Available online at www.sciencedirect.com
SCIENCE DIRECT®
Pattern Recognition Letters 27 (2006) 301–307
Pattern Recognition Letters
www.elsevier.com/locate/patrec

An ensemble-driven *k*-NN approach to ill-posed classification problems

Mingmin Chi, Lorenzo Bruzzone *

DIT, University of Trento, Via Sommarive 14, I-38050 Trento, Italy

Available online 12 October 2005

Abstract

This paper addresses the supervised classification of remote-sensing images in problems characterized by relatively small-size training sets with respect to the input feature space and the number of classifier parameters (ill-posed classification problems). An ensemble-driven approach based on the *k*-nearest neighbor (*k*-NN) classification technique is proposed. This approach effectively exploits semilabeled samples (i.e., original unlabeled samples labeled by the classification process) to increase the accuracy of the classification process. Experimental results obtained on ill-posed classification problems confirm the effectiveness of the proposed approach, which significantly increases both the accuracy and the reliability of classification maps.

Keywords: Ill-posed classification problems; Semisupervised classification; Semilabeled samples; Ensemble methods; *k*-nearest neighbor technique; Automatic classification; Remote sensing

1. Introduction

In recent years the development of remote-sensing technology has made it possible to greatly enhance our ability to monitor and manage natural resources and the environment. This technological improvement, together with the development of advanced automatic and supervised classification techniques, resulted in the possibility to produce reliable and accurate land-cover maps on a regular basis. However, one of the most critical problems relating to the supervised classification of remote-sensing images lies in the definition of a training set of proper size for an accurate learning of classifier parameters. Since the collection of ground-reference data is an expensive and complex task, in many cases the number of training samples is insufficient for a proper learning of classification systems. This results in ill-posed (or poorly-posed) classification problems (Jackson and Landgrebe, 2002; Baraldi et al., 2005) (also called small-size training set problems or problems affected by the Hughes phenomenon (Hughes, 1968; Shahshahani and Landgrede, 1994)), which involve classifiers with poor generalization capabilities. This issue is particularly critical when multisensor/multisource data sets or hyperspectral images are considered, because due to the intrinsic high dimension of the feature space, it is impossible to meet the requirements on the necessary number of training samples.

One possible way of addressing ill-posed classification problems is to include unlabeled (or semilabeled) samples in the training set. Though the positive effect of these patterns cannot be guaranteed, several theoretical (Shahshahani and Landgrede, 1994; Blum and Mitchell, 1998) and practical (Tadjudin and Landgrebe, 1996; Bennett and Demiriz, 1998; Fardanesh and Ersoy, 1998; Fung and Mangasarian, 1999; Joachims, 1999; Cohen et al., 2003; Kemp et al., 2003; Dundar and Landgrebe, 2004; Chi and Bruzzone, 2005) studies that have been carried out in the context of different applications (e.g. text classification (Blum and Mitchell, 1998), computer vision (Cohen et al., 2003), remote-sensing (Shahshahani and Landgrede, 1994; Jackson and Landgrebe, 2002; Chi and Bruzzone, 2005) show the validity and effectiveness of this kind of approach. At this point, it is important to mention the works relating

* Corresponding author. Tel.: +39 0461 882056; fax: +39 0461 882093.
E-mail addresses: chi@dit.unitn.it (M. Chi), lorenzo.bruzzone@ing.unitn.it (L. Bruzzone).

0167-8655/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2005.08.012

to the joint use of the expectation–maximization (EM) algorithm (Moon, 1996) with unlabeled samples to increase the reliability and the accuracy of the estimation of the statistical class distribution both in parametric (Shahshahani and Landgrede, 1994; Jackson and Landgrebe, 2002) and semiparametric (Blum and Mitchell, 1998; Cohen et al., 2003) classifiers. However, the use of unlabeled and semilabeled samples to solve ill-posed classification problems is not only limited to methods based on the EM algorithm. Other specific approaches based on nonparametric techniques (e.g., neural networks (Fardanesh and Ersoy, 1998), kernel classifiers (Dundar and Landgrebe, 2004), support vector machines (Bennett and Demiriz, 1998; Fung and Mangasarian, 1999; Joachims, 1999), and tree-based classifiers (Tadjudin and Landgrebe, 1996; Kemp et al., 2003) have been proposed in the pattern recognition literature to address this issue. Nonetheless, their application to remote-sensing problems is very limited at present.

In this paper, extending and reinforcing the work presented in (Chi and Bruzzone, 2005), we proposed a novel ensemble-driven approach based on the joint use of labeled and semilabeled samples for solving ill-posed classification problems. For the dual objective of developing a nonparametric method capable of dealing with multisensor/multisource data and studying the problem in the context of a well-established statistical classifier, we developed our approach through the well-known k-nearest neighbor (k-NN) technique. We chose this technique on account of the simplicity of this basic approach and its intrinsic inability to address ill-posed problems (which is due to the fact that it adopts local-based estimation procedure that significantly suffers from the very small number of available training samples). In this way, it is possible to assess the effectiveness of the proposed ensemble-driven method in a very critical condition. It is worth noting that this choice is different and complementary to the one considered in (Chi and Bruzzone, 2005), in which a multilayer perceptron neural network (MLPNN) classification algorithm (based on a global estimation procedure) was considered. The novelties of the proposed approach lie in the following:

(i) It defines an effective scheme for generating hybrid training sets[1] by the joint use of both labeled (training) and semilabeled samples;
(ii) It develops an ensemble-driven strategy for the proper exploitation of different classifiers based on semilabeled samples. This strategy (in proper assumptions) is capable of improving significantly the classification accuracy and stability in ill-posed problems;
(iii) It uses the k-NN classifier to address ill-posed classification problems.

[1] A hybrid training set is defined in this paper as a set that includes both original training samples and semilabeled samples.

In order to assess the effectiveness of the proposed approach, simulated ill-posed classification problems have been defined using multispectral Landsat Thematic Mapper images acquired on the Trentino area (Italy). Experimental results confirm that the presented method is capable of increasing both the accuracy and robustness of classification in ill-posed problems.

This paper is organized in four sections. In Section 2, the proposed ensemble-driven approach based on the k-NN technique is presented. Section 3 describes the data set used in the experiments and the results obtained by the proposed approach. Finally, Section 4 draws the conclusions of this work.

2. Ensemble-driven k-NN approach to ill-posed classification problems

2.1. k-nearest neighbor technique

The k-nearest neighbor technique is one of the simplest statistical nonparametric classifiers that have been studied extensively from both the theoretical and the practical point of view. This algorithm classifies each pattern based on the labels of the k closest training samples. This process can be modeled as a local estimation of the conditional posterior probabilities of classes based on the relative frequency of the class labels in a neighborhood (defined by the k closest training samples).

Let X be a d-dimensional feature vector, and $\Omega = \{\omega_1, \ldots, \omega_C\}$ the set of C land-cover classes that characterize the considered problem. Let T be a training set made up of B labeled samples. Given the pattern X_i, the k-NN estimate $\widehat{P}(\omega_j|X_i)$ of the conditional posterior probability $P(\omega_j|X_i)$ is obtained according to the analysis of the labels of the k samples (included in the training set T) closest to X_i (which define the neighbor N_i). The classification rule can be written as

$$X_i \in \omega_m \quad \text{if and only if}$$
$$\omega_m = \arg\max_{\omega_j \in \Omega} \left\{ \widehat{P}(\omega_j|X) \right\} = \arg\max_{\omega_j \in \Omega} \left\{ \frac{\text{number of patterns} \in \omega_j \text{ in } N_i}{k} \right\} \tag{1}$$

It is worth noting that the k-NN technique is intrinsically unsuitable to address ill-posed classification problems. However, in this paper, we analyze the proposed (general) approach in the context of this technique in order to assess its performance with a critical (though theoretically well established) classifier.

2.2. Ensemble-driven approach

One of the main problems relating to the use of semilabeled samples in the classification task lies in the risk of defining a "negative-feedback" loop, in which semilabeled patterns degrade the "knowledge" available to the classifier and consequently decrease classification accuracy (Jackson and Landgrebe, 2002). This risk is strongly related to the probability of including misclassified semilabeled samples in the hybrid training set, which, in turn, depends both on the reliability of the few training samples available and on the selection process of semilabeled samples. However, in the considered framework, information about the classification accuracy for each selected semilabeled sample is not available. Consequently, in order both to mitigate the possible negative effect of misclassified semilabeled samples and to reduce the probability of degrading the accuracy of the classification system, the proposed approach is defined in the context of an ensemble strategy.

The architecture of the proposed system is shown in Fig. 1, where T_{sl}^n refers to the subset of randomly selected semilabeled samples for the nth classifier, and T_o is the original small-size training set.

In the initial phase, only original labeled samples included in T_o are used to classify unlabeled patterns (k-NN$_0$ classifier). Subsequently several subsets of semilabeled samples are randomly selected from the classification maps obtained by applying the k-NN technique to the original image. In particular, two different steps are carried out:

(1) Individual classifiers to be included in the ensemble are generated using randomly selected semilabeled samples;
(2) Classifier results are combined to produce a final classification map.

These steps are described in the following:

Step 1: Selection of semilabeled samples and ensemble design.

We propose to define many different subsets of semilabeled samples T_{sl}^n $(n = 1, 2, \ldots, N-1)$ (according to a random procedure) and then to insert them in the original training set. This results in $N-1$ classifiers to be considered in the ensemble, which exploits $N-1$ different hybrid training sets together with k-NN$_0$. In greater detail, the hybrid training set of the nth classifier of the ensemble is derived as:

$$T_h^n = T_o \cup T_{sl}^n \tag{2}$$

It is worth noting that the hybrid training sets T_h^n $(n = 1, 2, \ldots, N-1)$ used for the different members of the ensemble are defined according to an iterative procedure, i.e., the subset T_{sl}^n of semilabeled samples given as input to the k-NN$_n$ is selected according to the classification results obtained from the k-NN$_{n-1}$ classifier. In other words, unlike in Shahshahani and Landgrede, 1994, Jackson and Landgrebe, 2002, at each iteration we discard the semilabeled samples included in the hybrid training set defined at the previous iteration. This is possible thanks to the specific ensemble-based architecture considered, and it results in the following two main advantages: (i) it avoids the possible accumulation of wrong information conveyed from misclassified semilabeled samples; (ii) it exploits the information present in the whole considered image.

A critical parameter to consider in defining the hybrid training sets concerns the choice of the proportions between original training patterns and semilabeled samples. On the one hand, the use of a small number of semilabeled samples may require a very large number of classifiers to be included in the ensemble to increase classification accuracy. On the other hand, if the number of semilabeled samples is

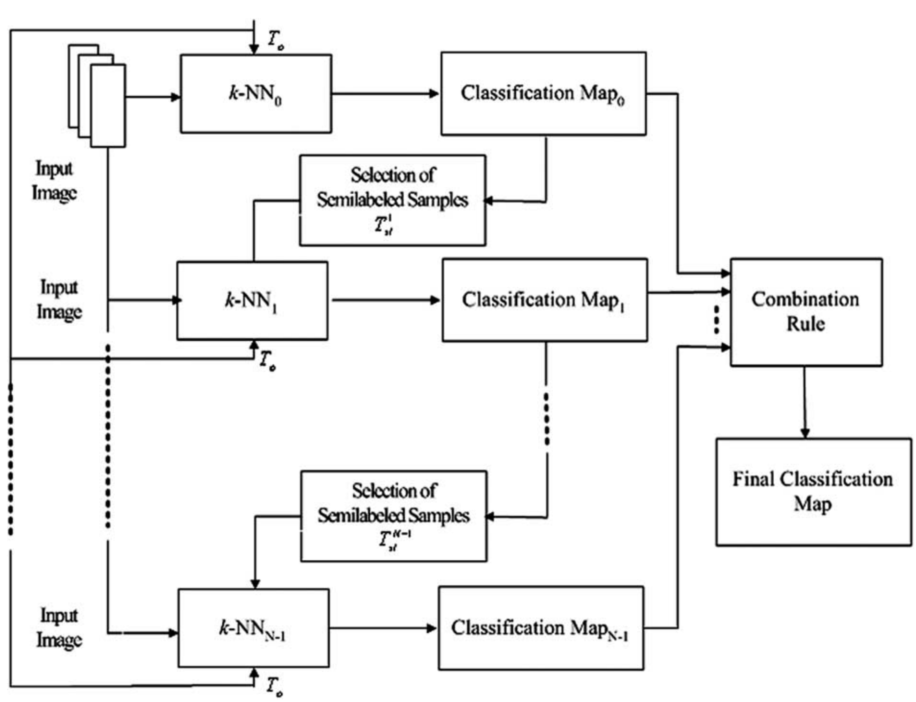

Fig. 1. Block scheme of the proposed ensemble-driven k-NN classification approach.

too large, it may give rise to a less controlled and constrained process, and consequently a degraded accuracy would be produced by the classifiers (if the error rate associated with the selected semilabeled samples is high). In the proposed approach, in order to reduce the possible negative effects involved by wrongly classified semilabeled samples, the hybrid training set of each classifier is defined considering a balanced number of semilabeled samples and original training patterns. As regards the selection of the semilabeled samples, the following conditions should be satisfied at each iteration:

(1) They should not belong to the original training set.
(2) They should have a uniform spatial distribution (in order to obtain a good spatial representation of the class distribution in the entire analyzed image).
(3) They should be highly reliable (classification labels associated with high values of the estimated local conditional posterior probability should be selected).

It is worth noting that when considering semilabeled samples, thanks to the greater number of resulting training samples, it is possible to increase the values of k in the k-NN classifiers and therefore to improve the reliability of the estimation and classification tasks. In our approach, the k value is automatically set according to the following empirical rule:

$$k = \begin{cases} m, & \text{if } m \text{ is odd} \\ m-1, & \text{if } m \text{ is even} \end{cases} \qquad m = \frac{1}{2}(n_{\max} + n_{\min})^{1/2} \tag{3}$$

where $n_{\max}$, $n_{\min}$ are the numbers of samples of the majority and minority classes, respectively.

Step 2: Combination rule.

Once the ensemble of classifiers is defined, the final classification map can be obtained according to any standard combination rule. In particular, since few training samples are available, we propose to consider only unsupervised combination strategies, such as the majority voting scheme (Bauer and Kohavi, 1999). This scheme associates the generic pattern X_i to the class ω_j if the majority of the N classifiers included in the ensemble decides for that class.

It is worth noting that the iterative random selection of semilabeled samples results in different hybrid training sets, which define significantly different classifiers. These classifiers are expected to incur in uncorrelated errors, thus satisfying a mandatory condition to obtain an effective multiple classifier system.

Another important remark concerns the relationship between the proposed method and bagging (Breiman, 1996). In our method, as in bagging, we define different bootstraps of the training set and exploit an ensemble-based architecture. However, unlike bagging, the hybrid training sets defined in the proposed approach are based on the use of semilabeled samples, which drive the definition of each member of the ensemble to reduce the small-size training problem.

3. Experimental results

In this section, the experimental results obtained by the proposed ensemble-driven approach are given. The considered data set is made up of a Landsat 5 Thematic Mapper image acquired on the Trentino area (northern Italy). For this data set, a large number of labeled samples were collected from ground-reference data. The labeled samples were divided into a training set and a test set. In order to simulate ill-posed classification problems, sub-sampling (with different rates) was applied to the training set. In greater detail, from 4549 original patterns, 10, 20, 30, 40, 50, 100 and 200 labeled samples were randomly selected (see Table 1), while maintaining as much as possible the prior class probabilities of the entire training set (with the constraint of having at least one sample for each class). In all data sets seven features and six land-cover classes were considered in the analysis. Hence, we are clearly dealing with an ill-posed complex classification problem (e.g., when the size of the training set is 10, 20, 30 and 40, the minority class has only one training pattern).

To assess the effectiveness of the proposed ensemble-driven approach, we considered test samples as unlabeled patterns, so that after classification semilabeled samples were randomly extracted from them. Three trials were carried out for each training set (with different sizes) and then the average overall accuracies were computed.

In the experiments, when the number of training samples was less than or equal to 50, k was set to 1 for the initial classifier, given the very limited number of training patterns available (only one for minority classes). In the following iterations, the k value was automatically set according to (3) (see Section 2). For the generic nth classifier included in the ensemble ($n = 1, \ldots, N-1$), the subsets T_{sl}^n of semilabeled samples were extracted from the test samples classified by k-NN$_{n-1}$ (these patterns were selected taking into account the constraints described in Section 2). A number of semilabeled samples that was double of that of the original training samples was included in each hybrid training set.

Fig. 2(a) shows the behavior of overall classification accuracy versus the number of classifiers included in the

Table 1
Distribution of training and test patterns in the seven considered simulated ill-posed classification problems

Land-cover classes	Number of test pixels	Number of training pixels						
		Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7
Conifers	1155	3	7	10	15	18	38	75
Trees	681	2	5	8	10	12	25	51
Grass	336	2	5	6	8	10	19	39
Water	84	1	1	1	1	2	3	6
Urban	104	1	1	2	2	3	5	10
Rocks	113	1	1	3	4	5	10	19
Overall	2473	10	20	30	40	50	100	200

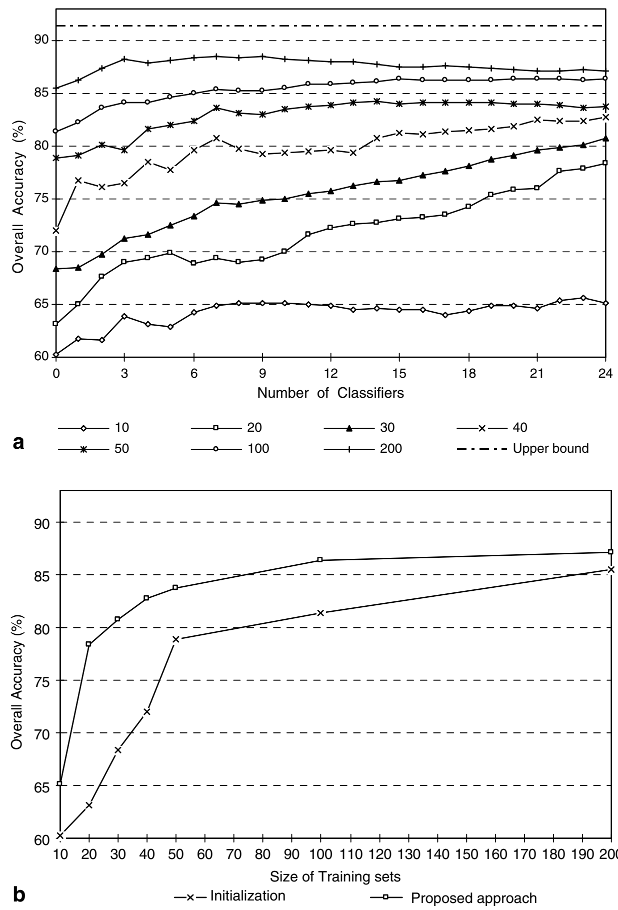

Fig. 2. (a) Overall classification accuracy (average on three trials) versus the number of classifiers included in the ensemble for different considered simulated data sets (the upper bound refers to overall accuracy with the original training set made up of 4549 samples) and (b) comparisons between initial accuracy (obtained by k-NN$_0$ on the initial training set) and the accuracy obtained by the proposed approach using 25 classifiers versus the number of initial training samples.

ensemble for the different simulated datasets. For comparison purposes, the upper bound of the accuracy on the test set (obtained using all 4549 training patterns originally available) is also given. On analyzing the diagram, one can observe that in general the proposed ensemble-driven approach significantly increased overall classification accuracy. If we consider the best case (i.e., set 2 defined by 20 training patterns), overall accuracy obtained on the test set using only the initial labeled samples was 63.12%, while the accuracy provided by the ensemble with 25 k-NN classifiers was 78.34% (accuracy sharply increased by 15.22%). More in general, on analyzing Fig. 2(b) (which reports overall accuracy versus the number of initial training samples for the standard k-NN$_0$ classifier and the proposed approach), we can observe a different behavior of the proposed approach for training sets of a different size. In greater detail, in all simulated data sets the presented method increased classification accuracy. However, this increase is not significant in the two extreme cases (i.e., set 1 and set 7 which are respectively made up of 10 and 200 original training samples). For set 1 (which led to an increase in accuracy of 4.80%), this depends on the fact that the number of original training samples is too small to let the ensemble capture the complexity of the classification problem. In set 7 (which produced an increase in accuracy of 1.55%), the reason is that 200 training samples are already sufficient for a reasonably good modeling of the classification problem. For all other cases, a significant improvement in the classification result was achieved with the proposed approach, which increased overall accuracy over a range between 5.9% and 15.22%. It is worth noting that the accuracy provided by the ensemble increased also when a significant amount of misclassified semilabeled samples were included in the hybrid training sets. This was possible thanks: (i) to the defined selection strategy (which avoids accumulating wrong information conveyed from misclassified semilabeled samples); and (ii) to the diversity of the classifiers achieved by the adopted random bootstrapping process.

Two more interesting observations can be made from an analysis of Fig. 2(a): (i) as expected, on increasing the number of initial training samples, the number of classifiers to be included in the ensemble to obtain the convergence of overall accuracy decreases (from more than 30 classifiers with 20 training patterns to three classification algorithms with 200 samples); and (ii) different numbers of initial training samples result in different overall accuracy values at convergence. Concerning the former, in general the definition of the number of classifiers needed to reach convergence is a complex problem, which depends on the initial classification accuracy, the size of the initial training set and the number of randomly selected semilabeled samples included in each hybrid training set. (It is worth noting that the number of selected semilabeled samples affects not only the convergence rate, but also the accuracy and stability of the method).[2] As regards the latter, it shows that the proposed approach cannot recover information that is not present in the original training set; it can only mitigate the effect of the small number of training samples by exploiting the structural regularized distributions of classes (and of the related patterns) in the feature space.

On comparing the results obtained in this paper with those yielded by the MLPNN based architecture presented in (Chi and Bruzzone, 2005), interesting conclusions can be drawn on the effectiveness of the proposed ensemble-driven approach. In particular, in spite of the fact that the MLPNN and k-NN classification techniques are based on significantly different principles (the learning of the MLPNN classifier is based on the minimization of a global error function that depends on all the training samples,

[2] On the one hand, if very few semilabeled samples are used, the approach requires the definition of ensembles made up of a very large number of classifiers for reaching convergence (this affects the complexity of the system design phase). On the other hand, if a large number of semilabeled samples are used, the risk of defining a negative feedback mechanism increases significantly (the stability of the method decreases).

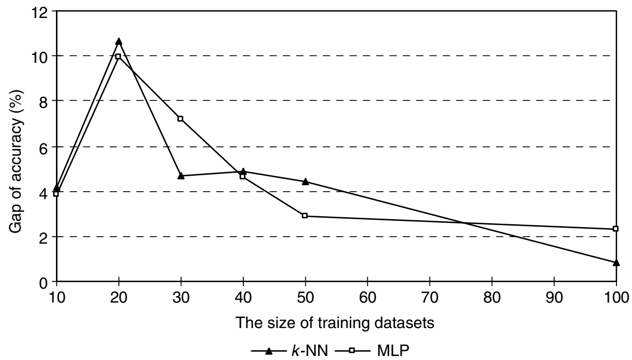

Fig. 3. Gap of accuracy between the standard supervised classifiers and the proposed k-NN approach (presented in this paper) and the MLPNN approach (presented in (Chi and Bruzzone, 2005)) versus the size of original training sets.

while the k-NN classifier accomplishes classification according to a local analysis of a neighborhood of the pattern to be analyzed in the feature space) and in spite also of the fact that they provide different classification accuracies, the relative increase in accuracy obtained with the proposed classification architecture is very similar (see Fig. 3). In greater detail, as expected, even if the overall accuracy of the k-NN was inferior to that of the MLPNN (see Chi and Bruzzone, 2005) due to the very challenging and critical ill-posed problem addressed (which is very complex for the local-based estimation of the k-NN algorithm), the increase in accuracy obtained with the proposed ensemble-driven k-NN method was very similar to that obtained with the ensemble-driven MLPNN technique. This points out the validity of the general architecture of the method, and confirms its robustness also in presence of classifiers intrinsically unsuitable to address ill-posed classification problems.

It is worth noting that the use of a very small number of unrepresentative training patterns may also establish a "negative feedback" loop, which may degrade the accuracy of the proposed classification approach. However, we expect such a situation to correspond to very poor initial training sets, which cannot be used for any kind of reliable supervised or semisupervised classification procedure (in other words, it seems that in these cases the classification problem cannot be solved with any supervised or semisupervised nonparametric approach).

4. Discussion and conclusions

In is paper, an ensemble-driven approach to the classification of remote-sensing data in ill-posed problems has been proposed, which extends and reinforces the work presented in (Chi and Bruzzone, 2005). In particular, an architecture made up of an ensemble of classifiers has been presented. In this architecture classification algorithms are developed in the context of the k-NN technique. Each member of the ensemble is derived by exploiting the hybrid training sets made up of a balanced number of labeled and semilabeled training patterns. In greater detail, after an initial classification using the original training set alone, a subset of semilabeled samples randomly selected from the classification map is exploited to bootstrap a new hybrid training set for the next classifier to be inserted in the ensemble. This process is iterated until the desired number of classifiers is included in the pool of classification algorithms. Then, the final classification map is achieved by combining the classification results provided by the members of the ensemble according to an unsupervised majority voting strategy. It is worth stressing that by exploiting semilabeled samples, the proposed approach drives the definition of the ensemble with an iterative inductive process applied to initial training samples. Experimental results, obtained on multispectral remote-sensing data (in the context of simulated ill-posed classification problems) confirmed the effectiveness of the proposed approach. In particular, this approach sharply increased both the accuracy and the stability of obtained classification results. These results are consistent with those obtained on the same data set with a completely different MLPNN based method (see Chi and Bruzzone, 2005). This confirms the validity of the proposed approach even further.

It is worth noting that unlike other ensemble methods based on the re-sampling of training data distributions, the proposed approach can be used with any parametric and nonparametric classification technique without requiring any specific constraint on the model of training data distributions.

As a final remark, it is important to point out that the performance of the proposed technique strongly depends on the set of initial training samples available. If the training samples are noisy and not representative of the true data distribution, the proposed technique may lead to unsatisfactory results. However, this condition is intrinsically unavoidable when ill-posed classification problems are considered. In order to better investigate this problem, we plan to carry out an intensive experimental validation as future developments of this work by simulating many different initial training sets for each size and statistically analyzing the probability to establish positive and negative feedbacks. In addition, special attention will be devoted both to the selection of the most "reliable" classifiers (based on semilabeled samples) to be included in the ensemble and to the use of different nonparametric techniques (e.g. support vector machines (Vapnik, 1999)) in the context of the proposed architecture.

References

Baraldi, A., Bruzzone, L., Blonda, P., 2005. Quality assessment of classification and cluster maps without ground truth knowledge. IEEE Trans. Geosci. Remote Sensing 43 (4), 857–873.

Bauer, E., Kohavi, R., 1999. An empirical comparison of voting classification algorithms: bagging, boosting, and variants. J. Mach. Learning 36 (1/2), 105–139.

Bennett, K.P., Demiriz, A., 1998. Semisupervised support vector machines. In: Kearns, M.S., Solla, S.A., Cohn, D.A. (Eds.), Advances in Neural Information Processing Systems, vol. 10. MIT Press, pp. 368–374.

Blum, A., Mitchell, T., 1998. Combining labeled and unlabeled data with co-training. In: Proceedings of the 11th Annual Conference on Computational Learning Theory, pp. 92–100.

Breiman, L., 1996. Bagging predictors. Mach. Learning 26 (2), 123–140.

Chi, M., Bruzzone, L., 2005. A semilabeled-sample-driven bagging technique for ill-posed classification problems. IEEE Geosci. Remote Sensing Lett. 2 (1), 69–73.

Cohen, I., Sebe, N., Cozman, F.G., Cirelo, M.C., Huang, T.S., 2003. Learning Bayesian network classifiers for facial expression recognition both labeled and unlabeled data. In: IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, vol. 1, pp. 595–601.

Dundar, M.M., Landgrebe, D.A., 2004. A cost-effective semisupervised classifier approach with kernels. IEEE Trans. Geosci. Remote Sensing 42 (1), 264–270.

Fardanesh, M.T., Ersoy, O., 1998. Classification accuracy improvement of neural network classifiers by using unlabeled data. IEEE Trans. Geosci. Remote Sensing 36 (3), 1020–1025.

Fung, G., Mangasarian, O.L., 1999. Semisupervised support vector machines for unlabeled data classification. Data Mining Institute Technical Report 99-05.

Hughes, G.F., 1968. On the mean accuracy of statistical pattern recognition. IEEE Trans. Inform. Theory IT-14, 55–63.

Jackson, Q., Landgrebe, D.A., 2002. Adaptive Bayesian contextual classification based on markov random fields. IEEE Trans. Geosci. Remote Sensing 40 (11), 2454–2463.

Joachims, T., 1999. Tranductive inference for text classification using support vector machines. In: Proceedings of the International Conference on Machine Learning (ICML).

Kemp, C.C., Griffiths, T.L., Stromsten, S., Tenenbaum, J.B., 2003. Semisupervised learning with trees. Adv. Neural Inform. Process. Systems 16.

Moon, T.K., 1996. The expectation–maximization algorithm. IEEE Trans. Signal Process. 13 (6), 47–66.

Shahshahani, B.M., Landgrede, D.A., 1994. The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon. IEEE Trans. Geosci. Remote Sensing 32 (5), 1087–1092.

Tadjudin, S., Landgrebe, D.A., 1996. A decision tree classifier design for high-dimensional data with limited training samples. In: Geosci. Remote Sensing Symposium (IGARSS '96), Remote Sensing, International, vol. 1, pp.790–792.

Vapnik, V.N., 1999. The Nature of Statistical Learning Theory, third ed. Springer-Verlag, Berlin.