



Fusion of Hybrid Optimization Algorithm and Fuzzy set for enhancing Information Retrieval using clustering

B. Gomathi*
Research & Development Centre,
Bharathiar University, Coimbatore,
Tamil Nadu, India.
gomathiphd2k15@gmail.com

P. Sakthivel
Department of Electronics and Communication Engineering
Anna University, Chennai
Tamil Nadu, India

ABSTRACT

Fusion is the concept which deals with combination and the aim is to improve the overall performance. This paper proposes a novel approach named EPSOCS where Particle Swarm Optimization and cuckoo search are combined and fuzzy set is used for enhancing information retrieval. Clustering is accomplished and fuzzy c means algorithm is used for retrieval of information. This approach Enhanced Particle Swarm Optimization and Cuckoo Search combines advantages of PSO and cuckoo with fusion of fuzzy set and clustering. The evaluation of the proposed algorithm shows that the new approach exhibits good optimization ability with fast convergence speed leading to efficient information retrieval.

Keywords: EPSOCS, Fuzzy C Means, PSO, Cuckoo Search, Swarm Optimization, Fusion

1. Introduction

Data is available enormously and the volume of data is increasing at a very high speed. Fusion is a methodology used in many areas of technology since they increase the performance rate. Fusion technique can be used along with information retrieval to improve the efficiency rate of data retrieved. Clustering is key feature that plays vital role in knowledge discovery. Fuzzy logic fused with clustering yields better results. Fuzzy C Means algorithm for clustering is an efficient way and this is commonly used if the data is fuzzy. There are various similarity measures and similarity computation is done using cosine similarity.

The degree of membership is directly proportional to the distance between data object to the cluster centers. The only pitfall when using Fuzzy c-means is randomly selected centre points reaching the local optimal solution. To overcome this issue swarm intelligence that are bio inspired can be applied. This paper focuses on hybridization of Particle swarm optimization and cuckoo search and the novel method proved to be efficient than applied individually.

This scope of the paper is to deliver a novel method called Enhanced Particle Swarm Optimization with Cuckoo Search EPSOCS which combines advantages of PSO and CS along with fusion of fuzzy set and clustering. Section 2 deals with Related work and section 3 deals with comparative study of genetic

algorithm and about the Proposed method, section 4 describes about results based on metrics followed by conclusion.

2. Related Work

Data mining is always a flourishing field because data analysis can be easily accomplished using clustering approaches. Fuzzy clustering is of key importance in data analysis as most of the real time system prefer fuzzy than hard clustering. The membership value ranges between 0 and 1 in fuzzy and data falls between these values[1-2]. Fusion technique was applied by BogdanDit et al., [3] for combining information retrieval with link analysis algorithm to upgrade feature location in software. This is a fusion model for feature location and the new feature location techniques are based on integration of textual, dynamic, and web mining or link analysis algorithms applied to software. Swarm Intelligence is based on inspiration received from biological systems[4-5].

Every optimization algorithm is based on various inspiring agents. Ant Colony Optimization is based on behaviour of ants. Inspiring agent for Artificial Bee colony is honey bee. Particle Swarm Optimization is based on social behaviour of birds[6]. Cuckoo search is based on reproduction strategy followed by cuckoo and it is also a metaheuristic algorithm[7]. CS is a new approach and the popularisation of this algorithm is its implementation is easy. Hence CS is applied to solve many real world problem like scheduling problem etc. A detailed review of different Information retrieval technique is done and compared with proposed model. The proposed methodology combines PSO with CS and the enhanced hybrid optimization algorithm yield better result than used individually. In information retrieval similarity computation is major task to find relevancy and after careful study of various distance measure cosine similarity is used. The proposed model is a fusion of hybrid optimization algorithm with fuzzy set for efficient information retrieval.

3. Proposed System Architecture

This paper provides step by process by systematically splitting the work into three phases. During the first phase association among data is done and the statistical measure used to calculate distance is cosine similarity. In the second phase clusters are formed and evaluated using Fuzzy C Means algorithm. In the final phase Fusion process is done by combining hybrid optimization techniques with fuzzy data. The proposed system architecture is depicted in figure 1.

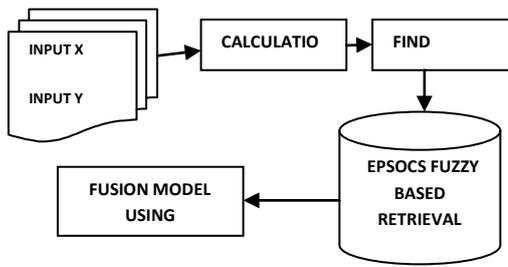


Figure1 Model of Proposed System Architecture

The Input Module deals about the data that needs to retrieved and this module is related to corpus creation. The proposed methodology uses health care data from UCI repository. Similarity computation is calculated and cosine similarity is used. The need for similarity computation is to find whether data belongs to same category. In this paper we use Fuzzy C means as many existing approaches proved that FCM provides better result compared to K-Means. The key factor is it permits data residing in multiple clusters with different membership values to participate in cluster analysis.

The objective function is given by equation (1)

$$K_m = \sum_{a=1}^D \sum_{b=1}^N \mu_{ab}^m \|x_a - c_b\|^2 \tag{1}$$

Step 1: Cluster membership values are initialised at random example μ_{ab}

Step 2: Cluster distance are calculated.

$$c_b = \frac{\sum_{a=1}^D \mu_{ab}^m x_a}{\sum_{a=1}^D \mu_{ab}^m} \tag{2}$$

Step 3: calculate the objective function.

$$\mu_{ab} = \frac{1}{\sum_{j=1}^N \left(\frac{\|x_a - c_b\|}{\|x_a - c_j\|} \right)^{\frac{2}{m-1}}} \tag{3}$$

Step 4: Repeat steps 2-4 till objective function is minimised or for specified number of maximum iteration.

The evaluation is based on precision and recall value of the retrieved document.

$$\text{Precision} = \frac{|\{\text{relevant document}\} \cap \{\text{Retrieved Document}\}|}{|\{\text{Retrieved Document}\}|} \tag{4}$$

The above equation 4 represents precision value.

The recall value is given by the equation 5.

$$\text{Recall} = \frac{|\{\text{relevant document}\} \cap \{\text{Retrieved Document}\}|}{|\{\text{Relevant Document}\}|}$$

The below figure 2 represents Precision and Recall value using Fuzzy C Means algorithm.

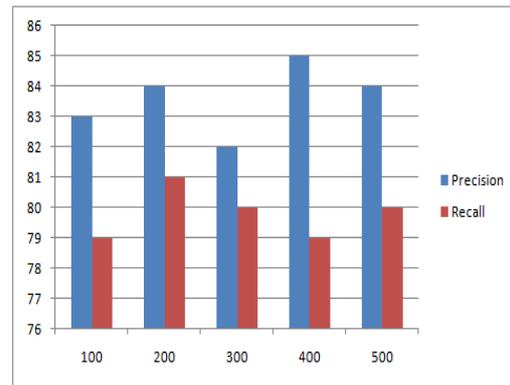


Figure 2 Precision and Recall value using Fuzzy C Means

3.1 Fusion of Fuzzy weight based system with Hybrid optimization

In this phase fuzzy based retrieval is performed using hybridization concept. The proposed method combines particle swarm optimization with cuckoo search and this enhancement on fuzzy based system proved to be efficient than used individually. In this paper, hybridization of Particle Swarm Optimization and Cuckoo search is done and classification based weight assignment with fuzzy approach is used in retrieving the information. Weights are calculated based on the importance such as very high, medium and related information and this is given in the figure 3.

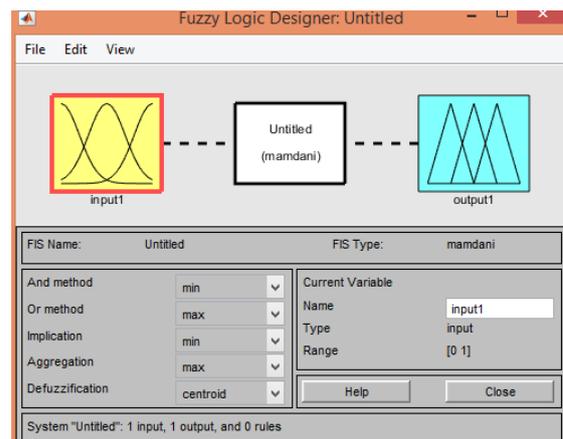


Figure 3 Fuzzy representation

Particle Swarm Optimization algorithm was discovered by James Kennedy and Russell Eberhart in the year 1995. It is based on the intelligence and movement of swarms. In this algorithm the work of agents also called as particles is to find the best solution in search space. The agents also referred as swarm must adjust flying and also they should consider about flying made by other agents. So personal best and global best position to be maintained. The advantages of PSO is the absence of selection operation normally used in genetic algorithms, and evolutionary programming. PSO also avoid crossover operation and it does not considers survival of fittest. Cuckoo search is based on the behavior of cuckoo. It is simple and easy to implement. The logic in CS is every cuckoo lays one egg and the egg is thrown at random in nest. Only the best eggs will be forwarded to the next level. The total number of nest is fixed and the host bird discover the egg based on probability. For finding new solution levy flight and random walk is used. The usage of levy flight in local and global search plays a significant rule in cuckoo search[8].

3.2 Algorithm for Information Retrieval

Step 1: Read the input data

Step 2: From that information, the relation between the information's are identified

Step 3 :Measure the distance between the information using the cosine similarity

Step 4: Based on the distance similarity between the information is identified

Step 5:Form the cluster using Fuzzy C means clustering approach with hybrid optimization

4. Results

Different IR techniques are compared with the new EPSOCS method on the benchmark dataset . Comparison of different Information Retrieval technique with the novel method is given in the table 4.1.

IR Technique	Precision	Recall
Vector Space Model	.75	.74
Latent Semantic Indexing	.83	.82
Latent Dirichlet Allocation	.861	.84
PSOCS weighted based Fuzzy Retrieval	.935	.87

Figure 4.1 Comparison of IR Techniques

The above table clearly depicts that our new approach EPSOCS Weight based Fuzzy Retrieval enhances the efficiency of information retrieval.

5. Conclusion

In this paper fusion technique is deployed by combining fuzzy set with hybrid algorithm. The integration of Particle swarm

optimization with cuckoo search fetches better result than used individually. EPSOCS has the advantages of fast convergence speed, strong searching ability, and the capability to solve the problem of multidimensional continuous space optimization by using test functions. In future bat algorithm may be fused with EPSOCS in order to get better result.

REFERENCES

- [1] Pal, N. R., Bezdek, J. C., and Tsao, E. C.-K. 1993, "Generalized clustering networks and Kohonen's self-organizing scheme", IEEE Trans. Neural Netw. 4, 549-557.
- [2]Bezdek, J. C. 1981, " Pattern Recognition With Fuzzy Objective Function Algorithms", Plenum Press, New York, NY.
- [3] Bogdan Dit, Meghan Revelle, Denys Poshvanyk, "Using Data Fusion and Web Mining to Support Feature Location in Software", 18th IEEE International Conference on Program Comprehension, 2010, pp.14-23.
- [4] Mohd Nadhir Ab Wahab, Samia Nefti-Meziani, and Adham Atyabi" A comprehensive Review of Swarm Optimization Algorithms", Published online 2015 May doi: 10.1371/journal.pone.0122827
- [5] Ying-Chih Wu, Wei-Ping Lee , Ching-Wei Chien,"Modified the Performance of Differential Evolution Algorithm with Dual Evolution Strategy", 2009 International Conference on Machine Learning and Computing IPCSIT vol.3 (2011)
- [6] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization: an overview," Swarm Intelligence, vol. 1, no. 1, pp. 33-57, 2007.
- [7] X.-S. Yang and S. Deb, "Cuckoo search via Lévy flights," in Proceedings of the World Congress on Nature Biologically Inspired Computing, pp. 210-214, IEEE, Coimbatore, India, December 2009
- [8] G. Zumofen, J. Klafter, and M.-F. Shlesinger, "Lévy flights and Lévy walks revisited," in Anomalous Diffusion From Basics to Applications: Proceedings of the XIth Max Born Symposium Held at Łądek Zdrój, Poland, 20-27 May 1998, vol. 519 of Lecture Notes in Physics, pp. 15-34, Springer, Berlin, Germany, 1999.