Scientific Review – Engineering and Environmental Sciences (2021), 30 (2), 354–364
Sci. Rev. Eng. Env. Sci. (2021), 30 (2)
Przegląd Naukowy – Inżynieria i Kształtowanie Środowiska (2021), 30 (2), 354–364
Prz. Nauk. Inż. Kszt. Środ. (2021), 30 (2)
http://iks.pn.sggw.pl
DOI 10.22630/PNIKS.2021.30.2.30

Firas N. Al-MASHHADANI[1], Qusay S. ALSAFFAR[2], Ibrahim A. Al-JADIR[3]

[1] Ministry of Education – Iraq, General Directorate of Vocational Education
[2] Ministry of Higher Education and Scientific Research – Iraq
[3] Murdoch University, PhD holder from the School of Engineering and Information Technology

An enhanced krill herd optimization technique used for classification problem

Key words: optimization, simulated annealing, standard krill herd

Introduction

Document classification (DC) has become one of the most important subjects in data mining and also it has gained a wide attention in the last few decades due to its important applications in various life aspects such as document filtering and dissemination, information retrieval, enterprise portals, business intelligence, text and summarization, etc. Consequently, vast document clustering algorithms have been suggested through the years, and an extensive survey of the most recent advances in document classification can be found in Kowsari et al. (2019). A couple of issues are pivotal for all the document classification algorithms which are: (i) the way that the discriminating features are extracted, and (ii) the way that the new coming document to be classified a according to the extracted feature in (i) the latent semantic indexing (LSI) (Yee, 2020) is considered as one of the most familiar representation and feature extraction method used for DC. Despite of the fact that the LSI is guaranteed to be handy in different text-related tasks, its objective is to discover the highly representative textual attributes, but it doesn't focus on the highly discriminative textual attributes in representing documents. Thus, the LSI would not be the best choice in the discrimination process of document into different classes. On the other hand, the linear discriminant analysis (LDA) (Li, Shao, Yin & Liu, 2019) is a different feature extraction method that which is famous in the pattern recognition mediums. Its objective is based on finding the best features in terms of discriminability via the maximization of the ratio of the between inter class distances-to-the intra class distance of any set of data subject to a supervised learning criteria. The intra class means the distance of attributes within the class while the inter class represents the distance of the attributes within different classes. Currently, a number of studies have reported that the document data might be located in a non-linear sub-manifold. Nonetheless, each one of the LSI, and the LDA would only be considering to the Euclidean configuration. They don't successes to find out the actual configuration, if the document data resides on a non-linear sub-manifold incorporated within the document space. In terms of the DC, the selection of the classifier is different key factor after the extraction of the important documents' attributes. Presently, the neural networks, Bayesian algorithm, the k-nearest neighbor (KNN) classifier, and the support vector machines (SVM) are extensively utilized in a wide range of text-based applications but with the vast increment of the online availability of text documents on the Web environment, these methods have become insufficient to be capable of dealing with that huge information increase. The problem of classifying document automatically becomes a real challenge because of the large number of documents, the number of features and restricted awareness of human method to classify documents attributed to a number of factors (Jaya, Aulia, Hardi, Tarigan & Lydia, 2019). Thus, these traditional classification methods that used the exhaustive search in the search space become impractical solutions (Schaetti, 2019). In the last few years the KH method has invented as a global optimization method. It has considered as a novel metaheuristic inspired from nature herding process. It has allured text mining community. This algorithm has been used in a successful manner with different complicated optimization problems since its invention (Hofmann, Haskell, Klinck & Lascara, 2004; Gandomi & Alavi, 2012; Bolaji, Al-Betar, Awadallah, Khader & Abualigah, 2016). However, the usage of this algorithm for the DC in the regard of text mining is not yet well investigated. Therefore, in this paper we mainly focus on applying this method on document classification.

In this section an explanation of the proposed method based on the above information is explained in details as follows:

In this paper, the KH is used as a global search for DC problem which is used as an efficient way for an inter class-discrimination and is used for the attributes discrimination forming out the new classes while the simulated annealing (SA) approximation method is used as a krill modifier. The KH method is relatively a new global search. It simulates the behavior of the krill herd members in movement and foraging using Lagrange model (Hofmann et al., 2004). It is firstly proposed by Gandomi and Alavi (2012). As is mention before, global search methods are derivative-free methods, that is, they are stochastic-based. Thus, the KH can be robust in the exploration of the search space while it is weak in search of the local areas within the search space. Therefore, it is highly skeptical to be trapped in local optima like many other global search methods. However, the advantage of the KH is that it has a few number of control parameters. More specifically, it has only one control parameter which is the time interval for the krill adjustment. By the combination of both the SA and the KH and producing the enhanced version of the KH method the we name our method the EKH, it becomes more likely to get a more robust method of document classification that trades-off between the exploration and the exploitation of the search space.

Krill herd algorithm

The KH algorithm is based on the idea of krill individual's movements and foraging. The objective function of the KH is calculated by measuring the minimum distance of each krill from the food source, at the same time, the entire herd density is also taken into account. The krill position is determined mainly by three factors, first its distance from the food, the impact of movement generated by other krill individuals and the krill physical diffusion. These three factors can be mathematically represented as:

- The movement of the KH that induced by other individuals from the herd with the aim of keeping the swarm as dense as possible.
- Forage activities.
- Random diffusion.

Lagrange model

Lagrange model can be used to generalize *n*-dimensional search space. The predation can remove krill from the herd and that reduces the density of the krill swarm. Also, that disturbs the way of the swarm to the food source. This is considered the initialization of the swarm. According to the three factors, a Lagrange model can be generated as is shown in Eq. (1). The time dependent position of krill individuals is selected according to the random diffusion controlled by N_i, foraging activity F_i, and movement made by the existence of the adjacent krills.

Foraging activity is made according to particular food criteria and is such that krill individuals move in slower motion and turn more frequently in regions of higher food densities. The existence/ /absence of neighboring krills might be dispersing, if the local food densities become very concentrated, or coalescing if food densities become very diluted, in a respective manner. Predation effect on krill individuals is incorporated and impacts the swarm structure by eliminating some individuals (Hofmann et al., 2004).

$$\frac{dX_i}{dt} = N_i + F_i + D_i \tag{1}$$

where:

N_i – parameter determining of the motion induced by other krill individuals,

F_i – foraging activity,

D_i – diffusion of the i^{th} krill.

Motion caused by other krills (N_i). Krill individuals strive to keep the swarm condense and move as one unit due to the impact of their mutual effects (Hofmann et al., 2004). The estimation of the swarm direction α_i can be determined by measuring the local swarm density, a target swarm density, and a repulsive swarm density. Therefore, in Eq. (2) these three densities are used to calculate the motion induced by other krill individuals. For the first parameter N_i it can be updated by applying Eq. (2).

$$N_i^{\text{new}} = N^{\max} \cdot \alpha_i + N_i^{\text{old}} \cdot \omega_n \tag{2}$$

$$\alpha_i = \alpha_i^{\text{local}} + \alpha_i^{\text{target}} \tag{3}$$

where:
$N^{\max}$ – maximum speed induced,
N_i^{old} – final motion induced,
ω_n – inertia weight in the range [0, 1],
α_i^{local} – effect of the surrounding krill individuals on the motion of a particular krill,
α_i^{target} – effect of the best krill that has the best fitness value.

The effect of the surrounding krills α_i^{local} can be obtained from the following equations:

$$\alpha_i^{\text{local}} = \sum_{j=1}^{NN} \hat{K}_{ij} \hat{X}_{ij} \tag{4}$$

$$\hat{X}_{ij} = \frac{x_j - x_i}{\left\| x_j - x_i \right\| + \varepsilon} \tag{5}$$

$$\hat{K}_{ij} = \frac{K_i - K_j}{K^{\text{worst}} - K^{\text{best}}} \tag{6}$$

where:
X_i, X_j – locations of i^{th} and j^{th} krills,
ε – small positive value added to denominator to prevent any singularities,
k_i, k_j – fitness scores of the i^{th} and the j^{th} krills,
k^{worst}, k^{best} – worst and best fitness scores in the swarn.

The target effect that determines the effect of the best krill that has the best fitness value can be calculated as follows:

$$\alpha_i^{\text{target}} = C^{\text{best}} \cdot \hat{K}_{i,\text{best}} \cdot \hat{X}_{i,\text{best}} \tag{7}$$

where:
$\hat{K}_{i,\text{best}}$ – computed as same as $\hat{k}_{ij}$, still the fitness value of j^{th} krill individual k_j, is substituted by the best fitness value,
$\hat{X}_{i,\text{best}}$ – computed as same as $\hat{X}_{ij}$, but X_j is substituted by the X_{best} that represents the best fitness value,
C^{best} – effective coefficient of the krill that has the best fitness value to the i^{th} krill; it be calculated as follows:

$$C^{\text{best}} = 2\left(rand + \frac{1}{I^{\max}} \right) \tag{8}$$

where:
$rand$ – random number in the range [0, 1],
I – iteration counter,
$I^{\max}$ – number of iterations.

Foraging motion (F_i). This factor can be computed in terms of food location and previous experience of where food was located. The foraging motion can be calculated as follows:

$$F_i = a \cdot \left(\beta_{ik}^{\text{food}} + \beta_{ik}^{\text{best}} \right) \tag{9}$$

where:
β^{food} – food attraction that is used to attract krills to global optimum,
β^{best} – effect of the current best krill individual and a is a constant value.

Physical diffusion. The physical diffusion is an arbitrary process which is computed as a function of the maximum diffusion speed and a random directional vector as follows:

$$D_i = D^{\max} \cdot \delta \tag{10}$$

where:

$D^{\max}$ – maximum diffusion speed,
δ – random directional vector with the values in the range [1, –1].

Motion process of krill herd. According to the N_i, F_i, and D_i, the krill positions can be calculated during the time interval Δt as is shown in the following equation.

$$x_i\left(t+\Delta t\right) = x_i\left(t\right) + \Delta t \frac{\mathrm{d}X_i}{\mathrm{d}t} \tag{11}$$

After the position updates of the krill individuals, the reproduction mechanisms are used, which are the crossover and the mutation.

Simulated annealing (SA)

It is inspired by the melting process of metals (Merendino & Celebi, 2013). After a particular heating degree, the metal liquid is cooled gradually, until it stabilizes by a decreased mobility of its atoms. This process continues until least energy is required. Because of the heating process allows atoms to move freely, the cooling process will have the time to let atoms to rearrange themselves to attain to the lower energy state (Wang, Guo, Gandomi, Alavi & Duan, 2013).

Like other approximation methods, this method has a local-based nature, it gradually improve the current state (solution) by small movements. The SA generates neighboring solutions and continues to generate other neighboring solutions in the same path that minimizes the objective function. In the exploration of the solution space, SA can accept the possibility of the worst solution in special cases to skip the local minima. More particularly, in every single iteration, the current solution x with an objective function value $f(x)$, there is a neighboring solution $\grave{x}$. Iteratively, the difference between the fitness of the current solution $f(x)$ and the neighboring solution $f(\grave{x})$ is used to calculate the probability of acceptance or rejection of the neighboring solution as follows:

$$\Delta f = f(\grave{x}) - f(x) \tag{12}$$

$$P_s = \exp\left(-\frac{\Delta f}{k \cdot T}\right) \tag{13}$$

$$T = T \cdot c \tag{14}$$

where:
P_s – probability of acceptance,
Δ – difference between the current and the neighboring solution,
k – Boltzmann constant,
T – initial temperature.

After the production of P_s probability, it is compared against a random value selection between zero and one (r). If $P_s > r$ the solution $\grave{x}$ will be accepted. Finally, T is decreased in every single iteration. In our proposed SA local search we used the single the Markov chain protocol (SMC) cooling schedule due to its equal efficiency in comparison to the Markov multi-chain protocol (Otranto, 2005).

Proposed simulated annealing krill herd

In this section an explanation of the proposed method based on the above information is explained in details as follows:

Document representation

Structured data contents the major sources for many text mining tasks. Nonetheless, many applications have unstructured data such as document collections. Therefore, to deal with unstructured data, text mining tasks must be performing many pre-processing steps for computing a structured model for the mining tasks. The pre-processing steps followed in this paper are listed below (Uysal & Gunal, 2014):

- Documents preparation: it involves the selection of the appropriate set of documents.
- Tokenization: it involves the identification of the words that would be pre-processed.
- Stop word removal: many words are not related to the classification tasks and these words must be deleted in this step. Often these words are commonly used in any text such as prepositions or articles.
- Stemming: here, each word is originated to its original grammatical form, by taking out the root of each word. For instance, the root form of "playing" is "play";
- Luhn-cut: this step is responsible of elimination of terms for which the frequency is less than a pre-defined threshold value.
- Weighting: is used for minimizing or maximizing the words contributions on mining tasks. For instance, a famous technique is used in this paper which is the term frequency-inverse document frequency (TF-IDF) (Christian, Agus & Suhartono, 2016).

The outcome of applying the pre-processing steps is a vector space model (Uysal & Gunal, 2014), also named as document-term-matrix. Here, each row is representing a document, while each column is representing an attribute, and each value represents a frequency (TF), or the term impact in that very document. Therefore, text mining methodologies like the classification would be employed in the structured document corpus. In addition, the majority of the information visualization applications need structured data, and for that reason, those pre-processing steps are suitable for the visualization of the document corpus (Saad, 2010).

Simulated annealing krill herd

For the native KH algorithm, because the search is only performed by random walks, the convergence cannot be expected easily. To enhance the performance of the native KH, the genetic reproduction operators have been combined to the native KH. Using the KH with the GA operators performed the best in comparison to the original version as is seen in publication by Abualigah, Khader and Hanandeh (2019). The KH uses three motions explained earlier, and that can guarantee the convergence to some extent with some simple optimization problems. Nonetheless, things could be harder, and convergence could be unreachable with more complicated optimization problems such as document clustering. The KH can fail to discover the optimality with the document clustering as it is an NP-hard optimization problem.

In the current method, another component is added to the original KH that can modify the exploitation phase. This is the SA that resembles the local search in the schemes. As is explained earlier, the SA is based on the idea of the melting metals. In effect, in this paper, this physical concept of annealing metals under a particular degree is combined with the idea of krill herding. This is to generate a breed of modified individuals whose capability of exploitation of the search space is expected to exceed the ordinary individuals. The main distinction between the EKH and the KH is in the capability of the EKH to modify the best ordinary members in the herd to outperform themselves after the use of the SA concept.

To correct the premature convergence resulted from the slight changes at the end of the optimization process; the use of the SA can lift the curve where there might be stagnation points during the search in final stages. This can potentially suppress the possibilities of obtaining weak solutions (Wang et al., 2013).

The mechanism of the SA in that context is based on the acceptance of a small number of ordinary krill individuals with a low probability named a transition probability. This acceptance probability technique allows the population to be more diversified and that in return makes the premature convergence more preventable. Also, that would help the search to span wider areas in the solutions space leading to discover more promising areas (Mirjalili & Lewis, 2016). The proposed method can be summarized as shown in Algorithm 1.

ALGORITHM 1. The proposed hybrid krill herd and simulated annealing binary feature selection method

```
Begin
Step 1: Initialize. The generations counter $t$, the population-size $P$ of $N$ krill, the diffusion speed $D_{\max}$, the foraging speed $V_f$, and max speed $N_{\max}$, Boltzmann constant $k$, initial temperature $T_0$, cooling factor $\alpha$, and an acceptance thresholding  value $f_n$.
Step 2: Fitness computing. Compute the fitness for each individual krill according to the initial location of krills.
Step 3: While $t$ less than maximum generation number do
    arrange the entire krill herd based on their fitness.
    Store the best krill so far.
    for $i$ := 1 to $N$ (all krill population) do
        conduct three motions as described earlier.
        Updating locations for the krill individual $i$ according to the simulated annealing operator (Algorithm 2).
        Calculating the fitness value for each single krill according to its new location $X_i$+1.
    end for $i$
    Substituting the best krill individual with the worst one.
    Sorting all the krill according to their fitness and find the current best.
    $t$ := $t$ + 1;
Step 4: end for while
Step 5: best solution is found
End
```

ALGORITHM 2. Simulated annealing for krill individuals selection

```
Begin
    $T := \alpha * T$;
    $\Delta := f(X'_i) - f(X_i)$;
    # Acceptance if become better
    If ($-\Delta f > f_n$) is true then do
        $X_i+1 := X'$;
    end if
    % Acceptance  with a lower probability if not become better
    If ($\Delta f <= f$ and $\exp(-\Delta f$ div $(k\ X_T)) > r$) do
        $X_i+1 := X'$;
    End if
End
```

The limited performance of the traditional KH method on with many scientific and engineering problems, made it necessary to modify the existing KH method in this paper we used the simulated annealing because of its capability to conduct the search in narrow search areas. In our proposed method, first, the traditional KH method is used to nominate promising candidates solutions set. Later, a krill selection using the simulated annealing method is applied. This method is considered as an operator incorporated with the traditional KH method. This operator has greedy technique and acceptance a not very good individual with a relatively small probability. This technique is used for accepting well-fit candidates solutions in order to enhance its reliability and efficiency and to solve optimization problems of a global numerical nature. Moreover, the simulated annealing operator is not only accepting modifications that enhance the fitness function. Rather it also retains some modifications in not very good individuals with a lower probability that is not ideally optimal. This could improve the diversity of the krill population; enhance the explorative aspect of the KH method, as well as avoiding prematurity of convergence. Moreover, this proposed method could speed up the global convergence ratio with no loss of the robust and strong features of the traditional KH method.

Parameter settings and evaluation measures

In this section an explanation of the test results is given. The parameters concerning the KH are used as follows, $V_f = 0.02$, the diffusion speed $D_{\max} = 0.005$, the maximum induction $N_{\max} = 0.01$, initial temp $T_0 = 1.0$, acceptance number $Accept_{\max} = 15$, Boltzmann's factor $k = 1$, cooling parameter $\alpha = 0.95$, and threshold of acceptance $f_n = 0.01$. All these parameters are replicated from the original KH paper.

After setting the of parameters, two evaluation measures were used to evaluate the results one is the F-measure external measure while the other classification error value is used as an internal measure which also serves as a clustering fitness function. Unlike the external measures such as those used in Forsati, Keikha and Shamsfard (2015), the internal measure using the classification error is also used.

Datasets and test results

In this paper, five datasets were used which are diversified according to their number of classes, documents size, subjects and documents number. Table 1 shows the details of the used datasets that shows the number of instances (documents) and number of features for each dataset.

In this section, results obtained by EKH are described on five datasets shown in Table 1. A comparison of EKH is performed against a set of other clustering methods. Namely, these methods are the differential evolution (DE) (Opara & Arabas, 2019), the global artificial bee colony and Levenberq-Marquardt (GABC-LM) (Shah et al., 2017), the k-means, the harmony search (HS) (Sridharan & Komarasamy, 2020) and the standard KH (Gandomi & Alavi, 2012). For the GABC-LM the same parameter settings used in the original paper are also used in this paper. In the tests, 50 times each algorithm is run. This number is used by many other papers in this area.

Tables 2 and 3 report the values of the external and internal evaluation (fitness error values). The k-nearest neighbor has been used as a fitness function. The comparison of these methods is conducted using the comparing using the five datasets shown in Table 1. In Table 2 the EKH achieved better results in comparison to the other methods regarding the F-measure. This indicates that the use of the modified version of KH improved the performance of the classification as can be seen with the majority of datasets.

When it comes to the classification error, this criterion is used to measure the

TABLE 1. Datasets

Dataset	D#	#Classes	Instances	Attributes
6-event crimes	D-1	6	223	3 864
Classic3	D-2	3	3 893	13 310
TDT	D-3	53	6 738	1 445
Pair-20news	D-4	2	1 071	9 497
Reuters	D-5	8	4 195	6 738

TABLE 2. The classification results using the F-measure

Dataset	DE	GABC-LM	k-Means	HS	KH	EKH
6 Events crimes	0.7545	0.8265	0.6801	0.7697	0.8459	0.9689
Classic 3	0.9465	0.9428	0.7509	0.9049	0.9804	0.9956
TDT 5	0.6760	0.6617	0.0698	0.9194	0.7095	0.8728
Pair of 20news	0.0036	0.9896	0.6755	0.9816	0.9854	0.9987
Reuters	0.5865	0.5341	0.0698	0.5899	0.8551	0.9819

TABLE 3. The classification accuracy

Dataset	DE	GABC-LM	HS	KH	EKH
6 Events crimes	0.3039	0.2222	0.7188	0.7240	0.1232
Classic 3	0.4605	0.2606	0.7075	0.7438	0.1455
TDT 5	0.3450	0.3827	0.8214	0.6883	0.2894
Pair of 20news	0.2130	0.3450	0.7419	0.8449	0.1454
Reuters	0.2830	0.8306	0.7322	0.7345	0.2346

internal coherency of the resulted classes with no need to deal with the class labels. The classification error measures the distances between documents within the class. The error score should be minimized in contrast to the F-measure. The error rate and the F-measure relationship have been hardly studied. Idealistically, the error rate should be reduced while the F-measure increases. However, this is not always true. It is likely to happen that the internal measure remains the steady or slightly fluctuates whereas the F-measure changes largely. This case can give a positive indication as the F-measure change can show the positive or negative algorithm's performance despite the steadiness (or the slight changes) of the error rate. This case is clear in Table 3 where almost all the error rate results where nearly equal with all methods. At the same time the F-measure values shown in Table 3 are different among the algorithms. Therefore, despite the fact that with almost all the datasets, the error rate results are not equal.

Conclusions

In this paper, we propose the krill herd classification method named as the EKH. This method combines the KH with the SA approximation method. The SA is utilized because of its capability to enhance the solutions produced by the KH in a local manner using the synergy of the melting metals. The test results indicate that the EKH outperformed other document classification methods: these methods are the DE, the GABC-LM, the k-means, the HS, and the KH.

References

Abualigah, L.M., Khader, A.T. & Hanandeh, E.S. (2019). Modified krill herd algorithm for global numerical optimization problems. In S.K. Shandilya (eds.), *Advances in nature-inspired computing and applications* (pp. 205-221). Cham: Springer.

Mirjalili, S. & Lewis, A. (2016). The whale optimization algorithm. *Advances in Engineering Software*, *95*, 51-67.

Bolaji, A.L.A., Al-Betar, M.A., Awadallah, M.A., Khader, A.T. & Abualigah, L.M. (2016). A comprehensive review: Krill Herd algorithm (KH) and its applications. *Applied Soft Computing*, *49*, 437-446.

Christian, H., Agus, M.P. & Suhartono, D. (2016). Single document automatic text summarization using term frequency-inverse document frequency (TF-IDF). *ComTech: Computer, Mathematics and Engineering Applications*, *7*(4), 285-294.

Forsati, R., Keikha, A. & Shamsfard, M. (2015). An improved bee colony optimization algorithm with an application to document clustering. *Neurocomputing*, *159*, 9-26.

Gandomi, A.H. & Alavi, A.H. (2012). Krill herd: a new bio-inspired optimization algorithm. *Communications in Nonlinear Science and Numerical Simulation*, *17*(12), 4831-4845.

Hofmann, E.E., Haskell, A.E., Klinck, J.M. & Lascara, C.M. (2004). Lagrangian modelling studies of Antarctic krill (*Euphausia superba*) swarm formation. *ICES Journal of Marine Science*, *61*(4), 617-631.

Jaya, I., Aulia, I., Hardi, S.M., Tarigan, J.T. & Lydia, M.S. (2019). Scientific documents classification using support vector machine algorithm. *Journal of Physics: Conference Series*, *1235*(1), 012082. https://doi.org/10.1088/1742-6596/1235/1/012082

Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L. & Brown, D. (2019). Text classification algorithms: a survey. *Information*, *10*(4), 150. https://doi.org/10.3390/info10040150

Li, C.N., Shao, Y.H., Yin, W. & Liu, M.Z. (2019). Robust and sparse linear discriminant analysis via an alternating direction method of multipliers. *IEEE Transactions on Neural Networks and Learning Systems*, *31*(3), 915-926.

Merendino, S. & Celebi, M.E. (2013). A simulated annealing clustering algorithm based on center perturbation using Gaussian mutation. In Ch. Boonthum-Denecke, G.M. Youngblood (eds.), *Proceedings of the 26th International Florida Artificial Intelligence Research Society Conference* (pp. 456-461). Palo Alto: AAAI Press.

Otranto, E. (2005). The multi-chain Markov switching model. *Journal of Forecasting*, *24*(7), 523-537.

Opara, K.R. & Arabas, J. (2019). Differential evolution: a survey of theoretical analyses. *Swarm and Evolutionary Computation*, *44*, 546-558.

Saad, M.K. (2010). *The impact of text preprocessing and term weighting on Arabic text classification.* Gaza: Computer Engineering, the Islamic University.

Schaetti, N. (2019). Behaviors of reservoir computing models for textual documents classification. *2019 International Joint Conference on Neural Networks (IJCNN)*, *2019*, 1-7.

Shah, H., Tairan, N., Mashwani, W.K., Al-Sewari, A.A., Jan, M.A. & Badshah, G. (2017). Hybrid global crossover bees algorithm for solving boolean function classification task. In *International Conference on Intelligent Computing* (pp. 467-478). Springer: Cham.

Sridharan, K. & Komarasamy, G. (2020). Sentiment classification using harmony random forest and harmony gradient boosting machine. *Soft Computing*, *24*(10), 7451-7458.

Uysal, A.K. & Gunal, S. (2014). The impact of preprocessing on text classification. *Information Processing & Management*, *50*(1), 104-112.

Wang, G.G., Guo, L., Gandomi, A.H., Alavi, A.H. & Duan, H. (2013). Simulated annealing-based krill herd algorithm for global optimization. *Abstract and Applied Analysis*, *2013*, 213853. https://doi.org/10.1155/2013/213853

Yee, C.W. (2020). *Retrieving semantically relevant documents using Latent Semantic Indexing.* Yangon: Unversity of Computer Studies.

Summary

An enhanced krill herd optimization technique used for classification problem. In this paper, this method is intended to improve the optimization of the classification problem in machine learning. The EKH as a global search optimization method, it allocates the best representation of the solution (krill individual) whereas it uses the simulated annealing (SA) to modify the generated krill individuals (each individual represents a set of bits). The test results showed that the KH outperformed other methods using the external and internal evaluation measures.

Authors' address:
Firas N. Al-Mashhadani – corresponding author
(https://orcid.org/0000-0003-4537-7832)
Ministry of Education
D.G. Vocational Education
City Baghdad
Iraq
e-mail: firas.naziyah@vocational.edu.iq