

REVIEW OF NATURE INSPIRED ALGORITHMS WITH DATA MINING

Neha Kathuria, Sunita Beniwal

Guru Jambheshwar University Of Science and Technology, Hisar, Haryana.

Abstract : The presence of large amount of data and need for turning such a huge data into knowledge has made data mining an important field of research. The huge amounts of data growing at a fast rate, makes it impossible for humans to analyse and extract knowledge from such data. For solving real-world optimization problems, optimization tools are required. Many Nature Inspired Algorithms namely Swarm Optimization, Firefly Algorithm, and Cuckoo search etc. have been used for mining. This paper presents a survey of some nature inspired algorithm which have been used for mining.

Keywords: Data Mining, Nature Inspired Algorithms, Classification, Clustering.

Introduction

Due to the presence of large amount of data and need for turning such a huge data into knowledge and useful information, Data Mining has secured an attention in this field. The problem arises with the large datasets or databases to extract the hidden information. The fast growing vast amount of data is stored in the large data repositories which is impossible for humans to comprehend without any powerful tool. "Necessity is the mother of invention." Necessity to uncover the hidden patterns and make the data 'information rich' attracted the attention towards the Data Mining. Data Mining is a promising field. Data Mining is a task of extracting and discovering the hidden interesting patterns from a huge amount of data. Data Mining is an essential step in the process of decision making and adding the information to our knowledge base. [1]

Data Mining

Data Mining is the process of solving the problems of evaluating the useful information/data already present in the large amount of database. The word Data Mining is actually a wrong name. Mining of data from knowledge should be referred to as 'knowledge

mining' but 'data mining' is somewhat small and also contain the word 'data' from which the knowledge is mining and hence, become popular. It is also known as knowledge extraction, data dredging, and knowledge discovery from data etc. Data Mining is used to uncover concealed patterns for evaluation. [1]

As the number of dimensions and the size of data increases, the data analysis needs to be performed. The process of extracting the knowledge from dataset is referred to as KDD (knowledge discovery from data). Data Mining is just a step in the process of KDD which work as a knowledge mining process from a huge amount of data. It is an automatic process of collecting the concepts, rules and patterns etc which is hidden in the dataset and help user in decision making after finding the interesting patterns. [2]

Following are the steps for the knowledge discovery:

- Data Cleaning
- Data Integration
- Data Selection
- Data Transformation
- Data Mining

- Pattern evaluation
- Knowledge Presentation [1]

Nature Inspired Algorithms

To solve real-world optimization problems, optimization tools are required. But these tools does not guarantee for the optimal solution. There are many problems for which no efficient algorithm is present and they have to solve by hit and trial method. Algorithms need a source of inspiration to develop and solve a problem. This inspiration is present in nature. Many Nature Inspired Algorithms namely Swarm Optimization, Firefly Algorithm, and Cuckoo search etc. are present. [3]

The Nature Inspired Algorithms can be classified into the following-

1. Swarm Intelligence based algorithm
2. Bio-inspired algorithm
3. Physics-chemistry based algorithms
4. Other algorithms

Swarm Intelligence based algorithm: In this algorithm, the population of agents is present which are interacting with each other. The individual agents are unintelligent but the whole system's behavior is intelligent. The examples of this algorithm are Ant Colony Optimization, Bacterial Foraging Optimization, and Cuckoo search etc. SI based algorithms are widely used and most popular among all.

Bio-inspired algorithm: A Bio-inspired algorithm is a wide area of algorithms and Swarm intelligence is a part of Bio-inspired algorithms. And Bio-inspired algorithms belong to a wide area of nature inspired algorithms. The examples of these are Differential evolution and brainstorm optimization etc.

Physics-chemistry based algorithms: Some of the algorithms are inspired by the physics and chemistry and they are not bio-inspired. As many fundamental laws of both the branches

are same, so we can't separate them and called physics-chemistry based algorithms. The examples are: central force optimization and gravitational search etc. [4]

The Nature Inspired algorithms are used with the aim to find out the global optimal solution to a problem. The factors affecting the solution of these algorithms are Exploration and Exploitation and they are common in all the nature inspired algorithms. Exploration is defined as finding the global optimal solution through randomly search of a solution space. Exploitation is referred to as finding the local optimal solution through the already explored space. The main theory behind the nature inspired algorithms is "survival of the fittest." [3]

Data Mining and Nature Inspired Algorithms

The Data Mining has been used with the nature inspired algorithms to solve the various problems associated in it. The algorithms like PSO, ACO are very popular in this field. The various data mining techniques can be used with nature inspired algorithms. Following are the techniques:

- Association
 - Classification
 - Clustering
1. Association in Data mining
The large amount of data is stored in the database for which the association is required to mine the interesting rules. P. Sehrawat [5] used the firefly algorithm with association rule mining. Firefly algorithm is used in swarm optimization algorithm. In this approach, the high quality association rules are explored without using the minimum confidence and support threshold. The approach was compared with similar previously existing technique and results are in favor of the

proposed approach due to less simulation time for complete population size.

2. Classification in Data mining

The classification technique in data mining is used with the Ant Colony Optimization technique. Rafael S. Parpinelli [6] proposed this approach for extracting classification rules. In this paper, the performance of CN2 (classification algorithm in data mining) was compared with the Ant Miner. The Ant Colony Algorithm showed better results due to high predictive accuracy and small rule list.

Data Mining can also be used in the cases where the comprehensibility, accuracy, and user interest is concerned.

3. Clustering in Data mining

In many Web applications, the data mining clustering algorithm has been widely applied. Tang Rui, Simon Fong [7] proposed an approach in which Clustering has been used with Particle Swarm optimization and Ant Colony optimization nature inspired algorithms i.e., C-PSO and C-ACO respectively. K-means clustering algorithm has been used with the web applications. The PSO and ACO are used with this technique to overcome the drawback of clustering algorithm to trap in the local minima. Swarm and

mutation behavior of PSO and ACO are used to help K-means to move out of local minima.

The other nature inspired algorithms were also used with data mining. Szymon Lukasik [8] studied the Flower Pollination Algorithm that is used with the clustering. First, the performance of FPA is evaluated and then the K-means clustering is used for the comparison. The clustering accuracy is high in the FPA base solution and is required in real time tasks.

Luciano D. S. Pacifico [9] introduced a new optimization model KGSO i.e., k-means with Group search optimization model. An improved GSO is used with evolutionary approach for data clustering. KGSO showed better results than PSO, K-mean, fuzzy C-mean and Also, it is robust.

Conclusion

Data is generated in large amounts nowadays. Extraction of knowledge from such large amounts of data is very tedious for humans. Nature inspired algorithms found application in many areas. They have been used for data mining. Various applications of nature inspired algorithms in clustering, classification and association rule mining have been discussed in this paper.

References

- [1] Jiawei Han, Micheline Kamber, and Jian Pei, *Data Mining concepts and techniques, 2nd edition*, Morgan Kaufmann, 2009.
- [2] S. Beniwal and J. Arora, "Classification and feature selection techniques in data mining," *International Journal of Engineering Research & Technology*, vol. 1, no. 6, pp. 1-6cc, 2012.
- [3] P. Agarwal and S. Mehta, "Nature-Inspired Algorithms: State-of-Art, Problems and Prospects", *International Journal of Computer Applications (0975 – 8887)*, Volume 100 – No.14, August 2014.
- [4] IztokFister Jr., Xin-She Yang, IztokFister, Janez Brest, and DusanFister, "A Brief Review of Nature-Inspired Algorithms for Optimization", *ELEKTROTEHNIŠKI VESTNIK 80(3): 1–7*, 2013, English Edition.
- [5] Poonam Sehrawat, Manju, Harish Rohil, "Association Rule Mining Using Firefly Algorithm", *International Journal of Latest Trends in Engineering and Technology*, Vol. 3, Issue 2, pp 263-270, November 2013.
- [6] Rafael S. Parpinelli, Heitor S. Lopes, and Alex A. Freitas, "Data Mining With an Ant Colony

Paper ID: UGC 48846-927

Optimization Algorithm, *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTING*, vol. 6, no. 4, pp 321-332, AUGUST 2002

[7] Tang Rui, Simon Fong, Xin-She Yang, Suash Deb, "Nature-inspired Clustering Algorithms for Web Intelligence Data", *IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, pp 147-153, 2012.

[8] Szymon Łukasik, Piotr A. Kowalski, Małgorzata Charytanowicz, Piotr Kulczycki, "Clustering using Flower Pollination Algorithm and Calinski-Harabasz Index", *IEEE Congress on Evolutionary Computation (CEC)*, pp 2724-2728, 2016

[9] Luciano D. S. Pacifico, Teresa B. Ludermir, "Group Search Optimization Method for Data Clustering", *IEEE Brazilian Conference on Intelligent Systems*, pp 342-347, 2014.
