Nature Inspired Recommender Algorithms for Collaborative Web based Learning Environments

Dinesh Kumar Saini
Faculty of Computing and Information Technology
Sohar University
Sultanate of Oman
Research Fellow and Adjunct Faculty, University of Queensland Australia

Lakshmi Sunil Prakash
Faculty of Computing and Information Technology
Sohar University
Sultanate of Oman

ABSTRACT
The design of recommender systems for various domains has been proposed based on the nature inspired algorithms. In this paper attempt is made to propose a Nature Inspired Algorithms based architecture for recommender system for web based learning environments. The paper also compares between the traditional recommender systems and the nature inspired algorithm recommender systems. Collaborative filtering is proposed for personalized recommendations; user and item attributes are used as filtration parameter. Attributes and rating of the user's similarity is used for collaborative filtering process. Hybrid collaborative filtering is proposed for user and item attribute that can alleviate the sparsity issue in the recommender systems. Traditional systems are studied in detail and all the possible limitations of the traditional systems are bought under attention.

General Terms
Computing, Nature, Algorithms, Web Science.

Keywords
Recommender Systems, web based educational environments, architecture, nature inspired algorithms, optimization, and software testing.

1. INTRODUCTION
The role of recommender systems for decision-making is gaining paramount importance as several domains are now having such systems as an integral component of their architectures [1]. The study of recommender systems was initiated in the mid-90s. Users are by and large familiar with websites like Amazon.com, Netflix, YouTube, It was observed that the magnitude and variety of information available on the internet was overwhelming for a great majority of the users and they were often perplexed when it came to selecting or making a choice or a choice set from a recommended group of items. The reason for incorporating recommender systems in a service or website is manifold.
Of primary importance is the need to

- Improve the efficiency of service offered.
- Attract more users to use the website or service.
- Understand the requirements of the user so that the contents of the system or service can be improved according to this parameter.
- Increase the volume of transactions and be an aggressive competitor in the online transactional systems environment.

Assess the contents available in the website based on ratings and rankings which translates or converts into information that will help recognize or discover the most preferred item in the item collection. Develop trust in the service that will in turn lead to users recommending the items in the service to others surfers, who share similar preferences or trust the recommendations made by this particular user.

Predicting the demand or next possible addition to the content repository by studying user patterns based on feedback from several user sessions in the website. A learner's activity is guided by Protus which is an intelligent web-based Programming Tutoring System .It is used for guiding the learner's activities and recommends relevant links and actions to him/her during the learning process. In [2] the authors discuss how Nutch's automated crawling and indexing techniques as well as standardized educational content indexing are used to build content profiles, and Web usage mining techniques (clustering and association rule mining) are used to build user profiles. Hybrid recommendations (content based filtering and collaborative based filtering) were used in the recommendation phase. The approach in this paper is towards filtering the learners accessing the system into clusters based on their learning styles and subjects of study. We also take into account the ratings earned by learners based on the number

2. TRADITIONAL RECOMMENDER SYSTEMS
Collaborative filtering systems face the problem of shilling. It is the term used to refer to the injection of fake user profiles into the rating database of a recommender system, with the intent of influencing the recommendation behavior of the system. In this the shilling problem will not arise as the learners will be having unique id generated at the time of course registration, the system will authenticate the user on the basis of their registration details at the institution.

Users expect collaboration based learning environments are required to be able to handle increasing number of users and learning items. However the real challenge lies in getting recommendations and ratings from users. This is called the data sparsity problem [3,4].

Table 1: Traditional Algorithms Comparison

Data Sparsity Algorithms	Descriptions
Singular Value Decomposition (SVD) [23],	a closely-related factor analysis technique remove unrepresentative or insignificant users or items to reduce the dimensionalities
Latent Semantic Indexing (LSI) SVD [5]	similarity between users is determined by the representation of the

	users in the reduced space
Principle Component Analysis (PCA),[6],	a closely-related factor analysis technique remove unrepresentative or insignificant users or items to reduce the dimensionalities
Eigentaste,[6]	Goldberg et al. developed which applies to reduce user-item dimensionality
hybrid collaborative filtering approach [7]	How to exploit bulk taxonomic information designed for exact product classification to address the data sparsity problem of CF recommendations, based on the generation of profiles via inference of super-topic score and topic diversification

3. TRADITIONAL RECOMMENDATION ALGORITHMS

The following are some of the traditional recommendation algorithms that have been developed, these include

- collaborative filtering [3,4],
- content-based analysis [5],
- spectral analysis [6,7] and
- Iterative self-consistent refinement [8, 9].

What most traditional collaboration filtering algorithms have in common is that they are based on similarity, either of users or items or both[8]. Such approach is under high risk of providing poor coverage of the space of relevant items. As a result, with recommendations based on similarity rather than difference, more and more users will be exposed to a narrow band of popular items. Although it seems more accurate to recommend popular items than niche ones, being accurate is not enough [10]. Diversity and novelty are also important criteria of algorithmic performance. The diversity-accuracy dilemma becomes one of the main challenges in recommender system.

These algorithms face similar problems like

The tasks for which collaborative filtering can be performed are [3,7]

1. Suggest items in the data set which the user may find interesting

2. Create a group of users who share the same interest

3. Suggest a recurring set of similar set of items that a user may find interesting

4. Suggest details about a selected item..

5. To group results of previous searches and predict recommendations for future

4. REASONS FOR NEW ALGORITHMS IN RECOMMENDER SYSTEMS

The large scale of data in recommender systems is a major reason for the need to move away from the traditional algorithms which include the collaborative algorithms (Pearson's coefficient.

Nature Inspired Algorithms have been very popular in recent years as they have been able to provide simple and effective meta-heuristic solutions to complicated problems in the real-world

Several Bee Colony algorithms have been proposed based on the foraging behaviour which includes the food searching and searching for new nest behaviours of bees.

Table 2: New Algorithms in Recommender Systems

Bee Inspired Algorithm	**Essence**	**Application**
Bees System (BS) Algorithm [9]	Collects maximum nectar from the hives in the bee trajectory.	Tested on travelling salesman problem. Produced good results
Bee Colony Optimization (BCO) [10]	It determines the route to be taken taking into consideration the distance and demand at various nodes in the route.	Vehicle routing Problem
Honey Bee Algorithm[11]	Honey bee colonies are self-organised in that they have reach the food source with the help of other bees involved in the same activity	Dynamic allocation of internet sources
Beehive Algorithm[12]	Based on the local information that a short distance bee agent collects in a food searching zone	Applied to routing in wired computer networks.
Ant Colony Algorithm [15]	Based on the pheromone secretion of ants which helps to create a trail for the ants coming after.	In VRS to help vehicles find the least congested path
Bat inspired Algorithm[14]	Echolocation property of bats	Identifying the correct object and discriminating between objects in a search routine.

5. PROPOSED WORK BASED ON BEE COLONY ALGORITHM

In a bee colony, the queen bee can be compared to a highly rated user. All the other bees in the bee colony are prone to the influence of this queen bee. In the same way, learners who have high success rates influence the learning decisions of other learners in the group. Each cluster can be compared to a bee colony with its own queen bee.

5.1. Contents in the Learning Management System

The components of the LMS are divided into Learners, Instructors and Learning items.

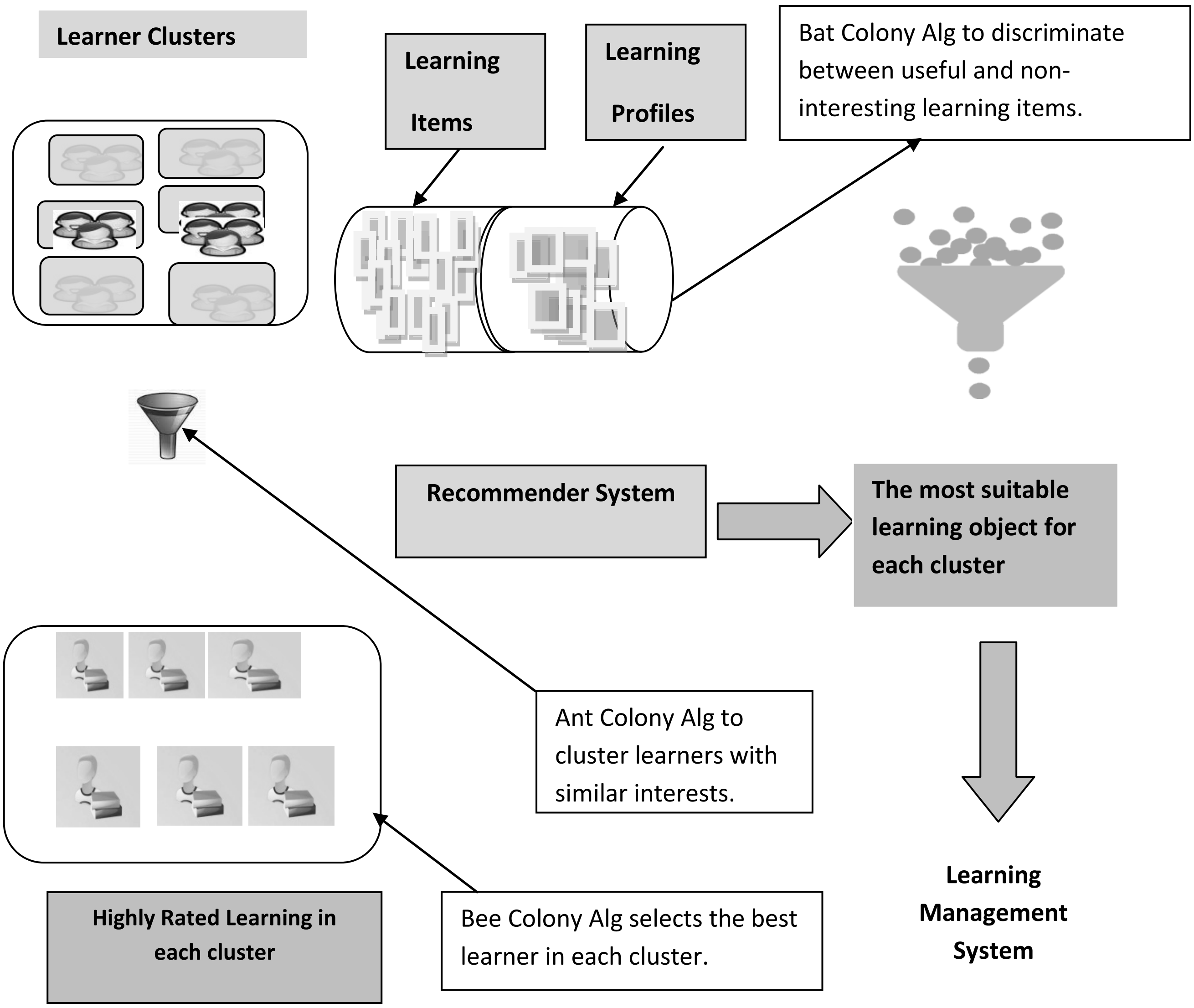

Figure 1

In Figure 1, we discuss the three algorithms which determine the recommender's ability

to provide the most optimised search results to its users.

The Ant Colony Algorithm [18] is required by the recommender system to cluster similar learners. These clusters have dimensions such as learning style, and subject interest. Once the learning style and subject interest are gleaned from the learner profiles, then a trail is created for other users with similar interests to be clusters together on the basis of these two traits.

Similarly the Bee Colony algorithm helps to identify the learner with best ratings on the basis of the recommender systems calculations of access time and assessment scores of the learners. This helps to filter the best learner in each cluster.

While the Bat Algorithm helps to discriminate between the useful learning objects and others which are not useful , so that the highly rated learner in the cluster is now able to receive the best recommendations for his /her learning module.

The Learning Management System consists of the following entities:

Course Name

- *Subject*
- *Course Coordinator*
- *Course Description*
- *Course Learner Profile* – Advanced , Intermediate , Beginner
- *No of learners is denoted by N*

Each course will have **learning items** .Its attributes will be as follows

- *Learning item_id* – unique identifier
- *Learning item_type* – assessment item, learning material, group assignment etc.
- *Learning_outcome* – expected learning outcome achieved after completing the learning item.
- *Learning item_filetype* - audio, video, presentation, word document.
- *Content* – advanced, intermediate, beginner
- *Suggested for Learning_style* – Using Vark Learning Styles[7] - Verbal, aural, visual, logical, kinaesthetic ,solitary or social
- *Frequency of use* (F_q) - total score of accesses earned by the item during the duration of the module.

♦ The learner group is categorised by the learning style preferences collected from the learner profile.

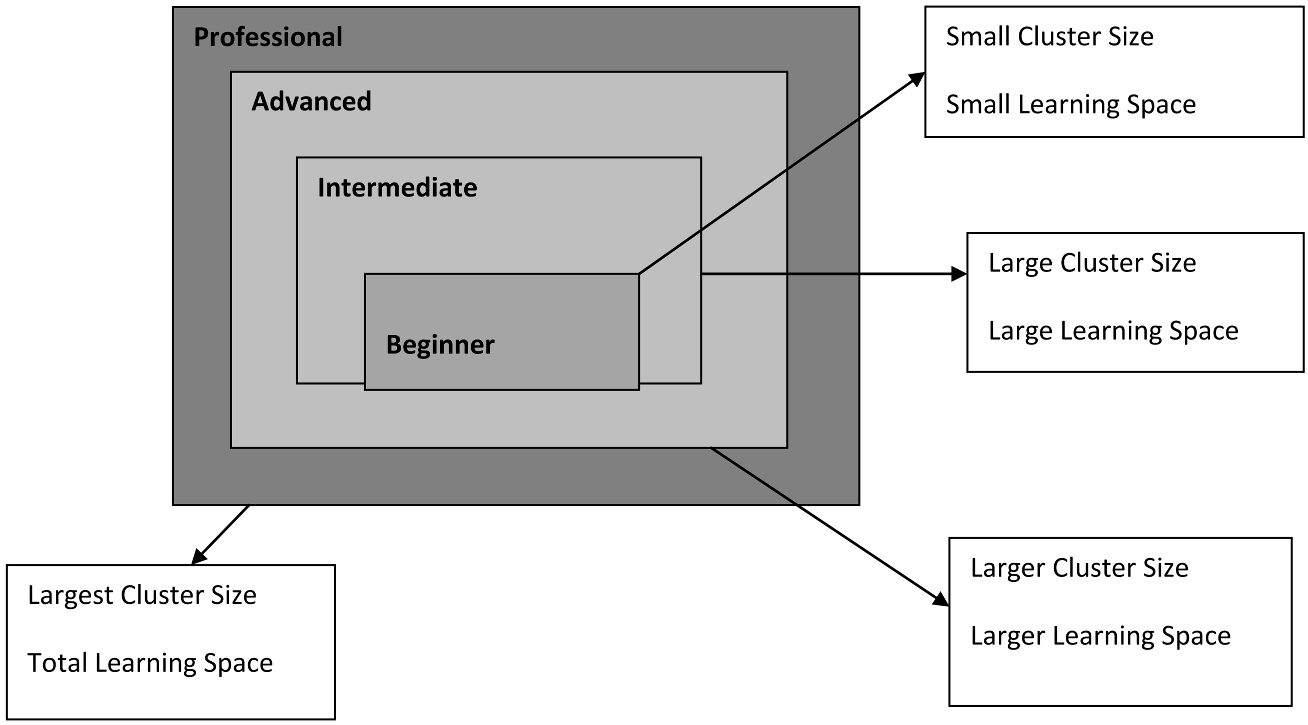

Figure 2: Learner Groups and Space

Recommendations_earned (LR_n)- calculated by the recommender system on the basis of learner access and duration of use.

- *Item_Rating* (IR_n) - ratings earned by the item, calculated by frequency of access by top-rated learners and recommendations earned.

Each learner will have the following key attributes

- *Learner_id* – student registration number.
- *Learning_style* (L_s)- Verbal, aural, visual, logical, kinaesthetic ,solitary or social
- *Assessment_result* (R) – achieved by the learner on completion of a module.
- *Learner_rating* (Lr_n) – ratings earned by the learner on the basis of assessment results.
- *Learner Cluster (LC)* – category or categories to which the learner may belong

While clustering learners by the learning style, we also need to deliver the most suitable learning content to the learner. Normally suitability of content is measured by the nearest neighbor algorithm or Pearson's coefficient, however using The suitability of the content can be assessed by the recommendations of the learners who score higher assessment results; this learner becomes the learner with the highest learner rating. According to the QBE algorithm, the queen bee is the learner with the most authority to lead the group, in this manner the recommender system can suggest to each learner the most suitable items for his study based on the recommendation ratings earned by each item

The recommender based learning systems will not suffer from sparsity problems if the system can rate any item by the

$$Lr_n = \sum L_s + \sum (Lr_n) + \sum LR_n + \sum R_{recommended}$$

number of items that is available in the content database by the number of users accessing the item multiplied by the access times.

Similarly each learner profile will be having a rating once he completes the module depending on his/her performance in the assessment for that module.

6. THE ABB ALGORITHM

In this algorithm a user cluster is created based on the similarity in learning styles and similarity of subject interest.

Here the best performing learners for a module receive the highest ratings from the module or course coordinator. These top-rated learners are then filtered by their learning styles; these learning styles can be termed L_s

The Mean average recommendations earned R_n by the item are then calculated.

The Mean average ratings for the learner are also calculated across each assessment, $\mathbf{MLR_n}$

The Learning Style factor $\mathbf{L_s}$ influences the categorisation of learners into clusters.

$$\sum Fq + \sum LR_n + \sum IR_n = LR_n$$

$$LRn = \sum_{i=1}^{n} Fq + \sum_{i=1}^{n} LRn + \sum_{i=1}^{n} IRn$$

With time and duration of access, the recommendation earned and frequency of access

$$Lr_n = {}^{d}\!/_{dx}(F_q) + {}^{d}\!/_{dx}(IR_n) + \sum_{i=1}^{n} LR_n$$

Learning Rating, Lr_n =

Centroid distance $F_{2} = \sum_{i-1}^{N} j \in \{1,...K\} d(z_{i,mj})$

Variance Ratio criterion = F_4 = VRC = trace B /(K-1) / trace W/(N-K)

Intra and inter cluster distance = F_5 =

$\sum_{i=1}^{K} D_{\text{inter}}(c_i)w - D_{\text{intra}}(c_j)$, w is a parameter.

Dunn's, index

$$F_6 = DI/K = i \in K, j \neq 1\{\frac{\delta(c_i, c_j)}{k \in \{K(\Delta(c_k))\}},$$

where $\delta(c_i, c_j)$ = min {d(z_i, z_j) : $z_i \in c_i, z_j$}

7. TESTING THE RECOMMENDER SYSTEMS

Recommender systems are testing based on the accuracy and closeness of the recommendation suggested by the algorithm used. [19] The scope of the system will be tested the used of the best algorithm, assumptions made for learners, baseline documents, methodology adapted to designing the proposed systems, entry criteria. As shown in figure 3 concept and formulas will be the basis of the recommendation with structure and relations.

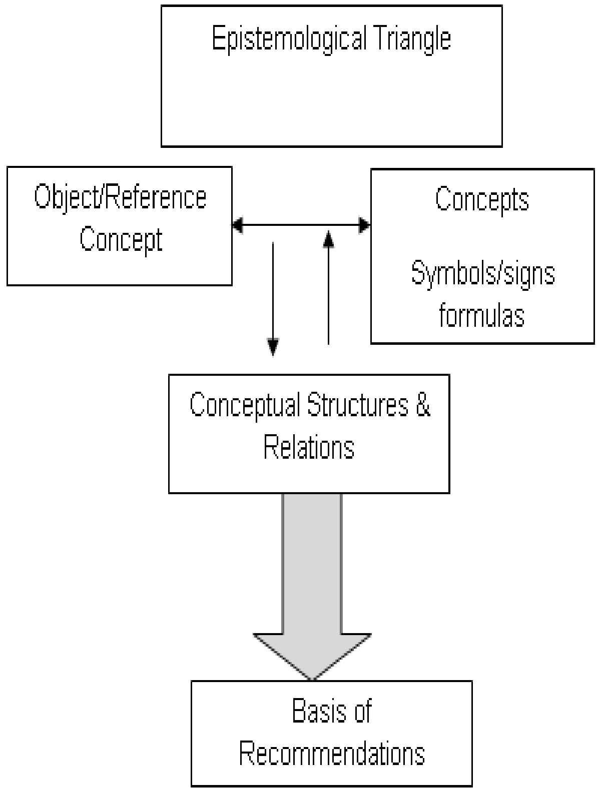

Figure3: Epistemological Triangle and recommender systems

7.1 Testing Process

For recommender systems we need to test how the systems adapt recommendation process, which algorithm comes closer to the expectations and preadaptation in the process. The systems need to be tested on sufficient explosion and for performance and accuracy [20,21]. The system must be tested for fault tolerance, prevention and forecasting of faults in the system is difficult to predict but it is still needed in the recommendation systems. Implementation of supervised learning mechanism in the recommendation systems is very much desired to that false recommendations can be minimized [22].

Context perspective in recommendation systems using qualitative research is very subjective and situations arising from the qualitative research are not easy to handle. Moreover, qualitative research methodologies are concerned with the opinions, experiences and feelings of individuals [16]. Testing such recommendations is not easy task but various testing techniques will be employed in the given situation. [23]

As shown in figure no 4 various testing strategies will be adopted for checking the accuracy and perfection of the system. Recommendation functions, GUI components, systems acceptance and accuracy will be tested and validated before adapting the particular algorithm for the recommendation system [17].

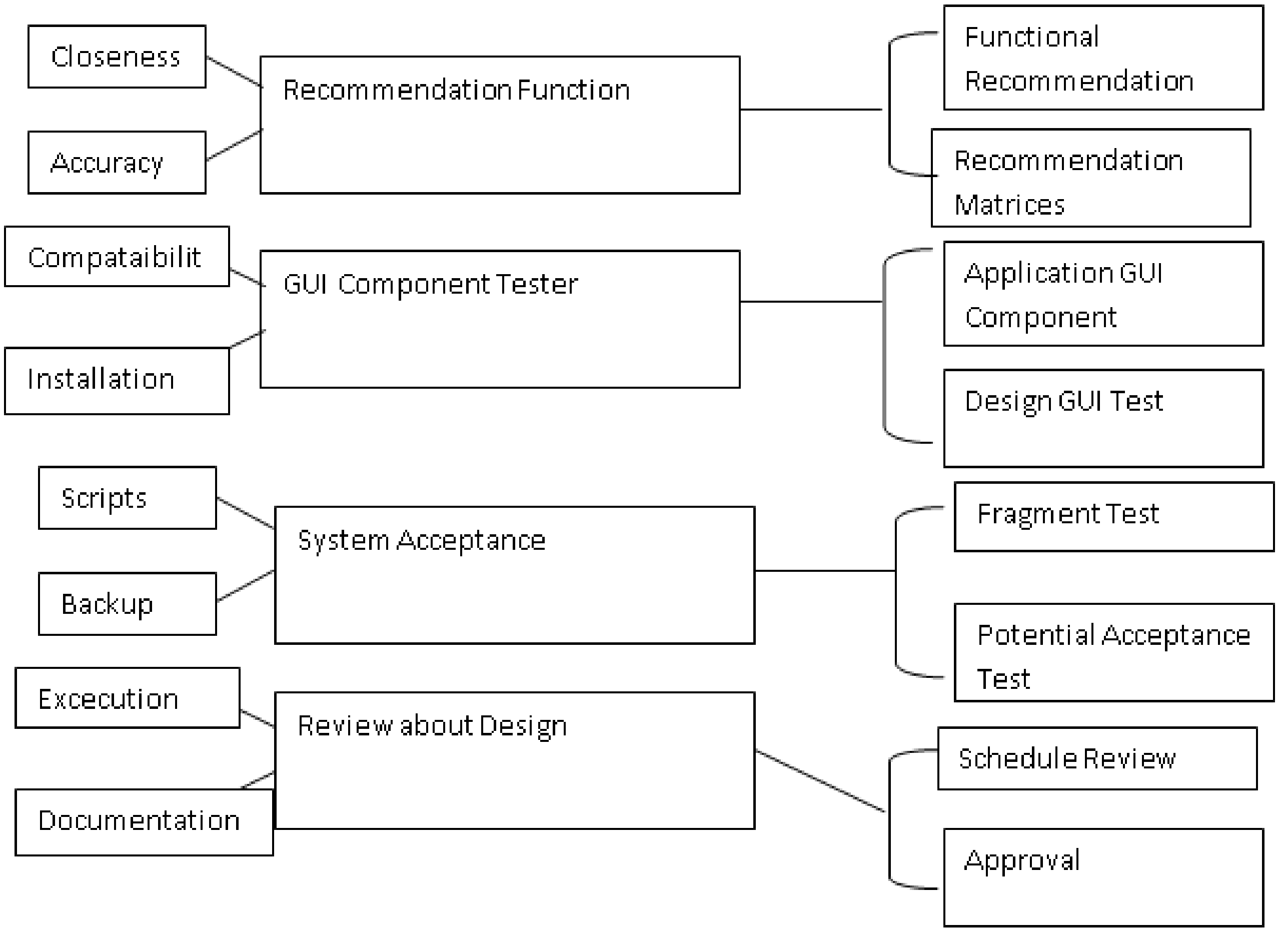

Figure 4: Software Testing Process for Recommender Systems

Table4: Software Test Cases for the recommender system

S.No	Test Case ID	Objective Id	Category	Condition	Expected Result	Actual Result	Req.ID
1	General Function	Performance and Functionality	Sponsor /development /Testing recommender	Which algorithm is better	Best Recommender	Accuracy	Which is better

8. CONCLUSIONS AND FUTURE WORK

In this paper we proposed recommender systems for various Knowledge domains based on nature inspired algorithms. Recommender systems architecture based on nature inspired algorithm is for web based learning environments. The paper also compares between the traditional recommender systems and the nature inspired algorithm recommender systems. Collaborative filtering is proposed for personalized recommendations; user and item attributes are used as filtration parameter.

Attributes and rating of the user's similarity is used for collaborative filtering process. Hybrid collaborative filtering is proposed for user and item attribute that can alleviate the sparsity issue in the recommender systems. This system need to be tested and validated that nature inspired algorithm perform better than traditional algorithms.

First Bee colony optimization algorithm was used to design and propose the recommendation systems, and it is suggested that can be integrated in the Learning content management systems.

9. REFERENCES

[1] Zhang, Fuzhi, and Quanqiang Zhou. "A Meta-learning-based Approach for Detecting Profile Injection Attacks in Collaborative Recommender Systems." Journal of Computers 7.1 (2012).

[2] Khribi, Mohamed Koutheaïr, Mohamed Jemni, and Olfa Nasraoui. "Toward a hybrid recommender system for e-learning personalization based on web usage mining techniques and information retrieval." World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education. Vol. 2007. No. 1. 2007.

[3] Prakash, Lakshmi Sunil, Dinesh Kumar Saini, and Narayana Swamy Kutti. "Integrating EduLearn learning content management system (LCMS) with cooperating learning object repositories (LORs) in a peer to peer (P2P) architectural framework." ACM SIGSOFT Software Engineering Notes 34.3 (2009): 1-

[4] Ludford, P.J., Cosley, D., Frankowski, D., Terveen, L.”Think Different: Increasing Online Community Participation Using Uniqueness And Group Dissimilarity”. Proceedings of the SIGCHI conference on Human factors in computing systems (2004),ACM Press: Vienna, Austria p. 631-638.

[5] Sunil Prakash, Lakshmi, Narayana Swamy Kutti, and A. S. M. Sajeev. "Review of challenges in content extraction in web based personalized learning content management systems." Proceedings of the 12th International Conference on Information Integration and Web-based Applications & Services. ACM, 2010.

[6] Goldberg,T.Roeder,D.Gupta,,Perkins,“Eigentaste: a constant time collaborative filtering algorithm,” Information Retrieval, vol. 4, no. 2, pp. 133–151, 2001.

[7] T. Landauer, M. Littman, and Bell Communications Research (Bellcore), “Computerized cross-language document retrieval using latent semantic indexing,” US patent no. 5301109, April 1994.

[8] K. Pearson, “On lines and planes of closest fit to systems of points in space,” Philosophical Magazine, vol. 2, pp. 559–572.

[9] Prakash, Lakshmi Sunil, and Dinesh Kumar Saini. "E-assessment for e-learning." Engineering Education: Innovative Practices and Future Trends (AICERA), 2012 IEEE International Conference on. IEEE, 2012.

[10] Fleming, N. (1995), “VARK – a guide to learning styles”, available at: www.varklearn.com/englis h/index.asp

[11] Billsus and M. Pazzani, “Learning collaborative information filters,” in Proceedings of the 15th International Conference on Machine Learning (ICML ’98), 1998.

[12] P. Lucic, and D. Teodorovic, “Bee system: Modeling Combinatorial Optimization Transportation Engineering Problems by Swarm Intelligence,” Preprints of the TRISTAN IV Triennial Symposium on Transportation Analysis, Sao Miguel, Azores Islands, pp. 441-445, 2001

[13] D. Teodorovic, and M. Dell'Orco, “Bee Colony Optimization - A Cooperative Learning Approach to Complex Transportation Problems,” Advanced OR and AI Methods in Transportation, pp. 51-60, 2005

[14] S. Nakrani, and C. Tovey, “On Honey Bees and Dynamic Allocation in an Internet Server Colony,” Proceedings of 2nd International Workshop on the Mathematics and Algorithms of Social Insects, Atlanta, Georgia, USA, 2004

[15] H.F. Wedde, M. Farooq, and Y. Zhang, “BeeHive: An Efficient Fault-Tolerant Routing Algorithm Inspired by Honey Bee Behavior,”

[16] H.F. Wedde, M. Farooq, T. Pannenbaecker, B. Vogel, C. Mueller, J. Meth, and K. Jeruschkat, “BeeAdHoc: An energy efficient routing algorithm for mobile ad hoc networks inspired by bee behavior,” GECCO 2005, Washington DC, USA, 2005

[17] Yılmaz, S., E. Ugur Kucuksille, and Y. Cengiz. "Modified Bat Algorithm." Electronics & Electrical Engineering 20.2 (2014).

[18] Ant Colony, Optimization and Swarm Intelligence, Eds. M. Dorigo, LNCS 3172, Springer Berlin, pp. 83-94, 2004

[19] Liamputtong, P. Qualitative data analysis: Conceptual and practical considerations. Australian Journal of Health Promotion, 20(2), (2009). 133–139.

[20] WM Omar, DK Saini, M Hasan “Credibility of Digital Content in a Healthcare Collaborative Community”, Software Tools and Algorithms for Biological Systems, pp 717-724, 201.

[21] N Gupta, D Saini, H Saini “Class Level Test Case Generation in Object Oriented Software Testing”, Web Engineering Advancements and Trends: Building New Dimensions of Information Technology, pp203-210, 2010.

[22] Saini, Dinesh Kumar, Lakshmi Sunil Prakash, and M. Goyal. "Emerging information technology and contemporary challenging R & D problems in the area of learning: An artificial intelligence approach." Engineering Education: Innovative Practices and Future Trends (AICERA), 2012 IEEE International Conference on. IEEE, 2012.

[23] Lima, Salvador, and José Moreira. "A Semantic Framework for Touristic Information Systems." Cases on Open-Linked Data and Semantic Web Applications (2013): 132.