Evolving Committees of Support Vector Machines*

D. Valincius[1,3], A. Verikas[1,2], M. Bacauskiene[1], and A. Gelzinis[1]

[1] Department of Applied Electronics, Kaunas University of Technology, Studentu 50, LT-51368, Kaunas, Lithuania
[2] Intelligent Systems Laboratory, Halmstad University, Box 823, S-30118 Halmstad, Sweden
[3] UAB "Elinta", Pramones pr. 16E, LT-51187 Kaunas
donatas.valincius@elinta.lt, antanas.verikas@ide.hh.se, mabaca@ktu.lt, adas.gelzinis@ktu.lt

Abstract. The main emphasis of the technique developed in this work for evolving committees of support vector machines (SVM) is on a two phase procedure to select salient features. In the first phase, clearly redundant features are eliminated based on the paired t-test comparing the SVM output sensitivity-based saliency of the candidate and the noise feature. In the second phase, the genetic search integrating the steps of training, aggregation of committee members, and hyper-parameter as well as feature selection into the same learning process is employed. A small number of genetic iterations needed to find a solution is the characteristic feature of the genetic search procedure developed. The experimental tests performed on five real world problems have shown that significant improvements in correct classification rate can be obtained in a small number of iterations if compared to the case of using all the features available.

1 Introduction

Aggregating outputs of multiple predictors into a committee output is one of the most important techniques for improving prediction accuracy [1,2,3]. An efficient committee should consist of predictors that are not only very accurate, but also diverse in the sense that the predictor errors occur in different regions of the input space [4,5]. Manipulating training data set, using different architectures, and employing different subsets of variables are the most popular approaches used to achieve the diversity. To promote diversity of neural networks aggregated into a committee, Liu and Yao [6,7] proposed the so-called *Negative correlation learning* approach, according to which, all individual networks in the committee are trained simultaneously, using an error function augmented with a correlation penalty term. In [8], aiming to find a trade-off between the accuracy and diversity

* We gratefully acknowledge the support we have received from the agency for international science and technology development programmes in Lithuania (EUREKA Project E!3681).

P. Perner (Ed.): MLDM 2007, LNAI 4571, pp. 263–275, 2007.

of committee networks, the approach was extended by integrating into the same learning process also the feature selection step. However, to assess and control diversity of predictors and to find the trade-off between the accuracy and diversity is not a trivial task [9,10]. For instance, feature selection may influence the quality of a committee in several ways, namely by reducing model complexity, promoting diversity of committee members, and affecting the trade-off between the accuracy and diversity of committee members. Therefore it seems promising to integrate the steps of training, hyper-parameter and feature selection, and aggregation of members into a committee into the same learning process and to use the prediction accuracy to assess the quality of the committee.

This paper is concerned with such an approach to evolving committees of support vector machines for classification. The main emphasis of the paper is on feature selection for classification committees. A large variety of feature selection techniques have been proposed for a single predictor [11,12], ranging from the sequential forward selection or backward elimination [13,14], sequential forward floating selection [15] to the genetic [16] or tabu search [17]. However, works on feature selection for classification or regression committees are very scarce [5]. It has been demonstrated that even simple random selection of feature subsets may be an effective technique for increasing the accuracy of classification committees [18,19].

One needs to assess the feature saliency when selecting features. The Predictor output sensitivity [20,21,22,23] is the most popular measure used to assess the saliency. Eq. 1 exemplifies such a measure [20,21]

$$\Upsilon_i = \frac{1}{QP} \sum_{j=1}^{Q} \sum_{p=1}^{P} \left| \frac{\partial y_{jp}}{\partial x_{ip}} \right| \tag{1}$$

where y is the predictor output, Q is the number of outputs, P is the number of training samples, and x_{ip} is the ith component of the pth input vector $\mathbf{x}_p$. However, a saliency measure alone does not indicate how many of the candidate features should be used. Therefore, some of feature selection procedures are based on making comparisons between the saliency of the candidate and the noise feature [20,21]. Nonetheless the usefulness of such comparisons, the measure does not have direct relation to the prediction error.

The procedure developed in this work for evolving classification committees consists of two phases. In the first phase, clearly redundant features are eliminated based on the paired t-test comparing the saliency of the candidate feature and the noise feature in a single classifier. Then, in the second phase, the genetic search integrating the steps of training, aggregation of committee members into a committee, search for the optimal hyper-parameter values, and selection amongst the remaining features into the same learning process is employed. The committee prediction accuracy is the measure used to assess the committee quality in the genetic search. A small number of genetic iterations needed to find a solution is the characteristic feature of the genetic search procedure developed. The rationale of using the first phase of the procedure is to reduce the computation time needed for the genetic search. If the computation time is not a

problem, the first phase of the procedure can be skipped. We use an SVM as a committee member in our tests. However, other types of classifiers can also be utilized.

2 Procedure

The procedure for evolving classification committees is summarized in the following steps.

1. Augment the input vectors with one additional noise feature.
2. Train the model.
3. Calculate the saliency score Γ_i,
$$\Gamma_i = \frac{\Upsilon_i}{\max_{l=1,...,N} \Upsilon_l}, \quad i = 1, ..., N \tag{2}$$
where Υ_i is given by (1) and N is the number of features.
4. Repeat Steps 2 to 3 K times using different random data partitioning into training, validation and test sets.
5. Eliminate features the saliency of which, do not exceed the saliency of the noise feature. Use the paired t-test to compare the saliency values.
6. Choose the number of committee members L. Construct a chromosome characterizing feature inclusion/noninclusion, regularization and kernel parameters of all the committee members. More details on the chromosome definition are given in Section 2.3.
7. Perform the genetic search.
8. The committee is given by the parameters encoded in the "best" chromosome found during the genetic search.

2.1 The Paired t-Test

To assess the equality of the mean saliency of ith feature μ_{Γ_i} and the noise μ_{Γ_n} the paired t-test is defined as suggested in [21]: `Null Hypothesis` $\mu_{D_i} = 0$, `Alternative Hypothesis` $\mu_{D_i} > 0$, where $\mu_{D_i} = \mu_{\Gamma_i} - \mu_{\Gamma_n}$. To test the null hypothesis, a t^* statistic
$$t^* = \frac{\overline{D}_i}{S_{\overline{D}_i}} \tag{3}$$
is evaluated, where $\overline{D}_i = K^{-1} \sum_{j=1}^{K} D_{ij}$, $D_{ij} = \Gamma_{ij} - \Gamma_{nj}$, Γ_{ij} and Γ_{nj} are the saliency scores computed using (2) for the ith and the noise feature, respectively, in the jth loop, and
$$S_{\overline{D}_i} = \sqrt{\frac{\sum_{j=1}^{K} (D_{ij} - \overline{D}_i)^2}{K(K-1)}} \tag{4}$$
Under the null hypothesis, the t^* statistic is t distributed. If $t^* > t_{crit}$, the hypothesis that the difference in the means is zero is rejected, where t_{crit} is the critical value of the t distribution with $\nu = K - 1$ degrees of freedom for a significance level of α: $t_{crit} = t_{1-\alpha,\nu}$.

2.2 The SVM Output Sensitivity, an Example

The output of a support vector machine $y(\mathbf{x})$ is given by:

$$y(\mathbf{x}) = \sum_{j=1}^{N_s} \alpha_j^* d_j \kappa(\mathbf{x}_j, \mathbf{x}) + b \tag{5}$$

where N_s is the number of support vectors, $\kappa(\mathbf{x}_j, \mathbf{x})$ is a kernel, d_j is a target value ($d_j = \pm 1$), and the threshold b and the parameter α_j^* values are found as a solution to the optimization problem defined by the type of SVM used. In this work, we used the 1-norm soft margin SVM [24]. The parameters α_j satisfy the following constrains:

$$\sum_{j=1}^{N_s} \alpha_j y_j = 0, \quad \sum_{j=1}^{N_s} \alpha_j = 1, \quad 0 \le \alpha_j \le C, \quad j = 1, ..., N_s \tag{6}$$

with C being the regularization constant.

For a Gaussian kernel $\kappa(\mathbf{x}_j, \mathbf{x}_k) = \exp\{-||\mathbf{x}_j - \mathbf{x}_k||^2/\sigma\}$, where σ is the standard deviation of the Gaussian, having the jth input vector $\mathbf{x}_j$ presented to the input, the derivative of the output with respect to the ith feature is given by:

$$\frac{\partial y(\mathbf{x}_j)}{\partial x_{ij}} = -\frac{2}{\sigma} \sum_{k=1}^{N_s} \alpha_k^* d_k (x_{ij} - x_{ik}) \exp\Big\{ - \sum_{n=1}^{N} \frac{(x_{nj} - x_{nk})^2}{\sigma} \Big\} \tag{7}$$

2.3 Genetic Search

Chromosome design, initial population generation, evaluation, selection, crossover, mutation, and reproduction are the issues to consider when designing a genetic search algorithm. We divide **the chromosome** into sections and each section into parts. The number of sections is equal to the number of committee members L. There are three parts in each section. One part encodes the regularization constant C, one the kernel width σ, and the third one encodes the inclusion/noninclusion of features. The binary encoding scheme has been adopted in this work. Fig. 1 illustrates the chromosome structure, where NC and $N\sigma$ stand for the number of bits used to encode the regularization constant C and the kernel width σ, respectively and N is the number of features.

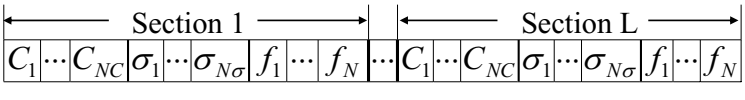

Fig. 1. The structure of the chromosome consisting of L sections

To generate the **initial population**, information obtained from the first feature selection phase, namely, the values of C and σ, and the maximum number of features, is exploited. The maximum number of features allowed for one

committee member is equal to the number of features determined in the first phase. In the initial population, the features are masked randomly and values of the parameters C and σ are chosen randomly from the interval $[C_0 - \Delta C, C_0 + \Delta C]$ and $[\sigma_0 - \Delta\sigma, \sigma_0 + \Delta\sigma]$, respectively, where C_0 and Δ_0 are the parameter values obtained from the first phase.

The **fitness function** used to evaluate the chromosomes is given by the correct classification rate of the validation set data. In this study, the committee output was obtained by averaging the outputs of committee members. To distinguish between more than two classes, the one vs one pairwise-classification scheme has been used.

The **selection process** of a new population is governed by the fitness values. A chromosome exhibiting a higher fitness value has a higher chance to be included in the new population. The selection probability of the ith chromosome p_i is given by

$$p_i = \frac{r_i}{\sum_{j=1}^{M} r_j} \tag{8}$$

where r_i is the correct classification rate obtained from the model based on the ith chromosome and M is the population size.

The **crossover operation** for two selected chromosomes is executed with the probability of crossover p_c. If a generated random number from the interval [0,1] is larger than the crossover probability p_c, the crossover operation is executed. Crossover is performed separately in each section of a chromosome. In the "feature mask" and two parameter parts of each section, the crossover point is randomly selected and the corresponding parts of two chromosomes selected for the crossover operation are exchanged at the selected point.

The **mutation operation** adopted is such that each gene is selected for mutation with the probability p_m. The mutation operation is executed independently in each part of each chromosome section. If the gene selected for mutation is in the feature part of the chromosome, the value of the bit representing the feature in the feature mask is reversed. To execute mutation in the parameter part of the chromosome, to choices are possible: i. to reverse the value of the bit in the parameter representation determined by the selected gene; ii. to mutate the value of the offspring parameter determined by the selected gene by $\pm\Delta\gamma$, where γ stands for C or σ, as the case may be. The mutation sign is determined by the fitness values of the two chromosomes, namely the sign resulting into a higher fitness value is chosen. The way of determining the mutation amplitude $\Delta\gamma$ is somewhat similar to that used in [25] and is given by

$$\Delta\gamma = w\beta(\max(|\gamma - \gamma_{p1}|, |\gamma - \gamma_{p2}|)) \tag{9}$$

where γ is the actual parameter value of the offspring, $p1$ and $p2$ stand for parents, $\beta \in [0, 1]$ is a random number, and w is the weight decaying with the iteration number:

$$w = k(1 - t/T) \tag{10}$$

where t is the iteration number, k is a constant, and T is the total number of iterations.

In the **reproduction process**, the newly generated offspring replaces the chromosome with the smallest fitness value in the current population, if a generated random number from the interval [0,1] is larger than the reproduction probability p_r or if the fitness value of the offspring is larger than that of the chromosome with the smallest fitness value.

3 Experimental Investigations

In all the tests, we run an experiment 30 times with different random partitioning of the data set into `<Learning>`, D_l, `<Validation>`, D_v, and `<Test>`, D_t data sets. The mean values and standard deviations of the correct classification rate presented in this paper were calculated from these 30 trials. The parameter values used in the genetic search have been found experimentally. The following values worked well in all the tests: $p_c = 0.05$, $p_m = 0.02$, and $p_r = 0.05$.

3.1 Data Used

To test the approach we used five real-world problems. Data characterizing four of the problems: *US congressional voting records problem*, *The diabetes diagnosis problem*, *Wisconsin breast cancer problem*, and *Wisconsin diagnostic breast cancer problem* are available at: `www.ics.uci.edu/~mlearn/`. The fifth problem concerns classification of laryngeal images [26].

Laryngeal images. The task is to automatically categorize colour laryngeal images (images of vocal folds) into the *healthy*, *nodular*, and *diffuse* decision classes [26]. Fig. 2 presents characteristic examples from the three decision classes considered.

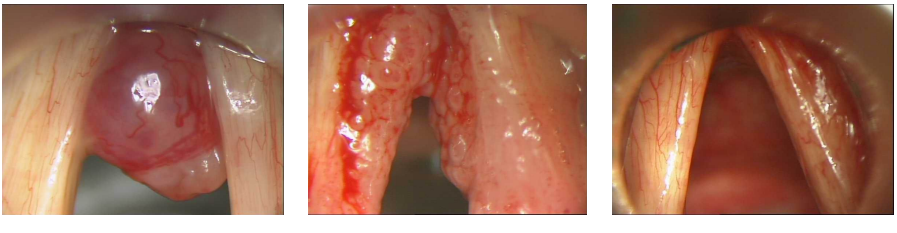

Fig. 2. Images from the *nodular* (left), *diffuse* (middle), and *healthy* (right) classes

Due to a large variety of appearance of vocal folds, the categorization task is sometimes difficult even for a trained physician. Fig. 3 provides an example of such a task. The image placed on the right-hand side of the figure comes from the *nodular* class, while the other two are taken from the *healthy* vocal folds. In this case, the only discriminative feature is the slightly convex vocal fold edges in the upper part of the image coming from the *nodular* class.

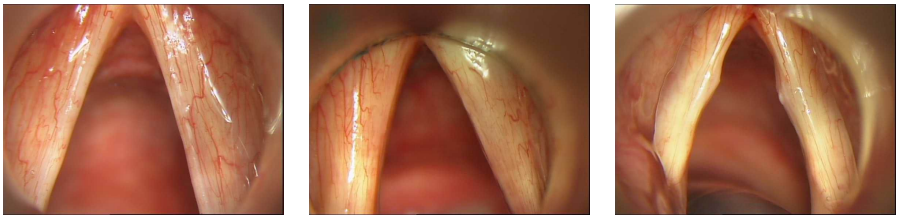

Fig. 3. Three examples of laryngeal images

Aiming to obtain a comprehensive description of laryngeal images, multiple feature sets exploiting information on image colour, texture, geometry, image intensity gradient direction, and frequency content are extracted [27]. Image colour distribution, distribution of the image intensity gradient direction, parameters characterizing the geometry of edges of vocal folds, distribution of the spectrum of the Fourier transform of the colour image complex representation (two types of the frequency content based features), and parameters calculated from multiple co-occurrence matrices are the feature types used to describe laryngeal images [27]. A separate SVM is used to categorize features of each type into the decision classes. The final image categorization is then obtained based on the decisions provided by a committee of support vector machines. In this work, there were 49 images from the *healthy* class, 406 from the *nodular* class, and 330 from the *diffuse* class. Out of the 785 images available, 650 images were assigned to the set D_l.

3.2 Results

First, the average test data set correct classification rate obtained from a single SVM without any involvement of the designing procedure proposed was estimated. The optimal values of the regularization constant C and the kernel width σ have been selected experimentally. Table 1 presents the average test data set correct classification rate obtained for the first four data sets from a single SVM when using all the original features in the classification process. The number of classes and the number of features available are also given in the table. In the parentheses, the standard deviation of the correct classification rate is provided. The average test data set correct classification rate obtained when using a separate SVM for each type of features extracted from the laryngeal images is shown in Table 2.

In the next experiment, we studied the effectiveness of the feature selection procedure applied to single SVMs. Table 3 summarizes the results of the test concerning the first four problems. Apart from the average test data set correct classification rate obtained using the selected features, the table also provides the number of selected features and the number of genetic iterations required to achieve the solution. The number of features eliminated in the first selection phase has been equal to 1, 1, 6, and 12 for the *Diabetes*, *WBCD*, *Voting*,

Table 1. The average test data set correct classification rate obtained for the different data sets from a single SVM when using all the original features

Data set	Number of Classes	Number of features	Classification rate
Diabetes	2	8	76.87 (1.60)
WBCD	2	9	96.86 (0.79)
Voting	2	16	95.49 (1.03)
WDBC	2	30	97.23 (1.01)

Table 2. The average test data set correct classification rate obtained when using a separate SVM for each type of features extracted from the laryngeal images

Feature type	Number of classes	Number of features	Classification rate
Gradient	3	1000	52.30 (5.80)
Co-occurrence	3	42	83.63 (3.17)
Frequency (F1)	3	180	83.38 (3.43)
Frequency (F2)	3	40	78.02 (3.04)
Geometrical	3	18	69.19 (3.48)
Colour	3	50	91.80 (2.69)

and *WDBC* databases, respectively. Observe that the first two problems are characterized by 8 and 9 features, respectively. Thus, there are very few clearly redundant features. The larger number of features eliminated in the first phase for the other two problems significantly speeds up the genetic search executed in the second phase.

Table 3. The average test data set correct classification rate obtained for the different data sets from a single SVM when using the selected features

Data set	Average number of selected features	Average number of iterations	Classification rate
Diabetes	4	8	77.64 (1.50)
WBCD	6	7	97.20 (0.75)
Voting	3	12	96.30 (0.96)
WDBC	17	20	98.06 (0.73)

As it can be seen from Table 1 and Table 3, for all the databases, the average correct classification rate obtained from the single SVMs trained on the selected feature sets is higher than that achieved using all the features available. The number of genetic iterations needed to achieve the solutions is very small. The

number of attempts made to make the crossover operation during one genetic iteration is equal to the population size, which was set 50 in all the tests. Fig. 4 provides two graphs plotting the correct classification rate as a function of the number of genetic iterations for the *WDBC* and *Voting* databases. For each genetic iteration, the performance of the best (*max*), the average (*mean*) and the worst (*min*) population member is shown in Fig. 4. The performance achieved by the best member at the end of the search procedure is also shown.

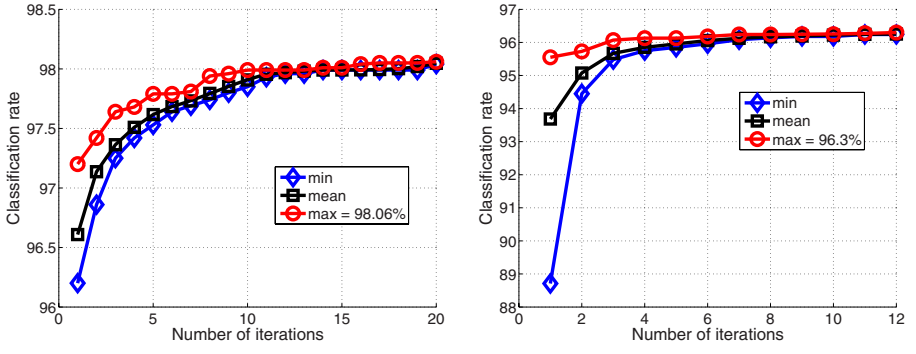

Fig. 4. The test data set correct classification rate obtained from a single SVM as a function of the number of genetic iterations for the Wisconsin diagnostic breast cancer (*left*) and the US congressional voting records (*right*) data sets

Table 4. The average test data set correct classification rate obtained for the different types of features extracted from laryngeal images when using a separate SVM for each type of selected features

Feature type	Average number of selected features	Average number of iterations	Classification rate
Gradient	362	17	83.65 (4.40)
Co-occurrence	28	13	85.48 (3.63)
Frequency (F1)	78	37	89.68 (2.36)
Frequency (F2)	29	13	79.56 (3.47)
Geometrical	10	13	72.12 (3.53)
Colour	42	13	92.74 (2.58)

The results obtained for the different feature sets characterizing the laryngeal images are summarized in Table 4. The number of features eliminated in the first feature selection phase ranged from 5 to over 400. As it can be seen from Table 2 and Table 4, a considerable improvement in classification accuracy has been obtained using the proposed SVM designing approach. The number of

features chosen is considerably lower than that presented in Table 2, especially for the *Gradient* and *Frequency* (*F1*) feature types. On average, a very small number of genetic iterations was required to find the solutions. Fig. 5 provides two graphs plotting the correct classification rate as a function of the number of genetic iterations for the two types of frequency features. For each genetic iteration, the performance of the best (*max*), the average (*mean*) and the worst (*min*) population member is shown.

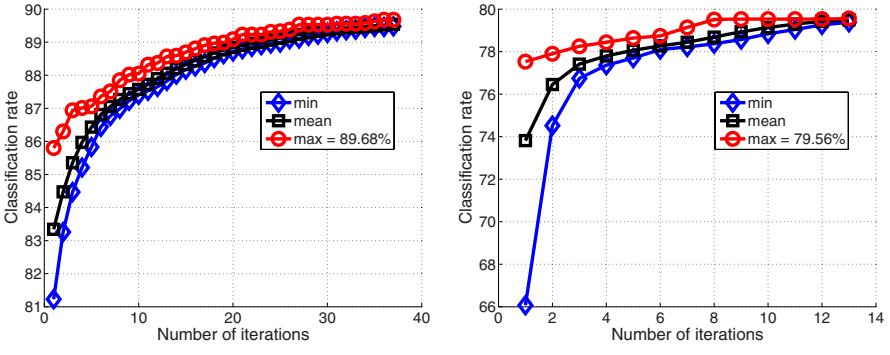

Fig. 5. The test data set correct classification rate obtained from a single SVM as a function of the number of genetic iterations for the two types of frequency features extracted from the laryngeal images

In the last experiment, the effectiveness of the feature selection procedure applied to SVM committees has been studied. Table 5 summarizes the results of the experiment.

Table 5. The average test data set correct classification rate obtained for the different data sets from a committee when using the selected features

Data set	Average number of selected features	Average number of iterations	Classification rate
Diabetes	5	8	77.66 (1.50)
WBCD	5	14	97.27 (0.59)
Voting	6	37	96.62 (0.79)
WDBC	9	20	98.31 (0.46)
Laryngeal	95	8	95.04 (1.88)

All the committees were made of six members. All six members of the committees built for solving the first four problems used the same initial feature set. Each member of the committee built for solving the Laryngeal problem utilized a different feature set—one of the six available types. The average test data set

correct classification rate, the average number of features used by one committee member, and the number of iterations needed to obtain the solution are given in Table 5. As it can be seen from Table 5, the technique developed is capable of evolving accurate classification committees in a small number of genetic iterations. The relatively large average number of features used by the "laryngeal" committee is due to the large number of "gradient" features selected. Fig. 6 provides two graphs plotting the test data set correct classification rate obtained from the committees as a function of the number of genetic iterations for the Laryngeal (*left*) and the Wisconsin diagnostic breast cancer (*right*) problems.

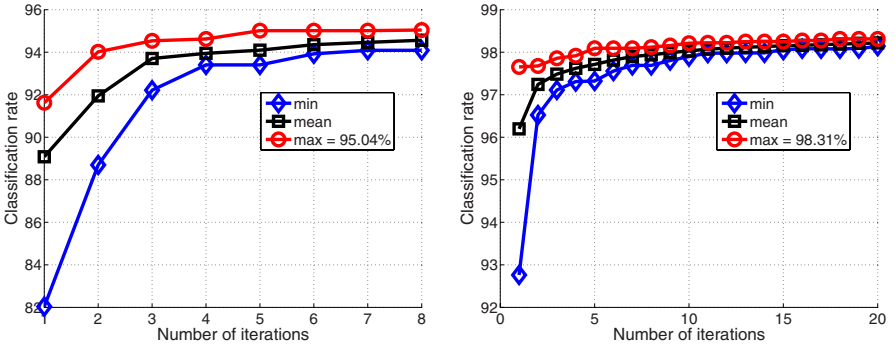

Fig. 6. The test data set correct classification rate obtained from the committee as a function of the number of genetic iterations for the Laryngeal (*left*) and the Wisconsin diagnostic breast cancer (*right*) data sets

4 Conclusions

A technique for evolving committees of support vector machines has been presented in this work. The main emphasis of the technique is on selection of salient features. Elimination of clearly redundant features in the first phase of the procedure developed speeds up the genetic search executed in the second phase of the designing process. The genetic search integrating the steps of training, aggregation of committee members, and hyper-parameter as well as feature selection into the same learning process allows creating effective models in a small number of genetic iterations. The experimental tests performed on five real world problems have shown that considerable improvements in classification accuracy can be obtained using the proposed SVM designing approach.

References

1. Gader, P.D., Mohamed, M.A., Keller, J.M.: Fusion of handwritten word classifiers. Pattern Recognition Letters 17, 577–584 (1996)
2. Liu, C.L.: Classifier combination based on confidence transformation. Pattern Recognition 38, 11–28 (2005)

3. Verikas, A., Lipnickas, A., Malmqvist, K., Bacauskiene, M., Gelzinis, A.: Soft combination of neural classifiers: A comparative study. Pattern Recognition Letters 20, 429–444 (1999)
4. Krogh, A., Vedelsby, J.: Neural network ensembles, cross validation, and active learning. In: Tesauro, G., Touretzky, D.S., Leen, T.K. (eds.) Advances in Neural Information Processing Systems, vol. 7, pp. 231–238. MIT Press, Cambridge (1995)
5. Bacauskiene, M., Verikas, A.: Selecting salient features for classification based on neural network committees. Pattern Recognition Letters 25, 1879–1891 (2004)
6. Liu, Y., Yao, X.: Ensemble learning via negatine correlation. Neural Networks 12, 1399–1404 (1999)
7. Liu, Y., Yao, X., Higuchi, T.: Evolutionary ensembles with negative correlation learning. IEEE Trans on Evolutionary Computation 4, 380–387 (2000)
8. Bacauskiene, M., Cibulskis, V., Verikas, A.: Selecting variables for neural network committees. In: Wang, J., Yi, Z., Zurada, J.M., Lu, B.-L., Yin, H. (eds.) ISNN 2006. LNCS, vol. 3971, pp. 837–842. Springer, Heidelberg (2006)
9. Kuncheva, L.I., Whitaker, C.J.: Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. Machine Learning 51, 181–207 (2003)
10. Brown, G., Wyatt, J., Harris, R., Yao, X.: Diversity creation methods: a survey and categorisation. Information Fusion 6, 5–20 (2005)
11. Kudo, M., Sklansky, J.: Comparison of algorithms that select features for pattern classifiers. Pattern Recognition 33, 25–41 (2000)
12. Verikas, A., Bacauskiene, M.: Feature selection with neural networks. Pattern Recognition Letters 23, 1323–1335 (2002)
13. Mao, K.Z.: Orthogonal forward selection and backward elimination algorithms for feature subset selection. IEEE Trans Systems, Man, & Cybernetics—Part B: Cybernetics 34, 629–634 (2004)
14. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. Machine Learning 46, 389–422 (2002)
15. Pudil, P., Novovicova, J., Kittler, J.: Floating search methods in feature selection. Pattern Recognition Letters 15, 1119–1125 (1994)
16. Yu, S., Backer, S.G., Scheunders, P.: Genetic feature selection combined with composite fuzzy nearest neighbor classifiers for hyperspectral satellite imagery. Pattern Recognition Letters 23, 183–190 (2002)
17. Zhang, H., Sun, G.: Feature selection using tabu search method. Pattern Recognition 35, 701–711 (2002)
18. Ho, T.K.: The random subspace method for constructing decision forests. IEEE Trans Pattern Analysis and Machine Intelligence 20, 832–844 (1998)
19. Tsymbal, A., Puuronen, S., Patterson, D.W.: Ensemble feature selection with simple Bayesian classification. Information Fusion 4, 87–100 (2003)
20. Priddy, K.L., Rogers, S.K., Ruck, D.W., Tarr, G.L., Kabrisky, M.: Bayesian selection of important features for feedforward neural networks. Neurocomputing 5, 91–103 (1993)
21. Steppe, J.M., Bauer, K.W.: Improved feature screening in feedforward neural networks. Neurocomputing 13, 47–58 (1996)
22. Acir, N., Guzelis, C.: Automatic recognition of sleep spindles in EEG via radial basis support vector machine based on a modified feature selection algorithm. Neural Computing & Applications 14, 56–65 (2005)
23. Evgeniou, T., Pontil, M., Papageorgiou, C., Poggio, T.: Image representations and feature selection for multimedia database search. IEEE Trans Knowledge and Data Engineering 15, 911–920 (2003)

24. Shawe-Taylor, J., Cristianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press, Cambridge, UK (2004)
25. Leung, K.F., Leung, F.H.F., Lam, H.K., Ling, S.H.: Application of a modified neural fuzzy network and an improved genetic algorithm to speech recognition. Neural Computing & Applications 16 (2007)
26. Verikas, A., Gelzinis, A., Bacauskiene, M., Uloza, V.: Integrating global and local analysis of colour, texture and geometrical information for categorizing laryngeal images. International Journal of Pattern Recognition and Artificial Intelligence 20, 1187–1205 (2006)
27. Verikas, A., Gelzinis, A., Valincius, D., Bacauskiene, M., Uloza, V.: Multiple feature sets based categorization of laryngeal images. Computer Methods and Programs in Biomedicine 85, 257–266 (2007)