# AN EFFICIENT ROBUST HYPERHEURISTIC CLUSTERING ALGORITHM

MOHAMMAD BABRDEL BONAB

UNIVERSITI TEKNOLOGI MALAYSIA

AN EFFICIENT ROBUST HYPERHEURISTIC CLUSTERING ALGORITHM

MOHAMMAD BABRDEL BONAB

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Computer Science)

Faculty of Computing
Universiti Teknologi Malaysia

MAY 2016

*To my beloved father, mother, wife and my son*

# ACKNOWLEDGMENTS

In preparing this thesis, I was in contact with many people, researchers, and academicians. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main thesis supervisor, PM. Dr. Siti Zaiton Mohd Hashim, for encouragement, guidance, and critics.

I am also indebted to Ministry of Science and Technology of Malaysia (MOSTI) and Universiti Teknologi Malaysia (UTM) for funding my Ph.D. study. Ministry of Science, Research and Technology of Iran also deserves special thanks for generous support of my Ph.D. study.

My fellow Malay and Iranian friends should also be recognized for their support. My sincere appreciation also extends to all my family, colleagues and others who have provided assistance at various occasions. Their views and tips are useful indeed. Unfortunately, it is not possible to list all of them in this limited space. I am grateful to all my family members.

## ABSTRACT

Observations on recent research of clustering problems illustrate that most of the approaches used to deal with these problems are based on meta-heuristic and hybrid meta-heuristic to improve the solutions. Hyperheuristic is a set of heuristics, meta- heuristics and high-level search strategies that work on the heuristic search space instead of solution search space. Hyperheuristics techniques have been employed to develop approaches that are more general than optimization search methods and traditional techniques. In the last few years, most studies have focused considerably on the hyperheuristic algorithms to find generalized solutions but highly required robust and efficient solutions. The main idea in this research is to develop techniques that are able to provide an appropriate level of efficiency and high performance to find a class of basic level heuristic over different type of combinatorial optimization problems. Clustering is an unsupervised method in the data mining and pattern recognition. Nevertheless, most of the clustering algorithms are unstable and very sensitive to their input parameters. This study, proposes an efficient and robust hyperheuristic clustering algorithm to find approximate solutions and attempts to generalize the algorithm for different cluster problem domains. Our proposed clustering algorithm has managed to minimize the dissimilarity of all points of a cluster using hyperheuristic method, from the gravity center of the cluster with respect to capacity constraints in each cluster. The algorithm of hyperheuristic has emerged from pool of heuristic techniques. Mapping between solution spaces is one of the powerful and prevalent techniques in optimization domains. Most of the existing algorithms work directly with solution spaces where in some cases is very difficult and is sometime impossible due to the dynamic behavior of data and algorithm. By mapping the heuristic space into solution spaces, it would be possible to make easy decision to solve clustering problems. The proposed hyperheuristic clustering algorithm performs four major components including selection, decision, admission and hybrid metaheuristic algorithm. The intensive experiments have proven that the proposed algorithm has successfully produced robust and efficient clustering results.

# ABSTRAK

Pemerhatian terhadap penyelidikan terkini berkaitan dengan masalah pengelompokan menunjukkan bahawa kebanyakan pendekatan yang menangani masalah ini menggunakan meta-heuristik dan hibrid meta-heuristik untuk menyelesaikan masalah tersebut. Hiperheuristik adalah satu set heuristik atau strategi carian peringkat tinggi yang berfungsi pada ruang carian heuristik dan bukannya ruang carian penyelesaian. Teknik hiperheuristik telah dibangunkam untuk membangunkan pendekatan yang lebih umum daripada kaedah carian pengoptimuman dan teknik tradisional yang biasa. Dalam beberapa tahun kebelakangan ini, kebanyakan kajian telah memberi tumpuan kepada algoritma hiperheuristik untuk mencari suatu algoritma hiperheuristik yang umum. Idea utama kajian ini adalah untuk membangunkan teknik yang dapat memberi tahap kecekapan dan prestasi yang sesuai dalam mencari suatu kelas tahap heuristik asas yang sesuai untuk pelbagai jenis masalah kombinasi pengoptimuman. Pengelompokan adalah satu kaedah tanpa pengawasan dalam pengumpulan data dan pengiktirafan corak. Walau bagaimanapun, sebahagian besar algoritma pengelompokan adalah kurang stabil dan sangat sensitif kepada parameter input. Kajian ini mencadangkan algoritma berkelompok hiperheuristik yang efisen dan teguh untuk mencari penyelesaian terbaik dan cuba menjadikannya algoritma umum untuk domain masalah kelompok yang berbeza. Tujuan pendekatan pengelompokan adalah untuk mengurangkan ketidaksamaan semua titik pada sesuatu kelompok dengan menggunakan kaedah hiperheuristik dari pusat graviti kelompok berkenaan dengan kekangan kapasiti dalam setiap kelompok. Pemetaan antara ruang adalah salah satu teknik yang hebat dan digunakan secara meluas dalam semua bidang saintifik, kebanyakan algoritma yang ada boleh bekerjasama dengan ruang yang ada di mana dalam situasi ini ianya amat sukar dan kebanyakannya agak mustahil untuk dilihat berdasarkan tingkahlaku data dan algoritma. Dengan menggunakan pengelompokan heuristik dalam penyelesaian ini, secara tidak langsung ianya memudahkan keputusan diambil untuk menyelesaikan masalah pengelompokan. Algoritma yang dicadangkan melakukan empat komponen utama termasuk mekanisma seleksi, keputusan, penerimaan dan algoritma hibrid meta heuristik. Eksperimen intensif yang dijalankan membuktikan algoritma yang dicadangkan berjaya menghasilkan keputusan pengkelompokan yang teguh dan efisen.

# TABLE OF CONTENTS

# LIST OF ALGORITHMS

# LIST OF FIGURES

## LIST OF TABLES

# LIST OF ABBREVIATIONS

| | | |
|---|---|---|
| GA | - | Genetic Algorithm |
| PSO | - | Particle Swarm Optimization |
| BA | - | Bees Algorithm |
| ABC | - | Artificial Bee Colony |
| HS | - | Harmony Search |
| SA | - | Simulated Annealing |
| DE | - | Differential Evolution |
| TS | - | Tabu Search |
| ACO | - | Ant colony optimization algorithms |
| HBMO | - | Honey-Bees Mating Optimization |
| ICA | - | Imperialist Competitive Algorithm |
| ACO-SA | - | Ant colony optimization-Simulated Annealing |
| PSO-SA | - | Particle swarm optimization-Simulated Annealing |
| H.H | - | HyperHeuristic |
| L.L.H | - | Low-Level Heuristic |
| NFE | - | Number of Function Evaluation |
| NBS | - | Number of Best Solution |
| NWS | - | Number of Worst Heuristic |
| EXE_TIME | - | Execution Time |
| CMC | - | Contraceptive Method Choice |
| UCI | - | University of California Irvine |
| CCIA | - | Cluster Center Initialization Algorithm |
| HHCA | - | HyperHeuristic Clustering Algorithm |
| CCIA-SAGA-K | - | Cluster Center Initialization Algorithm-Simulated Annealing and Genetic Algorithm with k-means |

# LIST OF APENDIX

# CHAPTER 1

# INTRODUCTION

## 1.1 An Overview

Clustering approaches have received attention in several study fields like biology, medicine, engineering and data analysing fields (Niknam, Taherian Fard *et al.* 2011). The main goal of clustering approaches are to collect data points. Clustering is the process of grouping data in similar groups. The k-means approach is one of the most widely-used clustering approach is one of main algorithms used for analysis of unsupervised data. However, the k-means algorithm results are depend on the initialization and converge towards the local optimum. In order to overcome obstacles due to local optimum, many studies have reported on clustering-related works (Wang, Zhang *et al.* 2007, Kao, Zahara *et al.* 2008, Niknam and Amiri 2010). This thesis presents a new and efficient hyperheuristic algorithm based on a proposed online genetic clustering learning method, thus advancing the heuristic selection method for optimum clustering solutions. The new hyperheuristic clustering algorithm (HHCA) was tested on different datasets and its performance was compared with several meta-heuristic algorithms such as Honey Bee Mating Optimization (HBMO), Simulated Annealing (SA), ant colony optimization (ACO), Tabu Search (TS), Artificial Bee Colony (ABC) particle swarm optimization (PSO), Genetic Algorithm (GA), and K-means algorithm (Wang, Zhang *et al.* 2007, Kao, Zahara *et al.* 2008, Kuo, Suryani *et al.* 2013).

For decades, a large quantity of raw data has been collected from various application areas such as health care systems, telecommunications, science and business (Dolnicar 2003, Bewley, Shekhar *et al.* 2011). The volume of such data has increased exponentially because of the widespread use of various technologically sophisticated devices for the gathering of scientific data from different fields. Many scientists have applied data mining techniques to explore large amounts of data instances in a wide-variety of applications for instances in scheduling and planning, finance, sales and marketing. However, several data mining tasks differ when used for various purposes.

Clustering is the process of categorizing unlabelled data according to their similarity. In cluster analysis, each class of data is called a ‘cluster’ and it consists of data instances which are similar within a cluster and dissimilar between other clusters (dissimilar between the objects of other groups and similar among themselves). As a result, clustering techniques are powerful exploratory approaches for the extraction of a pattern in the data. Many difficulties are encountered in general clustering techniques when it comes to the analysis of the data pattern due to the similarity measurement and the optimum cluster centres (Kao, Zahara *et al.* 2008). Hence, this work looked into improving the solutions by proposing a hyperheuristic algorithm.

## 1.2 Background of the Research

Clustering techniques are data analysis tools that are utilized for categorizing data with similar attributes. Cluster analysis has been applied in the data mining and machine learning tasks such as the unsupervised classification (Omran, Salman *et al.* 2006) and summation of data (Ng and Wong 2002) . The main objective in data clustering is to detect the natural categories of observations. Data clustering methods have been applied in several fields such as telecommunications networks, financial investments (fraud detection, credit card data, interest rates, stock prices and indexes), nuclear science, medicine (several diagnostic information), clustering of coals, local model development, discovery of classes in DNA dinucleotides, process monitoring, data compression and qualitative interpretation, analysis of chemical compounds, manufacturing (troubleshooting and process optimization) and radar scanning (Krishna and Murty 1999, Zhang, Wong *et al.* 1999, Maulik and Bandyopadhyay 2000, Sung and Jin 2000, Hee-Su and Sung-Bae 2001, Bandyopadhyay and Maulik

2002, Ng and Wong 2002, Shelokar, Jayaraman *et al.* 2004, Laszlo and Mukherjee 2006, Laszlo and Mukherjee 2007, Kao, Zahara *et al.* 2008, Nguyen and Cios 2008, Niknam, Firouzi *et al.* 2008, Žalik 2008, Firouzi, Sadeghi *et al.* 2010, Niknam and Amiri 2010, Zou, Zhu *et al.* 2010).

Generally, data clustering techniques have been used when large data need to be stored. Cluster analysis can be divided into partitional or hierarchical clustering. This study focused on partitional cluster analysis, and specifically, a popular and common partitional clustering technique known as the k-means algorithm. The k-means algorithm is a process of categorizing data into groups so that the objects in each class have a maximum similarity, while having a minimum dissimilarity with other classes. The dissimilarity is specifically based on the feature values of the objects. Distance measures are commonly utilized.

The k-means has its roots in several areas comprising image segmentation, machine learning, neural networks, statistics, and biology such as fraud detection, disease diagnosis, time series predictions, financial statement fraud, shareholder value predictions, traffic predictions, sensor networks (Bewley, Shekhar *et al.* 2011), business and marketing, medical imaging (Bewley and Upcroft 2013), analysis of antimicrobial activity, social network analysis, crime analysis, educational data mining, and mathematical chemistry (Basak, Magnuson *et al.* 1988, Kao, Zahara *et al.* 2008, Nguyen and Cios 2008, Žalik 2008). Despite significant improvements up to now in groups of data for a wide range of application domains, the k-means method still suffers from various disadvantages. The k-means objective function is not convex and it is confined to a local optimum.

As a result, there exists a possibility of trapping to local optima, in the minimization of the fitness function (Firouzi, Sadeghi *et al.* 2010). Consequently, the results of the k-means technique depend heavily on the initial state and initial cluster centres that are randomly selected.

To overcome these disadvantages, many clustering approaches, according to evolutionary algorithms for instance TS, BA, PSO, HBMO, SA, ABC and ACO have been presented. The Table 1.1 summarizes the previous researches related to the current research.

**Table 1.1:** Related works on clustering

| Clustering methods (Author / Year) | Summary | Future work / Limitation |
|---|---|---|
| "An Improved Animal Migration Optimization Algorithm for Clustering Analysis" (Ma, Luo *et al.* 2015) | "Propose a new evolutionay based algorithm based on the improved animal migration optimization to deal with clustering algorithm" | fall into local optima easily, sensitive to data behavior and no good for high dimensional datasets |
| "A Hybrid Monkey Search Algorithm for Clustering Analysis" (Chen, Zhou *et al.* 2014) | "Introduce an algorithm according to the monkey algorithm and artificial bee colony operator" | Sensitive to parameters and sensitive to noise and outliers, limited for use of heuristics |
| "Artificial Bee Colony algorithm, A novel clustering approach:" (Karaboga and Ozturk 2011) | "Propose a clustering algorithm inspired by foraging behaviour of a honey-bee swarm" | trapping into local optimum, senstive to initialization and parametrs |
| "An efficient hybrid algorithm according to modified ICA and K-means for data clustering" (Niknam, Taherian Fard *et al.* 2011) | "presents a new hybrid evolutionary algorithm according to K-means and modified ICA for clustering of data" | sensitive to noise and outliers, and parameters setting, limited to use of heuristic algorithm |
| "A hybridized approach to data clustering" (Kao, Zahara *et al.* 2008) | "A combined algorithm according to mixing Nelder–Mead simplex search, the K-means algorithm, and particle swarm optimization" | Problem on parameters setting, limited for use of heuristics, trapping into local optimum still exist |
| "Cluster center initialization algorithm for K-means clustering" (Khan and Ahmad 2004) | "Performance of iterative approaches that converges to numerous local optima depend highly on initial state intial centers" | Problem on finding outlier data, sensitive to parameters of algorithm and less efficiency and computational expensive |

The Figure 1.1, gives a summary of the current research. For instance, , Ma et al. proposed an improved algorithm for cluster analysis according to the Improved Animal Migration Optimization (IAMO) algorithm that uses a population updating process and a new migration process by organizing a living area to find optimum cluster centres. However, the performance and results of the Improved Animal Migration Optimization (IAMO) algorithm are greatly affected by the size of the living area. The Improved Animal Migration Optimization (IAMO) algorithm produces the best performance for the Animal Migration Optimization (AMO) algorithm but it suffers from several drawbacks in that it is sensitive to initialization (parameter setting), it cannot be used for high dimensional datasets, and it is sensitive to outliers and noisy data (Ma, Luo *et al.* 2015). Chen et al. introduced a combined clustering algorithm according to the monkey search algorithm and artificial bee colony (ABC) algorithm, which the algorithm uses the artificial-bee-colony search operator for the clustering of data. According to the simulation results, the algorithm gives a better performance than the basic monkey search algorithm for the solving of clustering problems, but suffers from sensitivity to parameters, noise and outliers, and limitations on the use of the heuristics (Chen, Zhou *et al.* 2014). Karaboga and Ozturk introduced an algorithm for the clustering of data based on the ABC algorithm that simulates the behavior of a swarm of honey bees (Karaboga and Ozturk 2011). The artificial bee-colony optimization method was presented by Karabogaa in 2005 (Karaboga 2005) for the optimization of numerical problems. However, the algorithm is hampered by initialization and parameter setting, and is easily affected by local optima. Niknaam et al. proposed a combined algorithm according to the k-means approach and a modified imperialist competitive algorithm (ICA). The article proposed a new mutation operator to improve the performance of the imperialist competitive method. The algorithm has several drawbacks such as premature convergence, falling into local minima, sensitivity to noise and outlier data, and is limited to the use of heuristics (Niknam, Taherian Fard *et al.* 2011). Kao et al. introduce a combined method based on a combination of the Nelder-Mead simplex search, partial swarm optimization and a genetic algorithm (Kao, Zahara *et al.* 2008). However, the algorithm is still subject to parameter adjustments, tapping into a local optimum, and is limited to the use of heuristics. Some approaches attempted to select the initial cluster centres appropriately through the use of certain tricks (Khan and Ahmad 2004). Khan and Ahmad proposed an approach for selecting the initial cluster centres because the performance of the iterative algorithm is highly dependent on the initial cluster centre in order to escape from falling into the local optimum. The algorithm is based on individual attributes and similar patterns. Some of the drawbacks of this algorithm are that it has a problem in finding outlier data, is sensitive to the parameters of the algorithm, is less efficient and is computationally expensive.

Most meta heuristic approaches such as Genetic Algorithm, Simulated Annealing, etc., are usually very slow in solving optimization problems. Recently, researchers have introduced new algorithms like BA, ABC, ACO and lately, a hybrid version of evolutionary algorithms (MICA, k-NM-PSO, ACO-PSO, etc.) has emerged in the search for optimum solutions, which not only produce better results in comparison with other evolutionary algorithms but also converge faster (Krishna and Murty 1999, Firouzi, Sadeghi *et al.* 2010).

However, the evolutionary-based algorithm (meta-heuristic and a hybrid of meta-heuristics) also suffers from several drawbacks including limited hybridization, sensitivity to data parameters, no routine approach to hybridization, sensitivity to random initialization, possibility of getting stuck in local optima, and sensitivity to the behaviour of algorithms.

To overcome these drawbacks, a robust clustering algorithm based on a hyperheuristic algorithm (HHCA) was proposed according to the performance of a population-based simulated annealing algorithm combined with a genetic clustering algorithm. The algorithm has been used in hyperheuristic algorithms to search in the heuristic space for an optimal and suitable low-level heuristic methods (Burke, Kendall *et al.* 2003, Misir 2012, Mısır, Verbeeck *et al.* 2013).

A hyperheuristic algorithm is a heuristic search algorithm which looks for an automated process, often by the inception of a machine learning strategy and a selection process to combine, generate and adapt several simple heuristics to solve computational search problems efficiently. The goal of a hyperheuristic algorithm is to reduce the domain knowledge in the search strategies (Ross, Marín-Blázquez *et al.* 2004, Bilgin, Özcan *et al.* 2007, Poli and Graff 2009, Qu and Burke 2009, Burke, Gendreau *et al.* 2013, Pillay 2013, Sabar, Ayob *et al.* 2015). The resulting method must be fast and cheap for implementing, should be robust enough to handle a wide range of problems from different types of domains and should require less expertise in the heuristic approach or problem domain.

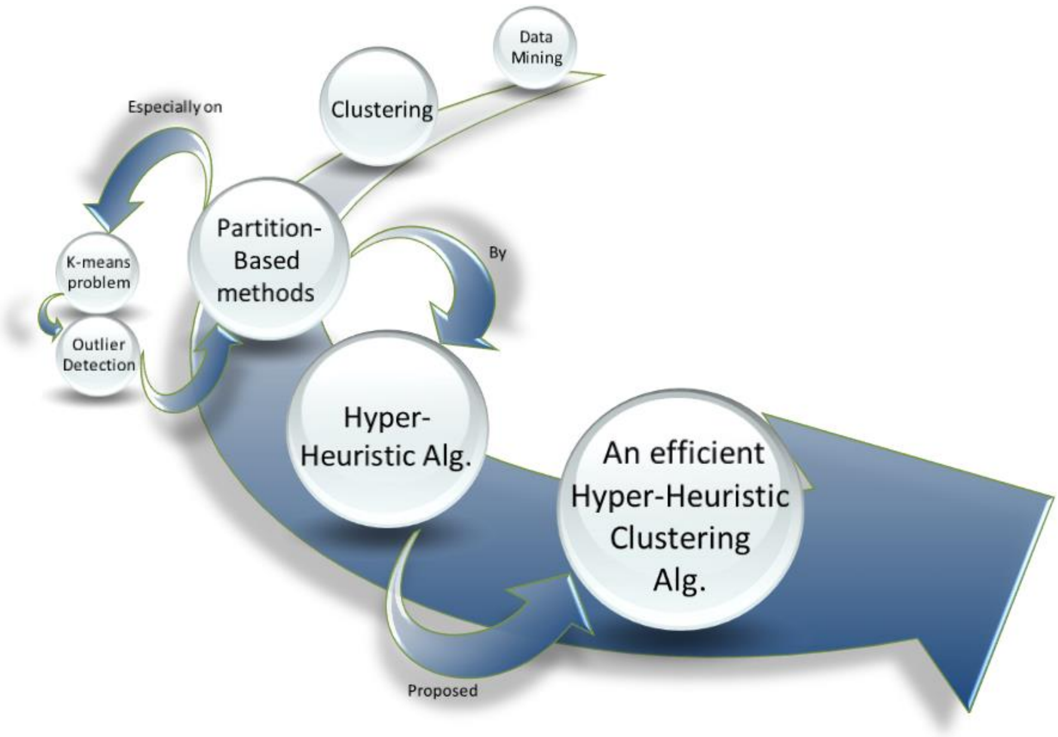


**Figure 1.1:** overview on research

Figure 1.1 gives an overview of the development of the algorithm and the steps taken in the current research. To develop the HHCA algorithm, some pre-requisites had to be taken into consideration. The first pre-requisite was a set of easy and non-parameterized low-level heuristics, which were used to search in the solution space and were placed in the heuristics pool.

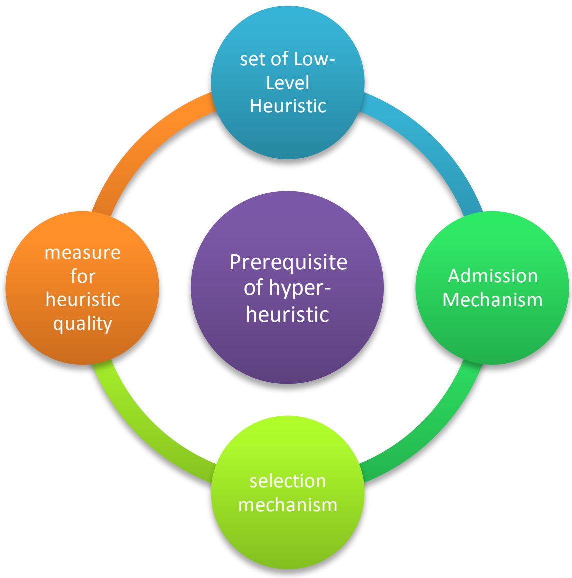

**Figure 1.2:** Prerequisite of hyperheuristic

The second prerequisite was to measure the quality of the heuristic in order to evaluate the low-level heuristics. The next prerequisite was to have a selection mechanism in the hyperheuristic algorithm that would be able to select a sequence of low-level heuristics that would make the greatest improvement on the solutions. The final prerequisite was to be able to move to acceptance in order to try to choose the most suitable and best solutions during the optimization process (Admission Mechanism). Figure 1.2 lists the prerequisites for the hyperheuristic in this research.

The proposed method incorporated four prerequisites: (1) the introduction of a new algorithm for the cluster analysis based on the hyperheuristic algorithm; (2) a modified learning algorithm based on the learning vector quantization (LVQ); (3) a proposed new acceptance scenario to accept newly discovered solutions; and (4) a proposed low-level heuristics to search within the solution domain.

## 1.3 Problem Statement

Three main problems were addressed in this study. The first problem is the limitations of meta-heuristic and hybrid meta-heuristic based clustering algorithms in the search for solutions within the solution space. It has been proven that existing meta-heuristic based clustering algorithms outperform traditional clustering algorithms, but these frequently have limitations, thus resulting in the use of several combinations of algorithms. This has made it necessary to have a hyperheuristic clustering algorithm without any limitations and with a dynamic section for the setting of parameters in order to increase the power of exploration and exploitation within the solution space.

The second problem is the absence of algorithm for interpreting and validating the heuristics during clustering process. In some cases, it is difficult to decide whether the used heuristic and its performance in one hybrid algorithm are good enough because the theories underlying some techniques are not very elaborate. In order to evaluate the performance of the heuristic algorithms used, a hyperheuristic clustering algorithm is used to achieve the optimum solutions and results.

## 1.4 Research Questions

The following research questions have been formulated in order to analyse the problems of clustering algorithms.

1. Which strategy (i.e. heuristic and Meta-heuristic algorithms, hybrid of meta-heuristics, and hyperheuristic algorithms) is appropriate for solving partitioning-based clustering problems?

2. Which criteria (i.e. execution time, number of function evaluations, number of new best solutions found) should be used to compare the heuristics?

3. Which selection method (i.e. elitist selection, random selection, tournament selection, and roulette wheel selection) is appropriate for selecting a suitable heuristic?

4. Which solution representation (i.e. continuous solution representation or discrete solution representation) is suitable for representing the solutions for the earlier mentioned problems?

5. Which model (i.e. dynamic programming, linear and non-linear programming) should be used to solve the earlier mentioned problems?

## 1.5 Aim of the Research

The aim of this study *was to propose a new, robust hyperheuristic clustering algorithm that can produce an efficient and high quality performance across various low-level heuristic sets in solving generic clustering problems in order to minimize the dissimilarity between all objects of a cluster from the centre of gravity of the cluster with respect to the capacity constraints in each cluster, such that each element is allocated to only one cluster (hard-clustering).* In addition, the purpose of this study was also to contribute to the combined meta-heuristic algorithms and hyperheuristic search algorithm to find *the optimum cluster centre* by minimizing the distance between the objects and the cluster centres, and *improving the scale of the clustering* on the *large dataset* and finding *the optimum results* for the model from the data.

## 1.6 Research Objectives

The objectives of this research were defined based on the literature review, background of the study, and the statement of the problem. The main objectives of the current research were as follows:

1. To propose efficient and robust hyperheuristic based on meta and heuristic algorithms by optimizing the initialization and setting of parameters adaptively.

2. To obtain optimum cluster centre by introducing low level heuristics to achieve better results for increasing the performance of hyperheuristic algorithm.

3. To validate stability and high performance of the proposed hyperheuristic clustering algorithm by identifying the optimum cluster centers using standard criteria.

## 1.7 Scope of the Research

This research was confined to the following scopes. The first scope was about the meta-heuristic and hyperheuristic algorithm, while the second scope was about the data clustering technique for this problem.

1. This study used the k-means algorithm and the partition-based clustering algorithm, which was used as a partitioning-based clustering algorithm.

2. The mixed and individual meta-heuristics were used in this study.

3. Evolutionary algorithms and clustering methods were applied for this problem (GA, PSO, BA, ABC, HS, SA , DE and K-means)

4. Nineteen low-level heuristic algorithms were used to deal with clustering problems, where seventeen of them were existing heuristics and two of them were proposed heuristics.

5. The standard case studies, artificial datasets and industrial images were used in order to validate the efficiency of the proposed methods and the standard datasets available from the UCI library.

## 1.8 Significance of the Research

Despite significant improvements in the analysis of data for a wide range of application areas up to now, these methodologies still need to be integrally merged and combined with other intelligence methods. Many experts from the fields of operational research, artificial-intelligence and computer science have acknowledged the need to develop automated systems to replace the roles of humans in such circumstances.

The goal of a hyperheuristic algorithm is to reduce the amount of domain knowledge by using the abilities of low-level heuristics and the capabilities of high-level heuristics simultaneously in the search strategies. The resulting method should be fast and cheap to implement, should be robust enough to handle a wide-range of problems from different types of domains and it require less expertise in either the heuristic approach or the problem domain. One of the aims of hyperheuristic algorithms is to increase the level of popularity of decision support strategies, perhaps at the expense of reduced solutions qualities when compared to tailor made meta-heuristic strategies. A robust hyperheuristic has been proposed in order to reduce the gap between hyperheuristic based methodologies and tailor-made designs.

In today's data environment, it seems important to minimize the similarity between clusters and to find the best representation for each cluster simultaneously in order to obtain high-quality results, and to increase the similarity at the same time. By implementing this approach, both of these goals (high quality results and maximum similarity) can be achieved and satisfied at the same time.

One of the most important motivations in studying hyperheuristics is to create and build systems that can handle group of problems instead of solving just one problem. Hyperheuristics use heuristics (or meta-heuristics) to choose heuristics (or meta-heuristics).

In hyperheuristics, the high-level approaches, depending upon the current state of the problem or the search conditions elects which low-level heuristic should be used at any given time. A hyperheuristic can generate new heuristics based on the used algorithms. Hyperheuristic methods can be categorized in to two most important classes, the first being *heuristics to choose heuristics*, while the second is *heuristics to generate heuristics*. In Figure 1.3 shows the summary of the justifications.

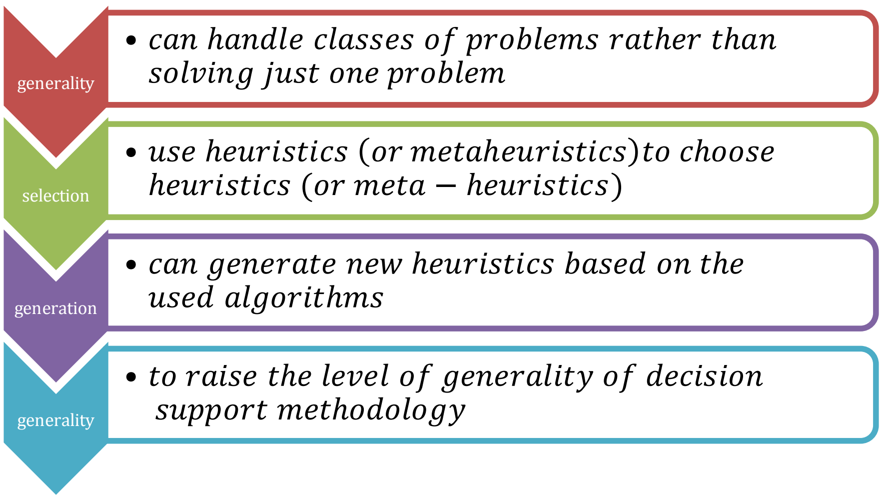


**Figure 1.3:** Summary of justifications

The proposed methods have been tested by various sample problems. In addition, it should be noted that calculated results showed the efficiency and capability of the proposed solutions. Although the use of a meta-heuristic algorithm for data clustering with the k-means clustering method takes into consideration the problem of sensitivity to initial values, yet the risk of getting trapped in local optimality threatens the algorithm. The hyperheuristic algorithm is a global optimization method that is appropriate for overcoming the mentioned problem. In this study, a proposed hyperheuristic method was developed by taking advantage of the low-level heuristics based on the proposed algorithm, in which the clustering of the data was selected properly.

## 1.9 Structure of the Thesis

This thesis consists of seven chapters, with the structure of the dissertation being given as follows:

Chapter 1: This is the introduction, which gives an overview of the development of the methods and techniques that are applied in cluster analysis, the background of the study and the common problems that are usually encountered in cluster analysis. It also consists of the problem statement, the research questions, the aims of the research, the research objectives, the scope of the research, and the significance of the research and the justification for the thesis.

Chapter 2: This is the literature review, which is made up of three main parts based on clustering, meta-heuristic and hyperheuristic algorithms that explored the concept of clustering methods, heuristic, meta-heuristic and hyperheuristic algorithms, the validation of clusters, and the interpretation and detection of optimum cluster centers. This chapter also contains a review of previous works related to clustering, meta-heuristic and hyperheuristic algorithms.

Chapter 3: This chapter presents the Research Methodology, which explains the approach that was taken to solve clustering problems, and gives a detailed description of the proposed hyperheuristic clustering algorithm. In addition, the experimental schemas and procedures are also discussed in this chapter.

Chapter 4: This chapter, titled ‘Proposed Hyperheuristic Clustering Algorithm and Hybrid Algorithms’, describes the basic and the main proposed algorithms, and gives a detailed description of the proposed hybrid and hyperheuristic clustering algorithms.

Chapter 5: This chapter presents an analysis of the results obtained on several datasets (i.e. artificial datasets and benchmark datasets) and image data (i.e. industrial and benchmark) with several criteria (i.e. accuracy, precision, F-measure, G-measure, variance of solutions, standard deviation, Rand index, etc.). This chapter also discusses in detail the simulation results for each dataset.

Chapter 6: This chapter, titled ‘Conclusions and Future Work’, provides a summary of the work, the contribution of the research, its extension and suggestions for future works, and the final remarks.

# REFERENCES


Abbasgholipour, M., M. Omid, A. Keyhani and S. S. Mohtasebi (2011). Color image segmentation with genetic algorithm in a raisin sorting system based on machine vision in variable conditions. *Expert Systems with Applications* 38(4), 3671-3678.

Afshar, A., O. Bozorg Haddad, M. A. Mariño and B. J. Adams (2007). Honey-bee mating optimization (HBMO) algorithm for optimal reservoir operation. *Journal of the Franklin Institute* 344(5), 452-462.

Ahmed, L. N., E. Özcan and A. Kheiri (2015). Solving high school timetabling problems worldwide using selection hyper-heuristics. *Expert Systems with Applications* 42(13), 5463-5471.

Al-Zoubi, M. d. B., A. Hudaib and B. Al-Shboul (2007). A fast fuzzy clustering algorithm. Proceedings of the 6th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases, Corfu Island, Greece.

Arthur, D. and S. Vassilvitskii (2007). k-means++: the advantages of careful seeding. Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. New Orleans, Louisiana, Society for Industrial and Applied Mathematics, 1027-1035.

Bai, R., J. Blazewicz, E. K. Burke, G. Kendall and B. McCollum (2011). A simulated annealing hyper-heuristic methodology for flexible decision support. *4OR* 10(1), 43-66.

Bai, R., E. K. Burke, G. Kendall and B. McCollum (2006). A simulated annealing hyper-heuristic for university course timetabling problem. *Abstract) PATAT* 6.

Bandyopadhyay, S. and U. Maulik (2001). Nonparametric genetic clustering: comparison of validity indices. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* 31(1), 120-125.

Bandyopadhyay, S. and U. Maulik (2002). An evolutionary technique based on K-Means algorithm for optimal clustering in RN. *Information Sciences* 146(1–4), 221-237.

Basak, S. C., V. Magnuson, G. Niemi and R. Regal (1988). Determining structural similarity of chemicals using graph-theoretic indices. *Discrete Applied Mathematics* 19(1), 17-44.

Basu, S., I. Davidson and K. Wagstaff (2008). Constrained clustering: Advances in algorithms, theory, and applications, CRC Press.

Berkhin, P. (2006). A Survey of Clustering Data Mining Techniques. Grouping Multidimensional Data. J. Kogan, C. Nicholas and M. Teboulle, Springer Berlin Heidelberg**:** 25-71.

Bewley, A., R. Shekhar, S. Leonard, B. Upcroft and P. Lever (2011). Real-time volume estimation of a dragline payload. Robotics and Automation (ICRA), 2011 IEEE International Conference on, IEEE.

Bewley, A. and B. Upcroft (2013). Advantages of Exploiting Projection Structure for Segmenting Dense 3D Point Clouds. Australian Conference on Robotics and Automation.

Bezdek, J. C. (1981). Pattern recognition with fuzzy objective function algorithms, Kluwer Academic Publishers.

Bezdek, J. C. and N. R. Pal (1998). Some new indexes of cluster validity. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* 28(3), 301-315.

Bianchi, L., M. Dorigo, L. Gambardella and W. Gutjahr (2009). A survey on metaheuristics for stochastic combinatorial optimization. *Natural Computing* 8(2), 239-287.

Bilgin, B., E. Özcan and E. E. Korkmaz (2007). An experimental study on hyper-heuristics and exam timetabling. Practice and Theory of Automated Timetabling VI, Springer: 394-412.

Blum, C. and A. Roli (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys (CSUR)* 35(3), 268-308.

Brucker, P. and S. Knust (2012). Resource-Constrained Project Scheduling. Complex Scheduling, Springer**:** 117-238.

Burke, E., M. Hyde, G. Kendall, G. Ochoa, E. Özcan and J. Woodward (2010). A Classification of Hyper-heuristic Approaches. Handbook of Metaheuristics. M. Gendreau and J.-Y. Potvin, Springer US. 146**:** 449-468.

Burke, E., G. Kendall, D. Landa Silva, R. O'Brien and E. Soubeiga (2005). An ant algorithm hyperheuristic for the project presentation scheduling problem. Evolutionary Computation, 2005. The 2005 IEEE Congress on.

Burke, E., G. Kendall, J. Newall, E. Hart, P. Ross and S. Schulenburg (2003). Hyper-heuristics: An emerging direction in modern search technology. *International series in operations research and management science*, 457-474.

Burke, E. K., M. Gendreau, M. Hyde, G. Kendall, G. Ochoa, E. Ozcan and R. Qu (2013). Hyper-heuristics: a survey of the state of the art. *J Oper Res Soc* 64(12), 1695-1724.

Burke, E. K., M. Hyde, G. Kendall and J. Woodward (2010). A Genetic Programming Hyper-Heuristic Approach for Evolving 2-D Strip Packing Heuristics. *Evolutionary Computation, IEEE Transactions on* 14(6), 942-958.

Burke, E. K. and G. Kendall (2005). Search methodologies, Springer.

Burke, E. K., G. Kendall and E. Soubeiga (2003). A tabu-search hyperheuristic for timetabling and rostering. *Journal of Heuristics* 9(6), 451-470.

Burke, E. K., S. Petrovic and R. Qu (2002). Case-based heuristic selection for examination timetabling. Proceedings of the SEAL, Citeseer.

Burke, E. K., J. D. L. Silva and E. Soubeiga (2005). Multi-objective hyper-heuristic approaches for space allocation and timetabling. Metaheuristics: Progress as Real Problem Solvers, Springer**:** 129-158.

Černý, V. (1985). Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of Optimization Theory and Applications* 45(1), 41-51.

Chen, X., Y. Zhou and Q. Luo (2014). A Hybrid Monkey Search Algorithm for Clustering Analysis. *The Scientific World Journal* 2014, 16.

Cheng, T. W., D. B. Goldgof and L. O. Hall (1998). Fast fuzzy clustering. *Fuzzy Sets and Systems* 93(1), 49-56.

Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 17(8), 790-799.

Chun-Wei, T., S. Huei-Jyun and C. Ming-Chao (2012). A hyper-heuristic clustering algorithm. Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on.

Cormen, T. H. (2009). Introduction to algorithms, MIT press.

Cowling, P., G. Kendall and H. Limin (2002). An investigation of a hyperheuristic genetic algorithm applied to a trainer scheduling problem. Evolutionary Computation, 2002. CEC '02. Proceedings of the 2002 Congress on.

Cowling, P., G. Kendall and E. Soubeiga (2001). A hyperheuristic approach to scheduling a sales summit. Practice and Theory of Automated Timetabling III, Springer**:** 176-190.

Cowling, P., G. Kendall and E. Soubeiga (2001). A Hyperheuristic Approach to Scheduling a Sales Summit. Practice and Theory of Automated Timetabling III. E. Burke and W. Erben, Springer Berlin Heidelberg. 2079**:** 176-190.

Cowling, P. and E. Soubeiga (2000). Neighborhood structures for personnel scheduling: a summit meeting scheduling problem. Proceedings of the 3rd International Conference on the Practice and Theory of Automated Timetabling, by EK Burke, W. Erben.

Davies, D. L. and D. W. Bouldin (1979). A cluster separation measure. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*(2), 224-227.

Denzinger, J., M. Fuchs and M. Fuchs (1996). High performance ATP systems by combining several AI methods, Citeseer.

Diday, E. (1987). The symbolic approach in clustering and relating methods of data analysis: The basic choices. 1. Conference of the International Federation of Classification Societies.

Diday, E. and J. Simon (1980). Clustering analysis. Digital Pattern Recognition, Springer**:** 47-94.

Diday, E. and J. C. Simon (1980). Clustering Analysis. Digital Pattern Recognition. K. Fu, Springer Berlin Heidelberg. 10**:** 47-94.

Dolnicar, S. (2003). Using cluster analysis for market segmentation-typical misconceptions, established methodological weaknesses and some recommendations for improvement.

Dowsland, K. A., E. Soubeiga and E. Burke (2007). A simulated annealing based hyperheuristic for determining shipper sizes for storage and transportation. European Journal of Operational Research 179(3), 759-774.

Dubes, R. C. (1987). How many clusters are best?-an experiment. Pattern Recognition 20(6), 645-663.

Dubes, R. C. (1987). How many clusters are best? - An experiment. Pattern Recognition 20(6), 645-663.

Dubes, R. C. (1993). Cluster analysis and related issues. Handbook of pattern recognition & computer vision. C. H. Chen, L. F. Pau and P. S. P. Wang, World Scientific Publishing Co., Inc. 3-32.

Dudoit, S. and J. Fridlyand (2002). A prediction-based resampling method for estimating the number of clusters in a dataset. Genome biology 3(7), research0036.

Dunn, J. C. (1973). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics* 3(3), 32-57.

Eckersley, A. (2007). Novel Knowledge Based and Heuristic Approaches to University Timetabling, Citeseer.

Ester, M., H.-P. Kriegel, J. Sander and X. Xu (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. KDD.

Fahim, A. M., A. M. Salem, F. A. Torkey and M. A. Ramadan (2006). An efficient enhanced k-means clustering algorithm. *Journal of Zhejiang University SCIENCE A* 7(10), 1626-1633.

Fathian, M. and B. Amiri (2008). A honeybee-mating approach for cluster analysis. *The International Journal of Advanced Manufacturing Technology* 38(7-8), 809-821.

Fathian, M., B. Amiri and A. Maroosi (2007). Application of honey-bee mating optimization algorithm on clustering. *Applied Mathematics and Computation* 190(2), 1502-1513.

Firouzi, B., M. S. Sadeghi and T. Niknam (2010). A new hybrid algorithm based on PSO, SA, and K-means for cluster analysis. *International journal of innovative computing, information and control* 6(7), 3177-3192.

Fisher, H. and G. L. Thompson (1963). Probabilistic learning combinations of local job-shop scheduling rules. *Industrial scheduling* 3, 225-251.

Fowlkes, E. B. and C. L. Mallows (1983). A method for comparing two hierarchical clusterings. *Journal of the American statistical association* 78(383), 553-569.

Furtuna, R., S. Curteanu and F. Leon (2012). Multi-objective optimization of a stacked neural network using an evolutionary hyper-heuristic. *Applied Soft Computing* 12(1), 133-144.

Gan, G., C. Ma and J. Wu (2007). Data clustering: theory, algorithms, and applications, Siam.

García-Martínez, C., M. Lozano and F. J. Rodríguez-Díaz (2012). A simulated annealing method based on a specialised evolutionary algorithm. *Applied Soft Computing* 12(2), 573-588.

Glover, F., G. Kochenberger and A. Gary (2003). Handbook of Metaheuristics, volume 57 of International Series in Operations Research & Management Science, Kluwer Academic Publishers.

Golberg, D. E. (1989). Genetic algorithms in search, optimization, and machine learning. *Addion wesley* 1989.

Gratch, J. and S. Chien (1996). Adaptive problem-solving for large-scale scheduling problems: A case study. *Journal of Artificial Intelligence Research*, 365-396.

Gratch, J., S. Chien and G. DeJong (1993). Learning search control knowledge for deep space network scheduling. Proceedings of the Tenth International Conference on Machine Learning.

Guha, S., N. Koudas, A. Marathe and D. Srivastava (2004). Merging the results of approximate match operations. Proceedings of the Thirtieth international conference on Very large data bases-Volume 30, VLDB Endowment.

Güngör, Z. and A. Ünler (2007). K-harmonic means data clustering with simulated annealing heuristic. *Applied Mathematics and Computation* 184(2), 199-209.

Güngör, Z. and A. Ünler (2008). K-Harmonic means data clustering with tabu-search method. *Applied Mathematical Modelling* 32(6), 1115-1125.

Gwiazda, T. D. (2006). Crossover for single-objective numerical optimization problems, Tomasz Gwiazda.

Halkidi, M., Y. Batistakis and M. Vazirgiannis (2001). On Clustering Validation Techniques. *Journal of Intelligent Information Systems* 17(2-3), 107-145.

Han, J., M. Kamber and J. Pei (2006). Data mining: concepts and techniques, Morgan kaufmann.

Han, L. and G. Kendall (2003). An investigation of a tabu assisted hyper-heuristic genetic algorithm. Evolutionary Computation, 2003. CEC'03. The 2003 Congress on, IEEE.

Haupt, R. L. and S. E. Haupt (2004). Practical genetic algorithms, John Wiley & Sons.

Hee-Su, K. and C. Sung-Bae (2001). An efficient genetic algorithm with less fitness evaluation by clustering. Evolutionary Computation, 2001. Proceedings of the 2001 Congress on.

Hruschka, E. R. and N. F. Ebecken (2003). A genetic algorithm for cluster analysis. *Intelligent Data Analysis* 7(1), 15-25.

Ippoliti, E. (2015). Reasoning at the frontier of knowledge: Introductory essay. Heuristic Reasoning, Springer**:** 1-10.

Jain, A. K., M. N. Murty and P. J. Flynn (1999). Data clustering: a review. *ACM Comput. Surv.* 31(3), 264-323.

Jain, K. and V. V. Vazirani (2001). Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and Lagrangian relaxation. *Journal of the ACM (JACM)* 48(2), 274-296.

Jiang-She, Z. and L. Yiu-Wing (2004). Improved possibilistic C-means clustering algorithms. *Fuzzy Systems, IEEE Transactions on* 12(2), 209-217.

Juang, L.-H. and M.-N. Wu (2010). MRI brain lesion image detection based on color-converted K-means clustering segmentation. *Measurement* 43(7), 941-949.

Juntao, W. and S. Xiaolong (2011). An improved K-Means clustering algorithm. Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on.

Kanungo, T., D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu (2002). An efficient k-means clustering algorithm: analysis and implementation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24(7), 881-892.

Kanungo, T., D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu (2004). A local search approximation algorithm for k-means clustering. *Computational Geometry* 28(2–3), 89-112.

Kao, Y.-T., E. Zahara and I. W. Kao (2008). A hybridized approach to data clustering. *Expert Systems with Applications* 34(3), 1754-1762.

Karaboga, D. (2005). An idea based on honey bee swarm for numerical optimization, Technical report-tr06, Erciyes university, engineering faculty, computer engineering department.

Karaboga, D. and C. Ozturk (2011). A novel clustering approach: Artificial Bee Colony (ABC) algorithm. *Applied Soft Computing* 11(1), 652-657.

Khan, S. S. and A. Ahmad (2004). Cluster center initialization algorithm for K-means clustering. *Pattern Recognition Letters* 25(11), 1293-1302.

Kirkpatrick, S., D. G. Jr. and M. P. Vecchi (1983). Optimization by simmulated annealing. *science* 220(4598), 671-680.

Kiss, O. (2006). Heuristic, methodology or logic of discovery? Lakatos on patterns of thinking. *Perspectives on Science* 14(3), 302-317.

Krishna, K. and M. N. Murty (1999). Genetic K-means algorithm. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* 29(3), 433-439.

Krishnapuram, R. and J. M. Keller (1993). A possibilistic approach to clustering. *Fuzzy Systems, IEEE Transactions on* 1(2), 98-110.

Kuo, R. J., E. Suryani and A. Yasid (2013). Automatic Clustering Combining Differential Evolution Algorithm and k-Means Algorithm. Proceedings of the Institute of Industrial Engineers Asian Conference 2013. Y.-K. Lin, Y.-C. Tsao and S.-W. Lin, Springer Singapore**:** 1207-1215.

Kuo, R. J., H. S. Wang, T.-L. Hu and S. H. Chou (2005). Application of ant K-means on clustering analysis. *Computers & Mathematics with Applications* 50(10–12), 1709-1724.

Lange, T., V. Roth, M. L. Braun and J. M. Buhmann (2004). Stability-based validation of clustering solutions. *Neural computation* 16(6), 1299-1323.

Laszlo, M. and S. Mukherjee (2006). A genetic algorithm using hyper-quadtrees for low-dimensional k-means clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 28(4), 533-543.

Laszlo, M. and S. Mukherjee (2007). A genetic algorithm that exchanges neighboring centers for k-means clustering. *Pattern Recognition Letters* 28(16), 2359-2366.

Lehrbaum, A. and N. Musliu (2012). A New Hyperheuristic Algorithm for Cross-Domain Search Problems. Learning and Intelligent Optimization. Y. Hamadi and M. Schoenauer, Springer Berlin Heidelberg**:** 437-442.

Levine, E. and E. Domany (2001). Resampling method for unsupervised estimation of cluster validity. *Neural computation* 13(11), 2573-2593.

Liu, B., L. Wang and Y.-H. Jin (2008). An effective hybrid PSO-based algorithm for flow shop scheduling with limited buffers. *Computers & Operations Research* 35(9), 2791-2806.

Ma, M., Q. Luo, Y. Zhou, X. Chen and L. Li (2015). An Improved Animal Migration Optimization Algorithm for Clustering Analysis. *Discrete Dynamics in Nature and Society* 2015, 12.

Maulik, U. and S. Bandyopadhyay (2000). Genetic algorithm-based clustering technique. *Pattern Recognition* 33(9), 1455-1465.

Meilă, M. (2003). Comparing Clusterings by the Variation of Information. Learning Theory and Kernel Machines. B. Schölkopf and M. Warmuth, Springer Berlin Heidelberg. 2777**:** 173-187.

Michalski, R. S. and R. E. Stepp (1983). Automated Construction of Classifications: Conceptual Clustering Versus Numerical Taxonomy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* PAMI-5(4), 396-410.

Mirkin, B. (1998). Mathematical classification and clustering: From how to what and why, Springer.

Misir, M. (2012). Intelligent hyper-heuristics: a tool for solving generic optimisation problems. *status: published*.

Mısır, M., K. Verbeeck, P. Causmaecker and G. Berghe (2013). A new hyper-heuristic as a general problem solver: an implementation in HyFlex. *Journal of Scheduling* 16(3), 291-311.

Misir, M., K. Verbeeck, P. De Causmaecker and G. Vanden Berghe (2011). A new hyper-heuristic implementation in HyFlex: a study on generality. Proceedings of the 5th Multidisciplinary International Scheduling Conference: Theory & Application.

Misir, M., K. Verbeeck, P. De Causmaecker and G. Vanden Berghe (2012). Design Principles and Performance Analysis of a Selection Hyper-heuristic across Multiple Problem Domains. *status: accepted*.

Mısır, M., K. Verbeeck, P. De Causmaecker and G. Vanden Berghe (2012). An Intelligent Hyper-Heuristic Framework for CHeSC 2011. Learning and Intelligent Optimization. Y. Hamadi and M. Schoenauer, Springer Berlin Heidelberg**:** 461-466.

Mu-Chun, S. and C. Chien-Hsing (2001). A modified version of the K-means algorithm with a distance based on cluster symmetry. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 23(6), 674-680.

Ng, M. K. and J. C. Wong (2002). Clustering categorical data sets using tabu search techniques. *Pattern Recognition* 35(12), 2783-2790.

Nguyen, C. D. and K. J. Cios (2008). GAKREM: A novel hybrid clustering algorithm. *Information Sciences* 178(22), 4205-4227.

Nie, F., T. Tu, M. Pan, Q. Rong and H. Zhou (2012). K-Harmonic Means Data Clustering with PSO Algorithm. Advances in Electrical Engineering and Automation. A. Xie and X. Huang, Springer Berlin Heidelberg. 139**:** 67-73.

Niknam, T. (2009). An efficient hybrid evolutionary algorithm based on PSO and HBMO algorithms for multi-objective Distribution Feeder Reconfiguration. *Energy Conversion and Management* 50(8), 2074-2082.

Niknam, T. and B. Amiri (2010). An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis. *Applied Soft Computing* 10(1), 183-197.

Niknam, T., B. Amiri, J. Olamaei and A. Arefi (2009). An efficient hybrid evolutionary optimization algorithm based on PSO and SA for clustering. *Journal of Zhejiang University SCIENCE A* 10(4), 512-519.

Niknam, T., B. B. Firouzi and M. Nayeripour (2008). An efficient hybrid evolutionary algorithm for cluster analysis. World Applied Sciences Journal, Citeseer.

Niknam, T., E. Taherian Fard, N. Pourjafarian and A. Rousta (2011). An efficient hybrid algorithm based on modified imperialist competitive algorithm and K-means for data clustering. *Engineering Applications of Artificial Intelligence* 24(2), 306-317.

Omran, M. G., A. Salman and A. P. Engelbrecht (2006). Dynamic clustering using particle swarm optimization with application in image segmentation. *Pattern Analysis and Applications* 8(4), 332-344.

Osman, I. H. and N. Christofides (1994). Capacitated clustering problems by hybrid simulated annealing and tabu search. *International Transactions in Operational Research* 1(3), 317-336.

Ozcan, E., B. Bilgin and E. E. Korkmaz (2008). A comprehensive analysis of hyper-heuristics. *Intelligent Data Analysis* 12(1), 3.

Özcan, E., M. Mısır, G. Ochoa and E. K. Burke (2012). A Reinforcement Learning: Great-Deluge Hyper-Heuristic. *Modeling, Analysis, and Applications in Metaheuristic Computing: Advancements and Trends: Advancements and Trends*, 34.

Pal, S. K. and D. D. Majumder (1977). Fuzzy sets and decision making approaches in vowel and speaker recognition. *IEEE Transactions on Systems, Man, and Cybernetics* 7(8), 625-629.

Papadimitriou, C. H. and K. Steiglitz (1982). Combinatorial optimization: algorithms and complexity, Prentice-Hall, Inc.

Patané, G. and M. Russo (2001). The enhanced LBG algorithm. *Neural Networks* 14(9), 1219-1237.

Pearl, J. (1984). Heuristics: intelligent search strategies for computer problem solving.

Pillay, N. (2012). Evolving hyper-heuristics for the uncapacitated examination timetabling problem. *Journal of the Operational Research Society* 63(1), 47-58.

Pillay, N. (2012). A study of evolutionary algorithm selection hyper-heuristics for the one-dimensional bin-packing problem. *South African Computer Journal* 48, 31-40.

Pillay, N. (2013). A comparative study of hyper-heuristics for solving the school timetabling problem. Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference, ACM.

Poli, R. and M. Graff (2009). There Is a Free Lunch for Hyper-Heuristics, Genetic Programming and Computer Scientists. Genetic Programming. L. Vanneschi, S. Gustafson, A. Moraglio, I. De Falco and M. Ebner, Springer Berlin Heidelberg. 5481**:** 195-207.

Qu, R. and E. K. Burke (2009). Hybridizations within a graph-based hyper-heuristic framework for university timetabling problems. *Journal of the Operational Research Society* 60(9), 1273-1285.

Ramos, V. and F. Muge (2004). Image colour segmentation by genetic algorithms. *arXiv preprint cs/0412087*.

Rattadilok, P., A. Gaw and R. S. Kwan (2005). Distributed choice function hyper-heuristics for timetabling and scheduling. Practice and Theory of Automated Timetabling V, Springer**:** 51-67.

Rivest, R. L. and C. E. Leiserson (1990). Introduction to algorithms, McGraw-Hill, Inc.

Ross, P., J. G. Mar, #237, n-Bl, #225, zquez, S. Schulenburg and E. Hart (2003). Learning a procedure that can solve hard bin-packing problems: a new GA-based approach to hyper-heuristics. Proceedings of the 2003 international conference on Genetic and evolutionary computation: PartII. Chicago, IL, USA, Springer-Verlag, 1295-1306.

Ross, P., J. G. Marín-Blázquez and E. Hart (2004). Hyper-heuristics applied to class and exam timetabling problems. Evolutionary Computation, 2004. CEC2004. Congress on, IEEE.

Ryser-Welch, P. and J. F. Miller (2014). A review of hyper-heuristic frameworks. Proceedings of the Evo20 Workshop, AISB.

Sabar, N. R., M. Ayob, G. Kendall and Q. Rong (2015). Automatic Design of a Hyper-Heuristic Framework With Gene Expression Programming for Combinatorial Optimization Problems. *Evolutionary Computation, IEEE Transactions on* 19(3), 309-325.

Schiminovich, S. (1971). Automatic classification and retrieval of documents by means of a bibliographic pattern discovery algorithm. *Information Storage and Retrieval* 6(6), 417-435.

Segura, C., G. Miranda and C. León (2011). Parallel hyperheuristics for the frequency assignment problem. *Memetic Computing* 3(1), 33-49.

Shelokar, P. S., V. K. Jayaraman and B. D. Kulkarni (2004). An ant colony approach for clustering. *Analytica Chimica Acta* 509(2), 187-195.

Shin, K. and A. Abraham (2006). Two Phase Semi-supervised Clustering Using Background Knowledge. Intelligent Data Engineering and Automated Learning – IDEAL 2006. E. Corchado, H. Yin, V. Botti and C. Fyfe, Springer Berlin Heidelberg. 4224**:** 707-712.

Shouqiang, W., Z. Daming and Z. Sheng (2008). Randomized algorithm with constant approximation for k-means based on the least cluster size. Intelligent Control and Automation, 2008. WCICA 2008. 7th World Congress on.

Song, M. and S. Rajasekaran (2005). Fast k-Means Algorithms with Constant Approximation. Algorithms and Computation. X. Deng and D.-Z. Du, Springer Berlin Heidelberg. 3827**:** 1029-1038.

Sörensen, K. (2015). Metaheuristics—the metaphor exposed. *International Transactions in Operational Research* 22(1), 3-18.

Storer, R. H., S. D. Wu and R. Vaccari (1992). New search spaces for sequencing problems with application to job shop scheduling. *Management science* 38(10), 1495-1509.

Sun, L.-X., F. Xu, Y.-Z. Liang, Y.-L. Xie and R.-Q. Yu (1994). Cluster analysis by the K-means algorithm and simulated annealing. *Chemometrics and Intelligent Laboratory Systems* 25(1), 51-60.

Sung, C. S. and H. W. Jin (2000). A tabu-search-based heuristic for clustering. *Pattern Recognition* 33(5), 849-858.

Talbi, E.-G. (2009). Metaheuristics: from design to implementation, John Wiley & Sons.

Terashima-Marín, H. and P. Ross (1999). Evolution of constraint satisfaction strategies in examination timetabling. Proceedings of the Genetic and Evolutionary Computation Conference (GECCO99, Citeseer.

Theodoridis, S., A. Pikrakis, K. Koutroumbas and D. Cavouras (2010). Introduction to Pattern Recognition: A Matlab Approach: A Matlab Approach, Access Online via Elsevier.

Tibshirani, R. and G. Walther (2005). Cluster validation by prediction strength. *Journal of Computational and Graphical Statistics* 14(3), 511-528.

Van Leeuwen, J. (1990). Handbook of theoretical computer science: Algorithms and complexity, Elsevier.

Vazirani, V. V. (2013). Approximation algorithms, Springer Science & Business Media.

Wagstaff, K., C. Cardie, S. Rogers, S. Schr, #246 and dl (2001). Constrained K-means Clustering with Background Knowledge. Proceedings of the Eighteenth International Conference on Machine Learning, Morgan Kaufmann Publishers Inc., 577-584.

Wang, Y.-J., J.-S. Zhang and G.-Y. Zhang (2007). A dynamic clustering based differential evolution algorithm for global optimization. *European Journal of Operational Research* 183(1), 56-73.

Wegener, I. (2005). Complexity theory: exploring the limits of efficient algorithms, Springer Science & Business Media.

Wilson, D. R. and T. R. Martinez (1997). Improved heterogeneous distance functions. *arXiv preprint cs/9701101*.

Wolpert, D. H. and W. G. Macready (1997). No free lunch theorems for optimization. *Trans. Evol. Comp* 1(1), 67-82.

Xu, J., Q. Lu and Z. Liu (2012). Combining classification with clustering for web person disambiguation. Proceedings of the 21st international conference companion on World Wide Web. Lyon, France, ACM, 637-638.

Yager, R. R. and D. P. Filev (1994). Approximate clustering via the mountain method. *Systems, Man and Cybernetics, IEEE Transactions on* 24(8), 1279-1284.

Yang, F., T. Sun and C. Zhang (2009). An efficient hybrid data clustering method based on K-harmonic means and Particle Swarm Optimization. *Expert Systems with Applications* 36(6), 9847-9852.

Yao, H., Q. Duan, D. Li and J. Wang (2013). An improved -means clustering algorithm for fish image segmentation. *Mathematical and Computer Modelling* 58(3–4), 790-798.

Yusoff, I. A. and N. A. M. Isa (2011). Two-dimensional clustering algorithms for image segmentation. *WSEAS transactions on computers* 10(10), 332-342.

Žalik, K. R. (2008). An efficient k′-means clustering algorithm. *Pattern Recognition Letters* 29(9), 1385-1391.

Zhang, C., P. M. Wong and O. Selinus (1999). A comparison of outlier detection methods: exemplified with an environmental geochemical dataset. Neural Information Processing, 1999. Proceedings. ICONIP '99. 6th International Conference on.

Zou, W., Y. Zhu, H. Chen and X. Sui (2010). A Clustering Approach Using Cooperative Artificial Bee Colony Algorithm.

## LIST OF PUBLICATIONS

1. Mohammad Babrdel Bonab, Siti Zaiton Mohd Hashim, Nor Erne Nazira Bazin,and Ahmed Khalaf Zager Alsaedi (2015). An Effective Hybrid of Bees Algorithm and Differential Evolution Algorithm in Data Clustering. *Mathematical Problems in Engineering, Hindawi*. **(Published, ISI and SCOPUS Indexed, Impact Factor: 1.082)**.

2. Mohammad Babrdel Bonab, Siti Zaiton Mohd Hashim, Ahmed Khalaf Zager Alsaedi, and Ummi Raba'ah Hashim (2015). Modified K-means Combined with Artificial Bee Colony Algorithm and Differential Evolution for Color Image Segmentation. *Advances in Intelligent Systems and Computing, Springer*. **(Published, Book Series, ISI and SCOPUS Indexed)**.

3. Mohammad Babrdel bonab, Siti Zaiton Mohd Hashim, Nor Erne Nazira Bazin (2014). Data Analysis by Combining the Modified K-Means and Imperialist Competitive Algorithm. *Jurnal Teknologi, Penerbit Universiti Teknologi Malaysia*. **(Published, SCOPUS Indexed)**.

4. Mohammad Babrdel bonab, and Siti Zaiton Mohd Hashim (2015). Image Segmentation with Genetic Clustering Using Weighted Combination of Particle Swarm Optimization. *Proceedings of the 14th International Conference on Applied Computer and Applied Computational Science (ACACOS '15),* Kuala Lumpur, Malaysia, April 23-25, 2015. **(Published, SCOPUS Indexed)**.

5. Mohammad Babrdel bonab, and Siti Zaiton Mohd Hashim (2014). Improved k-means clustering with Harmonic-Bee algorithms. *Proceedings of the Fourth World Congress on Information and Communication Technologies (WICT*

*2014)*, Malacca, Malaysia, December 08-10, 2014. **(Published, SCOPUS Indexed, BEST STUDENT PAPER AWARD)**.

6. Mohammad Babrdel Bonab, Siti Zaiton Mohd Hashim, Ahmed Khalaf Zager Alsaedi, and Ummi Raba'ah Hashim (2014). Modified K-means Combined with Artificial Bee Colony Algorithm and Differential Evolution for Color Image Segmentation. *Proceedings of the 4th International Neural Network Society Symposia Series (INNS-CIIS 2014)*, Bandar Seri Begawan, Brunei, 7-9 November 2014. **(Published, SCOPUS Indexed)**.

7. Mohammad Babrdel bonab, and Siti Zaiton Mohd Hashim (2014). Population based simulated annealing to solving the kmeans problem combined with genetic algorithm. *Proceedings of the 2014 IEEE Symposium on Computers & Informatics (ISCI 2014)*, Kota Kinabalu, Malaysia, 28-29 September 2014. **(Published, SCOPUS Indexed)**.

8. Mohammad Babrdel bonab, and Nor Erne Nazira Bazin (2013). Data Clustering by Combining the K-means and ICA Algorithms. *Proceedings of 4th International Graduate Conference on Engineering, Science and Humanities (IGCESH 2013)*, Johor Bahru, Malaysia, 16-7 April 2013. **(Published)**

9. Ummi Raba'ah Hashim, Azah Kamilah Muda, Siti Zaiton Hashim, Mohammad Babrdel Bonab (2013). Rotation Invariant Texture Feature Based on Spatial Dependence Matrix for Timber Defect Detection. *Proceeding of the 13th International Conference on Intelligent Systems Design and Applications (ISDA 2013)*, Bangi, Malaysia, 8-10 December 2013. **(Published, SCOPUS Indexed)**.

10. Mohammad Babrdel bonab, and Siti Zaiton Mohd Hashim (2014). An efficient hybrid approach based on simulated annealing with genetic clustering to solving the k-means problem. Paper presented at PARS2014 -

Soft Computing Seminar (SoCoS 2014), UTM, Skudai, Malaysia, 24-25 June 2014.