

SHUFFLED FROG LEAPING ALGORITHM AND FEATURE SELECTION FOR IMPROVING RECOGNITION RATE OF PERSIAN HANDWRITTEN DIGITS CLASSIFIER

N. Ghanbari¹, M. Heidari²

^{1,2}Department of Electrical Engineering, University of Zabol, Zabol, Iran.
najme.ghanbari@gmail.com¹

Received 24/07/2017 – Accepted 08/09/2017

DOI: 10.15628/holos.2017.6144

ABSTRACT

In this paper, Shuffled Frog Leaping Algorithm is used to improve the recognition rate of Persian handwritten digits. In proposed approach, the effective features in increasing the recognition rate are selected using the Binary Shuffled Frog Leaping Algorithm (BSFLA). By selecting the most suitable features from among all extracted features, the recognition rate is improved and

computational costs are also decreased. The fitness function in BSFLA is the number of errors in the Fuzzy classifier which its minimum value is desired. The results indicate that Shuffled Frog Leaping algorithm (SFLA) is more efficient.

KEYWORDS: Persian handwritten digits recognition, Shuffled Frog Leaping algorithm (SFLA), features selection.

1 INTRODUCTION

One of the major issues in pattern recognition is the recognition of letters, digits, and words. The research in this field has been started 60 years ago and still continues. In different studies, various methods have been applied for recognition including statistical, structural, neural network and fuzzy methods. Also, many studies have been conducted in the field of handwritten digits' recognition. In a study on the recognition of Persian handwritten digits, characteristics of places, Bays classification, and Markov chain are used (Nafisi, 199). The final recognition of Persian handwritten digits has been conducted by using fuzzy methods in Johari et al's study (Johari & Razavi, 2001). In another study, the shape matching algorithm and classification of the nearest distance from the representative (s) of each class have been used for recognition of Persian handwritten digits (Darvish, Kabir, & Khosravi, 2005). In his research, Khosravi (Khosravi, 2006) used the classification by neural network and combination of three features of characteristics of places, improved gradient, krish to recognize digits. Nahvi et al. (Miri, 2013; Nahvi, 2008) used the neural network and combination of two-class-ten classification to recognize the Persian handwritten digits. In another study, Miri et al. (Miri, 2013) used K-mean clustering and Particles Swarm Optimization to improve the recognition rate. Also, using optimization algorithms in solving the various problems have been dramatically increased and Day by day, the number of these algorithms is added.

Shuffled Frog Leaping Algorithm (SFLA) firstly used by Eusuff and Lansey in 2003 to determine the optimum size of new pipes in development of network of pipes (Eusuff & Lansey, 2003); for instance, this algorithm was used in (Bi Jami, 2011; Ranjkesh, 2013; Seyedi, 2011; Shahriari Kahshki, 2011). In this study, Shuffled Frog Leaping Algorithm (SFLA) is used to increase the recognition rate. The binary version of Shuffled Frog Leaping Algorithm (SFLA) can be used to select the optimum features for handwritten digits' recognition system. Each frog is an answer in the problem space which has a position. Some features are considered in any position (corresponding to 1 value) and some feature are ignored (corresponding to zero value). Finally, the most suitable features are selected to increase the recognition rate. In this paper, the used database, feature extraction, SFLA algorithm, type of fuzzy classifier is briefly explained in section 2. In section 3, the two proposed methods for increasing the recognition rate of Persian handwritten digits are described. Section 4 reports the obtained results and section 5 provides the conclusion.

2 MAIN POINTS

2.1 Database

Recently, many studies (e.g., (Miri, 2013) and (Ghanbari, 2010) references) done in the field of handwritten digits' recognition used Hoda database which is a standard database and has been presented by Khosravi and Kabir (Khosravi & Kabir, 2007). This database has two large subsets for letters and digits; the digits' database is used in this study.

This database has 102352 digits in which 60000 digits are considered as training samples, 20000 digits as test samples and 22352 digits as remaining samples. The remaining digits are used differently in different issues.

2.2 Feature Extraction

Selecting the feature extraction method is one of the most important and effective steps in a pattern recognition system. Different features such as zoning features, Geometric moments, Zernike moments, instant descriptors, invariant moments, histogram display, projections, features of characteristics of places, krish, border profile, gradients, gradient histogram, etc. have been used (Darvish et al., 2005; Johari & Razavi, 2001; Nafisi, 199; Rashedi, Nezamabadi-Pour, & Saryazdi, 2009, 2010; Razavi, 2001). In this study, zoning features are used for recognition due to the less computational complexity.

2.3 Fuzzy model

There are 10 rules in the introduced Fuzzy system in reference (Johari & Razavi, 2001); in which each rule recognizes each digit. An example of these rules can be seen in the reference (Johari & Razavi, 2001). The Fuzzy set $A_{i,j}$ is obtained from training samples (Johari & Razavi, 2001). The algebraic multiplication is utilized for “and” operator.

2.4 Shuffled Frog Leaping Algorithm (SFLA)

Shuffled Frog Leaping Algorithm (SFLA) was used for the first time by Eusuff and Lansey in 2003 to determine the optimum size of new pipes in the development of network of pipes (Eusuff & Lansey, 2003). This algorithm is a population-based algorithm which acts according to the laws of probability and random search. This algorithm is inspired by the frogs' behavior in finding food in the wetlands. In solving the engineering problems, positions that have higher fitness have more food; the aim of frogs is also to find more food. Like other population-based algorithms, there are initial population of frogs that are randomly dispersed in the search space. In SFLA, each frog is an answer in the problem space. In this algorithm, X_i is the position of i frog at the D dimensional space and $f(X_i)$ is the amount of food in the frog position which is the same as its fitness value.

$$X_i = [X_{i,1}, X_{i,2}, \dots, X_{i,D}] \quad (1)$$

SFLA as the other collective intelligence algorithm is an iterative algorithm that the following steps are performed in each iteration.

- 1- Population Members are sorted according to their fitness value in a way that the best frog is placed in index 1.
- 2- Frogs are divided into m group called as memplexes. Each memplexes has n frog. The division is as the first frog is the member of the first memplexes, second frog member of the second memplexes and m frog member of memplexes. Then, $m+1$ frog is the member of first memplexes and so on continues until finally there will be n frog in each memplexes.
- 3- After the division of frogs, each memplexes repeats the following procedure to the number of problem's iteration. X_b and X_w are the position of best and worst frog in terms of fitness value. Then, a candidate position is obtained using the following formulas.

$$D = \text{Rand} \cdot (X_b - X_w) \quad (2)$$

$$X'_w = X_w + D \quad , \quad D_{\min} \leq D \leq D_{\max} \quad (3)$$

After obtaining the position X'_w , its fitness value is measured. If the new fitness value is better than the previous position, X_w is moved to the new position; otherwise the equations 2 and 3 rerun with the exception that X_g , the position of best frog among all memplexes, is used instead of the position X_b . If after this stage, the value of X'_w is not better than X_w , this frog randomly be initialized again.

After running the stage 3 to the number of problem iteration, condition of the end of algorithm is checked. If this condition is not fulfilled, all frogs fall in one category and the stages 1 to 3 reruns.

In this study, the binary version of SFLA algorithm is used. In this method, the equations 2 and 3 are changed to the equations 4 and 5.

$$X'_w = X_w \oplus D \quad (4)$$

$$D = (C_1 \cdot (X_b \oplus X_w) + C_2 \cdot (X_g \oplus X_w)) \quad (5)$$

Where C_1 and C_2 are random binary vectors. Details of this method are provided in (Gomez-Gonzalez, Ruiz-Rodriguez, & Jurado, 2014).

Figure 1 illustrate the flow-chart of BSFLA.

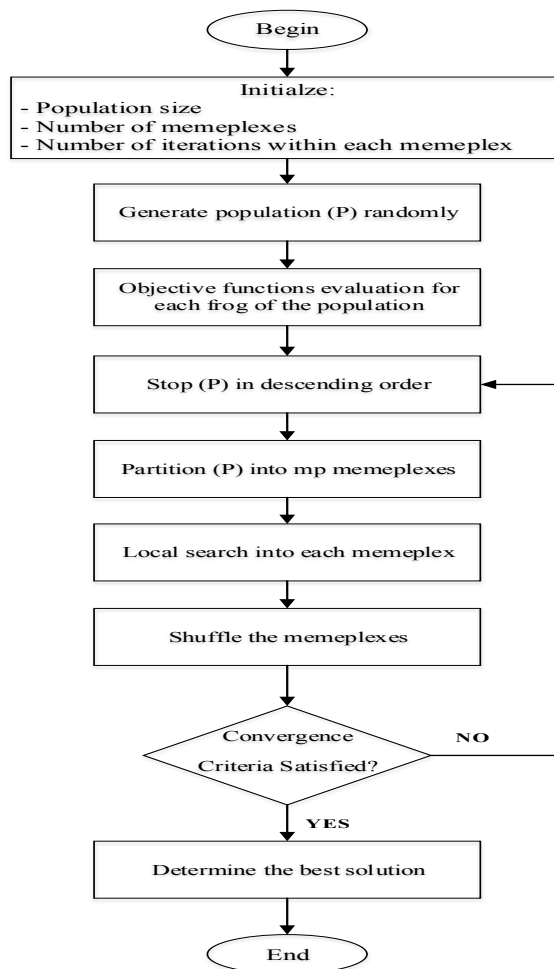


Figure 1. Flow- chart of the proposed BSFLA algorithm(Gomez-Gonzalez et al., 2014)

3 SUGGESTION FOR IMPROVING RECOGNITION RATE

In the proposed method, the effective features on increasing the recognition rate are selected using the binary version of Shuffled Frog Leaping Algorithm.; this has been done using other binary population-based algorithms such as genetics, particles swarm, and gravitational search (Ghanbari, 2010). The aim of this study is to indicate the effectiveness of the BSFLA algorithm. In this method, 64 features are extracted from each digit by zoning method. Then, the appropriate features are selected to have higher recognition rate from among these features. The initial population of memeplexes is randomly generated. Each memeplexes can be an answer to the problem which is defined as a chain of zero and one (the value of zero stands for non-participation of the feature and one stands for participation of the feature). Then, the fitness value is calculated for each memeplexes. At each step, the best memeplexes is selected as a response. If the stop condition has been fulfilled (the number of algorithm iterations), the memeplexes with the lowest fitness function is selected as the feature vector. Initial population in our algorithm has approximately 60 or 100 memeplex. Recognition rate significantly improves through this method. The number of features also decreases from 64 to 34 (for example). Results have been shown in Table 1. As well as, the obtained results confirmed that BSFLA has better performance than the Binary Genetic Algorithm (BGA) used.

To recognize digits, the zoning features are used due to the less computational complexity.

The steps of proposed method:

1. Extraction of digits features (zoning method)
2. Applying the fuzzy method in reference (Johari & Razavi, 2001)
3. Selecting the top features by BSFLA algorithm
4. Reapplying the method of reference (Johari & Razavi, 2001), with obtained features from step3

4 RESULTS

As mentioned earlier, the applied method in reference (Johari & Razavi, 2001) was used for digits' recognition. At first, the applied method was implemented by using the zoning and Hoda database. Then, the feature selection was carried out using the BSFLA algorithm. The method (Johari & Razavi, 2001) was implemented again with new obtained features. The results related to the proposed method were reported in the table 1. Proposed method increased the recognition rate which confirm the proper performance of SFLA algorithm in solving the problems of pattern recognition (here digits' recognition). Also, one parameter called the generalized strength was reported in table which is defined as follows:

$$\text{Generalized strength} = \frac{(\text{the difference between the recognition rate of training and test digits})}{(\text{the recognition rate of training digits})}$$

The convergence of BSFLA algorithm of proposed method is given in Figure 2.

Recognition algorithm was also optimized with the binary genetic algorithm (BGA) to illustrate the performance of BSFLA. The obtained results confirmed that BSFLA has better performance than the binary Genetic Algorithm (BGA) used.

Results have been shown in Table 1. Table 2 shows the performance of matrix on the test samples with BSFLA. Also, figure 2 show the convergence of BSFLA in decreasing the error phase classification.

Table 1. Recognition rate in, different scenarios

Recognition method	Number of features	Recognition rate of test digits	Number of digits error related to test step	Recognition rate of training digits	Number of digits error related to training step	Generalized strength
method of reference (Johari & Razavi, 2001) with zoning feature	64	80.19	3962	86.21	8276	93.02
feature selection(zoning) with BGA algorithm with 20 chromosomes and the number of iterations 100	34	84.32	3136	89.65	6207	94.05
feature selection(zoning) with BSFLA algorithm with 60 memplexes and the number of iterations 200	34	84.86	3028	90.22	5870	94.06
feature selection(zoning) with BSFLA algorithm with 100 memplexes and the number of iterations 150iteration	34	85.22	2955	90.17	5898	94.51

Table 2. The final performance of recognition system on test samples with BSFLA

	0	1	2	3	4	5	6	7	8	9	The number of error	The percentage of error
0	1851	51	2	3	5	68	9	9	0	2	149	7.450
1	116	1713	91	1	11	0	39	0	0	29	287	14.35
2	0	119	1471	139	73	0	50	56	6	86	529	26.45
3	6	17	144	1698	50	0	66	5	2	12	302	15.10
4	2	26	40	165	1600	61	48	9	11	38	400	20.00
5	67	52	4	1	10	1835	11	1	17	2	165	8.25
6	4	3	108	22	26	4	1439	81	4	309	561	28.05
7	1	14	37	11	34	5	74	1820	0	4	180	9.00
8	6	67	2	0	8	67	31	0	1810	9	190	9.50
9	5	54	0	0	5	4	97	0	27	1808	192	9.60

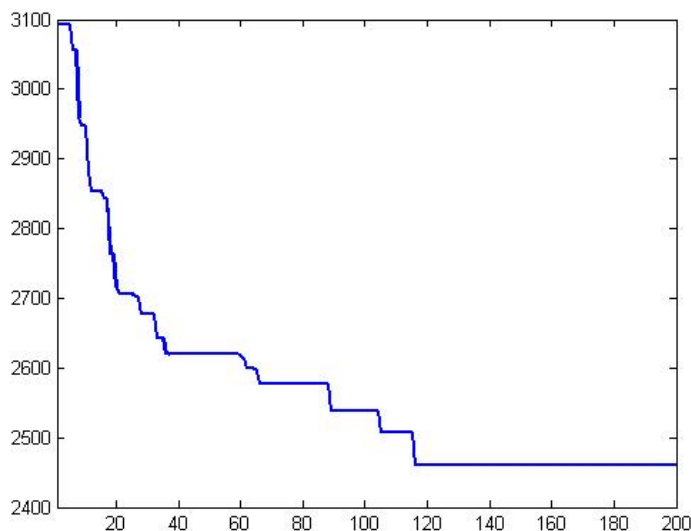


Figure 2. Convergence of BSFLA at feature selection in order to Reduction of Classification Error

5 CONCLUSION

In this paper, the effectiveness of relatively new algorithm BSFLA for recognition system of Persian handwritten digits was studied. The results confirmed the effectiveness of this algorithm. The used Fuzzy system in this study is a simple system with a few numbers of rules (ten Fuzzy rules for ten digits) and has the lower recognition rate compared to the classifiers such as neural network. One can use data clustering for each class and write more rules to increase the recognition rate of Fuzzy classifier (of course, the algorithm will be slow down by increasing the

Fuzzy rules, for this reason, we have avoided this). It is natural that as the recognition rate of Fuzzy classifier improves, higher recognition rate is achieved by selecting the optimal features. Therefore, the purpose of this study was to demonstrate the effectiveness of BSFLA algorithm on the digits' recognition system which is obtained. Also, the results of recognition are without any preprocessing and post-processing operations, hence the obtained recognition rate is a good rate. It is necessary to say that using BSFLA algorithm is time consuming, but this algorithm runs only in training step of Fuzzy classifier; in testing step, the results of training step is just used and do not need to run it again. By spending more cost in training step, the better results will be obtained in the testing step and lower computational costs needed.

The combination of classifiers is an interesting discussion on pattern recognition issues. Suggestion for further researches is to use the combination of classifiers for the purpose of increasing the recognition rate. To combine the classifiers, more features will be used. These features shall be relatively independent from each other and the recognition rate should be increased by their appropriate combination. As mentioned above, another suggestion is to improve the applied Fuzzy classifier.

REFERENCES

- Bi Jami, E. E., Akbar;Askari, Javad. (2011). Problem Solving of Economic Distribution of Nonconvex Charge Using New Hybrid Frog Leaping Algorithm. *Journal of Intelligent Systems in Electrical Engineering*, 2(3), 35-52.
- Darvish, A., Kabir, E., & Khosravi, H. (2005). Application of Shape Matching in Persian Handwritten Digits Recognition. *Modares Technology and Eng*, 22, 37-47.
- Eusuff, M. M., & Lansey, K. E. (2003). Optimization of water distribution network design using the shuffled frog leaping algorithm. *Journal of Water Resources planning and management*, 129(3), 210-225.
- Ghanbari, N. R., S.Mohammad; Nabavi Karizi, S. Hassan. (2010). *Using GSA algorithm in feature selection and features weighting for improving the recognition rate of Persian handwritten digits by Fuzzy classifier*. Paper presented at the Sixteenth Annual International Conference of Computer Society of Iran, Iran.
- Gomez-Gonzalez, M., Ruiz-Rodriguez, F., & Jurado, F. (2014). Probabilistic optimal allocation of biomass fueled gas engine in unbalanced radial systems with metaheuristic techniques. *Electric Power Systems Research*, 108, 35-42.
- Johari, V., & Razavi, M. (2001). *Fuzzy Recognition of Persian Handwritten Digits*. Paper presented at the Proc. 1st Iranian Conf. on Machine Vision and Image Processing, Birjand.
- Khosravi, H. (2006). *Recognition of Persian Handwritten Letters and Digits in MA National Exam Registration Forms*.
- Khosravi, H., & Kabir, E. (2007). Introducing a very large dataset of handwritten Farsi digits and a study on their varieties. *Pattern recognition letters*, 28(10), 1133-1141.
- Miri, I. R., S. Mohammad; Sadri, Javad. (2013). Improving the Recognition Rate of Persian handwritten digits Fuzzy Classifier by K-means Clustering and Particle Swarm optimization. *Journal of Soft Computing and Information Technology*, 2(1).
- Nafisi, H. Z. K., Ehsan Allah. (199). *Recognition of Persian Handwritten Digits*. Paper presented at the Second Conference on Electrical Engineering, Iran.
- Nahvi, M. R., Mahdi; Ebrahim pour, Reza; Kabir, Ehsan Allah. (2008). *Combination of Two-Class Classifiers for Recognition of Persian Handwritten Digits*. Paper presented at the Sixteenth Conference on Electrical Engineering.

- Ranjesh, S. (2013). Improved Frog Leaping Algorithm Using Cellular Learning Automata. *International Journal of Engineering-Transactions A: Basics*, 27(1), 1.
- Rashedi, E., Nezamabadi-Pour, H., & Saryazdi, S. (2009). GSA: a gravitational search algorithm. *Information sciences*, 179(13), 2232-2248.
- Rashedi, E., Nezamabadi-Pour, H., & Saryazdi, S. (2010). BGSA: binary gravitational search algorithm. *Natural Computing*, 9(3), 727-745.
- Razavi, S. M. S. Y., Hadi; Kabir, Ehsan Allah. (2001). *Selecting Feature for Persian Handwritten Digits Recognition by genetic algorithms*. Paper presented at the Seventh Annual Conference of Computer Society of Iran, Iran.
- Seyedi, I. M. F., Malihe; Barati, Mohammad; Nezam Abadi pour, Hossein. (2011). A New Approach for positioning Multi-mode Problems Using Improved Frog Leaping Algorithm. *Journal of Intelligent Systems in Electrical Engineering*, 2(1), 45-56.
- Shahriari Kahshki, M. Z., Maryam. (2011). Hybrid Learning Algorithm for Designing Fuzzy Wavelet Networks for Functions Approximation, Real-time Identification and Control of Nonlinear Systems. *K. N. Toosi University of Technology*, 5(1), 14-26.