OPTIMIZED SEARCH SPACE USING PARTICLE SWARM OPTIMIZATION: A REVIEW

Kripashankar Bhopche, Anurag Jain
Dept of Computer Science, RITS Bhopal, M.P., India

ABSTRACT

Data mining an non-trivial extraction of novel, implicit, and actionable knowledge from large data sets is an evolving technology which is a direct result of the increasing use of computer databases in order to store and retrieve information effectively .It is also known as Knowledge Discovery in Databases (KDD) and enables data exploration, data analysis, and data visualization of huge databases at a high level of abstraction, without a specific hypothesis in mind. The Optimization is a dire need for a huge amount of data processing. So that optimization is a challenging issue in data mining. It seems to be that there are many different approaches has been proposed by authors in order to optimize the results. Partial swarm optimization and genetic algorithms are some sort of approach which can be used for optimization. This paper is a review of optimization and its technique.

KEYWORDS: *Optimization, Swarm Optimization, Genetic Algorithm*

I. INTRODUCTION

Data mining (DM) is a non-trivial and useful information to extract a large amount of data that can be collected in many and various fields of science, business and engineering. DM [1] is part of a broader framework, known as knowledge discovery in databases (ECD), which covers a complex process of data preparation for modeling knowledge. In this process, and management techniques and algorithms are the real tools that analysts at his disposal to the Franciscans reasons is the second known correlation in the data. It can be found to extract data as an essential step in the knowledge discovery process. The data is generally pre-processed by the data cleaning, data integration, and data selection, and convert the data and prepared for the operation of the task. Which is nothing more dry DM extension techniques, DM now make important contributions in the critical areas of science and traditional surveys such as astronomy and high energy physics, biology and medicine, that has always provided a rich source of applications for data miners.

In computer science the optimization is a approach by which the selection is of data element will be the best element in the provided alternatives. Here the number of data wills large so that there is need to be used the efficient data. It can be happened with help of the optimization approached. There are many optimization approaches are available but the figure 1 shows some of them.

II. PARTICLE SWARM OPTIMIZATION

Particle swarm optimization (PSO) is a stochastic optimization technology on the basis of population and developed by Dr. Eberhart and Dr. Kennedy in 1995, inspired by the social behavior of birds[10]

PSO shares many similarities with evolutionary computation techniques such as Genetic Algorithms (GA). The system is initialized with a population of random solutions and searches for optima by updating generations. However, unlike GA, PSO has no evolution operators such as crossover and mutation. In PSO, the potential solutions, called particles, fly through the problem space by following the current optimum particles. The detailed information will be given in following sections [11].

Compared to GA, the advantages of PSO are that PSO is easy to implement and there are few parameters to adjust. PSO has been successfully applied in many areas: function optimization, artificial neural network training, fuzzy system control, and other areas where GA can be applied.

Inspired by the flocking and schooling patterns of birds and fish, Particle Swarm Optimization (PSO) was invented by Russell Eberhart and James Kennedy in 1995. Originally, these two started out developing computer software simulations of birds flocking around food sources, and then later realized how well their algorithms worked on optimization problems.

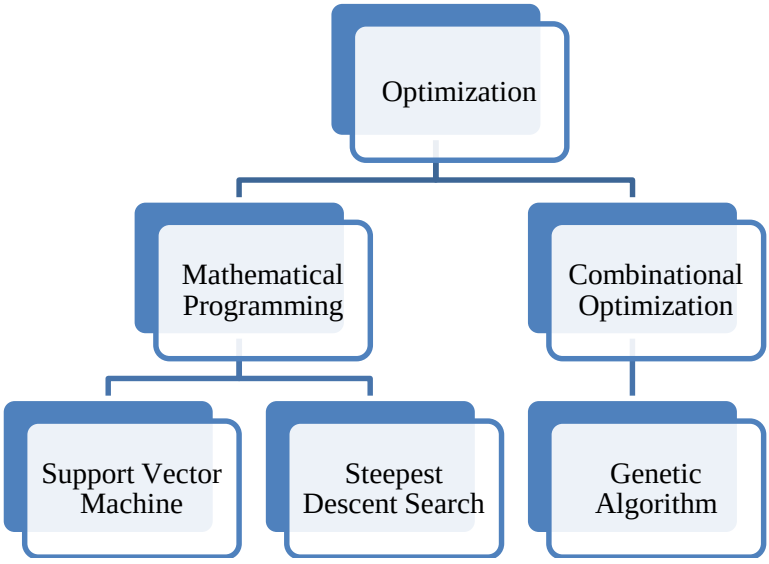

Figure 1 Optimization Classification

Particle Swarm Optimization might sound complicated, but it's really a very simple algorithm. Over a number of iterations, a group of variables have their values adjusted closer to the member whose value is closest to the target at any given moment. Imagine a flock of birds circling over an area where they can smell a hidden source of food. The one who is closest to the food chirps the loudest and the other birds swing around in his direction. If any of the other circling birds comes closer to the target than the first, it chirps louder and the others veer over toward him. This tightening pattern continues until one of the birds happens upon the food. It's an algorithm that's simple and easy to implement [10,11].
The algorithm keeps track of three global variables:

- The Target value or in other words condition
- Global best value or here Best indicating which particle's data is currently closest to the Target
- Stopping value indicating when the algorithm should stop if the Target isn't found

Each particle consists of:

- Data representing
- A Velocity value
- A personal best value

Figure 2 shows the flow graph of particle swarm optimization. Here the pseudo code of the PSO has been attached in a simple manner.

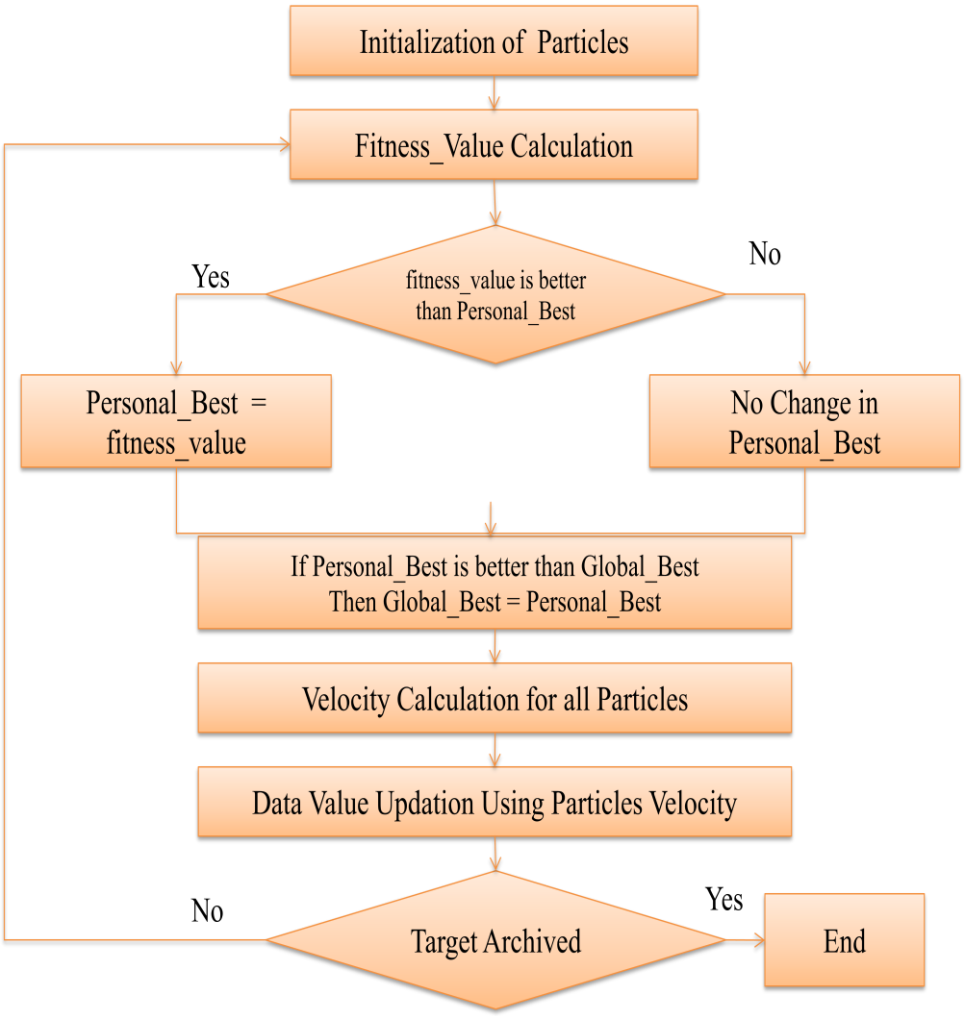

Figure 2. Flow of swarm optimization

```
For all particles
{
        Initialization of particle
}

Do for maximum iterations
{
        For all particles
        {
                Calculation of fitness_value
                If the fitness_value is better than Personal_Best
        {
        Set Personal_Best = fitness_value
}
If Personal_Best is better than Global_Best
{
        Set Global_Best = Personal_Best
}
}

For all particles
{
```

```
        Calculate particle Velocity
        Use Global_Best & Velocity to update particle Data
}
```

III. GENETIC ALGORITHM

Genetic algorithms (also known as evolutionary algorithms) use evolutionary principles to design systems that perform specific functions. Among other areas, computer programming with genetic algorithms is used to create programs and data structures that are continually redefined and refined to accomplish a particular state or purpose.

Evolution is an immensely rich and powerful method of creating and designing systems. Its achievements are results of a few simple concepts: sexual recombination, mutation and natural selection, all of which are recursed over many generations (Koza, et. al., 2003). Each generation is composed of the many qualities, and more, of the many generations before it. The span of generations is a search space of all possible designs. In order to speed up the process of natural selection, genetic algorithms can be run on isolated subpopulations, and after some period of time, the surviving populations are mixed together and further algorithms are performed.

Sudo Code for GA Approach

```
for each population members
        sum += FitnessValue
end

for each population members
   probability = sum of probabilities + (fitness / sum)
   sum of probabilities += probability
en

do till new population is full
    do this twice
       number = Random between 0 and 1
     for all members of population
        if number > probability but less than next probability
            then you have been selected
     end for
    end
end do
```

Specialty of the proposed method can more clearly be stated as follows:

A. It performs either equally good or better than many of the existing methods.
B. Its accuracy is more when applied on real and large dataset.
C. It is very simple and light because GA is used to search the optimal subset of attributes besides being used for searching the optimal techniques for attribute selections amongst the available ones.

A genetic algorithm [8] is a type of searching algorithm. It searches a solution space for an optimal solution to a problem. The algorithm creates a "population" of possible solutions to the problem and lets them "evolve" over multiple generations to find better and better solution. Algorithm is started with a set of solutions (represented by chromosomes) called population. Solutions from one population are taken and used to form a new population. Cycle of the Algorithm: The algorithm operates through a simple cycle.

- Creation of a population of strings.
- Evolution of each string.
- Selection of the best string.

- Genetic manipulation to create a new population of strings.

IV. RELATED WORK & PROBLEM FORMULATION

The traditional variable weighting methods suffer from unbalanced phenomenon: the view with more variables will play more important role than the view with less variables. In the two-level variable weighting method, the view weights will be only determined in the view level, while the variable weights will be only determined in a view. Therefore, the two levels of variable weights will eliminate the unbalanced phenomenon and compute more objective weights. There are lots of research paper has been studied but their scummy has been shown in the tabular format

citation	Author Name	Title	Publication Year	Methodology
4	Xiaojun chen xiaofei xu	Automated two-level variable weighting clustering algorithm for multiview data	2013 IEEE	The author has been used two real-life data sets to investigate the properties of two types of weights in TW-k-means and investigated the difference between the weights of TW-k-means and the individual variable weighting method
5	T.Velmurugan	Performance based analysis between k-Means and Fuzzy C-Means clustering algorithms for connection oriented telecommunication data	2014 Elsevier	These algorithms are implemented by means of practical approach to analyze its performance based on their computational time the telecommunication data is the source data for this analysis
6	Enmei Tu,Longbing Cao	A novel graph-basedk-means for nonlinear manifold clustering and representative selection	2014 Elsevier	The author has been proposed a graph –based k-means algorithm GKM which bears the simplicity of classic k-means while incorporating global information of data geometric distribution
7	Grigorios Tzortzis	The minmax k-means clustering algorithm	2014 Elsevier	The author has been proposed the Minmax k-means algorithm a method that assigns weights to the clusters relative to their variance and optimize a weighted to the clusters relative to their variance and optimize a weighted version of the k-means objective
8	M.A .Rahman M.Z Islam	A hybrid clustering technique combining a novel genetic algorithm with k-means	2014 Elsevier	The author has been proposed a novel GA based clustering technique that the is capable of automatically finding the right number of clusters and identifying the right genes through a novel initial population selection approach

In Gen Cluster two-level variable weighting method, high complexity of the genetic algorithms including Gen Cluster can cause problems for clustering data sets with huge number of records. A possible solution to this problem can be as follows:

(1) Take a random sample of a manageable number of records (as many as possible) into a sample data set,
(2) apply Gen Cluster on the sample data set and get the best chromosome,
(3) use the genes of the best chromosome as the initial seeds of the K-Means algorithm which is applied on the whole (not just the sample) data set.
Due to the low complexity of K-Means its application on the whole data set should not be a problem.
In future we suggest to use Particle swarm optimization (PSO) technique behalf of GA [6] for optimized clustering . Particle swarm optimization (PSO) is a population based stochastic optimization technique developed by Dr. Eberhart and Dr. Kennedy in 1995, inspired by social behavior of bird flocking or fish schooling. PSO shares many similarities with evolutionary computation techniques such as Genetic Algorithms (GA). The system is initialized with a population of random solutions and searches for optima by updating generations. However, unlike GA, PSO has no evolution operators such as crossover and mutation. In PSO, the potential solutions, called particles, fly through the problem space by following the current optimum particles.
The previous technique aims to achieve better quality clusters without requiring any user inputs such as the number of clusters k. GenClust uses GA to avoid the user input on k while achieving high quality clusters. Moreover, it uses 30 default radii values (r) ranging from 0.0001 to 0.2, where the domain of rx value is [0, 1]. The minimum number of records within the radius of a seed T = 1% by default. Since the radii values and T are interrelated it.
PSO based variable weighted clustering algorithm for multi view data will optimized the clustering technique in effective and efficient way. The rules discovered are generally with high accuracy, generalization and comprehensibility.

V. CONCLUSION

Data mining an non-trivial extraction of novel, implicit, and actionable knowledge from large data sets is an evolving technology which is a direct result of the increasing use of computer databases in order to store and retrieve information effectively .It is also known as Knowledge Discovery in Databases. A data mining software does not just change the presentation, but discovers previously unknown relationships among the data. The information on which the data mining process operates is contained in a historical database of previous interactions. . Some techniques such as association rules are unique to data mining, but most are drawn from related fields such as machine learning or pattern recognition. Due to large dataset of optimization has been needed. This paper gives an idea of optimization algorithm by which the efficient result can be fetched.

REFERENCES

[1] B.N. Lakshmi., G.H. Raghunandhan "A conceptual Overview of Data Mining" Proceedings of the National Conference on Innovations in Emerging Technology-2011 Kongu Engineering College, Perundurai, Erode, Tamilnadu, pp.27-32. India.17 & 18 February, 2011.
[2] Han J. and M. Kamber (2000), Data Mining: Concepts and Techniques, Academic Press, San Diego, CA.
[3] Thair Nu Phyu "Survey of Classification Techniques in Data Mining" Vol I Imecs 2009, March 18 - 20, 2009, Hong Kong.
[4] Xiaojun Chen, Xiaofei Xu, Joshua Zhexue Huang, and Yunming Ye "TW-k-Means: Automated Two-Level Variable Weighting Clustering Algorithm for Multi-view Data" in IEEE Transactions On Knowledge And Data Engineering, Vol. 25, No. 4, April 2013
[5] Enmei Tu , Longbing Cao , Jie Yang , Nicola Kasabov "A novel graph-based k-means for nonlinear manifold clustering and representative selection" in Elsevier transaction of Neuro computing 143 (2014) 109–122
[6] T. Velmurugan "Performance based analysis between k-Means and Fuzzy C-Means clustering algorithms for connection oriented telecommunication data" in Elsevier transaction of Applied Soft Computing 19 (2014) 134–146
[7] Grigorios Tzortzis "The Min Max k-Means clustering algorithm" in Elsevier transaction of Pattern Recognition 47 (2014) 2505–2516

[8] Cheng-Huang Hung , Hua-Min Chiou b, Wei-Ning Yang "Candidate groups search for K-harmonic means data clustering" in Elsevier transaction of Applied Mathematical Modelling 37 (2013) 10123–10128

[9] Md Anisur Rahman, Md Zahidul Islam "A hybrid clustering technique combining a novel genetic algorithm with K-Means" in Elsevier transaction of Knowledge-Based Systems 71 (2014) 345–365.

[10] Ahmed A. A. Esmin, Rodrigo A. Coelho, Stan Matwin, "A review on particle swarm optimization algorithm and its variants to clustering high-dimensional data", Springer Artificial Intelligence Review June 2015, Volume 44, Issue 1, pp 23-45

[11] Abraham A, Das S, Konar A "Kernel based automatic clustering using modified particle swarm optimization algorithm", Proceedings of the 9th annual conference on genetic and evolutionary computation GECCO'07, ACM Press, pp 2–9, ISBN 978-1-59593-698-1

AUTHORS BIOGRAPHY

Kripa Shankar Bopche has completed his engineering from Corporate institute of science & technology, Bhopal in information technology in 2012. Currently he is doing his Master of Technology from RITS Bhopal.