

ЖАРАТЫЛЫСТАНУ ЖӘНЕ ТЕХНИКАЛЫҚ ҒЫЛЫМДАР

NATURAL AND TECHNICAL SCIENCES

IRSTI 28.23.02

N. Anefiyayev¹, D. Anefiyayeva², A. Talasbek³
^{1,2,3}Suleyman Demirel University, Kaskelen, Kazakhstan

DESIGNING A RECOMMENDATION SYSTEM FOR SPECIALIZED COURSES FOR THE UNIVERSITY

Abstract. Each university has compulsory subjects and there are subjects that the student must choose for themselves. This choice affects the further path of the student because he chooses not only the subject that he will study but also what he will work within the near future. I know from myself how difficult it is to choose a subject yourself. The recommendation engine is one of the most popular artificial intelligence applications, attracting many researchers from all over the world. From the moment we switched to the Internet, the recommendation system has become widely used in our daily life, even when we do not notice it. Many machine learning techniques can be used to implement a recommendation system, but in this work, we consider the KNN method for classification.

Keywords: Recommender system, machine learning, subject selection, university, KNN.

Аңдатпа. Әрбір университетте міндетті пәндер және студент өзі таңдауы керек пәндер бар. Бұл таңдау студенттің әрі қарайғы жолына әсер етеді, өйткені ол оқитын пәнді ғана емес, жақын болашақта не істейтінін де таңдайды. Пән таңдаудың қаншалықты қиын екенін өз басымнан білемін. Ұсыныс қосымшасы бүкіл әлемнен көптеген зерттеушілерді қызықтыратын ең танымал жасанды интеллект қосымшаларының бірі болып табылады. Ғаламторды іске қосқан сәттен бастап ұсыныс беру жүйесі күнделікті өмірде біз байқамай қалсақ та кеңінен қолданыла бастайды. Ұсыныс жүйесін енгізу үшін қолдануға болатын көптеген машиналық оқыту әдістері бар, алайда бұл жұмыста классификация үшін KNN әдісін қарастырамыз.

Түйін сөздер: Ұсыныс жүйесі, машиналық оқыту, пән таңдауы, университет, KNN.

Аннотация. В каждом университете есть обязательные предметы и есть предметы, которые студент должен выбрать сам. Этот выбор влияет на дальнейший путь студента, потому что он выбирает не только предмет, который будет изучать, но и то, чем он будет заниматься в ближайшее время. Я знаю по себе, как сложно выбрать предмет самостоятельно. Рекомендационная система — одно из самых популярных приложений искусственного интеллекта, привлекающее множество исследователей со всего мира. С того момента, как мы перешли в Интернет, система рекомендаций стала широко использоваться в нашей повседневной жизни, даже когда мы этого не замечаем. Существует множество методов машинного обучения, которые можно использовать для реализации рекомендательной системы предмета по выбору, но в этой работе мы рассматриваем метод KNN.

Ключевые слова: рекомендательная система, машинное обучение, выбор предмета, вуз, KNN.

I. Introduction

Recommended systems (RS) are widely used in many areas of modern man. They can also be used in the educational field. In every university where there is a credit system, there are compulsory subjects and electives. The list of courses that a student must choose is made up of a number of optional subjects, while the required course must be taken by students. Compulsory subjects are mainly aimed at closing the degree; on the other hand, elective subjects are more focused on a narrower specialization. This choice affects the whole life because the course passed will determine what the future profession will be. It can be concluded that this decision is one of the most important. Based on this, the creation of an automatic system, of course, recommendations within the walls of the university are relevant. Analyze and highlight the main aspects that affect the successful delivery of the subject and recommend based on them, the student when choosing. The main goal of my research is to create a recommendation system for students, which will help in choosing subjects based on the results obtained when passing a psychological test. This recommendation system will be created using the latest machine learning algorithms technology. The remainder of the paper is laid out as follows. Section II discusses related course and recommender system efforts that employ diverse techniques. The suggested recommender system for optional course suggestions is described in depth in Section III. The experimental results and conclusion are described in Section IV.

II. Literature review

In this paper, methods were shown that included recurrent neural networks (RNNs), which were used to predict actions while retaining past observed behavior. For a recommendation system, this means an overly narrow set of sentences devoid of intuitive foresight, which puts the user in a filter bubble.

This article has eliminated this problem. To predict course outcomes that are given for a student to choose from that will suit their interests. By building a set of models based on the course catalog (BOW) and another built on the enrollment histories (course2vec). Compared the performance of the two models using validation datasets. Comparing the two models, the results using user research show a dramatic lack of interest in RNN guidelines. But machine learning of the course2vec model showed the best results in tasks when validated offline. [1].

This article presents several methods that were used for course recommendation systems. With the help of this system, students can find out their approximate results in advance, which will further help them in choosing a subject. There is also a table and a comparative analysis of the effectiveness of the educational data of each of the methods. Then the best method was chosen. And then the final version Of the recommendation system has been drawn up, which can already be applied in practice. [2].

Course recommendation systems are created to help students choose courses from a wide variety of subjects. However, the student's choice may depend not only on his interest; teachers, peers, and other things have an impact on them. Interpersonal relationships, the assessment text, and the existing rating matrix in the "user-element" format constitute a multimodal data structure from several sources; therefore, to develop recommendations based on these disparate qualities requires method for systematically combining data. Therefore, this article proposed a hybrid recommendation model that combines structured network functions with the use of graph neural networks and interactive student actions with factorization of tensors. First, utilizing student ratings, a structured graphical learning evaluation network is developed to characterize students, courses, comment text, grades, and interpersonal relationships. By studying the pupils' relationship structure, the neural network with random walks is then utilized to build a vectorized representation of them. Finally, Bayesian Probability Tensor Factorization is used to investigate and forecast student scores in topics they have not attended, recognizing these personalized traits as the third dimension of the rating tensor. With smaller prediction error and improved recommendation accuracy, the suggested technique beats other current matrix factorization models and neural networks (xSVD ++, RTTF, DSE). [3].

Pakistani universities offer credit programs of all kinds to applicants. Pakistan Virtual University - Developed with the latest information technology. This university offers a choice of a large number of professions and specialties. Each program has subjects that the student must complete. The chosen course that suits the competencies and interests greatly affects the final grade (CGPA) of the student. In this work, a system for recommending subjects in the Virtual University was created. The system has been tested on presently 470 courses are accessible, including simulated data from 2,600 people. As a result, it was proved that the grades are influenced by: the average score of the student in the

courses already studied and the average score in the similar courses of the student. The implemented system and its accuracy were evaluated using the mean absolute error for 100 observations. The MAE was in the acceptable range. [4].

This document describes a hybrid RS that uses Content-Driven Filtering (CBF) and Collaborative Filtering (CF) to propose the best relevant courses to students based on several factors linked to both student and course information. The genetic algorithm (GA) was created to find the best RS configuration, which takes into account the most significant criteria and the remaining factors. The pilot project employed real data from the University of Cordoba's computer science program, comprising data collected from students over three academic years and based on 2,500 records from 95 individuals and 63 courses. The examination of the most acceptable criteria for course suggestion is demonstrated by experimental findings; the experimental findings show the relevance of employing a hybrid model that includes both student and course information to increase the dependability of suggestions, as well as better performance compared to earlier models.. [5].

Consider all the methods that were used in the works that are described above, as well as to conduct a comparative analysis and show the results of the work. The method is proposed consists of several stages: (1) after reading all the resources, identify what were the goals and what subjects were considered in the work (2) what was the strategy and what technologies were chosen to achieve the goal (3) what saw the obstacles and what they plan to do in the future. To identify methods that will help me in further work, I brought out all the algorithms used, as well as the results of the work in Table 1.

Table 1. Related works

No	Research	Goals and objectives	Strategy /Approach	Performance	Limitation and Future Work
1	Zachary A. Pardos and Weijie Jiang 2020[1]	Design serendipity for university recommendation system	Course2vec, multifactor course2vec and bag-of-words model	Unexpectedness, novelty, variety, and the major measure of serendipity were all greater in the diversity-based algorithms than in the	L: Students have to rely on the semantics of the course description. FW: Additional information, such as latent semantics, should be added to

				non-diversity algorithms.	course descriptions.
2	H. Thanh-Nhan, N. Thai-Nghe, H. Nguyen 2016[2]	Building the course recommendation systems.	KNN each for student and course; MF and BMF with RMSE measure	Because the BMF method has the least root mean square error, it is employed in course recommendation systems.	FW: To make better use of relational data, take a multi-relational strategy.
3	Yifan Zhu, Hao Lu, Ping Qiu, Kaize Shi, James Chambua, Zhendong Ni 2020[3]	Network-based offline course recommendation with graph learning and tensor factorization for heterogeneous teaching assessment	SCD, DSE, RTTF, TENTF	When compared to PMF techniques (RTTF, DSE) and SVD-based methods, the proposed model obtains the lowest MAE and RMSE.	Simplified computation method and lower the amount of time spent training and forecasting
4	Aleem Akhtar 2020[4]	Recommender system for Virtual University in course selection process	Collaborative Filtering, KNN, VU-CRS algorithm	The difference between the expected and real values stayed within 10%. For 100 observations, the MAE came out to be 5.12, which is a reasonable mistake.	More tailored findings can be obtained using a psychological test-based method.
5	A. Esteban, A. Zafra, C. Romero 2020[5]	A system with genetic optimization was developed	Proposed hybrid RS CBF with clustering user-based &	Using a hybrid approach with multiple criteria are	Add limitations to course recommendations to filter

		using a hybrid multi-criteria recommendation system.	item-based CF MCSeCF	significantly better than rest models (RMSE values)	them; expand with more degrees
--	--	--	----------------------	--	--------------------------------

Based on all of the above, say with confidence that the creation of a recommendation system for the university is very relevant at the moment. The articles that were presented to us will be very helpful in our work since they showed what goals to pursue in our work, what data to collect, what algorithms to use in our work. Analysis of the Table 1 shows that all work is aimed at creating a system that can prompt the student as accurately as possible which subject will be of interest to him based on his choice of previous subjects, the average grade for them, as well as the teacher who teaches the subject.

III. Method and results

Data structure

As a basis, we took data from an open source[7]. They consist of 8016 rows and 7 columns: course id, the course title, duration in seconds, course release date, course description, assessment status, and whether the course is retired. In Table 3, you can see the first 5 entries sorted by the date the course was released.

Table 3. Courses data first 5 rows

	CourseId	Course Title	Duration InSeconds	Release Date	Description	AssessmentStatus	IsRetired
0	identity-access-management-AWS-roles-groups	Identity and Access Management on AWS: Roles a...	9015	2019-02-25	Using individual users for all access control ...	None	no
1	AWS-s3-implementing	Implementing Amazon S3 Storage on AWS	12329	2019-02-23	AWS S3 is one of the most fundamental services...	None	no

2	nunit-moq-mocking	Mocking with Moq and NUnit	6780	2019-02-22	Writing effective unit tests can be tricky whe...	None	no
3	secure-coding-using-components-known-vulnerabilities..	Secure Coding: Using Components with Known Vul...	4414	2019-02-22	Do you know if old components you are using are...	None	no
4	identity-access-management-aws-users	Identity and Access Management on AWS: Users	6179	2019-02-22	At the core of Identity and Access Management ...	None	no

Data Pre-processing

After receiving the data, one of the key steps is data pre-processing. Data pre-processing increases the quality of data and makes it easier to get useful conclusions from it. Cleaning and arranging raw data to make it acceptable for creating and training machine learning models is referred to as data preprocessing in machine learning. My data went through the following steps:

- Drop Nan Values from all columns
- Replace 'll and dash (-) with empty space (" ")
- Remove all characters except alphanumeric
- Remove stopwords
- Replace words with vectors using TfidfVectorizer

The KNN (k-means) method is a basic supervised machine learning technique that may be used to handle classification issues. The KNN method merely saves the information during the training phase, then when new data is received, it is classified into a cluster that is very comparable to the new data. We divided all university subjects into certain clusters based on the description of the subject. Items in a given cluster are contiguous. Next, the student describes

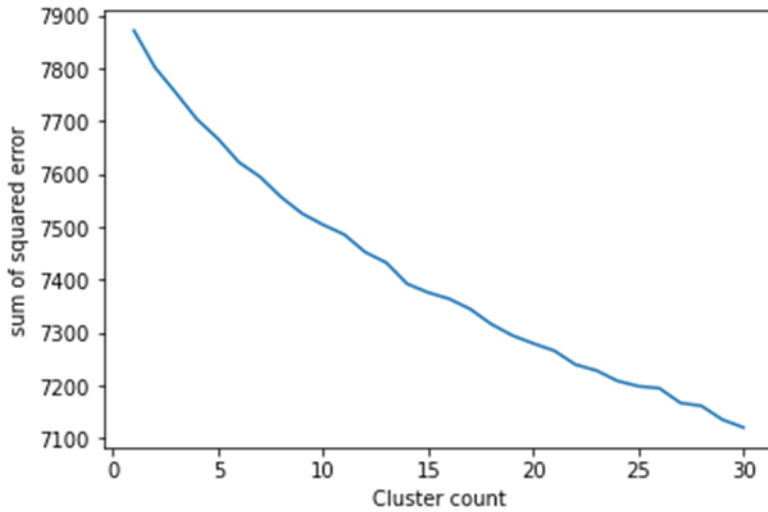
his predisposition, for example, he likes to draw and invent characters. Here our recommender system will offer him items from the first cluster, such as 3D modeling, unity, or photoshop. In the table (Table 4) below, I compare the input parameters of the previous researcher and my own and compare our results. After the model was fitted with vector data, it created 30 different clusters, and Table 5 provides the first four clusters with the top seven terms. Pic. 1 illustrates how the sum of squared errors is decreasing slightly while cluster count is increasing.

Table 4. Compare results

Name parament	Previous result	My results
Cluster count	8	30
Maximum iteration	100	500
Different centroid initializations	8	15
Sum of squared errors	7563.58	7120.59
Model accuracy score	0.725	0.836

Table 5. First 4 cluster top 7 terms

Cluster0	Cluster1	Cluster2	Cluster3
photoshop	wireless	windows	application
illustrator	network	directory	web
design	troubleshoot	server	game
3d	security	active	play
unity	cisco	2012	management
effects	ccna	administration	javascript
create	information	sql	build



Picture 1. Improving sum-of-squared-errors by the cluster count

IV. Conclusion and limitation

Based on all of the above, say with confidence that the creation of a recommendation system for the university is very relevant at the moment. This paper will help students find a course according to their ability simply by writing their precautions. Since our method determines a more detailed cluster and issues the corresponding rates. We also improved the previous method and got a more detailed clustering model. This model works well when it is necessary to predict the category of courses that have a fairly complete description of the subject, as well as a description of the student's key abilities since the connectivity and the sum of the square errors depend on this.

References

- 1 Z. A. Pardos, W. Jiang, "Designing for Serendipity in a University Course Recommendation System", In Proceedings of the 10th International Conference on Learning Analytics and Knowledge (LAK '20), pp 350-359, 2020
- 2] H. Thanh-Nhan, N. Thai-Nghe, H. Nguyen, "Methods for building course recommendation systems", In Eighth International Conference on Knowledge and Systems Engineering (KSE), 2016
- 3 Y. Zhu, H.Lu, P.Qiu, K.Shi, J.Chambua, Z.Ni, "Heterogeneous teaching evaluation network based offline course recommendation with graph learning and tensor factorization", In Neurocomputing 415, pp 84–95, 2020
- 4 A.Akhtar, "Implementation of Course Recommender System for Virtual University of Pakistan", In Computers in Human Behavior, 2018

- 5 A. Esteban, A. Zafra, C. Romero, “Helping university students to choose elective courses by using a hybrid multi-criteria recommendation system with genetic optimization”, In Knowledge-Based Systems, 2020
- 6 C. De Medio, C. Limongelli , F.Sciarrone & M.temperini , “MoodleREC: A recommendation system for creating courses using the moodle elearning Platform”, In Computers in Human Behavior, 2019
- 7 Pluralsight [Online] <http://api.pluralsight.com/api-v0.9/courses> [Last update available 07.03.2022]