Intelligent Automation And Soft Computing, 2020
Copyright © 2020, TSI® Press
Vol. 26, no. 2, 269–280
https://doi.org/10.31209/2019.100000150

# Personalized Nutrition Recommendation for Diabetic Patients Using Optimization Techniques

**Bhavithra Janakiraman[1], Saradha Arumugam[2]**

[1] Department of Computer Science and Engineering, Dr. Mahalingam College of Engineering and Technology, Pollachi, Tamilnadu 642003, India
[2] Department of Computer Science and Engineering, Institute of Road and Transport Technology, Erode, Tamilnadu 638316, India

## ABSTRACT
Personalization in recommendation system has been emerging as the most predominant area in service computing. Collaborative filtering and content based approaches are two major techniques applied for recommendation. However, to improve the accuracy and enhance user satisfaction, optimization techniques such as Ant Colony and Particle Swarm Optimization were analyzed in this paper. For theoretical analysis, this paper investigates web page recommender system. For experimentation, Diabetic patient's health records were investigated and recommendation algorithms are applied to suggest appropriate nutrition for improving their health. Experiment result shows that Particle Swarm Optimization outperforms other traditional methods with improved performance and accuracy.

KEYWORDS: Ant Colony Optimization, Clustering, Diabetics, Nutrition Recommendation, Particle Swarm Optimization, Personalization, Recommendation System.

## 1 INTRODUCTION

INFORMATION retrieval is the process where user-relevant information will be extracted from data-ware house which is linked with enormous amount of data sources. ŞuleGunduz-Oguducu (2006) have proposed that, as information on the web is increasing day by day, recommendation systems are introduced with the motivation of reducing the search burden of end users. In today's information era, search engines are accessed frequently by web user for their regular activities. ŞuleGunduz-Oguducu (2006) have proposed that rather than using traditional searching methods, researchers employ artificial intelligence in search engines to optimally retrieve the required information

Recommendation System (RS) is one of the major research areas under data mining that predicts and suggests relevant information based on the requirement for end users. Recommendation system thus plays a vital role in information retrieval by reducing the delay faced by end users while searching and thereby satisfying them by suggesting relevant documents. Personalization in RS aims in providing tailored search results in order to increase user satisfaction by creating specific user profile for each user. For example, during web page recommendation, Murat Goksedef (2010) proposes that, web-user profiles are created by examining the search patterns and navigation history and analyzing their interest. Features or attributes are those which stores specific characteristics of end-user. User profiles will be populated through identified attributes that are relevant to the application. Those attributes identified for profile specification can be modified based on the application where recommendation is applied. The recommender system identifies the similarity among the user profiles by comparing the user profiles. Top "k" profiles that are highly ranked based on the similarity with current active user (AU) will be considered for further analysis. The value of "k" depends upon the recommendation engine, which can be fine-tuned to achieve desirable outcome. The web pages that are visited by these top "k" users will be recommended for the current AU.

The main goal of this paper is to reduce the complexity of handling data through clustering and to increase the accuracy during recommendation through personalization. This simplifies the users' work on giving their preferences (e.g., the user interests and

CONTACT Bhavithra Janakiraman bavi.rr@gmail.com

their personal information). The accuracy level while predicting options during recommendation will be increased by applying optimization techniques and machine learning processes. This paper investigates the role of clustering and optimization in recommendation systems. Fitness function applied in Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO) is adapted suitably, enabling the algorithm to find the best suggestion and hence increase the accuracy. The performance of recommendations before and after applying optimizations has been analyzed. To theoretically investigate the effectiveness of the proposed system, personalization in web page recommendation was analyzed. Web user profiles comprising of eight user specific attributes and two content specific attributes were identified. Recommendations for AU will be suggested based on the pages visited by similar user profiles. To test the effectiveness of the proposed idea, personalized nutrition recommendation has been experimented for Diabetic patients. Unique profile was developed for each diabetic patient based on their past history of blood-glucose level readings. The profile comprises of thirty attributes, the former ten attributes specifies the generic aspects of a patient. Later twenty attributes are mostly related to their blood glucose level. The complete details of attributes with description are narrated in next coming sections. Based on the similarity among the individual profiles, successful nutrition therapy underwent by nearest (best matching) and optimal patient (profile) will be recommended for each patient. Various test cases were conducted to analyze the effectiveness of proposed recommendation techniques.

The remaining portion of the paper is organized as follows: Section II discusses the related work of this paper. Section III discusses about the overall system architecture which narrates about construction of user profiles and clustering using Naïve Bayes probabilistic algorithm. In section IV, optimization algorithms such as ACO and PSO are analyzed to improve personalization. In Section V, experimental setup for testing the proposed algorithms for nutrition recommendation for diabetic patients was discussed along with evaluation metrics, results and inferences observed. Section VI concludes the paper with scope for further improvement.

## 2 RELATED WORK

RECOMMENDATION system commonly uses collaborative filtering algorithm in order to effectively predict web pages that will be likely to be clicked by an active user. User's with similar interest are grouped together to form clusters. Similarities among the profiles are computed to identify the collaboration among users. Recommendation algorithm will be employed on the clusters of users and suggesting the web pages that are visited by the users who were highly synchronized towards the active user. In order to identify effective grouping of web users, their search interest has to be obtained. Two major types of collecting user's interest are explicit approach and implicit approach. In explicit approach, user's star ratings and feedbacks are analyzed to gather the interest on a web page. Through implicit approach, user's interest is collected through indirect means such as time spend by an user on a web page, click rate, session duration, etc.., These indirect opinions are summarized as rating by a user towards the corresponding web page. Such ratings collected by the RS algorithms include both implicit and explicit opinions. Collection of explicit ratings is simple approach, as user opinions are mostly in direct nature. On the other hand, determining implicit feedback is a challenging task. Based on the accessing time and nature of visits user's interest towards the item has to be optimally identified.

For any online recommendation process, web log files are collected from the users' browsing history, consisting of IP address, date & time of visiting the web pages, method URL/protocol, status, received bytes etc. Web page contents and keywords are then extracted from such processed log files. User profiles are constructed for each user based on implicit/ explicit feedbacks and keywords. These user profiles are usually represented as matrix format. Machine learning algorithms are employed along with RS algorithms to estimate the similarity among user profiles. When new user enters a search query, personalized recommendations are predicted and suggested based on similarity among new user profile and existing user profiles.

Freddy L´ecu´e (2010) proposes that semantic content based approach is another effective recommendation process where semantic similarities between web pages are analyzed. Today many researchers try to combine semantic similarity within the content and collaborative based approaches to improve efficiency. For analyzing the semantic content, user search pattern which are collected from the past history and their personal information acts as implicit and explicit inputs respectively. In many such recommendation systems, explicit inputs that include user's name, user id, area of interest, page likes, feedbacks, etc.., are not considered to be mandatory for predicting web pages that could be further recommended. Kazuyoshi Yoshii, et.al (2006) proposed that, latent user preferences are considered where the content alone is not sufficient to find out the interests about the user. Hence overall ratings of web pages are also considered to include unobservable preferences to enhance the recommendation.

Alexandrin Popescul Lyle, et.al (2001) proposed unified and hybrid framework that combines both content and collaborative based approaches along with latent information. The data which present sparsely that are same as the users, interests is hence recommended. Alexandrin Popescul Lyle, et.al (2001) proposed a generative probabilistic model that incorporates three-way co-occurrence data among users, items and item content which combines both content and collaborative approach. In three-way aspect model users are classified based on the document they access along with the latent variables. Here, the core topic that generates the document retrieval has been considered for computing latent variables. Along with these techniques, k-Nearest Neighbors are used to find the most relevant document which is to be recommended to the new user. Byron Bezerra,et.al ( 2004) and Katja Niemann, et.al, (2015) used such types of recommendation systems to handle the data among the sparse environment

Weihui Dai, et.al (2009), Pablo Loyola, et. al (2012) and Xiao-Feng Xie, et.al (2002) proposed Ant Colony Optimization (ACO) which is a probabilistic based model from the family of Swarm Intelligence. ACO employ meta-heuristic methods of optimizations to solve computational problems. It is based on real world phenomena followed by ants in search of its food. The field of bio-inspired computing customizes the phenomena followed by the biological creatures for solving computational problems. In addition to ACO, Deepa.S.N, et.al (2011) suggests that Particle Swarm Optimization (PSO) is another efficient optimization model from the family of swarm intelligence which adapts natural intelligence for computing. The collective behavior of self-organizing particles has been modified suitably for solving computational problems. PSO is also meta-heuristic algorithm, containing a set of algorithms which is used to define heuristic methods. Deepa.S.N, et.al (2011) suggests that PSO defines a Fitness Function (FF) or Objective Function (OF) expresses the core functionality of research to be optimized. FF could be either in maximization or minimization phenomena.

Effective nutrition therapy plays a vital role for diabetic patient's health improvement. American Diabetes Association (2008) suggests that appropriate nutrition intake helps in preventing diabetics, managing the blood glucose in appropriate level or at least reduces the risk factor of irregular body-insulin secretion. American Diabetes Association (2008) proposes that nutrition recommendation helps for patients in three levels of diabetic prevention namely (i) Primary prevention which helps in preventing from diabetics, (ii) Secondary prevention in order to prevent from further complications, and (iii) Tertiary prevention to prevent mortality Nutrition recommendation also helps in providing nutrition education and self-management by patients themselves. Nutrition therapy aims innutrition recommendations that promote healthy eating and assist in achieving appropriate glucose, lipid, and blood pressure levels. The following section describes the overall system architecture for enhancing personalized recommendation using optimization techniques.

## 3 SYSTEM ARCHITECTURE

THE main objective of this paper is to apply optimization techniques to improve the performance of recommendation system. The overall architecture of the proposed idea has been depicted in figure 1. The proposed system functions under two central modules namely "offline process" and "online process". During offline process, cluster of user profiles are formulated Naïve Bayes probabilistic (NP) approach. These clusters are populated with similar user profiles. The algorithms for similarity matching and clustering using NP algorithms are described in forth coming sections. Such clusters are compared with current active user profile, which results in one best matching cluster. Profiles grouped under the best matching cluster acts nearest neighbors. From the dynamic set of matching nearest neighbors, top "k" neighbors are shortlisted. Finding optimal "k" value decides the success of optimization algorithms that runs during online process. These final "k" user profiles are the seed input for optimization algorithms. Comparing the query form online active user and "k" profiles, items are predicted and recommended for online user.

### *3.1 Construction of User Profile*

Data preprocessing is a stage in which the information about the end-users are collected and analyzed for personalization. In preprocessing stage, the log file is cleansed by removing unwanted information such as inappropriate and incomplete entries. Finally, the user profile is constructed by identifying user relevant attributes. The attributes identified for profile specification can be modified based on the application where recommendation is applied. In this research work, proposed algorithms are implemented for web page recommendation and as an extension it has been tested for nutrition recommendation for diabetic patients. For Web page recommendation, Abirami. S, et.al (2017) suggests that web-user specification has been profiled comprising of ten features that were classified as Usage-Based features (8 attributes) and Content-Based features (2 attributes). For Nutrition recommendation, diabetic patient specification has been recorded based on thirty features which are again classified as Generic features (10 attributes) and Blood Glucose-Specific features (20 attributes).

In order to adapt the contribution of each feature for developing user profile, Weights ($\beta$) were assigned for each feature. In the proposed system, minimum value for $\beta$ is fixed as 1.0 and maximum value ranges up to 2.0. Adding appropriate weights to the corresponding features results in enhanced accuracy, as weighted significant features produces reminiscent user profiles. The attributes used for profiling along with weight ($\beta$) assignment is discussed in Table 1. Similarly for diabetic patient profiling, the identified attributes were assigned with weight ($\beta$) which is deliberated in Table 2.

**Table 1.** List of user profile attributes for web page recommendation.

| Attribute Name | Weight (β) | Description |
|---|---|---|
| TOP | 1.00 | Time on Page: Total time spent on a web page by the corresponding user |
| TOS | 1.00 | Time on Site: Total time spent on a web site (a set of web pages) by the corresponding user |
| ATP | 1.75 | Average Time at this Page: Average time spent for any web page pi by active user (Average of all sessions for certain threshold) |
| BR | 1.75 | Bounce Rate: Access rate of a web page pi between all the corresponding sessions |
| ER | 1.50 | Exit Rate: Number of times the corresponding web page pi is the end page of that session |
| CR | 1.75 | Conversion Rate: The Ratio between total sessions of web usage by the active user to the total number of sessions that contains the web page pi |
| NOV | 1.50 | Number of Visitors: Denotes the priority of a web page which is computed based on the total number of visitors for pi. |
| APR | 2.00 | Average Page Rank: Average time spent by the user on webpage pi x number of times pi is accessed by different users |
| SK | 1.75 | Top Similar Keywords: Search keywords with maximum frequency computed for each ranked page pi. |
| ASM | 2.00 | Average Similarity between Keywords: Semantic similarity of top keywords between each neighboring user and Active User (AU). |

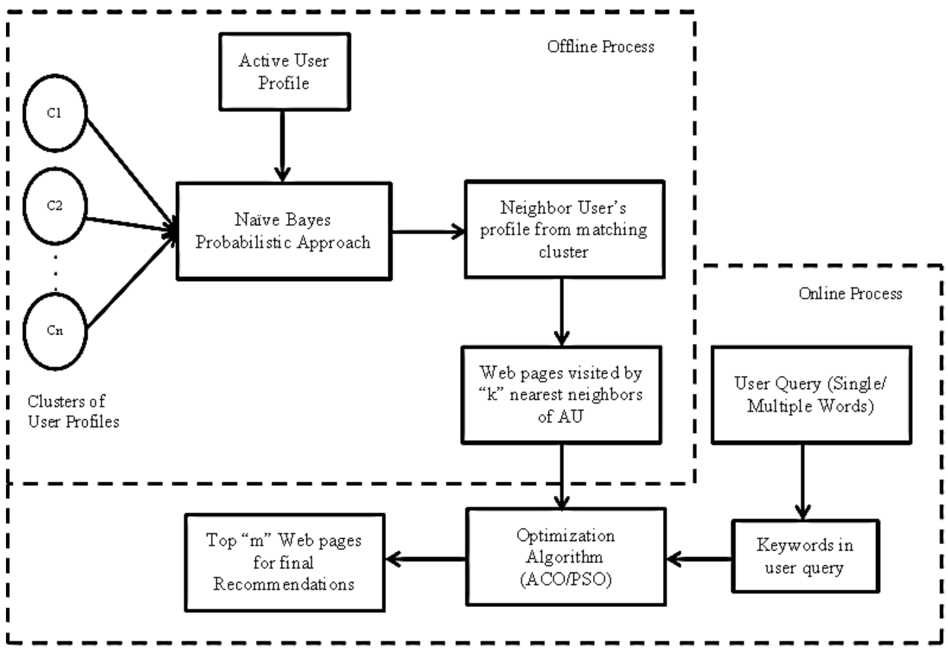


**Figure 1.** Architecture of personalized recommendation using optimization techniques.

**Table 2.** List of user profile attributes for nutrition recommendation for diabetic patients.

| Attribute Name | Weight (β) | Description |
|---|---|---|
| AG | 1.50 | Age of the patient |
| GN | 1.00 | Gender (Male/Female/ Transgender) |
| WT | 1.75 | Weight in pounds |
| HT | 1.00 | Height in inches |
| BMI | 1.75 | Body Mass Index in Units |
| HB | 2.00 | Blood Hemoglobin level in units |
| HDL | 1.50 | High-density lipoprotein in units |
| LDL | 1.50 | Low-density lipoprotein in units |
| BP | 1.75 | Normalized blood pressure level |
| PL | 1.00 | Pulse rate of the patient |
| RID | 2.00 | Regular insulin dose given to the patient |

| Attribute Name | Weight (β) | Description |
|---|---|---|
| NPH | 2.00 | NPH insulin dose given to the patient |
| ULID | 2.00 | UltraLente insulin dose given to the patient |
| BGM1 | 1.50 | Unspecified blood glucose measurement (Sample 1) |
| BGM2 | 1.50 | Unspecified blood glucose measurement (Sample 2) |
| pB_BGM | 2.00 | Pre-breakfast blood glucose measurement |
| PB_BGM | 1.75 | Post-breakfast blood glucose measurement |
| pL_BGM | 2.00 | Pre-lunch blood glucose measurement |
| PL_BGM | 1.75 | Post-lunch blood glucose measurement |
| pSu_BGM | 2.00 | Pre-supper blood glucose measurement |
| PSu_BGM | 1.75 | Post-supper blood glucose measurement |
| pSn_BGM | 2.00 | Pre-snack blood glucose measurement |
| HYP | 1.75 | Hypoglycemic symptoms seen in the patient |
| TMI | 1.50 | Typical meal ingestion |
| MTMI | 1.50 | More-than-usual meal ingestion |
| LTMI | 1.50 | Less-than-usual meal ingestion |
| TEX | 1.75 | Typical exercise activity |
| MTEX | 1.75 | More-than-usual exercise activity |
| LTEX | 1.75 | Less-than-usual exercise activity |
| USE | 1.50 | Unspecified special event |

### 3.2 Clustering using Naïve-Bayes Probabilistic Approach

NAÏVE Bayes algorithm applies probabilistic based class conditional independence approach for clustering items. Meghna Khatri (2012) and Kebin Wang, et.al (2011) proposes that feature vectors of known items are used to train the system during clustering. Mustansar Ali Ghazanfar, et. Al (2004) proposes that one of the promising aspects of Naïve-Bayesian algorithm is the independency among each feature. It can effectively consider all the features that are extracted from the users' log file which helps to increase the efficiency of the recommendation to the active use.

In the current work, we have extracted the ten features from each user profile in order to train the recommendation system. When any Active User (AU) enters the search query, the profile dataset of N users who has semantically similar query in past history will be extracted. The profile attributes of the active user is also extracted in parallel as shown in Table 3. Naïve Bayes Probabilistic (NP) algorithm is then applied to cluster these profiles and assign AU into a cluster that contains users whose profiles are similar to AU. Algorithm 1 discusses the steps involved in NP approach. Finally, NP identifies nearest neighbors of AU.

**Table 3.** **Profile dataset of N users with similar search query given by AU**

| Features in User Profile | Weight (β) | User 1 Profile | User 2 Profile | User 3 Profile | … | User N Profile | Active User (AU) |
|---|---|---|---|---|---|---|---|
| UID | NA | 841 | 7895 | 87 | … | 785 | 7999 |
| TOP (In Sec) | 1.75 | 140 | 126 | 195 | … | 183 | 169 |
| TOS (in Sec) | 1 | 158 | 139 | 187 | … | 176 | 153 |
| ATP (in Sec) | 1 | 58 | 12 | 18 | … | 43 | 37 |
| BR (in %) | 1.75 | 0.58 | 0.49 | 0.68 | … | 0.61 | 0.54 |
| ER (in %) | 1.5 | 0.38 | 0.23 | 0.53 | … | 0.41 | 0.42 |
| CR (in %) | 1.75 | 0.0417 | 0.0256 | 0.0528 | … | 0.0394 | 0.0423 |
| NOV (in Nos) | 1.5 | 14 | 6 | 3 | … | 9 | 7 |
| APR (in Nos) | 2 | 6 | 1 | 3 | … | 4 | 8 |
| SK (in Nos) | 1.75 | 254 | 69 | 124 | … | 176 | 185 |
| ASM (in Nos) | 2 | 158 | 69 | 85 | … | 248 | 173 |
| Cluster ID | NA | 3 | 2 | 1 | … | 2 | x |

**Algorithm 1: NP algorithm for clustering**

1. Compute the probability of each cluster's occurrence within the profile dataset as
   P(cluster_x) = Number_of_cluster_x / N
2. Computing Probability Matrix
   For each attribute i = 1 to s
   For all user profiles p = 1 to N
   a. Compute the probability of attribute i contributing in classifying the profile p within cluster_x
   b. Populate the probability matrix (as shown in table 4, where three clusters are considered as an example)

3. Identifying the cluster_id for AU
   For each attribute i = 1 to s of AU's profile
   For all clusters c = 1 to M
   a. Compute the probability of mapping AU under each cluster x = 1 to M supported by attributes 1 to 10.

$$P(\text{cluster=x}|\text{AU}) = \prod_{i=1}^{s} P(\text{Attribute}_i|\text{ Cluster=x}) \times P(\text{cluster\_x}) \times \beta$$

   b. Assign AU to the cluster that has maximum probability
4. End Algorithm

**Table 4.Probability Matrix derived from profile dataset mapped to three clusters**

| P(Cluster) | Cluster_1 | Cluster_2 | Cluster_3 |
|---|---|---|---|
| P(TOP \| Cluster) | 0.36 | 0.25 | 0.39 |
| P(TOS \| Cluster) | 0.40 | 0.18 | 0.42 |
| P(ATP \| Cluster) | 0.45 | 0.25 | 0.30 |
| P(BR \| Cluster) | 0.25 | 0.39 | 0.36 |
| P(ER \| Cluster) | 0.42 | 0.35 | 0.23 |
| P(CR \| Cluster) | 0.36 | 0.32 | 0.32 |
| P(NOV \| Cluster) | 0.33 | 0.45 | 0.22 |
| P(APR \| Cluster) | 0.58 | 0.23 | 0.19 |
| P(SK \| Cluster) | 0.60 | 0.16 | 0.24 |
| P(ASM \| Cluster) | 0.48 | 0.35 | 0.17 |

## 4 OPTIMIZATION IN RECOMMENDATION SYSTEM

### *4.1 Ant Colony Optimization*

ANT Colony Optimization (ACO) is a probabilistic based model from the family of Swarm Intelligence. ACO employ meta-heuristic methods of optimizations to solve computational problems. It is based on real world phenomena followed by ants in search of its food. Initially ants move in a random manner during the search of their food. While such navigation, they eject a special substance called "pheromone" on their way to food and back to the nest.

Other ants following the initial set of ants will never move in random order, instead they follow based on the concentration of the pheromone ejected by ants those reached the food. Weihui Dai, et.al (2009), Pablo Loyola, et. al (2012) and Xiao-Feng Xie, et.al (2002) suggest that the concentration of pheromone ejected by the ants that migrates in all other directions opposite to the food will be gradually reduced Thus all ants are attracted in a path that optimally reaches the food and back to the nest.

The field of bio-inspired computing customizes the phenomena followed by the biological creatures. In this paper, we have implemented the feature of ACO for optimizing the accuracy and delay in prediction of web pages for recommending to the current AU. The functionality of applying ACO in web page recommendation is depicted in Algorithm 2.

Algorithm 2: Applying ACO for Optimizing Web Page Recommendation

1. Offline Process:
   a. Identify the web pages ('n') visited by 'k' nearest neighbors obtained through NP algorithm
   b. Identify the keywords in the web pages ('n') visited by 'k' nearest neighbors and keywords searched by AU using the user profiles
   c. Compute similarity between AU keywords and keywords in web pages
      For each AU keyword j = 1 to m
      For each webpage i= 1 to n
      For each keyword k=1 to si in each webpage i

$$\text{dij} = \sum_{j=1}^{m}\left[\sqrt{\sum_{k=1}^{si}\left(Key_j - Key_k\right)^2}\right]$$

   d. Update Pheromone value (P) as shown in Table 5
      For each query keyword j = 1 to m
      For each webpage i= 1 to n

$$\text{Pij} = Max\left(d1j, d2j, d3j, \ldots dij\right) - dij$$

2. Online Process:
   a. When the user inputs the search query with single/multiple words (m keywords)
   b. Select the web pages that has largest pij value for the corresponding keywords
   c. Recommend the selected web pages.
3. End Algorithm

**Table 5.** Pheromone Table for Recommendation

| Web Pages | Keyword_1 | Keyword_2 | … | Keyword_m |
|---|---|---|---|---|
| WebPage_1 | P11 | P12 | … | P1m |
| WebPage_2 | P21 | P22 | … | P2m |
| WebPage_3 | P31 | P32 | … | P3m |
| … | … | … | … | … |
| WebPage_n | Pn1 | Pn2 | … | Pnm |

### 4.2 *Particle Swarm Optimization*

PARTICLE Swarm Optimization (PSO) is another efficient optimization model from the family of Swarm Intelligence which adapts natural intelligence for computing. The collective behavior of self-organizing particles has been modified suitably for solving computational problems. PSO is also Meta heuristic algorithm, containing a set of algorithms which is used to define heuristic methods. PSO defines a Fitness Function (FF) which is also called as Objective Function which expresses the core functionality of research to be optimized. Renato A. Krohling (2004) suggests that, FF could be either in maximization or minimization phenomena. The core idea of PSO is to find an optimal solution for FF by searching within a population of potential solutions. PSO is initialized with a population of random solutions and it gradually searches for optimal one through various generations. PSO also defines a boundary for searching optimal solutions, termed as Search Space (SS). Two key aspects are involved in PSO while finding optimal solutions namely, Social Behaviour and Cognitive Behaviour. The Social Behaviour determines how particle behaves when compared globally (around search space) leading towards Global Best Solution (GBS). The Cognitive Behaviour determines how particle behaves among themselves (local group of particles) leading towards Local Best Solution (LBS). In each generation of PSO, new Velocity and Position of particles (candidate solutions) will be computed, which makes the generation reaching towards optimal solution.

The third contribution of this paper is to implement PSO for web page recommendation and to check whether this optimizes the performance when compared to traditional recommendations systems and ACO algorithm. The following Algorithm 3 describes the idea of implementing PSO during recommendation.

**Algorithm 3: Applying PSO for Optimizing Web Page Recommendation**

1. Identify the web pages ('n') visited by 'k' nearest neighbors obtained through NP algorithm
2. Identify the keywords in the web pages ('n') visited by 'k' nearest neighbors and keywords searched by AU using the user profiles
3. Initialize constants as ω = 0.3, c1=0.2, c2=0.2, r1 =1, r2 = 1, population = n (web pages), velocity for n particles = 0.
4. Assume each webpage as individual particle in the cluster search space. Initialize the position of particles as random values within 100.
5. Compute the Fitness Function (FF) as the following steps
   For each AU keyword j = 1 to m
   For each webpage i= 1 to n
   For each keyword k=1 to si in each webpage i

$$\text{Min.F(dij)} = \sum_{j=1}^{m}\left[\sum_{k=1}^{si}\left|Key_j - Key_k\right|^{si}\right]^{1/si}$$

6. Evaluate FF as stated in step 4 and check the value of FF that is best among "n" particles.
   a. If best found, stop the algorithm and go to step 9.
   b. Else, proceed with step 7
7. Update Velocity and Position of the particle (web page) as:
   Vi = ωVi-1 + c1r1(pbest – pi) + c2r2 (gbest – pi)
   Where
   Pbest is local best – the minimum FF among 'n' particles in the current iterations
   gbest is global best - the minimum FF among 'n' particles from first to the current iteration
   Pi = Pi-1 +Vi
8. With new position for "n" particles, move to step 5.
9. Recommend those particles (web pages) that are top best solutions among n particles (web pages)
10. End Algorithm.

## 5 EXPERIMENTS

### 5.1 Diabetes data set

EXPERIMENTS were conducted by employing diabetes patient records obtained from UCI Machine Learning Repository. Diabetes patient records were obtained from two sources: an automatic electronic recording device and paper records as suggested by Dua. D, et.al (2017). The automatic device had an internal clock to timestamp events, whereas the paper records only provided "logical time" slots (breakfast, lunch, dinner, bedtime). For paper records, fixed times were assigned to breakfast (08:00), lunch (12:00), dinner (18:00), and bedtime (22:00). Thus paper records have fictitious uniform recording times whereas electronic records have more realistic time stamps. Total records of 70 diabetic patients were considered for analysis. For each patient, around 350 readings were recorded for investigation. Diabetes files consist of four fields per record.

The schema of this dataset is: {Date, Time, Code, Value} as suggested by Dua. D, et.al, (2017). Where, Date in MM-DD-YYYY format, represents the date when test was performed to the corresponding patient. Time is represented in XX:YY format, Code contains the specific numeric code for the attributes.

The following are code, attribute pair used during experiments: {33, Regular insulin dose}, {34, NPH insulin dose}, {35, UltraLente insulin dose}, {48, Unspecified blood glucose measurement (Sample 1)}, {57, Unspecified blood glucose measurement (Sample 2)}, {58, Pre-breakfast blood glucose measurement}, {59, Post-breakfast blood glucose measurement}, {60, Pre-lunch blood glucose measurement}, {61, Post-lunch blood glucose measurement}, {62, Pre-supper blood glucose measurement}, {63,Post-supper blood glucose measurement}, {64, Pre-snack blood glucose measurement}, {65,Hypoglycemic symptoms}, {66,Typical meal ingestion}, {67, More-than-usual meal ingestion}, {68, Less-than-usual meal ingestion}, {69, Typical exercise activity}, {70,More-than-usual exercise activity}, {71, Less-than-usual exercise activity}, {72,Unspecified special event} (Dua. D, et.al, 2017). Value field stores the corresponding attribute measurement for each patient. A sample dataset is shown in Table 6 (Dua. D, et.al, 2017). For experimentation and analysis, the diabetic patient dataset is divided into eight samples of equal size with 50 records as mentioned in table 7.

**Table 6.Diabetes patient record (Sample) [Source: UCI Machine Learning Repository]**

| Date | Time | Code | Value |
|---|---|---|---|
| 04-21-1991 | 9:09 | 58 | 100 |
| 04-21-1991 | 9:09 | 33 | 9 |
| 04-21-1991 | 9:09 | 34 | 13 |
| 04-21-1991 | 17:08 | 62 | 119 |
| 04-21-1991 | 17:08 | 33 | 7 |
| 04-21-1991 | 22:51 | 48 | 123 |
| 04-22-1991 | 7:35 | 58 | 216 |
| 04-22-1991 | 7:35 | 33 | 10 |
| 04-22-1991 | 7:35 | 34 | 13 |
| 04-22-1991 | 13:40 | 33 | 2 |
| 04-22-1991 | 16:56 | 62 | 211 |
| 04-22-1991 | 16:56 | 33 | 7 |

**Table 7.Datasets sampled for experimentation and analysis**

| Sample Category | Description |
|---|---|
| Sample 1 | Without any conditions, 50 patients from diabetic log were selected randomly |
| Sample 2 | Uniform sampling was performed to select patients with Unspecified blood glucose measurement |
| Sample 3 | After clustering, 50 patients with Pre-breakfast blood glucose measurement greater than 170 was selected |
| Sample 4 | After clustering, 50 patients with Post-breakfast blood glucose measurement greater than 220 was selected |
| Sample 5 | Top 50 patients with maximum Post-breakfast blood glucose measurement were selected |
| Sample 6 | Least 50 patients with minimum Post-breakfast blood glucose measurement were selected |
| Sample 7 | Top 50 patients who perform More-than-usual exercise activity were selected |
| Sample 8 | Top 50 patients who perform Less-than-usual exercise activity were selected |

**Table 8.Contingency table used to compute Precision and Recall**

| Category | Description |
|---|---|
| True Positive (TP) | The web pages that are recommended were relevant |
| False Positive (FP) | The web pages that are recommended were irrelevant |
| True Negative (TN) | The web pages that are not recommended were irrelevant |
| False Negative (FN) | The web pages that are not recommended are relevant |

### 5.2 Evaluation Metrics

METRICS such as F1-Measure, Miss-Rate (MR), Fallout Rate (FR) and Matthews Correlation were used to analyze the performance of proposed algorithm as suggested by Schröder, G, et.al, (2011). Contingency table as shown in table 8 is used to compute Precision, Recall, Miss Rate, Fallout Rate and Matthews Correlation.

$$\text{Precision}=\frac{TP}{TP+FP} \tag{1}$$

$$\text{Recall}=\frac{TP}{TP+FN} \tag{2}$$

$$\text{F1}=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{3}$$

$$\text{Miss Rate}=\frac{FN}{TP+FN} \tag{4}$$

$$\text{Fallout Rate}=\frac{FP}{FP+TN} \tag{5}$$

$$\text{Matthews Correlation}=\frac{(TP\times TN)-(FP\times FN)}{\sqrt{(TP+FN)\times(FP+TN)\times(TP+FP)\times(FN+TN)}} \tag{6}$$

### 5.3 Results and Inferences

Experiments were conducted using the seven samples of dataset running under three algorithms Collaborative Filtering (CF), Naïve Bayes Probability with ACO (NP_ACO), Naïve Bayes Probability with PSO (NP_PSO). The graphs that measure F1-Measure, Miss Rate (MR), Fallout Rate (FR) and Matthews Correlation (MC) were shown in figure 2, figure 3, figure 4 and figure 5 respectively. The results clearly depicts that the proposed Naïve Bayes Probabilistic Model with Particle Swarm Optimization has shown improved F1-measure, hence the accuracy is highly maintained. Meanwhile the Miss Rate has been considerably reduced when compared to Naïve Bayes with Ant Colony Optimization technique and traditional collaborative approach in all data samples. The proposed algorithm also shows improvement in classification accuracy, hence while testing under all data samples Matthews Correlation was improved much better for proposed optimization based machine learning classification approach.

Inferences observed from the evaluation metrics suggests that accuracy of NP_PSO algorithm shows 9.96% improvement when compared to the accuracy of NP_ACO algorithm and 12.93% improvement when compared with the accuracy obtained by CF algorithm. On an average, the Miss Rate and Fallout rates of NP_PSO algorithm have been considerably reduced by 5.43% and 5.88% respectively when compared to other algorithms. When analyzing the Matthews Correlation, the proposed NP_PSO algorithm has drastic improvement with 15.54% when comparing with NP_ACO and CF algorithms.

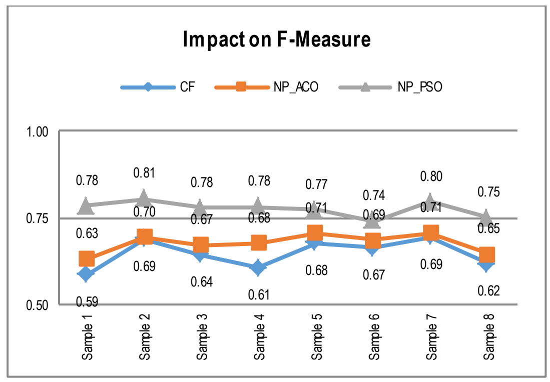


**Figure 2.** Analyzing F1-measure tested with various sample datasets.

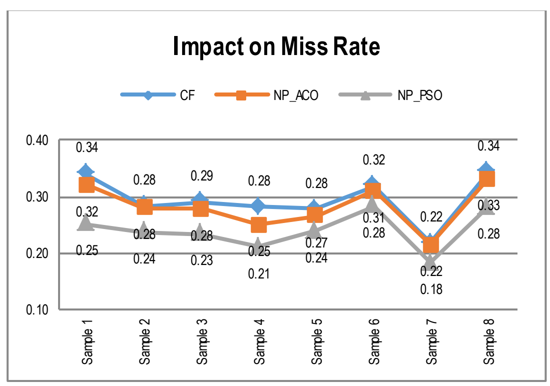


**Figure 3.** Analyzing Miss Rate tested with various sample datasets.

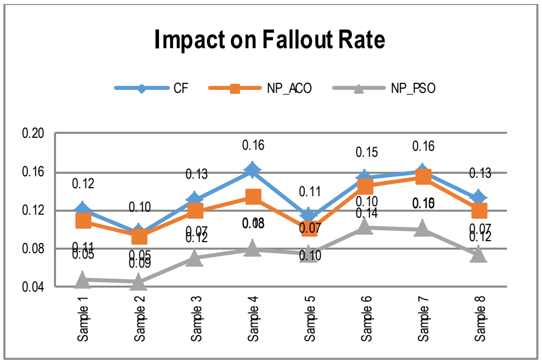


**Figure 4.**Analyzing Fallout Rate tested with various sample datasets.

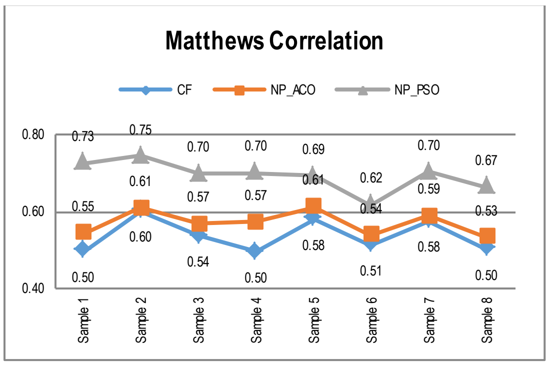


**Figure 5.**Analyzing Matthews Correlation tested with various sample datasets.

## 6 CONCLUSION

IN this paper, a novel approach for improving personalization in recommendation system was proposed. The paper investigates the efficiency of the proposed algorithm in web page recommendation and nutrition recommendation systems. User profiles for web based users and diabetic patient comprising of specific attributes were recorded. Naive Bayes Probabilistic approach was applied to cluster the profiles based on best match among the profile attributes. In addition, the effectiveness of applying optimization algorithms for personalized recommendation was also analyzed in this paper. The fitness function of Ant Colony and Particle Swarm Optimization techniques were adapted to find optimal suggestion for end user. The effectiveness of proposed algorithms was theoretically investigated using web page recommendation system. Experiments were conducted for recommending nutrition for diabetic patients with eight categories of test samples. Results infer that, the Particle Swarm Optimization outperforms when compared to traditional collaborative filtering approach and ACO algorithms with improved F1-Measure. The Miss Rate and Fallout Rates were also found to be decreased, hence enhancing the accuracy. Matthews Correlation value is found to be improved while applying PSO based optimization rather than ACO. To further enhance the accuracy and effectiveness, hybrid optimization techniques could be applied based on the nature of user profiles.

## 7 REFERENCES


Abirami, S., Bhavithra, J., Saradha, A. (2017), "A Web Page Recommendation using Naive Bayes Algorithm in Hybrid Approach", *IEEE International Conference on Science, Technology, Engineering and Management*.

Aditi Shrivastava, NitinShukla(2012), "Extracting Knowledge from User Access Logs", *International Journal ofScientific and Research Publications*, 2(4), 1-4.

Alexandrin Popescul Lyle H. Ungar, David M. Pennock, Steve Lawrence, (2001), "Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments", *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, 437-444.

American Diabetes Association (2008), "Nutrition Recommendations & Interventions for Diabetes: A position statement of the American Diabetes Association", *Diabetes Care*, 31(1), 61-78.

Amit Vishwakarma, Kedar Nath Singh, (2014), "A Survey on Web Log Mining Pattern Discovery", *International Journal of Computer Science and Information Technologies*, 5 (6), 7022-7031.

Bhavithra, J. & Saradha, A. (2018), "Personalized web page recommendation using case-based clustering and weighted association rule mining", *Cluster Computing*, *Springer*, https://doi.org/10.1007/s10586-018-2053-y,1-12.

Byron Bezerra and Francisco de A. T. E Carvalho, (2004), "A Symbolic Hybrid Approach to Face the New User Problem in Recommender Systems", *Springer Verlag Berlin Heidelberg*, 1011–1016.

D. Ciobanu, C. E. Dinuca (2012), "Predicting the next page that will be visited by a web surfer using Page Rank algorithm", *International Journal of Computers And Communications*, 6(1), 60-67.

Daxa k. Patel (2014), "A Retrieval Strategy for Case-Based Reasoning using USIMSCAR for Hierarchical Case", *International Journal of Advanced Engineering Research and Technology*, 2(2), 65-69.

Deepa S.N., G. Sugumaran. (2011), "Model order formulation of a multivariable discrete system using a modified particle swarm optimization approach", *Elsevier: Swarm and Evolutionary Computation*, 1(4), 404-212.

Dimitrios Pierrakos, Georgios Paliouras, "Personalizing Web Directories with the Aid of Web Usage Data", *IEEE Transactions on*

*Knowledge and Data Engineering*, 22(9), 1331-1344.

Dua, D. and Karra Taniskidou, E. (2017). UCI Machine Learning Repository. Retrieved from http://archive.ics.uci.edu/ml].

Freddy L´ecu´e, (2010), "Combining Collaborative Filtering and Semantic Content-based Approaches to Recommend Web Services", *IEEE Fourth International Conference on Semantic Computing*, 200-205.

G. Schröder, M. Thiele, and W. Lehner (2011), "Setting Goals and Choosing Metrics for Recommender System Evaluations," in *Proceedings of the Second Workshop on User-Centric Evaluation of Recommender Systems and their Interfaces*.

Gediminas Adomavicius, Alexander Tuzhilin, (2005), "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734-749.

J. Bobadilla, F. Ortega, A. Hernando, A. Gutiérrez (2013),"Recommender Systems Survey",*Elsevier, Universidad Politécnica de Madrid, Ctra. De Valencia, Spain*, pp.109-132.

John Z. Sun, Dhruv Parthasarathy, and Kush R. Varshney (2014), "Collaborative Kalman Filtering for Dynamic Matrix Factorization", *IEEE Transactions on Signal Processing*, 62(14), 3499-3509.

Jonathan L. Herlocker and Joseph A. Konstan, Loren G. Terveen, and John T. Riedl, (2004), "Evaluating Collaborative Filtering Recommender Systems", *ACM Transactions on Information Systems*, 22(1), 5-53.

Katja Niemann and Martin Wolpers (2015), "Creating Usage Context-Based Object Similarities to Boost Recommender Systems in Technology Enhanced Learning", *IEEE Transactions on Learning Technologies*, 8(3), 274-285.

Kazuyoshi Yoshii, Masataka Goto, Kazunori Komatani, Tetsuya Ogata, Hiroshi G.O kuno, (2006), "Hybrid Collaborative and Content-based Music Recommendation Using Probabilistic Model with Latent User Preferences", *7th International Conference on Music Information Retrieval*, 296-301.

Kebin Wang and Ying Tan, (2011), "A New Collaborative Filtering Recommendation Approach Based on Naïve Bayesian Method", *International Conference in Swarm Intelligence, Springer Advances in Swarm Intelligence*, 218–227.

Lina Yao, Quan Z. Sheng, Member, IEEE, Anne. H.H. Ngu, Jian Yu, and Aviv Segev (2015), "Unified Collaborative and Content-Based Web Service Recommendation", *IEEE Transactions On Services Computing*, 8(3), 453-466.

Marion J Franz, Jackie L Boucher, Alison B Evert (2014), "Evidence-Based Diabetes Nutrition Therapy Recommendations are Effective: The Keyis Individualization", *Dovepress: Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, 7, 65-72

Meghna Khatri, (2012), "A Survey of Naïve Bayesian Algorithms for Similarity in Recommendation Systems", *International Journal of Advanced Research in Computer Science and Software Engineering*, 2, 217-219.

Michael, G. Noll. (2006), "AOL Research Publishes 650,000 User Queries", Retrieved from:URL:http://www.michael-noll.com/blog/2006/08/07/aol-research-publishes-500k-user-queries/

Mingming Jiang, Dandan Song, Lejianliao, Feida Zhu, (2015), "A Bayesian Recommender Model for User Rating and Review Profiling", *Tsinghua Science and Technology*, 20(6), 634-643.

Murat Goksedef, ŞuleGunduz-Oguducu (2010), "Combination of Web page recommender systems", *Elsevier Journal on Expert Systems with Applications*, 37(4), 2911–2922.

Mustansar Ali Ghazanfar and Adam Pru¨gel-Bennett, (2010), "An Improved Switching Hybrid Recommender System Using Naïve Bayes Classifier and Collaborative Filtering", *Proceedings of the International Multiconference of Engineers and Computer Scientists*, 1, 1-10.

Mustansar Ali Ghazanfar and Adam Prugel-Bennett (2010). "A Scalable, Accurate Hybrid Recommender System",,*IEEE Third International Conference on Knowledge Discovery and Data Mining*, 94-98.

Pablo Loyola, Pablo E.Roma, JuanD.Vela´squez. (2012), "Predicting Web User Behavior Using Learning-Based Ant Colony Optimization", *Elsevier Engineering Applications of Artificial Intelligence*, 25(5), 889-897.

Pearl Pu, Li Chen,Rong Hu (2012), "Evaluating Recommender Systems from the User's Perspective: Survey of the State of the Art", *Springer Science*, 22, 317–355.

Renato A. Krohling. (2004), "Gaussian Swarm: A Novel Particle Swarm Optimization Algorithm", *IEEE Conference on Cybernetics and Intelligent Systems*, 372-376.

Schröder, G., Thiele, M., Lehner, W. (2011), "Setting Goals and Choosing Metrics for Recommender System Evaluations", *Second Workshop on User-Centric Evaluation of Recommender Systems and Their Interfaces*.

Şule Gunduz - Oguducu, M. Tamer Ozsu, (2006), "Incremental Click-Stream Tree Model: Learning from New Users for Web Page Prediction", *Distributed Parallel Databases, Springer Science*, 19, 5–27.

Thi Thanh Sang Nguyen, Hai Yan Lu, and Jie Lu (2014), "Web-Page Recommendation Based on Web Usage and Domain Knowledge", *IEEE Transactions on Knowledge and Data Engineering*, 26(10), 2574-2587.

Tranos Zuva, Sunday O. Ojo, Seleman M. Ngwira and Keneilwe Zuva (2012), "A Survey of Recommender Systems Techniques, Challenges and Evaluation Metrics", *International Journal of Emerging Technology and Advanced Engineering*, 2(11), 382-386.

Weihui Dai, Shouji Liu and Shuyi Liang (2009), "An Improved Ant Colony Optimization Cluster Algorithm Based on Swarm Intelligence", *Journal of Software*, 4(4), 299-306.

Xi Chen, Xudong Liu, Zicheng Huang, and Hailong Sun, (2010), "A Scalable Hybrid Collaborative Filtering Algorithm for Personalized Web Service Recommendation", *IEEE International Conference on Web Services*, 9-16.

Xiao-Feng Xie, Wen-Jun Zhang, Zhi-Lian Yang. (2002), "A Dissipative Particle Swarm Optimization", *IEEE Congress and Evolutionary Computation*, 1456-1461.

Zheng Lu, Hongyuan Zha, Xiao kang Yang, Weiyao Lin, ZhaohuiZheng (2013), "A New Algorithm for Inferring User Search Goals with Feedback Sessions", *IEEE Transactions on Knowledge and Data Engineering*, 25(3), 502-513.

Zibin Zheng, Hao Ma, Michael R. Lyu, Fellow, and Irwin King, (2011), "QoS-Aware Web Service Recommendation by Collaborative Filtering", *IEEE Transactions on Service Computing*, 4(2), 140–152.

## 8 NOTES ON CONTRIBUTORS

**Ms. J. Bhavithra**, completed her Bachelor of Engineering during 2003 and Master of Engineering during 2005, specialized in Computer Science and Engineering. She is currently pursuing research under Information and Communication Engineering specialized in Web Page Recommendation systems and Machine Learning Techniques. She is currently working as Assistant Professor (Senior Scale) in the department of Computer Science and Engineering at Dr. Mahalingam College of Engineering and Technology, Pollachi, Coimbatore (DT), Tamilnadu, India.

**Dr. A. Saradha**completed her doctorate in Information and Communication Engineering specialized in the field of Pattern Recognition. She is currently working as Professor and Head in the Department of Computer Science and Engineering, Institute of Road and Transport Technology, Erode, Tamilnadu, India. She has more than 26 years of teaching and research experience. Her research interests are Web Mining, Human Computer Interaction and Pattern Recognition