Social spider optimisation algorithm for dimension reduction of electroencephalogram signals in human emotion recognition

Abdullah Yousef Al-Qammaz*, Farzana Kabir Ahmad, Yuhanis Yusof

School of Computing, University Utara Malaysia, 06010 Sintok, Kedah, Malaysia
*Corresponding author E-mail: abdullah_yousef@ahsgs.uum.edu.my

Abstract

Due to some limitations of current heuristics and evolutionary algorithms, this paper proposed a new swarm based algorithm for feature selection method called Social Spider Optimization (SSO-FS). In this research, SSO-FS is used in the EEG-based emotion recognition model as searching method to find optimal feature set to maximize classification performance and mimics the cooperative behaviour and mechanism of social spiders in nature. This proposed feature selection method has been tested on DEAP EEG dataset with six subjects and compared with the most popular heuristic algorithms such as GA, PSO and ABC. The results show that the SSO-FS provides a remarkable and comparable performance compared to other existing methods. Whereby, the max accuracy obtained is 66.66% and 70.83%, the mean accuracy obtained is 55.51$\pm$7.17 and 60.97$\pm$8.38 for 3-level of valence emotions and 3-level of arousal emotions classification respectively.

Keywords: *Electroencephalogram; Human emotion recognition; Social spider optimization; Swarm intelligence methods.*

1. Introduction

Brain computer interface (BCI) enables the computer to be interacted with other electronic devices over EEG signals. Currently, BCI has been widely used in numerous areas of life such as in health, neuro-marketing, criminal and entertainment. Nowadays, emotion recognition based on EEG signals attracts many researchers' attentions since emotions are certainly a vital aspect and crucial requirement in human life such as for learning, making decision and communication [1-2]. As a result, various emotion recognition computational techniques are used to detect the human emotions. In the past years, large numbers of features are generated in most applications. Such number of variables often lead to curse dimensionality problem and finding optimal subset of features in a complex data like EEG remain a challenging task [1-4]. Due to this issue, several features selection-based techniques are developed to address the problem of reducing irrelevant, redundant variables, correlated and highly dimensional data which are a burden on challenging tasks. Most popular, researchers have been used heuristics algorithms to determine and select the optimal features such as GA, PSO and ABC algorithm [4-7]. However, such algorithms have some drawbacks such as in GA, dynamically managing of data sets is complicated, and it is not a straight ideal to solve restriction optimization problems. On another side, PSO and ABC suffer from ticklish flaws problem, like premature convergence and the erroneous exploration–exploitation balance [8-10]. Due to such limitations of current heuristics and evolutionary algorithms, this paper proposed a new swarm based algorithm for feature selection method called Social Spider Optimization (SSO-FS). In this study, SSO-FS is applied in EEG-based emotion recognition model as searching method to find optimal feature set to maximize classification performance and mimics the cooperative behaviour and mechanism of social spiders in nature. This proposed feature selection method has been tested on DEAP EEG dataset with six subjects and compared with the most popular heuristic algorithms such as GA, PSO and ABC. To the best of researcher's knowledge, SSO has not yet used in EEG-based recognition field.

2. Social spider optimization

A heuristic algorithm such as GA and recently, evolutionary algorithms such as PSO and ABC have widely used and shown a remarkable performance in several studies and they become popular. In the last few years, the science revolution in computing has led to provide much evolutionary algorithm that inspired from the nature. One of these algorithms has recently proposed by [9] and named as social spider optimization (SSO) algorithm.

SSO algorithm is a population based algorithm inspired form the behaviour of social spider colony. Social spider colony is mainly composed of two major components that include social members and communal web. The social member is mainly divided into males and females. In the social spider colony, the number of female members reaches in range between 65% and 90% of the whole number of the members in the social colony. Meanwhile, the number of male members is hardly reaching at 30% of the total members of the colony.

The actual web of social spider acts as transference media, which mainly related to vibrations produced from the spiders. Each single spider on the web keeps a position as well as the quality (fitness) of the solution which is depending on the objective function. In spite that the spiders can easily and freely move on the web, they cannot depart from the web. Whenever a spider moves towards a position, it produces a vibration that is disseminated on the web. Each single vibration of individual spider holds information, and other spiders on the web can recognize it upon receiving the vibration.

In the SSO algorithm, the female spiders show like (attraction) or dislike (repulsion) state to other spiders according to their vibration based on the weight and their positions. Male spiders in contrast are divided into two main parts, (i) one is dominant spiders, (ii) and the other one is non-dominant spiders. The dominant spiders have

better fitness characteristics than non-dominant spiders and usually the female spiders present an attraction to the dominant spiders for mating. The dominant male spider can mate with more than on female within a specific range to produce new offspring, and the mating between them allows the information to be exchanged.

In SSO algorithm, the search space is represented by the communal web and each position of spider on this search space represents a solution, whereby the weight of each spider is represented by fitness value of its solution. One of the unique characteristics of the social spiders is the extremely female biased populations. The algorithm starts the initialization phase by identifying the number of spiders on the web (population (s)) of N spider members which is including both female ($f_i, i = 1,2, \dots, N_f$) and male ($m_i, i = 1,2, \dots, N_m$). The female number N_f is selected randomly in the range between 65% and 90% as in the following equation:

$$N_f = floor[(0.9 - rand. 0.25). N] \tag{1}$$

where $rand$ is a random number in range between [0, 1] and $floor$ maps a real number to be an integer number. Meanwhile, the remains spiders allocated to be male (N_m) and it is calculated as following equation:

$$N_m = N - N_f \tag{2}$$

The number of spiders in the web remains unrevised. Therefore, a restricted memory size must be assigned to store the spiders' information. The spider's positions are randomly produced inside search space along with their fitness values computed and stored. Therefore, the female spider position f_i as well as the male spider position m_i are generated randomly between the lower initial parameter bound p^{low} and the upper initial parameter bound p^{high} as in the following two equations:

$$f_{i,j}^0 = p_j^{low} + rand . \left(p_j^{high} - p_j^{low} \right) \tag{3}$$
$$m_{i,j}^0 = p_j^{low} + rand . \left(p_j^{high} - p_j^{low} \right) \tag{4}$$

where p^{low} the lower initial parameter bound and p^{high} is the upper initial parameter bound.

In the SSO algorithm the weight of every individual (spider) represents the solution quality, whereby the function value (w_i) of each spider i is calculated as the following equation and it is received for everyone.

$$w_i = \frac{J(s_i) - worst_s}{best_s - worst_s}, \tag{5}$$

where $J(s_i)$ the fitness value is obtained of the spider position s_i, and the values best and worst are the maximum and the minimum values of the solution for the population respectively (maximization problem).

The information among the colony members is transmitted through the communal web and encoded as a small vibration. The vibrations depend on the weight and distance of the spider which is generating it. The transmitted information (vibrations) perceived by the individual i from member j are modelled as follow.

$$Vib_{i,j} = w_j . e^{-d_{i,j}^2} \tag{6}$$

where the $d_{i,j}$ is the Euclidian distance between the spiders i and j. There are three special relationships of the vibrations between any pair of individuals as follows:

i. Vibrations $Vibc_i$. It is between the individual i and the member c (s_c), which is the nearest member to i with a higher weight can be defined as follow:

$$Vibc_i = w_c . e^{-d_{i,c}^2} \tag{7}$$

ii. Vibrations Vibb$_i$. It is between the individual i and the member b (s_b), which is the best member in the population S can be defined as follow:

$$Vibb_i = w_b . e^{-d_{i,b}^2} \tag{8}$$

iii. Vibrations $Vibf_i$. It is between the individual i and the nearest female individual f(s_f) can be defined as follow:

$$Vibf_i = w_f . e^{-d_{i,f}^2} \tag{9}$$

In the mataing behaviour of the social spiders, the female spiders present an attraction or repulsion movment over other irrespective of gender. These movement depends on several random phenomena. A uniform random number r_m is generated within the range [0,1]. If r_m is smaller than a threshold PF that is experimentally deteremined by user, then an attraction movement is generated. Otherwise, a repulsion movement is produced as in the following:

$$f_i^{t+1} =$$
$$\begin{cases} f_i^{t+1} + \alpha.Vibc_i.(s_c - f_i^t) + \beta.Vibb_i(s_b - f_i^t) + \delta.(rand - 0.5) \\ \qquad\qquad with\ probability\ PF \\ f_i^{t+1} + \alpha.Vibc_i.(s_c - f_i^t) - \beta.Vibb_i(s_b - f_i^t) + \delta.(rand - 0.5) \\ \qquad\qquad with\ probability\ 1 - PF \end{cases}$$
$$(10)$$

where 'α, β, δ and rand' are random numbers between [0,1] whereas 'k' represents the iteration number. The individual s_c and s_b represent the nearest member to i that holds a higher weight and the best individual of the entire population S respectively.

The male spider with a weight bigger than the median value of the male population is considered as dominant (D) spider. The other males with weights under the median are considered as non-dominant (ND). The median weight is indexed by $N_f + m$. The position of the male spider can be modeled as follows:

$$m_i^{t+1} =$$
$$\begin{cases} m_i^t + \alpha.Vibf_i.\left(s_f - m_i^t \right) + \delta.(rand - 0.5) & if\ w_{N_{f+i}} > w_{N_{f+m}} \\ m_i^t + \alpha.\left(\frac{\sum_{h=1}^{N_m} m_h^t.w_{m+h}}{\sum_{h=1}^{N_m} w_{N_{f+h}}} - m_i^t \right) \end{cases}$$
$$(11)$$

where the individual represents the nearest female individual to the male member i whereas ($\sum_{h=1}^{N_m} m_h^t.w_{m+h} / \sum_{h=1}^{N_m} w_{N_{f+h}}$) correspond to the weighted mean of the male population M and t is iteration number. The mating in a social spider colony is performed by the dominant males and the female members. Whereby, when a dominant male m$_g$ spider locates a set E$_g$ of female members within a specific range r (range of mating) which is calculated as follows:

$$r = \frac{\sum_{j=1}^n (p_j^{high} - p_j^{low})}{2.n} \tag{12}$$

The spider holding a heavier weight are more likely to influence the new product. The influence probability Ps$_i$ of each member is assigned by the roulette wheal method as in the following equation:

$$Ps_i = \frac{w_i}{\sum_{j \in T^k} w_j}, \tag{13}$$

where i $\in T^g$, T^g is new subset of new spider that generated from the the male subset and female memebres subsets from mating.

3. Methodology

As the methodology of this study, briefly, the extracting features have been performed using DWPT [11] algorithm whereby four

frequency bands include θ, α, β and γ have been extracted from 6 subjects include subject 7, 8, 10, 16, 19 and 20 of the pre-processed DEAP [9] dataset provided by [12]. These features were extracted from 32 channels and 40 trials; this resulting 7680 features (128 feature vectors) and entropy value for each have been calculated. Moreover, DEAP [12] database provide four emotional states based on valence and arousal dimension include LALV, HALV, LAHV and HAHV whereby L = low, H = high, A = Arousal, V = Valence. However, like study of [13], this study mapped the emotion labels into 3-level of valence and arousal which resulting 9 emotion states include distressed, miserable, depressed, excited, neutral, calm, happy, pleased and relax.

Basically, three experiments have been performed; (i) emotion classification with the proposed SSO-FS. (ii) Emotion classification with feature selection using GA, PSO and ABC algorithms have separately implemented and (iii) emotion classification without feature selection where the whole extracted features are input into the classifier. On another side, the least square support vector machine (LS-SVM) proposed by [4, 14-16] has been used for classification. Whereby, the objective function of the optimization process has consisted of LSSVM along with Cross-validation (CV) with 10-folds and classification accuracy is calculated as follows to be as a fitness value:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN},$$

where TP, TN, FP and FN are the true positive points, negative points, false positive points and negatives points respectively.

This section describes the developed SSO-FS that performed in the context of this study. First, the parameters of SSO-FS setup as illustrated in Table 1 where 'SN' is the number of the spider in the colony, 'lb' and 'ub' is the lower bound and upper bound respectively, CB_{th} is the "code to binary" threshold (the threshold used to code the real number to binary), 'dims' is the dimension of spider and Max_{itr} is max iteration.

Table 1: Parameter setup for SSO-FS

Parameter	SN	lb	ub	dims	CB_{th}	Max_{itr}
Number	40	0	1	128	0.5	100

The proposed SSO-FS has been performed as in the following steps:

3.1. Step 1: Initialization

- Randomly generate initial population of spider colony S, female and male spiders with position (solution) for each randomly in the range [0,1] (Equations 1, 2, 3 and 4)
- Codes the continuous value of spider position which is equal to the length of position dimension (dims), from real number to binary based on CBth; if value bigger than threshold, set to 1 (selected), otherwise, set to 0 (not selected).
- Check if spider selects at least one solution (one of its position value equal '1'). If not, then generate new position and repeat the coding.
- Evaluate each spider position, and calculate the weight for each spider (Equation 5).

3.2. Step 2: Female movement

- Calculate the vibration $Vibc_i$ and $Vibb_i$. Equations (7 and 8).
- A random number generating for each female spider in range (0,1); if a random number of each female spider is less than PF, then female change its position by doing attraction. Otherwise, female change its position by doing apulsion (Equation 10).

3.3. Step 3: Male movement

- Calculate the median male spider **M.**
- Calculate the vibration $Vibf_i$ (Equation 9).
- Determine dominant and non-dominant male. Dominant male changes its position by being attracted by the nearest female and non-dominant males change its position by moving to the middle of dominant male position (Equation 11).

3.4. Step 4: Mating

- The mating is performed between male and female spiders within a specific distance. If offspring is produced, then the mating is done and a survive operation is performed. Otherwise, the mating operation is cancelled (Equation 12 and 13).

3.5. Step 5: Stopping criteria

The searching process is ended when the maximum cycle is met, and then best solution is produced, otherwise, back to step 2.

4. Results and discussion

This section provide discussion on empirical findings results of the undertaken experiments over 20 normal runs with 100 iterations for each run. Table 3 illustrates the EEG-based emotion classification performance of the experiments; where number of feature vectors (#of FV) of maximum accuracy, minimum (min), maximum (max), mean accuracy and standard deviation (SD) have been obtained. Upon the obtained results, it is shown that feature selection has improved the classification performance compare with when no feature selection performed.

On another side, it can be noticed that SSO-FS outperforms GA, PSO and ABC in minimum, maximum, mean, standard deviation. The performance of SSO-FS proofs the capability of the social spider to adaptively search the large feature space and its remarkable capability in balancing between exploration and exploitation to some extent. In another hand, SSO-FS has not superior with minimum accuracy and has shown a same maximum accuracy with other algorithms in arousal classification. This can be expressed due to the slow down convergence towards the optimal solution.

Table 2: EEG-based emotion classification performance without feature selection and by four features selection algorithms

Algorithm		# of FV	Min %	Max %	mean %	SD
Without selection	V	128	23.61	56.94	41.77	5.09
	A	128	33.33	68.06	53.34	5.06
GA	V	58	41.67	63.89	50.88	3.88
	A	78	48.61	70.83	58.25	4.25
PSO	V	68	43.06	65.28	55.48	4.84
	A	60	50	70.83	60.15	4.46
ABC	V	57	45.83	65.28	54.56	4.34
	A	61	50	70.83	60.05	3.90
SSO-FS	V	68	49.85	66.66	55.51	3.75
	A	71	48.61	70.83	60.97	3.84

5. Conclusion and future work

To assess the performance of the proposed features selection (SSO-FS), this study performed several different experiments to be compared as benchmark algorithms include GA, PSO and ABC. SSO-FS is basically developed to enhance the performance in searching with a large and complex feature space of EEG signals. The results show that the SSO-FS provides a remarkable and comparable performance with other existing methods and it ensures its searching capability with complex feature space.

On another side, regardless of that SSO have bestead exploration-exploitation balance compare with other existing algorithms, it is observed that SSO algorithm suffer from some drawbacks; (i) the exploration part in SSO seems to be weak. This is because the spiders in SSO are classified by gender, whereby male spiders responsible for exploration and female spiders responsible for exploitation. But, the number of male spiders is very few compare with female spiders; (ii) in somehow, SSO forces itself to be far from the optimal solution, to avoid rapid convergence, at the event that an accelerate convergence is may needed during searching process. Hence, SSO algorithm can be improved in these parts and it considered as future work.

Acknowledgement

References

[1] Liu Y & Sourina O (2013), EEG databases for emotion recognition. *Proceedings of the IEEE International Conference on Cyberworlds*, pp. 302–309.

[2] Kim MK, Kim M, Oh E & Kim SP (2013), A review on the computational methods for emotional state estimation from the human EEG. *Computational and Mathematical Methods in Medicine* 2013, 1–13.

[3] Kumar S & Lopez D (2014), A review on feature selection methods for high dimensional data. *International Journal of Engineering and Technology* 8, 669–672.

[4] Al-Qammaz AY, Ahmad FK & Yusof Y (2016), Optimisation of least squares support vector machine technique using genetic algorithm for electroencephalogram multi-dimensional signals. *Jurnal Teknologi* 78, 107–115.

[5] Hsu WY & Hu YP (2015), Artificial bee colony algorithm for single-trial electroencephalogram analysis. *Clinical EEG and Neuroscience* 46, 119–125.

[6] Yavuz G & Aydin D (2016), Angle modulated artificial bee colony algorithms for feature selection. *Applied Computational Intelligence and Soft Computing* 2016, 1–7.

[7] Wang J & Zhao Y (2014), EEG signal classification with feature selection based on one-dimension real valued particle swarm optimization. *Proceedings of the International Conference on Mechatronics, Control and Electronic Engineering*, pp. 310-314.

[8] Binitha S & Sathya SS (2012), A survey of bio inspired optimization algorithms. *International Journal of Soft Computing and Engineering* 2, 137–151.

[9] Cuevas E, Cienfuegos M, Zaldívar D & Pérez-Cisneros M (2013), A swarm optimization algorithm inspired in the behavior of the social-spider. *Expert Systems with Applications* 40, 6374–6384.

[10] Pereira LA, Rodrigues D, Ribeiro PB, Papa JP & Weber SA (2014), Social-spider optimization-based artificial neural networks training and its applications for Parkinson's disease identification. *Proceedings of the IEEE 27th International Symposium on Computer-Based Medical Systems*, pp. 14–17.

[11] Wali MK, Murugappan M & Ahmmad B (2013), Wavelet packet transform based driver distraction level classification using EEG. Mathematical Problems in Engineering 2013, 1–10.

[12] Koelstra S, Muhl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, Pun T, Nijholt A & Patras I (2012), Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing* 3, 18–31.

[13] Jirayucharoensak S, Pan-Ngum S & Israsena P (2014), EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation. *Scientific World Journal* 2014, 1–10.

[14] Suykens JA & Vandewalle J (1999), Least squares support vector machine classifiers. *Neural Processing Letters* 9, 293–300.

[15] Mustaffa Z & Yusof Y (2014), LSSVM parameters tuning with enhanced artificial bee colony. *International Arab Journal of Information Technology* 11, 236–242.

[16] Lai L & Liu J (2014), Support vector machine and least square support vector machine stock forecasting models. *Computer Science and Information Technology* 2, 30–39.