

# Information Retrieval in Data Mining with Soft Computing Algorithms

S. Surya <sup>[1]</sup>

Ph.D Research Scholar,  
PG and Research,

Department of Computer Science and Applications,  
Vivekanandha College of Arts and Sciences for  
Women (Autonomous), Elayampalayam,  
Tiruchengodu (DT), Namakkal (DT), Tamil Nadu, India.

Dr. P. Sumitra <sup>[2]</sup>

Assistant Professor,  
PG and Research,

Department of Computer Science and Applications,  
Vivekanandha College of Arts and Sciences for  
Women (Autonomous), Elayampalayam,  
Tiruchengodu (DT), Namakkal (DT), Tamil Nadu, India.

**Abstract-** In recent years IR-Information Retrieval has shown its development of indexing and searching useful information from the collection of database. Web is the universal repository to store information such as documents, text, music and images in huge collections. This information has to be retrieved by the user from the WWW with the irrelevant information. In this regard finding useful information is becoming more complicated, therefore proper information retrieval has to be taken care. Applying soft computing techniques like ANN, genetic algorithm, ant colony algorithm, PSO and differential evolution can overcome the problem of retrieving useful information. This paper will give an overview of soft computing technique for information retrieval.

**Keywords:** *Information Retrieval, Soft Computing, ANN, Genetic Algorithm, Ant Colony Algorithm, Differential Evolution, PSO*

## I. INTRODUCTION

IR is a wide area concerned for searching information, documents, metadata of document, relational databases on the World Wide Web. The ultimate goal for IR is to find relevant documents for the information required from the huge document. Classification of document, categorizing, modelling, data visualizing, system architecture, filtering, etc, are the recent research carried out in current years. Information retrieval has been utilized in many areas such as library records, scientific publication, books, journals and general data with respect to user's expectation. Retrieving useful information from the irrelevant data is highly complicated. Sorting from the huge amount of data and getting relevant information is known to be data mining. In general data mining is applied in organizations by business analyst and financial analysts and increasingly utilized in the field of science for extracting information from huge set of data.

Various techniques in soft computing like ANN, genetic algorithm, PSO and ant colony algorithm are used for the efficient retrieval of information [2, 6]. This paper explores the various soft computing techniques used for information retrieval.

## II. SOFT COMPUTING

Methodologies that are designed mathematically with a mix of modelling and enabling solutions in real world problems are soft computing. The main goal of soft computing is for exploiting. To exploit imprecision, approximate reasoning and uncertainty data in order to accomplish low margined results with tractability and robustness is the main aim of this technology [1]. Soft computing is variant with hard computing. Problems that is hard to answer for a solution is implemented using optimization technique which is followed in soft computing. The process of regulating the given inputs for finding min or max results is said as optimization. Soft computing utilizes various techniques, some of them are neatly explained below.

## III. ANT COLONY ALGORITHM

The born of AA-Ant Algorithm is derived from natural bionic algorithm. M. Dorigo has initially proposed this algorithm with the main aim in finding the optimal solution by using ants as information transmission [4]. That is the reason this algorithm is called as ACO-Ant Colony Optimization algorithm and AS-Ant System is applied in artificial ants. This algorithm best suits for graph based problems. ACO is applied only after the transformation of optimization problem to the problem that finds the better path in the weighted graph. Solutions are built incrementally by artificial ants' movement on the graph. This solution constructing process is speculative and hence influenced by pheromone model. ACO is suitable for travelling salesperson problem. In general ants have the capability to find the nearest path. With respect to environment change, ants can search the new paths. To search food ant have a special secretion known to be pheromone is utilized. More pheromone is left to increase the choice of probability, when more number of ants chooses the same path [2].

## IV. GENETIC ALGORITHMS (GA)

GA – Genetic Algorithm is one of the soft computing approaches. It is a subset of fuzzy logic and AI. The main idea in GA is inferring various optimization problems relevant to real life applications. The principles stimulated

by natural genetics in evolving solutions for a problem is followed in GA [3]. Fields such as climatology, climatology, automated manufacturing and design, games theory, biomedical engineering, biomedical engineering and code-breaking utilizes genetic algorithm. Maintaining the population of chromosomes is the basic idea of GA. The population is evolved for a positive iterative procedure of competitive and controlled variation. Generation word is mentioned for each state of population. Chromosomes at every generation are associated with a fitness value that specifies the quality of solution with chromosome's value. Selection of chromosomes for the new generation is decided by the fitness value. Crossover and mutations are the genetic operators used in creation of new chromosomes.

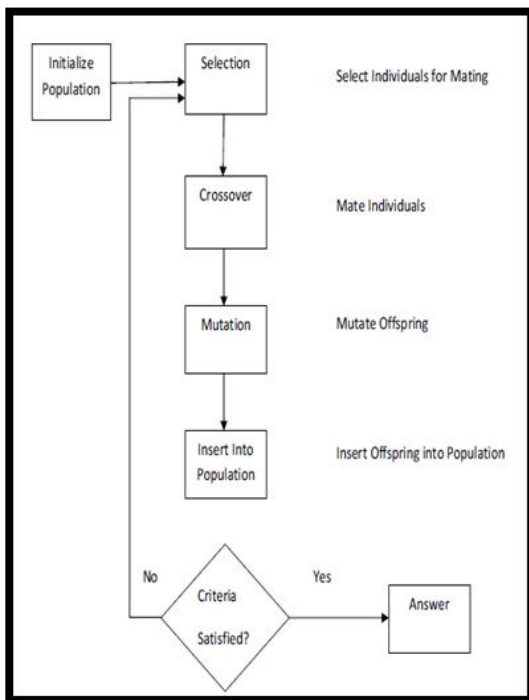


Fig.1 Mechanism of Genetic Algorithm

## V. PSO – PARTICLE SWARM OPTIMIZATION

It is a different evolutionary method which does not use crossover and mutation filtering operation. With the help of search procedure the members of the whole population is maintained. GA has a flaw that they do not keep information about the best solution amongst the community but they provide good solution [5]. With the composition of candidate solution and fitness and its velocity evaluated are maintained in each particles position. In addition best fitness value is remembered. Hence finally best fitness value of all particles in the swarm is maintained and is so called as global best fitness. Then the candidate that achieves this fitness is said as global best candidate solution or global best solution. This algorithm follows the below steps:

1. Location, velocity and population are initialized.
2. Individual particle fitness is evaluated (Pbest).
3. Highest fitness individuals are kept in track (Gbest).

4. Based on Gbest and Pbest location modification is carried in velocity.
5. Particle position is updated.
6. If the condition is satisfied terminate the process or else step 3 is continued.

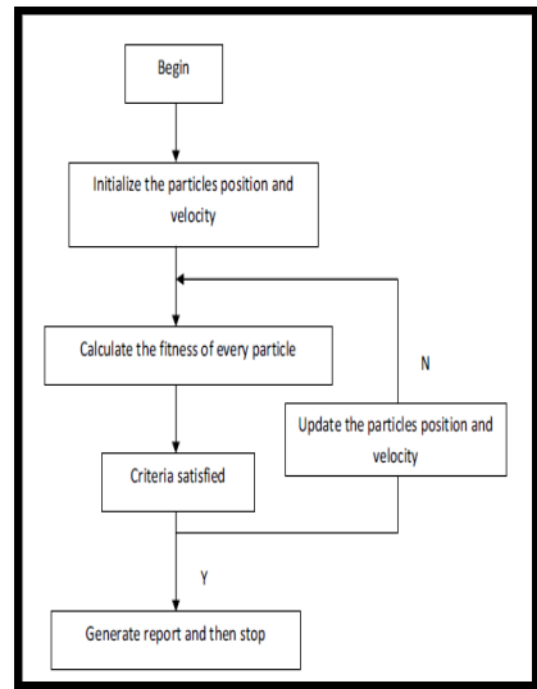


Fig. 2 PSO – Flow Chart

## VI.CONCLUSION

Data mining for information retrieval focus their research mainly on visualization techniques and discovery algorithms. Discovering patterns is easier but finding the patterns those results with useful information required by the user is badly low. To prevent user from the uninterested patterns, more techniques that identify useful patterns are introduced. Soft computing techniques and algorithms plays vital role in this areas. Some of those algorithms like genetic algorithm, ant colony algorithm and particle swarm optimization have been clearly overviewed in this paper.

## REFERENCES

- [1] [http://shodhganga.inflibnet.ac.in/bitstream/10603/10161/1/11/\\_chapter%203.pdf](http://shodhganga.inflibnet.ac.in/bitstream/10603/10161/1/11/_chapter%203.pdf)
- [2] Namrata Nagpal, "Applying Soft Computing Techniques in Information Retrieval", IJAEMS, Vol.4, Issue.5, May 2018.
- [3] Dogan Ibrahim, "An overview of soft computing", 12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS 2016, 29-30 August 2016, Vienna, Austria.
- [4] Marco Dorigo, Gambardella, Luca Maria, "Ant colonies for the traveling salesman problem." Biosystems, 1997, 43(2): 73-81.
- [5] Md. Abu Kausar, Md. Nasar & Sanjeev Kumar Singh, "Information Retrieval using Soft Computing: An Overview", IJSER, Vol. 4, Issue. 4, April 2013.
- [6] A. Roshdi, A. Roohparvar, "Review: Information Retrieval Techniques and Applications", International Journal of Computer Networks and Communications Security, Vol. 3, No. 9, September 2015.

## AUTHORS PROFILE



*SURYA.S*<sup>[1]</sup> received her M.Phil Degree in Computer Science from Indo-American College (Thiruvalluvar University)

Cheyyar, in the year 2013 and received her M.C.A Degree in Computer Applications

from Kirshnasamay College of Engineering and Technology (Anna university of Tiruchirappalli), Cuddalore in the year 2011. She is doing her Part time External Ph.D (Data Mining with Soft Computing) in PG and Research Department of Computer Science and Applications Vivekananda College of Arts and Sciences for Women (Autonomous), Elayampalayam, Tiruchengode-637205, Tamil Nadu, India. . She is presently working as an Assistant Professor in Department of Computer Applications, Kamban College of Arts and Science for Women, Tiruvannamalai (DT)-606603, Tamil Nadu, India. She published one paper in International conference and one papers in National Conference. She published one UGC Approved Journal and one IEEE Journal paper. Her research areas include Data Mining and Soft Computing.



*Dr.P.SUMITRA*<sup>[2]</sup> received her Ph. D Degree in Computer Science from Mother Teresa Women's University, Kodaikannal, Tamil Nadu in the year 2013. She is presently working as an Assistant Professor in PG and Research Department of Computer Science and Applications,

Vivekanandha College of Arts and Sciences for Women, Elayampalayam, Tiruchengodu (TK), Namakkal (DT), TamilNadu, India. She is a life member of The Indian Science Congress Association. She is currently guiding 7 Ph.D Research Scholar and 1 M.Phil Research Scholar. Her research interests are in Image Processing, Soft Computing and Data Mining.