Hybrid Solution of K-Means with PSO and Cuckoo Search

Alper TADAY
Computer Engineering Department
TOBB University of Economics and Technology
Ankara,Turkey
st101101020@etu.edu.tr

Ecem Elvin ÇEVİK
Computer Engineering Department
TOBB University of Economics and Technology
Ankara,Turkey
st111101037@etu.edu.tr

Fırat TOP
Computer Engineering Department
TOBB University of Economics and Technology
Ankara,Turkey
st101101047@etu.edu.tr

ABSTRACT

This paper presents a k-means clustering approach for classifying and shows how the traditional k-means clustering algorithm can be modified to be used as a classifier algorithm. The proposed model comprises particle swarm optimization (PSO) with the traditional K-means algorithm to provide the requirements of a classifier. Also it combines cuckoo-search based evolutionary algorithm with some centroid-calculation heuristics.

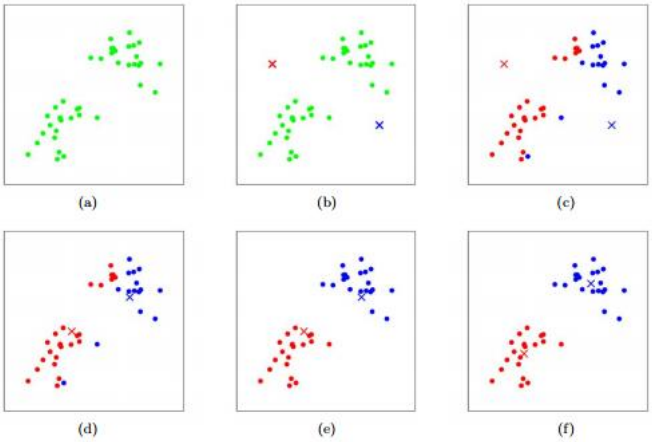
1. INTRODUCTION

This work promotes to the application of a popular clustering technique, k-means clustering, modified with the use of PSO. The classifier is projected using the proposed PSO based k-means clustering algorithm. This proposed algorithm states the optimal cluster centers based on train data set. Classification precision and false dismissals provide affirmation to the performance of the algorithm for both training and testing steps. The proposed clustering approach is implemented and the traditional k-means algorithm compared with the results.

In this work, PSO based k-means clustering algorithm is determined as a suitable tool for the design of classifier. This gives a brief outline of k-means clustering, use of PSO in improving the performance of k-means algorithm for classification.

2. K-MEANS

K-means is one of the simplest algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. [1]



The algorithm is composed of the following steps: [2]

1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the positions of the K centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

3. PARTICLE SWARM OPTIMIZATION

Particle swarm optimization (PSO) is a quantitative method that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. Elementally, PSO is a population based stochastic optimization technique. 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. [3]

4. CUCKOO SEARCH

It was inspired by the obligate brood parasitism of some cuckoo species by laying their eggs in the nests of other host birds (of other species). Some host birds can engage direct conflict with the intruding cuckoos.

Cuckoo search idealized such breeding behavior, and thus can be applied for various optimization problems. It seems that it can outperform other metaheuristic algorithms in applications.

An important advantage of this algorithm is its simplicity. In fact, comparing with other population or agent-based metaheuristic algorithms such as particle swarm optimization. There is essentially only a single parameter P_a in CS (apart from the population size n). Therefore, it is very easy to implement.

Cuckoo search uses the following representations:

Each egg in a nest represents a solution, and a cuckoo egg represents a new solution. The aim is to use the new and potentially better solutions (cuckoos) to replace a not-so-good solution in the nests. In the simplest form, each nest has one egg. The algorithm can be extended to more complicated cases in which each nest has multiple eggs representing a set of solutions.

CS is based on three idealized rules:

1. Each cuckoo lays one egg at a time, and dumps its egg in a randomly chosen nest;
2. The best nests with high quality of eggs will carry over to the next generation;
3. The number of available hosts nests is fixed, and the egg laid by a cuckoo is discovered by the host bird with a probability $P_a \in (0, 1)$. Discovering operate on some set of worst nests, and discovered solutions dumped from farther calculations. [4]

5. CONCLUSION

In this paper, we aim to present techniques and algorithms separately which improve clustering quality. These are k-means, PSO and Cuckoo Search. We determine to improve clustering algorithm by combining with these algorithms. We expect that PSO based k-means algorithm is more useful than traditional k-means. After finishing implementation of project, we will test our algorithm with some datasets to compare with traditional k-means clustering.

6. REFERENCES

[1] https://sites.google.com/site/dataclusteringalgorithms/k-means-clustering-algorithm

[2] http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html

[3] http://www.swarmintelligence.org/tutorials.php

[4] https://en.wikipedia.org/wiki/Cuckoo_search