Case Retrieval using Bhattacharya Coefficient with Particle Swarm Optimization

Dr. Poonam Yadav
Assistant Professor, D.A.V College of Engineering & Technology, Kanina, Haryana 123027, India
poonam.y2002@gmail.com

Abstract—now a day, health information management and utilization is the demanding task to health informaticians for delivering the eminence healthcare services. Extracting the similar cases from the case database can aid the doctors to recognize the same kind of patients and their treatment details. Accordingly, this paper introduces the method called H-BCF for retrieving the similar cases from the case database. Initially, the patient's case database is constructed with details of different patients and their treatment details. If the new patient comes for treatment, then the doctor collects the information about that patient and sends the query to the H-BCF. The H-BCF system matches the input query with the patient's case database and retrieves the similar cases. Here, the PSO algorithm is used with the BCF for retrieving the most similar cases from the patient's case database. Finally, the Doctor gives treatment to the new patient based on the retrieved cases. The performance of the proposed method is analyzed with the existing methods, such as PESM, FBSO-neural network, and Hybrid model for the performance measures accuracy and F-Measure. The experimental results show that the proposed method attains the higher accuracy of 99.5% and the maximum F-Measure of 99% when compared to the existing methods.

Keywords - *case retrieval, BCF, similarity measure, PSO, accuracy*

*****

I. INTRODUCTION

The health information search and retrieval methods become a vital research area in recent years. In the US, 70% of users of the search engine do a search for retrieving the information about the health problems [13] [1]. In biomedical domains, medical cases participate the exclusive and inimitable role for knowledge representation while analyzing the clinical experience. Anyhow, the consistent method for accessing the medical knowledge quality depends on the evidence model. The medical reports of the patient consist of details of different degrees of completeness in the company of huge variable array for every patient. It is impractical to investigate all the features of each patient in the case database by means of every measurement of a new case. Most of the features are unrelated for finding the similarity among the new case and the cases in the database. The description of medical cases with several features did not improve the results of the case retrieval process [14] [3].

Medical Case Retrieval (MCR) is defined as the process of determining the details of diseases or health records of patients which are applicable for the new patient's symptoms that is send as the query by the medical experts. In clinical decision support systems [15], MCR plays the significant role and employs the concept of case-based reasoning (CBR) [16] [17]. In case-based reasoning, the most similar cases are retrieved from the case database for a given query, and the processes, such as diagnosis and treatment are applied to the patient. Anyhow, in medical research and education, MCR is the important problem because it permits to choose the exciting cases and create the datasets for medical studies visiting the case-based criteria. The case database contains the description of the case, and symptoms of various diseases which are in the form of multimedia documents usually has medical images and structured text [2].

The CBR is defined as the problem-solving strategy in which the knowledge of the already practiced cases is utilized to attain the solutions, such as treatment and diagnosis for the new case. Now a day, the CBR health science applications are a more interesting topic in several areas, such as the conventional medical domain of the clinical decision support [18], bioinformatics [19], and so on [20] [21] and intelligence methods exploited in many applications [26] [27]. The cognitive science community was introduced the CBR approach, and it is utilized if the absents of the global knowledge. In CBR, the problem solution in the case database which is more similar to the new case is near to the new problem's solution [3] [32]. Besides, the information retrieval is assisted with advanced compression techniques [30] [31] and modern data transmission technologies [28] [29].

This paper introduces the case retrieval system named as H-BCF for retrieving the similar case from the patient's case database. Initially, the patient's case database is constructed with details of different patients and their treatment details. If the new patient comes for treatment, then the doctor collects the information about that patient and sends the query to the H-BCF [33] [34]. The H-BCF system matches the input query with the patient's case database and retrieves the similar cases. Here, the PSO algorithm is used with the BCF for retrieving the most similar cases from the patient's case database. Finally, the Doctor gives treatment to the new patient based on the retrieved cases.

ISSN: 2321-8169
822 – 829

The rest of the paper is organized as follows: Section II presents the literature review, and Section III presents the system model of the proposed case retrieval method. In Section III, the proposed case retrieval method is described. Results and discussions are provided in Section IV and Section V concludes the paper.

II. Review of Related Works

This section presents the review of the existing research works in medical case retrieval and the various challenges occur during the case retrieval process.

A. Motivation

Here, five research works in medical case retrieval are discussed and the advantages and disadvantages of each work are described. Shen Ying *et al*. [22] have presented the steps and tools for the integrated CBR and Multi-Agent System (MAS) to represent the applications of these methods in the clinical decision support system. The advantages of this study are that it enhances the knowledge gaining process from the clinical operations and minimize a lot of the unsuitable quotations of the medical databases. The existing medical databases need a dull sorting process which is inappropriate for the medical practice. The drawback of this method is that it is very difficult to investigate and retrieve the evolutionary contexts. Mario Taschwer [2] had explained the MCR as the process of retrieving information from the group of medical case descriptions in which the symptoms of patients are utilized as queries. The author uses the familiar techniques for retrieving the similar case depends on the document and query expansion and then, integrates these techniques with new algorithms to the matching process of documents and queries with the Medical Subject Headings (MeSH). MeSH terms were obtained from the query which was used by the query expansion methods, and the local feedback increases the performance of the MCR for retrieving the full text. Here, the single data set was responsible for the evaluation, and the optimization of parameters is responsible for the results which affect the results of this method.

Markatou *et al.* [3] enhance the Comparative Effectiveness Research (CER) performance by integrating the medical case and the indirect evidence of same patients. The major advantage is that this method easily determines the subgroups from the huge datasets. The drawbacks of this method are that it was affected by data limitations and the unsuitable similarity measures selection. David A. Hanauer *et al.* [4] have represented a Michigan's University by a full-text search engine for performing the Information Retrieval (IR) from documents kept in Electronic Health Records (EHRs). This method was very useful in several areas, such as operational, research enterprise, and clinical areas. The drawback of this method is that it did not retrieve the optimal cases. Klaar Vanopstal *et al.* [5] have organized an information retrieval research in a nursing students group with mixed linguistic and education level backgrounds. This method had two objectives: improving the bibliographic information retrieval and make the profile for worst, average, and best performers of the test. Here, the limitation is the problem occurs in the relevance judgment step.

B. Challenges

The major challenge in the designing process of the MCR system is that it should be designed to support all the general medical datasets [2]. In CBR, case representation is the basic problem. There are a number of challenges in CBR. They are what to keep in a case, how to determine the suitable structure for representing the elements which describe the content of case, and how to organize and index the case database for effectively retrieving the cases [3]. Even though a number of techniques are available for retrieving the similar cases from the case database, creating an effective method to retrieve the similar cases remains the challenging problem [23, 14] [3].

III. System Model

This section presents the system model of the proposed H-BCF for case retrieval operation. The main objective of the proposed H-BCF is to retrieve the similar cases from the patient's case database for a new patient and helps the doctors to provide the treatments to the new patient. Figure 1 depicts the system model of the proposed H-BCF for retrieving similar cases for the new patient. Here, the Doctor gives the patient details to the medical diagnosis system which compare these details with the patient case database and retrieve the similar cases. Then, the doctor provides the treatment for the new patient. The patient's case database consists of details about numbers of the patient, like height, weight, blood pressure, and so on. Every case is represented as an attribute-value pair and also represented by a problem description and the corresponding solution description. If the new patient comes for the treatment, the Doctor collects the details of the new patient and communicated with the medical diagnosis system which checks the new patient's detail with the patient's case database. Then, the case which is similar to the new patient's details is retrieved from the patient's case database, and the treatment of the corresponding patient is provided to the new patient.

 ISSN: 2321-8169
822 – 829

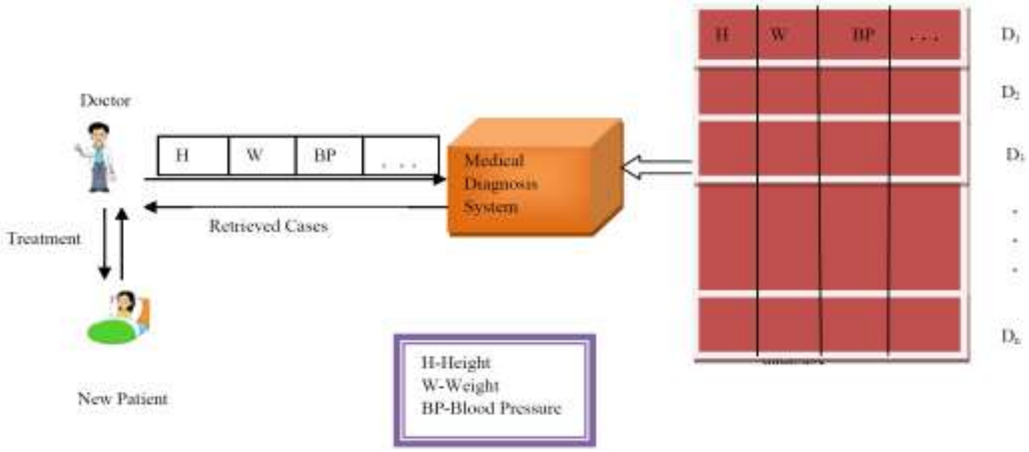


Figure 1. System model of H-BCF case retrieval

IV. PROPOSED H-BCF FOR RETRIEVING THE SIMILAR CASES FROM THE PATIENT'S CASE DATABASE

This section presents the proposed H-BCF for retrieving the similar cases from the patient's case database. The proposed system uses the BCF function and PSO algorithm for retrieving the similar cases. Figure2 shows the block diagram of the proposed H-BCF for retrieving the similar cases from the patient's case database. The information about patients, such as height, weight, blood pressure, and so on are collected and stored in a patient's case database. The patient's case database consists of a number of cases, and each case is represented as an attribute value. If the new patient comes for treatment, then the doctor collects the information about that patient and sends the query to the H-BCF. The H-BCF system matches the input query with the patient's case database and retrieves the similar cases. Here, the PSO algorithm is used with the BCF for retrieving the most similar cases from the patient's case database. Finally, the Doctor gives treatment to the new patient based on the retrieved cases.

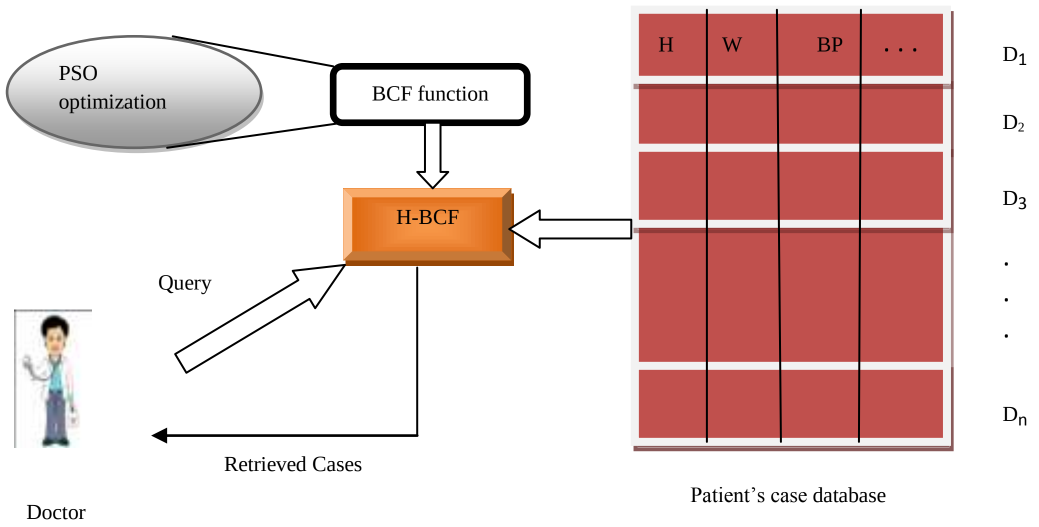


Figure 2. Block Diagram of the proposed case retrieval model

Let assume that, D be the patient's case database consists of details of the number of patients. The information about the patients, such as height, weight, blood pressure, and so on are stored in the database, and the treatment details are also stored in the patient's case database. The following equation represents the patient's case database,

$$D = \{D_i : 1 \le i \le n\} \tag{1}$$

where, D is the patient's case database and n is the number of cases stored in the database. Each case is represented as an attribute which is represented by the following equation,

$$D_i = \{b_j ; 0 \le j \le m\} \tag{2}$$

If the new patient visits the doctor then, the doctor collects the information about that patient and send the query to the medical diagnosis system. The query send by the doctor is represented as follows,

$$Q = \{q_k ; 0 \le k \le m\} \tag{3}$$

where, Q represents the query. q_k represents the number of queries. The query is matched with the patient's case database, and the similar cases are retrieved. Here, the similarity calculation is performed by the BCF.

A. Similarity function calculation using BCF

The similarity between the cases is calculated by the Bhattacharyya Coefficient in Collaborative filtering (BCF). The BCF amalgamates the global and local similarity to acquire the ultimate similarity value. Collaborative filtering (CF) is the victorious recommendation system [8-10] in recent years. The recommendation of items to the user is accomplished by examining the other users or other item's rating information. The reason for using the collaborative filtering is that it has several advantages, like domain independent and precise. The Bhattacharyya measure is employed in several areas, such as pattern recognition [11, 12], image processing, and signal processing. It calculates the similarity among two probability distributions [6].

The BCF integrates the local and global similarity to acquire the final similarity value. The query is evaluated with the cases presented in the patient's case database, and the cases which are similar to the query is retrieved from the database. The similarity among the query and the case which is already presented in the case database is calculated by the equation (3).

$$BCF(Q,D_i)=\alpha\left[\sum_{j=1}^{m}\sum_{k=1}^{m}BC(j,k)loc_{cor}(r_j,r_k)\right]+\beta\left[\sum_{j=1}^{m}\sum_{k=1}^{m}BC(j,k)loc_{med}(r_j,r_k)\right] \quad (4)$$

where, $BC(j,k)$ produces the global similarity information among the query k and the case attribute j.The local similarity takes the significant role, and it produces the local similarity information. α and β values are calculated by the PSO algorithm. If the query k and the case attribute j are similar to each other, then $BC(j,k)$ improves the local similarity among k and j. If the query k and the case attribute j are dissimilar to each other, then $BC(j,k)$ reduces the significance of local similarity among k and j. Here, the local similarity function is calculated by two functions, namely $loc_{cor}(\ )$ and $loc_{med}(\ )$. The $loc_{cor}(\ )$ is used to assess the correlation among the case attribute and the query which is represented by the following equation,

$$loc_{cor}(r_j,r_k)=\frac{(r_j-\bar{r})(r_k-\bar{r})}{\sigma_i,\sigma_j} \quad (5)$$

where, r_j represents the case attribute and r_k represents the query. σ_i is the standard deviation of the i^{th} case and σ_j is the standard deviation of the attribute j. Then, the local similarity is calculated by the $loc_{med}(\ )$ function.

$$loc_{med}(r_j,r_k)=\frac{(r_j-r_{med})(r_k-r_{med})}{\sqrt{\sum_{j=1}^{m}(r_j-r_{med})^2}\sqrt{\sum_{k=1}^{m}(r_k-r_{med})^2}} \quad (6)$$

where, r_{med} is the median of the case attributes. r_j represents the case attribute and r_k represents the query. The global similarity information is calculated by the function $BC(j,k)$. The BC similarity among the case attribute j and the query k is calculated as follows,

$$BC(j,k)=\sum_{f=1}^{z}\sqrt{\left(\overset{j}{P_f}\right)\left(\overset{k}{P_f}\right)} \quad (7)$$

where, $BC(j,k)$ produces the global similarity information among the query k and the case attribute j. The following equation calculates the value of $\overset{j}{P_f}$.

$$\overset{j}{P_f}=\frac{\#f^{\,j}}{n} \quad (8)$$

where, $\#f^{\,j}$ is the number of cases having f^{th} attribute value and n is the number of cases in the patient case database. $\overset{k}{P_f}$ can be calculated by the following equation,

$$\overset{k}{P_f}=\frac{\#f^{\,k}}{n} \quad (9)$$

where, $\#f^{\,k}$ represents the number of queries having the f^{th} attribute value and n is the number of cases in the patient case database.

k-retrieval: The input query is matched with the medical cases presented in the patient case database, and the BCF function computes the similarity between the query input and the medical cases presented in the case database. The BCF calculates the local and global similarity among the cases and integrates these similarity values to obtain the optimal similarity. Finally, the k numbers of similar cases having the most similarity are taken as the retrieved cases.

B. Optimal coefficients of BCF using Particle Swarm Optimization

Particle Swarm Optimization (PSO) is the non-linear function optimization algorithm introduced by James Kennedy and Russell Eberhart [7]. PSO is similar to the Genetic Algorithm in most of the cases, and it is inspired by the behavior of Birds flocking and Fish schooling. The number of similar cases is retrieved from the patient's case database, and the PSO algorithm selects the optimal case, and the treatment corresponding to that case is provided for the new patient. Initially, the population is generated with the random solutions, and the generations are updated to produce the optimal solutions. The solutions in PSO are called as particles, and the initial particles are moved over the problem space by following the current optimal particles. In each iteration, every particle is updated by the

 ISSN: 2321-8169
822 – 829

two best values namely, personal best (pbest) and global best (gbest). Then, the velocity and the position of the particles are updated. This process is repeated until the optimal solution reaches. The reason for choosing the PSO is that it is simple and it needs the adjustment of a small number of parameters. The PSO is used in several areas, like fuzzy system control, artificial neural network training, and function optimization and so on.

a) Solution Encoding: The weight values are assigned to every solution in the search space, and the size of the solution encoding is equal to the number of weights assigned. Here, every solution has two variables, namely α and β.

b) Fitness Calculation: The query input is given to the proposed H-BCF model, and the similar cases are retrieved from the patient's case database based on the weights assigned in the corresponding solution. Depending on the retrieved cases and the corresponding query the fitness is calculated by F-Measure.

c) Algorithm:

1. Initialization: In this step, the particles are initialized with random position and velocities on dimension H. Every particle contains the data which represents the possible solution, velocity value and the pbest value which represents the nearest particle's data which comes to the target at any time. The gbest is the current nearest particle's data to the target. In each iteration, the particles are updated by two best values, namely "pbest" and "gbest". If the pbest value of any particle comes nearer to the target than the gbest value then, the value of gbest will change. In each iteration, the gbest migrates nearer to the target till any one of the particles reaches the target.

2. Evaluation: Here, the fitness is calculated for every particle using F-Measure. The input query is matched with every case in the patient case database, and the matching cases are retrieved from the database. Then, the retrieved cases are evaluated with the cases presented in the training data set. The cases which have better values are considered as the fitness value.

3. Updating: If the fitness value is better than the current pbest value then, the fitness value is considered as the new pbest value otherwise the previous pbest is maintained. Then, the pbest value of the best particle is assigned to the gbest value. After finding the pbest and gbest value, the velocity and position of the particles are updated. The following equation updates the velocity of the particle,

$$v[\ ] = v[\ ] + l1 * r(\ ) * (pbest[\ ] - present[\ ]) + l2 * r(\ ) * (gbest[\ ] - present[\ ]) \tag{10}$$

where, $v[\ ]$ represents the velocity of the particle, $present[\ ]$ represents the current particle, $r(\ )$ is the random number between 0 and 1, and $l1, l2$ represent the learning factors. Then, the data values of each particle are updated by the velocity of the particle. Generally, $l1 = l2 = 2$. The following equation updates the position of the particle,

$$present[\ ] = present[\ ] + v[\ ] \tag{11}$$

where, $v[\ ]$ represents the velocity of the particle and $present[\ ]$ represents the current particle.

4. Termination: The steps mentioned above are repeated until the target solution reaches.

V. RESULTS AND DISCUSSION

This section presents the experimental results of the proposed method and the analysis of the proposed method with the existing case retrieval methods, such as PESM [25], FBSO-neural network [25], and Hybrid model [25] for two different data sets, namely breast cancer and breast cancer wins.

A. Experimental Setup

Platform: The proposed case retrieval technique is experimented in a personal computer with 2GB RAM and 32-bit OS. The proposed method is implemented using MATLAB 8.2.0.701 (R2013b).

Datasets used: For experimentation, two datasets such as breast cancer and breast cancer wins are taken from the UCI machine learning repository [24] for experimenting the proposed method.

Evaluation metrics: Evaluation metrics considered for analyzing the performance of the proposed method are accuracy and F-measure. Accuracy is defined as the proportion of accurately classified instances to the whole tested instances. The accuracy becomes false if the testing data consists of a disproportional number of cases. This problem is avoided through the use of F-Measure.

$$Accuracy = \frac{Number\ of\ correctly\ reterived\ cases}{Total\ number\ of\ cases} \tag{12}$$

$$F - Measure = \frac{2\,(S \bullet R)}{S + R} \tag{13}$$

where, S represents the precision and R represents the recall.

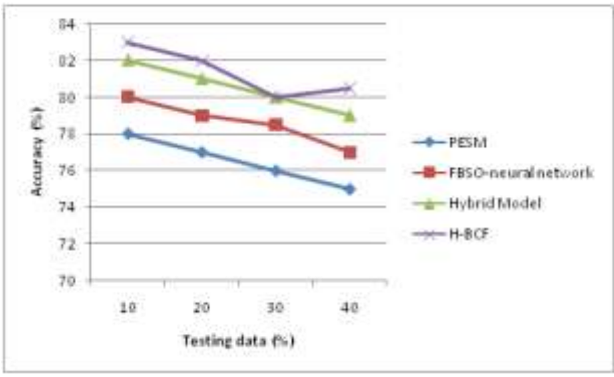
B. Performance Analysis

Here, the performance of the proposed method is analyzed with the existing methods, such as PESM, FBSO-neural network, and Hybrid model for the performance measures accuracy and F-Measure. Here, the performance is analyzed for two different data sets.

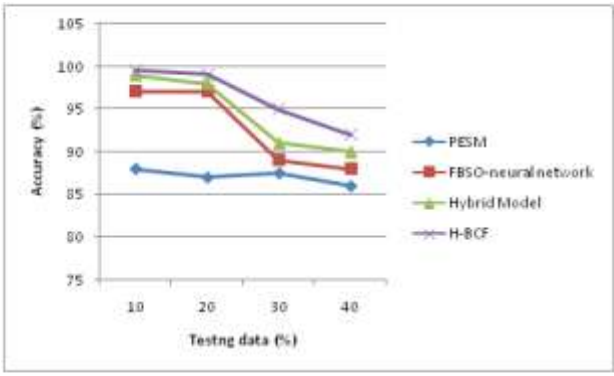
a) Accuracy

Figure 3 shows the accuracy of the proposed H-BCF method and the existing methods, such as PESM, FBSO-neural network, and Hybrid model for different size of testing data. Figure 3 (a) shows the accuracy of the proposed H-BCF method and the existing methods, such as PESM, FBSO-neural network, and Hybrid model for different size of testing data in breast cancer dataset. When the size of the testing data is 10%, the accuracy obtained by the proposed H-BCF is 83% on the other hand, the accuracy attained by the existing methods, such as PESM, FBSO-neural network, and Hybrid model is 78%, 80%, and 82% respectively. For the testing data size of 20%, the accuracy of the existing methods, such as PESM, FBSO-neural network, and Hybrid model is 77%, 79%, and 81% respectively while the proposed method has the accuracy of 82%. The accuracy of the proposed method is 80% when the testing data size is 30% while for the same data size the accuracy obtained by the existing methods, such as PESM, FBSO-neural network, and Hybrid model is 76%, 78.5%, and 80% respectively. When the size of the testing data is 40%, the accuracy attained by the existing methods, such as PESM, FBSO-neural network, and Hybrid model is 75%, 77%, and 79% respectively, on the other hand, the proposed method has the accuracy of 80.5% .

Figure 3 (b) shows the accuracy of the proposed H-BCF method and the existing methods, such as PESM, FBSO-neural network, and Hybrid model for different size of testing data in breast cancer wins dataset. When the size of the testing data is 10%, the accuracy obtained by the proposed H-BCF is 99.5% on the other hand, the accuracy attained by the existing methods, such as PESM, FBSO-neural network, and Hybrid model is 88%, 97%, and 99% respectively. For the testing data size of 20%, the accuracy of the existing methods, such as PESM, FBSO-neural network, and Hybrid model is 87%, 97%, and 98% respectively while the proposed method has the accuracy of 99%. The accuracy of the proposed method is 95% when the testing data size is 30% while for the same data size the accuracy obtained by the existing methods, such as PESM, FBSO-neural network, and Hybrid model is 87.5%, 89%, and 91%respectively. When the size of the testing data is 40%, the accuracy attained by the existing methods, such as PESM, FBSO-neural network, and Hybrid model is 86%, 88%, and 90%respectively, on the other hand, the proposed method has the accuracy of 92%. From figure 3, it can be shown that the proposed method has the higher accuracy than the existing methods, such as PESM, FBSO-neural network, and Hybrid model for different data size in breast cancer data set.



a) Breast Cancer



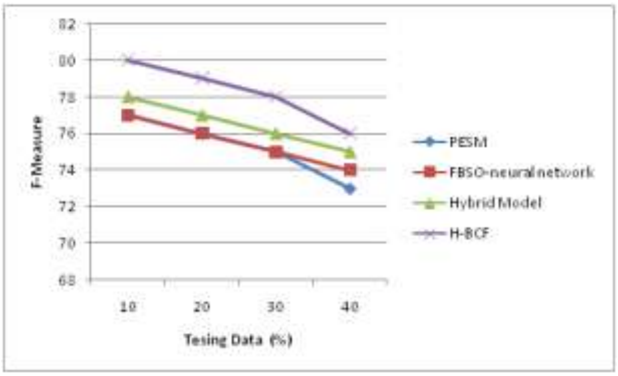
b) Breast Cancer Wins

Figure 3. Illustration of accuracy of the proposed H-BCF and the existing methods, such as PESM, FBSO-neural network, and Hybrid model

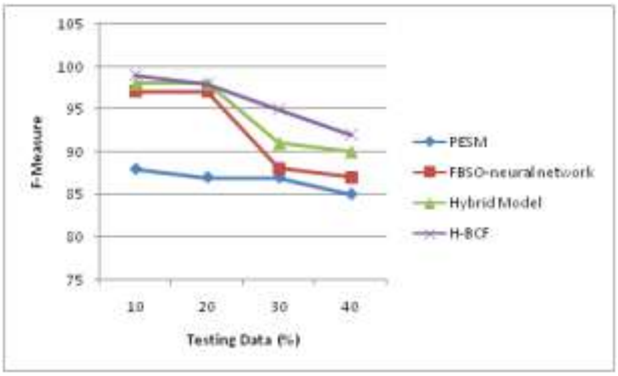
b) F-Measure

Figure 4 shows the F-Measure of the proposed H-BCF method and the existing methods, such as PESM, FBSO-neural network, and Hybrid model for the size of testing data 10%, 20%, 30%, and 40%. Figure 4 (a) shows the F-Measure of the proposed H-BCF method and the existing methods, such as PESM, FBSO-neural network, and Hybrid model for different size of testing data in breast cancer dataset. When the size of the testing data is 10%, the F-Measure obtained by the proposed H-BCF method is 80% while the F-Measure obtained by the existing methods, such as PESM, FBSO-neural network, and Hybrid model is 77%, 77%, and 78% respectively. The proposed method has the F-Measure of 79% when the size of the testing data is 20%, on the other hand, the F-Measure attained by the existing methods, such as PESM, FBSO-neural network, and Hybrid model is 76%, 76%, and 77% respectively for the same data size. The F-Measure attained by the proposed method is 78% for the testing data size of 30% while the F-Measure attained by the existing methods, such as PESM, FBSO-neural network, and Hybrid model is 75%, 75%, and 76% respectively for the testing data size of 30%. When the testing data size is 40%, the F-Measure of the existing methods, such as PESM, FBSO-neural network, and Hybrid model is 73%, 74%, and 75% respectively, on the other hand, the proposed H-BCF method has the F-Measure of 76%.

Figure 4 (b) shows the F-Measure of the proposed H-BCF method and the existing methods, such as PESM, FBSO-neural network, and Hybrid model for different size of testing data in breast cancer wins dataset. When the size of the testing data is 10%, the F-Measure obtained by the proposed H-BCF method is 99% while the F-Measure obtained by the existing methods, such as PESM, FBSO-neural network, and Hybrid model is 88%, 97%, and 98% respectively. The proposed method has the F-Measure of 98% when the size of the testing data is 20%, on the other hand, the F-Measure attained by the existing methods, such as PESM, FBSO-neural network, and Hybrid model is 87%, 97%, and 98% respectively for the same data size. The F-Measure attained by the proposed method is 95% for the testing data size of 30% while the F-Measure attained by the existing methods, such as PESM, FBSO-neural network, and Hybrid model is 87%, 88%, and 91% respectively for the testing data size of 30%. When the testing data size is 40%, the F-Measure of the existing methods, such as PESM, FBSO-neural network, and Hybrid model is 85%, 87%, and 90% respectively, on the other hand, the proposed H-BCF method has the F-Measure of 92%. From the figure 4, it can be shown that the proposed H-BCF method has the maximum F-Measure when compared to the existing methods, such as PESM, FBSO-neural network, and Hybrid model.



a) Breast Cancer



b) Breast Cancer Wins

Figure 4. Illustration of F-Measure of the proposed H-BCF and the existing methods, such as PESM, FBSO-neural network, and Hybrid model

VI. Conclusion

This paper introduces the method called H-BCF for retrieving the similar cases from the patient's case database. Initially, the patient's case database is constructed with the medical details of different patients and their treatment details. If the new patient comes for treatment, then the doctor collects the information about that patient and sends the query to the H-BCF. The H-BCF system matches the input query with the patient's case database and retrieves the similar cases. Here, the PSO algorithm is used with the BCF for retrieving the most similar cases from the patient's case database. Finally, the Doctor gives treatment to the new patient based on the retrieved cases. The performance of the proposed method is analyzed with the existing methods, such as PESM, FBSO-neural network, and Hybrid model for the performance measures accuracy and F-Measure. The experimental results show that the proposed method attains the higher accuracy of 99.5% and the maximum F-Measure of 99% when compared to the existing methods.

References

[1] Pavel Pecina, Ondrej Dusek, Lorraine Goeuriot, Jan Hajic, Jaroslava Hlavácová,Gareth J.F. Jones, Liadh Kelly, Johannes Leveling, David Marecek, Michal Novák, Martin Popel, Rudolf Rosa, Ales Tamchyna, and Zdenka Uresová, "Adaptation of machine translation for multilingual information retrieval in the medical domain," Artificial Intelligence in Medicine, vol. 61, no. 3, pp. 165–185, July 2014.

[2] Mario Taschwer, "Textual Methods for Medical Case Retrieval," Technical Report No. TR/ITEC/14/2.01, Institute of Information Technology (ITEC), Alpen-Adria-University at Klagenfurt, Austria, May 2014.

[3] M. Markatou, P. Kuruppumullage Don, J. Hu, F. Wang, J. Sun, R. Sorrentino, and S. Ebadollahi, "Case-based reasoning in comparative effectiveness research," IBM Journal of Research and Development, vol. 56, no. 5, pp. 4:1 - 4:12, 2012.

[4] David A. Hanauer, Qiaozhu Mei, James Law, Ritu Khanna, and Kai Zheng, "Supporting information retrieval from electronic health records: A report of University of Michigan's nine-year experience in developing and using the Electronic Medical Record Search Engine (EMERSE)," Journal of Biomedical Informatics, vol. 55, pp. 290–300, June 2015.

[5] Klaar Vanopstal, Joost Buysschaert, Godelieve Laureys, and Robert Vander Stichele, "Lost in PubMed. Factors influencing the success of medical information retrieval," Expert Systems with Applications, vol.40, no. 10, pp. 4106–4114, August 2013.

[6] Bidyut Kr. Patra, Raimo Launonen, Ville Ollikainen, Sukumar Nandi, "A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data," Journal of Knowledge-Based Systems archive, vol. 82, no. C, pp. 163-177, July 2015.

[7] James Kennedy and Russell Eberhart, "Particle Swarm Optimization," In proceedings of the IEEE International Conference on Neural Networks, 1995.

[8] G. Adomavicius, A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," IEEE Transaction on Knowledge and Data Engineering, vol. 17, no. 6, pp.734–749, 2005.

[9] J. Bobadilla, F. Ortega, A. Hernando, A. Gutirrez, "Recommender systems survey," Knowledge-Based Systems, vol. 46, pp. 109–132, 2013.

[10] X. Su, T.M. Khoshgoftaar, "A survey of collaborative filtering techniques," Advances in Artificial Intelligence, 2009.

[11] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," IEEE Transaction on Communication Technology, vol. 15, no. 1, pp. 52–60, 1967.

[12] F. Nielsen, S. Boltz, "The Burbea-Rao and Bhattacharyya centroids," IEEE Transaction on Information Theory, vol. 57, no. 8, pp. 5455–5466, 2011.

[13] Fox S., "Health Topics: 80% of internet users look for health information online," Technical Report. Pew Research Center, 2011.

[14] H. Cao, G. Melton-Meaux, M. Markatou, and G. Hripcsak, B, "Use abstracted patient-specific features to assist an information theoretic measure to assess similarity between medical cases," Journal of Biomedical Informatics, vol. 41, no. 6, pp. 882–888, December 2008.

[15] K. Kawamoto, C. A. Houlihan, E. A. Balas, and D. F. Lobach, "Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success," British Medical Journal, vol. 330, no. 7494, 2005.

[16] A. Aamodt and E. Plaza. "Case-based reasoning: Foundational issues, methodological variations, and system approaches," AI Communications, vol. 7, no. 1, pp. 39-59, 1994.

[17] S. Begum, M. Ahmed, P. Funk, N. Xiong, and M. Folke. "Case-based reasoning systems in the health sciences: A survey of recent trends and developments," IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 41, no. 4, pp.421-434, July 2011.

[18] L. Fritsche, A. Schlaefer, K. Budde, K. Schroeter, and H. H. Neumager, B, "Recognition of critical situations from time series of laboratory results by case-based reasoning," Journal of American Medical Informatics Association, vol. 9, no. 5, pp. 520–528, 2002.

[19] B. Yao and S. Li, "BANMM4CBR: a case-based reasoning method for gene expression data classification," Algorithms Molecular Biology, vol. 5, p. 14, January 2010.

[20] S. V. Pantazi, J. F. Arocha, and J. R. Moehr, "Case-based medical informatics," BMC Medical Informatics and Decision Making, vol. 4, p. 19, 2004.

[21] X. Song, S. Petrovic, and S. Sundar, "A case-based reasoning approach to dose planning in radiotherapy," In Proceedings of 7th International Conference on Case-Based Reasoning Workshop, Northern Ireland, pp. 348–357, 2007.

[22] Shen Ying, Colloc Joël, Jacquet-Andrieu Armelle, and Lei Kai, "Emerging medical informatics with case-based reasoning for aiding 4 clinical decision in multi-agent system,"

[23] G. Melton-Meaux, S. Parsons, F. P. Morrison, A. Rothschild, M. Markatou, and G. Hripcsak, B, "Inter-patient distance metrics using SNOMED-CT ontology structure," Journal Biomedical Informatics, vol. 39, no. 6, pp. 697–705, December 2006.

[24] Datasets from UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/.

[25] Poonam Yadav, "Case Retrieval Algorithm Using Similarity Measure and Adaptive Fractional Brain Storm Optimization for Health Informaticians," Arabian Journal for Science and Engineering, vol. 41, no. 3, pp. 829–840, March 2016.

[26] K. S. S. Rao Yarrapragada and B. Bala Krishna, "Impact of tamanu oil-diesel blend on combustion, performance and emissions of diesel engine and its prediction methodology",Journal of the Brazilian Society of Mechanical Sciences and Engineering, pp. 1-15,2016.

[27] Rao Yerrapragada. K. S.S, S.N.Ch. Dattu .V, Dr. B. Balakrishna,"Survey of Uniformity of Pressure Profile in Wind Tunnel by Using Hot Wire Annometer Systems",International Journal of Engineering Research and Applications, vol.4(3), pp. 290-299, 2014.

[28] Kavita Bhatnagar and S. C. Gupta, "Investigating and Modeling the Effect of Laser Intensity and Nonlinear Regime of the Fiber on the Optical Link", Journal of Optical Communications, 2016.

[29] K Bhatnagar and SC Gupta, "Extending the Neural Model to Study the Impact of Effective Area of Optical Fiber on Laser Intensity", International Journal of Intelligent Engineering and Systems, vol.10, 2017.

[30] B.S. Sunil Kumar, A.S. Manjunath, S. Christopher, "Improved entropy encoding for high efficient video coding standard", Alexandria Engineering Journal, In press, corrected proof, November 2016.

[31] BSS Kumar, AS Manjunath, S Christopher," Improvisation in HEVC Performance by Weighted Entropy Encoding Technique" Data Engineering and Intelligent Computing, 2018.

[32] S Chander, P Vijaya, P Dhyani,"Fractional lion algorithm– an optimization algorithm for data clustering" ,Journal of computer science, 2016.

[33] P. Vijaya, Satish Chander,"Fuzzy Integrated Extended Nearest Neighbour Classification Algorithm for Web Page Retrieval", Proceeding of the International Conclave on Innovations in Engineering and Management, 2016.

[34] P Yadav and RP Singh, "An Ontology-Based Intelligent Information Retrieval Method For Document Retrieval", International Journal of Engineering Science, 2012.