

Review

Not peer-reviewed version

Improved E-learning Based Recommender Systems: A Survey

Oswa Maryam , [Humaira Ashraf](#)^{*} , [Noor Jhanjhi](#)

Posted Date: 21 December 2023

doi: 10.20944/preprints202312.1672.v1

Keywords: E-learning; Learning Analytics; Machine Learning; MOOCs; Recommender Systems



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

Improved E-Learning Based Recommender Systems: A Survey

Oswa Maryam ¹, Humaira Ashraf ^{1,*} and NZ Jhanjhi ²

¹ Department of Computer Science and Software Engineering, International Islamic University, Islamabad, Pakistan; (oswa.maryam@gmail.com; humaira.ashraf@iiu.edu.pk)

² School of Computer Science, SCS, Taylor's University, Subang Jaya, Malaysia(noorzaman.jhanjhi@taylors.edu.my)

* Correspondence: author: Humaira Ashraf (e-mail: humaira.ashraf@iiu.edu.pk)

Abstract: Online learning is gaining massive popularity with time. The e-learning platforms operate differently from traditional educational institutions and hence need different strategy for course recommendations. This survey aims to cover the major emerging research areas in e-learning recommender systems. Our study covers different areas of research including graph-based methodologies, ITS, query optimization, content-based, and collaborative filtering, big data, and association rules mining. This survey aimed to explore all major emerging directions of recommender systems in education. This study analyzed existing literature in all of the areas mentioned before and performed an objective-based analysis. A brief performance analysis was also done for the researchers where the values were comparable. Limitations of existing research were also identified and studied to shed some light on future directions.

Keywords: E-learning; learning analytics; machine learning; MOOCs; recommender systems

1. INTRODUCTION

A recommender system is used to filter information according to the user's needs. They play a key role in improving the user's experience and are being used in different fields like e-commerce and entertainment. Recommender systems are generally used to help the user identify relevant information from massive data chunks hence helping the user [1] to make a better decision. One of the very early recommender systems presented by Goldberg et al in the early nineties used collaborative filtering to recommend information to the user. The model presented in the earlier systems is still being used in some hybrid systems for generating recommendations in various fields. The current interest of research in recommender systems lies in fields like e-commerce [2], music [3], movie recommendation [4] specifically Netflix [5]. Education is also an emerging area for recommendations. Most recommendation systems from the past focused on higher education [6] because universities offer several electives and students need counseling to choose appropriate courses. Some of the existing work also considered kindergarten to 12th grade education to recommend career path for young students [7]. Online learning is another subarea where recommender systems can play a key role in improving the learning experience of the students [8]. The increasing trend of e-learning has introduced the concept of Massive Open Online Courses (MOOCs). MOOCs are a massive collection of free resources available online. COronaVirus Disease (COVID-19) pandemic, that hit most of the world in the year 2020 is still forcing educational institutions to shut down to contain the Severe Acute Respiratory Syndrome COronaVirus 2 (SARS-

COV-2) [9]. The sudden shift from traditional classroom to online mode has also dramatically increased the number of new enrollments in MOOCs [10]. More students turning towards online learning has increased the demand of recommender system for MOOCs.

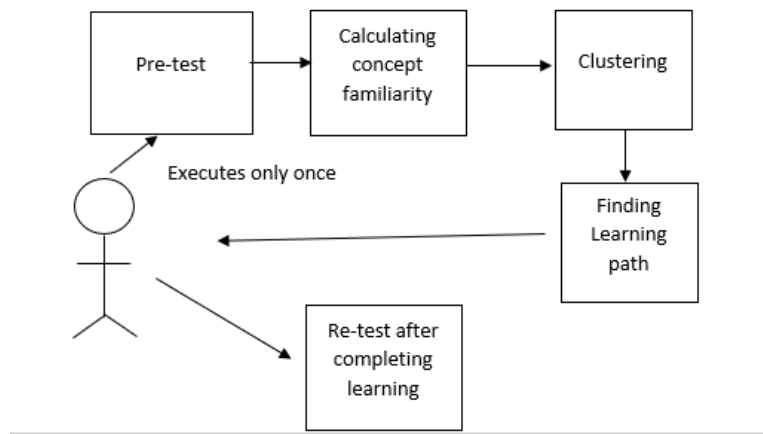


Figure 1. Hybrid methodology to improve learning path recommendation [11].

Recommendation systems for e-learning are different from traditional recommenders because, in traditional education, all courses have specified length, difficulty level and pre-requisites. Furthermore, the dropout rate of students is also low in traditional education as compare to e-learning platforms [12]. Grade prediction and prior knowledge can play a role in generating recommendations in e-learning but other factors like ratings of a course, rating of instructor, relevancy of course, difficulty level are also important. Massive online courses (MOOCs) are playing a great role in revolutionizing e-learning experience but it comes with some limitations like lack of course sequencing and limited personalization [13].

The current trends of a study conducted in this area are more focused on hybrid techniques rather than using just one methodology. The hybrid techniques help in overcoming the limitation of the older techniques. In one such study, Rawat et al. tried to overcome the limitations of collaborative filtering caused by the sparsity of data by utilizing Jena rules using Resource Description Framework (RDF) [14]. Similarly, Nabizadeh et al focused on improving the quality of learning while considering time constraints by allowing the user to choose the first item then utilizing a graph-based recommender that considers both the available time of the users and their prior knowledge [15]. Another research conducted by Rafiq et al utilized queries given by the user to recommend both the relevant courses and the answers to the questions that were asked by a similar level of understanding [11].

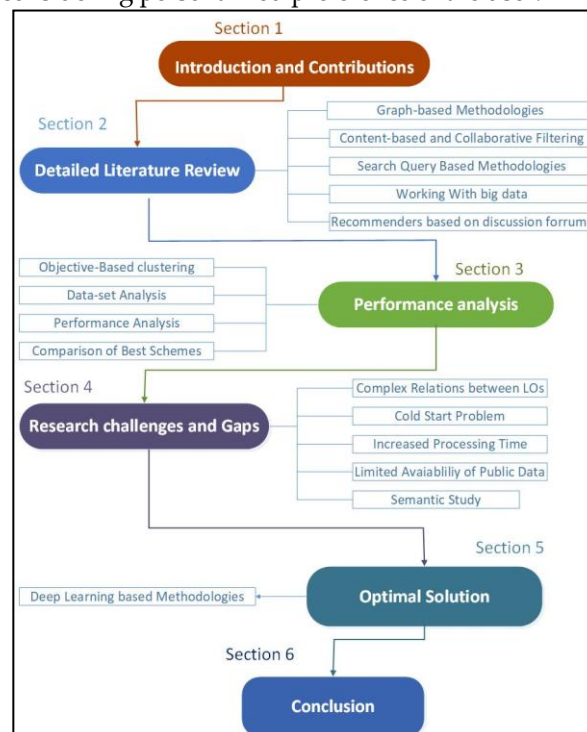
One Research conducted by Niknam & Thulsirman used a hybrid methodology to design an effective recommender system. They utilized fuzzy c-means (FCM) for clustering and Ant colony optimization (ACO) for generating effective learning path recommendations (LPR) for the students which is graphically explained in Figure 1. [16]. In Figure 1, it is displayed that pre-test taken from the user will determine initial understanding of the concepts. Re-tests will be taken side by side to improve the quality of recommendations.

The existing literature in this survey considered different dimensions of the subject under study. This study included but is not limited to MOOC courses. The exiting surveys are summarized in Table 1. They include a study conducted by Mu considered deep learning approaches for all kinds of recommender systems that mostly including music, movies, and books recommender systems [17]. This study did not include existing researches for online course recommender systems.

Table 1. Existing surveys.

| Authors | The topic of the survey conducted | Limitations |
|---------|--|--|
| [17] | Recommender systems using deep learning in the existing literature | This survey was limited only to the deep learning techniques and left major areas like content-based and collaborative filtering |
| [18] | Survey on existing ITS | It only covered ITS while left all other techniques used for recommendations |
| [19] | Surveys on swarm intelligence | It only included researches using swarm intelligence |
| [20] | Recommending experts in a field | Limited scope of query-based search and information retrieval. |
| [21] | Ontology-based recommender systems in e-learning. | Studied only ontology-based models that mainly focus on personalization |

Another survey conducted by Nabizadeh, Leal, et al focused on LPR in intelligent tutoring Systems (ITS). The purpose of their study was the analysis of the long-term goals of learners in existing literature [18]. Another study [19] included the swarm intelligence methodologies for generating recommendations for e-learning platforms. A study conducted by Nikzad-Khasmakhi et al analysed the work done in area of expert recommendation system [20]. The expert recommendation system finds experts based on the query entered by the user while utilizing Natural Language Processing and Information Retrieval methodologies. The study conducted by [21] reviewed work done in ontology-based models to find most relevant study material while considering personalized preference of the user.

**Figure 1.** Hybrid methodology to improve learning path recommendation [11].

Our study majorly includes the recommendation systems that are designed solely for e-learning and does not include other areas like e-commerce and entertainment. The study also includes various techniques and methodologies used for e-learning recommendations. This study will help readers understand different trends of recommender systems in e-learning.

The key contributions of this survey include:

- Exploring existing methodologies to design an effective recommender system in an e-learning environment
- Analyzing existing methodologies and comparing them to identify the most promising area for the future.
- Identify knowledge gaps in the existing literature and identify future challenges.
- Identifying the objective of each research and comparing objectives of existing literature.
- Evaluate the performance of existing methodologies and compare them with each other to identify the one methodology that is outperforming others

The different sections of this study are depicted in Figure 2 that include section 1 that introduces the topic, section 2 that consists of all the literature covered in this study, section 3 contain a comparative analysis of objectives of different researches and comparison of comparable values, section 4 includes gaps and future challenges while section 5 concludes the study with some discussion in detail.

2. Survey Methodology

This survey was conducted on research articles published between the year 2018 and 2020. Query was formulated including keywords recommendation or recommender system, grade prediction and future performance. The data was retrieved from ACM, IEEE, Springer, Google Scholar and Science Direct. The retrieved papers included recommender systems for both online and offline course. We screened the research papers to include only the papers that focused on e-learning and online courses. The resultant relevant papers mostly used different techniques to utilize dataset obtained from log data of various LMS.

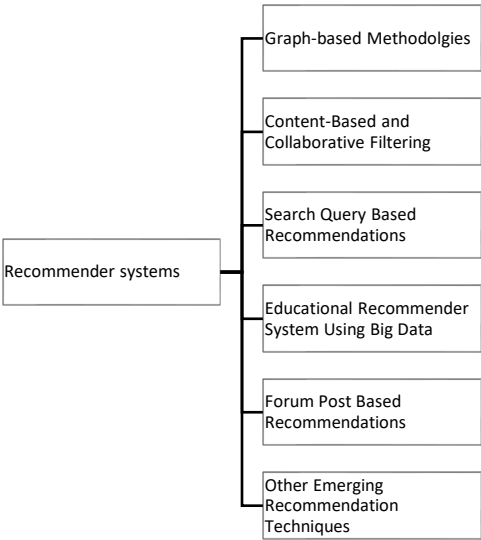


Figure 3. Recommendation techniques for educational recommender system.

The papers relevant recommender systems were further divided into six sub-sections based on the techniques used and were studied separately as shown in Figure 3. There were a few papers that

did not fall under any of the major categories so we included them in a separate section and named it as other emerging recommendation techniques in education.

3. Literature Review

Recommender system in online learning is an emerging area of interest due to the increasing demand for online courses [22]. Different techniques are being used to not only improve the recommendations from an available set of courses but recommendations can also be generated for content available within a course. The following sections will discuss this detail to provide better insight to the readers.

3.1. Graph-Based Techniques

Graph-based techniques focus on organizing content in a graphical structure in form of edges and vertices. Traversing through a graph can help in the identification of a path that can provide learners a road map that they can follow to improve their performance. The graph-based technique can effectively recommend content within a course as well hence increasing the efficiency of the students.

One such study conducted by Nabizadeh et al focused on maximizing the performance of learners' results while taking into account the limited time that a student can spare for a course. They designed a framework that allows users to choose the first item then a graph-based recommender designed using the available time of the users and their prior knowledge based on the item they had chosen first. [15]

Their proposed model focuses on covering all concepts instead of completing all lessons. In the proposed methodology, a two-layered graph is constructed in which one layer consists of lessons while the other layer consists of Learning Objects (LOs). Cluster means and cluster median is used to create groups of slow, fast, and medium learners while matrix factorization estimates the time expected to be taken by the user to complete a specific LO. The path that takes minimum time is recommended to the user.

The experimental results indicated that the cluster median algorithm performed well in most cases by showing around 90% accuracy. The model was implemented, and an online evaluation was also performed. The results showed that the control group performed well only in initial lessons while the experimental group showed average performance in all lessons.

Another similar study was conducted to provide a personalized guideline to the students in the selection of a course by using several techniques like content analysis, assessment, and behavioral analysis [23]. This study is based on Bayesian knowledge tracing (BKT). BKT is an statistical algorithm that helps to infer the current knowledge state of students [24]. BKT is used in educational data mining and learning analytics to model and analyze student learning. The algorithm is based on Bayesian probability theory and is used to estimate the probability that a student has mastered a particular skill or concept based on their responses to a series of questions or exercises. This helps to predict if they have learnt a skill or not.

Four variations of BKT were used for knowledge, tracing in the proposed model including Coarse BKT, multi-grained BKT, Fine BKT, and Historical BKT [23]. Experiments were performed to predict dropout rate and a t-test was conducted on the results which indicated that the multi-grained BKT showed the best performance due to its ability to handle hierarchal structure.

Another study conducted by Wan & Nu focused on the personalization of recommendations using LOs [25]. For existing users, their proposed model stores a profile that can be updated with time while for new users, the system offers a survey that helps in constructing a user's profile.

Recommendations are generated by constructed ontology-based graph using the user's profile and similarity between LOs. The model also considers learners' learning styles and learning goals while recommending material.

An Intelligent Tutoring System (ITS) was proposed by Muangprathub, Boonjing, and Chamnongthai. The focus of this research is to adapt the system according to the requirements of the user based learning style. The system combines information of the learner including knowledge state and learning styles with the existing knowledge repositories along with various teaching strategies [26]. The information about the course was extracted in form of objectives, contents, exercises, and tests. A personalized path was generated for each user utilizing the knowledge-based graph using the co-occurrence of different objects. To test the proposed methodology, a web-based application was deployed, and the performance of the system was measured. The results showed that the users who interacted with the proposed model performed significantly better with a mean score of 10 as compared to the control group with a mean score reaching 5.

Similar research conducted by Zhou et al utilized the Long Short-Term Memory (LSTM) model of deep learning along with personalized path generation [27]. LSTM networks were introduced by Hochreiter and Schmidhuber [28]. LSTM networks are designed to overcome the vanishing gradient problem in standard RNNs. LSTM networks have a unique memory cell that allows them to selectively store and access information over extended time periods.

To overcome the cold-start problem, their methodology proposed a separate path where sufficient prior information was not available to generate recommendations [27]. The users are also divided into five main clusters based on their interaction with the platform and then the learning effect is calculated using the LSTM model. The results of their proposed methodology indicated that using clustering along with the LSTM model showed the best performance with a precision value of around 0.45. The value of precision for LSTM after clustering increases as the number of learners' increases hence showing its effectiveness for large amounts of data.

One research conducted by Niknam & Thulasiraman used hybrid methodology using both clustering and graph-based method to design an effective recommender system. This study focuses on Learning Path Recommendation (LPR) using Ant Colony Optimization (ACO). Ant colony optimization was originally called Ant system and was proposed by M. Dorigo et al. [29] ACO is a kind of swarm intelligence and is a metaheuristic optimization algorithm inspired by the behavior of ants searching for food. The algorithm is often used to solve combinatorial optimization problems. In ACO, a colony of artificial ants iteratively constructs a solution to the problem by moving through a graph or network of possible solutions. In the research conducted by [16] the learning objects are mentioned as concepts that are extracted by using the concept Map tool. In their proposed methodology, learners take a pre-assessment test to identify the concepts familiar to the user. Fuzzy C-means (FCM) clustering is used to assign learners to a cluster. FCM was first introduced by Dunn [30]. The algorithm is based on the principles of fuzzy logic and set theory. It is an extension of the K-means algorithm that allows for soft clustering that allows data point can belong to multiple clusters with varying degrees of membership. FCM was chosen in the research [16] because it doesn't create rigid clusters but creates softer boundaries. ACO algorithm helps in finding the most effective path for new learners by using data of old learners of the same clusters. The progress of the learners is continuously monitored and change in recommendation is made as and when required by continuously updating the graph. They implemented their algorithm using java and a user interface was created to test it for real-world data. Fifty learners participated in the study for three weeks. In the pre-assessment test, both groups scored an average of 30 marks while the experimental group performed better in the post-test with an average of 65 marks while the control group students achieved an average of 58 marks.

Some researches in this area are based on query-based system where the user enters a query and system generate a list of recommendation based on that query. One such study was conducted by Shi et al. in which their proposed model used a graphical structure to store LOs and then recommended path according to the profile of a user [31]. The knowledge graph used in this research utilized LOs and recorded the semantic data regarding the relationship between these LOs. Their proposed methodology constructed a multi-dimensional knowledge graph that stores class-wise hierarchical structure that includes basic knowledge, algorithm, and tasks. All these classes include LOs but from different perspectives. Their framework initially inputs a query from user and then collects the target LOs after tokenizing the input query. The semantic relationship between LOs were utilized to generate a more effective path. A system can generate multiple paths for one query. In that case, the recommendation algorithm considers the weight of features for all learning paths while assigning the highest score to the most popular feature. The system recommends the path with the highest score.

They conducted experiment including test and control group. The test group showed much higher user satisfaction with scores reaching up to 0.9 while the control group showed an average reaching up to 0.45.

The research is graph based technology mostly utilized learning objectives to either recommend a path or to generate query response. The Table 2 presents a comparison of results indicating how different methodologies utilized different evaluation measures depending upon the nature of problem and proposed solution. Some methodologies that used similar methodologies such as [16] and [26] indicated that clustering along with swarm intelligence can increase learning gains.

Table 2. Result compasion of graph-based methodologies.

| Reference | Methodology | Performance measure | Results |
|-----------|-------------------------|--------------------------|---------|
| [15] | Cluster mean | Accuracy | 90% |
| [16] | FCM and ACO | Learning gain | 33 |
| [23] | Historical BKT | Area under curve | 0.7695 |
| | | P-value | 0.0003 |
| [25] | Ontology based graph | Average Computation time | 0.07s |
| [26] | Formal concept analysis | Learning gain | 10.63 |
| [27] | LSTM | Precision | 0.45 |
| [31] | Knowledge graph | User satisfaction | 0.9 |

3.2. Content-Based and Collaborative Filtering

Content-based filtering and collaborative filtering are mostly used together in hybrid systems because they help to overcome each other's limitation when they are used together. Some existing literature used both techniques together while utilized another technique as well to further enhance the performance.

3.1.1. Content-Based Filtering

Content-based filtering is a type of recommendation systems. It generates recommendations to users based on their past behavior and preferences. It works by analyzing the characteristics of items that a user has already liked or interacted with and recommending items that are similar to those items. One of the benefits of content-based filtering is that it can provide personalized recommendations to users even if they are new to the platform. The earliest content based systems can be traced back to mid-nineties when researchers developed a recommendation system called Fab, which used item features and user profiles to recommend movies [32] while another system called

GroupLense was developed to recommend news items [33]. That recommender used hybrid system by utilizing both content based filtering and collaborative filtering.

This section of survey will cover the existing research that included content-based filtering along with collaborative filtering to improve recommendation results. Once such a study was conducted by Xiao et al. that utilized content and collaborative filtering for relevant recommendations for MOOC courses [34]. Their proposed framework picks one of the techniques depending upon a threshold value. If the threshold value is met, then the algorithm will take the path towards collaborative filtering but if the dataset does not have enough previous ratings to meet the defined threshold then the algorithm will take the path of content-based filtering.

Another research conducted by Neamah & El-Ameer proposed a framework that utilized the course description, user's profile, and ratings to determine the most relevant course for a user [35]. They made it mandatory for the instructor to enter a course description. The user's preference and profile are then compared with the course description to determine relevant courses. On the other hand, the user's profiles are also updated based on the courses they picked.

Their proposed methodology consists of two phases. The first phase uses K-nearest neighbour (KNN) classifier to determining the most relevant courses based on the user's profile and this portion of the proposed methodology will be executed for every recommendation. The second phase is optional and is executed only when enough ratings are available in the database for the relevant courses. In the second phase, the Naïve Bayes classifier is used to recommend courses based on the user's rating.

The results proposed methodology were measured in terms of precision, recall, and f-measures. KNN showed .43 precision, .50 recall, and .46 f-measure value while Naïve Bayes showed 0.75 precision, 0.50 recall, and 0.60 f-measures. Apparently, Naïve Bayes showed better overall precision but it should be considered that naïve Bayes is executed only for relevantly complete data hence showing better results. This also indicates the importance of user's rating in course recommendations as an important factor for online course recommendation.

The online learning platforms are different from traditional education institutions as they do not have a predefined sequence in which different courses should be taken. This can cause a problem for learners especially the beginner who does not know which course should be taken first. To solve this problem research was conducted by Tarus et al that utilized a hybrid recommender system that included content-based collaborative filtering and sequence pattern mining [36]. This helps in identifying a sequence in which courses should be studied.

Their model considered learner's profile that included both demographics data and the ontology-based data created by recording user's interaction with the system, LOs that were derived from the learning material and content, weblogs that stored user's interaction with the system. The proposed framework used both content and context-aware course recommendations by utilizing the weblogs, LOs and user profiles. The generalized sequential mining (GSM) algorithm was used for identifying the recommended sequence of courses. GSM is a data mining technique used to discover frequent sequential patterns in a sequence database. It was first introduced in 1993 by Agrawal et al [37]. GSM first converts database into a set of generalized sequences hence generating a sequence of item sets. Each item set in a represents a set of items that occur together in a transaction or event.

The result of the hybrid methodology proposed by [36] was compared with individual algorithms and the proposed hybrid model showed the best precision and recall values as compared to any of the three methodologies (CF, CB, or GSM) with 0.35 precision and 0.45 recall.

Another similar research conducted by Klačnja-Milićević et al used Apriori to find pre-requisites and utilized different learning styles while recommending online courses [38]. Their proposed methodology utilized tag-based Collaborative filtering or tagging which is an application of

collaborative filtering. Tagging-based techniques are natural language processing methods that involve labeling words or phrases in text with tags that represent their part of speech, named entity, sentiment, or other characteristics. In their proposed method, the learners will write tags according to their requirements.

Their proposed framework was implemented on an online learning management system called Protus which is a programming tutor. Initially, the user completes registration which includes a survey that helps in the identification of learning style. Every newly registered user is divided into a cluster based on his/her learning styles and the courses will be presented to the user according to his/her learning style. The tags were created by the user so the results of the proposed solution indicated that only 55% were covered when they were matched with tags created by experts. In the future, some suggested tags can also be utilized keeping in view the experts' generated tags to ensure vast coverage of LOs.

Another study conducted by De Medio et al introduced another dimension to the existing research by considering the perspective on instructors in course recommendation [39]. They proposed a teacher-centric recommender system that helps instructors in identifying courses similar to those they are interested in teaching. The proposed MoodelREC system utilizes popular LO repositories to crawl LOs. Then both content-bases similarity and collaborative filtering are used to rank the LOs. The instructor enters a query in the provided interface which is utilized to generate recommendations. Their proposed methodology was implemented and tested by comparing the predicted rating with the actual rating of the user. The results indicated a precision of 0.22 and a recall of 0.39. Hence indicating a lot of room for work in this area in the future.

3.1.2. Collaborative Filtering

Collaborative filtering is a type of recommendation algorithm that utilizes the behavior and preferences of multiple users in order to recommend items to a specific user. One of the very early recommender systems using collaborative filtering was presented by Goldberg et al in the early nineties. The presented model presented is still being used in some hybrid systems for generating recommendations in various fields [1] and is based on the idea to collect and analyze data from users having similar preferences in the past and considering that they are likely to have similar preferences in the future. Collaborative filtering uses the ratings or feedback provided by users to predict their preferences for items they have not yet interacted with. There are two main types of collaborative filtering: user-based and item-based. User-based collaborative filtering recommends items to a user based on the ratings of similar users. Item-based collaborative filtering recommends items to a user based on the similarity of the items to items the user has already rated. Collaborative filtering cannot be applied alone due to its limitations like the cold start problem. Different methodologies like learning styles prediction, content-based filtering, association rules mining, sequential pattern mining or deep learning are used along with collaborative filtering to overcome its limitations.

One such study conducted by H. Zhang et al combined a deep belief network with collaborative filtering for more relevant recommendations [40]. For this research, they used a dataset from real users in StarC platform. The dataset after pre-processing is fed into both supervised and unsupervised Deep Belief Networks (DBN). DBN is a type of Artificial Neural Network (ANN) that is composed of multiple layers of interconnected neurons. It is mostly used for unsupervised learning tasks such as feature extraction and dimensionality reduction and was initially introduced by [41]. The proposed methodology in [40] utilized ratings of users and used them for supervised DBN while unsupervised DBN doesn't consider the user's rating. They used Restricted Boltzmann machines (RBM) for constructing neural networks for unsupervised learning. RBM is a type of artificial neural network

that is used for unsupervised learning tasks [42]. The results showed a minimum RMSE of 67.48% for their proposed methodology hence outperforming the existing methodologies [40].

The dropout rate in MOOCs is higher and greatly depends on user satisfaction. Hence the dropout rate can be reduced by increasing the user's satisfaction. One such study conducted by Pang et al utilized collaborative filtering along with learning series techniques to maximize user satisfaction [43]. They proposed a methodology that is based on Recommendations with Learner's Neighbour and Learning Series (RLNLS). The RLNLS model used inverse distance weight to calculate the past achievement of the user which can be used to predict their future achievements. Collaborative filtering technique was used to find the behaviour of neighbor learners while Hawkes process was used to predict the learning intensity. The recommendations hence generated utilized both the information regarding neighboring learners and the capacity of the target learner. The RLNS model showed the most accurate results with a precision value of 0.34, recall value of 0.38, and f-measure value of 0.35.

The data that is being collected from online learning platforms is getting sparse with time. This can compromise the efficiency of collaborative filtering algorithms. A study conducted by Rawat et al. tried to overcome the limitations of collaborative filtering caused by the sparsity of data [13]. Their research used collaborative filtering while utilizing Jena rules using resource description framework (RDF). Their proposed methodology first collects information from the user using Moodle then calculated its RDF representations so it can be further utilized by Jena rules engine. The Jena Rules engine is a forward-chaining rule engine that allows developers to define rules to infer new knowledge from existing RDF data. KNN is applied to further refine the results and then final recommendations are generated. The recommendation list is evaluated and then is presented to the user.

For the evaluation of the proposed methodology, they used RMSE, precision-recall, and f-measures. The active users' data showed results of 0.25 RMSE, 0.98 precision, 0.90 recall, and 0.93 value for f-measures. The average users' data showed results of 0.48 RMSE, 0.97 precision, 0.84 recall, and 0.90 value for f-measures. The non-active users' data showed results of 0.51 RMSE, 0.64 precision, 0.27 recall, and 0.38 value for f-measures.

Research conducted by Obeidat, Duwairi, and Al-Aiad proposed a model that utilizes the courses opted by the student in the past to recommend similar courses to the new students. [22]. They utilized collaborative filtering along with clustering and association rules mining for better course recommendations. They utilized sequence mining that helped to determine prerequisites and in recommending courses in the correct sequence. For clustering, they used k-means algorithm that uses Euclidean distance to groups, similar students, together. Apriori algorithm was used for association rules mining while SPADE (Sequential PAttern Discovery using Equivalence classes) algorithm was used for sequence rules mining. Apriori algorithm works by first finding all frequent itemsets in the dataset [44] while SPADE is used to discover clusters in high dimensional dataset [45]. The coverage, support, and confidence were considered while ranking the generated association rules. The highest-ranked rules were utilized to generate recommendations. An analysis is performed on the prediction of future grades. If students are predicted to score higher, then the course is recommended to the students otherwise it is excluded from the list of recommendations.

The results indicated that 5 is the most optimal value for the number of clusters. Both Apriori and SPADE algorithms showed better coverage of courses after clustering as compared to recommendations generated without utilizing clustering. Apriori showed 60% coverage with clustering while without clustering it showed 48% coverage. Similarly SPADE algorithm showed 56% coverage with clustering while without clustering it showed 51% coverage.

Another study conducted by Murad et al. explored the recommendation of study material within a course at early stages [46]. This research mainly focused on improving the performance of students

in an online course by recommending to them the most relevant study material in time. Their proposed solution uses the profiles of old students to predict and recommend relevant study material to the new students. Their proposed framework combines collaborative filtering with rule-based filtering and KNN classification to create a hybrid system. The recommendation hence generated will be context-sensitive and will depend on the choice of the user. The results showed that adding context sensitivity showed the same Mean Square Error (MSE) as the other baseline methodologies that did not utilize context-sensitive information. This indicated the ability of the proposed methodology to add personalization in a system without compromising the quality of recommendations.

On an online platform, different courses with similar titles may include the content of varying difficulty level. One such research was conducted by Segal et al. that focused on calculating the relevance of a course based on its difficulty level [47]. Their proposed EduRank method used collaborative filtering along with social voting to evaluate the relevance of a course. The algorithm first collects data from students and assign ranking based on the answers. For each question, the social choice methodology is applied and for this purpose, the Copeland method is chosen. The Copeland method helps in collecting votes that label a question as difficult, not difficult or tie. Item response theory and Bayesian knowledge tracing were used for calculating item difficulty. If questions were solved quickly then they were easy, it took more attempts than were difficult. AP correlation matrix was used to evaluate the proposed method and perform a comparison with competing methodologies.

The results indicated that the AP correlation for the EduRank algorithm was much higher as compared to the competing methodologies like SVD, EngenRank, and topic-based ranker (TBR) indicating its better performance.

3.1.3. Matrix Factorization

Matrix factorization is a type of collaborative filtering technique that is most commonly used in recommender systems. Its use in educational recommendation system is also prevalent. One such study conducted by Symeonidis & Malakoudis used multi-modal matrix factorization for course recommendation in MOOCs platform [48]. The proposed methodology presented a new dimension of user and course skill set. This matrix calculates either a course provides a particular skill set that a user is looking for. The most frequent items in the neighborhood are ranked the highest and they appear higher on the list while recommending courses to the user. For calculating the similarity between the users, the proposed technique used cosine similarity.

Their results indicated that the user-item matrix generated only by utilizing ratings showed higher RMSE as compares to the matrix that included course content as well as the ratings. The results indicated that ratings with course content and student's skill set showed minimum RMSE value.

The CF and CB methodologies are widely used in recommendation systems but may use a different methodology to design a hybrid solution. The comparison of existing methodologies are compared with each other in Table 3 that indicates that most of the existing research in this area such as [13], [35] and [36] used precision, recall and RMSE or MSE for evaluation. Some researches such as [22] and [38] also utilized coverage as evaluation measures. The results indicated that using association rules mining along with CF or CB increased precision and recall and decreased RMSE. Association rules mining also improved results for coverage as well.

Table 3. Comparison of results for content-based and collaborative filtering techniques.

| Reference | Methodology | Performance measure | Results |
|-----------|-----------------------------|---------------------|---------|
| [13] | CF and jena rules using RDF | Precision | 0.98 |
| | | Recall | 0.90 |

| | | | |
|------|--|-----------------------------------|-------|
| | | RMSE | 0.25 |
| [22] | Apriori with clustering | Coverage | 60% |
| | Apriori without clustering | | 48% |
| | SPADE with clustering | | 56% |
| | SPADE without clustering | | 51% |
| [35] | KNN and naïve bayes | Precision | 0.75 |
| | | Recall | 0.50 |
| [36] | Hybrid (CF, CB and GSM combined) | precision | 0.35 |
| | | Recall | 0.45 |
| [38] | Tagging and Apriori | Coverage | 55% |
| [39] | CB and CF | Precision | 0.22 |
| | | Recall | 0.39 |
| [40] | DBN and RBN | RMSE | 0.67 |
| [43] | RLNLS | Precision | 0.34 |
| | | Recall | 0.38 |
| [46] | CF and KNN | MSE | 21.4 |
| [47] | Copeland method and Bayesian knowledge tracing | AP correlation (higher is better) | 0.77 |
| | | NDPM (lower is better) | 0.28 |
| [48] | Matrix factorization | RMSE | 0.491 |

3.3. Search Query Based Recommendations

Searching relevant material by entering a search string is a popular method to fetch relevant material on the internet. Due to the complex nature of an educational setting and an increasing number of similar online courses, there is an emerging need to interpret the query entered by the user according to his or her personal needs. One such research conducted by Lalitha and Sreeja proposed a method that majorly focuses on the older audience who are self-learning and do not have access to any recommendations from teachers or instructors [49]. Their research aimed to design a system that can accommodate learners who have no idea from where they should start learning. Content based filtering, collaborative filtering, and web mining were used to identify appropriate learning material on the internet which was later narrowed down according to the personal needs of the learners. Clustering is done using KNN to group similar learners in one cluster to maximize personalization. Once a user enters a query, the recommender system consider information extracted from the learner's profile and the nearest neighbor. The recommendations are generated based on similarity checks.

Another similar study conducted by Rafiq et al utilized the intelligent query optimization technique by introducing multiple agents to optimize generated recommendations [11]. Their methodology focused on prompting the user to enter information and to also give feedback regarding generated recommendations. Association rules mining is utilized for generating the most relevant recommendations. The results of their proposed methodology showed around 90% accuracy, precision, and recall values that fluctuated a little as the number of lectures were increased but all values stayed up to 89%.

The research covered in this section utilized Association rules mining, KNN and knowledge based filtering. Their results are compared in Table 4.

Table 4. Comparison of results for query based techniques.

| Reference | Methodology | Performance measure | Results |
|-----------|-----------------------------------|---------------------|---------|
| [11] | Association rules mining | Accuracy | 90% |
| [49] | KNN and knowledge based filtering | Recall | ---- |

3.4. Educational Recommender System Using Big Data

The increasing number of online learning platforms and learners is generating massive amounts of raw data that can be utilized to generate recommendations for future learning. Big data refers to extremely large and complex data sets whose processing is beyond the ability of traditional data processing and management tools. There are several existing recommender systems in online learning that utilize big data to generate more effective recommendations. One such study was done by Zhang et al proposed a course recommendation model that utilized big data to generate association rules in a distributed environment to generate the most relevant related courses that can help in generating ranked list recommended courses for online platforms [50]. They utilized Hadoop Distributed File System (HDFS) while using Spark for data processing for association rules mining and Sqoop to import and export datasets. Course recommender used the Apriori algorithm to generate recommendations of the most relevant courses.

Another similar study conducted by Dahdouh et al also utilized HDFS and Spark for big data analysis [51]. The purpose of their study was to discover the most relevant course by implementing association rules mining on big data. For this purpose, they used Spark ML libraries to implement the FP-Growth algorithm. For course recommendations, they analyzed the behavior of learners who interacted with the system in past by utilizing course enrollment and log data. The same model was studied by using more traditional systems like weka and R. The comparison with SparkML with Weka and R indicated that SparkML took minimum execution time while execution time was maximum for Weka.

The data collected from online learning platforms is not only massive but also is very diverse in nature due to its access to a global audience. Education is a subjective term so qualitative analysis is also important in this area to generate better recommendations. The research conducted by Hou et al utilized the feedback of the users to improve the model [52]. They proposed a methodology that divides all courses in different clusters. When users enter the systems, they are recommended courses based on their demographic data. For users who have visited the system earlier, a learning profile is also utilized that includes information regarding courses taken in past. A personalized tree is constructed for each user and courses are recommended based on that personalized tree. If the user is satisfied with the recommendation, the recommender system is rewarded for it and it learns it as a success. This information can help the system in future recommendations. Similarly, unsatisfactory feedback will generate regret and the system will also learn and update its recommendations for the future. The resultant recommendation not only covers personalization but also the course sequencing by observing the course selected in a sequence by past students.

Research conducted using big data mostly utilized Association rules mining and reinforcement learning. Different methodologies are compared in Table 5.

Table 5. Comparison of results for methodologies using big data.

| Reference | Methodology | Performance measure | Results |
|-----------|------------------------|-----------------------------|--|
| [50] | Apriori | Analysis of generated rules | ---- |
| [51] | FP-growth | Execution time | 34s |
| [52] | Reinforcement learning | Average regret comparison | Decrease in regret as rounds increased |

3.5. Recommendations Using Forum Posts

Just like online courses, online forums are another educational resource which is widely popular among the learner's especially when it comes to fetching quick response. Several online forums provide answers to popular questions. The recommender system in this scenario can help in extracting information from these discussions. One such study conducted by Porcel et al utilized a hybrid technique based on fuzzy linguistics and ontology-based approach [53]. Fuzzy logic is a type of mathematical logic that deals with approximate reasoning [54]. Fuzzy linguistics does not views linguistic expressions as either true or false. It is helpful in capturing the complexity of natural language. An ontology is a formal specification of a shared conceptualization of a domain that provides a common vocabulary and a shared understanding of the entities, concepts, and relationships between them. The purpose of their research [53] was to identify the interest of the users and present recommendations based on their interests. They proposed a framework called sharing notes that is a web-based forum designed to facilitate students by providing them personalized material. Sharing notes utilized collaborative and content-based filtering to rate the material while the fuzzy linguistics model was implemented to perform the qualitative study on the communication that takes place on the forum between different users. Ontologies are used to map the relationship between users and their preferred material. Trust is also calculated for the resources posting material. The final recommendations are generated after obtaining ratings from the trusted users. Sharing notes was tested with the help of 425 users amongst those 88% of the users show satisfaction with their proposed recommender system.

Another similar study conducted by Albatayneh, Ghauth, and Chua focused on the discussion forums of online learning platforms [55]. The previous methodologies were based just on negative or positive ratings and new recommendations were generated using content-based filtering. Their proposed methodology added the layer of semantic analysis as well. Semantic analysis is the process of understanding the meaning of language by analyzing the meaning of individual words, their relationships to other words in a sentence, and the overall meaning of the text. They implemented and tested their model in a form of a case study. The results showed that semantic based content filtering technique with negative ratings showed a minimum MAE of 0.25.

Another forum-based study conducted by Ntourmas et al pointed out the limitation of platform independence in existing research [56]. They aim to solve this limitation by increasing the transferability of the system to make it applicable on other platforms as well. They proposed a solution that performs semantic analysis on messages posted on discussion forums and implemented this methodology on two MOOCs. For this purpose, they divided posts into three categories, content-related posts, logistic-related posts and posts on which no action is required. Support Vector Machine (SVM) was utilized in the end to generate recommendations for relevant posts. SVM is based on the idea of finding a hyperplane that separates the data points into different classes. The data points that

lie closest to the hyperplane are called support vectors [57]. The results of their proposed methodology indicated up to 65% accuracy, 0.59 precision, 0.65 recall, and 0.61 value for f-measure.

Most of the research in forum post discussion context used semantic analysis because in forum post discussion understanding the language and emotion part is important. Methodologies like fuzzy linguistics [53] and semantic analysis [55] are the main focus in this area.

Table 6. Comparison of results for forum post discussion.

| Reference | Methodology | Performance measure | Results |
|-----------|---|---------------------|---------|
| [53] | Fuzzy linguistics and ontology-based approach | User satisfaction | 88% |
| [55] | Semantic analysis and collaborative filtering | MAE | 0.25 |
| [56] | Semantic analysis and SVM | Precision | 0.59 |
| | | Recall | 0.65 |
| | | Accuracy | 65% |

3.6. Other Emerging Recommendation Techniques in Education

Apart from the techniques and methodologies mentioned in previous literature, some new areas are being explored. One such study conducted by G et al focused on the group of young adults from the generation Z who have different learning styles as compared to older generations [58]. Their proposed methodology takes the data of online courses as input and ranks them into various difficulty levels. Similarly, students create their profiles and their data is also extracted through the log files that record the data of a student interacting with the system.

Another research conducted by Dwivedi, Kant, and Bharadwaj proposed a methodology to optimize the course sequencing which can be a very challenging task for online course recommendations [59]. They collected data from the learner's profile and used variable-length genetic algorithms for learning path recommendations utilizing data of both new and old students. Collaborative filtering was used to find the popularity matrix which was later utilized by a genetic algorithm for personalized course sequencing.

Another research conducted by Lin, Li, and Lian aimed to solve the problem of goal-oriented generation of recommended courses by using a convex optimization algorithm [60]. Their study focused on university students as their audience and proposed a data-driven approach based on student's ratings and performance. Their proposed modified clustering algorithm that embeds optimization technique showed more than 90% accuracy by outperforming the existing techniques [76–89].

Other methodologies that used different approached along such as genetic algorithm, convex optimization and SVM along with collaborative filtering utilized different evaluation measure. Their results are compared in Table 7.

Table 7. Comparison of results for other emerging approach.

| References | Methodology | Performance measure | Results |
|------------|---|---------------------|--------------------------|
| [58] | SVM | User satisfaction | Satisfactory performance |
| [59] | Collaborative filtering and genetic algorithm | Learner's score | 90% |
| [60] | Convex optimization algorithm | Accuracy | 90% |

3.7. Critical Analysis

All of the literature reviews discussed so far have been summarized in Table 8 which includes techniques and limitations of major researches. The limitation of few types of research was covered by newer researches in the existing literature and they are included in this research. For example faced cold start problem [15] was overcome by utilizing content-based filtering along with collaborative filtering [14].

Table 8. Critical analysis.

| Author | Technique | Limitations and future work |
|--------|---|--|
| [14] | Their research used collaborative filtering while utilizing Jena rules using resource description framework (RDF). After utilizing the RDF and Jena rules engine, an enriched rating matrix will be generated. KNN is applied to further refine the results to generate final recommendations. The recommendation list is evaluated and then is presented to the user | Their proposed methodology utilized manual method to update RDF factbase and this method can be automated in future. [14] |
| [15] | Their proposed model focuses on covering all concepts instead of completing all lessons. Their proposed method also monitors the learner's progress with the passage to time and if the recommended path is not showing satisfactory results, then the recommender recommends auxiliary LOs. The auxiliary LOs are the highest ranked LOs which has not been seen by the user | Their proposed methodology faced coldstart problem for initial LOs that can be solved in future work by automatically selecting the first LO. [15] |

| | | |
|------|---|---|
| [11] | <p>Their methodology focused on prompting user to enter information and to also give feedback regarding generated recommendations. Association rules mining is utilized for generating most relevant recommendations.</p> | <p>In future, Swarm intelligence and genetic algorithm can be utilized to improve results [11].</p> |
| [16] | <p>Their research used hybrid methodology using both clustering and graph-based method to design an effective recommender system. This study focusses on LPR using ACO. Fuzzy C-means (FCM) clustering is used to assign learner to a cluster.</p> | <p>The proposed FCM and ACO are computationally intensive. They can be parallelized in future for more effective performance. In future, the same model can be tested for large audience with more diverse demographics [16].</p> |
| [22] | <p>They utilized collaborative filtering along with clustering and association rules mining for better course recommendations. They also utilized sequence mining that helped in recommending courses in correct sequence. For clustering, they used k-means algorithm that uses Euclidean distance to groups similar students together. Apriori algorithm was used for association rules mining while SPADE algorithm was used for sequence rules mining. If students are predicted to score higher, then the course is recommended to the students otherwise it is excluded from the list of recommendations.</p> | ---- |
| [23] | <p>They utilized several techniques like content analysis, assessment and behavioral analysis. Four variations of BKT were used for knowledge, tracing in the proposed model including Coarse BKT, multi-grained BKT, Fine BKT and Historical BKT .</p> | <p>In future, words embeddings can be used to construct knowledge graph instead of word co-occurrence to improve the quality of constructed graphs [23].</p> |
| [25] | <p>This study is based on the concept of self-organizing LOs. The recommendation is based on both user's explicit requirements and implicit needs that are captured by</p> | <p>The one case study was not enough to test their proposed method. In future more case studies can be conducted to verify the</p> |

| | | |
|------|--|---|
| | analyzing user's behaviour. Recommendations are generated by constructed ontology based graph using user's profile and similarity between LOs. | effectiveness of their proposed methodology [25]. |
| [26] | An ITS was proposed in this research that aimed to adapt the system according to the user's requirements based on his or her learning style. A personalized path was generated for each user utilizing the knowledge-based graph using co-occurrence of different objects. | Their proposed model has been tested only on limited set of students. More extensive evaluation can be performed on proposed model in future. |
| [27] | Their methodology proposed a separate path when there is no sufficient prior information to avoid cold start problem. The users are also divided into five main clusters based on their interaction with the platform and then learning effect is calculated using LSTM model. | In future, data streams of learners can be utilized to improve the generated learning paths. |
| [31] | Their proposed methodology constructed a multi-dimensional knowledge graph that stores class wise hierarchical structure that includes basic knowledge, algorithm and tasks. Their framework inputs a query form user, tokenize it and then collects the target LOs. It is possible for a system to generate multiple paths for one query. In that case the system recommends the path with highest score. | In future this system can be implemented for subjects other than computing and user's profile can also be considered in further researchers. |
| [34] | Their methodology utilized content and collaborative filtering for relevant recommendation for MOOC courses. Their framework picks one of the techniques depending upon a threshold value. If threshold is met then collaborative filtering is used otherwise the algorithm will take the path of content-based filtering. | ---- |

| | | |
|------|---|---|
| [35] | <p>Their proposed methodology consists of two phases. The first phase uses KNN classifier to determining most relevant courses based on user's profile and this portion of proposed methodology will be executed for every recommendation. The second phase is optional and is executed only when enough ratings are available in the database for the relevant courses. In second phase, naïve Bayes classifier is used to recommend courses based on user's rating.</p> | <p>Even though one algorithm showed better precision for course recommendation but still the overall f-measure was not showing satisfactory results. Other hybrid algorithms like association rules mining, deep learning and collaborative filtering can be considered to improve results in future.</p> |
| [36] | <p>A hybrid recommender system was proposed by them that included content based collaborative filtering and sequence pattern mining to help in identifying a sequence in which courses should be studied. The proposed framework used both content and context aware course recommendations by utilizing the weblogs, LOs and user's profile. GSM algorithm was used for identifying the recommended sequence of courses.</p> | <p>Emerging tools of machine learning and artificial intelligence can be utilized in future to generate more effective recommendations [36].</p> |
| [38] | <p>Their proposed methodology utilized tag based Collaborative filtering. In their proposed method, the learners will write tags according to their requirements.</p> | <p>In future, some suggested tags can also be utilized keeping in view the experts' generated tags to ensure vast coverage of LOs. [38]</p> |
| [39] | <p>They proposed MoodleREC teacher centric system that utilized popular LO repositories to crawl LOs. These recommendations can be used to design a new course hence the proposed mode provides an outline that an instructor can utilize to design a new course.</p> | <p>There are no standard LOs and LOs may differ in different LORs. Semantic solutions can be used to solve this problem in future. [39]</p> |
| [40] | <p>They combined deep belief network with collaborative filtering for more relevant recommendations. RBM are used for constructing neural networks for unsupervised learning. The unsupervised</p> | --- |

| | | |
|------|--|---|
| | learning RMB model assigns initial values to the back-propagation model and hence eliminate the drawback of traditional method. | |
| [43] | They proposed a methodology that is based on RLNLS model. The RLNLS model utilized the learner's interest and capacity to predict their behaviour in future. Collaborative filtering technique was used to find behaviour of neighbour learners while Hawkes process was used to predict the learning intensity. | In future, feature reduction can be performed to extract data only from relevant variables to maximize efficiency of the algorithm [43]. |
| [46] | Their proposed framework combines collaborative filtering with rule-based filtering and KNN classification to create a hybrid system. The recommendation hence generated will be context sensitive. | Their proposed methodology does not consider the authenticity of the user posting course. It can be improved in future by analyzing the profile of instructor. |
| [47] | Their proposed EduRank method used collaborative filtering along with social voting to evaluate the relevance of a course. The recommended questions are ranked in descending order based on their Copeland score while item response theory and Bayesian knowledge tracing were used for calculating item difficulty. AP correlation matrix was used to evaluate the proposed method and perform comparison with competing methodologies. | Their framework only considered questions and their solution time for difficulty-based ranking while not adding any solution to identify or help weak students. Work can be done in future to help students in coping with their weaknesses while utilizing their strengths to maximize learning. |
| [48] | They used multi-modal matrix factorization for course recommendation in MOOCs platform. The proposed methodology calculates either a course provides particular skill set that a user is looking for or not. The most frequent items in the neighborhood are ranked the highest. | The precision results showed between 1% to 3% accuracy for sparse datasets [48]. The lack of available ratings is affecting the accuracy of their proposed model and it be overcome in future by using hybrid techniques. |
| [49] | They proposed a method that focuses on the older audience who are self-learning and do | --- |

| | | |
|------|--|--|
| | not have access to any recommendations from teacher or instructor. Content based filtering, collaborative filtering and web mining were used to identify appropriate learning material on the internet which was later narrowed down according to personal needs of the learners. Clustering is done using KNN to group similar learners in one cluster to maximize personalization. | |
| [50] | They proposed Course recommender model that utilized big data to generate association rules in distributed environment. They utilized Hadoop Distributed File System (HDFS) while using Spark for data processing for association rules mining and Sqoop to import and export dataset. Course recommender used Apriori algorithm to generate recommendations of most relevant courses. | For future work, they recommended using deep learning for better utilization of big data in generating course recommendations [50]. |
| [51] | The purpose of their study was to discover most relevant course by implementing association rules mining on big data. They used Spark ML libraries to implement FP-Growth algorithm. For course recommendations, they analyzed the behaviour of learners who interacted with system in past by utilizing course enrollment and log data. | In future, big data can be utilized to analyze real time data |
| [52] | They utilized the feedback of the users and improved the model by using the concept of regret to improve recommendations for the future. | In future work can be done to handle the dynamically increasing courses dataset. (Hou et al. 2018) |
| [53] | They proposed a framework called sharing notes that is a web based forum designed to facilitate students by providing them personalized material. Sharing notes utilized collaborative and content based filtering to | In future their own specific metrics can be used to perform analysis of social networks, and incorporate them in new recommendation approaches [53]. |

rate the material while fuzzy linguistics model was implemented to perform qualitative study on communication between different users. Ontologies were used to map the relationship between users and their preferred material. The final recommendations are generated after obtaining ratings from the trusted users. Sharing notes was tested with the help of 425 users amongst those 88% of the users shows satisfaction with their proposed recommender system.

| | | |
|------|--|--|
| [55] | Their proposed methodology added the layer of semantic analysis along with ratings to improve the performance of generated recommendations. The semantics matrix was separately constructed and was used by utilizing cosine similarity. | In future difficulty level can also be added to generated more relevant recommendations [55]. |
| [56] | Their proposed solution performed semantic analysis on messages posted on discussion forums. Text pre-processing was performed only on content logistic related posts and semantic similarity calculations were used to performed semantic analysis. SVM was utilized in the end to generate recommendations for relevant posts. | In future, other classification models like LSTM can be utilized to improve accuracy [56]. |
| [58] | The proposed methodology takes the data of online courses as input and rank them into various difficulty levels. Similarly, students create their profile and their data is also extracted through the log files that records the data of a student interacting with the system. | Their proposed model did not consider the rating matrix that can include information about how other users interact with the recommender system. |
| [59] | They collected data from the learners profile and used variable length genetic algorithm for learning path recommendations. | In future work can be done to calculate trust for different users and use this value in collaborative filtering [59]. |

Collaborative filtering was used to find the popularity matrix which was later utilized by genetic algorithm for personalized course sequencing.

| | | |
|------|---|--|
| [60] | Their solution considered both formatted data like courses and unformatted data like web sources, documents, papers and publications. They performed clustering on the statistics of data rather than on raw data hence improving the processing speed. | In future, deep learning algorithms can also be explored |
|------|---|--|

Table 9. Objective based clustering.

[illegible]

[illegible]

Table 10. Brief comparison of datasets.

| Refere nce | Dataset | Source/platform |
|---------------|--|--|
| [14] | Retrieved dataset from MOODLE server | MOODLE |
| [15] | Mooshak and ENKI | ENKI and Mooshak |
| [11] | data is collected from multiple online sources like emails, blog, reports, phone calls and social sites | Scrapping of publically available data |
| [16] | Dataset was created | UM Learn (University of Manitoba Learning Management System) Tools: a third-party open-source quiz tool called "Savsoft Quiz v3.0" |
| [22] | Harvard and MIT dataset | HarvardX-MITx Person-Course Academic Year 2013 De-Identified dataset |
| [23] | MOOCs | Coursera courses from Peking University (2 courses) |
| [25] | Own dataset | two courses taught by the authors |
| [26] | Own dataset | Own dataset |
| [27] | Junyi Academy -- http://www.junyiacademy.org/ a Chinese e-learning website which is established on open-source code released by Khan Academy (https://www.khanacademy.org/) in 2012 | Haw-Shiuan Chang, Hwai-Jung Hsu and Kuan-Ta Chen, "Modeling Exercise Relationships in E-Learning: A Unified Approach," International Conference on Educational Data mining (EDM), 2015 (Link: https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=1198) |
| [31] | Crawled data from different educational websites (sources not mentioned) | Own dataset |
| [34] | Shanghai Lifelong Learning Network, | Shanghai Lifelong Learning Network, |
| [36] | LMS of university (name of university not mentioned) | - |
| [38] | An educational dataset, consisting of 120 learners, high school students, participants of the Center for Young Talents project, supported by the Schneider Electric DMS NS company, Novi Sad | ---- |
| [39] | (2018). SOFIA, Teachers Formation Platform. http://www.istruzione.it/pdgg/ Last | Sofia |

| | | |
|------|--|--|
| [40] | Own dataset | starC MOOC platform of Central China Normal University |
| [43] | http://jclass.pte.sh.cn | Mic-video Platform of ECNU |
| [46] | BOL repository including records from 2017 to 2018. variables include students' personal information, learning activity scores, and perception and confidence scores toward each mandatory course | BINUS online learning (BOL) |
| [47] | Koedinger, K.R., Baker, R., Cunningham, K., Skogsholm, A., Leber, B., Stamper, J., 2010. A data repository for the edm community: the pslc datashop. Handbook of educational data mining. pp. 43–55 (https://pslcdatashop.web.cmu.edu/KDDCu p) | KDD cup 2010 and an unpublished data of schools name K12 dataset by authors |
| [48] | http://merlot.org | Multimedia Educational Resource for Learning and MACE project (Real world object based access to architecture learning material–the mace experience. In: Proceedings of the World Conference on Educational Multimedia and Hypermedia (EDMEDIA 2010), Toronto, Canada (June 2010) (2010)) and Online Teaching (MERLOT) |
| [49] | Own dataset | Own dataset |
| [50] | IBM dataset for Apriori algorithm, T10I4D100K (http://fimi.ua.ac.be/data/), T25I10D10K (http://www.philippe-fourrier.com/spmf/index.php?link=datasets.php) (generated by a random iteration of the database), Harvard and MIT published the edX learning data in 2012–2013, HMXPC13_DI_v2_5–14–14.csv (https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/26147) | IBM and edX |

| | | |
|------|---|---|
| [51] | Moodle. Open-source learning platform Moodle.org. https://moodle.org/ . | operational database of ESTenLigne platform that is built on LMS Moodle |
| [52] | Created their own dataset | Created their own dataset |
| [53] | http://sharingnotes.ujen.es/ | Sharing notes |
| [55] | Created their own dataset | Created their own dataset |
| [56] | MOOCs | MOOC |
| [58] | Source not mentioned | One public one private dataset (source not mentioned) |
| [59] | Created their own dataset | Created their own dataset |
| [60] | data were gathered from several local Universities (Shandong University of Finance and Economics and Shandong University) | Local universities dataset |

4. Comparative Analysis

The previous section discussed the existing literature in the recommender system for e-learning in detail. This section will focus on comparing the existing researches with one another to identify the strength of existing techniques.

4.1. Objective Based Clustering

We will go into the detail of objective based analysis of our literature review in this section. The previous sections indicated that there are a variety of objectives and directions that can be followed to enhance the effectiveness of teaching-learning experience on e-learning platforms. The main objectives of our overall literature are summarized in Table 9 which indicates that most of the present literature focus on personalization and accuracy. Some researchers use precision, recall and f-measure while other use Root Mean Square Error (RMSE) as well. Learning path recommendation is also a direction while some researchers considered execution time and memory utilization as a measure of effective recommender as well.

4.2. Dataset analysis

Since most of the research conducted for recommender systems in e-learning area uses wide variety of datasets. The detailed descriptions of various datasets along with links and references where applicable is listed in Table 10. Some researches like [55] constructed their own dataset while some researchers like [50] used public data set for MOOCs. While LMS datasets of universities is also utilized by some researches like [36] and [47]

4.3. Performance analysis

The objective based clustering has helped us in identification of common objectives. It is also important to compare the performance of these methodologies in order to identify the methodologies that

are outperforming others in the same domain. We can observe from Table 11 that most of the researches focused on measuring precision, recall, f-measure, RMSE as an accuracy measure. Some researches like [26] focused on execution time as well. The performance analysis table is constructed for comparable values and survey results may vary from scenario to scenario and cannot provide a uniform assessment hence they are excluded from the Table 5. The results show that methodology proposed by [14] shows best precision, recall and f-measure values.

Table 11. Performance analysis.

| Reference | Year | Accuracy | precision | Recall | f-measure | Execution time |
|-----------|------|-------------|-----------|--------|-----------|----------------|
| [14] | 2020 | | 0.98 | 0.90 | 0.93 | |
| [27] | 2018 | 0.45 | | | | |
| [35] | 2018 | | 0.75 | 0.50 | 0.60 | |
| [36] | 2018 | | 0.35 | 0.45 | | |
| [50] | 2018 | | | | | 60s |

4.4. Comparison of Best Scheme From Each Category Existing Schemes

The existing literature includes various schemes being used for different objectives. A brief comparison of some of the major existing schemes is summarized in Table 12 that shows that performance for most of the existing schemes were measured in terms of accuracy. Results for active user showed better accuracy as compared to non-active users [14] while full course recommendation showed less accuracy as compared to step by step recommendation in some methodologies [26]. In some cases performance stayed constant with increasing size of dataset [52] while in some cases is slightly dropped [14].

Table 12. Brief comparison of schemes.

| Reference | Methodology | Results | Limitations |
|-----------|--|--|--|
| [14] | Combined jena rules with collaborative filtering | Showed 0.97 precision for active user and 0.64 precision value for non-active users. | Their proposed methodology utilized manual method to update RDF factbase and this method can be atomized in future. [14] |
| [11] | Association rules mining combined with multiple agents | The accuracy of their proposed model initially increased when the number of lectures increased but start to decrease once number of lectures reached 40. Overall the accuracy value stayed above 90% | In future, Swarm intelligence and genetic algorithm can be utilized to improve results [11]. |

| | | | |
|------|---|--|--|
| [27] | Hybrid methodology including clustering and LSTM | Their proposed model showed up to 45% precision for full course recommendation while up to 75% precision for ten-step recommendation | In future, data streams of learners can be utilized to improve the generated learning paths. |
| [52] | Reinforcement learning | The proposed algorithm showed around 0.85 value for reward even for massive dataset. | In future work can be done to handle the dynamically increasing courses dataset. (Hou et al. 2018) |
| [56] | Neural networks embedding were to calculate semantic similarity along with SVM. | The accuracy was measured in terms of transferability where world history trained dataset when tested on python course dataset showed 62% accuracy while when model trained on python was tested on world history, it showed 65% accurate results. | In future, other classification models like LSTM can be utilized to improve accuracy [56]. |

5. Research Challenges and Gaps

The existing research in the area of educational recommendation specifically for online learning platforms is of various kinds. All types of recommender system work for a particular objective and area and every area has challenges of its own.

5.1. The Complex Relationship Between LOs

The graph-based methodologies are mostly used in ITS or within course recommendation like [25]. The graph-based methodologies that recommend different courses mostly face cold start problems like [18]. Some query-based methodologies like [27] utilized graph-based methodology for query optimization but did not consider the complex relationship between different LOs.

5.2. Cold Start Problem

The content based filtering was utilized in various researches like [34] which proposed two-phases of research. The phase that included only content based filtering showed less accurate results as compared to the one that combines collaborative filtering with content based filtering [38] and [31]. Some researchers utilized association rules mining [36] while others used sequence pattern mining [35]. The limitations in existing research vary for different methodologies. The content based filtering excludes the item and user based ratings hence ignoring the rankings generated from old users. The collaborative filtering technique has a cold start problem.

5.3. Increased Processing Time

Association rules mining and sequence pattern mining can significantly slow down the system. Recommendations using big data is also gaining popularity as the volume of online learning platform data is increasing. The existing researches needs enhancements to handle ever increasing amounts of data [52] and real-time data [51]. Both of these tasks can be challenging due to the complexity of currently used methodologies. Light weight model can be utilized to reduce processing time in future.

5.4. Limited Availability of Public Data

Query-based recommendations is another emerging area that has been explored by [49] and [11]. The future enhancement can be achieved in this area by utilizing the profile of both the learner and the instructor who is posting the course while generating recommendations. This direction can face challenges due to limited availability of public data.

5.5. Sematic Study

Forum-based discussion presents some emerging trends in research that can be utilized in semantic study to recommend courses in future [53]. Online learning platforms have a discussion forum that can be effective in analyzing the quality of the course [55]. Semantic study is an important area that can be explored using deep learning [56].

6. Optimal solution

The previous section described various problems like cold-start problems and un-availability of datasets. Since the research in this area is scattered and vary from solution to solution hence it won't be possible to design one optimal solution that works for all kinds of problems, for example recommendations generated by ITS utilizes contents within a course that cannot be used by a generic course recommender system. That is why we shall look for multiple optimal solution instead of one that works for all scenarios.

In the existing literature there are some researchers used graph based methodologies to capture knowledge base [31]. Deep learning algorithms can be utilized in future to capture more complex relationships between LOs as they are being used in other areas of text analysis [61].

The existing methodologies also has cold-start problem in some papers [34] which has been overcome by combining content-based filtering with collaborative filtering. Using LSTM models along with collaborative and content based filtering can help in solving cold-start problems. [27]

Existing researches mostly construct their own datasets due to little availability of public datasets. Datasets for future researches can be constructed by scraping the online resources. MOOCs and online question answers forums like stack-overflow are good sources for scraping data. Video content can also be utilized to construct dataset by utilizing subtitles [23]. Increase in the size of dataset will further require optimized solutions for complex problems that can be resolved by utilizing deep reinforcement learning [52].

Research in the area of semantic analysis is scarce for recommender system. Hence further work can be done to improve existing methodologies by combining sematic study with deep learning algorithms. Forum discussion are rich source of such datasets that can be explored in future. Deep learning algorithms can help in finding context in semantic analysis of such dataset. In conducting a comprehensive survey on improved E-learning based recommender systems, our research builds upon foundational insights presented in [62–75]

7. Proposed Solution

The proposed solution for an optimized recommender system cannot be universal as the nature of problem can vary. Here we considered a solution for MOOCs where we get diverse audience from all around the world searching for different material. In this scenario, same keyword may match the title of

various courses hence evaluating description of course is important. Since anyone can post material in MOOCs platform so evaluating the quality of the content in this scenario can be another hurdle. The size of catalog for online courses is increasing and is expected to increase in future hence making the task of generating optimized recommendation more challenging.

The proposed solution summarized in Figure 4 tries to tackle all of the above mentioned challenges by utilizing course rating, comments and descriptions for sentiment analysis, content-based filtering, collaborative filtering and association rules mining. Use of deep learning will improve the results of recommendations while association rules mining can provide a more in depth insight about the relationship between keywords and selection of courses.

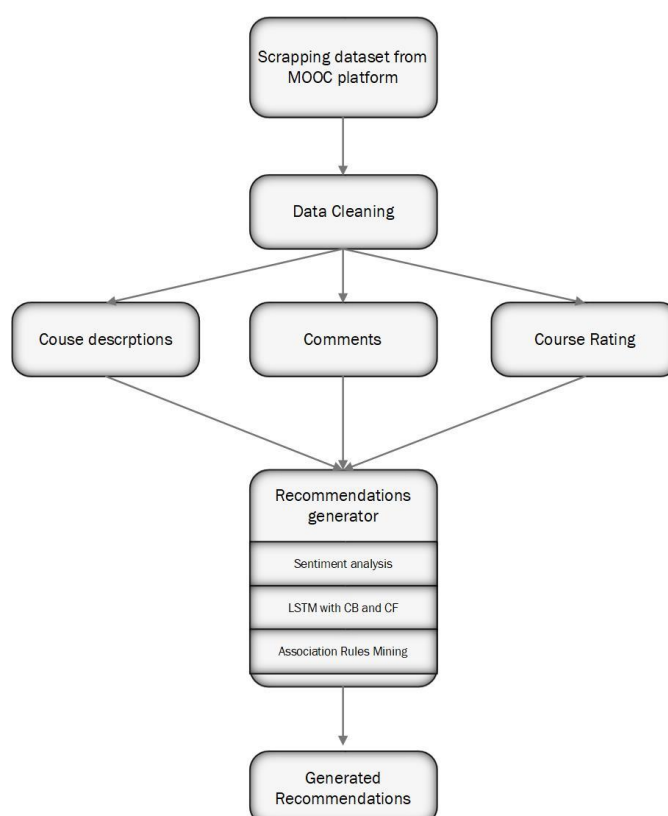


Figure 4. Proposed methodology.

8. Conclusion

Recommender systems are gaining popularity in various fields like e-commerce and entertainment. Its use in education is also getting popular with the increasing popularity of online learning platforms. The existing literature in educational recommender system include areas like graph-based methodologies, ITS, query optimization, content based and collaborative filtering, big data and association rules mining. This survey aimed to explore emerging directions of recommender systems in education. This study analyzed existing literature and performed objective-based analysis. Performance analysis was also considered where the values were comparable. Limitations of existing researches were also identified and studied to identify research gaps for future directions.

In the future, the directions that are appearing to be gaining popularity are big data analysis and hybrid recommendation systems. There is an increasing trend of shifting to online learning mode instead of traditional classroom which has paced up due the COVID-19 pandemic. The increasing demand of online learning platforms has also increased the demand of MOOCs and a lot more people are turning towards MOOCs as an alternate source of learning [10]. This as a result, has increased the need of effective recommender systems. More diverse group of people are opting to learn online which demands more personalized recommendations that considers the requirements of particular region along with other

factors. Instructors' profiles on online learning platforms can also be considered to ensure their authenticity. Semantic analysis on online discussion forums of e-learning platforms can also be explored in future to generate more effective recommendations.

References

- Goldberg D, Nichols D, Oki BM, Terry D (1992) Using collaborative filtering to weave an information tapestry. *Commun ACM* 35:61–70
- Schafer B, Konstan J, Riedl J (1999) Recommender Systems in E-Commerce. 1st ACM Conference on Electronic Commerce, Denver, Colorado, United States. <https://doi.org/10.1145/336992.337035>
- Kim H-T, Kim E, Lee J-H, Ahn CW (2010) A recommender system based on genetic algorithm for music data. In: 2010 2nd International Conference on Computer Engineering and Technology. pp V6-414-V6-417
- Azaria A, Hassidim A, Kraus S, Eshkol A, Weintraub O, Netanel I (2013) Movie recommender system for profit maximization. In: Proceedings of the 7th ACM conference on Recommender systems. Association for Computing Machinery, New York, NY, USA, pp 121–128
- Gomez-Urbe CA, Hunt N (2016) The Netflix Recommender System: Algorithms, Business Value, and Innovation. *ACM Trans Manage Inf Syst* 6:13:1-13:19
- Ragab AHM, Mashat AFS, Khedra AM (2012) HRSPCA: Hybrid recommender system for predicting college admission. In: 2012 12th International Conference on Intelligent Systems Design and Applications (ISDA). pp 107–113
- Natividad MCB, Gerardo BD, Medina RP (2019) A fuzzy-based career recommender system for senior high school students in K to 12 education. In: IOP Conference Series: Materials Science and Engineering. IOP Publishing, p 012025
- Estrela D, Batista S, Martinho D, Marreiros G (2017) A Recommendation System for Online Courses. In: Advances in Information Systems and Technologies. pp 195–204
- UNICEF (2021) COVID-19 and School Closures: One year of education disruption. In: UNICEF DATA. <https://data.unicef.org/resources/one-year-of-covid-19-and-school-closures/>. Accessed 16 Sep 2021
- Impey C, Formanek M (2021) MOOCS and 100 Days of COVID: Enrollment surges in massive open online astronomy classes during the coronavirus pandemic. *Social Sciences & Humanities Open* 4:100177
- Rafiq MS, Jianshe X, Arif M, Barra P (2021) Intelligent query optimization and course recommendation during online lectures in E-learning system. *J Ambient Intell Human Comput*. <https://doi.org/10.1007/s12652-020-02834-x>
- Gupta S, Sabitha AS (2019) Deciphering the attributes of student retention in massive open online courses using data mining techniques. *Educ Inf Technol* 24:1973–1994
- Gamage D, Fernando S, Perera I (2015) Factors leading to an effective MOOC from participants perspective. In: 2015 8th International Conference on Ubi-Media Computing (UMEDIA). pp 230–235
- Rawat B, Samriya JK, Pandey N, Wariyal SC (2020) Enriching 'user item rating matrix' with resource description framework for improving the accuracy of recommendation in E-learning environment. *Materials Today: Proceedings*. <https://doi.org/10.1016/j.matpr.2020.09.701>
- Nabizadeh AH, Gonçalves D, Gama S, Jorge J, Rafsanjani HN (2020) Adaptive learning path recommender approach using auxiliary learning objects. *Computers & Education* 147:103777
- Niknam M, Thulasiraman P (2020) LPR: A bio-inspired intelligent learning path recommendation system based on meaningful learning theory. *Education and Information Technologies* 25:3797–3819
- Mu R (2018) A survey of recommender systems based on deep learning. *Ieee Access* 6:69009–69022
- Nabizadeh AH, Leal JP, Rafsanjani HN, Shah RR (2020) Learning path personalization and recommendation methods: A survey of the state-of-the-art. *Expert Systems with Applications* 159:113596
- Peška L, Tashu TM, Horváth T (2019) Swarm intelligence techniques in recommender systems - A review of recent research. *Swarm and Evolutionary Computation* 48:201–219
- Nikzad-Khasmakhi N, Balafar MA, Reza Feizi-Derakhshi M (2019) The state-of-the-art in expert recommendation systems. *Engineering Applications of Artificial Intelligence* 82:126–147
- George G, Lal AM (2019) Review of ontology-based recommender systems in e-learning. *Computers & Education* 142:103642
- Obeidat R, Duwairi R, Al-Aiad A (2019) A collaborative recommendation system for online courses recommendations. In: 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML). IEEE, pp 49–54
- Zhang M, Zhu J, Wang Z, Chen Y (2019) Providing personalized learning guidance in MOOCs by multi-source data analysis. *World Wide Web* 22:1189–1219

24. Corbett AT, Anderson JR (1994) Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction* 4:253–278
25. Wan S, Niu Z (2018) An e-learning recommendation approach based on the self-organization of learning resource. *Knowledge-Based Systems* 160:71–87
26. Muangprathub J, Boonjing V, Chamnongthai K (2020) Learning recommendation with formal concept analysis for intelligent tutoring system. *Heliyon* 6:e05227
27. Zhou Y, Huang C, Hu Q, Zhu J, Tang Y (2018) Personalized learning full-path recommendation model based on LSTM neural networks. *Information Sciences* 444:135–152
28. Hochreiter S, Schmidhuber J (1997) Long Short-Term Memory. *Neural Computation* 9:1735–1780
29. Dorigo M, Maniezzo V, Colomi A (1996) Ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 26:29–41
30. Well-Separated Clusters and Optimal Fuzzy Partitions: *Journal of Cybernetics*: Vol 4, No 1. <https://www.tandfonline.com/doi/abs/10.1080/01969727408546059>. Accessed 18 Feb 2023
31. Shi D, Wang T, Xing H, Xu H (2020) A learning path recommendation model based on a multidimensional knowledge graph framework for e-learning. *Knowledge-Based Systems* 195:105618
32. Balabanović M, Shoham Y (1997) Fab: content-based, collaborative recommendation. *Commun ACM* 40:66–72
33. Resnick P, Iacovou N, Suchak M, Bergstrom P, Riedl J (1994) GroupLens: an open architecture for collaborative filtering of netnews. In: *Proceedings of the 1994 ACM conference on Computer supported cooperative work*. Association for Computing Machinery, New York, NY, USA, pp 175–186
34. Xiao J, Wang M, Jiang B, Li J (2018) A personalized recommendation system with combinational algorithm for online learning. *Journal of Ambient Intelligence and Humanized Computing* 9:667–677
35. Neamah AA, El-Ameer AS (2018) Design and Evaluation of a Course Recommender System Using Content-Based Approach. In: *2018 International Conference on Advanced Science and Engineering (ICOASE)*. IEEE, pp 1–6
36. Tarus JK, Niu Z, Kalui D (2018) A hybrid recommender system for e-learning based on context awareness and sequential pattern mining. *Soft Computing* 22:2449–2461
37. Agrawal R, Imieliński T, Swami A (1993) Mining association rules between sets of items in large databases. In: *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*. Association for Computing Machinery, New York, NY, USA, pp 207–216
38. Klačnja-Milićević A, Ivanović M, Vesin B, Budimac Z (2018) Enhancing e-learning systems with personalized recommendation based on collaborative tagging techniques. *Appl Intell* 48:1519–1535
39. De Medio C, Limongelli C, Sciarrone F, Temperini M (2020) MoodleREC: A recommendation system for creating courses using the moodle e-learning platform. *Computers in Human Behavior* 104:106168
40. Zhang H, Huang T, Lv Z, Liu S, Yang H (2019) MOOCRC: A highly accurate resource recommendation model for use in MOOC environments. *Mobile Networks and Applications* 24:34–46
41. Hinton GE, Osindero S, Teh Y-W (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18:1527–1554
42. Rumelhart DE, McClelland JL (1987) Information Processing in Dynamical Systems: Foundations of Harmony Theory. In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*. MIT Press, pp 194–281
43. Pang Y, Liao C, Tan W, Wu Y, Zhou C (2018) Recommendation for MOOC with learner neighbors and learning series. In: *International Conference on Web Information Systems Engineering*. Springer, pp 379–394
44. Agrawal R, Srikant R (1994) Fast Algorithms for Mining Association Rules. pp 487–499
45. Kaufman L, Rousseeuw P (1990) *Finding Groups in Data: An Introduction To Cluster Analysis*. Wiley, New York ISBN 0-471-87876-6. <https://doi.org/10.2307/2532178>
46. Murad DF, Heryadi Y, Isa SM, Budiharto W (2020) Personalization of study material based on predicted final grades using multi-criteria user-collaborative filtering recommender system. *Educ Inf Technol* 25:5655–5668
47. Segal A, Gal K, Shani G, Shapira B (2019) A difficulty ranking approach to personalization in E-learning. *International Journal of Human-Computer Studies* 130:261–272
48. Symeonidis P, Malakoudis D (2019) Multi-modal matrix factorization with side information for recommending massive open online courses. *Expert Systems with Applications* 118:261–271
49. Lalitha TB, Sreeja PS (2020) Personalised Self-Directed Learning Recommendation System. *Procedia Computer Science* 171:583–592
50. Zhang H, Huang T, Lv Z, Liu S, Zhou Z (2018) MCRS: A course recommendation system for MOOCs. *Multimedia Tools and Applications* 77:7051–7069
51. Dahdouh K, Dakkak A, Oughdir L, Ibriz A (2019) Large-scale e-learning recommender system based on Spark and Hadoop. *Journal of Big Data* 6:1–23

52. Hou Y, Zhou P, Xu J, Wu DO (2018) Course recommendation of MOOC with big data support: A contextual online learning approach. In: IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, pp 106–111
53. Porcel C, Ching-López A, Lefranc G, Loia V, Herrera-Viedma E (2018) Sharing notes: An academic social network based on a personalized fuzzy linguistic recommender system. *Engineering Applications of Artificial Intelligence* 75:1–10
54. Zadeh LA (1974) The Concept of a Linguistic Variable and its Application to Approximate Reasoning. In: Fu KS, Tou JT (eds) *Learning Systems and Intelligent Robots*. Springer US, Boston, MA, pp 1–10
55. Albatayneh NA, Ghauth KI, Chua F-F (2018) Utilizing learners' negative ratings in semantic content-based recommender system for e-learning forum. *Journal of Educational Technology & Society* 21:112–125
56. Ntourmas A, Daskalaki S, Dimitriadis Y, Avouris N (2021) Classifying MOOC forum posts using corpora semantic similarities: a study on transferability across different courses. *Neural Comput & Applic.* <https://doi.org/10.1007/s00521-021-05750-z>
57. Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. In: *Proceedings of the fifth annual workshop on Computational learning theory*. Association for Computing Machinery, New York, NY, USA, pp 144–152
58. G S, M P, Dubey P, Drolia AR, S S (2020) Subjective Areas of Improvement: A Personalized Recommendation. *Procedia Computer Science* 172:235–239
59. Dwivedi P, Kant V, Bharadwaj KK (2018) Learning path recommendation based on modified variable length genetic algorithm. *Education and information technologies* 23:819–836
60. Lin J, Li Y, Lian J (2020) A novel recommendation system via L0-regularized convex optimization. *Neural Computing and Applications* 32:1649–1663
61. Kim T, Yun Y, Kim N (2021) Deep Learning-Based Knowledge Graph Generation for COVID-19. *Sustainability* 13:2276
62. Shahid, H., Ashraf, H., Javed, H., Humayun, M., Jhanjhi, N. Z., & AlZain, M. A. (2021). Energy optimised security against wormhole attack in iot-based wireless sensor networks. *Comput. Mater. Contin.*, 68(2), 1967-81.
63. Wassan, S., Chen, X., Shen, T., Waqar, M., & Jhanjhi, N. Z. (2021). Amazon product sentiment analysis using machine learning techniques. *Revista Argentina de Clínica Psicológica*, 30(1), 695.
64. Almusaylim, Z. A., Zaman, N., & Jung, L. T. (2018, August). Proposing a data privacy aware protocol for roadside accident video reporting service using 5G in Vehicular Cloud Networks Environment. In *2018 4th International conference on computer and information sciences (ICCOINS)* (pp. 1-5). IEEE.
65. Ghosh, G., Verma, S., Jhanjhi, N. Z., & Talib, M. N. (2020, December). Secure surveillance system using chaotic image encryption technique. In *IOP conference series: materials science and engineering* (Vol. 993, No. 1, p. 012062). IOP Publishing.
66. Humayun, M., Alsaqer, M. S., & Jhanjhi, N. (2022). Energy optimization for smart cities using iot. *Applied Artificial Intelligence*, 36(1), 2037255.
67. Kaur, R., Verma, S., Jhanjhi, N. Z., & Talib, M. N. (2021, August). A comprehensive survey on load and resources management techniques in the homogeneous and heterogeneous cloud environment. In *Journal of Physics: Conference Series* (Vol. 1979, No. 1, p. 012036). IOP Publishing.
68. Alotaibi, A. F. (2021). A comprehensive survey on security threats and countermeasures of cloud computing environment. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(9), 1978-1990.
69. Bashir, I. R. A. M., Hamid, B. U. S. H. R. A., Jhanjhi, N. Z., & Humayun, M. A. M. O. O. N. A. (2020). Systematic literature review and empirical study for success factors: client and vendor perspective. *J Eng Sci Technol*, 15(4), 2781-2808.
70. Ponnusamy, V. (Ed.). (2016). *Biologically-Inspired Energy Harvesting through Wireless Sensor Technologies*. IGI Global.
71. Ponnusamy, V., Jung, L. T., Ramachandran, T., & Zaman, N. (2017, April). Bio-inspired energy scavenging in wireless ad hoc network. In *2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT)* (pp. 1-5). IEEE.
72. Abbas, S. F., Shahzad, R. K., Humayun, M., Jhanjhi, N. Z., & Alamri, M. (2019). SOA Issues and their Solutions through Knowledge Based Techniques—A Review. *Int. J. Comput. Sci. Netw. Secur.*, 19(1), 8-21.
73. Humayun, M., Jhanjhi, N. Z., Talib, M. N., Shah, M. H., & Suseendran, G. (2021). Cybersecurity for Data Science: Issues, Opportunities, and Challenges. *Intelligent Computing and Innovation on Data Science: Proceedings of ICTIDS 2021*, 435-444.
74. Kumar, P., Kumar, R., Aljuhani, A., Javeed, D., Jolfaei, A., & Islam, A. N. (2023). Digital twin-driven SDN for smart grid: A deep learning integrated blockchain for cybersecurity. *Solar Energy*, 263, 111921.

75. 75 Mukherjee, D., Ghosh, S., Pal, S., Akila, D., Jhanjhi, N. Z., Masud, M., & AlZain, M. A. (2022). Optimized Energy Efficient Strategy for Data Reduction Between Edge Devices in Cloud-IoT. *Computers, Materials & Continua*, 72(1).
76. 76 E. Ndashimye, N. I. Sarkar, and S. K. Ray, "A Multi-criteria based handover algorithm for vehicle-to-infrastructure communications," *Computer Networks*, vol. 185, no. 202152, Article ID 107652, 2020
77. 77 Ray, S. K., Pawlikowski, K., & Sirisena, H. (2009). A fast MAC-layer handover for an IEEE 802.16 e-based WMAN. In *AccessNets: Third International Conference on Access Networks*, AccessNets 2008, Las Vegas, NV, USA, October 15-17, 2008. Revised Papers 3 (pp. 102-117). Springer Berlin Heidelberg.
78. 78 Srivastava, R. K., Ray, S., Sanyal, S., & Sengupta, P. (2011). Mineralogical control on rheological inversion of a suite of deformed mafic dykes from parts of the Chottanagpur Granite Gneiss Complex of eastern India. *Dyke Swarms: Keys for Geodynamic Interpretation: Keys for Geodynamic Interpretation*, 263-276.
79. 79 Ray, S. K., Sinha, R., & Ray, S. K. (2015, June). A smartphone-based post-disaster management mechanism using WiFi tethering. In *2015 IEEE 10th conference on industrial electronics and applications (ICIEA)* (pp. 966-971). IEEE.
80. 80 Chaudhuri A, Ray S (2015) Antiproliferative activity of phytochemicals present in aerial parts aqueous extract of *Ampelocissus latifolia* (Roxb.) planch. on apical meristem cells. *Int J Pharm Bio Sci* 6(2):99-108
81. 81 Hossain, A., Ray, S. K., & Sinha, R. (2016, December). A smartphone-assisted post-disaster victim localization method. In *2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)* (pp. 1173-1179). IEEE.
82. 82 Airehrour, D., Gutierrez, J., & Ray, S. K. (2018). A trust-based defence scheme for mitigating blackhole and selective forwarding attacks in the RPL routing protocol. *Journal of Telecommunications and the Digital Economy*, 6(1), 41-49.
83. 83 Ray, S. K., Ray, S. K., Pawlikowski, K., McInnes, A., & Sirisena, H. (2010, April). Self-tracking mobile station controls its fast handover in mobile WiMAX. In *2010 IEEE Wireless Communication and Networking Conference* (pp. 1-6). IEEE.
84. 84 Dey, K., Ray, S., Bhattacharyya, P. K., Gangopadhyay, A., Bhasin, K. K., & Verma, R. D. (1985). Salicyladehyde 4-methoxybenzoylhydrazone and diacetylbis (4-methoxybenzoylhydrazone) as ligands for tin, lead and zirconium. *J. Indian Chem. Soc.:(India)*, 62(11).
85. 85 Airehrour, D., Gutierrez, J., & Ray, S. K. (2017, November). A testbed implementation of a trust-aware RPL routing protocol. In *2017 27th International Telecommunication Networks and Applications Conference (ITNAC)* (pp. 1-6). IEEE.
86. 86 Ndashimye, E., Sarkar, N. I., & Ray, S. K. (2016, August). A novel network selection mechanism for vehicle-to-infrastructure communication. In *2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)* (pp. 483-488). IEEE.
87. 87 Ndashimye, E., Sarkar, N. I., & Ray, S. K. (2020). A network selection method for handover in vehicle-to-infrastructure communications in multi-tier networks. *Wireless Networks*, 26, 387-401.
88. 88 Hussain, S. J., Ahmed, U., Liaquat, H., Mir, S., Jhanjhi, N. Z., & Humayun, M. (2019, April). IMIAD: intelligent malware identification for android platform. In *2019 International Conference on Computer and Information Sciences (ICCIS)* (pp. 1-6). IEEE.
89. 89 Shahid,H.,Ashraf,H.,Ullah,A.,Band,S.S.&Elnaffar,S.Wormholeattackmitigationstrategiesandtheirimpact onwireless sensor network performance: A literaturesurvey. *International Journal of Communication Systems* 35, e5311(2022). URLhttps://onlinelibrary.wiley.com/doi/abs/10.1002/dac.5311. https://onlinelibrary.wiley.com/doi/pdf/10.1002/dac.5311.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.