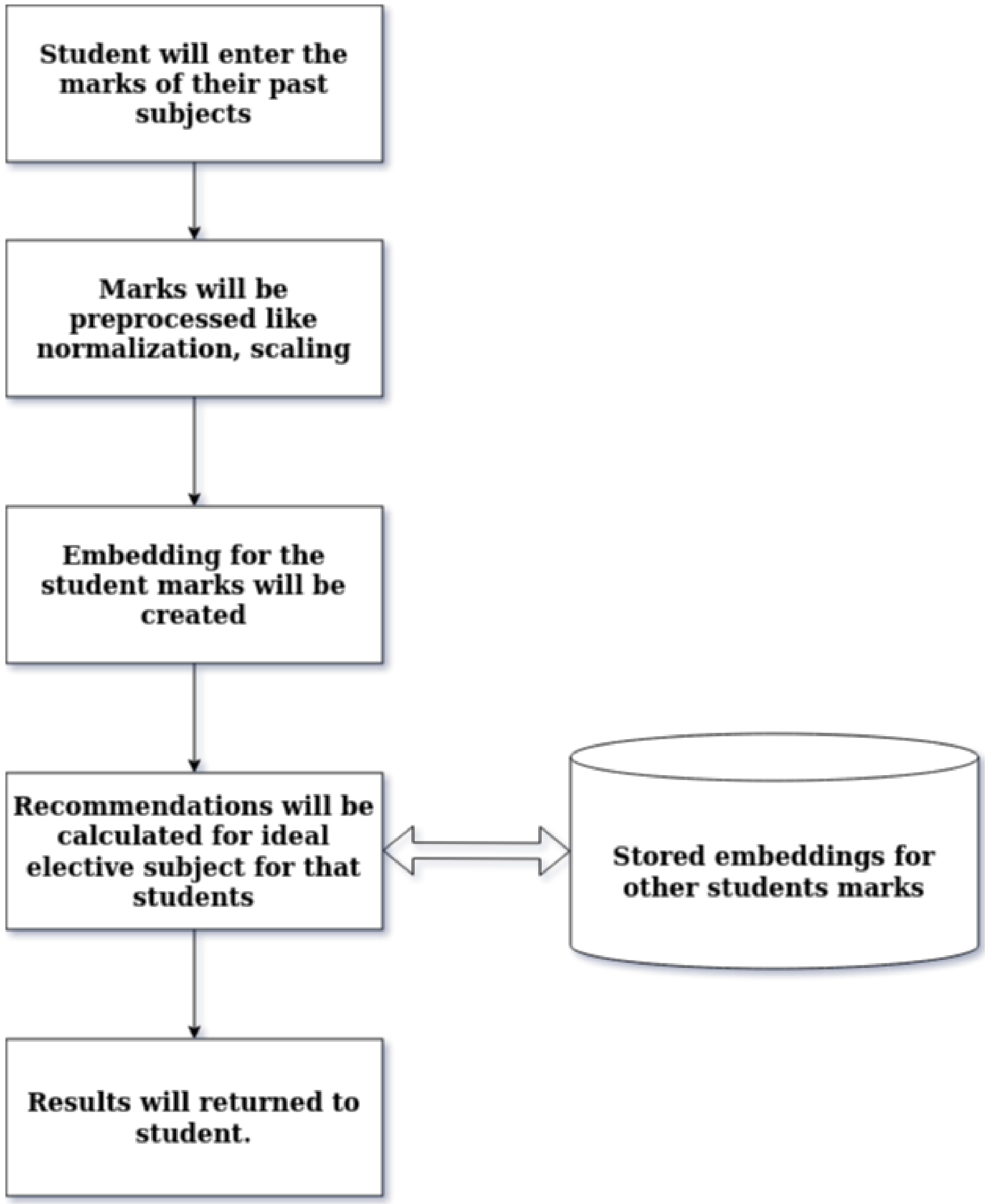
PROJECT OUTCOME

The primary objective of this project is to develop a recommendation system that can provide personalized elective subject recommendations to students based on their academic performance in their past curriculum subjects. This recommendation system will enable students to make informed decisions about which elective subjects to choose for the upcoming semester, based on their academic strengths and interests. This will ultimately result in higher academic achievement, improved student satisfaction, and more effective allocation of educational resources.

The research paper will document the entire process of developing the elective subject recommendation system in detail. It will include an explanation of each step involved in the development process, such as data pre-processing, recommendation calculation, and user interaction. Additionally, the paper will compare and analyse different algorithms used in the development of the recommendation system and the performance of the system on different datasets.

The research paper will provide insights into the effectiveness of the recommendation system in terms of accuracy, speed, and usability. The evaluation of the recommendation system will be based on the system's ability to generate personalized recommendations for each student, its ability to adapt to changing student needs, and its ability to provide feedback and adjust recommendations based on user feedback. This research paper will serve as a valuable resource for educators, administrators, and researchers interested in developing effective recommendation systems for higher education.

FLOW CHART


Student will enter the marks of their past subjects
Marks will be preprocessed like normalization, scaling
Embedding for the student marks will be created
Recommendations will be calculated for ideal elective subject for that students
Stored embeddings for other students marks
Results will returned to student.

GANT CHART

Task	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12
Requirement analysis and feasibility check												
Designing												
Coding												
Testing and deployment												

REFERENCES

1. Ardiansyah, F. M., Solim, H., & Gunawan, A. A. (2023). Level of user satisfaction with the current you tube recommendation system. *Procedia Computer Science, 216*, 442-452.

2. Wang, Y., Gao, M., Ran, X., Ma, J., & Zhang, L. Y. (2023). An improved matrix factorization with local differential privacy based on piecewise mechanism for recommendation systems. Expert Systems with Applications, 216, 119457.

3. Tang, H., Zhao, G., He, Y., Wu, Y., & Qian, X. (2023). Ranking-based contrastive loss for recommendation systems. Knowledge-Based Systems, 261, 110180.

4. Zhao, L. T., Wang, D. S., Liang, F. Y., & Chen, J. (2023). A recommendation system for effective learning strategies: An integrated approach using context-dependent DEA. Expert Systems with Applications, 211, 118535.

5. dos Santos Freitas, M. M., Barbosa, J. R., dos Santos Martins, E. M., da Silva Martins, L. H., de Souza Farias, F., Lourenço, L. D. F. H., & e Silva, N. D. S. (2022). KNN algorithm and multivariate analysis to select and classify starch films. Food Packaging and Shelf Life, 34, 100976.

6. Fopa, M., Gueye, M., Ndiaye, S., & Naacke, H. (2022). A parameter-free KNN for rating prediction. Data & Knowledge Engineering, 142, 102095.

7. Meire, M., Coussement, K., De Caigny, A., & Hoornaert, S. (2022). Does it pay off to communicate like your online community? Evaluating the effect of content and linguistic style similarity on B2B brand engagement. Industrial Marketing Management, 106, 292-307.

8. Liu, M., Wang, Q., Yi, J., Zhang, H., Liu, J., & Hu, W. (2023). Influence of vegetation type and topographic position on volumetric soil water content dynamics and similarity among surface and deep soil layers. International Soil and Water Conservation Research, 11(1), 183-196.