

## Research Article

### Firefly Algorithm with Artificial Neural Network for Time Series Problems

Mohammed Alweshah

Prince Abdullah Bin Ghazi Faculty of Information Technology,

Al-Balqa Applied University, Salt, Jordan

**Abstract:** Time series classification is a supervised learning method maps the input to the output using historical data. The primary objective is to discover interesting patterns hidden in the data. For the purpose of solving time series classification problems used the multi-layered perceptrons Artificial Neural Networks (ANN). The weights in the ANN are modified to provide the output values of the net, which are much closer to the values of the preferred output. For this reason, several algorithms had been proposed to train the parameters of the neural network for time series classification problems. This study attempts to hybrid the Firefly Algorithm (FA) with the ANN in order to minimize the error rate of classification (coded as FA-ANN). The FA is employed to optimize the weights of the ANN model based on the processes. The proposed FA-ANN algorithm was tested on 6 benchmark UCR time series data sets. The experimental results have revealed that the proposed FA-ANN can effectively solve time series classification problems.

**Keywords:** Artificial neural networks, firefly algorithm, time series problems

#### INTRODUCTION

Time series data are produced, continuous and are processed within a wide range of application domains in different fields such as economics, engineering, science, medical and sociology. There has been an explosion of interest within the data mining community in pattern discovery, segmenting, clustering and classifying time series (Fu, 2011; Keogh *et al.*, 2001; Lin *et al.*, 2003, 2004). In recent years, the popularity of time series data has been growing rapidly.

This type of data occurs every second, every minute, hourly and daily in several applications. This kind of data produces massive volumes of data stamped with time and this time is one of the essential aspects of many real-life data which people want to use to predict movements or trends in data. The areas that have occupied the attention of most time series data mining research are classification, clustering, detection and rule discovery. Time series data mining refers to data mining performed on temporal data. The primary objective is to discover interesting patterns hidden in the data (Fu, 2011).

Many research studies have been conducted on issues related to time series data, for instance, time series classification. However, in the time series classification problem there are labels for each time point. A classifier is trained using the values and labels from the time points of the training set. Given a new time series, the task is to label all time points. This is known as sequential supervised learning and is related to prediction (Fu, 2011).

However, studies on the use of time series classification in actual applications is lacking in the literature. Artificial Neural Network (ANN) is a widely used tool for the prediction of time series that has been used in several different applications. ANN has also been applied in time series applications for prediction and classifications (Hüsken and Stagge, 2003).

In this study, we investigate the search capability of the Firefly Algorithm (FA) in finding optimal values for the weights of the ANN. To our knowledge, this is the first attempt where FA is experimented with ANN for time series classification problems. The ANN is employed first to obtain the initial solution and later the weights of the ANN will be adjusted by the FA in order to handle the time series data problem and minimize the error rate.

#### LITERATURE REVIEW

**Firefly algorithm:** A population based FA was developed by Yang (2009) for solving optimization problems. Short and the rhythmic flashing light produced by fireflies are the motivational aspect of the FA, The flashing lights of the fireflies enable them to attract each other and assist them to, find their mates, reach their prey and protect themselves from their predators, by creating a sense of fear in the minds of enemies (Apostolopoulos and Vlachos, 2011). It is noteworthy that, the less bright fireflies are easily attracted by the bright fireflies. This procedure can be created as an optimization algorithm, since the flashing light can be designed to be coordinated with the fitness

function to be optimized. The firefly algorithm follows three rules:

- Fireflies must be unisex
- Lighter firefly is attracted towards the randomly moving brighter fireflies
- The brightness of every firefly symbolizes the quality of the solutions

The diversity in the Firefly Algorithm (FA) optimization is depicted by the random movement component, whilst the intensification is unconditionally manipulated by the attraction of various fireflies and the strength of attractiveness. As opposed to the other meta-heuristics, the association between exploration and exploitation in FA are relatively inter-connected; this might be a significant factor for its success, in solving multi-objective and multi-modal optimization problems.

Lukasik and Żak (2009) have employed FA for addressing continuously constrained optimization tasks and has consistently outperformed the particle swarm optimizer. Yang (2009) has employed and compared firefly algorithm with particle swarm optimization, for various test functions and has obtained better results in terms of efficiency and success rate, as against particle swarm and genetic algorithm. The transmitting ability of FA gives better and quicker convergence, towards the optimality. A similar work has been carried out by Yang and Deb (2010) and experimental results have revealed that, firefly algorithm has relatively outperformed other approaches. Sayadi *et al.* (2010) has proposed firefly with local search for minimizing makespan in permutation flow scheduling problems. The initial results have indicated that, the proposed method has outshined the ant colony algorithm. Gandomi *et al.* (2011) has applied firefly algorithm for mixed variable structural optimization problems. The investigations have revealed that, firefly has performed better than particle swarm optimization, genetic algorithm, simulated annealing and harmony search algorithms. Ultimately another study on firefly algorithm can be found Apostolopoulos and Vlachos (2011). The success of the firefly motivates this present study to further investigate its performance over the time series problems.

## METHODOLOGY

**Firefly algorithm with Artificial Neural Networks (ANN):** Rumelhart *et al.* (1986) have proposed the multi-layered perceptrons (Artificial Neural Networks) for the purpose of solving time series classification problems. Zhang (1999) has reviewed current works in Artificial Neural Networks (ANNs). A multi-layer neural network comprises of a huge volume of units (neurons), which are linked with each other in a pattern of connections. Generally the units in a net are divided into three classes:

- Input units (receive information to be processed)
- Output units (which, presents the outcomes of the processing)
- Hidden-units (which, lie between the input and output units)

Feed-forward ANNs enables one-way transmission of the signals i.e., from input to output. Initially, the network is trained on a collection of combined data, to map the input-output. The weights of the associations between neurons are then set and the network are utilized to identify the classifications of a new set of data.

Throughout the process of time series classification, the signal at the input units completely distributed via the net, to establish the initial values at all the output units. Every single input unit has an initial value, which signifies some external feature of the net. Later, each and every input unit transmits its initial value to all the associated hidden units. All these hidden units compute their own initial value and the signals are then sent to the output units. The initial value of each receiving unit is computed in accordance with simple activation function. The function chunks the contributions of all sending units, in which, the contribution of a unit is identified as the weight of the connection between the sending and receiving units, multiplied by the initial value of the sending unit. Normally this value is then further customized, such as, by modifying the initial sum to a value between 0 and 1 and/or by establishing the initial value to zero, until a threshold level of that sum is achieved. The ANN relies upon three fundamental aspects such as, input and activation functions of the unit, network architecture and the weight of each input connection. Provided that, the 1st two aspects are set, the performance of the ANN is characterized by the present values of the weights. The weights of the net to be trained are originally fixed to arbitrary values and then samples of the training set are continuously revealed to the net. The values for the input of a sample are entered to the input units and the output of the net is compared with the preferred output for this sample. Later, all the weights in the net are marginally modified to provide the output values of the net, which are much closer to the values of the preferred output. A lot of multi-layered perceptrons algorithms are available, for training networks (Neocleous and Schizas, 2002).

In this study, the FA is employed to optimise the weights of the ANN model, denoted as FA-ANN, to obtain the optimal parameter settings for training the network of ANN and to minimize the error rate.

The pseudo code of the FA is shown in Fig. 1. At the first, it generates the initial population of candidate solutions for the given problem (here, the weights of the ANN). After that, it calculates the light intensity for all fireflies and finds the attractive firefly (best candidate) within the population. Then, calculate the attractiveness and distance for each firefly to move all fireflies towards the attractive firefly in the search space.

```

Algorithm: FA-ANN.
Begin
Generate the initial solution randomly
Evaluate each individual in the population f(x) based on error rate.
Find the best solution from the population
While (stopping criterion satisfied)
  For i = 1 to n do
    For j = 1 to n do
      If (f(xj) < f(xi) )
        Calculate attractive fireflies by Eq.1
        Calculate the distance between each fireflies i and j by Eq.2
        Move all firefly (xi) to the best solution (xj) by Eq.3
      End if
    End for j
  End for i
  Moves best solution randomly by Eq.4
  Find the best solution from the new population
End while
Return best (TP), (TN), (FP), and (FN)
End of the algorithm.
    
```

Fig. 1: Pseudo code for FA-PNN

Finally, the attractive firefly moves randomly in the search space. This process is repeated until a termination criterion is met i.e., the maximum number of generations is reached.

Normally, the quality of the time series classification is measured based on the error rate that can be calculated based on True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) as shown in Fig. 1 (Al-Obeidat *et al.*, 2010; Yeh *et al.*, 2007).

The procedure starts from an initial population of randomly generated individuals. The quality of each individual is calculated using Eq. (1) and the best solution among them is selected.

In FA, the form of attractiveness function of a firefly is depicted by the following:

$$\beta(r) = \beta_0 \exp(-\gamma r^2) \tag{1}$$

where,

- r = The distance between any two fireflies
- $\beta_0$  = The initial attractiveness at r = 0 and set to 1 in this study
- $\gamma$  = An absorption coefficient which controls the decrease of the light intensity and also set to 1 in this study

The distance between any two fireflies i and j, at positions  $x_i$  and  $x_j$ , respectively, can be defined as a Cartesian or Euclidean distance as follows:

$$r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^d (x_{i,k} - x_{j,k})^2} \tag{2}$$

where, d is the dimensionality of the given problem.

The movement of a firefly i which is attracted by a brighter firefly j is represented by the following equation:

$$x_i = x_i + \beta_0 \cdot \exp(-\gamma r_{ij}^2) \cdot (x_j - x_i) + \alpha \cdot \left( rand - \frac{1}{2} \right) \tag{3}$$

$$x_i = x_i + \alpha * \left( rand - \frac{1}{2} \right) \tag{4}$$

In Eq. (3), the first term is the current position of a firefly, the second term is used for considering the attractiveness of a firefly (attractive firefly), towards the intensity of light by neighbouring fireflies and the third term is used for the random movement of a firefly (random part), when it lacks the brighter ones. The coefficient  $\alpha$  is a randomization parameter determined by the problem of interest, while rand is a random number generator consistently distributed in the space (0, 1) (Hung *et al.*, 1997). In Eq. (4), the movement of the best candidate is done randomly.

### EXPERIMENTAL RESULTS

The proposed algorithm is implemented using MATLAB and simulations are performed on an Intel Pentium 4, 2.8 GHz computer. We execute 10 independent runs for each datasets.

**Benchmark datasets:** This experiment is performed on 6 datasets that can be freely downloaded from the UCR Time Series Classification/Clustering Homepage: [www.cs.ucr.edu/~eamonn/time\\_series\\_data](http://www.cs.ucr.edu/~eamonn/time_series_data). The data

Table 1: UCR time series data characteristics

No.	Name	Number of classes	Size of training set	Size of testing set	Time series length
1	Gun-point	2	50	150	150
2	Wafer	2	1000	6174	152
3	Lightning-2	2	60	61	637
4	ECG	2	100	100	96
5	Yoga	2	300	3000	426
6	Coffee	2	28	28	286

Table 2: Parameter settings

Parameter	Value
Population size (# of fireflies)	50
Number of iterations	100
Initial attractiveness ( $\beta_0$ )	1
Absorption coefficient ( $\gamma$ )	1

Table 3: Error rate (%) for ANN and hybrid between firefly and ANN

Datasets	TP	FP	TN	FN	ANN	FA-ANN
Gun-point	68	8	65	9	0.113	0.080
	69	7	69	5		
Wafer	5480	19	656	9	0.005	0.004
	5481	18	656	9		
Lightning-2	27	6	19	9	0.245	0.197
	31	2	18	10		
ECG	59	5	29	7	0.120	0.080
	59	5	33	3		
Yoga	1097	296	1398	209	0.168	0.161
	1105	288	1411	196		
Coffee	11	2	10	5	0.250	0.107
	13	0	12	3		

Table 4: Results comparison between FA-ANN and state-of-the-art approaches

No.	Data sets name	1-NN Euclidean distance	1-NN best warping window DTW	1-NN DTW, no warping window	FA-ANN
1	Gun-point	0.087	0.087	0.093	0.080
2	Wafer	0.005	0.005	0.020	0.004
3	Lightning-2	0.246	0.131	0.131	0.197
4	ECG	0.120	0.120	0.230	0.080
5	Yoga	0.170	0.155	0.164	0.161
6	Coffee	0.250	0.179	0.179	0.107

contains 6 time series data sets, which come from different domains (Table 1).

All the UCR data sets are categorized as having similar complexity to real-world data sets with the data sets based on several criteria. All the 6 benchmark UCR time series data sets have a moderate to high time series length that ranges from 96 to 637 time series length.

**Parameter settings:** Table 2 shows the parameters for the proposed algorithm which were determined after some preliminary experiments.

**Experimental results:** Table 3 presents the comparison of the error rate (%) between FA-ANN and ANN time series classification techniques with 6 datasets. The results clearly indicate that the hybrid method (FA-ANN) has outperformed the ANN algorithm on all datasets. For example, in the Gun-Point dataset the ANN has achieved 11.33% error rate, while the proposed FA-ANN obtained 0.08% of error rate. It is due to capability of the FA which incorporated into ANN to find the optimal weights for the ANN and consequently increase the performance of the ANN. This is believed that fireflies come together more closely around the optimal solution. In other words, it has good exploitation capability and can find better solutions as many candidates (fireflies) are gathered near optimal solution.

**Comparison with state-of-the-art:** Table 4 shows the comparison of the results of FA-ANN and other available approaches in terms of error rate classification using 6 datasets. The best results are presented in bold.

The experimental results indicate that the proposed hybrid method (FA-ANN) outperforms other approaches on four out of six datasets (i.e., Gun-Point,

Wafer, ECG and Coffee). FA-ANN is able to classify the Wafer with error rate of 0.004%. This capability is supported by the feature of the attractiveness i.e., the density of the light that caused the fireflies to be brighter (is determined by the value of the objective function) and attract to the location of near optimal solutions.

## CONCLUSION

In this study a hybrid method based on the ANN and FA is proposed for solving time series classification problems. Initial solutions are generated at random using ANN and the improvement is carried out by the FA that tries to optimize the weights of the ANN. Experiments results using 6 benchmark UCR time series data sets show that the proposed FA-ANN outperforms the ANN on all datasets. Further comparison with other approaches in the literature shows that the hybrid method is able to minimize the error rate with new best results on 4 out of 6 datasets. As an extension of this study, further investigation will be devoted to validate the hybridization between FA with local search algorithm for the purpose of creating a balance between the exploration and exploitation during the optimization process and to avoid the premature convergence.

## REFERENCES

- Al-Obeidat, F., N. Belacel, J. Carretero and P. Mahanti, 2010. Automatic Parameter Settings for the PROAFTN Classifier using Hybrid Particle Swarm Optimization. In: Farzinder, A. and V. Keselj (Eds.), Canadian AI 2010. LNAI 6085, Springer-Verlag, Berlin, Heidelberg, pp: 184-195.

- Apostolopoulos, T. and A. Vlachos, 2011. Application of the firefly algorithm for solving the economic emissions load dispatch problem. *Int. J. Combinat.*, Article ID 523806, pp: 1-23.
- Fu, T.C., 2011. A review on time series data mining. *Eng. Appl. Artif. Intel.*, 24(1): 164-181.
- Gandomi, A.H., X.S. Yang and A.H. Alavi, 2011. Mixed variable structural optimization using firefly algorithm. *Comput. Struct.*, 89: 2325-2336.
- Hung, H.M.J., R.T. O'Neill, P. Bauer and K. Kohne, 1997. The behavior of the p-value when the alternative hypothesis is true. *Biometrics*, 53(1): 11-22.
- Hüsken, M. and P. Stagge, 2003. Recurrent neural networks for time series classification. *Neurocomputing*, 50: 223-235.
- Keogh, E., K. Chakrabarti, M. Pazzani and S. Mehrotra, 2001. Dimensionality reduction for fast similarity search in large time series databases. *Knowl. Inform. Syst.*, 3(3): 263-286.
- Lin, J., E. Keogh, S. Lonardi and B. Chiu, 2003. A symbolic representation of time series, with implications for streaming algorithms. *Proceeding of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, pp: 2-11.
- Lin, J., E. Keogh, S. Lonardi, J.P. Lankford and D.M. Nystrom, 2004. Visually mining and monitoring massive time series. *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp: 460-469.
- Łukasik, S. and S. Żak, 2009. Firefly Algorithm for Continuous Constrained Optimization Tasks. In: Nguyen, N.T., R. Kowalczyk and S.M. Chen (Eds.), *ICCCI 2009*. LNAI 5796, Springer-Verlag, Berlin, Heidelberg, pp: 97-106.
- Neocleous, C. and C. Schizas, 2002. Artificial Neural Network Learning: A Comparative Review. In: Vlahavas, I.P. and C.D. Spyropoulos (Eds.), *SENT 2002*. LNAI 2308, Springer-Verlag, Berlin, Heidelberg, pp: 300-313.
- Rumelhart, D.E., G.E. Hinton and R.J. Williams, 1986. Learning representations by back-propagating errors. *Nature*, 323(6088): 533-536.
- Sayadi, M.K., R. Ramezani and N. Ghaffari-Nasab, 2010. A discrete firefly meta-heuristic with local search for makespan minimization in permutation flow shop scheduling problems. *Int. J. Ind. Eng. Comput.*, 1(1): 1-10.
- Yang, X.S. 2009. Firefly algorithms for multimodal optimization. *Lect. Notes Comput. Sci.*, 5792: 169-178.
- Yang, X.S. and S. Deb, 2010. Eagle Strategy Using Lévy Walk and Firefly Algorithms for Stochastic Optimization. In: Gonzalez, J.R. *et al.* (Eds.), *Nature Inspired Cooperative Strategies for Optimization (NICSO, 2010)*. Springer, Berlin, 284: 101-111.
- Yeh, W., W. Chang and C. Chiu, 2007. Using the novel discrete particle swarm optimization for mining breast cancer database. *Proceeding of International Conference on Business and Information (BAI, 2007)*. San Francisco, CA.
- Zhang, G., 1999. Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *Eur. J. Oper. Res.*, 116(1): 16-32.