ISSN 2063-5346

A HYBRID PERSONALIZED MOOC COURSE RECOMMENDATION MODEL

Sujatha.U[1] Gunasundari.R[2]

Article History: Received: 01.02.2023 Revised: 07.03.2023 Accepted: 10.04.2023

Abstract

With the rapid growth in online learning platforms, learners find it challenging to identify courses that meet their learning preferences. A personalized recommendation system is needed to identify and suggest courses according to their learning needs. The proposed recommendation model consists of, a chatbot for capturing users' interest, and web crawlers to extract the course information from Massive Open Online Course (MOOC) sites and recommend courses that match the learner's requirements. Beautiful Soup is used for web scraping for creating the custom MOOC dataset. Pre-processing of the dataset is carried out with the Natural Language Tool Kit (NLTK). Term Frequency–Inverse Document Frequency(TF–IDF), a word embedding model is used for every pre-processed data to extract the features and information. The extracted features act as the input of the Long Short-Term Memory Network (LSTM) to categorise the user's preference message and provide random response from the list of trained responses. Cosine similarity, content- based filtering responsible for selecting information based on learner keywords and filtering of elements, with the Neo4j graph-based model used for recommending courses.

Keywords— *Personalised Recommendation, Chatbot, Web scraping, Cosine similarity, Neo4j***.**

Research Scholar Department of Computer Science Karpagam Academy of Higher Education, Coimbatore,India sujathau@gmail.com

[2]*Professor and Head, Department of Computer Applications Karpagam Academy of Higher Education*, Coimbatore,India gunasoundar04@gmail.com

DOI:10.31838/ecb/2023.12.s1-B.382

I. INTRODUCTION

Chatbots are digital assistants that interact with users via natural languages to provide assistance in scheduling appointments, booking tickets, and all daily activities. In the field of education[1], Chatbots function as service assistants to help in improving the student learning process. The main challenge for learners is choosing courses of their preference from many MOOC providers. To overcome this issue, a recommendation model is developed with suggestions received from users through the chatbot. A chatbot can assist in choosing courses from online learning platforms by exploring available options from MOOC sites[2]. Various types of chatbots are available depending on how they are used.

Cosine similarity is content-based filtering used for the recommendation model with Natural Language Processing (NLP)[3]. The recommender system interaction is built through a conversation mechanism to collect the learner's course preferences.
A conversational chatbot for a recommendation of courses isused to identify the preferences of user and recommend t through dialogues [4]. This improves the preference of the recommendation system through dialogues to address the cold-start problem of collaborative filtering with better user interaction.

II. RELATED WORK

In this work a hybrid approach of combining web scraping, and chatbot with an algorithm for online learning course recommendation is analysed. A series of works had been focused on applying cosine similarity and knowledge graphsalgorithms to recommender systems.

Chan Chun Ho etal.[5] designed an EASElective chatbot to receive user preferences and recommend elective courses. A learning buddy chatbot was used to guide learners in all listening activities by KheFoonHew etal.[6]. An Intelligent chatbot is designed with LDA Bayesian statistical method byYassine Benjelloun Touimi etal.[7] to answer learners' questions through knowledge extraction.

Shabina etal.[8] used web scraping with beautiful soup to getrelevant news articles over the web for creating an Information Retrieval System. An approach to extract metadata of Human activity recognition [9] for applying NLP is proposed by Alamet al.

Rajesh Shrivastava Question Answering chatbot assists users questioner regarding kidney diseases using cosine similarity and TF-IDF[10]. With the Airbnb dataset, a conversational Artificial Intelligence graph-based recommendation algorithm is implemented in SMARTREC by Sudha Vijayakumar[11].

Without identifying the preferences of learners, the effectiveness of MOOC courses does not provide significant outcomes [12]. It opens up more research opportunities in creating a personalized recommendation system.

III.

IV. RECOMMENDATION SYSTEM

Shopping platforms keep on giving recommendations with similar products we search for or music applications create a new playlist according to your preference, it's a good exampleof recommendation systems. A system designed to recommend things to the user based on many different facts orchoices. Recommender systems are basically classified as Collaborative Recommender systems, Content-basedrecommender systems, and hybrid recommender systems.

The Collaborative approach, aggregates recommendations of objects, between the users on the basis of their ratings, and generates inter-user comparisons. Content-based recommender system learns new user's interests based on the features present, It's a keyword-specific recommender system, where keywords are used to describe the items[13]. Ahybrid recommender system combines the features of any twosystems.

This work proposes a model of personalized recommendation with a cosine similarity algorithm and knowledge graph in Neo4j for MOOC course recommendations. The custom dataset used in this paper is web scraped from MOOC sites. The whole system was implemented using 'Python programming language' with a beautiful soup package and Neo4j.

V. WEB SCRAPING

It has automatic web crawlers to obtain large amounts ofdata from websites. Unstructured data from MOOC sites areconverted into structured data in a spreadsheet or a database [14]. With the popularity of MOOCs growing, courses are updated

very frequently. Rather than relying on the traditionaldataset, the information from the MOOC website such as courses offered in different subjects, university of course offerings, and the characteristics of a typical course, reviews, skill set taken into the dataset using web scraping.

VI. PROPOSED MODEL OF RECOMMENDED SYSTEM

To address the cold start problem of collaborative filtering [15], the proposed personalized recommendation model uses content-based recommendation with a knowledge graph and cosine similarity as shown in the Fig.1.

Fig.1. Proposed recommended model

In this section, we define the main components of a graph- based recommendation approach with three main elements. The main elements are

1. User Profile extraction represents all the knowledge requests through the chatbot with the LSTM algorithm, a type of recurrent neural network learns order dependence in sequence prediction problems [16].
2. Course data extraction mechanism through web scraping with the beautiful soup package and selenium tool.
3. A reasoning mechanism to infer the knowledge representation in the form of the knowledge graph and cosine similarity

The course recommendation of the knowledge graph is compared with the cosine similarity model. These techniques are used to provide appropriate recommendations for courses to the learners.

A. KnowledgeExtraction

The first phase of the model uses a chatbot for user profile extraction through a chatbot using deep learning techniques. A conversational system is created to extract users' preferences for courses through interaction and the information is stored in the form of a JSON file [17]. The chatbot is trained on the dataset which contains intents, patterns, and responses. A Deep learning algorithm, LSTM isused to classify the user's message and the learner will get a random response from the list of responses. The chatbot is created using NLTK, Keras, and Python packages, The user preference will be stored in the JSON file format. All data is pre-processed before applying to the deep learning model. First, it undergoes tokenizing, it's a process of breaking the whole text into small parts like words. Then it iterates the patterns, tokenizes the sentence and appends each word in thewords list, then applies the lemmatization. A Deep neural network model is created with three layers used to predict from the pattern it belongs to the user's choice. The model isiterated for 200 epochs, which is a measure of training the neural network for one complete cycle.

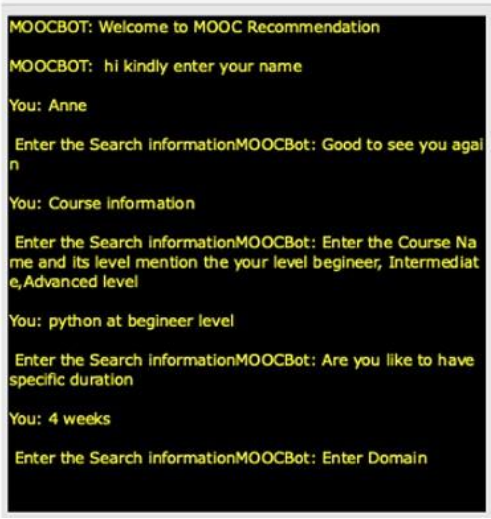

Fig.2. Chatbot-Receiving user response

Chatbot captures user preferences as shown in Fig.2.and the accuracy of the chatbot reaches one with reduced lossis given in Fig 3.

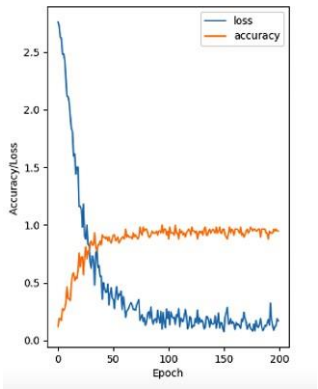

Fig .3. Accuracy of the Chatbot

The proposed recommended system uses the web-scraping methodology with beautiful soup and selenium implementedin the python framework to create the custom dataset from MOOC sites. MOOCs offer accessible and affordable remote learning opportunities to learners. Coursera is one of the leading MOOC providers which offers courses from theworld's leading universities. Data about the courses such as course name, university, course URL, rating of the course, Course duration, and specialization details as shown in Table 1 are stored in the dataset from Coursera and the Edx platforms.

TABLE 1. Details in the Custom Dataset

MOOC Provider	No of Courses scraped	Data scraped
Coursera	8174	Course name University/Industry Level Ratings
Edx	975	Average hours takenURL Skills Specialization Domain Language and subtitles

Web scraping with Python Beautiful Soup library that makes web scraping by traversing the DOM dynamically scraps thedata. With Beautiful, Soup Selenium is used to scrape the Coursera, edX websites for information on all courses offered.

B. *Cosine Similarity model of course selection*

Content-based filtering takes input from the users, checks for user history/past behavior, and recommends a list of similar courses. Cosine Similarity is a machine learning technique used to find the similarity between courses using two non-zero vectors with the formula.1 given here.

$$\text{cosine}(v,w) = \frac{v.w}{|v||w|} + \frac{\sum^{N} v\&w\&}{\cdot \sum_{i}^{N} vi2 \cdot \sum^{N} wi2} \quad (1)$$

TF-IDF is a process of determining the relevancy of a word in a series or corpus is to a text. If the meaning of a word increases in proportion to how many times it appears in the text assigns word frequency as weighted text [19]. The recommendation algorithm results in a range of values between 0 to 1, 0 means not similar and higher values near 1represent similar documents.

Algorithm Recommendation (WebscrapedCourses W, User preferences U)
Input: List of words from W and U
Result: List of courses with high similarity valueBegin:
W and U <- List of unique words Lfor words in L:
for W in words:
Compute TF-IDF value using the formulaReturn weighted text
Calculate the similarity values using the formulareturn the similarity matrix

Fig.4.demonstrates the working of the algorithm when searching for a machine learning course with python from MOOC providers. A list of twenty courses recommended withtheir similarity values is listed in the graphical representation

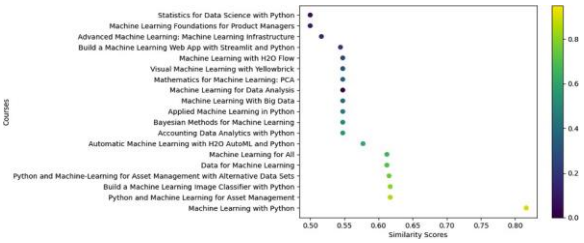

Fig.4.Cosine Similarity values

C. Graph data science model in Neo4j

Graph Neural Network focuses on learning to exploit knowledge to provide better performance with neighbourhood node aggregation.Neo4j is a NoSQL,schema-free graph database. It stores data in the nodes and edges areused to represent the relationship between the nodes. Cyber Query language is used to interact with the graph database [20].Neo4j constructs a graph with all the courses offered byall the universities and as shown in Fig 5.

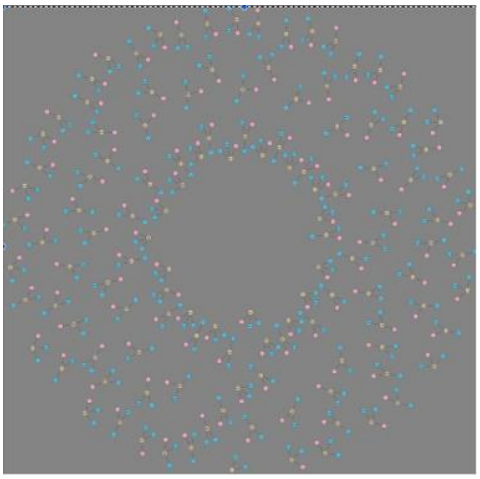

Fig.5. Graphical representation of All Courses.

Learner can't identify the required courses from Fig.5. By filtering out according to the user preference received from thechatbot, the course is recommended. As an example, Python Courses offered by the university of Sydney with the rating and difficulty level of the courses are listed out in Fig.6 through the knowledge graph which helps learners to identify the courses using cosine similarity

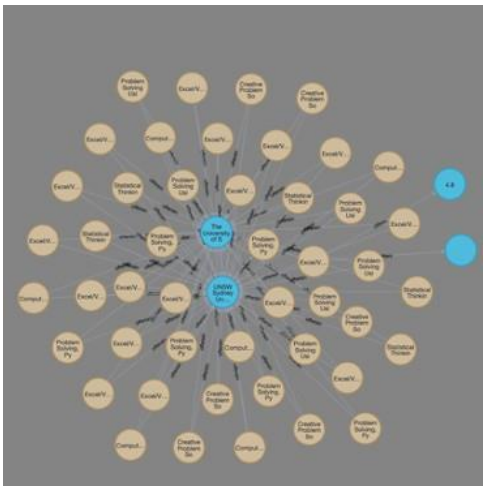

Fig 6. Recommended Courses

VII. CONCLUSION

In this paper, we proposed a hybrid algorithm, to offer personalized MOOC Course recommendation services. It integrates the preferences of learners and course information from multiple MOOC providers. We adopt the TF-IDF-based cosine similarity model and graph algorithms with Neo4j to recommend courses. In future, we will investigate the effectiveness of the knowledge graph in the MOOC recommendation model with an optimized algorithm and scalable algorithm.

REFERE
NCES

[1] Koivisto M,"Experiences on Creating Personal Study Plans with Chatbots. In: Guralnick, D., Auer, M.E., Poce, A. (eds) Innovative Approaches to Technology-Enhanced Learning for the Workplace and Higher Education", TLIC 2022. Lecture Notes in Networks and Systems, vol 581. Springer, Cham. 2022,https://doi.org/10.1007/978- 3-031-21569-8_18

[2] María Consuelo Saiz-Manzanares, Raul Marticorena-Sanchez, Luis Jorge Martín-Anton, Irene Gonzalez Diez, Leandro Almeida,"Perceived satisfaction of university students with the use of chatbots asa tool for self-regulated learning",Heliyon,2023,e12843,ISSN 2405-8440.

[3] https://doi.org/10.1016/j.heliyon.2023.e12843.

[4] Shijun Li, Wenqiang Lei, Qingyun Wu, Xiangnan He, Peng Jiang, and Tat-Seng Chua. 2021,"Seamlessly Unifying Attributes and Items: Conversational Recommendation for Cold-start Users". ACM Trans. Inf. Syst. 39, 4, Article 40 (October 2021), 29 pages. https://doi.org/10.1145/3446427

[5] V. Garg and R. Tiwari, "Hybrid massive open online course (MOOC) recommendation system using machine learning," International Conference on Recent Trends in Engineering, Science & Technology - (ICRTEST 2016), Hyderabad, 2016, pp. 1-5, doi: 10.1049/cp.2016.1479.

[6] C. Chun Ho, H. L. Lee, W. K. Lo and K. F. A. Lui, "Developing a Chatbot for College Student Programme Advisement," 2018 International Symposium on Educational Technology (ISET), Osaka, Japan, 2018, pp. 52-56, doi: 10.1109/ISET.2018.00021.

[7] Hew, K.F., Huang, W., Du, J. et al.," Using chatbots to support student goal setting and social presence in fully online activities: learner engagement and perceptions", J Comput High Educ (2022). https://doi.org/10.1007/s12528-022-09338-x

[8] Benjelloun Touimi, Y., Hadioui, A., El Faddouli, N. & Bennani, S,"Intelligent Chatbot-LDA Recommender System. International Journal of Emerging Technologies in Learning (iJET)", 15(20), 4-20. Kassel, Germany: International Journal of Emerging Technology in Learning. Retrieved January 17, 2023 from https://www.learntechlib.org/p/218336/.

[9] Shabina, Chawla, S.,"Collaborative Approach Toward Information Retrieval System to Get Relevant News Articles Over Web: IRS-Web" In: Buyya, R., Hernandez, S.M., Kovvur, R.M.R., Sarma, T.H. (eds) Computational Intelligence and Data Analytics. Lecture Notes on Data Engineering and Communications Technologies, vol 142. Springer, 2023,Singapore. https://doi.org/10.1007/978-981-19-3391-2_35

[10] Alam, G., McChesney, I., Nicholl, P., Rafferty, J.,"An Approach to Extract and Compare Metadata of Human Activity Recognition (HAR) Data Sets." In: Bravo, J., Ochoa, S., Favela, J. (eds) Proceedings of the International Conference on Ubiquitous Computing & Ambient Intelligence (UCAmI 2022). UCAmI 2022. Lecture Notes in Networks and Systems, vol 594. Springer, Cham. https://doi.org/10.1007/978-3- 031-21333-5_71

[11] Shrivastava, R., Pujahari, A., Singh, S.P., Bhowmik, T,"Efficient Question Answering in Chatbot Using TF-IDF and Cosine Similarity" In: Garg, D., Kumar, N., Iqbal, R., Gupta, S. (eds) Innovations in Information and Communication Technologies. Algorithms for Intelligent Systems. Springer, 2023, Singapore. https://doi.org/10.1007/978-981-19-3796-5_3

[12] S. Vijayakumar and J. G. Jetcheva, "SMARTREC: A Conversational Recommendation System Using Semantic Knowledge Graphs," 2022 IEEE Eighth International Conference on Big Data Computing Service and Applications (BigDataService), Newark, CA, USA, 2022, pp. 53- 60, doi: 10.1109/BigDataService55688.2022.00016.

[13] U.Sujatha, R.Gunasundari,"Student's preferences and factors affecting learning in Online Education", International Journal of Mechanical Engineering 7 (4), 1350-1355, April 2022.

[14] https://python.plainenglish.io/recommendation-systems-content- based-recommendation-d8f9da5bfac

[15] Mazilu, M.C. ,"Web Scraping and Ethics in Automated Data Collection". In: Ciurea, C., Boja, C., Pocatilu, P., Doinea, M. (eds) Education, Research and Business Technologies. Smart Innovation, Systems and Technologies, vol 276. Springer, 2022 Singapore. https://doi.org/10.1007/978-981-16-8866-9_24

[16] Anwar, T., Uma, V., Hussain, M.I. et al," Collaborative filtering and kNN based recommendation to overcome cold start and sparsity issues: A comparative analysis", Multimed Tools Appl 81, 35693–35711(2022). https://doi.org/10.1007/s11042-021-11883-z

[17] https://data-flair.training/blogs/python-chatbot-project

[18] Baena-Perez, R., Ruiz-Rube, I., Mota, J.M. et al. Visual authoring of virtual reality conversational scenarios for e-learning. Univ Access Inf Soc (2022). https://doi.org/10.1007/s10209-022-00934-3

[19] https://towardsdatascience.com/introduction-to-recommender- systems-1-971bd274f421

[20] https://neo4j.com/