A K-Nearest Neighbour Algorithm-Based Recommender System for the Dynamic Selection of Elective Undergraduate Courses

Adewale Opeoluwa Ogunde*, Emmanuel Ajibade

Department of Computer Science, Redeemer’s University, Ede, Nigeria

Email address: ogunde@run.edu.ng (A. O. Ogunde)
*Corresponding author

To cite this article:

Received: October 15, 2019; Accepted: November 14, 2019; Published: November 21, 2019

Abstract: The task of selecting a few elective courses from a variety of available elective courses has been a difficult one for many students over the years. In many higher institutions, guidance and counsellors or level advisers are usually employed to assist the students in picking the right choice of courses. In reality, these counsellors and advisers are most times overloaded with too many students to attend to, and sometimes they do not have enough time for the students. Most times, the academic strength of the student based on past results are not considered in the new choice of electives. Recommender systems implement advanced data analysis techniques to help users find the items of their interest by producing a predicted likeliness score or a list of top recommended items for a given active user. Therefore, in this work, a collaborative filtering-based recommender system that will dynamically recommend elective courses to undergraduate students based on their past grades in related courses was developed. This approach employed the use of the k-nearest Neighbour algorithm to discover hidden relationships between the related courses passed by students in the past and the currently available elective courses. Real-life students’ results dataset was used to build and test the recommendation model. The new model was found to outperform existing results in the literature. The developed system will not only improve the academic performance of students; it will also help reduce the workload on the level advisers and school counsellors.

Keywords: Collaborative Filtering, Elective Undergraduate Courses, K-nearest Neighbour Algorithm, Recommender Systems

1. Introduction

In many educational institutions, undergraduates usually face a difficult time in choosing electives for many reasons. The number of available electives are sometimes many. There is no direct method of determining whether the student will pass the course. There is also no way the students can also assess their capacity to pass a specific number of extra elective courses. This paper, therefore, proposes a recommender system that would dynamically advise undergraduate students on which electives they are to pick for optimum academic performance. Recommender systems help with searching for suitable web resources, recommend the right solutions to improve students’ knowledge or analyse data obtained from quizzes and provide feedback to instructor to modify a quiz [1]. These systems can enhance the standard and process of teaching and learning. The collaborative filtering recommendation techniques have been proven to provide satisfying recommendations to users. They rely on the experiences of similar users, i.e., users who share the same preferences. Most of collaborative filtering-based algorithms use k-nearest-Neighbourhood algorithm to recommend items. The nearest-Neighbour users are those that exhibit the most substantial relevance to the target user. The k-Nearest Neighbours algorithm (or k-NN for short) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on
whether k-NN is used for classification or regression [2]. In this work, the k-Nearest Neighbours algorithm was deployed to find correlations and hidden knowledge from students’ previous grades. The discovered knowledge was used as a knowledge-base to recommend the elective courses to students. The proposed system will help students to dynamically choose the most appropriate electives they can do in a semester for better performance.

2. Recommender Systems

A recommender system is an information filtering system technique that attempts to recommend information items (books, movies, courses, etc.) that are likely to fit a user’s taste. They try to predict the rating or preference that a user would give to an item. A recommender system can be used to filter any sort of information, and with any kind of preference. The main methods recommender system utilizes to propose an item to a user include analyzing user data, extracting useful information, and finally predicting items to users. The availability of practically limitless online information makes recommender system absolutely necessary [3]. A lot of research work has been carried out yet the challenges in the field of recommender models persist, and there is plenty of scope for improvement. Aside this, recommender systems have been successful in many fields including the educational environment. Recommender systems have helped with searching for suitable web resources, recommend the right solutions to improve students’ knowledge and in general, enhance the teaching and learning processes. The use of efficient and accurate recommendation techniques is essential for a system that will provide good and useful recommendation to its users. This explains the importance of understanding the features and potentials of different recommendation techniques.

Content-Based filtering recommendation system, which recommends items that are similar in content to items the user has liked in the past or matched to attributes of the user, was applied to solve the problem presented in this paper. It usually bases its predictions around the information provided by the user and ignore other contributions from other users. The content-based filtering system can be defined as methods that provide recommendations by comparing representations of content describing an item to representations of content that interests the user [4]. Systems implementing a content-based recommendation approach analyse a set of documents and/or descriptions of items previously rated by a user, and build a model or profile of user interests based on the features of the objects rated by that user.

2.1. K-nearest Neighbour

K-nearest Neighbour is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition already in the beginning of the 1970’s as a non-parametric technique. K-NN can be used for both classification and regression predictive problems. It is a non-parametric method. In k-NN classification, the output is a class membership. An object is classified by a majority vote of its Neighbours, with the object being assigned to the class most common among its k nearest Neighbours (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest Neighbour. Whenever we have a new point to classify, we find its K nearest Neighbours from the training data. The similarity metrics do not consider the relation of attributes which result in inaccurate distance and then impact on classification precision. The wrong classification due to presence of many irrelevant attributes is often termed as the curse of dimensionality [5].

In k-NN regression, the output is the property value for the object. This value is the average of the values of its k nearest Neighbours. Both for classification and regression, it can be useful to assign weight to the contributions of the Neighbours, so that the closer Neighbours contribute more to the average than the more distant ones. For example, a common weighting scheme consists in giving each Neighbour a weight of 1/d, where d is the distance to the Neighbour.

2.2. Other Related Works

Four algorithms suitable for the course recommendation were proposed by Hana [1]. The first algorithm searches for the most frequently enrolled courses. The second algorithm utilizes similarities of students based on courses of their interests. The third algorithm recommends courses of students’ favorite teachers. The last algorithm calculates the social ties among students and selected courses which were attractive to students’ friends. A new collaborative recommendation system that employed association rules algorithm to recommend university elective courses to a target student based on what other similar students have taken was proposed by Al-Badarenah, and Alsakran [2]. The experiments showed that association rule is a popular tool for making recommendation to a target student. Through their experiments, it was noticed that the patterns of influence of different parameters on the performance of the system. The confidence and match of a rule have a significant impact on the performance, but the highest confidence or match may not be the best choice. By choosing a relatively high confidence or match, better performance can be achieved. Recommender systems open new opportunities for retrieving personalized information on the Internet [6]. It also helps to alleviate the problem of information overload which is a widespread phenomenon with information retrieval systems and enables users to have access to products and services which are not readily available to users on the system. Their work discussed the two traditional recommendation techniques and highlighted their strengths and challenges with diverse kind of hybridisation strategies used to improve their performances. Various learning algorithms used in generating recommendation models and evaluation metrics used in measuring the quality and
performance of recommendation algorithms were discussed. Systems that made only top-n recommendations have also been proposed in the literature [7, 8]. Two collaborative filtering approaches for predicting the grade a student will obtain in different courses based on their performance in earlier courses was compared by Sanjog and Sharma [9]. They used the collaborative filtering approach to recommend grades students will get in a course if they select some elective courses out of the range of available elective courses [9]. Experimental evaluation using the collaborative filtering approach on a real-life data set showed good results.

Two essential works deployed neural networks to build relationships between users and products, considering matrix factorisation for determining user preferences [10, 11]. An auto-encoder based collaborative filtering technique which employed original and partially perceived vectors was proposed by Ouyang et al. [12]. Another research focused on YouTube recommendations using a Multilayered Perceptron technique, which produced good results [13]. A related work that recommended elective subjects based on the neural network and association rules using the Rapid Miner Tool was presented by Samrit and Thomas [14]. Their system worked on real database obtained from the university. It analysed the past behaviour of students concerning their elective subject choices. More explicitly, it formalises association rules that were implicit before. The accuracy of the system and other performance metrics for the system were not presented in the study.

Collaborative metric learning, which maximised the distance between users and their disliked items and minimised the gap between the users and their liked items was proposed by Hsieh et al. [15]. A survey of the literature of course recommender systems for course selection was carried out by Unelsrod [16]. One of the goals was to develop and test a recommender system that uses a combination of collaborative filtering and content-based filtering to give recommendations. It was achieved by using memory-based collaborative filtering with k-nearest Neighbour search to find closely correlated users, and content filtering based on the most popular courses with older students and which teachers the user has shown a preference for.

Other recommender systems have been built to solve problems in several domains such as online libraries, e-commerce sites, LinkedIn, Google, Facebook, Twitter and other social media [17-19]. A very recent research also developed a system that recommended potential industrial training organizations to students using the J48 algorithm and a decision tree-based model [20]. The Placement predictor and course recommender system tool was developed to improve the quality of education and enhance school resource management using C4.5 algorithm, which gave a much better prediction than any other classification algorithms used in the work [21].

In a very recent work [22], a course recommender system that aims to improve students’ career readiness by suggesting relevant skills and courses based on their unique career interests was presented. The course recommender system used content-based filtering and an ensemble learning algorithm of k-means clustering and TF-IDF to suggest relevant skills and courses based on students’ career interests. More information, analysis and comparison of several methods used for building course recommendation systems from educational data set was presented by Thanh-Nhan et al. [23].

3. Methodology
3.1. Data Collection and Description

The data used in this work was collected from the Department of Computer Science, Redeemer’s University. The data consist of student grades in both compulsory and elective courses for five academic sessions, that is, 2010/2011 to 2015/2016 session. The data was made up of the students’ grades in which A was the maximum grade, and F was the minimum grade. The status of course can either be compulsory (C) or elective (E). The list of both elective and compulsory courses was also obtained. The attributes described here are shown in Table 1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Possible Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Compulsory Course</td>
<td>A, B, C, D, E and F</td>
</tr>
<tr>
<td>E</td>
<td>Elective Course</td>
<td>A, B, C, D, E and F</td>
</tr>
<tr>
<td>GR</td>
<td>Grade</td>
<td>A, B, C, D, E and F</td>
</tr>
</tbody>
</table>

To simplify the work for analysis, domain values were associated with each of the attributes used. The domain values for the variables defined for the present investigation are as follows:

1. C - Compulsory Course
2. E - Elective Course
3. CMP - Computer
4. GST - General Study
5. CHE - Chemistry
6. STA - Statistics
7. MAT - Mathematics
8. PHY - Physics
9. S - Score
10. GR – Grade

The final grade is represented by the values A, B, C, D, E and F. These values are expanded below.

A. Values from 70 and above.
B. Values ranging from 60 to 69.
C. Values ranging from 50 to 59.
D. Values ranging from 45 to 49.
E. Values ranging from 40 to 44.
F. Values below 40.

The data to be trained is a collection of students results for five academic sessions in various courses (2011/2012, 2012/2013, 2013/2014, 2014/2015 and 2015/2016 sessions). All Sessions consist of students’ results in both elective and compulsory courses. The detailed characteristics of the students’ results obtained are shown in Table 2. Altogether, a total of ten thousand six hundred and one data instances were
obtained.

Table 2. Data Characteristics.

<table>
<thead>
<tr>
<th>Session</th>
<th>Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011/2012</td>
<td>1506</td>
</tr>
<tr>
<td>2012/2013</td>
<td>1785</td>
</tr>
<tr>
<td>2013/2014</td>
<td>2400</td>
</tr>
<tr>
<td>2014/2015</td>
<td>2340</td>
</tr>
<tr>
<td>2015/2016</td>
<td>2570</td>
</tr>
<tr>
<td>TOTAL</td>
<td>10,601</td>
</tr>
</tbody>
</table>

3.2. Design of the Elective Course Recommender System

The recommender system, in this work, was designed to contain a knowledge-base that has accumulated experience and a set of rules for applying the knowledge-base to each particular situation that was described. Collaborative filtering recommendation approach was applied to analyse the data. This filtered information by using the recommendations of other students that passed the same elective courses to predict future cases. This was premised on the idea that people who passed certain elective courses in the past are likely to pass similar elective courses in the future. The collaborative filtering system used in this work applied the Neighbourhood-based technique. In the Neighbourhood-based approach, a number of elective courses were selected based on their similarity to the active student. A prediction for the active course was made by calculating a weighted average of the ratings of the selected courses. Using $k$-Nearest Neighbour classification, the training dataset was used to classify each member of the "target" dataset. The structure of the data is that there is a classification (categorical) variable of interest and a number of additional predictor variables.

WEKA, a popular data mining tool, was used to generate the recommendation model and the $k$-nearest Neighbour decision tree. The primary data used for the analysis was imported from a database in CSV format and converted to ARFF format, which is suitable for analysis by WEKA. Data preprocessing was done using the filtering algorithm. The selected data in CSV format was pre-processed by the WEKA ARFF viewer. The converted data was stored as "reco.arff". A sample of the converted data is shown in figure 1. The field attributes of figure 1 are serial number, semester, session, coursecode, course status, score and grade.

The $k$-Nearest Decision tree was generated from the stored. arff data format in WEKA explorer. The attributes in the data include semester, session, coursecode, status, score and grades. All these were used to generate rules for elective course recommendation. The $K$-Nearest rules generated using WEKA, and the results classifier are shown in figures 2 and 3 respectively.

A classifier with cross-validation option (10 folds) was used to classify the data and result is shown in figure 2. Figure 3 shows the results of the classifier. Correctly classified data instances were 95.6469% while incorrectly classified data instances were 4.3531%. This proved a high level of accuracy for the model generated. Kappa statistic measures the agreement of prediction with the actual class. In figure 2, the kappa statistic was shown to be 0.9467, which means a higher agreement between the predicted values and the actual class values. Mean Absolute Error (MAE), which is a measure of the average magnitude of the errors in a dataset without considering their direction, was 0.0146. The small value of the MAE indicated that the error associated with this classification was very minimal. Root Mean Squared Error (RMSE) was 0.1178, which is a measure of the average magnitude of the error. The RMSE is expected to be greater or equal to the MAE, which is the case in this work. This implied a
variance in the individual errors of the dataset. The True Positive (TP) value signifies the correct predictions. TP had an average value of 0.956, which implied correct prediction for majority of the class labels. The False Positive (FP) value is the number of instances predicted positive but was actually negative. In this work, FP had an average value of 0.010, which implied very few cases of incorrect predictions. The Recall was an average value of 0.956, which was an indication that majority of the class instances predicted to be positive was actually positive. The average value of Precision, in this work, was 0.956, which corroborated the fact that many cases that were predicted positive were positive. In this work, the average value of the F-measure was 0.956, which implied that the recall and the precision were evenly weighted. The average value of the MCC was 0.947. This indicated that the model was predicting with a very high level of accuracy.

4. Implementation and Results

The rules generated by the classifier were used as a knowledge-base for building a front-end web-based recommendation system, which was thoroughly tested to ensure that it complies with the rules. Some of the development tools include Adobe Dreamweaver and Sublime Text. The Admin is the user that performs administrative activities on the system such as editing, adding and deleting of courses from the system. The home page of the elective course recommendation system is shown in figure 4.

![Figure 4. Home Page.](image)

The login page for the recommender system is shown in figure 5. This page allows a student to either present his/her login credentials consisting of the student’s email address and the student’s password or head to the sign-up page to create an account.

![Figure 5. Login Page.](image)

The flowchart in figure 6 describes the steps a student must follow to view recommendations concerning the choice of elective courses.

The student needs to register or log in if already registered. If the login is valid, the student uploads his/her past results into the recommender system after which the student selects his/her department, level and the current semester in which the recommendation is needed. After that, the students can now view the recommendations of elective courses he/she can register in the current semester.

The developed recommendation system contains several useful pages to students. Some of them are: Update Past Results Page, View Recommendation Page, Recommendation Page etc. The Update Past Result page (figure 7) enables the user to update his/her past results and when they are done the system stores the information for future use. This increases the knowledge in the knowledge-base from time to time and improves the quality of recommendations with time.

![Figure 3. Result of the Classifier.](image)
The Prepare Recommendation (figure 8) helps the student to select his/her details and the current semester in which the recommendation is needed and after that ‘View Recommendation’ button is clicked.

When the ‘view recommendation’ button is clicked, the application triggers the recommendation engine based on the value inputted, and results of the recommendation are presented to the user (figure 9).

It was observed that grade F had the highest correct precision (89.7%), while grade B had the least correct precision (70.4%). This implies that the recommendation system is working at a very high level of accuracy in predicting all the grades.
4.2. Overall Accuracy Percentage of the Model

Table 4 shows the accuracy percentage for the k-nearest Neighbour classification of all the data instances. The overall accuracy percentage of the model is 95.6469%. This outperformed all the related works described in section two and further confirmed that the model for the developed recommender system is very efficient and predicting at a very high level of accuracy.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Correctly Classified Instances</th>
<th>Incorrectly Classified Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-Nearest Neighbour</td>
<td>95.6469%</td>
<td>4.3531%</td>
</tr>
</tbody>
</table>

5. Conclusions and Future Works

A study was conducted on the dynamics of recommending elective courses to undergraduate students based on the past performances of other students in the course and the results of the student under study using a machine learning technique. From the class-wise accuracy, it was clear that the true positive rates for the grades A, B, C, D, E and F were 79.4%, 70.4%, 82.7%, 71.0%, 82.7% and 89.7% respectively. A decision tree-based model from which prediction rules were extracted and implemented in a front-end web application. It could be concluded that this model with similar functionalities can aid schools in effectively monitoring students’ progress. The university can advise students on steps to take to ensure they come out successful in their examinations. The effectiveness and efficiency of this machine learning application have made it possible to solve real-world problems in the form of decision making in higher institutions. It is highly recommended that institutions should start adopting similar systems in their educational network to vastly improve educational experiences and render timely and effective decision-making solutions to students. Future works will include experiments using other types of decision tree algorithms. A comparative analysis of the performances of the algorithms would be carried out. Finally, data from a single discipline was used for the analysis in this work; future works will consider the development of a generic system with data from many disciplines.

Acknowledgements

The authors would like to appreciate the department of computer science, for providing the data used in executing this research.

References


