International Journal of Soft Computing 11 (1): 18-25, 2016
ISSN: 1816-9503

Towards Better Classification Using Improved Particle Swarm Optimization Algorithm and Decision Tree for Dengue Datasets

[1, 2]B. Renuka Devi, [3]K. Nageswara Rao and [4]S. Pallam Setty
[1]JNT University, Kakinada, India
[2]Department of CSE, VLITS, Vadlamudi, Guntur, India
[3]Department of CSE, PSCMR College of Engineering, Vijayawada, India
[4]Department of CS&SE, Andhra University, Visakhapatnam, India

Abstract: This study presents a novel methodology based on Particle Swarm Optimization algorithm to model a new classification system. In the creation of classifier, feature selection frequently used to remove in appropriate and noisy features asto retrieve relevant features. Physically developing a feature set can be very time taking and expensive attempt. PSO is an intelligent search methodology that employs a population of individuals prevailing within a multi-dimensional space. This study employs the correlation between the attributes as the fitness function to Particle Swarm Optimization algorithm. The proposed approach is applied to Clinical Dengue Datasets that retrieve optimal features and the obtained results shows the accuracy and validity of the approach. The proposed methodology is analyzed on dengue data set that is downloaded from http://www.ncbi.nlm.nih.gov/gds. The data set contains 18 attributes of 1275 patients.

Key words: Datasets, noisy, PSO, patients, time

INTRODUCTION

Nowadays, the development of the high-throughput technologies has resulted in exponential progress of very large volumes of databases with respect to both dimensionality and sample size. Efficient and effective management of these databases becomes the increasing challenging. Traditionally manual management of these datasets are practically impossible. Hence, data mining and machine learning techniques were developed rapidly to automatically discover knowledge and recognize patterns from these data. The collected databases is usually associated with a high level of noise due to various reasons such as deficiency in the technologies that includes the information and on the basis of the information itself. Certainly, mining suitable knowledge and patterns from such enormous and noisy data is a challenging job.

Numerous forms of data mining technique are developed in the nature to handle the huge and noisy data. Dimensionality reduction is one of the most widespread techniques to remove noisy or unrelated data and redundant features. Dimensionality reduction techniques can be categorized primarily into feature extraction and feature selection. Feature extraction approaches project features into a novel feature space with minor dimensionality and the freshly created features are frequent amalgamations of original features. The feature selection methodologies intents to choose a small subset of features that diminish redundancy and increases significance to the target such as the class labels in classification. Both feature extraction and feature selection are proficient inrefining learning performances, lowering computational complication, structuring improved generalizable prototypes and declining necessary storage.

Feature selection selects a subset of features from the original feature set deprived of lacking any alteration and preserves there a limplications of the original features. In this sense, feature selection is higher incase of enhanced readability and interpretability (Masaeli *et al*., 2010). A feature selection method comprises of four elementary stages (Liu and Yu, 2005) namely, subset generation, subset evaluation, stopping criterion and result validation. In the initial stage, a candidate feature subset will be selected depending on a specified search approach that is given next stage to calculater endering to certain assessment principles. The outcome that best suits the estimation measure will be selected from all the candidates that have been calculated after the stopping criterion are encountered. In the ultimate stage, the selected subset will be certified using domain knowledge or an authentication set.

Feature selection for classification: In the data mining classification systems, the input data frequently cover a

Corresponding Author: B. Renuka Devi, JNT University, Kakinada, India

lot of relevant and irrelevant features but only some part of them are correlated features for the classification. Existence of huge amount of unrelated features will generate a dimension calamity and decrease the signal to noise ration. In many classification problems, it is difficult to learn good classifiers prior to eliminate these undesirable features due to the huge size of the data. Some of the very good classifiers are not able to avoid the effects of huge amount of unrelated or redundant features on the classification consequences. Reducing the amount of unrelated or redundant features can extremely reduce the running time of the learning algorithms and produces a more common classifier. Large number of methodologies are proposed to swiftly diminish the computational competences for over high dimension but if lesser number of tremendously valuable features are chosen as a feature subset, some pretty simple classifiers yields worthy results. Thus, feature selection is crucial for cultivating the classification efficiency.

A general feature selection for classification framework is demonstrated in Fig. 1. Feature selection mainly affects the training phase of classification. After generating features, instead of processing data with the whole features to the learning algorithm directly, feature selection for classification will first perform feature selection to select a subset of features and then process the data with the selected features to the learning algorithm.

The feature selection phase might be independent of the learning algorithm, like filter models or it may

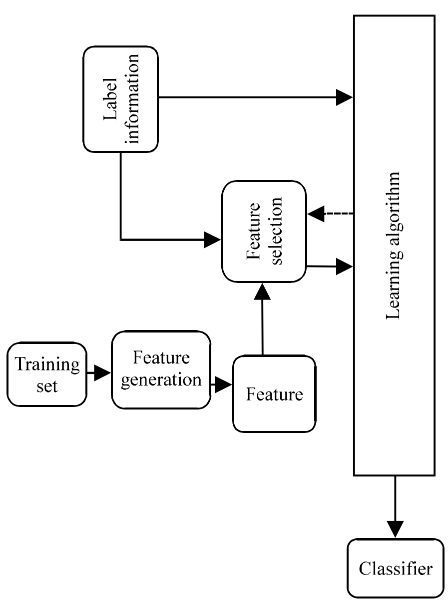

Fig. 1: General feature selection for classification framework

iteratively employ the performance of the learning algorithms to estimate the worth of the chosen characteristics, like wrapper models. With the finally selected features, a classifier is induced for the prediction phase. Feature selection for classification tries to select the negligibly sized subset of characteristics according to the following strategy:

- The classification accuracy does not considerably decrease
- The subsequent class dissemination is as near as likely to original class distribution when simply the values for the selected features are given

In order to improve the performance of the classification system and the classification accuracy, a novel methodology is presented in this study. The proposed methodology implements an effective feature selection procedure depending on improved Particle Swarm Optimization (PSO). This method is mainly interested in reducing the high dimensional data of the large Dengue Databases to a lower dimensional data and then classify the retrieved datasets accordingly. The feature selection technique depending on the improved particle swarm optimization is employed to extract the correlated and relevant attribute using a dependent criteria for dimension reduction. Then, the traditional Decision Tree Classification Technique is applied on the retrieved relevant features for classification of dengue datasets.

MATERIALS AND METHODS

According to whether the training set is labelled or not feature selection algorithms can be classified into supervised (Song *et al*., 2007; Weston *et al*., 2003), unsupervised (Dy and Bradley, 2004; Mitra *et al*., 2002) and semi-supervised feature selection (Zhao and Liu, 2007; Xu *et al*., 2010). There are numerous feature selection methodologies that aid to decrease the number of essential features requisite for classification. Genetic algorithms such as in the methodology presented in is one of the most widespread methods. The PSO has been applied in amalgamation with SVM. The new feature selection methods are constantly increasing to tackle the specific problem with different strategies:

- To ensure a better behavior of feature selection using an ensemble method (Saeys *et al*., 2008; Bolon-Canedo *et al*., 2012)

- Combining with other techniques such as tree ensemble (Tuv *et al*., 2009) and feature extraction (Vainer *et al*., 2011)
- Reinterpreting existing algorithms (Sun and Li, 2006; Sun *et al*., 2008)
- Creating a new method to deal with still unresolved problems (Chidlovskii and Lecerf, 2008; Loscalzo *et al*., 2009)
- To combine several feature selection methods (Zhang *et al*., 2008; El Akadi *et al*., 2011)

This methodology employs a diverse variation of PSO and use merely the traditional binary SVM classifier. Other PSO grounded methodology has been suggested to accomplish feature extraction from hyperspectral data. The PSO that is hybridized with neural network is also proposed to perform regression. This is the initial method in application of PSO and SVM for hyperspectral feature selection with the help of dedicated kernel augmented for spectral classification. Daamouche *et al*. (2013) proposed the use of PSO to select for classification the most informative features obtained by morphological profiles. Particle Swarm Optimization (PSO) (Kennedy and Eberhart, 1995; Shi and Eberhart, 1998) is a comparatively current evolutionary computing technique depending on swarm intelligence. PSO has been employed as an active technique in several areas including feature selection (Unler and Murat, 2010; Liu *et al*., 2011).

In recent times, PSO has attained more attention for resolving feature selection problems. The rough set can handle inaccuracy, uncertainty and vagueness. Wang suggested a filter feature selection methodologies depending on an enriched binary PSO and rough set theory. The goodness of a particle is assigned as the requirement degree among class labels and selected features which is measured by rough set. Chakraborty (2008) matches the performance of PSO with Genetic algorithm in a filter feature selection procedure with a fuzzy set grounded fitness function. The outcomes illustrates that PSO executes better compared to GA in case of the classification performance. Huang and Dun (2008) also suggest a similar feature selection methodology employing two versions of PSO.

Mohemmed *et al*. (2009) suggested a hybrid method known as PSO AdaBoost that integrates PSO with an AdaBoost context for face detection. This targets to explore for the finest feature subset and define the decision thresholds of AdaBoost concurrently which hurries the training procedure and rises the correctness of feeble classifiers in AdaBoost.

Advantages of PSO for feature selection and classification: The Particle Swarm Optimization algorithm will be promising for knowledge discovery using classification method due to:

- The mechanism of PSO has inherent advantages relative to Evolutionary algorithms. PSO has memory to hold the finest location of distinct particle and also of the complete swarm. The preliminary population is preserved throughout since there are no crossover or mutation operations applied to the population. So, PSO has stoutercapability to discovery the best classification rules
- PSO encompasses only one simple evolutionary operator that makes the procedure proficient both in estimation speed and memory constraint. Numerous measures such as a balanced compound fitness function, self-adaptive scale control of particle swarm are taken to improve the performance of the algorithm
- PSO has simple concept along with the fact that it can be employed in a few lines of code. Furthermore, PSO also has a memory of past iterations. On the other hand, in the GA, if a chromosome is not selected, the information contained by it is lost. Without a selection operator PSO may waste resources on inferior individuals. PSO may enhance the search capability for finding an optimal solution

The proposed approach: In this study, a novel methodology is proposed to find the best optimal features from the large database. In this approach, the particle Swarm Optimization algorithm is used as Random Selection algorithm to efficiently explore a vast search space which is frequently required in case of attribute selection and choose attributes to maximize the probability of desired classification. The algorithm describes the proposed methodology step wise. The proposed methodology is broadly implemented in two phases (Fig. 2). They are:

- Phase 1: feature selection using Particle Swarm Optimization (PSO) algorithm and Seperability Correlation Measure (SCM)
- Phase 2: classification of selected attributes using Decision Tree (DT) algorithm

Feature selection using Particle Swarm Optimization algorithm: The proposed feature selection method in this study endeavors to diminish the dimensionality of the data by choosing only the characteristics that are essential for a precise classification. PSO is a kind of

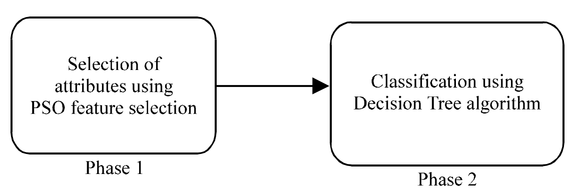

Fig. 2: Structure of proposed methodology

evolutionary optimization methodologies that has exposed to accomplish well in a variety of applications. PSO as a feature selection technique define how applicable it is for choosing features from database. Since, PSO is aimed to search through continuous spaces, it needs to be discretized for use in feature selection. The role of fitness function is to estimate the effectiveness of a certain particle of rule in contradiction to the sample data set.

Particle Swarm Optimization algorithm: A particle swarm optimization is demonstrated as the imitation of the communal behavior of bird flocks (Kennedy and Eberhart, 1995). PSO is easy to implement and has been effectively functioned to resolve a varied collection of optimization problems. Thus, because to its easiness and effectiveness in directing huge search spaces for optimal solutions and its dominance with further Evolutionary algorithm techniques (El Beltagi *et al*., 2005) PSO algorithm is engaged in this research to improve an effectual procedure to enhance feature selection problem (Fig. 3).

PSO algorithm is an intelligent optimization algorithm inferring the bird swarm performances which was proposed by psychologist (Kennedy and Eberhart, 1995). The particle swarm optimization is more objective and easy to perform healthier, it is functioned in various regions such as the function optimization, the neural network training, the fuzzy system control, etc. PSO is initialized with apopulation of individuals. Every individual is considered as apoint in an S-dimensional space. The ith particle isdenoted as:

$$X_i = (x_{i1}, x_{i2}, x_{i3}, ..., x_{is})$$

The finest preceding location, i.e., pbest gives the best fitness value of any particle as:

$$P_i = (p_{i1}, p_{i2}, p_{i3}, ..., p_{is})$$

The global best particle is denoted by 'gbest'. The velocity for particle i is:

$$V_i = (v_{i1}, v_{i2}, v_{i3}, ..., v_{is})$$

The particles are deployed according to the following equations:

$$v_{id} = w \times v_{id} + c1 \times rand() \times (p_{id} - x_{id}) + c2 \times Rand() \times (p_{id} - x_{id}) \quad (1)$$

$$x_{id} = x_{id} + v_{id} \quad (2)$$

where, w is the inertia weight, the acceleration constants c1 and c2 in Eq. 1 signify the weighting of the stochastic acceleration terms that attract every particle in the direction of pbest and gbest positions, rand() and Rand() are two random functions in the range [0, 1].

Particle's velocities on each dimension are restricted to a maximum velocity V_{max}. The traditional PSO is essentially established for continuous optimization complications. To execute feature selection, the standard PSO Model desires to be stretched so as to work with binary information. In precise, the search space D may be a finite group of states and the fitness function fbe a discrete function. Numerous kinds of discrete and binary PSO are offered in the literature (Karthi *et al*., 2009; Shi and Eberhart, 1998).

The proposed feature selection methodology endeavors to decrease the dimensionality of the features by choosing simply the features that are essential for an exact classification. PSO is a kind of evolutionary optimization algorithm that are proved to work fine in a range of applications.

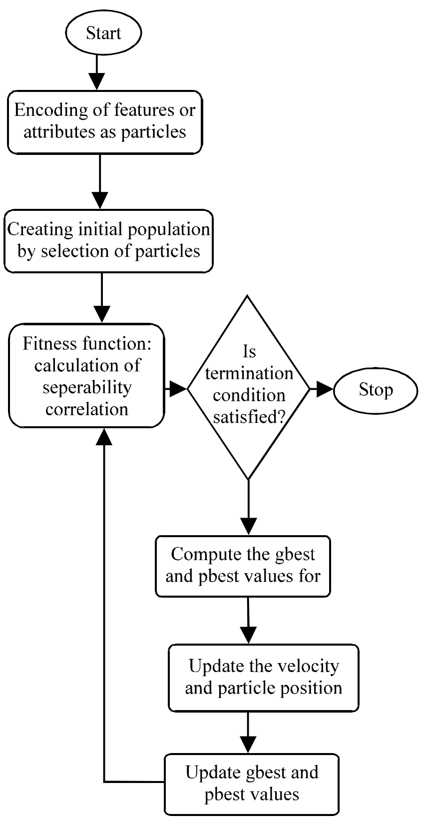

Fig. 3: Flow chart of proposed methodology

Seperability correlation measure as fitness function in the proposed methodology: The likelihood for accurate classification is high when the variations of characteristics among diverse classes are high. Hence, to recognize a subsection of features that can maximize the seperability among classes is a necessary goal for fitness calculation. The correlation measure is employed to merge with the class seperability measure. The significant stages of features is assessed using Seperability-Correlation Measure (SCM) by joining the attribute-class correlation and class seperability measure as follows:

$$R_k = \chi S_k + (1-\chi)C_k$$

where, χ is a weight parameter; $1 \geq \chi \geq 0$ and χ is determined empirically. The finest choice of χ should lead to a subset of attributes that help to move towards the maximum classification accuracy.

The class seperability is measured by the intra class distance, i.e., the distance of patterns within class S_w and the interclass distance, i.e., the distance among patterns of diverse classes S_b. The higher S_b is and the lesser S_w is the improved the seperability of the data set is. Thus, the ratio of S_w and S_b can be used to quantize the difference of the classes: the smaller the ratio, the finest is the seperability:

$$S_k = \frac{S_{wk}}{S_{bk}}$$

And the normalization of S_k is given by:

$$\overline{S}_k = \frac{S_k\text{-}\min(S_k)}{\max(S_k)\text{-}\min(S_k)}$$

The correlation measure (C_k) amid the variations in characteristics and their equivalent variations in class labels are measured when positioning the prominence of attributes. This correlation straightly associations features with class labels. The class labels are different for two diverse patterns and the dissimilarities of attributes in the two patterns are measured to be the affecting reason for the different class labels and should be weighted positively; if the class labels are the similar, the variants in the attributes are unrelated in determining the classes and should be weighted negatively:

$$\overline{C}_k = \frac{C_k\text{-}\min(C_k)}{\max(C_k)\text{-}\min(C_k)}$$

is the normalization of C_k. The significant stages of attributes are categorized by means of the values of R_k. The larger the magnitude of R_k, the more significant the kth attribute.

Classification of selected attribute using Decision Tree (DT) algorithm: The decision trees are influential and prevalent tools for classification and prediction. It exemplify rules which can be understood by humans and employed in knowledge system such as database. Decision tree classifier is a simple and broadly used classification technique. It uses a straight forward idea to resolve the classification problem. Construction of an optimum decision tree is significant problem in decision tree classifier. In broad, numerous decision trees are created from a specified group of attributes. When some of the trees are more correct compared to others, finding the best tree is computationally not possible due to the exponential magnitude of the search space.

Nevertheless, several proficient algorithms have been established to build a sensibly accurate, suboptimal, decision tree in a reasonable amount of time. These procedures typically employ a greedy approach that constructs a decision tree by performing a sequence of locally optimum choice of which attribute to employee for subdividing the data. The traditional decision tree classification technique is performed on the obtained optimized attributes on each sample which generates decision tree rules for the attributes of training data. Then, the generated rules are given to testing data for classification. The classification accuracy is calculated for the testing data. This decision tree recurrently partitions a data set into lesser divisions depending on the trials applied to one or more attributes at individual node of the tree.

In the proposed methodology, the attributes of the training dataset are the initial population that is encoded to numerical string chromosomes to use in Particle Swarm Optimization algorithm. The fitness function used in this methodology is the seperability correlation coefficient measure. The fitness value for the chromosomes are determined using this function. The particle swarm optimization operation, i.e., the updating of pbest and gbest values and moving the particle or attribute in the direction of the velocity of best attributes in order to obtain the best particle and global values. The termination for the algorithm will attain if maximum number of generations has been reached or if there is no changes to the population best fitness for specified number of times of generations. Algorithm 1 and Fig. 3 describes the Particle Swarm Optimization algorithm for feature selection using seperability correlation measure implemented. Only those operations will be fit to perform swarm intelligence operations that are having lower correlation coefficient. That is lower the correlation coefficient the higher is the fitness value.

Algorithm 1:

1. A dataset with M number of samples and N number of attributes in every samples considered.
2. The N number of attributes or features of each sample are encoded into numerical particles that are uniformly distributed.
3. The initial population by selecting the parental chromosomes from the dataset is generated.
4. The fitness value to update particles position calculated using the seperability correlation measure given by:

$$R_k = \chi S_k + (1-\chi)C_k$$

$$\text{And fitness value} = 1\text{-}R_k$$

5. If termination condition is reached go to step 9 else go to step 6.
6. If a particle's current position is enhanced than its previous best position, update it.
7. Determine the best particle or attribute according to the particle's previous best positions.
8. Update particles' velocities using:

$$v_{id} = w\times v_{id} + c1\times rand\times(p_{id}\text{-}x_{id}) + c2\times Rand\times(p_{id}\text{-}x_{id})$$

9. Move particles to their new positions using:

$$x_{id} = x_{id} + v_{id}$$

10. Rank of all the selected and fittest attributes of each sample obtained from genetic operations.
11. Select top K number of optimized attributes of each sample from the ranked ones and give this as input to the decision tree classifier algorithm.
12. Decision tree classifier classifies all the samples into classes depending on the optimized attributes as targets.

RESULTS AND DISCUSSION

The experimental analysis for the proposed study is carried out using the datasets available at http://www.ncbi.nlm.nih.gov/gds. In the study, the experiment is done on the total of 2,175 patient's records, each having 17 attributes along with the target class. Figure 4 represents the interface for running the proposed methodology where the information about the number of features, sample testing and training information, no of optimized features needed are taken from the user.

The Traditional Decision Tree algorithm and the proposed seperability correlation measure based fitness function in Genetic algorithm and decision tree classification accuracy and error rate obtain. Table 1 and Fig. 5 represent the accuracy and error rate of Traditional Decision Tree and Genetic Algorithm combined with decision tree for different optimized features selected by the user and how optimized selected features vary with obtain fitness values.

Table 1 clearly represents that the classification accuracy for the proposed methodology is more when compared to the traditional approach and the error rate is less for the proposed approach compared to traditional approach. Figure 6 and 7 represents the decision trees for traditional and proposed approach, respectively.

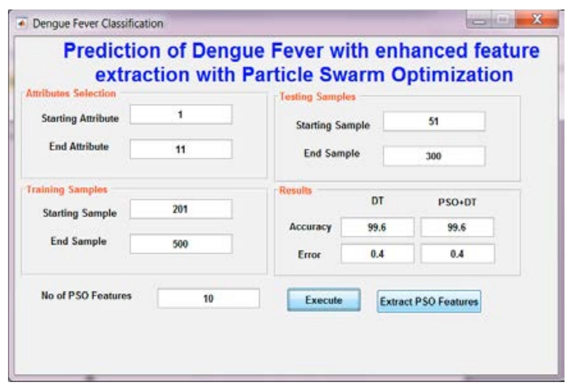

Fig. 4: Interface designed to run proposed algorithm

Table 1: Decision tree and proposed methodology classification accuracy and error rate

Attribute selection		Training samples		Testing samples		Features		DT results		PSO+DT results	
Start	End	Start	End	Start	End	Only DT	GA+DT	Accuracy	Error	Accuracy	Error
1	3	201	500	51	300	3	2	81.20	18.80	99.60	0.40
1	4	201	500	51	300	4	3	81.20	18.80	99.60	0.40
1	5	201	500	51	300	5	4	81.20	18.80	99.60	0.40
1	6	201	500	51	300	6	5	81.20	18.80	99.60	0.40
1	7	201	500	51	300	7	6	81.20	18.80	99.60	0.40
1	8	201	500	51	300	8	7	81.20	18.80	99.60	0.40
1	9	201	500	51	300	9	8	81.20	18.80	99.60	0.40
1	10	201	500	51	300	10	9	81.20	18.80	99.60	0.40
1	11	201	500	51	300	11	10	99.60	0.40	99.60	0.40
1	12	201	500	51	300	12	11	99.60	0.40	99.60	0.40
1	13	201	500	51	300	13	11	99.60	0.40	99.60	0.40
1	11	201	500	51	300	11	2	99.60	0.40	99.60	0.40

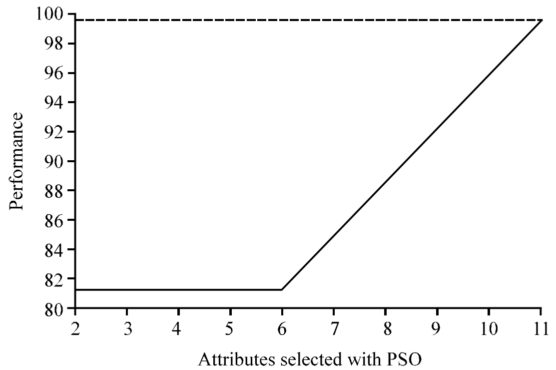

Fig. 5: Performance vs. selected attribute for decision tree classifier and proposed methodology

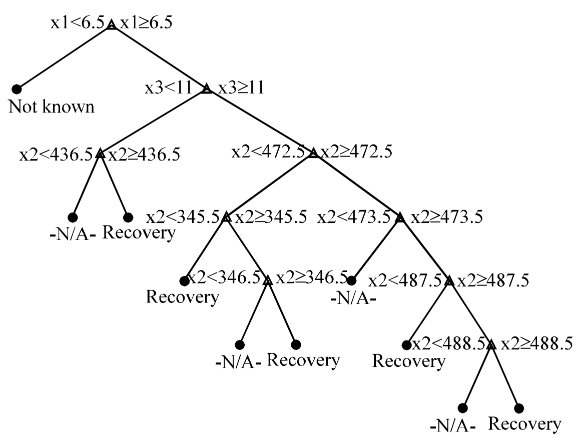

Fig. 6: Decision tree for 3 attributes

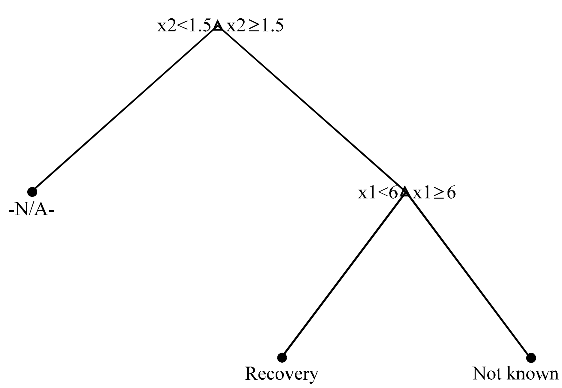

Fig. 7: Decision tree for 3 attributes selected by GA selected out of 5 supplied attributes

CONCLUSION

The Genetic algorithm is one of the soft computing technique employed in this study for optimized feature selection. The classification of the patients information is done using proposed classification system that consider the seperability correlation coefficient measure for the correlating the attributes. The experimental results of the proposed study shows a better classification accuracy compared to the existing methodologies and also shown that there is a considerable decrease in the error rate of the proposed compared to the existing classification system. From the experimental analysis also inferred that our methodology is effective and efficient with respect to the number of correctly classified patterns.

REFERENCES

Bolon-Canedo, V., N. Sanchez-Marono and A. Alonso-Betanzos, 2012. An ensemble of filters and classifiers for microarray data classification. Pattern Recognition, 45: 531-539.

Chakraborty, B., 2008. Feature subset selection by particle swarm optimization with fuzzy fitness function. Proceedings of the 3rd International Conference on Intelligent System and Knowledge Engineering, Volume 1, November 17-19, 2008, Xiamen, pp: 1038-1042.

Chidlovskii, B. and L. Lecerf, 2008. Scalable feature selection for multi-class problems. Proceedings of the European Conference Machine Learning and Knowledge Discovery Databases, September 15-19, 2008, Antwerp, Belgium, pp: 227-240.

Daamouche, A., F. Melgani, N. Alajlan and N. Conci, 2013. Swarm optimization of structuring elements for VHR image classification. Geosci. Remote Sensing Lett., 10: 1334-1338.

Dy, J.G. and C.E. Bradley, 2004. Feature selection for unsupervised learning. J. Mach. Learn. Res., 5: 845-889.

El Akadi, A., A. Amine, A. El Ouardighi and D. Aboutajdine, 2011. A two-stage gene selection scheme utilizing MRMR filter and GA wrapper. Knowledge Inform. Syst., 26: 487-500.

El Beltagi, E., T. Hegazy and D. Grierson, 2005. Comparison among five evolutionary-based optimization algorithms. Adv. Eng. Inform., 19: 43-53.

Huang, C.L. and J.F. Dun, 2008. A distributed PSO-SVM hybrid system with feature selection and parameter optimization. Applied Soft Comput., 8: 1381-1391.

Karthi, R., S. Arumugam and K.R. Kuma, 2009. A novel discrete particle swarm clusteringalgorithm for data clustering. Proceedings of the 2nd Bangalore Annual Compute Conference, January 9-10, 2009, India.

Kennedy, J. and R. Eberhart, 1995. Particle swarm optimization. Proceedings of the IEEE International Conference on Neural Networks, Volume 4, November 27-December 1, 1995, Australia, pp: 1942-1948.

Liu, H. and L. Yu, 2005. Toward integrating feature selection algorithms for classification and clustering. IEEE Trans. Knowl. Data Eng., 17: 491-502.

Liu, Y., G. Wang, H. Chen, H. Dong, X. Zhu and S. Wang, 2011. An improved particle swarm optimization for feature selection. J. Bionic Eng., 8: 191-200.

Loscalzo, S., L. Yu and C. Ding, 2009. Consensus group stable feature selection. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, June 28-July 1, 2009, France, pp: 567-576.

Masaeli, M., J.G. Dy and G.M. Fung, 2010. From transformation-based dimensionality reduction to feature selection. Proceedings of the 27th International Conference on Machine Learning, June 21-24, 2010, Haifa, pp: 751-758.

Mitra, P., C.A. Murthy and S.K. Pal, 2002. Unsupervised feature selection using feature similarity. IEEE Trans. Pattern. Anal. Mach. Intell., 24: 301-312.

Mohemmed, A.W., M. Zhang and M. Johnston, 2009. Particle swarm optimization based adaboost for face detection. Proceedings of the IEEE Congress on Evolutionary Computation, May 18-21, 2009, Trondheim, pp: 2494-2501.

Saeys, Y., T. Abeel and Y. Van de Peer, 2008. Robust feature selection using ensemble feature selection techniques. Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases, September 15-19, 2008, Antwerp, Belgium, pp: 313-325.

Shi, Y. and R. Eberhart, 1998. A modified particle swarm optimizer. Proceedings of the World Congress on Computational Intelligence and IEEE International Conference on Evolutionary Computation, May 4-9, 1998, Anchorage, AK., pp: 69-73.

Song, L., A. Smola, A. Gretton, K.M. Borgwardt and J. Bedo, 2007. Supervised feature selection via dependence estimation. Proceedings of the 24th International Conference on Machine Learning, June 20-24, 2007, Corvallis, OR., pp: 823-830.

Sun, Y. and J. Li, 2006. Iterative RELIEF for feature weighting. Proceedings of the 23rd International Conference on Machine Learning, June 2006, New York, pp: 913-920.

Sun, Y., S. Todorovic and S. Goodison, 2008. A feature selection algorithm capable of handling extremely large datadimensionality. Proceedings of the 8th SIAM International Conference on Data Mining, June 2008, Atlanta, GA., pp: 530-540.

Tuv, E., A. Borisov, G. Runger and K. Torkkola, 2009. Feature selection with ensembles, artificial variables and redundancy elimination. J. Mach. Learn. Res., 10: 1341-1366.

Unler, A. and A. Murat, 2010. A discrete particle swarm optimization method for feature selection in binary classification problems. Eur. J. Operat. Res., 206: 528-539.

Vainer, I., S. Kraus, G.A. Kaminka and H. Slovin, 2011. Obtaining scalable and accurate classification in large-scale spatio-temporal domains. Knowledge Inform. Syst., 29: 527-564.

Weston, J., A. Elisseeff, B. Scholkopf and M. Tipping, 2003. Use of the zero norm with linear models and kernel methods. J. Mach. Learn. Res., 3: 1439-1461.

Xu, Z., R. Jin, J. Ye, M. Lyu and I. King, 2010. Discriminative semi-supervised feature selection via manifold regularization. Proceedings of the 21th International Joint Conference on Artificial Intelligence, June 21-July 8, 2010, Germany, pp: 1033-1047.

Zhang, Y., C. Ding and T. Li, 2008. Gene selection algorithm by combining reliefF and mRMR. BMC Genomics, Vol. 9. 10.1186/1471-2164-9-S2-S27.

Zhao, Z. and H. Liu, 2007. Semi-supervised feature selection via spectral analysis. Proceedings of SIAM International Conference on Data Mining, June 20-24, 2007, USA.