FEATURE SELECTION AND ENHANCED KRILL HERD ALGORITHM FOR TEXT DOCUMENT CLUSTERING

LAITH MOHAMMAD QASIM ABUALIGAH

UNIVERSITI SAINS MALAYSIA

2018

FEATURE SELECTION AND ENHANCED KRILL HERD ALGORITHM FOR TEXT DOCUMENT CLUSTERING

by

LAITH MOHAMMAD QASIM ABUALIGAH

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

March 2018

ACKNOWLEDGEMENT

Beginning this Ph.D. thesis has been a life-changing experience for me and I would not have achieved it without the guidance and support of many people. I must declare many external donations from individuals who extended a helping hand throughout this study.

I am thankful to Allah SWT for giving me strength to finish this study. I am also grateful to my supervisor, Professor Dr. Ahamad Tajudin Khader from the School of Computer Sciences at Universiti Sains Malaysia, for his wise counsel, helpful advice, connected support and supervision throughout the duration of this study. I am also thankful to my co-supervisor, Dr. Mohammed Azmi Al-Betar from the Department of Information Technology, Al-Huson University College, Al-Balqa Applied University, for his assistance. I am also thankful to Dr. Essam Said Hanandeh from the Department of Computer Information System, Zarqa University, for his assistance.

My family deserves special thanks. Words cannot express how grateful I am to my father, mother, and brothers for all of the sacrifices that they have done for me. Finally, I thank all of my friends who encouraged me throughout the duration of this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABC	Ant Colony Optimization
ASDC	Average Similarity of Documents Centroid
BCO	Bee Colony Optimization
BKHA	Basic Krill Herd Algorithm
BPSO	Binary Particle Swam Optimization
CS	Cuckoo Search
DDF	Detailed Document Frequency
DDR	Detailed Dimension Reduction
DF	Document Frequency
DFTF	Document Frequency with Term Frequency
DR	Dimension Reduction
DTF	Detailed Term Frequency
FE	Feature Extraction
FF	Fitness Function
FS	Feature Selection
GA	Genetic Algorithm
HKHA	Hybrid Krill Herd Algorithm
HS	Harmony Search
IDF	Inverse Document Frequency
KH	Krill Herd
KHA	Krill Herd Algorithm
KHM	Krill Herd Memory
KI	Krill Individual
LFW	Length Feature Weight
MKHA	Modified Krill Herd Algorithm
NLP	Natural Language Processing
NP	Non-deterministic Polynomial-time
PSO	Particle Swam Optimization

TC Text Clustering

TD Text Document

TDCP Text document clustering problem

TF Term Frequency

TFSP Text feature selection problem

VSM Vector Space Model

WTDC Web Text Documents Clustering

PEMILIHAN FITUR DAN ALGORITMA KRILL HERD LANJUTAN UNTUK PENGKLUSTERAN DOKUMEN TEKS

ABSTRAK

Pengklusteran dokumen teks adalah satu tren baru dalam galian teks di mana dokumen-dokumen diasingkan kepada beberapa kluster yang koheren, di mana dokumen-dokumen dalam kluster yang sama adalah serupa. Dalam kajian ini, satu kaedah baru untuk menyelesaikan masalah pengklusteran dokumen teks dijalankan dalam dua peringkat: (i) Satu kaedah pemilihan fitur menggunakan algoritma optima kumpulan partikel dengan satu skima pemberat yang baru dan satu teknik pengurangan dimensi yang lengkap dicadangkan untuk mendapatkan satu subset baru fitur-fitur yang lebih bermaklumat dengan ruang berdimensi rendah. Subset baru ini digunakan untuk memperbaiki prestasi algoritma pengklusteran teks dalam peringkat berikutnya dan ini mengurangan masa pengiraannya. Algoritma pengklusteran min-k digunakan untuk menilai keberkesanan subset-subset yang diperolehi. (ii) Empat algoritma krill herd iaitu (a) algoritma krill herd asas, (b) algoritma krill herd yang telah diubahsuai, (c) algoritma krill herd hibrid, dan (d) algoritma hibrid pelbagai objektif krill herd, disarankan untuk menyelesaikan masalah pengklusteran teks; algoritma ini adalah penambahbaikan lanjutan kepada versi-versi yang terdahulu. Untuk proses penilaian, tujuh set data teks penanda aras digunakan dengan pencirian dan kesukaran yang berbeza. Keputusan menunjukkan bahawa kaedah yang dicadangkan dan algoritma yang diperolehi mencapai keputusan terbaik berbanding dengan kaedah-kaedah lain yang diutarakan dalam literatur.

FEATURE SELECTION AND ENHANCED KRILL HERD ALGORITHM FOR TEXT DOCUMENT CLUSTERING

ABSTRACT

Text document (TD) clustering is a new trend in text mining in which the TDs are separated into several coherent clusters, where documents in the same cluster are similar. In this study, a new method for solving the TD clustering problem worked in the following two stages: (i) A new feature selection method using particle swarm optimization algorithm with a novel weighting scheme and a detailed dimension reduction technique are proposed to obtain a new subset of more informative features with low-dimensional space. This new subset is used to improve the performance of the text clustering (TC) algorithm in the subsequent stage and reduce its computation time. The k-mean clustering algorithm is used to evaluate the effectiveness of the obtained subsets. (ii) Four krill herd algorithms (KHAs), namely, (a) basic KHA, (b) modified KHA, (c) hybrid KHA, and (d) multi-objective hybrid KHA, are proposed to solve the TC problem; these algorithms are incremental improvements of the preceding versions. For the evaluation process, seven benchmark text datasets are used with different characterizations and complexities. Results show that the proposed methods and algorithms obtained the best results in comparison with the other comparative methods published in the literature.

CHAPTER 1

INTRODUCTION

1.1 Background

With the growth of the amount of text information on Internet web pages and modern applications, in general, interest in the text analysis area has increased to facilitate the processing of a large amount of unorganized text information (Sadeghian & Nezamabadi-pour, 2015).

Text clustering (TC) is an efficient unsupervised learning technique used to deal with numerous text documents (TDs) without any foreknowledge of the class label of the document (Prakash, Hanumanthappa, & Mamatha, 2014). This technique partitions a set of large TDs into meaningful and coherent clusters by collating relevant (similar) documents in the same cluster based on its intrinsic characteristics (Cobos et al., 2014). The same clusters (groups) contain relevant and similar TDs. Meanwhile, different clusters contain irrelevant and dissimilar TDs (L. M. Abualigah, Khader, & Al-Betar, 2016a).

In the modern era, clustering is an important activity because of the size of text information on Internet web pages (Oikonomakou & Vazirgiannis, 2010). Clustering is used to determine relevant TDs and facilitate TD display by groups that share the same pattern and contents (Cobos et al., 2014). The TC technique is successfully utilized in many research areas to facilitate the text analysis process, such as data mining, digital forensics analysis, and information retrieval (Forsati, Mahdavi, Shamsfard, &

Meybodi, 2013).

Vector space model (VSM) is the most common model used in TC to represent each document; in this model, each term in the TDs is a feature (word) for document representation (Salton, Wong, & Yang, 1975; Yuan, Ouyang, & Xiong, 2013). The TDs are represented by a multi-dimensional space, in which the position value of each dimension corresponds to a term frequency (TF) value. The text features generated from different text terms, even in a small document, would be represented by hundreds and/or thousands of text features. Thus, TDs will have high-dimensional informative and uninformative features (i.e., irrelevant, redundant, unevenly distributed, and noisy features). These uninformative features can be eliminated using the feature selection (FS) technique (Bharti & Singh, 2016b; L. Zheng, Diao, & Shen, 2015).

FS techniques are nondeterministic polynomial time-hard optimization methods used to determine the optimal subset of informative text features and improve the performance of the TC method while maintaining the necessary text information (Bharti & Singh, 2016b; K.-C. Lin, Zhang, Huang, Hung, & Yen, 2016). Typically, these techniques are performed even without any foreknowledge of the class label of the document. Conventionally, these techniques are divided into three main types, namely, FS based on document frequency (DF), FS based on TF, and hybrid feature technique based on DF and TF (Y. Wang, Liu, Feng, & Zhu, 2015). Several text-based studies rely on FS methods, such as TC (L. M. Abualigah, Khader, & Al-Betar, 2016b), text classification (Z. Zheng, Wu, & Srihari, 2004), and data mining (K.-C. Lin et al., 2016). Recently, metaheuristic algorithms have been successfully used in the area of text mining to solve the text document clustering problems (TDCPs) and text feature selection problems (TFSPs) (BoussaïD, Lepagnot, & Siarry, 2013).

The application of FS techniques produces a new subset with numerous informative text features. However, the dimensionality is still high because all dimensions remain even after removing the uninformative features. The dimensional space of this subset must be reduced further to facilitate the TC process (Lu, Liang, Ye, & Cao, 2015). High-dimensional feature space has become a significant challenge to the TC domain because it increases the computational time while decreases the efficiency of TC techniques (van der MLJP & van den HH, 2009). Thus, a dimension reduction (DR) technique is necessary to produce a new low-dimensional subset of useful features (Diao, 2014; Esmin, Coelho, & Matwin, 2015; Sorzano, Vargas, & Montano, 2014a). This technique will reduce the computation time and improve the performance of the TC algorithm. The DR technique should eliminate useless text features; eliminate unnecessary, redundant, and noisy text features; preserve intrinsic information; and significantly reduce the dimension of the text feature space (Bharti & Singh, 2014b; Raymer, Punch, Goodman, Kuhn, & Jain, 2000).

1.2 Motivation and Problem Statement

Recently, unorganized TDs on Internet web pages and modern applications have increased exponentially, and the number of Internet users in the world has exceeded three billion[1]. These users face difficulties in obtaining the information that they need easily and neatly (Bharti & Singh, 2014b; Uğuz, 2011). The process of managing such a large TD is called TD clustering technique, which transforms a set of large unorganized TDs into coherent and similar groups, that is, clusters, which facilitate user

[1] https://en.wikipedia.org/wiki/List_of_countries_by_number_of_Internet_users

browsing and searching for information. From the literature on TC techniques, four main problems are identified and explained as follows:

First, TDs usually contain informative and uninformative text features. Uninformative features can confuse and mislead the TD clustering algorithm, thereby reducing the performance of the clustering algorithm (Bharti & Singh, 2014b, 2015b). Therefore, identifying and removing these uninformative text features can improve the performance of the clustering algorithm and reduce the computation time. The main drawback of these methods is focusing only on selecting a new subset of text features that rely on the existing weighting scheme (score) (Bharti & Singh, 2016b). The current weighting scheme has certain weaknesses in evaluating the features by computing the weight score for all features equally using one main factor (i.e., term frequency). Thus, the distinction between informative and uninformative text features over the document is insufficient (Ahmad, Abu Bakar, & Yaakub, 2015; Bharti & Singh, 2016b; Cobos, León, & Mendoza, 2010a; Moayedikia, Jensen, Wiil, & Forsati, 2015). The weighted score should be more accurate to facilitate the process of the text FS technique, where it plays the main role in the FS procedure by distinguishing between TD features by providing a high score to the more informative features.

Second, the high-dimensional feature space is one of the most critical weaknesses of TDs because it influences the process of TD clustering techniques by increasing the execution time and decreasing the performance of the TD clustering algorithm (Bharti & Singh, 2014b; Nebu & Joseph, 2016). The high-dimensional feature space contains the necessary (useful) and unnecessary (useless) text features. Thus, the DR technique reduces the dimensional feature space by pruning useless text features to improve the performance of the clustering algorithm. One of the possible ways to solve the high-dimensional feature space is the DF method. This method deals with the reduction process with fixed roles (DF of the feature) in making the decision to prune useless text features (Esmin et al., 2015; Tang, Shepherd, Milios, & Heywood, 2005; Yao, Coquery, & Lê Cao, 2012a). The fundamental premise of the DF method is impractical because the frequently occurring features are considered more important in the documents than the infrequently occurring features (Bharti & Singh, 2015b).

Third, the main advantage of the TC algorithm is its effectiveness in guaranteeing access to the accurate clusters. Over the past few years, a large proportion of researchers in the TC domain applied metaheuristic algorithms to solve the TDCPs. However, a major drawback of these algorithms is that it provides a good exploration of the search space at the cost of exploitation (Bharti & Singh, 2016a). Other problems are related to unsatisfactory outcomes, such as inaccurate clusters, and the behavior of the algorithms that were selected is inappropriate for the problem of the TC instances (Bharti & Singh, 2015a; Binu, 2015; Forsati, Keikha, & Shamsfard, 2015; G.-G. Wang, Gandomi, Alavi, & Deb, 2015). All available TC techniques based on metaheuristic algorithms still face these problems. Solving the TC problem using metaheuristic algorithms still need more in-depth investigation for several important reasons (Y. Guo, Li, & Shao, 2015; Mohammed, Yusof, & Husni, 2015; J. Wang, Yuan, & Cheng, 2015). However, these reasons can be justified by the "no free lunch" theorem (Wolpert, 2013; Wolpert & Macready, 1997).

Fourth, the core effectiveness of the TD clustering techniques relies on the similarity and distance functions of the TC algorithm. These functions are used in making the decision to partition the document into an appropriate cluster based on the similarity or distance value; these decisions affect the performance of the TD clustering algorithm (Rao, Ramakrishna, & Babu, 2016). Similarity and distance measurements are standard function criteria used in the TD clustering domain as an objective function. Nevertheless, the results of these measurements are different and lead to certain challenges because of the variance between the values of similarity and distance measures for the same document (L. M. Abualigah, Khader, & Al-Betar, 2016a; Forsati et al., 2013). Determining the appropriate objective function to deal with the large TDs is difficult (Mukhopadhyay, Maulik, & Bandyopadhyay, 2015; Mukhopadhyay, Maulik, Bandyopadhyay, & Coello, 2014). Multi-objective functions (multiple-criteria decision making) are currently used in several domains as an alternative technique to yield better results (George & Parthiban, 2015; Saha, Ekbal, Alok, & Spandana, 2014). However, for the TD clustering technique, multiple-criteria decision making is relatively unknown.

1.3 Research Objectives

The overall aim of this study is to develop an effective TD clustering method. The main objective is to show that the improved method can outperform the other comparative methods. This research has the following objectives:

- to find the best features:
  - to enhance the weight score of the terms for the text FS technique in order to improve the TD clustering;

- to improve the text FS technique for finding a new subset of more informative features to improve the TD clustering;
- to reduce the dimension of the feature space in the form of a low-dimensional subset of useful features to improve the TD clustering;

- to improve the text document clustering using krill herd algorithm:
  - to increase the effectiveness of the TD clustering technique and to reduce its errors;
  - to improve the global search ability and its speed of convergence;
  - to enhance the quality of initial solutions obtained by the local search strategy;
  - to increase the likelihood of obtaining an accurate decision (similarity value) between the document and clusters centroids in the k-mean clustering algorithm.

1.4 Contributions

After the research objectives are achieved, this study will have the following main contributions:

1. Introduced a new weighting scheme to provide a significant influence score for the informative text features within the same document. This scheme focuses on assigning a favorable term weight to facilitate the text FS technique and distinguishes among the features of the clusters by giving a high weight to essential features in the same document.

2. Adapted metaheuristic optimization algorithms (i.e., genetic algorithm (GA), harmony search (HS), and particle swarm optimization (PSO)) to find the best features at the level of each document using a new FS method.

3. Introduced a new detailed DR technique to reduce the dimensional space of text features based on the detailed term frequency (DTF) and detailed document frequency (DDF) of each feature compatible with the size of its effect on the document. The DDF of each feature at the level of all documents is compatible with the size of its effect on the documents in partnership with its DTF value.

4. Adapted the basic krill herd algorithm (BKHA) and tuning its parameters for the text document clustering problem.

5. The modified krill herd algorithm (MKHA) to improve the global search ability. These modifications occur during ordering of the basic KH operators where the crossover and mutation processes are invoked after updating the positions of the krill herd algorithm (KHA).

6. The hybrid krill herd algorithm with the k-mean algorithm (HKHA) as a new operator, which plays a basic role in the MKHA to improve the local search ability. Hybridization is used to enhance the capacity of the KHA for finding locally optimal solutions by taking the refining power of the k-means clustering algorithm.

7. Introduced a multi-objective function based on the local best concept for the k-mean algorithm to enhance the capacity of the KHA by achieving an accurate local search, called multi-objective hybrid krill herd algorithm (MHKHA).

1.5 Research Scope

This study covers the main TC preprocessing steps (i.e., text FS and DR techniques) and the metaheuristic algorithms (i.e., different versions of the proposed KHA) to deal with the TDCP. The methods proposed in this study are applied to a large amount of TDs as electronic pages (i.e., newsgroup documents appearing on newswires, Internet web pages, and hospital information), modern applications (technical reports and university data), and biomedical sciences (large biomedical datasets). Note, all the datasets used in this research have been written in English language. These TDs (datasets) are characterized by high-dimensional informative and uninformative text features (Bharti & Singh, 2014b, 2015b; L. Zheng et al., 2015). All of the proposed methods need the number of clusters as input parameter K. Determining the correct number of clusters for the given TD datasets is an important issue because the number of document clusters is an essential parameter in TC problems. Standard TD datasets with different sizes (i.e., number of documents, number of terms, and number of clusters), constraints, and complexities are used in the TC technique to evaluate the proposed methods.

1.6 Research Methodology

This section briefly discusses the stages of the research methodology, which are applied to achieve the research objectives for improving the TD clustering technique, as shown in Figure 1.1. The detailed description is provided in Chapter 4.

The first stage is modeling and adapting GA, HS, and PSO to solve the text FS problem (TFSP) with the novel weighting scheme and detailed DR technique. This

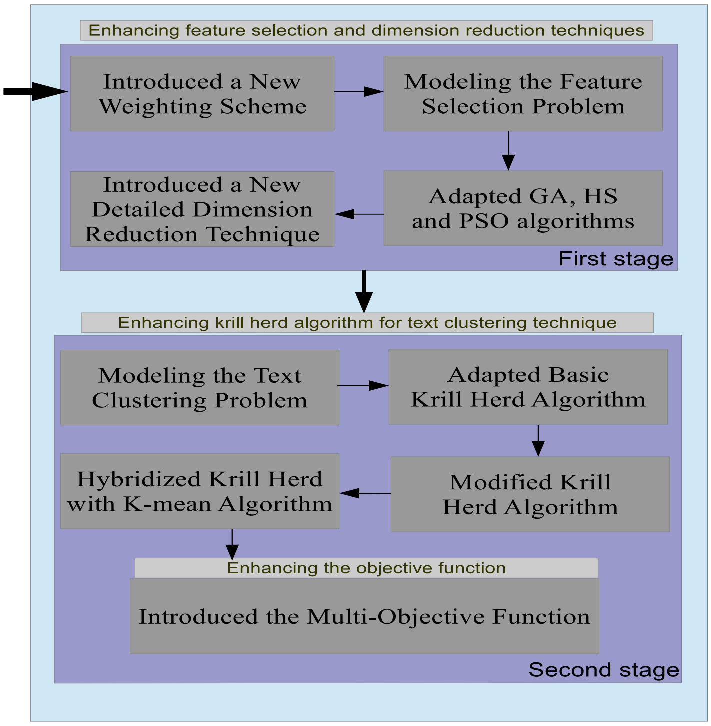

Figure 1.1: Research methodology.

stage facilitates the TC task to deal with a low-dimensional subset of informative text features, which reduce the computation time and improve the performance of the TD clustering algorithm.

The second stage is adapting the basic KH algorithm (BKHA) and tuning its parameters to solve the text DC problem (TDCP). Then, three versions of the BKHAs are modified (MKHAs) to improve the global (exploration) search ability. The three versions of the HKHA with the k-mean algorithm (MKHAs) are used to increase the performance of the TC technique by improving the local (exploitation) search ability. These hybrid versions used the results of the k-mean algorithm as the initial solutions in KHA to ensure balance between local exploitation and global exploration. Finally, a multi-objective function is applied to obtain an accurate TC technique by combining two standard measures (i.e., cosine similarity and Euclidean distance measurements).

The multi-objective function is the primary factor used to obtain an effective clustering method by deriving an accurate similarity value between the document and the cluster centroid.

1.7 Thesis Structure

The rest of this thesis organized as follows:

Chapter 2 (*Krill Herd Algorithm*): This chapter discusses the principles of the KHA. The analogy between the clustering technique and the optimization terms is provided. The steps of the KHA are described in detail.

Chapter 3 (*Literature Review*): This chapter provides an overview of the text pre-processing steps, TFSPs, and TDCPs with particular attention to TDs. This chapter also examines several methods used to deal with TFSP and TDCP. This chapter also presents a review of KHA in the areas of applications, modifications, and hybridizations across many fields.

Chapter 4 (*Proposed Methodology*): This chapter illustrates the modeling of TFSP and TDCP. This study also includes a comprehensive description of the adapted research methodology, including different weight schemes, metaheuristic algorithms for text FS, DR techniques, and KHAs for TD clustering, and the sequence of the procedures conducted.

Chapter 5 (*Experimental Results*): This chapter shows the experiments and results of all the proposed methods and presents the comparisons of each method with the others.

Chapter 6 (*Conclusion and Future Work*): This chapter provides the research conclusion and possible future works.

CHAPTER 2

KRILL HERD ALGORITHM

2.1 Introduction

Krill herd (KH) algorithm has a unique behavior to solve the text clustering problem. This algorithm was introduced by Gandomi and Alavi in the year 2012 to solve global optimization functions (Gandomi & Alavi, 2012). This section presents the modeling of the basic-krill herd algorithm (KHA) for the TDCP (L. M. Abualigah, Khader, Al-Betar, & Awadallah, 2016).

2.2 Krill Herd Algorithm

Krill herd (KH) is a swarm intelligence (SI) search algorithm based on the herding behavior of krill individuals (KIs). It is a population-based approach consisting of a huge number of krill, where each krill individual (KI) moves through a multi-dimensional space to search for close food and high-density herd (swarm). In KH as optimization algorithm, positions of KIs are considered as various design variables and the distance of the KI from the food is the objective function (Gandomi & Alavi, 2012; Mandal, Roy, & Mandal, 2014). The KH algorithm is considered in three categories: (1) Evolutionary algorithms (2) Swarm intelligence (3) Bacterial foraging algorithm (Bolaji et al., 2016).

2.3 Why the KHA has been Chosen for Solving the TDCP

The KH is a suitable algorithm for the TC technique according to: (i) the similarities between the behavior of the KHA and the behavior of the TD clustering technique, (ii) KH algorithm obtained better results in solving many problems in comparison with others common algorithms published in the literature.

The compatibility between KHA and TC involves searching for the closest food (closest centroid) and high density groups (similar groups) (Bolaji et al., 2016). Density is one of the main factors that influence the success of all the algorithms used to achieve coherence and similar groups. If documents in the same cluster are relevant, then density is high, and vice versa. If the KIs are close to the food, then density is high, and vice versa. Thus, the behavior of KIs is exactly the same as that of the TD clustering technique (both of them are a swarm).

With regard to the KHA, each KI (document) moves toward the best solution by searching for the herd (group) with high density (similar groups) and the closest food (closest centroid). These factors are used as objectives to lead each krill to an optimal herd around the food. With regard to the TC, each document moves toward the best solution by searching for the similar cluster centroid and the cluster with a high density. Moreover, these factors are used as objectives to lead each document to an optimal cluster around the closest centroid. The relationship between the behavior of KHA and the behavior of TD clustering is considered a strong feature in applying KHA to solve the TDCP.

2.4 Krill Herd Algorithm: Procedures

Due to the nature of this research, predation disperses KIs, leads to a decrease of the average krill density and distances of the KH from the food location. This process is the initialization phase in the KH algorithm. In the natural system, the objective function of each document is supposed to be the distance or similarity from the cluster centroid. The fitness function of each candidate solution is the total distance or similarity between all documents with clusters centroid. The KH algorithm has three main motion calculation to update individual positions; then it applies the KH operators, which is inspired by the evolutionary algorithm. The procedures sequence of the basic KH algorithm is shown in Figure 2.1.

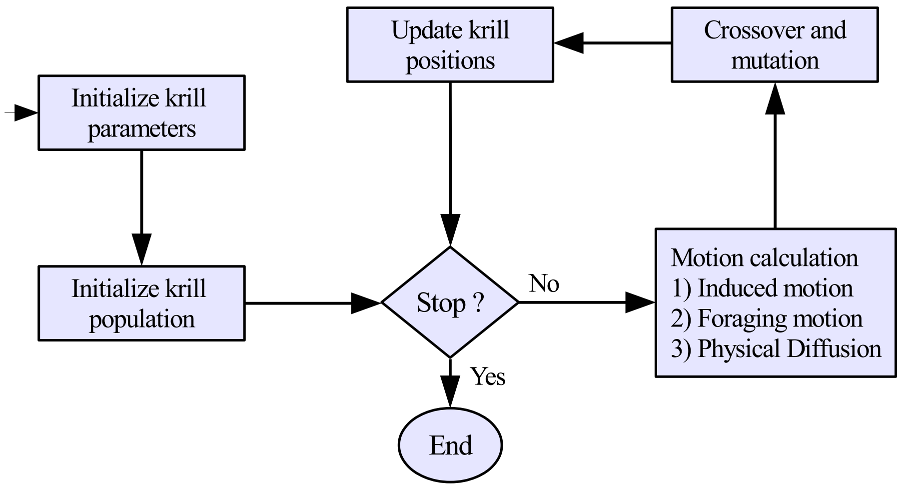

Figure 2.1: A flowchart of basic krill herd algorithm (Bolaji et al., 2016).

2.4.1 Mathematical Concept of Krill Herd Algorithm

The KH algorithm has three main steps to update the time-dependent position of each KI as follows:

- Movement induced by the presence of other KIs: only individual neighbors in the visual field that affects the KI moving.
- Foraging activity: the KIs search for food resources.
- Random diffusion: the net movement of each KI based on density regions (Gandomi & Alavi, 2012).

The i_{th} individual position is updated by the following Lagrangian model using Eq. (2.1).

$$\frac{dx_i}{dt} = N_i + F_i + D_i, \tag{2.1}$$

where for the krill i, N_i is the motion effect of the i_{th} individual from other KIs. This value is estimated from the local swarm density, a target swarm density, a repulsive swarm density, and the target direction which is effected by the best KI. F_i is the foraging motion for the i_{th} KI. This value estimated from the food attractiveness, food location, the foraging speed, the last foraging action or movement and the best fitness of the i_{th} krill so far. D_i is the physical diffusion for the i_{th} KI, where this value estimated from two factors: the maximum diffusion speed of the KIs and random direction (Gandomi, Talatahari, Tadbiri, & Alavi, 2013).

2.4.1(a) Movement Induced by other Krill Individuals

Movement induced is an illusion of visual perception in which a moving individual appears to move differently because of neighbors moving nearby in the visual field.

Theoretically, individuals try to keep the high density (Bolaji et al., 2016; G. Wang et al., 2014). The direction of movement induced is defined by Eq. (2.2).

$$N_i^{new} = N^{max}\alpha_i + \omega_n N_i^{old}, \tag{2.2}$$

where for krill i, N^{max} is the parameter for tuning the movement induced by other individuals, it is determined experimentally (see Table 5.11). α_i is estimated from the local swarm density by Eq. (2.3), ω_n is the inertia weight of the movement induced by other individuals' in range [0, 1], and N_i^{old} is the last change or movement produced.

$$\alpha_i = \alpha_i^{local} + \alpha_i^{target}, \tag{2.3}$$

where, the α_i^{local} is the effect of the neighbors in i_{th} individual movement, α_i^{target} is the target direction effected by the j_{th} KI. The effect of individual neighbors can be considered as an attractive or repulsive tendency between the KIs for a local search while the normalized values can be positive or negative (Bolaji et al., 2016; Gandomi & Alavi, 2012). The α_i^{local} is calculated by Eq. (2.4).

$$\alpha_i^{local} = \sum_{j=1}^{n} \widehat{K}_{i,j}\widehat{x}_{i,j}, \tag{2.4}$$

where, $\widehat{K}_{i,j}$ is the normalized value of the objective function vector for the i_{th} KI. $\widehat{x}_{i,j}$ is the normalized value of the related positions for the i_{th} KI. The $\widehat{K}_{i,j}$ is calculated by Eq. (2.5):

$$\widehat{K}_{i,j} = \frac{K_i - K_j}{K^{worst} - K^{best}}, \tag{2.5}$$

where, K_i is the objective function of i_{th} KI, K_j is the objective function of j_{th} neighbor ($j = 1, 2, ..., n$). n is the number of all KIs, K^{best} and K^{worst} are the best and worst objective function values of i_{th} individual. The $\widehat{x}_{i,j}$ is calculated by Eq. (2.6).

$$\widehat{x}_{i,j} = \frac{x_j - x_i}{\left\|x_j - x_i\right\| + \varepsilon}, \tag{2.6}$$

where, x_i is the current position, x_j is the position of j_{th} neighbor, $||x_j - x_i||$ is the vector normalization, it is used for calculating the neighbors of the i_{th} KI by Eq. (2.7), ε is a small positive number to avoid singularities (Jensi & Jiji, 2016; Mandal et al., 2014). The sensing distance is calculated by Eq. (2.7).

$$de_i = \frac{1}{5n} \sum_{j=1}^{n} \left\|x_i - x_j\right\|, \tag{2.7}$$

where, de_i is the sensing distance for the krill i. Note, if the distance value between two KIs is less than the current value, they are neighbors. Figure 2.2 illustrates the movement of the KIs and their neighbors.

The known target vector of each KI is the highest objective function. The effect of the best fitness on the j_{th} individual is calculated by Eq. (2.8). This procedure allows

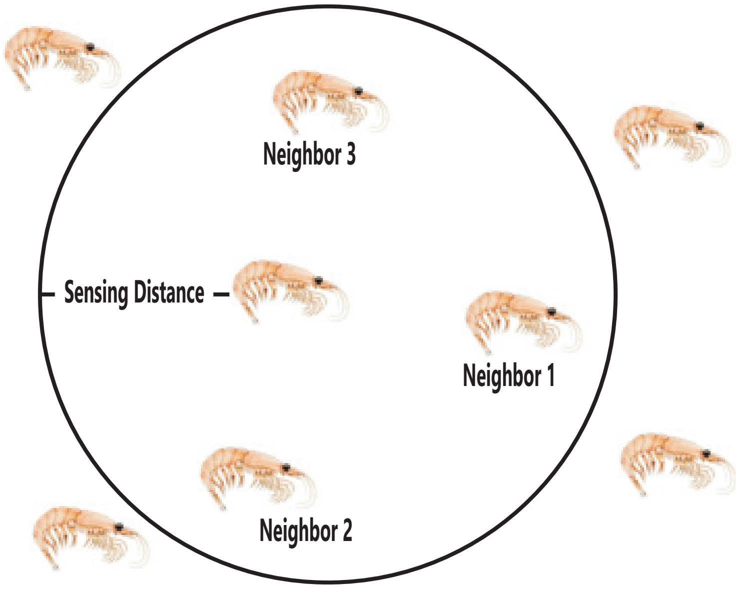

Figure 2.2: A schematic represents the sensing domain around a KI (Bolaji et al., 2016).

the solution to move towards the current best solution and is calculated by Eq. (2.8).

$$\alpha_i^{target} = C^{best} \widehat{K}_{i,best} \widehat{x}_{i,best}, \tag{2.8}$$

where,

$$C^{best} = 2\left(rand + \frac{I}{I_{max}}\right), \tag{2.9}$$

C^{best} is the coefficient of individuals, $\widehat{K}_{i,best}$ is the best objective function of the i_{th} KI, $\widehat{x}_{i,best}$ is the best position of the i_{th} KI, *rand* is a random number between [0, 1] for improving the local exploration; I is the current iteration number; I_{max} is the maximum number of iterations (Gandomi & Alavi, 2012).

2.4.1(b) Foraging Motion:

The foraging motion of KIs is estimated by two effects, namely, current food and old food location (L. M. Abualigah, Khader, Al-Betar, & Awadallah, 2016; Bolaji et al., 2016; Mandal et al., 2014). Food area or location is defined to attract KIs to the global optima possibly. The foraging motion for i_{th} individual is expressed by Eq. (2.10).

$$F_i = V_f \beta_i + \omega_f F_i^{old}, \tag{2.10}$$

where, V_f is the parameter for tuning the foraging speed, it is determined experimentally (see Table 5.11), β_i is the food location of the i_{th} KI by Eq. (2.11), ω_f is the inertia weight of the foraging speed in range [0, 1], and F_i^{old} is the last foraging motion.

$$\beta_i = \beta_i^{food} + \beta_i^{best}, \tag{2.11}$$

where, β_i^{food} is the food attractiveness of the i_{th} KI, it is calculated by Eq. (2.12). β_i^{best} is the best objective function of the i_{th} KI.

$$\beta_i^{food} = C^{food} \widehat{K}_{i,food} \widehat{x}_{i,food}, \tag{2.12}$$

where,

$$C^{food} = 2\left(1 - \frac{I}{I_{max}}\right), \tag{2.13}$$

$\widehat{K}_{i,food}$ is the normalized value of the objective function of the i_{th} centroid and $\widehat{x}_{i,food}$ is the normalized value of the i_{th} centroid position. The center of the individual's food for each iteration is calculated by Eq. (2.14).

$$x^{food} = \frac{\sum_{i=1}^{n} \frac{1}{K_i} x_i}{\sum_{j=1}^{n} \frac{1}{K_j}}, \tag{2.14}$$

where, n is the number of the KIs, K_i is the objective function of the i_{th} KI, and x_i is the i_{th} position value. The effect of the best objective function of the i_{th}KI is handled by using Eq. (2.15).:

$$\beta_i^{best} = \widehat{K}_{i,ibest} \widehat{x}_{i,ibest}, \tag{2.15}$$

where, $\widehat{K}_{i,best}$ is the best previous objective function of the i_{th} KI, $\widehat{x}_{i,food}$ is the best previous visited food position of the i_{th} KI. The movement induced by other individuals and the forging movement decrease with the increase in the time (iterations).

2.4.1(c) Physical Diffusion:

Physical diffusion is the net movement of each KI from a region of high density to a region of low density or vice versa. The better position of the KI is the less random direction. Physical diffusion values of individuals are estimated by two effects, namely, maximum diffusion speed (D_m) and random directional vector (δ) (L. M. Abualigah, Khader, Al-Betar, & Awadallah, 2016; Gandomi & Alavi, 2012; Jensi & Jiji, 2016; G. Wang et al., 2014). Physical diffusion for the i_{th} KI is determined by Eq. (2.16).

$$D_i = D^{max}\left(1 - \frac{I}{I_{max}}\right)\delta, \tag{2.16}$$

where, D^{max} is the parameter for tuning the diffusion speed, it is determined experimentally (see Table 5.11), and δ refers to the array that contains random values between [*-1, 1*]. I is the current iteration, I_{max} is max number of iterations.

2.4.1(d) Updating the Krill Individuals:

The movement of the i_{th} KI is influenced by the other KIs, foraging motion, and physical diffusion. These factors seek to obtain the best objective function for each KI. The foraging movement and the movement induced by other KIs include two global and two local strategies. These strategies are working in parallel to make KH a robust algorithm (Bolaji et al., 2016; Gandomi & Alavi, 2012; G. Wang et al., 2013). The individual positions updated towards the best objective function by Eq. (2.17).

$$x_i(I+1) = x_i(I) + \Delta t\frac{dx_i}{dt}, \tag{2.17}$$

where,

$$\Delta t = C_t \sum_{j=1}^{n} (UB_j - LB_j), \tag{2.18}$$

Δt is an important and sensitive constant computed by Eq. (2.18), and n is the total number of individuals. LB_j is the lower bound, UB_j is the upper bounds of the ith variables $(J = 1, 2, ..., n)$, and C_t is a constant value between [0, 2]. It works as a scale factor of the speed vector.

2.4.2 The Genetic Operators

Genetic algorithm (GA) is a stochastic meta-heuristic search method for the global solution in a large search space. This algorithm is inspired by the classical evolutionary algorithms (EA). The genetic operators encoded in a genome that performed in an unusual way that permits asexual reproduction that leads to the offspring. However, the sexual reproduction can swap and reorder chromosomes, giving birth to offspring which includes a cross breeding of genetic information from all parents. This operation is often called a crossover, which means swapping of the genetic information. To avoid premature convergence, the mutation operator is used to increase the diversity of the solutions (H. Chen, Jiang, Li, & Li, 2013; G.-G. Wang, Gandomi, & Alavi, 2014b). Genetic operators are incorporated into the KH algorithm to improve its performance (Bolaji et al., 2016; Gandomi & Alavi, 2012).

2.4.2(a) Crossover Operator of KH Algorithm:

The crossover operator is an effective procedure for global solutions. This procedure is controlled by a probability Cr by generating a uniformly distributed random value between [*0, 1*] (G.-G. Wang, Gandomi, & Alavi, 2014b). The mth component of $x_{i,m}$ is determined as the following:

$$x_{i,m} = \begin{cases} x_{p,m}, & \text{if } rand < Cr \\ x_{q,m} & \text{else} \end{cases} \tag{2.19}$$

.

$$Cr = 0.2\widehat{K}_{i,best}, \tag{2.20}$$

.

where, the crossover probability is determined by Eq. (2.19). p and q refer to the two solutions which are chosen for the crossover operator, $p, q \in \{1, 2, ..., i-1, i+1, ..., n\}$, the Cr increases with decreasing fitness function, $\widehat{K}_{i,best} = K_i - K^{best}$; K_i is the objective function value of the i_{th} KI, and K^{best} is the best objective function value of the i_{th} KI.

2.4.2(b) Mutation Operator of KH Algorithm:

The mutation operator is an effective strategy for a global solution. This strategy is controlled by a probability *Mu* (G. Wang et al., 2014). The mutation operator is determined as the following:

$$x_{i,m} = \begin{cases} x_{gbest,m} + \mu(x_{p,m} - x_{q,m}), & \text{if } rand < Mu \\ x_{i,m}, & \text{else} \end{cases} \tag{2.21}$$