The Meta-Heuristic Binary Shuffled Frog Leaping and Genetic Algorithms in Selecting Efficient Significant Features

Saeed Ayat[1], Mohammad Reza Mohammadi Khoroushani[2]

[1] *Associate Professor, Department of Computer Engineering and Information Technology, Payame Noor University, Iran*
[2] *M.Sc. student, Department of Computer Engineering and Information Technology, Payame Noor University, Esfahan, Iran*
[1] dr.ayat@pnu.ac.ir
[2] mr_mohammadi@of.iut.ac.ir (The corresponding author)

Abstract

Selecting the most suitable features among a collection of features to achieve accuracy, sensitivity and efficiency is considered as a big challenge in pattern recognition systems. In this study, the two binary genetic and the binary shuffled frog leaping evolutionary algorithms are evaluated with respect to efficient feature selection in a medical detecting system. The results point to the effectiveness of selection of the most suitable features through memetic Meta heuristic binary frog leaping in increasing the accuracy, sensitivity in detection and time saving in the Classification process against the genetic algorithm.

1. Introduction

Five types of lesions among 533 images are studied by Ballerini et al [1] who then introduced an image recognition method based on features significance by the features synthesis through genetic algorithm with 67 to 82% accuracy.

Efficient feature selection is among the issues regarding optimization; therefore, evolutionary algorithms in this area are highly applicable. During the last decades, the genetic algorithm as an evolutionary algorithm is investigated in many studies. In the recent years many complementary algorithms like Imperialist Competitive Algorithm (ICA), Particle Swarm Optimization (PSO), Ant Colony, shuffled frog leaping etc. have been introduced. Though the feature selection is of the NP-Hard issues and that every algorithm has its own limitations, and advantages and disadvantages, developing a coefficient algorithm or method effective in process time and performance is of essence. Accordingly, here the meta-heuristic binary shuffled frog leaping and genetic algorithms in selecting

efficient significant features is introduced on a medical detecting system to distinguish the skin lesion pigments and evaluating the system's efficiency and performance. The findings here are compared with the genetic algorithm. The results indicate that this binary hybrid algorithm with respect to optimization of similar selection systems with a high convergence speed and accuracy would contribute to feature selection.

2. Evolution algorithms

The genetic algorithm is replicated from the gradual genetic development in human body which is the most practical evolution algorithm regarding efficient feature selection optimization field. Some of the major studies conducted in this field consist of: [2,3,4,5]. The drawbacks of evolutionary algorithms consist of being trapped in local optimization and low convergence speed. The shuffled frog leaping approach is developed based on the group behavior of frogs during food hunt [6]. The advantages of this algorithm in optimization application are evaluated in [7] and [8]. The basic conceptions of most of the evolution algorithm are similar in a sense that: the concept of individual is chromosome and in frog sense it is (meme) in the memetic shuffled frog leaping algorithm.

In genetic algorithm, every chromosome consists of a collection of genes called memes. Gene, in genetic algorithm and memes in shuffled frog leaping are the so called features. In this study, illustrating a feature with "zero" value indicates its absence and illustrating a feature with "one" value indicates its presence in the final Collection. The Fitness function in genetic algorithm is like the objective function in shuffled frog leaping. In this study the Fitness function criterion is selected based on the linear combination of accuracy in recognition and the number of the most desirable features of the sub-Collection.

$$fitness(Fi) = \alpha \times CCR(Fi) + (1 - \alpha)length(Fi) \tag{1}$$

In this Eqn. Fi is a sub-collection of the selected features, correct classification rate (CCR) is the accuracy in recognition obtained through these features, and length is the number of the members in sub-memeplex and α is a number in the range of 0 and1. The value of α in this study is 0.8.

3. The process of detecting the skin pigment lesions

This issue, the main concern of this study, consists of five stages: 1) pre-process, 2) segmentation, 3) feature extraction, 4) feature selection 5) Classification. The first step is to promote the image quality. The detecting process is based on ABCDTP rule [9,10], developed for images based on the digital cameras advances where asymmetry, lesion edge irregularities, pigment variety, diameter, tissue and lesion descriptions are assessed. In this study 430 features are extracted and used from the surface of a lesion. At features selection stage the binary shuffled frog leaping algorithm and the genetic algorithm are evaluated. At the detecting stage the *k* nearest neighbor algorithm is used. In this study the skin lesions are classified in six groups and the shuffled frog leaping algorithm is considered better than the genetic algorithm, since the former is of high speed in convergence and selecting the lowest number of the most suitable features with no reduction in accuracy and sensitivity in diagnosis.

3.1. Selection of the most suitable content based features through binary genetic algorithm

Here, the binary genetic algorithm is used with the descriptions presented in Table 1.

3.2. Selection of the most suitable content based features through binary shuffled frog leaping algorithm

This algorithm corrects the frog memes as it advances. The frogs exchange data among themselves, thus, correcting their memes. As the features are improved the leaping of every frog is adjusted, hence a position change for every frog. The main distinguishing element of the memetic algorithms against other evolutionary algorithms is its ability in searching their locality.

Local searching provides the exchange of memes among the frogs in addition to providing the potential combined strategy in the exchange of population. After the evolutionary memetic stage is repeated for the designated turns, through combined process, the data among the memes are exchanged. This process guarantees that the desired evolution oriented towards any objective would not become bias. The local search and this process will continue till a designated convergence criterion is evaluated.

In the available evolutionary genetic and the binary shuffled frog leaping algorithms all evaluations are made through the hold-one-out (HOO) procedure; where every time one of the images is held back as the test Sample and the evolutionary algorithm is trained on the whole image collection. The parameters of the evolutionary genetic algorithm and the binary shuffled frog leaping algorithms are presented in Table 1.

Table 1. The evolutionary genetic algorithm and the binary shuffled frog leaping algorithms

Genetic	Shuffled frog leaping
Population	Population
Crossover displacement rate	Number of the Memeplexs
Genetic mutation rate	Number of the members in each Memeplexs
Integration method=single point	Number Of Sub-Memeplex (q)=8
Number of generations	Repetition rounds of the evolution
	Overall repetition rounds

3.3. Skin lesion detection

The last stage in this process is the assessment of similar input images based on the selected most suitable feature range from the previous steps. The most common manner in categorizing the image retrieval is the KNN. Here, the obtained ranges and the available range in the database are compared through Eqn. 2 and then the K of the similar input image, are recollected; here the value of K is 7.

$$d(I1, I2) = (\sum_{i=1}^{p} (f_{1,i} - f_{1,i})^2)^{\frac{1}{2}} \tag{2}$$

The distance criterion among both the images' feature ranges, where d is the distance and I_1 and I_2 are the subject images described through feature range of P dimension. The appropriation criterion is based on the most members in a group of the retrieved lesions. The above Eqn. is applied in computing the correct recognition pattern in the most suitable feature range selection and system evaluation stages through the test specimen.

4. Assessment

In this study 545 images are collected from verified internet sources collections of [11], [12] and 35 images collected from Esfahan province. Each one of the images is separately evaluated by four dermatology experts where the ones corresponding to the expert's detection and the provided diagnosis are selected and the doubtful ones are eliminated from the database. Out of this total 70% are applied in the most desirable feature selection and 30% are subject to test.

A system consisting of a Corei7 740Qm processor and a 6 Gigabit ram in an 8 Megabit cash memory is involved at all stages. In medical detecting systems, usually the sensitivity and accuracy criterion are used [13], likewise here.

Sensitivity: the rate of accuracy in detecting a type of lesion (Whether the lesion is a correct one of the correct group)

$$\text{Sensitivit } y = \frac{Tpi}{Tpi + Fpi} \tag{3}$$

Accuracy: the percentage of accurate detection of all skin lesion types (Accurate classification of the lesions of each group)

$$\text{Accuracy} = \frac{\sum_{i=1}^{5} Tpi}{\sum_{i=1}^{5} Tpi + Fpi} \tag{4}$$

The assessment of the reduction effect of the feature dimensions through binary evolutionary genetic algorithm and binary shuffled frog leaping on the detection of sensitivity and accuracy is expressed in Table 2.

Table 2. The assessment of the reduction effect of the feature dimensions through binary evolutionary genetic algorithm and binary shuffled frog leaping

	With Out Feature Selection		binary evolutionary genetic algorithm		binary shuffled frog leaping	
	Accuracy	Sensitivity	Accuracy	Sensitivity	Accuracy	Sensitivity
KNN	83.6%	85%	81%	80%	87%	86.6%

As observed, by a decrease in the number of the features not only the sensitivity and accuracy decrease, on the contrary they increase by 3.4 and 1.5%, respectively. This occurs when due to a decrease in the feature dimensions through genetic algorithm, the detection sensitivity and accuracy were subject to a decrease by 2.6 and 5%, respectively.

The assessment of the effect of the number of selected efficient features through binary genetic algorithm and the binary shuffled frog leaping algorithm during the whole process in relation to the non-reduced dimension of the feature is tabulated in Table 3.

Table 3. The skin lesion segmentation system results' assessment

Feature selection method	Tp	Ts	Tf	Tcf	Tt
Binary shuffled frog leaping algorithm			2.85 Sec	1.95 Sec	9.4 Sec
Genetic algorithm	2.4 Sec	2.2 Sec	2.9 Sec	2.2 Sec	9.7 Sec
No dimension reduction			4 Sec	3 Sec	11.6 Sec
Reduction in process time with respect to no dimension reduced state and after dimension reduced state through the binary shuffled frog leaping algorithm		28.75 %	35%	18%	
Reduction in process time with respect to no dimension reduced state and after dimension reduced state through the genetic algorithm		27.5 %	26.6%	16.3%	

In this table: Tp is the pre-process average, Tf is the feature extraction time average, Ts is the skin lesion spot Segmentation time average, Tc is the classification time average based on features and Tt is the total time average in detecting the skin lesion type.
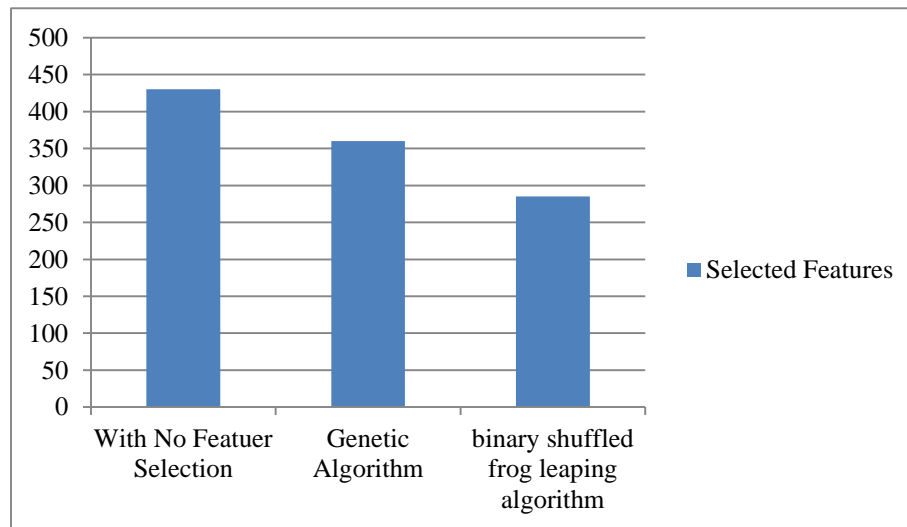

Fig 1. The diagram of all the extracted features from the image before and after the most desirable feature selection process through the binary genetic and binary shuffled frog leaping algorithms

Table 4. Number of the features before to selection process and after the most desirable feature selection process

Number of features	Method
430	No dimension reduction
360	After dimension reduction through binary genetic algorithm
285	After dimension reduction through binary shuffled frog leaping

5. Conclusion

With respect to the importance of the feature selection in issues regarding recognition patterns and the necessity of applying an efficient algorithm where process time and accurate performance are of essence the two evolutionary genetic and shuffled frog leaping algorithms in feature selection application are introduced and assessed in this study. For this purpose the binary structure is suggested to illustrate the memes with the objective of developing a sub-collection with fewer dimensions than

that of the original collection where detecting sensitivity and Accuracy would be scalable with that of the initial status (without feature selection). Based on the obtained results the shuffled frog leaping algorithms compared to its competitor the evolutionary genetic algorithm (Ballerini et al [1]) is of a high convergence speed, with more consistency and more appropriate feature selection ability. The most unique characteristic of the shuffled frog leaping algorithms is its ability in searching the locality which leads to better performance. The findings indicate that the number of features in relation to the initial 430 is reduced to 285 with no deduction in the sensitivity and accuracy, while both the latter have had an increase by 3.4 and 1.5%, respectively. This method accompanied with the obtained sensitivity and accuracy in this study can be adopted to evaluate types of other databases in medicine and non-medicine fields with different dimensions.

Acknowledgement

References

[1] L. Ballerini, X. Li, R. B. Fisher, and J. Rees, "A query-by-example content-based image retrieval system of non-melanoma skin lesions" presented at the Proceedings of the First MICCAI international conference on Medical Content-Based Retrieval for Clinical Decision Support, London, UK, 2010.

[2] M. Saberi, D. Safaai, "Feature Selection Method Using Genetic Algorithm For The Classification Of Small and High Dimension Data" in International Journal on IEEE Transaction On Pattern Analysis And Machine Intelligence, VOL. 23, NO. 11, 2005.

[3] H. Chouaib, O.R. Terrades, S. Tabbone, F. Cloppet, N. Vincent, "Feature selection combining genetic algorithm and Adaboost classifiers" in 19th International Conference on Pattern Recognition(IEEE), ICPR 2008, Tampa, FL, 2008.

[4] E. P. Ephzibah, "Cost Effective Approach on Feature selection Using Genetic Algorithms And Fuzzy Logic for Diabetes Diagnosi" , in International Journal on Soft Computing(IJSC), Volume 2,No1 , February 2011.

[5] S. Zanganeh, R. Javanmard, M. Ebadzadeh, "A Hybrid Approach for Features Dimension Reduction of Datasets using Hybrid Algorithm Artificial Neural Network and Genetic Algorithm-in Medical Diagnosis" in 4rd Iran Data Mining Conference (IDMC), 2010.

[6] M. Eusuff, K. Lansey, "Optimization of Water Distribution Network Design Using the Shuffled Frog Leaping Algorithm" in Journal of Water Resource Plan and Management, Vol. 3, pp. 10-25, 2003.

[7] M. Eusuff, "Shuffled frog-leaping algorithm: a memetic meta-heuristic for discrete optimization" in journal on Engineering Optimization, Volume 38(2), Pages.129-154, 2006.

[8] S. Chittineni, A. N. S. Pradeep, G. Dinesh, S. C. Satapathy, P. V. G. D. Prasad Reddy, "A parallel hybridization of clonal selection with shuffled frog leaping algorithm for solving global optimization problems (P-AISFLA)" Proceedings of the Second international conference on Swarm, Evolutionary, and Memetic Computing, Volume Part II, Visakhapatnam, Andhra Pradesh, India, Springer-Verlag: 211-222, 2011.

[9] M. R. Mohammadi Khoroushani, S. Mahzounieh, "Intelligent Skin Cancer Detection Software System based on the principles of telemedicine", in Journal of Hospital, Volume 5, Pages 55-60, April 2014.

[10] M. R. Mohammadi Khoroushani, S. Ayat, "Early Melanoma Skin Cancer Detection by Co-Occurrence Matrix and Gabor Filter", 11th National Conference on information and communication technology, Kish, Iran, 2013.

[11] P. T. a. M. Newton. (2005, 2013/14/01). Global Skin Atlas Available: http://www.globalskinatlas.com.

[12] G. S.A. (2013, 2013/14/01). Personalized learning and teaching resources for dermatologists today and tomorrow. Available at: https://www.dermquest.com.

[13] N. Z. Wen Zhu, Ning Wang, "Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS® Implementations ", in NESUG proceedings: Health Care and Life Sciences, 2010.