

Adaptive Resonance Theory Training Parameters: Pretty Good Sets

¹Taisir Mohammad Eldos and ²Abdulaziz Suleiman Almazayad

¹Department of Computer Engineering, Jordan University of Science and Technology,
Ar Ramtha, Irbid, Jordan

² Department of Computer Engineering, College of Computer Engineering and
Sciences, King Saud University, Riyadh 11451, Saudi Arabia

Abstract: Problem statement: ART1 artificial neural networks offer good tools for test clustering, where no expert is needed if the system is well trained. However, having no output reference for the input patterns makes it hard to judge the quality of the training. Moreover, the performance depends to a great extent on a set of training parameters. Designers follow some recommendations or depend on their expertise in finding good sets with no performance guarantees. Many methods were proposed; from greedy methods offering quick and acceptable solutions to evolutionary algorithms offering suboptimal sets of parameters. While the evolutionary algorithms are a good choice for quality, the computational cost is large even for an offline process; after all, computing resources are not for free. **Approach:** We introduced a method for selecting a set of parameters that yields a comparable performance and robust operation, with relatively low cost compared to the evolutionary methods. This method located a suitable set through repetitive portioning of the range, by considering the best subset for the next iteration. **Results:** Tests have shown that performance comparable with the computationally intensive evolutionary methods could be achieved in much less time. **Conclusion:** The repetitive portioning method for finding a good set of training parameters is very cost effective and yields good performance.

Key words: Adaptive resonance theory, pretty good set, artificial neural network, optimization

INTRODUCTION

Artificial Neural Network (ANN) is a complex interconnection of processors, called neurons, that uses a mathematical or statistical model for information processing and can be used as non-linear statistical data modeling or decision making tool. ANN structures can be used to find patterns in huge data sets or to model complex relationships between inputs and outputs. The power of these structures to infer a function from observations is particularly useful in applications where the complexity of the data or task makes the design of such a function by hand impractical. Application areas include function approximation, classification like pattern recognition (face identification, tumor detection) and sequence recognition (gesture, speech and handwritten text recognition), data processing like medical diagnosis, financial applications, data mining and others.

The artificial neural network concept has in fact evolved from a way to model the brain towards understanding how it works, to different structures and

training methods targeting different goals. Adaptive Resonance Theory (ART) is a class of ANN that is capable of self learning, ART1 type accepts binary inputs used primarily in pattern classification applications like text clustering (Carpenter and Grosberg, 1987), where documents are presented as binary strings characterizing the occurrences of features, including: taxonomy generation, topic extraction and search engines hits grouping, which are quite useful in many modern applications like hierarchical web search. Although the supervised Text Categorization (TC) is the best for such applications in terms of quality, it lacks adaptability, requires expert's intervention and occasional retraining (Fausett, 1993).

The ART1 based solutions are free from those drawbacks, but the design involves a set of parameters that have to be carefully selected to achieve a satisfactory performance. Typically, there are four parameters with recommended values to use, but for better results one may apply some heuristic based greedy algorithm to find reasonable values, or apply a time consuming search algorithm like genetic algorithm

Corresponding Author: Taisir Mohammad Eldos, Department of Computer Engineering,
College of Computer Engineering and Sciences, Al Kharj University,
Al Kharj 11942-0151 Saudi Arabia Tel: +966-59-946-3949 Fax: +966-1-548-6072

or simulated annealing, with no performance guarantee in all cases. The proposed method for parameters selection explores the space of training parameters to locate a set of high return in terms of the standard measures, with affordable amount of time.

The search for a practically good set of parameters requires working with a sample of the training data. Towards this, we modify one of those standard measures to guide the search process in a way that results in a robust design; one whose performance is independent of the training order and to enhance the selection of the clusters cardinality parameter in particular.

Figure 1 depicts the basic structure of an ART1 ANN (Al-Natsheh and Eldos, 2007); it consists of three groups of neurons; input neurons in layer S, interface neurons in layer X and supplementary neurons in layer Y. In addition, ART1 networks have reset neurons R and possibly other control neurons. Input patterns are presented to the input layer S, which sends its output signals to the X and R neurons.

An interface layer neuron X_i ($i = 1, \dots, n$), where n is the number of input units, is connected to each neuron of Y layer Y_j ($j = 1, \dots, m$), where m is the maximum allowed number of clusters, by two weighted pathways. Signals broadcast from neurons of layer X to neurons of layer Y over connections pathways with bottom-up weight matrix b_{ij} and from neurons of layer X to neurons of layer Y over connections pathways with top-down weight matrix t_{ji} ($i = 1, \dots, n$ and $j = 1, \dots, m$).

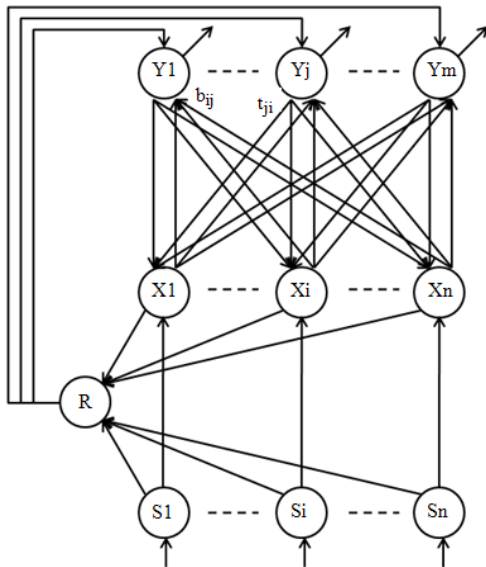


Fig. 1: ART1 structure

The neurons of layer Y are competitive; each neuron Y_j ($j = 1, \dots, m$) weight is initialized to 0 and then calculated by multiplying the interface layer and the bottom up weight matrix. However, some of the clustering neurons will be inhibited, to prevent the neuron from participating in any further computations during the presentation of the current input vector.

In addition to the maximum allowed number of clusters m and the vigilance parameter ρ , which are required for the learning process, two constants L and A_n are used to initialize the bottom-up weight matrix. So, our concern is to find the set of training parameters (L, a, m, ρ) that maximizes the performance of the ART1 network for a certain instance. Although the selection of the constants L and a can affect the outcome, the selection of m and ρ is more critical, as they have a significant impact on the performance, the trends as stated in (Fausett, 1993) are:

- Large ρ and Small m , provides stable cluster formation with few epochs of training and possibly numerous patterns with clustering
- Large ρ and Large m , provides stable cluster formation with few epochs but increase the input data order-insensitivity
- Small ρ and Small m , requires more epochs to stabilize with increased dependence on the training order

Al-Natsheh and Eldos (2007), the author suggested investigating the training order impact on the weights convergence. Two variations are used; few random scans and few scans with maximum diversification between adjacent input vectors. In the later, a random vector is considered then the farthest (using Hamming distance) is next and so on. The performance of those variations will be compared with a single pass using the same parameters selection method and against the conventional training, i.e., recommended set of parameters and single scan.

This study will review the work related to the ART1 performance enhancement methods and then present our method for locating a suitable set of parameters for the training process. We will then introduce the performance measures to be used in the evaluation and show an outline of the strategy of finding a pretty good set of parameters. And the end, we show the results compared with others and conclude with some hints on possible future work.

Related work: Al-Natsheh and Eldos (2007), reviewed the research conducted in this area and pointed that much of the focus was on the major issues like quality, space and time requirement and only little went to

consider application-independent architecture, learning algorithms and performance.

Fausett (1993), the author tests a simple ART1 network implementation and evaluates its text clustering quality on the Reuter data set by standard measures, using the K-means clustering quality as lower bound and supervised TC as upper bound to publish his results relatively and finding the best settings for ρ and m was performed by applying an incremental search.

He *et al.* (2003) and Russell and Norvig (1995), Adaptive Resonance Theory under Constraints ART-C, used dynamic variable value of vigilance parameter is applied, according to an extra constraint reset mechanism to the ART architecture. This concept was applied in ART 2A (He *et al.*, 2003) to produce ART 2A-C. It was examined by clustering gene expression data application. ART-C shows better performance compared to the k-means, Self-Organizing Map (SOM) and conventional ART. Zacharie (2007) proposed a real-time ART1 model for pattern recognition that preserves its previously learned knowledge while learning new input patterns using a parameter called the attentional vigilance parameter

Since ART was published, many approaches have been presented: improved ART1, Adaptive Hamming Net (AHN) in (Hung and Lin, 1995) and Fuzzy ART, which are optimized in terms of space and time. In AHN, ART clustering scheme as an optimization problem was solved by finding the best matching unit in time by 4 defined equations. The Symmetric Fuzzy ART (S-Fuzzy ART) network is presented as a possible improvement over Fuzzy ART.

Fuzzy ART is the best representative of the ART1 based network group. However, Fuzzy ART is sensitive to noise and outliers and input presentation (Baraldi and Alpaydin, 2002).

Recently, Cao and Wu (2004) have developed a very effective high-dimensional network called Projective ART (PART), based on the assumption that the model equations of PART, a large scale and singularly perturbed system of differential equations coupled with a reset mechanism, have quite regular computational performance.

Genetic algorithms have been used for a long time in optimizing certain types of ANN like the back-propagation (Rovithakis *et al.*, 2000; Han and May, 1996; Choi and Bluff, 1995; Wallrafen *et al.*, 1996; Yao, 1999), while (Lippmann, 1987; Massey, 2002) suggested methods to select the training parameters through trials to stop at a suitable set and parameter finding ART1 ANN in (Al-Natsheh and Eldos, 2007). In addition to the challenge that lies in finding a good

set of parameters for the genetic algorithm itself, the nature of this type of ANN does not lend itself straightforwardly to the optimization process because it lacks a reference for quality; hence some goodness measure had to be established. Simulated annealing is possibly a good alternative to genetic algorithms in terms of computational requirement although the quality and convergence is slightly less in most cases.

Yang and Yang (2008) the author addresses some drawbacks of certain networks, proposing a modified ART1 neural learning algorithm, in which the vigilance parameter is estimated by the data so that it is more efficient and reliable than Papaioannou and Wilson (2010) method for selecting a vigilance value. Isawa *et al.* (2009) the authors propose using variable vigilance parameters, where vigilance parameters are arranged for every category and varied according to the size of respective categories with learning and claimed more flexibility in classifying input data compared to the conventional Fuzzy ART. Chen *et al.* (2005), the author carried out a simulation case to analyze the ART1 architecture and the membrane equation of layer 2 to describe the oscillation possibility of the activities of the neurons and studied the influence of parameters setting on the behavior of L2 layer.

Parameters selection: The two weight matrices; bottom-up weights matrix (b_{ij}) and top-down weights matrix (t_{ji}) are initialized as follows:

$$b_{ij} = \frac{a.L}{L + n - 1}; i = 1:n, j = 1:m \quad (1)$$

$$t_{ji} = 1; i = 1:n, j = 1:m \quad (2)$$

Our guide to select a good set of parameters must be the performance metrics used at the end to judge the quality of clustering, which are the well known (FM) and (JAC) measures (Massey, 2003), defined as:

$$FM = \frac{A}{\sqrt{(A+B)(A+C)}} \quad (3)$$

$$JAC = \frac{A}{A+B+C} \quad (4)$$

Where:

A = The pair-wise number of true positives; number of document pairs expected to be grouped together and are indeed clustered together by the clustering algorithm

B = The pair-wise number of false positives; number of document pairs not expected to be grouped

together but are clustered together by the clustering algorithm

C = The pair-wise number of false negatives; number of document pairs expected to be grouped together but are not clustered together by the clustering algorithm

However, those measures do not reflect the parameter selection robustness; the impact of training set ordering on the performance. Practically, we are interested in a set of parameters that would give better performance regardless of the data set training order. In an early work (Al-Natsheh and Eldos, 2007) used a simple-to-compute fitness function that is proportional to the insensitivity to order and clustering capability as measured in terms of the number of input patterns that cannot be associated with a cluster.

Implementation:The fitness function F is equally affected by the insensitivity in the first term and the clustering capability in the second term:

$$F = \frac{1}{2} \left(\frac{x}{m} + \frac{m}{y+m} \right) \quad (5)$$

Where:

- m = The number of clusters
- x = The number of clusters in match in the two scans
- y = The number of input patterns that could not be clustered

A counter that increments if any of the following conditions is satisfied: Norm of input layer is 0 and winning cluster unit value is -1, which means either the input pattern has no features, or all cluster nodes are inhibited due to the reset activation caused by low vigilance.

In this study, we use the modifier ϵ , which is a factor that reflects on the standard clustering quality measure for the process of a pretty good set of parameters. It is an efficiency or utilization metric that is defined as:

$$\epsilon = \epsilon_c \cdot \epsilon_p \quad (6)$$

Where:

$$\epsilon_c = \frac{\text{Clusters} - \text{empty clusters}}{\text{Mum of clusters}} \quad (7)$$

$$\epsilon_p = \frac{\text{Patterns in all clusters}}{\text{Input patterns}} \quad (8)$$

Where:

ϵ_c = the clusters utilization; the ratio of non-empty clusters to the total number of clusters

ϵ_p = the pattern association; the ratio of mapped input patterns to the total patterns

The clusters utilization factor favors the parameters that lead to fewer unused clusters and the pattern association factor favors the parameters that lead to fewer orphans, i.e., input patterns without hosting clusters. Clearly, with a large number of inputs, few unmapped patterns will make no significant contribution to the robustness and this is one of the reasons why we carry out this process on a relatively small set of input patterns.

Regarding the ordering insensitivity, rather than having just two scans, forward and backward, we perform many scans at random ordering in ART1-Pr, or according to an ordering that yields the most diversified sequence in ART1-Pd. This is a deviation from the ART1-GEP which used only a forward and reverse ordering with no justification for a certain training order.

The competitive partitioning method used is outlined below; the ranges are split into 16 partitions by halving each dimension, then the most fit 4 partitions split further into 16 partitions each, making 48 sets to pick the best 4, the process continues until a stopping condition is met; less than 10% improvement in this implementation, although a faster and more deterministic approach can use the number of iterations:

- Form m clusters, using k-means method, where m the max cluster cardinality
- Pick q vectors at random from each cluster to form a small training data set (m×q) for parameters selection (if size of cluster is less than q borrow from another one)
- Split the range of the 4 parameters into halves and consider the middle of each as a case; this will generate $2^4 = 16$ parameter vectors
- Run the training algorithm using (m×q) data set using each of these combinations, with few random orderings per training session
- Compute the average A, B, C and compute the measure (FM or JAC) and multiply by the cluster utilization and input mapping percentage
- Consider the best 4 parameter sets in terms of the metric used in 5
- Redo the steps for each of the best 4 1/16 of the parameter ranges; this yields 48 out of which only the best 4 are to be pursued
- Continue until the improvement of the best metrics using the standard measures is below a certain

threshold or no improvement for the last few steps (partitions)

- Consider the best as the target set for the full training

The algorithm automatically maintains valid sets of parameters based on the recommended range for each (Lippmann, 1987; Massey, 2002), as it repeatedly splits the intervals in each dimension to generate the 16 subspaces for the next round, each of which further splits into another 16 subspaces, an exception is that when splitting along the m dimension, we only consider integer numbers by rounding up.

MATERIALS AND METHODS

Based on the study in (Al-Natsheh and Eldos, 2007), an enhanced collection of 2000 web text pages is prepared for training and testing, using top ranking outcomes from popular search engines in 15 different categories (news, finance, music, sports, games, travel, health, blogs and others). We applied a feature extraction procedure similar to (Lippmann, 1987; Massey, 2002; Massey, 2003; Cohen *et al.*, 2005; Davidov *et al.*, 2004; University of Waikato, 2005), to produce training and testing data sets. Each web page is represented by a string with a set of 2000 useful tokens with threshold frequencies of 3.

We compare the ART1-Px performance with that of the conventional ART1 with a recommended set based on the work of (Lippmann, 1987; Massey, 2002; Massey, 2003). The work will not be compared to that of ART1-GEP in (Al-Natsheh and Eldos, 2007) for few reasons; the work team has split and we have no access to the code, the data set has been augmented based on recommendations from some reviewers and the target of this work is to select a pretty good set of parameters at a much lower cost in comparison with the ART1-GEP.

Among the 2000 pages suite, we picked 500 pages at random to be used for training and only 120 of those pages were selected through k-means process that places the 500 input patterns in 30 clusters to select 4 patterns at random from each group.

Since the pages were picked from 15 different categories, we thought setting the initial maximum number of clusters to 30 would be fair, so the actual maximum for the training would be below that.

Table 1: ART1 training parameters range

	L	a	ρ	m
Lower limit	1	0	0	1
Upper limit	100	1	1	30

The range for each of these parameters is shown in Table 1, where the minimal vigilance computed as $\rho_{\min} = 1/N$, where N is the number of features (tokens) used to represent a document (Massey, 2003), which is 2000 in our tests. Because the repetitive interval halving considers the middle of the halves, we would start with 0 for ρ , so when it gets portioned many times it would be allowed to assume values as low as $1/N$. For each parameter set, we carry out three tests to evaluate its quality:

- ART1-Ps, the partition in hand and a single training order
- ART1-Pr, the partition in hand and 5 random orders per trial
- ART1-Pd, the partition in hand and 5 diversified orders per trial

RESULTS

We compare the performance of three variations of the proposed method against the conventional one. Table 2 shows the set of training parameters for the conventional ART1 using the method in (Massey, 2003) and the three variations of ART1-P. Clearly, no rule can be deduced from the empirical results; the ART1-P selected a higher ρ and smaller a factors, while the maximum number of clusters came more realistic with the ART1-P in the sense of expert knowledge.

Table 3 shows the performance of clustering using a test set of unseen 1500 input patterns, with three variations of the proposed method against the conventional one based on a recommended set of training parameters. Table 4 shows the comparative performance using the two well known measures to evaluate the clustering quality; JAC and FM.

According to the results detailed in Table 4, the ART1-P improved the clustering quality by up to 16% on the FM and up to 18% on the JAC measure.

Table 2: ART1 Training parameters selection

	L	a	ρ	m
ART1	2	0.5000	0.0250	27
ART1-Ps	11	0.3125	0.1250	23
ART1-Pr	14	0.3750	0.1875	21
ART1-Pd	16	0.4375	0.2188	21

Table 3: Performance comparison (1500 test patterns)

	A	B	C
ART1	1209	5326	6370
ART1-Ps	1347	5194	6208
ART1-Pr	1366	5088	6112
ART1-Pd	1387	5065	6117

Table 4: Performance comparison (1500 test patterns)

	FM	JAC	FM (%)	JAC (%)
ART1	0.172	0.094		
ART1-Ps	0.192	0.106	11.5	12.8
ART1-Pr	0.197	0.109	14.5	16.0
ART1-Pd	0.199	0.110	16.0	17.8

Table 5: Cluster utilization and input mapping

	Clusters	Unused	Unmapped
ART1	27	6	47
ART1-Ps	23	3	34
ART1-Pr	21	2	27
ART1-Pd	20	2	24

Unlike ART1-GEP, which offered a slight decrease in the true positives traded with a large decrease in the false positives, thus increasing the precision and decreasing the recall towards a better overall performance, the proposed method has shown improvement in both precision and recall.

Table 5 shows the effect of selecting the parameters with those methods, which resulted in better cluster utilization as less unused cluster and less orphans.

DISCUSSION

The proposed method for finding a practically good set of training parameters for the ART1 has been shown to improve the performance at a relatively small cost compared to the methods based on evolutionary algorithms as an upgrade of the recommended set method. This method achieves a good balance between the quality of the parameters in terms of the impact on performance and the time required to start the training.

CONCLUSION

The proposed method has improved the ART1 performance by carefully selecting a set of training parameters based on the application nature. It has consistently outperformed the conventional ART1 with recommended sets of parameters, on the two well known measures (FM and JAC).

The goodness of this approach to selecting the training parameters is also shown through the cluster utilization and the mapping capability; i.e., number of input patterns mapped. This also indicates that the order sensitivity is a factor that has to be considered towards better generalization; because order sensitive training cannot be justified. This method of selecting a set of parameters based on the selective subset of the training data set has shown consistent improvement in every aspect.

REFERENCES

- Al-Natsheh, H.T. and T.M. Eldos, 2007. Performance optimization of adaptive resonance neural networks using genetic algorithms. Proceeding of the IEEE Symposium on Foundations of Computational Intelligence, Apr. 1-5, IEEE Xplore Press, Honolulu, Hawaii, USA., pp: 143-148. DOI: 10.1109/FOCI.2007.372160
- Baraldi, A. and E. Alpaydin, 2002. Constructive feed forward ART clustering networks. IEEE Trans. Neural Networks, 13: 645-661. DOI: 10.1109/TNN.2002.1000131
- Cao, Y. and J. Wu, 2004. Dynamics of projective adaptive resonance theory model: The foundation of PART algorithm. IEEE Trans. Neural Networks, 15: 245-260. DOI: 10.1109/TNN.2004.824261
- Carpenter, G.A. and S. Grosberg, 1987. Massively parallel architecture for self-organizing neural pattern recognition machine. Comput. Vision Graph. Image Process., 37: 54-115. DOI: 10.1016/S0734-189X(87)80014-2
- Chen, Z., Z.X. Cai, Z.W. Li and X.H. Wang, 2005. The influence of ART1 parameters setting on the behavior of layer 2. Proceeding of 2005 International Conference on Machine Learning and Cybernetics, Aug. 18-21, IEEE Xplore Press, Guangzhou, China, pp: 4805-4808. DOI: 10.1109/ICMLC.2005.1527788
- Choi, B. and K. Bluff, 1995. Genetic optimization of control parameters of a neural network. Proceeding of the 2nd New Zealand International Conference on Artificial Neural Networks and Expert Systems, Nov. 20-23, IEEE Xplore Press, Dunedin, New Zealand, pp: 174. DOI: 10.1109/ANNES.1995.499466
- Cohen, S., E. Ruppim and G. Dror, 2005. Feature selection based on the shapley value. Proceeding of the International Joint Conference on Artificial Intelligence, (AI'05), IJCAI, Scotland, pp: 655-661. <http://www.ijcai.org/papers/0763.pdf>
- Davidov, D., E. Gabrilovich and S. Markovitch, 2004. Parameterized generation of labeled datasets for text categorization based on a hierarchical directory. Proceedings of The 27th Annual International ACM SIGIR Conference, (ISIGIRC'04), ACM Press, Sheffield, UK., pp: 250-257. DOI: 10.1145/1008992.1009036
- Fausett, L.V., 1993. Fundamentals of Neural Networks: Architecture, Algorithms and Applications. 1st Edn., Prentice Hall, USA., ISBN: 13: 978-0133341867, pp: 461.

- Han, S.S. and G.S. May, 1996. Optimization of neural network structure and learning parameters using genetic algorithms. Proceeding of the 8th IEEE International Conference on Tools with Artificial Intelligence, Nov. 16-19, IEEE Xplore Press, USA., pp: 200-206. DOI: 10.1109/TAL.1996.560452
- He, J., A.H. Tan and C.L. Tan, 2003. Self-organizing neural networks for efficient clustering of gene expression data. Proceeding of International Joint Conference on Neural Networks, July 20-24, IEEE Xplore Press, Portland, USA., pp: 1684-1689. DOI: 10.1109/IJCNN.2003.1223660
- Hung, C.A. and S.F. Lin, 1995. Adaptive hamming net: A fast-learning ART 1 model without searching. Neural Networks, 8: 605-618. DOI: 10.1016/0893-6080(94)00106-V
- Isawa, H., H. Matsushita and Y. Nishio, 2009. Fuzzy ART combining overlapped categories using variable vigilance parameters. Proceeding of the International Workshop on Nonlinear Circuits and Signal Processing, Mar. 1-3, NCSP, Waikiki, Hawaii, pp: 661-664. <http://nlab.ee.tokushima-u.ac.jp/nishio/Pub-Data/CONF/C323.pdf>
- Lippmann, R., 1987. An introduction to computing with neural nets. IEEE ASSP Mag., 4: 4-22. DOI: 10.1109/MASSP.1987.1165576
- Massey, L., 2002. Determination of clustering tendency with art neural networks. Proceeding of Recent Advances in Soft-Computing, (RASC'02), Nottingham, UK., pp: 1-1. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.58.5237>
- Massey, L., 2003. Evaluating quality of text clustering with ART1. Proceeding of the International Joint Conference on Neural Networks, July 20-24, IEEE Xplore Press, Portland, USA., pp: 1402-1407. DOI: 10.1109/IJCNN.2003.1223901
- Papaioannou, G. and J.M. Wilson, 2010. The evolution of cell formation problem methodologies based on recent studies (1997-2008): Review and directions for future research. Eur. J. Operat. Res., 206: 509-521. DOI: 10.1016/j.ejor.2009.10.020
- Rovithakis, G., M. Maniadakis and M. Zervakis, 2000. A genetically optimized artificial neural network structure for feature extraction and classification of vascular tissue fluorescence spectrums. Proceeding of the 5th IEEE International Workshop on Computer Architectures for Machine Perception, Sept. 11-13, IEEE Xplore Press, Padova, Italy, pp: 107-111. DOI: 10.1109/CAMP.2000.875964
- Russell, S.J. and P. Norvig, 1995. Artificial Intelligence: A Modern Approach. 1st Edn., Prentice Hall, USA., ISBN-13: 978-0131038059, pp: 960.
- University of Waikato, 2005. Weka 3: Data Mining Software in Java. University of Waikato. <http://www.cs.waikato.ac.nz/~ml/weka>
- Wallrafen, J., P. Protzel and H. Popp, 1996. Genetically optimized neural network classifiers for bankruptcy prediction. Proceeding of the 29th Annual International Conference on System Sciences, Jan. 3-6, IEEE Xplore Press, Wailea, HI., USA., pp: 419-426. DOI: 10.1109/HICSS.1996.495427
- Yang, M.S. and J.H. Yang, 2008. Machine-part cell formation in group technology using a modified ART1 method. Eur. J. Operat. Res., 188: 140-152. DOI: 10.1016/j.ejor.2007.03.047
- Yao, X., 1999. Evolving artificial neural networks. Proc. IEEE, 87: 1423-1447. DOI: 10.1109/5.784219
- Zacharie, M., 2007. Adaptive Resonance Theory 1 (ART1) neural network based horizontal and vertical classification of 0-9 digits recognition. J. Comput. Sci., 3: 869-873. <http://www.scipub.org/fulltext/jcs/jcs311869-873.pdf>