DOI: 10.21917/ijsc.2020.0298

DIET RECOMMENDATION FOR GLYCEMIC PATIENTS USING IMPROVED K-MEANS AND KRILL-HERD OPTIMIZATION

K. Renuka Devi, J. Bhavithra and A. Saradha
Department of Computer Science and Engineering, Dr. Mahalingam College of Engineering and Technology, India

Abstract

Maintaining nutrition for glycemic (diabetic) patients in order to retain the blood glucose level is one of the important activity to be followed. Stimulating the amount of carbohydrates, protein, vitamins, and minerals will result in a healthy diet. So, there is a necessity for recommendation of nutrition to those diabetic patients nowadays. Recommender Systems (RS) play a vital role in urging relevant suggestions to the users. To promote improvised and optimized results, Optimization technique plays a significant role in refining the parameters of chosen algorithm. To optimize and to upgrade the accuracy of recommendations, the system has been developed by implementing improved Krill-Herd algorithm. The system which clusters the profiles of diabetic patients using improved k-means clustering algorithm and results has been optimized using Improved Krill-Herd optimization algorithm. The performance will be analysed using different measures like Precision, Recall, F-measure, Accuracy, Matthews correlation, Fallout rate and Miss rate. The evaluation results show that the proposed system performs better and produces optimized results to the diabetic patients with minimum error rate.

Keywords:

Data Mining, Diabetes Patients, Recommender Systems, Clustering Algorithm, Improved K-Means, Krill Herd Optimization

1. INTRODUCTION

Data mining is the process of examining large set of databases to produce the appropriate information. In the context of recommender systems, data mining plays a vital role in making recommendations based on the analysis from large datasets [1].

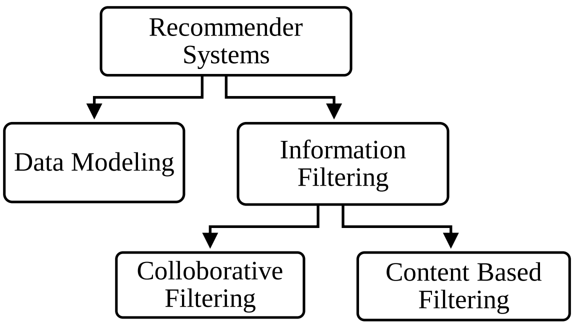

Fig.1. Categories of Recommender Systems

Recommender systems incorporate data mining techniques to make recommendations based on the knowledge gained from attributes of user such as their profiles, ephemeral data, etc. The algorithms include clustering, classification techniques, association rule mining, etc. The recommender systems are broadly classified into data modeling and information filtering system. The overview of recommender systems is shown in Fig.1.

1.1 DATA MODELING

To make predictions about user's interests, the system has to learn the user model [2]. The model contains the user's data and it is represented to make the match with the items that are contained in the collection [2]. The data modeling utilizes the past information of the users to model user profiles.

1.2 INFORMATION FILTERING

Most Recommender Systems make use of information filtering systems which focus on providing relevant items or information that the users are mostly interested or find useful [3]. These systems are widely used in e-commerce, healthcare industry etc. The information filtering basically contains two approaches namely collaborative filtering and content-based filtering.

1.3 COLLABORATIVE FILTERING

To recommend appropriate items to the users, the recommender systems widely utilize the concept of collaborative filtering. This method mainly relies on collecting and processing the user's information and predicts the items based on the similarity of other users. Collaborative filtering system is used to construct a structure based on the similarity of other users [1] [11]. Since the system is independent of contents that could be processed by machines, it is able to recommend the items without the knowledge of the items in the collection.

1.4 CONTENT BASED FILTERING

Content based filtering is another approach of information filtering system which is used to recommend the appropriate items to the users based on the user's preferred profile. This type of filtering system is mainly based on the item called keywords which are considered as the user's interest [2] [12]. It is used to obtain the items of similar properties and make recommendations based on the items the user liked in the past or likes in the present. In nutshell, it is used to recommend the items to the user based on the ratings given by the user in the past which is compared with the items in the collection and most appropriate items are recommended.

1.5 NEED FOR NUTRITION RECOMMENDATION

Maintaining nutrition for diabetic patients is one of the vital activities in preserving the levels of glucose in blood. Stabilizing the amount of carbohydrates, protein, fat with vitamins, minerals will promote a better diet. For those with diabetes, their nutritional needs are one of the utmost concern and hence are given extra attention. The diabetic patients are prone to heart disease than normal persons and also they suffer from depression [2] [13]. Type - 2 diabetes are kind of preventable and also they can be reversed. The balanced diet will boost up the energy and also improves mental ability. When there is a need to reduce glucose level in blood, the intake of carbohydrates is a vital factor which must be taken into consideration. In the context of maintaining a healthy diet, the reduction of weight also plays a major role. Losing 5% to 10% of weight can lower the blood sugar, blood pressure and cholesterol levels. Losing as well as reducing weight and having healthier diet results in reducing the blood glucose levels to lead healthy lifestyle.

2. RELATED WORK

Recommender systems are a vibrant class of information filtering system that consists of software tools and methodologies which provide suggestions to the user according to their need. Many popular Ecommerce sites widely use Recommender systems to recommend news, music, research articles, books, and product items. These systems utilize personal data, implicit values and local information from the Internet. With vibrant and rapidly growing web, website complexity is constantly increasing, making it more difficult for users to quickly locate the information based on their requirement. To overcome the problem of low efficiency and inaccurate recommendations while using collaborative filtering recommendation algorithms, Li [3] developed an improved hybrid collaborative filtering recommendation algorithm which is based on K-means [3] [14]. Data mining has a variety of algorithms, generally based on the characteristics of the data, classification of the data or finds the correlation between the data to be classified, and when using the collaborative filtering recommendation algorithm, it is based on the relationship between the users to find the user's nearest neighbor. This technique is used to reduce the amount of computation and to improve the efficiency of results.

Cluster categorical datasets is a method based on K-modes which has been proposed by Salema and Naouali [4]. Manhattan Frequency k-Means (MFk-M), a new categorical method which is based on partitions used to cluster the categorical attributes. It mainly relies on converting the categorical datasets into numeric values which utilizes the relative frequency from those attributes. To calculate the distance between the observations and centroids, the Manhattan distance and Euclidean distance were used as an evaluation measure [4] [15]. Euclidean distance is used to provide highly efficient and scalable results for numerical clustering and for the conversion process.

To improve the recommendation performance of an algorithm, Salema and Naouali [4] developed a new collaborative filtering recommendation algorithm based on dimensionality reduction and techniques based on clustering [4] [15]. The k-means algorithm and Singular Value Decomposition (SVD) are used to cluster the similar kind of users and to reduce the dimensionality between those users. The utilization of incremental learning will result in a sparsity problem.

Collaborative Filtering (CF) is a popularly known and widely used approach in recommender systems. CF is used to predict the interest of an individual based on the interests of other individuals [5]. Neighborhood-based Collaborative Filtering (NBCF) aims to identify the similarity between users (or) items which was given by Hu et al. [5]. Due to the increase in the number of users as well as the items, and because of high dimensional complexity such kinds of NBCF methods become impractical [5] [16]. The recommendation has been made by two step process; the clustering has been carried out in offline and top-N ranking done in online which is used to improve the scalability without reducing the quality of recommendations.

Clustering is a data mining method to divide the data objects into N number of clusters. Clustering algorithms are utilized in variety of domains such as online purchasing, bio-informatics etc., Improved k-means proposed by Khandare and Alvi [6] minimizes the gap between enhancement and scalability of algorithms [6] [17] [20] [23]. The proposed method selects the number of clusters and initial means that have further scope for improvements. This new method also calculates the initial centroids using arithmetic mean method instead of random selection. Due to this, the quality and number of iterations required for creating final clusters has improved to a great extent. This modified k-means algorithm has systematic and efficient method for finding initial centroids and allocating data points to clusters generated.

Krill-Herd (KH) is an algorithm of random and is inspired by nature for optimization which is used to solve multiple complex optimization problems. KH algorithm with harmony search has been proposed to improve the diversification search ability. Krill-Herd algorithm (KHA) proposed by Abualigah et al. [7] is based on the behaviour of krill individuals [7] [18] [21]. Krill herd gets tripped in the local search or premature convergence in the global search because of poor exploration.

To improve the accuracy of recommended items, Improved Krill-Herd algorithm has been developed. A clustering method for text along with a hybrid function, called MMKHA is proposed by Abualigah et al. [8]. This acts as an efficient method for clustering to obtain accurate results for the corresponding problems [8] [19] [22] [24]. Krill-Herd is a new swarm based optimization technique that simulates the behavior of a group of individual krills. It performs better than other optimization algorithms because it has higher potential than other methods. Moreover, it balances the process of exploration and exploitation searching techniques. Improved Krill-Herd performs robust search to handle high dimensional problems.

3. SYSTEM ARCHITECTURE

The diabetes data from the dataset are collected and pre-processed and then, the appropriate glucose for individual diabetic patients are extracted. Profile for each diabetic patient is generated. Users are clustered based on each patient's features such as age, height, blood glucose levels etc., by using Improved K-means clustering technique. On an active user query, Improved Krill-Herd based similarity is calculated between the active user query and other patient present in the cluster. Finally, the optimized results (nutrition) are recommended to the active user. The proposed system follows the process of collecting diabetes data from the dataset, processing them followed by clustering based on improved k-means algorithm and finally generating recommendations based on improved krill-herd algorithm. The proposed system is shown in Fig.2

3.1 PREPROCESSING OF DATA

The distinct profile for each diabetic patient is generated based on the blood glucose level which is fetched from the past history. The user profiles consist of thirty features, the first ten features specify the generic characteristic of a patient. The following twenty features are mostly associated with blood glucose level. Based on the similarity or equivalence among the individual profiles, the best nutrition therapy encountered by the closest (best match) and absolute patient (profile) will be recommended for each active user.

In order to analyze the correctness and effectiveness of the proposed recommendation techniques, various test cases are governed.

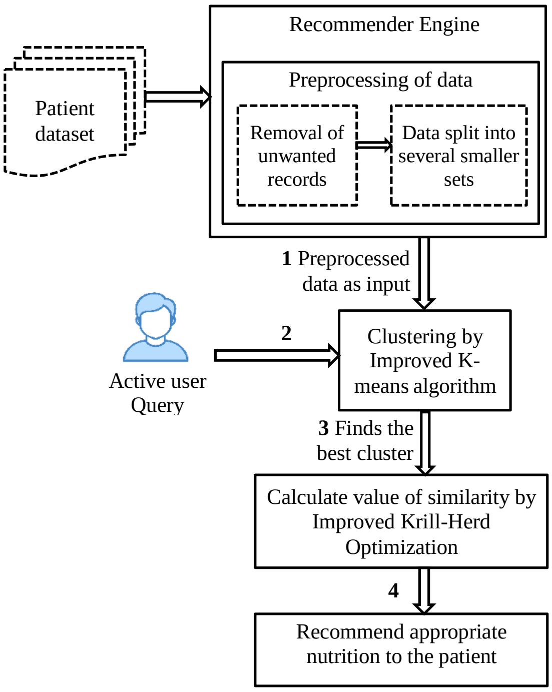

Fig.2. System Architecture

Experiments were conducted by analyzing patient records acquired from UCI Machine Learning Repository.

Step 1: The User's data from the dataset is uploaded.

a. The data which includes the following details {date, time, code, value}.

b. They are cleaned before pre-processing them through the following steps.

Step 2: Change code with feature name.

Step 3: Store each code in a file.

Step 4: Extract the recent glucose level.

From the former process, segregated user profiles are generated.

3.2 IMPROVED K-MEANS CLUSTERING

This process involves clustering deals with the concept of combining the similar entities into a single group. The objective of this unsupervised learning methodology is to analyze and discover the similarity between data points and to combine the same data points into a single group. The process of combining will help in profiling the variety of attributes belonging to different groups. This results in the discovery of new patterns which belong to different groups. The concept of clustering is used to minimize the data dimensionality when the user has to process ample number of attributes.

Improved K-means fall under the category of Point Assignment Algorithms which is an unsupervised learning algorithm. This algorithm finds the number of clusters dynamically rather than initializing statically. It also calculates the initial partitions (centroids) in a predominant way rather than random selection. The algorithm results in minimizing the number of iterations. Thus reduction in number of iterations results in the increase of cluster quality and decrease in the number of empty clusters.

The pre-processed data from data set contains the patient profiles that show the blood glucose levels. This data is the input for the Improved K-means algorithm for clustering. The algorithm consists of three phases. The first phase of algorithm is to find the *K* value (i.e.) the number of clusters dynamically. The second phase of algorithm is to calculate the initial partition where it divides the data points into sub-arrays. The third phase is to generate the final clusters. The distance between data objects is computed and checked against the initial centroid. If the distance is lesser or equal, then the data point remains in the same cluster. Otherwise, it is migrated to the appropriate cluster. This process is repeated until there is no change in the cluster mean.

3.3 OPTIMIZATION THROUGH IMPROVED KRILL HERD ALGORITHM

Improved Krill-Herd (KH) is an algorithm which is inspired by nature as it behaves in the nature of movement exhibited by krill individuals. It is noted as an efficient algorithm as it is used to solve several optimization problems based on different contexts that provide better solutions. It is known as swarm intelligence based search algorithm as it imitates the herding behavior of krill individuals. The Improved Krill-Herd algorithm consists of an objective function which is computed by the minimum distance of each krill from its food and the density of the krill herd.

Generally, to model the behavior of this algorithm, the following three operational processes are considered,

- Motion induced by other krill individuals.
- Foraging movement.
- Random physical diffusion.

The Improved Krill-Herd algorithm is considered as a robust search methodology because it contains both exploration and exploitation methods and the movement propagated by other krills.

When the active user enters the system, the appropriate active user's cluster is identified. In order to maintain the glucose level, the active user's features are compared with the neighbors present within the cluster. The objective function of Improved Krill-Herd picks the appropriate user for each feature using Eq.(1) [8].

$$\alpha_i^{\text{target}} = c^{best}\hat{k}_{i,best}\hat{x}_{i,best} \tag{1}$$

The nutrition is recommended by comparing the selected user's chemical composition with the default value.

4. RESULTS AND DISCUSSIONS

4.1 DATASET

Diabetes dataset is collected from UCI Machine Learning Repository [9], which has datasets covering several weeks to months' worth of outpatient care on 150 patients. The dataset includes Date, Time, Code and Value. The data is collected and pre-processed based on the value of each patient. The recent glucose level of each feature is extracted. For recommendation of nutrition, the data set is collected from USDA National Nutrient Database [10], wherein each record consists of nutrition along with the chemical composition for the nutrition. The data includes the following details {date, time, code, value}

- The date is in the format MM-DD-YYYY.
- The time corresponds to XX.YY format.
- Code denotes numeric value assigned to each insulin measurement.
- Value denotes the corresponding glucose level for a particular patient.

4.2 EXPERIMENT

The experimental result of the proposed system is compared with the existing system which utilizes K-means algorithm for clustering and threshold ranking algorithm for recommendation. The results obtained show that the proposed system performs better in terms of producing clusters and in making recommendations. The proposed approach utilizes optimization technique to produce more accurate results. The proposed system utilizes Improved K-means algorithm for clustering and Improved Krill-Herd algorithm for optimization. The proposed system utilizes less time for making recommendations and also produces more accurate results to the users.

5. PERFORMANCE EVALUATION

The quality and performance behavior of recommendation system is evaluated using several evaluation metrics. Through different kinds of test cases such as users, a comparison is made between the threshold ranking algorithm and Improved Krill-Herd optimization algorithm. The performance measures like precision, recall, F-measure and accuracy are found to have higher values for the proposed system than the existing system. The performance and accuracy are calculated using the following factors and metrics shown in Table.1.

Table.1. Evaluation factors

Factors	Description
True Positive (TP)	The number of correctly recommended nutrition
True Negative (TN)	The number of correctly not recommended nutrition
False Positive (FP)	The number of wrongly recommended nutrition
False Negative (FN)	The number of missed correct nutrition

Precision: The ratio of number of correctly recommended nutrition to the sum of number of correctly recommended nutrition and number of wrongly recommended nutrition from Eq.(2) [15].

$$Precision = TP/(TP+FP) \quad (2)$$

Recall: The ratio of the nutrition recommended and accessed by the user to the sum of the number of missed correct nutrition and number of correctly recommended nutrition from Eq.(3) [15].

$$Recall = TP/(TP+FN) \quad (3)$$

F-Measure: The harmonic mean of recall and precision. This is approximately the average of two when they are close together. F-Score will be high only when the precision and recall is high from Eq.(4) [15].

$$F\text{-}measure = 2(Precision * Recall)/(Precision + Recall) \quad (4)$$

Accuracy: The ratio of sum of correctly recommended nutrition to the sum of all the possibilities from Eq.(5) [15].

$$Accuracy = (TP+TN)/(TP+TN+FP+FN) \quad (5)$$

Matthews Correlation (MC): It is the combination of informedness and markedness from which single metric has been calculated using the Eq.(6) [15].

$$MC = \frac{(TP*TN)-(FP*FN)}{(TP+FN)*(FP+FN)*(TP+FP)*(FN+TN)} \quad (6)$$

Fallout rate: Fallout or false positive rate is measured as the ratio of prescribed nutrition that are not relevant to the total number of insignificant nutrition from Eq.(7) [15].

$$Fallout\ rate = FP/(FP+TN) \quad (7)$$

Miss rate: Probability of not being prescribed an appropriate diet. Miss rate or false negative rate is measured as the ratio of items not recommended but actually applicable to the total relevant nutrition Eq.(8) [15].

$$Miss\ rate = FN/(TP+FN) \quad (8)$$

5.1 RESULTS

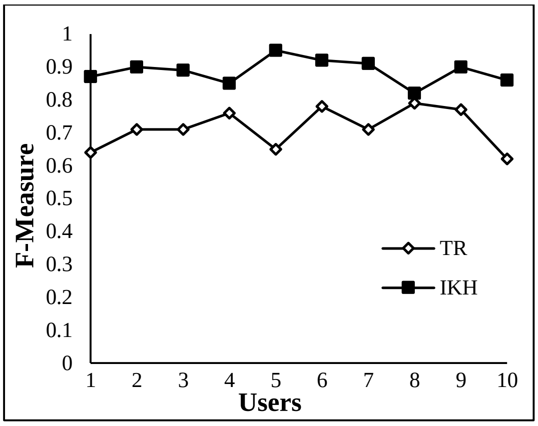

Fig.3. Comparison of F-measure values

Based on the experiment that was carried out for 10 different users according to their attributes such as height, weight etc., the F-measure of the proposed system is inferred to be higher than the existing system as the values of precision and recall are found to be higher, which is revealed by Fig.3. It is due to the fact that it recommends more number of relevant nutrition to the user. As a result, the performance metrics such as precision and recall have attained higher value.

The graph of Accuracy for threshold ranking algorithm and Improved Krill-Herd algorithm is shown in the Fig.4. Since the Accuracy depends on all the four evaluation factors, higher the value results in better performance. From Fig.4, it is inferred that Improved Krill-Herd has higher accuracy value than threshold ranking algorithm.

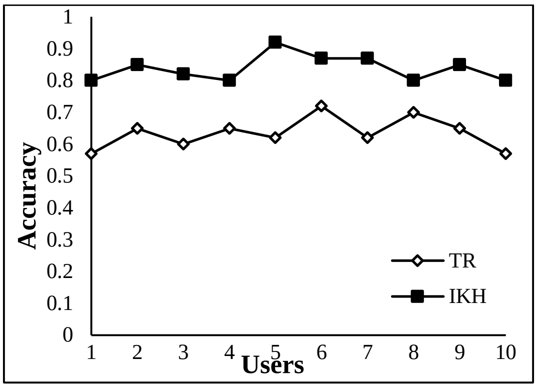

Fig.4. Comparison of Accuracy values

The graph of Matthews Correlation (MC), and the graph of accuracy for threshold ranking algorithm and improved Krill-Herd algorithm is shown in Fig.5. Since, MC relies on informed-ness and markedness, the result should be higher for positive and negative reliability. From the graph it is inferred that MC values are higher for Improved Krill-Herd than threshold ranking algorithm.

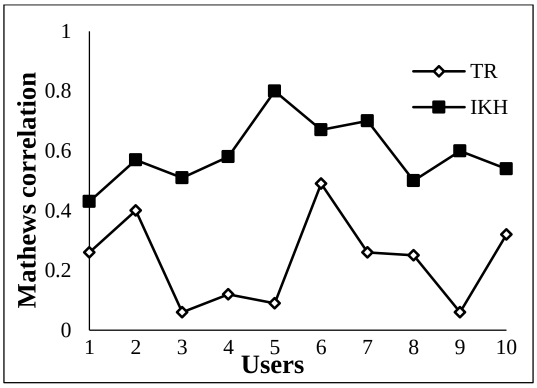

Fig.5. Comparison of Matthews correlation values

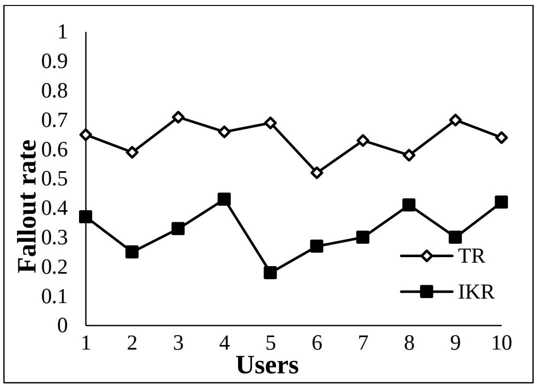

Fig.6. Comparison of Fallout rate values

The Fig.6 and Fig.7 show the comparison of fallout rate and miss rate for the Threshold ranking and Improved Krill-Herd algorithm. From the graph it is inferred that the error rates such as fallout and miss rates are identified to be maximum for the threshold ranking algorithm, whereas, it is minimum for improved krill-herd algorithm. This shows that the Improved Krill - Herd system has less false recommendations than the Threshold Ranking system.

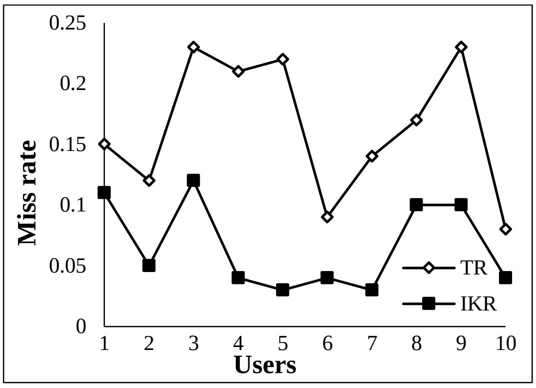

Fig.7. Comparison of Miss rate values

5.2 INFERENCES OBSERVED

The experimental results of the proposed Improved Krill - Herd based system is compared with the threshold ranking of the existing system. The results obtained show that the proposed system performs better in terms of time complexity and in producing clusters. The proposed system focuses on making recommendation of nutrition to the diabetic patients. The proposed system is an improved version of krill herd algorithm for providing better recommendations to the users. The experiment results inferred the following aspects for the system which utilizes Improved Krill-Herd algorithm.

- 17% of improvement in F measure
- 20% of improvement in Accuracy
- 36% of improvement in Matthews correlation
- 31% of reduction in Fallout Rate
- 10% of reduction in Miss Rate
- 5% of reduction in RMSE

The increase in F-measure values is due to the increase in precision values from Eq.(4). The increase in Accuracy values is due to decrease in True negative values from Eq.(5). The increase in Matthews correlation values is due to the fact of eliminating the false values from Eq.(6). The decrease in false positive and true negative values from Eq.(7) results in the minimization of Fallout rate value. The decrease in false negative values from Eq.(8) results in the reduction of miss rate values.

6. CONCLUSION

This paper focuses on describing nutrition recommendation system for diabetic patients by utilizing Improved K-means for clustering the data items and Improved Krill-Herd algorithm for optimization in order to deliver more accurate results to the users. The quality and performance of the system is evaluated using various measures like Precision, Recall, F-measure, Accuracy, Matthews correlation, Fallout rate and Miss rate. The results convey that the F-measure, Accuracy and Matthews correlation values are higher for the proposed system, whereas Fallout rate and miss rate are found to be lower for the proposed system. This shows that the recommendations made by Improved Krill-Herd algorithm give accurate results and also perform better than the threshold ranking algorithm. The values of the real time data sets are to be utilized in the future aspects.

REFERENCES

[1] Y. Koren and R.M. Bell, "*Advances in Collaborative Filtering*", Springer, 2011.

[2] Kunal Shah, "Unsupervised Recommender Systems: An Overview of Different Approaches to Recommendations", *Proceedings of IEEE International Conference on Innovations in Information, Embedded and Communication Systems*, pp. 1-4, 2017.

[3] Xiaofeng Li, "Improved Hybrid Collaborative Filtering Algorithm Based on K-Means", *Expert Systems with Applications*, Vol. 842, pp. 928-934, 2019.

[4] Semeh Ben Salema and Sami Naouali, "A Fast and Effective Partitional Clustering Algorithm for Large Categorical Datasets using a K-means Based Approach", *Computers and Electrical Engineering*, Vol. 37, No. 9, pp. 463-483, 2018.

[5] J. Hu, J. Liang, Y. Kuang and V. Honavar, "A User Similarity-Based Top-N Recommendation Approach for Mobile In-Application Advertising", *Expert Systems with Applications*, Vol. 111, pp. 51-60, 2018.

[6] Anand Khandare and A.S. Alvi, "Clustering Algorithms: Experiment and Improvements", *Computing and Network Sustainability*, Vol. 25, No. 12, pp. 263-271, 2017.

[7] Laith Mohammad Abualigah, Ahamad Tajudin Khader, Essam Said Hanandeh, and Amir H. Gandomi, "A Novel Hybridization Strategy for Krill-Herd Algorithm Applied to Clustering Techniques", *Applied Soft Computing*, Vol. 60, pp. 423-435, 2017.

[8] Laith Mohammad Abualigah, Ahamad Tajudin Khader and Essam Said Hanandeh, "Hybrid Clustering Analysis using Improved Krill-Herd Algorithm", *Applied Intelligence*, Vol. 48, No. 11, pp. 4047-4071, 2018.

[9] D. Dua and E. Karra Taniskidou, "UCI Machine Learning Repository", Available at: http://archive.ics.uci.edu/ml, Accessed on 2017.

[10] Methods and Application of Food Composition Laboratory, Available at: http://www.ars.usda.gov/Services/docs.htm?docid=8964.

[11] R.V. Rao, D.P. Rai and J. Balic, "A Multi-Objective Algorithm for Optimization of Modern Machining Processes", *Engineering Applications of Artificial Intelligence*, Vol. 61, pp. 103-125, 2017.

[12] Divya Pandove, "A Comprehensive Study on Clustering Approaches for Big Data Mining", *Proceedings of IEEE International Conference on Electronics and Communication System*, pp. 1333-1338, 2015.

[13] Jonathan Jeffrey, "Optimizing Web Structures using Web Mining Techniques", Springer, 2007.

[14] Daniel Kluver, M.D. Ekstrand and J.A. Konstan, "Rating-Based Collaborative Filtering: Algorithms and Evaluation", *Social Information Access*, Vol. 42, No. 8, pp. 344-390, 2018.

[15] Gunnar Schroder, Maik Thiele and Wolfgang Lehner, "Setting Goals and Choosing Metrics for Recommender System Evaluations", *Proceedings of Workshop at 5th ACM Conference on Recommender Systems*, pp. 53-61, 2011.

[16] M.V.B.T. Santhi, "Enhancing K-Means Clustering Algorithm", *International Journal on Computer Science and Technology*, Vol. 2, No. 4, pp. 235-246, 2012.

[17] Huanjing Wang and Jaason Van Hulse, "A Comparative Study of Threshold-based Feature Selection Techniques", *Proceedings of IEEE International Conference on Granular Computing*, pp. 499-504, 2010.

[18] X. Chen, X. Xu, J.Z. Huang and Y. Ye, "K Means: Automated Two-Level Variable Weighting Clustering Algorithm for Multiview Data", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 25, No. 4, pp. 932-944, 2013.

[19] G.Y. Zhang, C.D. Wang and D. Huang, "K-Means: Two Level Weighted Collaborative K-Means for Multiview Clustering", *Knowledge Based Systems*, Vol. 150, pp. 127-138, 2018.

[20] X.T. Zhang, "Multi-Task Multi-View Clustering", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 28, No. 4, pp. 3324-3338, 2016.

[21] L.M. Abualigah, A.T. Khader, E.S. Hanandeh and A.H. Gandomi, "A Novel Hybridization Strategy for Krill-Herd Algorithm Applied to Clustering Techniques", *Applications of Soft Computing*, Vol. 60, pp. 423-435, 2017.

[22] J. Bhavithra and A. Saradha, "Personalized Web Page Recommendation using Case-Based Clustering and Weighted Association Rule Mining", *Cluster Computing*, Vol. 22, No. 3, pp. 6991-7002, 2018.

[23] K. Renuka Devi, "Evaluation of Partitional and Hierarchical Clustering Techniques", *International Journal of Computer Science and Mobile Computing*, Vol. 8, No. 11, pp. 48-54, 2019.

[24] K. Renuka Devi, J. Bhavithra and A. Saradha, "Personalized Nutrition Recommendation for Diabetic Patients using Improved K-Means and Krill-Herd Optimization", *International Journal of Scientific and Technology Research*, Vol. 9, No. 3, pp. 1076-1083, 2020.