International Journal of Research in Advent Technology, Special Issue, NCRCEST 2019
E-ISSN: 2321-9637
Available online at www.ijrat.org

An Implementation of E-learning Recommendation System For Tracing Learner Behavior

Nidhi Joshi[1], Rajendra Gupta[2]

[1,2]*Department of Computer Science, Rabindranath Tagore University, Raisen, Madhya Pradesh, India*
nidhi11joshi@gmail.com, rajendragupta1@yahoo.com

Abstract: *:* A Huge Amount of E-Learning materials are now available in numerous forms, distributed over the network. In the online learning scenario, it is very difficult for users to choose suitable resources for their learning, without sufficient background knowledge. Mostly recommend system have been tried in E-Commerce to increase purchasing of goods, but rarely tried in E-Learning. The proposed work is intended to design a recommendation system. This paper suggests the use of web content mining techniques, web page pattern analysis with KNN and genetic search algorithms to build such an agent that could improve course material navigation as well as online learning process. The proposed recommendation system is developed for the e-learning systems to suggest the contents to the end users. There is specific learning style with each student or user differ from each other, so this work focuses on tracing learner's behavior. It accordingly suggests the appropriate learning material. The complete design is promising to suggest the user context based web content recommendations.
Keywords: Recommendation system, e-learning, E-commerce, KNN algorithm, genetic algorithm;

1. INTRODUCTION

The recommendation systems are basically a software technique by which different user behavioral attributes can be used for offering the most relevant to products and to the end user.In this context the user navigational behavior, interest and other factors can be involved for finding more accurate predictions [1]. In the presented work the recommendation system is the key area of investigation and system design.The proposed recommendation system is employed on the e-learning system,where the offered recommendation system deals with the web page, database user search and navigation history. The system recommends the accurate data for reading and learning on the basis of user context.

Now a days the e-learning systems have become more common among the students. The e-learning systems contains a significant amount of knowledge and learning data [2].Some of the contents are much similar to each other,so finding relevant content according to the user context is a complex task [3]. Therefore proposed work is a new model for designing and implementation.It evaluates the database content and user requirements. It is based on the factors, most likely contents which are recommended.

2. SYSTEM ARCHITECTURE

The main objective of the proposed work is to find an efficient and accurate technique for recommendation system.The technique works on the following key objectives [4].

1. **User behavior model applying the user's historical navigational pattern analysis:** the objective is defined for understanding the user requirements and the historical behavior of patterns.
2. **Utilizing the user interest, behavior and current context to optimize the recommendation needs:** the objectives of present work is to optimize the generated initial recommendation needs that fit in user context and requirements.
3. **Evaluation of proposed technique with respect to existing methodology of recommendation system design:** during literature, collection a number of techniques are evaluated. Among them, a promising model is selected for comparative performance study of the system over different performance factors.

In order to accomplish the given objectives the following system is proposed for design and implementation.The figure 1 shows the involved infrastructure and the actors of the system. The model contains the following components:

1. Learner or end user
2. Administrator
3. E-learning server (infrastructure)

A. Learner Or End User

The learner or end user is the entity, who gives us to use the developed system. That is the target user, for whom the entire system is crafted. The end user is responsible to create their accounts and find the contents from the learning site by browsing the topics or using the retrieval engine (using search system).

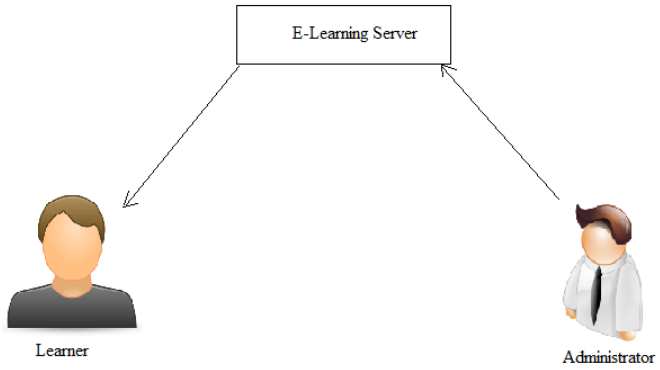

Figure 1 Basic overview of system

B. Administrator

The administrator is the entity who is responsible for managing the server resources and the e-learning web portal. Therefore an administrator is responsible for uploading contents, managing topics and their sub-topics with contents.

C. E-Learning Server

The e-learning server is the infrastructure server which is used to deploy the e-learning portal. In addition of that contains the algorithms and techniques which are used for analysis of data, their organization and also for retrieval for user query relevant records from the digital documents.

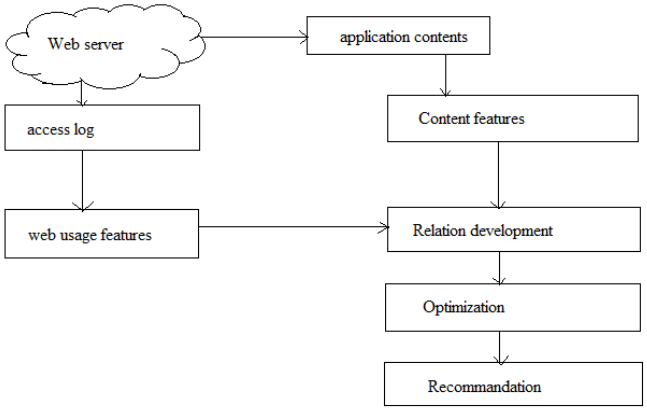

Figure 2 Server architecture

The proposed system architecture is demonstrated in figure 2. The given technique incorporate the different sources of data and user information for analyzing and processing the user request. Besides provides recommendations according to the user requirements. The following components are involved in this system.

Web server:A web server is used to host the applications. The web servers content the ability to process code blocks and return the web pages. The generated web pages by servers using the code blocks are accessible, using the HTTP or HTTPS protocols using the web browsers [5].

Access log: The web server is enable to report every event done in the server. Web server prepares different kinds of logs i.e. error log, access log, information log and others. The access log is also known as the web access server log or web usages log. Every event performed by the end user is written in these log files at the web servers i.e. resource request and server response. Here the server maintains the time stamp, request IP address, URL, protocols, and other server information [6]. A simple example of web access log example is given in table 1.

Example.com - - [06/Oct/2019:00:00:09 -0700] "GET / HTTP/1.1" 200 5258 http://www.example.com/academic_programs.htm "Mozilla/4.0 (compatible; MSIE 4.01; Windows 95)" example.com - - [06/Oct/2019:00:00:11 -0700] "GET /images/bullets/bsqs.gif HTTP/1.1" 200 489 http://www.example.com/ "Mozilla/4.0 (compatible; MSIE 4.01; Windows 95)"

Table 1 Access log example

Web usage features: Table 1 shows the attributes used in web access log file. All the features in access log file are not used for designing the recommendation system. Some of the features among them are need to be selected for utilizing in the proposed recommendation system. Therefore the following information is filtered from the entire log file.

1. User IP address
2. Time stamp
3. Requested URL
4. Method

Except these all the data from the web access log is remains as it is.

Application contents: The e-learning system contains a number of different topics and their contents. All the contents of this application are available in web document (web page) formats. When the retrieval and processing of user request for content is made .If search required topics, contents are fetched as search results.

Content features: The entire process handled in this modeling is provided for finding relevant content as well as recommending the contents which much relevant to current user's context. In this context to reduce the data processing efforts the content features from all the e-learning documents are extracted and organized with the target files. To select the features from the target document (web pages) the word frequency is used [7].The word frequency of the target word in a document is computed using the following formula:

$$Word\ Frequency = \frac{target\ word\ count}{total\ words\ in\ document} \ldots\ldots (1)$$

The computed word frequency is used to select the words from given document as feature set, therefore the target web document and their features are used in the following manner, as given in table 2.

Web pages	Features extracted
Datamining.jsp	Data, database, attributes ...
Computeranimation.jsp	Images, graphics,dimension..,

Table 2 Web content feature

Relation development: In this phase the extracted content features and the resource URL is used to find the keywords which are relevant to each other. The URLs which contains the keywords available in feature set is used for finding the related URLs and contents. The following algorithm is used for processing the data keywords and the target URLs.

Input: List of web page URLs WU_n, List of all extracted features F_m
Output: related links R_l

Process:

1. $for(i = 1; i \leq n; i++)$
   a. $temp = WU_i$
   b. $for(j = 1; j \leq m; j++)$
      i. $if\left(temp.contains(F_j)\right)$
         1. $R.add(WU_i, F_j)$
      ii. $end\ if$
   c. End for

2. Return R_l

Table 3 Relationship algorithm

The table 3 provides a list of URLs with the relevant keywords which indicates the keywords can be found in the given URL.

Moreover it is required to apply the search on the given system.Therefore we need to implement the search system in existing system architecture the extended architecture of the existing system is given in figure 3. The extended system architecture contains some additional components for finding accurate recommendations for the e-learning system. Therefore the entire data model is defined in two major modules:

1. **Search Results:**Relation development among the extracted keywords from the e-learning web pages and the target URLs are established using the algorithm given in table 3. The outcome of this process is a data structure which contains the following attributes as demonstrated in table 4.

URLs	Respected Keywords
/e_learning/3d.jsp	Computer animation
/e_learning/computer_graphics.jsp	Computer animation
/e_learning/image_formation.jsp	Computer animation
/e_learning/object_detection.jsp	Computer animation

Table 4 Relationship features

Query Keywords: The keywords which are used by the end user for finding the contents from the e-learning system,so the query contains a phrase or a set of words which is searched on the existing web page database.

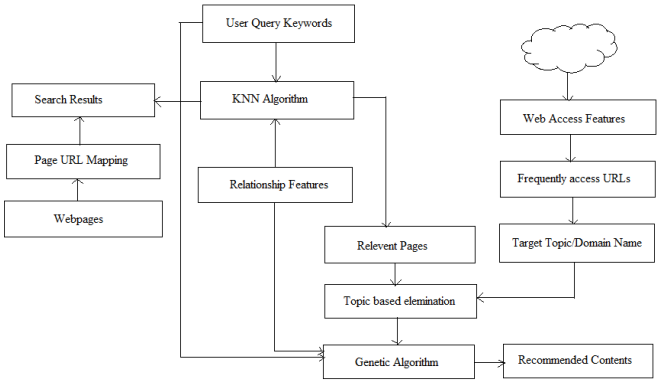

Figure 3 Extended system architecture

KNN (k-nearest neighbor) algorithm: The KNN algorithm is a distance based classification approach. This technique finds the differencebetween two given sequence of data i.e. one of them is query sequence and second is database sequence.By using distance computation it distinguish how many features are different from each other. If maximum sequence attributes are similar then the distance is calculated for minimum distance otherwise it obtain as higher [8]. For computing the distance among two sequences the following formula can be used.

$$D(x,y) = \sqrt{\sum_{i=1}^{n}(x_i^2 - y_i^2)}$$

Where $D(x,y)$ is the distance measured
$x_i$andy_i are the query sequence and database sequence
Hence the following process is used for finding the user input query sequence.

Input : relationship features R_l, user query keywords Q Output : relevant content page RP
Process: 1. $[PageURL, Keywords] = readRelationship(R_l)$ 2. $for(i = 1; i \leq keywords.length; i++)$ a. $D(Q, Keywords) = \sqrt{\sum_{i=1}^{M}(Query_i^2 - keywords_i^2)}$ b. $if(D \leq 0.25)$ i. $RP.Add(PageURL)$ c. End if 3. End for 4. Return RP

Table 5 KNN based search

This process returns the query relevant web pages from the e-learning application. Also enable navigation with the results the mapping between web page URL and search results data is performed. After that the search technique works according to the user input keywords.

On the other hand it is needed to suggest or recommend the web pages or web content for the user context. Therefore user web page recommendation is also employed with this methodology.

Recommendations: In order to recommend the web pages for the end user again the relationship features are used. The web pages which are obtained by the KNN algorithm are combined in this stage. Also the filtration on the data is performed. To filter the data the following processes are included with the system.

Web access log features: The web access log is previously introduced and the required features from the web log data are extracted. These web log features are used here for finding the user behavior and domain of interest. Therefore the frequency of the different web pages which are accessed by current user is evaluated. To compute the frequency of the user accessed web pages the following formula is used.

$$web\ page\ access\ frequency = \frac{C_A}{C_p}$$

$where\ C_A$is the count of a web page from a domain
$C_p$is the total number of web pages accessed

The top visited subjects/ domain / topics are identified from the web page frequency of visits. Using the identified topic names, the web pages of relationship features and the KNN search process are eliminated and remaining web pages are used in further for finding most suitable recommendations.

Optimization: After the elimination of web pages from the combined set of search results and relationship features the optimization process is applied. To optimize the recommendations the user query and the combined data is usedwith the genetic search process. The genetic

algorithm is a naturally inspired algorithm. It is works on the following four steps [9]:

1. **Selection:**the genetic algorithm is a works with the population. Here the term population is used for the initial solutions available for search. In our context the combined set of web pages which are filtered using subjects criteria is used as the initial population. The algorithm selects N/2 solution among them randomly for evaluation with the next steps.
2. **Crossover:** after random selection of the available solutions the cross over operation is performed. The crossover helps to design new set of solutions from the selected solutions. The newly generated solutions are evaluated in the next phase.
3. **Mutation:** the crossover based generated solutions are used with the mutation operator. Using these solutions are repaired and new solutions are tested for finding the fit solutions. If the solutions are repairable then it is repaired using the insertion, deletion, replacing the attributes otherwise the solutions are removed for next phase analysis.
4. **Update:**In this phase the fitness of all the solutions are estimated using the fitness function. The higher fitness values based solutions are preserved and remaining solutions are removed. The obtained fit solutions are updated as the population for next generation evaluation. Here the distance function is used as the fitness function of the algorithm [10].

Recommendations: The genetic algorithm continuously works on population to find the best fit solutions. After reaching the termination conditions the algorithm is stop working and generates the final outcomes. The outcome of genetic algorithm is given here (in fig 3) as the final recommended URLs for the end user.

3. CONCLUSION & FUTURE WORK

The proposed work is based on the E-learning recommendation system model based on KNN and GENETIC Search algorithm and user behavior preferences analysis. Every information related to the learner or user behavior is very important. So we keep the details of user's IP address, access log file to record the behavior of user. We have presented an approach to recommend the web contents to the e-learning web portals. This paper integrates the KNN and Genetic search algorithm to propose the new perspective of the recommendation method, which is an innovation idea. We have outlined a framework for modeling a recommender system which improve the online learning experience of user and expressiveness.

In terms of future prospects, performance study and comparative study of our recommendation model with the existing models will be done, through which it will be optimized more. By using this study, proposed model will be improved.

REFERENCES

[1] Paulo, A.; Alves, R.; André, C. D.; (2018): "Adaptive anomaly-based intrusion detection system using genetic algorithm and profiling", Security Privacy. John Wiley & Sons, Ltd.

[2] Merve, B. S.; Ferhat, Ö. Ç.; Ensar G.; (2018) "Detection of attack-targeted scans from the Apache HTTP Server accesslogs", Applied Computing and Informatics, 14, 28-36

[3] Annibale P.; Bogdan D.; Rocco O.; Massimiliano D. P.; Denys P.; Andrea D. L.; (2016): "Parameterizing and Assembling IR-based Solutions for SE Tasks using Genetic Algorithms", 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER), IEEE

[4] D.A. Adeniyi: Z. Wei; Y. Yongquan; (2016): "Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN)classification method",(Applied Computing and Informatics) 12, 90–108

[5] Tiago A. A.; Tiago P. S.; Igor S.; José M. G. H.; (2016): "Text normalization and semantic indexing to enhance Instant Messaging and SMS spam filtering", Knowledge-Based Systems 108 (2016) 25–32, 2016 Elsevier B.V

[6] D. A. Adeniyi; Z. Wei; Y. Yongquan: (2016) "Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN)classification method", Applied Computing and Informatics, 12, 90–108

[7] Francesco R.; Lior R. and Bracha S.: (2015) "Recommender Systems: Introduction and Challenges", Springer Science- Business Media New York, Recommender Systems Handbook, DOI 10.1007/978-1-4899-7637-6_1

[8] Lan U.; Damijana K.; Nina T. and Aleksander A.: (2015) "MOODLE E-LEARNING SYSTEM AND STUDENTS' PERFORMANCE IN HIGHER EDUCATION: THE CASE OF PUBLIC ADMINISTRATION PROGRAMMES", International Conference e-Learning, ISBN: 978-989-8533-40-1

[9] John D.; Katherine A. R.; Elizabeth J. M.; Mitchell J. N. and Daniel T. W.; (2013) "Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology", Psychological Science in the Public Interest 14(1) 4–58

[10] Nidhi J.; Rajendra G., (2019) "An E-Learning based Recommendation System for Semantic Web: A Survey", International Journal on Computer Science and Engineering (IJCSE), Vol. 11 No.01