

Improving Nearest Neighbor Classifier using Tabu Search and Ensemble Distance Metrics

Muhammad Atif Tahir and James Smith
School of Computer Science
University of the West of England
BS16 1QY, Bristol, United Kingdom

E-mail: {muhammad.tahir, james.smith@uwe.ac.uk}

Abstract

The nearest-neighbor (NN) classifier has long been used in pattern recognition, exploratory data analysis, and data mining problems. A vital consideration in obtaining good results with this technique is the choice of distance function, and correspondingly which features to consider when computing distances between samples. In this paper, a new ensemble technique is proposed to improve the performance of NN classifier. The proposed approach combines multiple NN classifiers, where each classifier uses a different distance function and potentially a different set of features (feature vector). These feature vectors are determined for each distance metric using Simple Voting Scheme incorporated in Tabu Search (TS). The proposed ensemble classifier with different distance metrics and different feature vectors (TS-DF/NN) is evaluated using various benchmark data sets from UCI Machine Learning Repository. Results have indicated a significant increase in the performance when compared with various well-known classifiers. Furthermore, the proposed ensemble method is also compared with ensemble classifier using different distance metrics but with same feature vector (with or without Feature Selection (FS)).

1 Introduction

The nearest-neighbor (NN) classifier has long been used in pattern recognition, exploratory data analysis, and data mining problems. Typically, the k nearest neighbors of a unknown sample in the training set are computed in order to predict the class label as the most frequent one occurring in the k nearest-neighbors. The 1NN classifier is well explored in the literature and has been proved to have good classification performance on a wide range of real-world

data sets [1, 2, 3].

Feature selection is useful technique for improving the classification accuracy of NN rule [7, 8]. The term feature selection refers to algorithms that select the best subset of the input feature set. These algorithms are used in the design of pattern classifiers that have three goals [9, 11]:

1. to reduce the cost of extracting features
2. to improve the classification accuracy
3. to improve the reliability of the estimation of performance

Tabu search (TS) has been applied to the problem of feature selection by Zhang and Sun [12]. In their work, the tabu search performs the feature selection in combination with an objective function based on Mahalanobis distance. This objective function is used to evaluate the classification performance of each subset of the features selected by the TS. Feature selection vector in TS is represented by a 0/1 bit string where 0 indicates the feature is not included in the solution while 1 indicates the feature is included. Their experimental results on *synthetic data* have shown that the tabu search not only has a high possibility to obtain the optimal or near-optimal solution, but also requires less computational effort than the other suboptimal and genetic algorithm based methods. Later, Tabu Search has been successfully applied in other feature selection problems [8, 13, 14].

Recently, Bao et al. [10] have proposed an ensemble technique to improve the accuracy of NN classifier. Multiple NN classifiers are combined using different distance functions. The premise is that ensemble classifiers will be more accurate than their consistent members provided that the errors of the individual classifiers are not correlated. Although Bao's et al. [10] ensemble approach does use different distance metrics, they use the same set of features, so it is possible that some errors will be common, arising from

features containing noise which have high values in certain samples. An alternative approach is proposed by Bay [15] that uses a single distance metric, but instead builds an ensemble where each member sees a different randomly selected subset of the features, and a simple voting scheme.

In this paper, a new algorithm is proposed that combines and extends these ideas. Following building on Bay's and Bao's work [10, 15], we explore the hypothesis that the overall ensemble accuracy can be improved if those choices of subsets arise from

- tabu search rather than random sampling and
- different distance metrics rather than single distance metric

Furthermore we hypothesise that these choices are best co-adapted, rather than learnt separately, as co-adaptation may permit implicit tackling of the problem of achieving ensemble diversity. In order to do this, and to distinguish the effects of different sources of benefits, a novel ensemble classifier is proposed in this paper that consists of multiple NN classifiers each using different feature subsets derived using tabu search. To increase the diversity, simple voting scheme is introduced in the cost function of Tabu Search. The proposed ensemble NN classifier (TS-DF/1NN) is then compared with various well-known classifiers.

The rest of the paper is organized as follows. Section 2 describes proposed multiple distance function ensemble classifier followed by experiments in section 3. Section 4 concludes the paper.

2 Proposed Ensemble Multiple Distance Function TS/1NN classifier (TS-DF/1NN)

In this section, we will discuss about the proposed ensemble multiple distance function TS/1NN classifier (TS-DF/1NN). Figure 1 shows the training phase of the proposed classifier. During each iteration, N random neighbors with *HammingDistance*¹ are generated for each distance metric using nearest neighbor (NN) classifier as cost function. M ¹ best neighbors are selected from N neighbors during each iteration. M^n solutions are then evaluated using simple voting scheme (SVS) to find feature vectors for various distance measure in which errors are not correlated. Thus, the feedback from the SVS allows tabu search to iteratively search for feature vectors that improves the classification accuracy. By using n distance functions, n feature vectors are obtained using TS in the training phase. The use of different distance functions, each with a potentially different set of features along with SVS has increased the likelihood

¹In this study $M = 2$, $n = 5$, and $N = \sqrt{F}$, where F = Total Number of Features

that the errors of the individual classifiers are not correlated. In the testing phase, multiple NN classifiers are combined using n distance functions and n different feature vectors as shown in Figure 2.

In the following subsections, Feature selection using TS and various distance metrics used in this paper are discussed as they are the heart of the proposed algorithm.

2.1 Distance Metrics

The following five distance metrics are used for NN classifier. All metrics are widely used in the literature.

- Squared Euclidean Distance: $E = \sum_{i=1}^m (x_i - y_i)^2$
- Manhattan Distance: $M = \sum_{i=1}^m (x_i - y_i)$
- Canberra Distance: $C = \sum_{i=1}^m (x_i - y_i) / (x_i + y_i)$
- Squared chord distance: $S_c = \sum_{i=1}^m (\sqrt{x_i} - \sqrt{y_i})^2$
- Squared Chi-squared distance: $C_s = \sum_{i=1}^m (x_i - y_i)^2 / (x_i + y_i)$

where x and y are the two input vectors and m is the number of features.

2.2 Feature Selection and Diversity using Tabu Search

Tabu Search (TS) was introduced by Glover [5, 6] as a general iterative metaheuristic for solving combinatorial optimization problems. TS is conceptually simple and elegant. It is a form of local neighborhood search. TS starts from initial solution, and then examine feasible neighboring solutions. It moves from a solution to its best admissible neighbor, even if this causes the objective function to deteriorate. To avoid cycling, solutions that were recently explored are declared forbidden or tabu for a number of iterations. The tabu list stores characterization of the moves with lead to those solutions. The tabu status of a solution is overridden when certain criteria (aspiration criteria) are satisfied. Sometimes intensification and diversification strategies are used to improve the search. In the first case, the search is accentuated in promising regions of the feasible domain. In the second case, an attempt is made to consider solutions in a broad area of the search space. The flow chart of TS algorithm is given in Figure 3.

The size of tabu list can be determined by experimental runs, watching for the occurrence of cycling when the size is too small, and the deterioration of solution quality when the

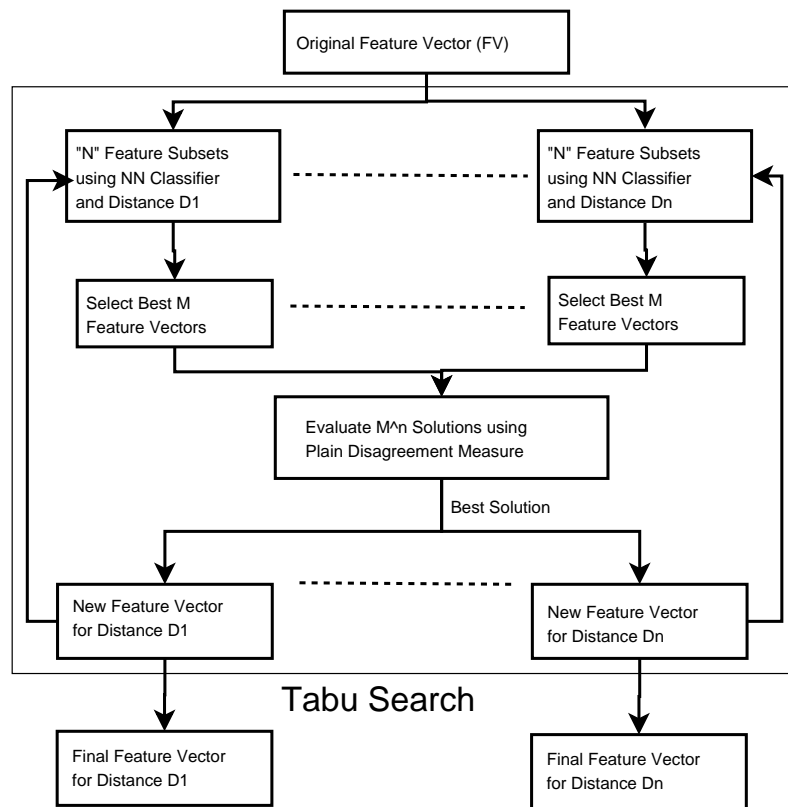


Figure 1. Training Phase of proposed TS-DF/1NN classifier.

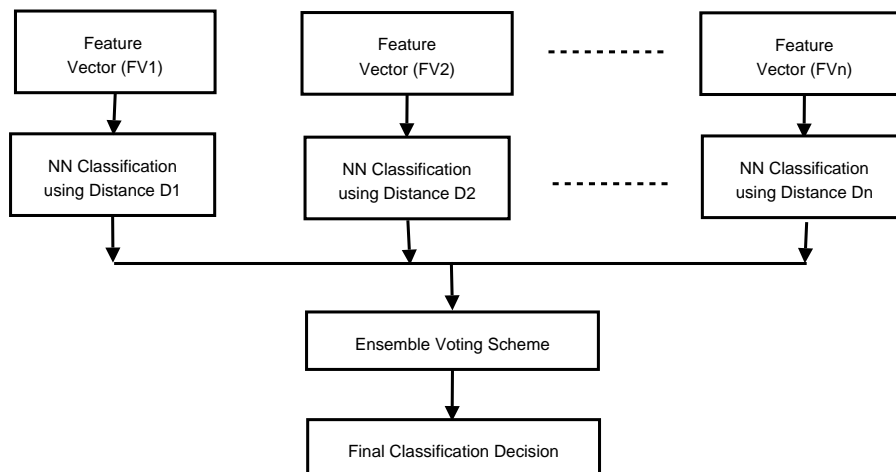


Figure 2. Testing Phase.

size is too large [16]. Suggested values of tabu list include Y, \sqrt{Y} (where Y is related to problem size, e.g. number of modules to be assigned in the quadratic assignment problem (QAP), or the number of cities to be visited in the travel salesman problem (TSP), and so on) [17].

Algorithm Short-Term-TS

Ω : Set of feasible solutions
 S : Current Solution
 S^* : Best admissible solution
 $Cost$: Objective function
 $N(S)$: Neighborhood of solution S
 V^* : Sample of neighborhood solutions
 T : Tabu list
 AL : Aspiration Level

Begin

1. Start with an initial feasible solution $S \in \Omega$.
2. Initialize tabu list and aspiration level.
3. For fixed number of iterations Do
4. Generate neighbor solutions $V^* \subset N(S)$.
5. Find best $S^* \in V^*$.
6. If move S to S^* is not in T Then
7. Accept move and update best solution.
8. Update tabu list and aspiration level.
9. Increment iteration number.
10. Else
11. If $Cost(S^*) < AL$ Then
12. Accept move and update best solution.
13. Update tabu list and aspiration level.
14. Increment iteration number.
15. End If
16. End If
17. End For

End

Figure 3. Algorithmic description of Tabu Search(TS) [17]

2.2.1 Objective Function:

Simple Voting Scheme is used in each instance of n classifiers. The objective function is the number of instances incorrectly classified using Simple Voting Scheme. The objective is to minimize Equation 1.

$$Cost = \sum_{i=1}^S C_i \quad (1)$$

where S is the number of samples, $C_i = 1$ if instance is classified incorrectly after simple voting in n classifiers else $C_i = 0$.

2.2.2 Initial Solution

The feature selection vector is represented by a 0/1 bit string where 0 indicates that the feature is not included in the solution while 1 indicates that it is. All features are included in the initial solution.

2.2.3 Neighborhood solutions

During each iteration, N random neighbors with *HammingDistance1*($HD1$) are generated for each distance metric using nearest neighbor (NN) classifier as cost function. Neighbors are generated by randomly adding or deleting a feature from the feature vector of size n . Among the neighbors, M best solutions are selected. M^n solutions are then evaluated using Equation 1 and the one with the best cost (i.e. the solution which results in the minimum value of Equation 1) is selected and considered as a new current solution for the next iteration. Figure 4 shows an example showing neighborhood solutions during one iteration. Lets assume that the cost of initial solution is 50, 48, and 47 using Distance Metrics 1, 2, and 3 respectively. $N = 4$ neighbors are then randomly generated for each distance metric using $HD1$. $M = 2$ best solutions are selected and $M^n = 2^3 = 8$ solutions are evaluated using ensemble cost function. The best solution is then selected for the next iteration.

2.2.4 Tabu moves

A tabu list is maintained to avoid returning to previously visited solutions. In our approach, if an ensemble solution (move) is selected at iteration i , then selecting the same ensemble solution (move) for T subsequent iterations (tabu list size) is Tabu.

2.2.5 Aspiration criterion

Aspiration criterion is a mechanism used to override the tabu status of moves. It temporarily overrides the tabu status if the move is sufficiently good. In our approach, if an ensemble solution is selected at iteration i and this move results in a best cost for all previous iterations, then that solution is selected even if that feature is in the tabu list.

2.2.6 Termination rule

The most commonly used stopping criteria in TS are

- after a fixed number of iterations.
- after some number of iterations when there has been no increase in the objective function value.
- when the objective function reaches a pre-specified value.

In this work, the termination condition is the fixed number of iterations.

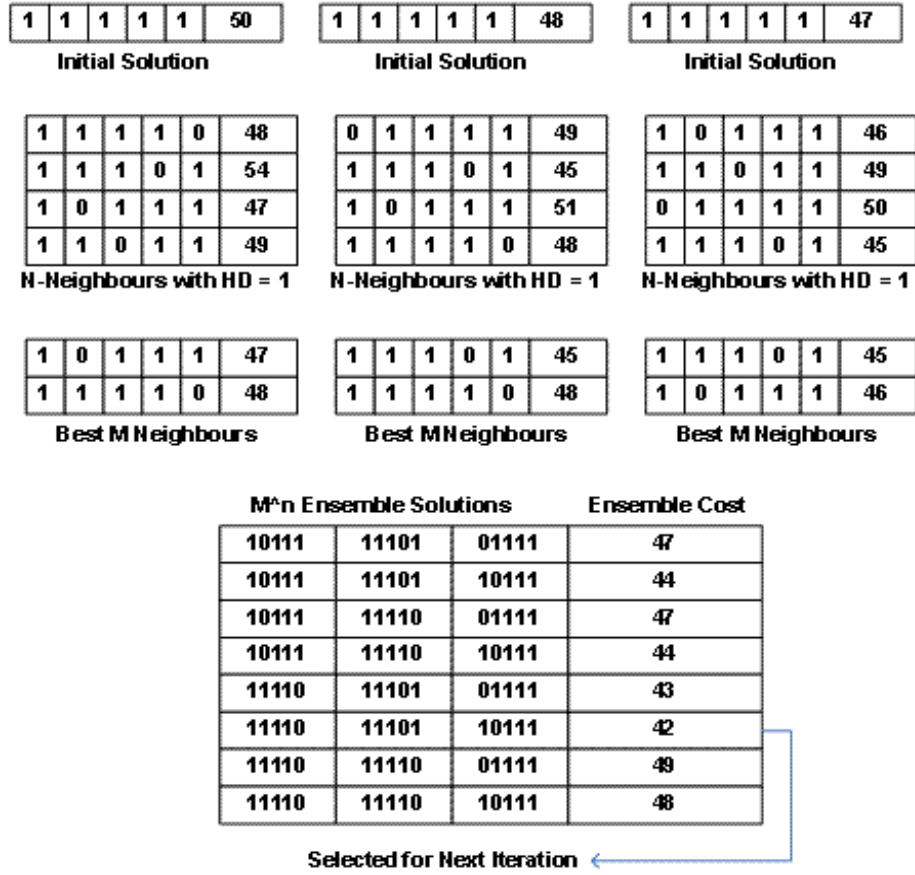


Figure 4. An example showing neighborhood solutions during one iteration in proposed tabu search method. $n = 3$, $N = 4$, and $M = 2$.

3 Experiments

To evaluate the effectiveness of our method, extensive experiments were carried out. Comparisons with several methods were also performed as will be shown in the following section.

3.1 Methods

The proposed (TS-DF/INN) algorithm is compared with the following methods. All methods are implemented using WEKA library [26].

- Decision Tree Method (C4.5): Decision tree is a classifier in the form of a tree structure, where each node is either a leaf node or a decision node [20, 3].
- Decision Table (DT): It uses a simple decision table majority classifier [21].

- Random Forest (RF): Classifier for constructing a forest of random trees [22].
- Naive Bayes Algorithm (NBayes): The Naive Bayes Classifier technique is based on Bayesian theorem. Despite its simplicity, Naive Bayes can often outperform numerous sophisticated classification methods [23].
- Bagging: A method for generating multiple versions of a predictor and using these to get an aggregated predictor [24]. C4.5 is used as base classifier.
- AdaBoost1: A meta-algorithm, and can be used in conjunction with many other learning algorithms to improve their performance [25]. C4.5 is used as base classifier.

In addition, we compare with the following variations of the proposed ensemble algorithms

1. DF/1NN: Ensemble Classifier using various distance metrics and without Feature Selection.
2. TS-DF/1NN₁: Ensemble classifier but with a common feature subset for various distance metrics derived using tabu search.

3.2 Data sets Descriptions and Experimental Setup:

We have performed a number of experiments and comparisons on several benchmarks from the UCI [4] in order to demonstrate the performance of the proposed classification system. A short description of the used benchmarks along with Tabu Search run-time parameters are mentioned in Table 1.

Table 1. Data sets Description. P = Prototypes, F = Features, C= Classes, T = Tabu List Size, N = Number of Neighborhood Solutions.

Name	P	F	C	T	N
Statlog Diabetes	768	8	2	3	3
Glass	214	9	6	3	3
Statlog Heart	270	13	2	4	4
Statlog Australian	690	14	2	4	4
Statlog Vehicle	846	18	4	5	5
Statlog German	1000	20	2	5	5
WDBC	569	32	2	6	6
Ionosphere	351	34	2	6	6
Sonar	208	60	2	8	8
Musk	476	166	2	13	13

The tabu list size and Number of Neighborhood Solutions are determined using the following equation:

$$T = N = \text{ceil}(\sqrt{F}) \quad (2)$$

where T is the Tabu List Size, N is the number of neighborhood solutions and F is the number of features.

In all data sets, B fold cross validation has been used to estimate error rates [18]. For B -fold CV, each data set is divided into B blocks using $(B-1)$ blocks as a training set and the remaining block as a test set. Therefore, each block is used exactly once as a test set. Each experiment was run 100 times using different random 10-CV partitions and the results were averaged over the 100 runs [19].

The number of iterations for FS using TS is 200 for all data sets which is chosen after preliminary experimentation.

3.3 Results and Discussion

Table 3.3 shows the comparison of classification accuracy (in %) between proposed TS-DF/1NN classifier and

others for different data sets. The proposed ensemble feature selection (FS) technique using TS/NN has achieved higher accuracy in all data sets except Diabetes. For Australian, German and Ionosphere data sets there is improvement of 1.98%, 5.06% and 0.4% respectively when compared with Random Forest Classifier. For Heart, there is an improvement of 3.3% when compared with Naive Bayes Classifier. For Vehicle, WDBC and Musk data sets, there is improvement of 0.5%, 0.76%, and 4.55% respectively when compared with AdaBoost. For Sonar, there is an improvement of 7.8% when compared with 1NN. Since Diabetes has only 8 features, the proposed algorithm unable to combine the benefits of Feature Selection and Ensemble Classifiers using different distance metrics. Figure 5 shows the standard deviation obtained over the 100 runs of random 10-fold cross validation of each data set for different algorithms. From the graph, it is clear that the standard deviation of the proposed classifier compares favorably with other algorithms.

Tables 3 shows the classification accuracy comparison when compared with the various variations of the proposed method. In all data sets, the proposed ensemble classifier has outperformed other methods. The performance of an ensemble improves when all of its classifiers use a common feature subset derived using tabu search and the cost function is the simple voting scheme. However, the performance of an ensemble is further increased when different feature subsets are used for different classifiers.

Figures 6 and 7 show the classification accuracy (%) vs number of iterations for Australian data set using one run of the solution search space using TS. The figure clearly indicates that TS focuses a good solution space. The proposed TS algorithm progressively zooms towards a better solution subspace as time elapses; a desirable characteristics of approximation iterative heuristics.

4 Conclusion

A new ensemble technique is proposed in this paper to improve the performance of nearest neighbor (NN) classifier. The proposed approach combines multiple NN classifiers, where each classifier uses a different distance function and potentially a different set of features (feature vector). These feature vectors are determined independently for each distance metric using Tabu Search (TS). To increase the diversity, simple voting scheme is introduced in the cost function of Tabu Search. The proposed ensemble distance TS/NN classifier is evaluated using various benchmark data sets from UCI Machine Learning Repository. Results have indicated a significant increase in the performance when compared with different well-known classifiers. Furthermore, the proposed TS based ensemble classifier progressively zoomed towards a better solution subspace as time

Table 2. Average Classification Accuracy (%) using different classifiers. DT = Decision Table. RF = Random Forest.

Data Set	C4.5	DT	RF	NBayes	Bagging	AdaBoost	1NN	TS-DF/1NN
Australian	84.34	84.71	86.09	77.14	86.04	84.94	79.62	88.07
Heart	78.24	82.32	80.17	83.98	80.47	79.15	75.68	87.28
Ionosphere	89.8	94.2	95.4	92.8	92.2	90.3	87.5	95.80
Vehicle	72.74	66.43	74.71	45.36	74.49	76.44	69.72	76.90
WDBC	93.59	93.32	95.88	93.31	95.35	96.06	95.38	96.82
Diabetes	74.29	74.11	74.73	75.62	76.0	72.39	70.31	74.5
Musk	82.72	80.82	87.81	73.88	88.21	89.98	85.58	94.53
Sonar	72.98	72.61	80.27	67.89	78.46	80.08	86.51	94.29
German	71.55	72.45	74.74	74.47	74.61	72.50	70.88	79.80

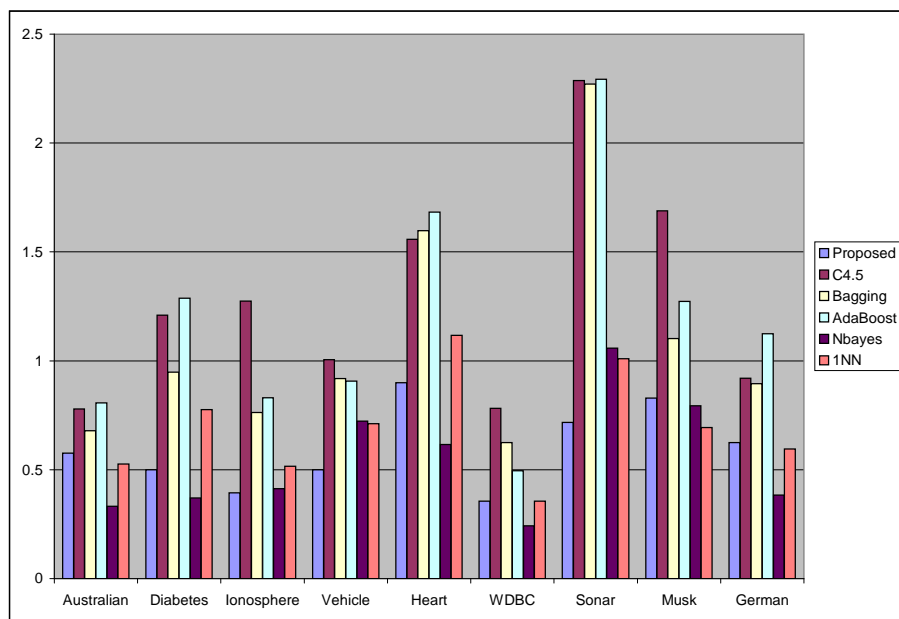


Figure 5. Standard deviation for different algorithms on various data sets.

Table 3. Classification Accuracy (%) using individual classifiers and various variations of the proposed classifier. M =Manhattan, E =Euclidean, C =Canberra, C_s = Chi-Squared, S_c = Squared-Chord.

Data Set	E	M	C	C_s	S_c	DF/NN ₁	TS-DF/NN ₂	TS-DF/NN
Australian	80.9	80.1	82.4	83.5	81.9	82.1	85.1	88.1
Heart	80.3	77.1	73.8	81.6	78.7	76.6	83.8	86.3
Vehicle	74.0	72.8	74.8	74.1	74.5	70.8	74.0	76.9
German	72.2	72.8	72.9	73.5	73.0	72.4	74.2	79.8
Diabetes	71.0	70.3	70.5	72.0	70.3	69.3	72.5	74.5
Breast Cancer	91.7	92.6	96.9	95.2	95.5	94.7	96.1	96.8
Ionosphere	93.6	92.1	93.2	92.8	93.0	90.3	95.1	96.3
Sonar	88.8	85.9	91.3	87.8	89.1	87.7	92.1	94.3
Musk	85.7	82.3	80.5	83.9	83.6	86.2	92.3	94.5

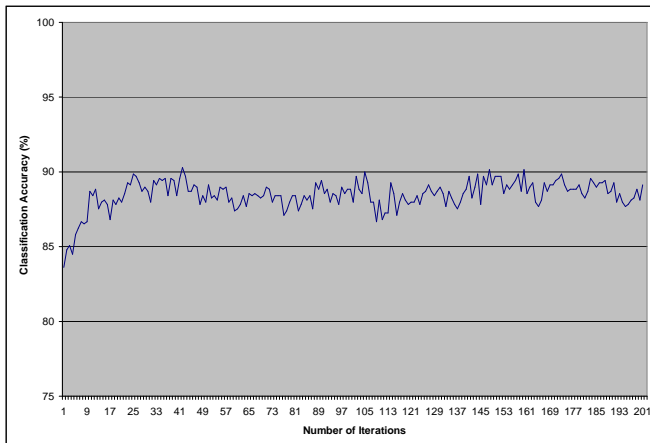


Figure 6. Error Rate vs Iterations for Australian Data set.

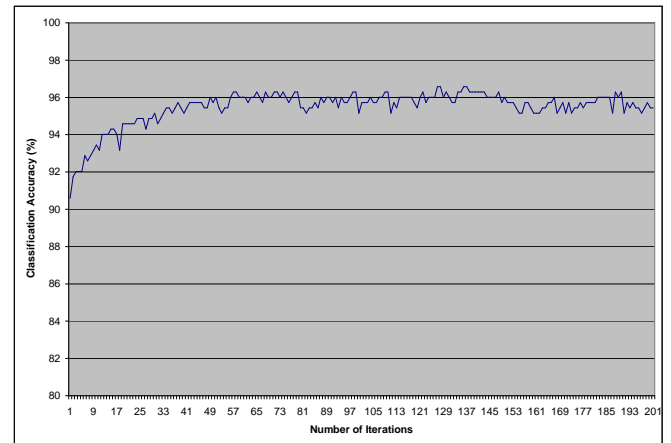


Figure 7. Error Rate vs Iterations for Ionosphere Data set.

elapsed, a desirable characteristic of approximation iterative heuristics.

Acknowledgement

This work is supported by the European Commission (Project No. STRP016429, acronym DynaVis). This publication reflects only the authors view.

References

- [1] T. M. Cover, and P. E. Hart. *Nearest Neighbor Pattern Classification*. IEEE Transactions on Information Theory. 13(1), 21–27, 1967.
- [2] C. Domeniconi, J. Peng, and D. Gunopulos. *Locally Adaptive Metric Nearest-Neighbor Classification*. IEEE Transactions on Pattern Analysis and Machine Intelligence. 24(9), 1281–1285, 2002.
- [3] D. Michie, D. J. Spiegelhalter and C. C. Taylor. *Machine Learning, Neural and Statistical Classification*. Ellis Horwood. 1994.
- [4] C. Blake, E. Keogh, and C. J. Merz. UCI Repository of machine learning databases, University of California, Irvine.
- [5] F. Glover. *Tabu search I*. ORSA Journal on Computing, 1(3), 190–206, 1989.
- [6] F. Glover. *Tabu search II*. ORSA Journal on Computing, 2(1), 4–32, 1990.
- [7] M. L. Raymer et al. *Dimensionality Reduction using Genetic Algorithms*. IEEE Transactions on Evolutionary Computation, 4(2), 164–171, 2000.
- [8] M. A. Tahir et al. *Feature Selection using Tabu Search for Improving the Classification Rate of Prostate Needle Biopsies*. IEEE Proceedings of the International Conference on Pattern Recognition, Cambridge, 2004.
- [9] A. K. Jain, and R. P. W. Duin, and J. Mao. *Statistical Pattern Recognition: A Review*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(1), 4–37, 2000.
- [10] Y. Bao and N. Ishii and X. Du. *Cobmining Multiple k-Nearest Neighbor Classifiers Using Different Distance Functions* Lecture Notes in Computer Science (LNCS 3177), 5th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2004), Exeter, UK.
- [11] M. Kudo and J. Sklansky. *Comparison of algorithms that select features for pattern classifiers*. Pattern Recognition. 33, 25–41, 2000.
- [12] H. Zhang and G. Sun. *Feature selection using tabu search method*. Pattern Recognition., 35, 701–711, 2002.
- [13] M. A. Tahir, A. Bouridane, and F. Kurugollu. *Simultaneous Feature Selection and Weighthing for Nearest Neighbor Using Tabu Search*. Lecture Notes in Computer Science (LNCS 3177), 5th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2004), Exeter, UK.

- [14] D. Korycinski, M. Crawford, J. W. Barnes, and J. Ghosh. *Adaptive feature selection for hyperspectral data analysis using a binary hierarchical classifier and tabu search*. Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2003.
- [15] S. D. Bay. *Combining Nearest Neighbor Classifiers Through Multiple Feature Subsets*. Proceedings of the Fifteenth International Conference on Machine Learning, 37–45, 1998.
- [16] F. Glover, E. Taillard, and D. de Werra. *A user's guide to tabu search*. Annals of Operations Research., 41, 3–28, 1993.
- [17] S. M. Sait and H. Youssef. *General Iterative Algorithms for Combinatorial Optimization*. IEEE Computer Society. 1999.
- [18] S. Raudys and A. Jain. *Small Sample Effects in Statistical Pattern Recognition: Recommendations for Practitioners*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(3), 252–264, 1991.
- [19] R. Paredes and E. Vidal. *Learning Weighted Metrics to Minimize Nearest-Neighbor Classification Error*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(7), 1100–1110, 2006.
- [20] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan-Kaufmann. 1993.
- [21] R. Kohavi. *The Power of Decision Tables*. Proceedings of the 8th European Conference on Machine Learning, 1995.
- [22] L. Breiman. *Random Forests*. Machine Learning, 45(1), 5–32, 2001.
- [23] R. Duda and P. Hart. *Pattern Classification and Scene Analysis*. Wiley, New York, 1973.
- [24] L. Breiman. *Bagging predictors*. Machine Learning, 24(2), 123–140, 1996.
- [25] Y. Freund and R. E. Schapire. *Experiments with a new boosting algorithm*. Proceedings of International Conference on Machine Learning, 148–156, 1996.
- [26] I. H. Witten and E. Frank.
- [27] , 2nd Edition, Morgan Kaufmann, San Francisco, 2005.