

Swarms Intelligence and Their Applications in Data Mining

Yazdani Hasan

Department of Computer Application Sri Satya Sai University of Technology & Medical Sciences, Bhopal, India.

Abstract: *Swarm intelligence system consists typically of a population of relatively simple agents interacting only locally with themselves and with their environment. Data mining deals with analysis of large-scale data. In this paper, we discuss about the applications of Swarm intelligence techniques for data mining applications.*

Keywords: *Swarm Intelligence, Data Mining, Clustering, classification.*

I. Introduction

Data mining techniques analyse huge datasets [1, 2]. Supervised learning and unsupervised learning are two important data mining problems. In supervised learning, the output is given. The task is to create a model from a given training data (data points with output) such that when a new data point without output comes, the model predicts the output of the data point. When the output is class, the problem is called classification problem. If the output is a real number, the problem is called a regression problem. Decision trees, Naïve Bayes classifiers, neural networks etc. are popular classifiers [1, 2]. In clustering, the output of data points are not provided, the task is to group data points in different clusters such that the data points in a cluster are more similar than the data points in different clusters. K-means clustering, Hierarchical clustering etc. are popular clustering methods [8, 9].

There are many examples of collectively complex behaviour for example social insects such as: ant' colonies, termites, bees, wasps ...etc., So, a swarm intelligence system consists typically of a population of relatively simple agents interacting only locally with themselves and with their environment, without having a global knowledge about their own state and of the state of the world [3-6]. In this paper, we will discuss about the applications of swarm intelligence in data mining.

The paper is organized in the following manner. The Section 2 has discussion about the data mining techniques. Section 3 discusses about swarm intelligence and its application in data mining. Section 4 has conclusion and future work.

II. Data Mining

Data Mining is an important research area. Data mining deals with data analysis of large scale data. There are many popular data mining techniques. We will discuss some of them [1,2].

- **Clustering**

In clustering, similar data points are assigned to the same cluster, whereas different data points are assigned to different clusters.

- **Classification**

In classification, we try to predict the class of a given data point. These classes are given in the training data. The training data is used to train a model. The model is used to predict a new data point depending upon the the features of the data point.

- **Regression**

Regression is a supervised learning problem in which the task is to predict the output with given inputs. The output is a real number. There are many regression methods such that linear regression methods, regression trees, support vector machines and neural networks which are used extensively for regression problems. Regression trees, support vector machines and neural networks can handle both linear and nonlinear regression problems.

- **Neural Networks**

Neural networks are very popular for regression problems. A neural network consists of one or many interconnected artificial neurons. They can approximate highly nonlinear regression functions. However, neural networks are slow and selection of proper hidden layers is a problem.

- **Support vector machines**

Support vector machines are useful for regression problems. Similar to neural networks they can approximate highly nonlinear problems. However, the selection of a proper kernel is very important for the accuracy of the support vector machines.

- **Ensembles**

Ensembles are a combination of multiple base models; the final classification depends on the combined outputs of individual models. Classifier ensembles have shown to produce better results than single models, provided the models are accurate and diverse.

III. ISwarm Intelligence And Its Applications In Data Mining

A swarm intelligence system consists typically of a population of relatively simple agents interacting only locally with themselves and with their environment, without having a global knowledge about their own state and of the state of the world. Particle Swarm Optimization (PSO) and Ant Colonies Optimization (ACO) are currently the most popular algorithms in the swarm intelligence domain. PSO is a population-based search algorithm and is initialized with a population of random solutions, called particles [7]. Unlike in the other evolutionary computation techniques, each particle in PSO is also associated with a velocity. Particles fly through the search space with velocities which are dynamically adjusted according to their historical behaviors. Therefore, the particles have the tendency to fly towards the better and better search area over the course of search process. PSO have been used for many data mining applications. PSO have shown excellent performance for clustering applications. PSO has also been used for Image classification.

Ant Colonies Optimization (ACO) algorithms were introduced around 1990 [8]. These algorithms were inspired by the behavior of ant colonies. Ants are social insects, being interested mainly in the colony survival rather than individual survival. Of interests is ants' ability to find the shortest path from their nest to food. This idea was the source of the proposed algorithms inspired from ants' behavior. The first application of ant systems for data mining was reported in [9], where the authors introduce the AntMiner algorithm for the discovery of classification rules. Researchers [10, 11] have used Ant Colonies Optimization (ACO) algorithms for the quadratic assignment problem. Ant Colonies Optimization (ACO) algorithms have been used for feature selection task [12, 13]. Yang and Kamel [14] propose a clustering ensemble approach using three ant colonies. ant-based clustering algorithm [15] has been very popular for clustering problems. Ant Colonies based algorithms have also been used for outlier detection.

IV. Conclusion

In this paper, we discussed about the basics of data mining and Swarm Intelligence. We also discussed about the applications of Swarms Intelligence. This is an emerging field. In future, Swarm Intelligence will have many new data mining applications.

References

- [1]. J. Han, M. Kamber and J. Pei, *Data Mining: Concepts and Techniques*, Morgan Kaufmann; 2011.
- [2]. I. H. Witten, E. Frank, M. A. Hall (Author) *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann; 2011.
- [3]. Beni, G., Wang, J.: *Swarm intelligence*. In: Proc. of the Seventh Annual Meeting of the, Tokyo, Japan (1989) 425–428.
- [4]. Beni, G.: *From swarm intelligence to swarm robotics*. In S,ahin, E., Spears, W., eds.: *Swarm Robotics: State-of-the-art Survey*. Lecture Notes in Computer Science 3342, Springer-Verlag (2005) 1–9.
- [5]. Bonabeau, E., Dorigo, M., Theraulaz, G.: *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press, New York, NY, USA (1999).
- [6]. Dudek, G., Jenkin, M., Milius, E., Wilkes, D.: *A taxonomy for multi-agent robotics*. *Autonomous Robots* 3 (1996) 375–397.
- [7]. Cao, Y., Fukunaga, A., Kahng, A.: *Cooperative mobile robotics: Antecedents and directions*. *Autonomous Robots* 4 (1997) 1–23.
- [8]. Kennedy, J.; Eberhart, R. (1995). "Particle Swarm Optimization". *Proceedings of IEEE International Conference on Neural Networks*. **IV**. pp. 1942–1948
- [9]. A. Colomi, M. Dorigo et V. Maniezzo, *Distributed Optimization by Ant Colonies*, actes de la première conférence européenne sur la vie artificielle, Paris, France, Elsevier Publishing, 134-142, 1991.
- [10]. Parpinelli RS, Lopes HS, Freitas AA (2001) An ant colony based system for data mining: Applications to medical data. In: Lee Spector, Goodman E, Wu A, Langdon WB, Voigt H, Gen M, Sen S, Dorigo M, Pezeshk S, Garzon M, Burke E (eds) *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, pages 791–797, San Francisco, California, USA, 7-11. Morgan Kaufmann.
- [11]. Maniezzo V (1998) Exact and approximate nondeterministic tree-search procedures for the quadratic assignment problem. *Research CSR 98-1*, Scienze dell'Informazione, Universit'a di Bologna, Sede di Cesena, Italy
- [12]. Maniezzo V, Colomi A (1999) The ant system applied to the quadratic assignment problem. *IEEE Transactions on Knowledge and Data Engineering*.
- [13]. Jensen R, Shen Q (2003) Finding Rough Set Reducts with Ant Colony Optimization. In *Proceedings of the 2003 UK Workshop on Computational Intelligence*, pp 15-22
- [14]. Jensen R, Shen Q (2005) Fuzzy-Rough Data Reduction with Ant Colony Optimization. *Fuzzy Sets and Systems*, Vol. 149, No. 1, pp. 5-20.
- [15]. Yang Y., Kamel M. (2003) Clustering ensemble using swarm intelligence. In: *IEEE Swarm Intelligence Symposium 2003*, pp. 65-71.
- [16]. Lumer E., Faieta B. (1994) Diversity and adaptation in populations of clustering ants. In: *Third Int. Conf. on Simulation of Adaptive Behavior: from animals to animats 3*, Cambridge, MA: MIT Press, pp. 499-508.