A Project Report

Preliminary Analysis using Clustering Techniques leading to a Game Theory based Expert System for Reactor Control

**Jayantha Nanduri**

**16XJ1A0564**

Under the supervision of

**Dr Arya Kumar Bhattacharya**

**SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF**

**SE 322: PROJECT TYPE COURSE**



****

**MAHINDRA ECOLE CENTRALE**

**COLLEGE OF ENGINEERING**

**HYDERABAD**

**(MAY 2019**)

CONTENTS

Title page……………………………………………………………….1

Acknowledgements………………………………………………..2

Certificate…………………………………………………………......3

Abstract………………………………………………………………....4

Introduction……………………………………………….……….....6

Problem Definition…………………………………………………..7

Background and Related Work…………………….…………….8

Implementation………………………………………………………10

Results……………………………………………………………………12

Conclusion………………………………………………………….…..13

References……………………………………………………….……..14

ACKNOWLEDGMENTS

I would first like to thank my supervisor Dr. Arya Kumar Bhattacharya, Dean of Research, Mahindra Ecole Centrale for giving me the opportunity and the resources to work on this project. I would also like to thank my project partner, Artika. Without such a passionate participator and team member, our project could not have been successfully conducted.

ABSTRACT

Our aim in this project is to analyse the various parameters of the dataset collected from an industrial reactor using statistical and Machine Learning techniques to extract relationships between variables, using Unsupervised (Mainly Fuzzy C-means) techniques, which later might serve as foundations of a Game Theory based Expert System. This system will be designed to help the reactor run in an optimised state throughout its work.

Initially, we worked with mock data which was put into codes of K-means and Fuzzy C- means algorithm, which we implemented from scratch, to test the efficiency of these in giving us the relations between the attributes. Results from both these algorithms were compared and it was found that since Fuzzy C results in giving us a “degree of belonging” of a certain instance in a cluster rather than forcefully pushing it into a cluster (like in the case of K-means), it is a better algorithm to work within our case.

Once satisfactory and accurate results of the mock data were obtained from our codes, actual data of various parameters of an industrial reactor that have been accumulated over some time were put into the same codes and were compared with results obtained using MATLAB. We found that different instances of the states of the industrial reactor were grouped according to the regimes in which they were operating, hence giving us a better understanding, of which set of parameters have to be adjusted accordingly according to a particular regime to give the best possible result.

This analysis that we gather now will later serve as an input to a game theory-based expert system as mentioned above, in which each of these control variables will be an agent which will intend to maximise its efficiency to give the best optimal results.

INTRODUCTION

Given an industrial reactor such as a blast furnace, is it important to know how to balance out the control and state variables (like wall temperatures, wall pressures etc.) to optimise productivity, quality and stability throughout the period for which it runs. Since, the reactor is active for very long periods (more than 15 years), maintaining the optimal state by making all the variables work in each other’s favour is a difficult task for a human. Therefore, an expert system based on game theory can be used to improve the functionality of the reactor.

The goal of an expert system is to run the reactor all the time at the highest feasible levels of the overarching objectives of stability, productivity and quality by imposing optimum control through suggested operator action, and occasionally through direct intervention.

We have worked with data consisting of 26 and 36 variables from an actual blast furnace and tried to classify this data such that there is high intra-class similarity and low inter-class similarity. We have used clustering as an approach because it helps to find natural groupings about objects. The clustering techniques we used were K-means and fuzzy C.

Why fuzzy logic?

The term **fuzzy** refers to things which are not clear or are vague. In the real world many times we encounter a situation when we can’t determine whether the state is true or false, their fuzzy logic provides a very valuable flexibility for reasoning. In this way, we can consider the inaccuracies and uncertainties of any situation.

In the Boolean system truth value, like the one K means uses, 1.0 represents absolute truth value and 0.0 represents absolute false value. Therefore, running out of data sets for K-means, we forcefully put a data point into a cluster if it was 0 for the rest. But in the fuzzy system, there is no logic for absolute truth and absolute false value. But in fuzzy logic, there is an intermediate value too present which is partially true and partially false. Using fuzzy, we found that different instances of the states of the industrial reactor were grouped according to the regimes in which they were operating, hence giving us a better understanding, of which set of parameters have to be adjusted accordingly according to a particular regime to give the best possible result.

We plan to use game theory later to analyse these variables.

A brief about game theory:It’s a branch of applied mathematics that provides tools for analysing situations in which parties, called players, make interdependent decisions. Hence, our system will have control variables where each tries to be in its best state, giving us the best productivity, stability and quality.

PROBLEM DEFINITION

Basic questions that our project targets:

1. How can the control and state variables of an industrial reactor be controlled so that the most feasible output is obtained throughout the lifecycle of the reactor?

Since we have used fuzzy logic, it has given us a probabilistic measure of how the variables are dependent on each other. Therefore, when we input these measures into the game theory model, it is expected to give results that haven’t originated from forced relationships, but from a most probable point of view.

1. Will game theory-based experts be efficient enough to implement in real life?

If trained enough and properly, the system must be ready to go out into the real world and help industries put their reactors to the best use.

1. The advantages of using fuzzy C means clustering to find relationships between variables.

As mentioned above, for the industrial data that we have been working on, Fuzzy C has given “a degree of belonging ” rather than just a belonging of a data point in a particular class. Hence it has proved to be a good unsupervised learning algorithm, suitable to group data based on similarity; in our case, the different regimes of operating a reactor.

BACKGROUND AND RELATED WORK

Some of the previous work in the same field:

**Comparative Analysis of K-Means and Fuzzy C-Means Algorithms**

Soumi Ghosh Department of Computer Science and Engineering, Amity University, Uttar Pradesh Noida, India Sanjay Kumar Dubey Department of Computer Science and Engineering, Amity University, Uttar Pradesh Noida, India

The raw, unlabeled data from the large volume of dataset can be classified initially in an unsupervised fashion by using cluster analysis i.e. clustering the assignment of a set of observations into clusters so that observations in the same cluster may be in some sense be treated as similar. The outcome of the clustering process and the efficiency of its domain application are generally determined through algorithms. There are various algorithms which are used to solve this problem. In this research work two important clustering algorithms namely, centroid-based K-Means and representative object-based FCM (Fuzzy C-Means) clustering algorithms are compared. These algorithms are applied and performance is evaluated based on the efficiency of clustering output. The number of data points as well as the number of clusters are the factors upon which the behavior patterns of both algorithms are analyzed. FCM produces close results to K-Means clustering but it still requires more computation time than K-Means clustering.

**A Comparative Study Between Fuzzy Clustering Algorithm and Hard Clustering Algorithm**

Diya Jyoti Bora1 Dr. Anil Kumar Gupta2 1Department Of Computer Science and Applications, Baramulla University, Bhopal, India 2Department Of Computer Science and Applications, Baramulla University, Bhopal, India

Data clustering is an important area of data mining. This is an unsupervised study where data of similar types are put into one cluster while data of other types are put into different clusters. Fuzzy C means is a very important clustering technique based on fuzzy logic. Also, we have some hard-clustering techniques available like K-means among the popular ones. In this paper, a comparative study is done between the Fuzzy clustering algorithm and the hard clustering algorithm.

**Comparison of- Means and Fuzzy-Means Algorithms**

Ankita Singh, MCA Scholar  
Dr Prerenal Mahajan, Head of department, Institute of information technology and management

Clustering is the process of grouping feature vectors into classes in the self-organizing mode. Choosing cluster centres is crucial to the clustering. In this paper, we compared two fuzzy algorithms: the fuzzy c-means algorithm and the fuzzy k-means algorithm. Fuzzy c-means   
the algorithm uses the reciprocal of distances to decide the cluster centres. The representation reflects the distance   
of a feature vector from the cluster centre but does not differentiate the distribution of the clusters [1, 10, and 11]. The fuzzy k means algorithm in data mining, is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean [10,11].

IMPLEMENTATION

We have worked on two unsupervised machine learning algorithms: K-means and Fuzzy C means. After implementation of both on various datasets, we figured fuzzy C is preferable because using the fuzzy logic, gives us a “degree of belonging” of a data point to a particular dataset.

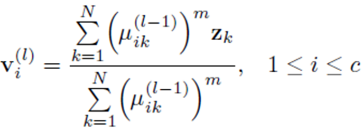
We used dummy data sets of 2 (X and Y) variables, and when our implementation of these algorithms gave correct results, we moved on to data from an actual blast furnace consisting of 26 and 36 variables whose number of classes couldn’t be found without some sort of machine learning. Hence, unsupervised algorithms were used.

Implementation of K-means:

1. Decide on a value for k where k is the number of classes.
2. Initialize the k cluster centres (randomly, if necessary).
3. Decide the class memberships of the N objects by assigning them to the nearest cluster centre.
4. Re-estimate the k cluster centers, by assuming the abovementioned memberships are correct.
5. Exit if none of the N objects changed membership in the last iteration. Otherwise, go to 3.

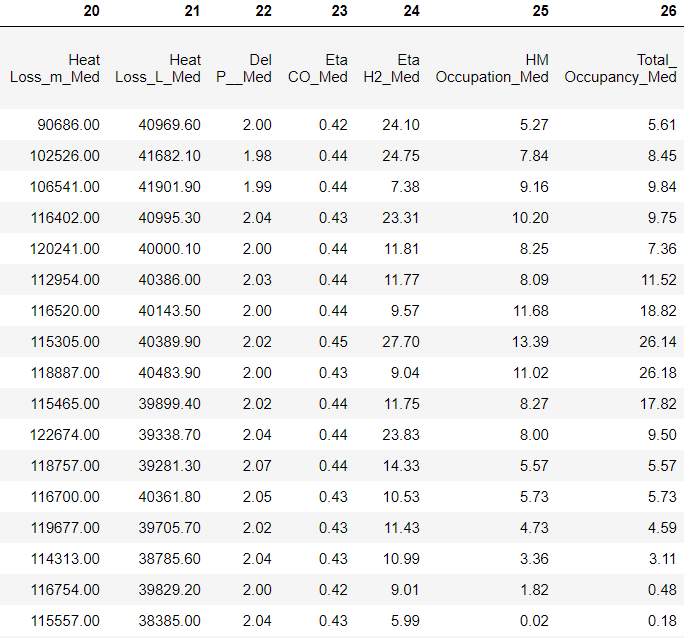
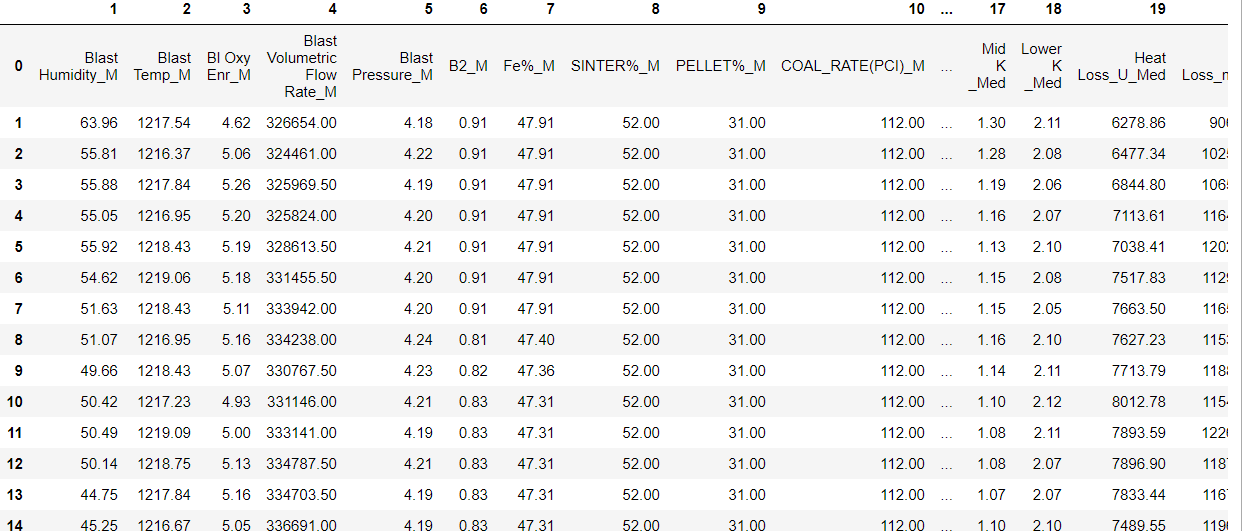
Implementation of Fuzzy C means:

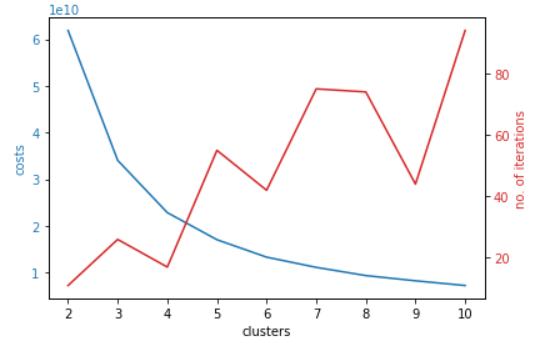
1. Read the data set *Z*, choose the number of clusters *c* as 1 < c < N, the fuzziness level *m > 0*, the error threshold *ε* and the norm-inducing matrix *A*
2. Initialize the partition matrix randomly, such that U(0) ∈ Mfc (U: Partition matrix)
3. Advance iteration loop, increment iteration number l
4. Compute the cluster means using

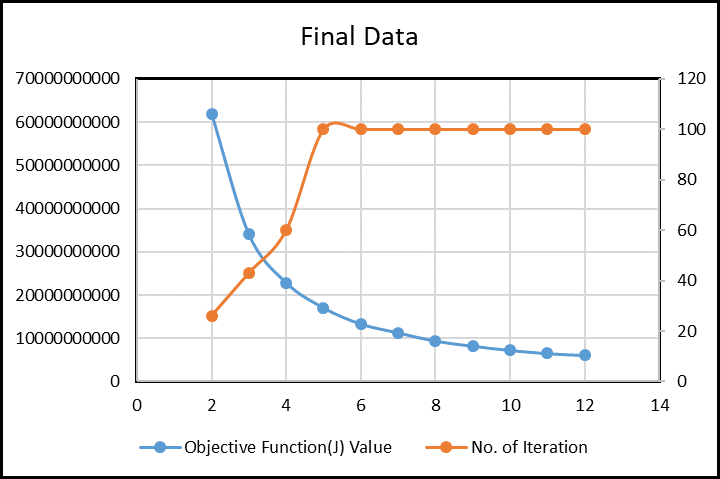


1. Compute the distances between the cluster and the data point.
2. Update the elements of the partition matrix.
3. Check. If True, Stop, and return to Step 2. *For calculation purposes, consider the value of this norm as the (abs of) largest element of this difference matrix*.

The final dataset used: (572X26)



Validation of results using Fuzzy C:



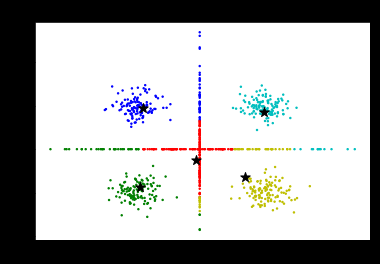
|  |
| --- |
| **Objective Function(J) Value** |
| 61900193928 |
| 34071528353 |
| 22865699212 |
| 17063589756 |
| 13325945967 |
| 11298188700 |
| 9377465372 |
| 8220807878 |
| 7235055952 |
|  |
|  |

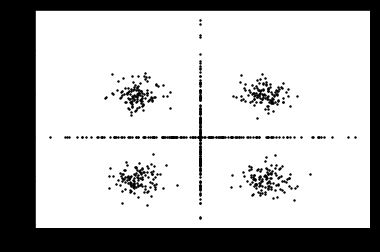
|  |
| --- |
| **Objective Function(J) Value** |
| 61900200851.14557 |
| 34071565859.667225 |
| 22865703712.293465 |
| 17063706917.322214 |
| 13325989217.903046 |
| 11135642195.70023 |
| 9377481269.285433 |
| 8256293029.997285 |
| 7250409097.810154 |
|  |
|  |

Matlab Results Fuzzy C results

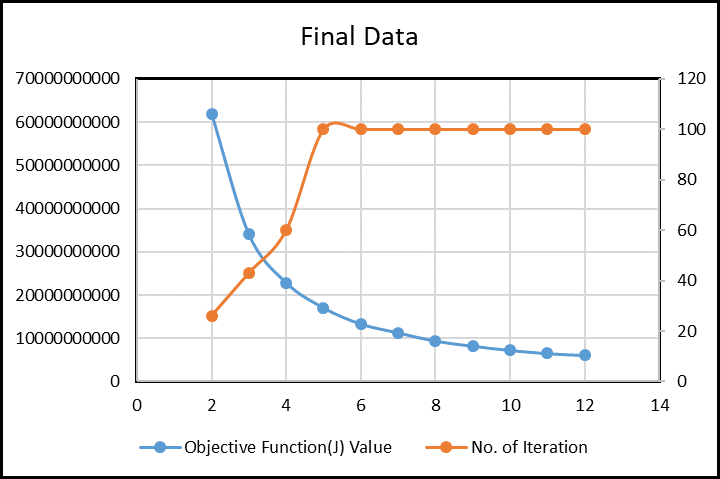
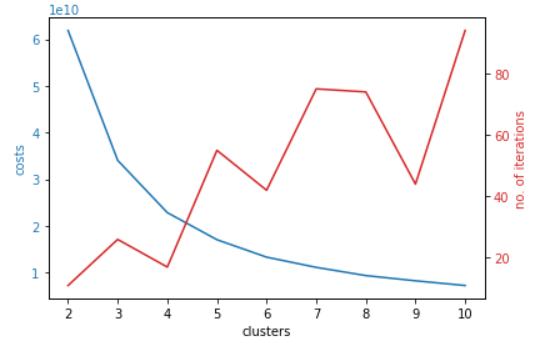
RESULTS

K-means results:





Fuzzy C means results:



CONCLUSIONS

We hope to implement the game theory-based expert system for optimising the states of the reactor and successfully use it in the real world. Also, the plan is to test the extremities of the fuzzy c means algorithm for such industrial data and work on the drawbacks, if any.

REFERENCES

1. Fuzzy C means K-means and Game Theory slides by Prof. Arya.
2. Book: Essentials of Game Theory by Kevin Leyton-Brown and Yoav Shoham.