2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Classification with rejection based on various SVM techniques

Wladyslaw Homenda*†, Marcin Luckner* and Witold Pedrycz‡§
*Faculty of Mathematics and Information Science, Warsaw University Technology, ul. Koszykowa 75, 00–662 Warsaw, Poland
†Faculty of Mathematics and Computer Science, University of Bialystok, ul. Sosnowa 64, 15-887 Bialystok, Poland
‡Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01–447 Warsaw, Poland
§Department of Electrical & Computer Engineering, University of Alberta, Edmonton T6R 2G7 AB Canada
e–mails: homenda@mini.pw.edu.pl, mluckner@mini.pw.edu.pl and wpedrycz@ualberta.ca

***Abstract*—The task of identifying *native* and *foreign* elements and rejecting *foreign* ones in the pattern recognition problem is discussed in this paper. Such the task is a nonstandard aspect of pattern recognition, which is rarely present in research. In this paper, ensembles of support vector machines solving two–classes and one–class problems are employed as classification tools and as basic tools for rejecting of foreign elements. Evaluation of quality of classification and rejection methods are proposed in the paper and finally some experiments are performed in order to illustrate acquainted terms and methods.**

I. Introduction

Rejection is an aspect of pattern recognition shallowly researched and rarely considered in practice. In standard attempt to pattern recognition, an object is classified to one of given classes. The set of classes is either fixed a priori (supervised problem), or is determined at the stage of a recognizer construction (unsupervised problem). In these cases, it is assumed that all classified elements belong to a fixed set of classes. However, this assumption is often insufficient in practice. It is proved in important practical applications that processed are not only elements of the fixed set of classes, but also elements not belonging to these classes. Let us use the names *native elements* for elements of given classes and *foreign elements* for ones not belonging to any given class. Example applications of rejecting foreign elements are: recognizing printed texts, manuscripts, music notations, biometric features, voice, speaker, recorded music, medical signals, images, etc, c.f. [1], [10].

The dissemination of technologies using pattern recognition increases the importance of identifying foreign elements. For example: in recognition of printed texts, foreign elements (blots, grease, or damaged symbols) appear in a negligible scale due to regular placement of printed texts' elements (letters, numbers, punctuation marks) and due to their good separability. These features of printed texts allow employing effective segmentation methods and filtering foreign elements. However, in recognition of such sources as recorded voice, biometric features, medical images, geodetic maps or music notation etc., problem of foreign elements is more important. Unlike printed text, such sources as, for instance, geodetic maps or music scores, contain elements placed irregularly and overlapping native elements. Such elements are hardly distinguishable by size and shape analysis.

Due to weak separability of foreign and native elements of recognized sources, segmentation criteria must be more tolerant than in the case of printed texts in order not to reject native elements at the stage of segmentation. In consequence, more foreign elements are subjected to stages of recognition following segmentation and should be eliminated then. The problem of analysis of foreign elements is highly important in such domains as analysis of medical signals and images, recognition of geodesic maps or music notations (printed and handwritten) and its importance will be increasing in future, c.f. [4].

The problem of rejecting foreign elements in recognition tasks is not present frequently in research and is rather rarely present in papers on pattern recognition. Assuming that classified elements are always native ones, i.e. they belong to one of recognized classes, and ignoring rejection of foreign elements is rather a standard attempt. Alike, papers describing practical applications of pattern recognition methods ignore the problem of foreign elements, what may come from insufficient theoretical research of this subject and limited abilities of existing rejection methods. There are significant exceptions, which show that the rejection problem cannot be disregarded, c.f. [5], [8].

The motivation of this study arises from discussion on classification with rejection option. As outlined above, up–to–date research and practice still need conducting further studies on new aspects in the domain of pattern recognition. It is expected that research in this area will overcome technological barriers and will increase effectiveness in areas mentioned above.

The paper is structured as follows. Related research and introductory remarks are presented in section II. In section III, various ensembles of classifications based on binary classifiers are shown. The discussion includes evaluation criteria of rejection and reclassification methods. Conclusions and directions of further research are presented in section IV.

II. Preliminaries

A. Support Vector Machine

The experiment described in this paper is a classification with Support Vector Machine (SVM). Here we give brief description of SVM necessary to introduce discussed concept of rejection. Technical details of SVMs can be found in [9].

Classification of elements is usually done on vectors of features (observed, measured etc.) rather than on elements themselves. In order to simplify description we will talk about features representing elements rather then about elements themselves. Let us recall that SVM can directly separate two classes of elements, which are linearly separable. However, using transformation of features into a highly–dimensional space, it is possible to separate linearly in ultimate space classes, which are linearly non separable in the space of features.

1) Two–classes SVM: Let us assume that there are two classes labeled -1 and +1. Hence, the space of features is included in R^d. Assume that vectors of features of training elements $x_i \in R^d$ are labeled as y_i, where y_i is either -1 or +1, for $i = 1, 2, \ldots, N$ and for N being the cardinality of the learning set. For a possibility of a non linear separation of the two classes, the space R^d is mapped into a space of higher dimension using so called kernel function $K(x, x')$. Then a SVM decision function is implemented as:

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{N} y_i * \alpha_i * K(x, x_i) + b\right) \tag{1}$$

where coefficients α_i and b are computed by maximization of the following convex quadratic programming (QP):

$$\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i * \alpha_j * y_i * y_j * K(x_j, x_i) \tag{2}$$

with the following constrains:

$$\bigwedge_{i \in \{1,2,\ldots,N\}} 0 \leqslant \alpha_i \leqslant C \ \wedge \ \sum_{j=1}^{N} \alpha_i * y_i = 0 \tag{3}$$

The regularization coefficient C in Equation 3 controls trade–off between margins and misclassification errors. In this work the regularization coefficient C is set to 1. As the kernel function $K(x, x')$ the Gaussian kernel is taken

$$\exp\left(-\frac{1}{d}||x - x'||^2\right), \tag{4}$$

where d is the number of features.

2) One–class SVM: In this study we employ one–class SVM classifier, which separates one class data from thee origin. Below, we give brief description of this class of SVMs while details ate presented in [7] Assume that $x_i \in R^d$ are vectors of features of training elements, for $i = 1, 2, \ldots, N$ and for N being the cardinality of the learning set. Assume also $K(x, x')$ is a kernel function. Then a one–class SVM decision function is implemented as:

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{N} \alpha_i * K(x, x_i) - \rho\right) \tag{5}$$

where α_i and ρ are obtained by maximization of the following convex quadratic programming (QP) problem:

$$\frac{1}{2}\Big(\sum_{i=1}^{N} \alpha_i * K(x, x_i)\Big)^2 - \rho - \sum_{i=1}^{N} \alpha_i \bigg(\Big(\sum_{i=1}^{N} \alpha_i * K(x, x_i)\Big) - \rho\bigg) \tag{6}$$

with constrains:

$$\bigwedge_{i \in \{1,2,\ldots,N\}} 0 \leqslant \alpha_i \leqslant \frac{1}{\nu N} \ \wedge \ \sum_{j=1}^{N} \alpha_i = 1 \tag{7}$$

where ν is an equivalent of the regularization coefficient C in Equation 3.

The classifiers were implemented using the libsvm library for Matlab [2].

B. Evaluation

Quality evaluation of classification with rejection requires non standard measures. Intuitively, it is important to measure how exact rejection procedure, i.e. how many foreign elements are classified as native ones and oppositely, how many native elements are rejected. Of course, measuring classification quality of native elements to proper classes is still of great importance.

For better understanding how quality of classification with rejection should be measured we adopt parameters and quality measures used in signal detection theory. Since these parameters are widely utilized, we do not refer to original sources, but of course do not claim to letting these factors on. The following confusion matrix a two classes problem is adopted in evaluating quality of classification with rejection in Table I. Parameters exposed in this Table were then used in defining several factors, which outline classification quality.

TABLE I. CONFUSION MATRIX FOR *classification with rejection* PROBLEM

	Classification as native	Classification as foreign
Native elements	True Positives (TP)	False Negatives (FN)
Foreign elements	False Positives (FP)	True Negatives (TN)

The parameters in Table I are numbers of elements of a testing set. They have the following meaning:

- TP – the number of native elements classified as native elements (no matter, if classified to correct class, or not),
- FN – the number of native elements incorrectly classified as foreign ones,
- FP – the number of foreign elements incorrectly classified as native ones,
- TN – the number of foreign elements correctly classified as foreign ones.

To describe quality of classification with rejection three sets of complimentary factors are used.

Classifier's overall performance for both categories of elements (native and foreign ones) is measured by the following two factors. These factors measure effectiveness of correctly classified native elements together with correct rejected foreign elements:

$$\begin{aligned} Accuracy &= \frac{TP + TN}{TP + FN + FP + TN} \\ Error &= \frac{FP + FN}{TP + FN + FP + TN} \end{aligned} \tag{8}$$

The next two factors evaluate influence of one category of elements classes at classification of native elements or rejection of foreign elements. Here are the factors for native elements:

$$Native\ Precision = \frac{TP}{TP+FP} \tag{9}$$
$$Native\ False\ Discovery\ Rate = \frac{FP}{TP+FP}$$

and the factors for foreign elements:

$$Foreign\ Precision = \frac{TN}{TN+FN} \tag{10}$$
$$Foreign\ False\ Discovery\ Rate = \frac{FN}{TN+FN}$$

The factors measuring effectiveness of classification inside one category of elements, i.e. in the category of native elements and in the category of foreign elements. In this case taken are proportions of correctly classified and incorrectly rejected native elements to all native ones:

$$Native\ Sensitivity = \frac{TP}{TP+FN} \tag{11}$$
$$Native\ Miss\ Rate = \frac{FN}{TP+FN}$$

and proportions of correctly rejected foreign elements and incorrectly classified foreign elements to all foreign ones:

$$Foreign\ Sensitivity = \frac{TN}{TN+FP} \tag{12}$$
$$Foreign\ Miss\ Rate = \frac{FP}{TN+FP}$$

Finally, the so called F–measure defines a balance between the precision and the sensitivity:

$$\text{F–measure} = 2 \cdot \frac{Precision \cdot Sensitivity}{Precision + Sensitivity}. \tag{13}$$

The above factors are pairs of complimentary factors. In order to increase quality of classification, Accuracy, Sensitivity and Precision should be maximized or, equivalently, Error, Miss Rate and False Discovery Rate should be minimized. Hence, there is no need to analyze all factors. It is sufficient to focus on one type of them, i.e. either on the ones to be maximized or on the ones to be minimized.

III. EXPERIMENT

In this study we test several architectures of binary classifiers and one–class classifiers on synthetic data. Support vector machines were used as classifiers. Dataset includes native elements split between two classes and foreign elements structured as described later in this section.

A. Dataset

The testing environment is defined in the bipolar unit interval $[-1, 1] \times [-1, 1]$. Two classes of native elements were generated using Gaussian distribution, 600 elements in each class. Centers (mean values) of these classes were located in points $(0.1, 0.1)$ and $(-0.3, -0.3)$ while standard deviations was set to 0.15 for each class. Therefore the classes overlap, as it can be seen in Figure 1.

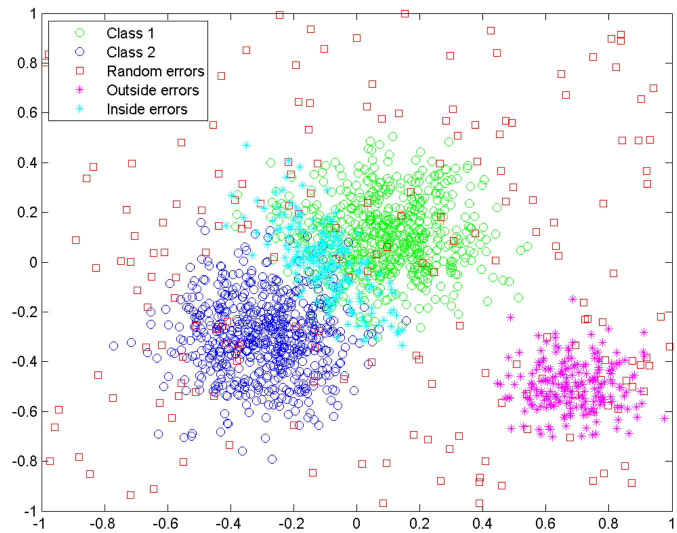

Fig. 1. Testing environment

Three different groups of foreign elements were added, 200 elements in each group. The first group includes uniformly distributed ones in the whole testing area. Therefore, foreign elements laid on both classes of native ones and outside them.

The second group contains 200 foreign elements located outside of both classes of native elements. This elements were generated with Gaussian distribution centered in the point $(0.6, -0.6)$ and with standard deviation equal to 0.1. As it can be seen in Figure 1, this group of elements is separated from both classes of native elements classes.

The third group includes 200 foreign elements located *between* both classes of native elements. This group was generated using a modified Gaussian distribution. The x coordinate was generated using the Gaussian distribution with the mean value equal to 0 and with the standard deviation equals to 0.1. Then, the y coordinate was generated as equal to $-x$ increased by a random value (uniformly distributed) from the interval $[0, 0.1]$. Finally, the x coordinate was decreased by 0.1. As the result, this group of foreign elements lays between both classes of native elements, c.f. Figure 1.

Test were done in two series. In the first one the whole set of elements was used as the learning set while testing set included randomly chosen 50% elements of every group of elements (i.e. 300 elements in each class of native elements and 100 ones in every group of foreign ones). Hence, the testing set is the subset of the learning set. In the second testing series, the whole set of elements was split to two equal parts. Therefore, learning and testing sets were disjoint.

B. Rejection methods

1) Global SVM rejection: The global rejection schema is shown in Figure 2. The dataset is split into two subsets of native and foreign elements first. Then, in the second step, the set of native elements is split up into proper classes.

This method needs the learning set to contain native elements split up among classes and foreign elements.

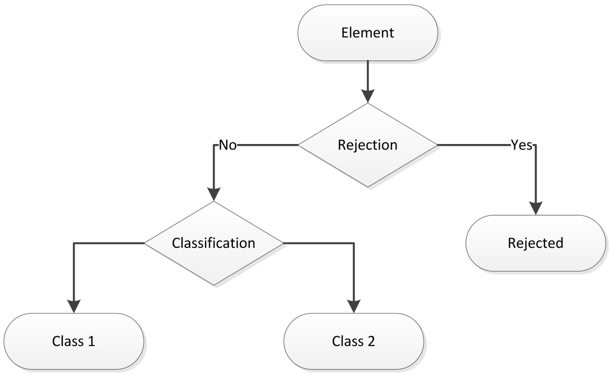

Fig. 2. Global rejection

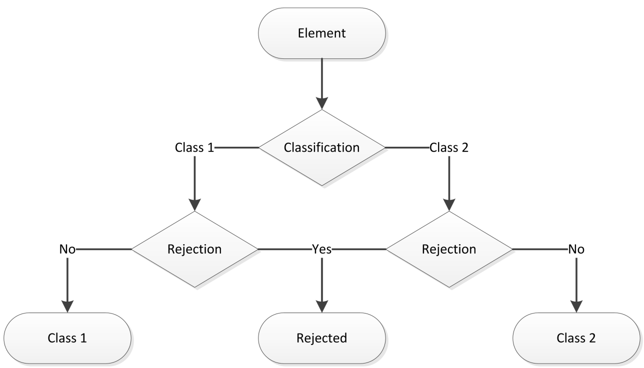

Fig. 3. Local rejection

2) Local SVM rejection: In Figure 3 the local rejection schema is shown. In the first step of classification, the dataset is split up among both classes of native elements. This means that all foreign elements fall into classes of native ones. In the next step of classification, foreign elements are separated from native ones.

This method needs the learning set to contain native elements divided among classes and foreign ones.

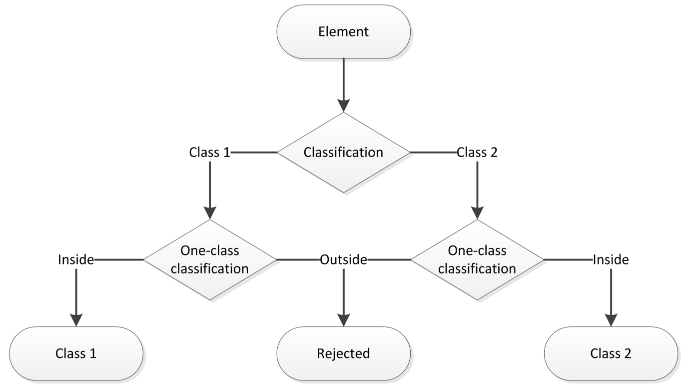

Fig. 5. One–class rejection

3) One–class SVM rejection: In Figure 5 the one–class rejection schema is shown. In the first step of classification, the dataset is split up among both classes of native elements. This means that all foreign elements fall into classes of native ones. In the next step of classification, one–class SVM is used to separate elements of the class from the rest of elements. This step is done for each class separately.

This method needs the learning set to contain native elements divided among classes to train the SVM classifier as well as the one–class SVM classifier. However, unlike the global and the local rejection this rejection method does not need the knowledge on distribution of foreign elements.

4) Distance rejection: The classifier was modified in order to produce probabilistic outputs. The method proposed in [6] was applied. Using a distance $f(x)$ between an element and decision boundary, a probability of correct classification was computed according to the formula:

$$p(x) = \frac{1}{1 + \exp(f(x))}, \tag{14}$$

The classification decision taken at the low probability level can be rejected as not credible. The low probability level

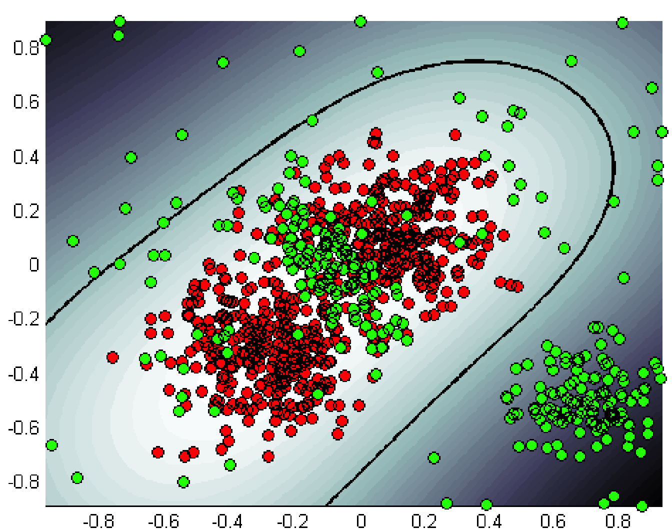

Fig. 4. Results of global rejection

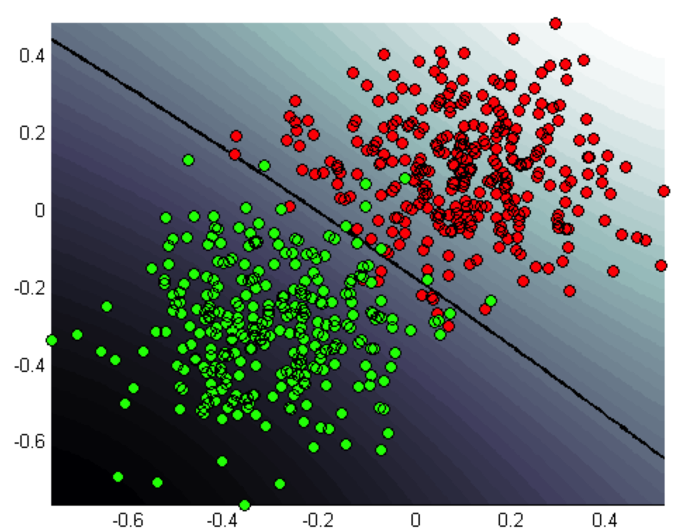

Fig. 6. Classifier for distance rejection

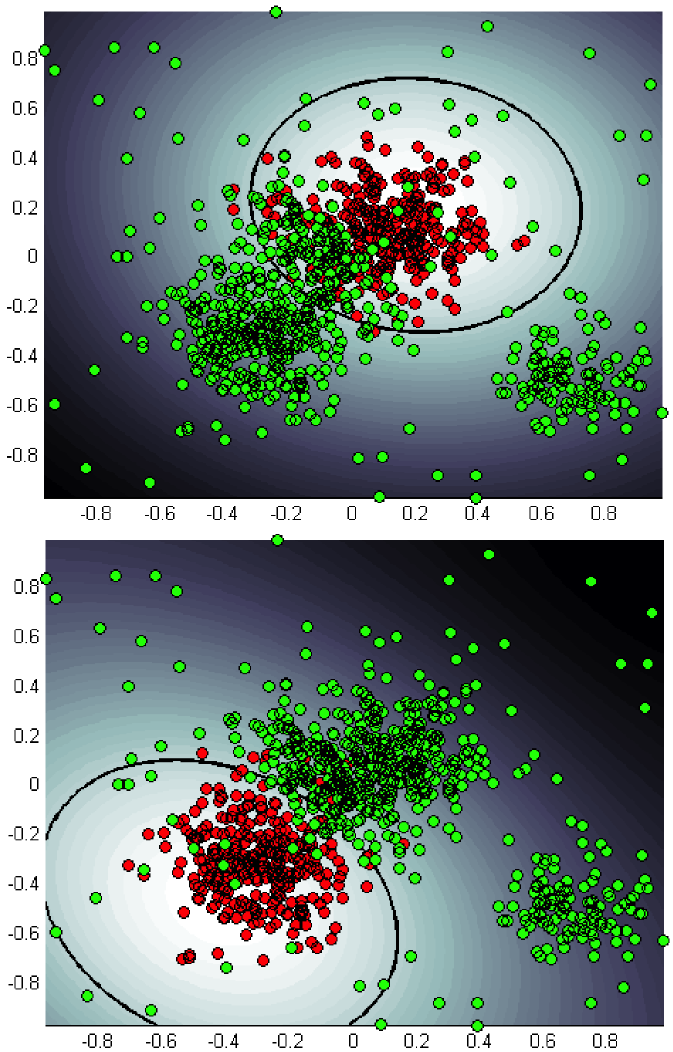

Fig. 7. Results of local rejection

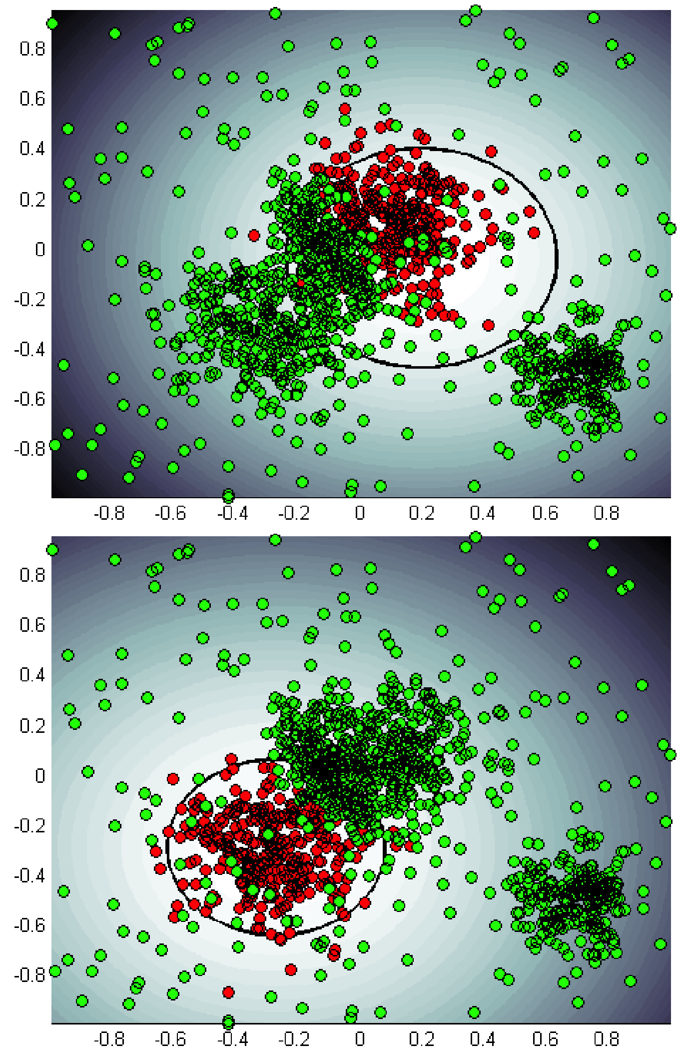

Fig. 8. Results of one–class rejection

is defined by a threshold. The threshold can be obtained by calculation of the reject trade–off for the learning sets [3].

This method needs the learning set to contain native elements divided among classes and foreign ones.

C. Results

Tests were performed in two series:

- the testing set was taken as a subset of the learning set (the half of the learning set was taken under consideration). This approach was aimed to outline differences between analyzed methods of classification with rejection without an influence of learning shortcomings,
- separated learning and testing sets (the whole dataset was split into two equal subsets) were used to simulate more realistic conditions.

In each series five models have been tested:

- classification without rejection. i.e. all elements fall into one of two classes of native elements,
- classification with a one–class SVM to reject all foreign elements classified as members of the class, but laying outside of the area occupied by the given class of native elements, c.f. Figure 5,
- classification in two stages, c.f. Figure 2. In the first stage a SVM classifier separates native elements and foreign ones. In the second stage, elements already classified as native, are assigned to proper classes,
- classification in two stages, c.f. Figure 3. Firstly, all elements, native and foreign, were distributed among both classes of native elements. Then, foreign elements were rejected for each class of native elements separately,
- all native and foreign elements classified to both classes of native ones. Then, foreign elements rejected based on distance to boundary of decision regions.

All models were tested on four testing datasets. Each dataset included two classes of native elements and selected foreign elements:

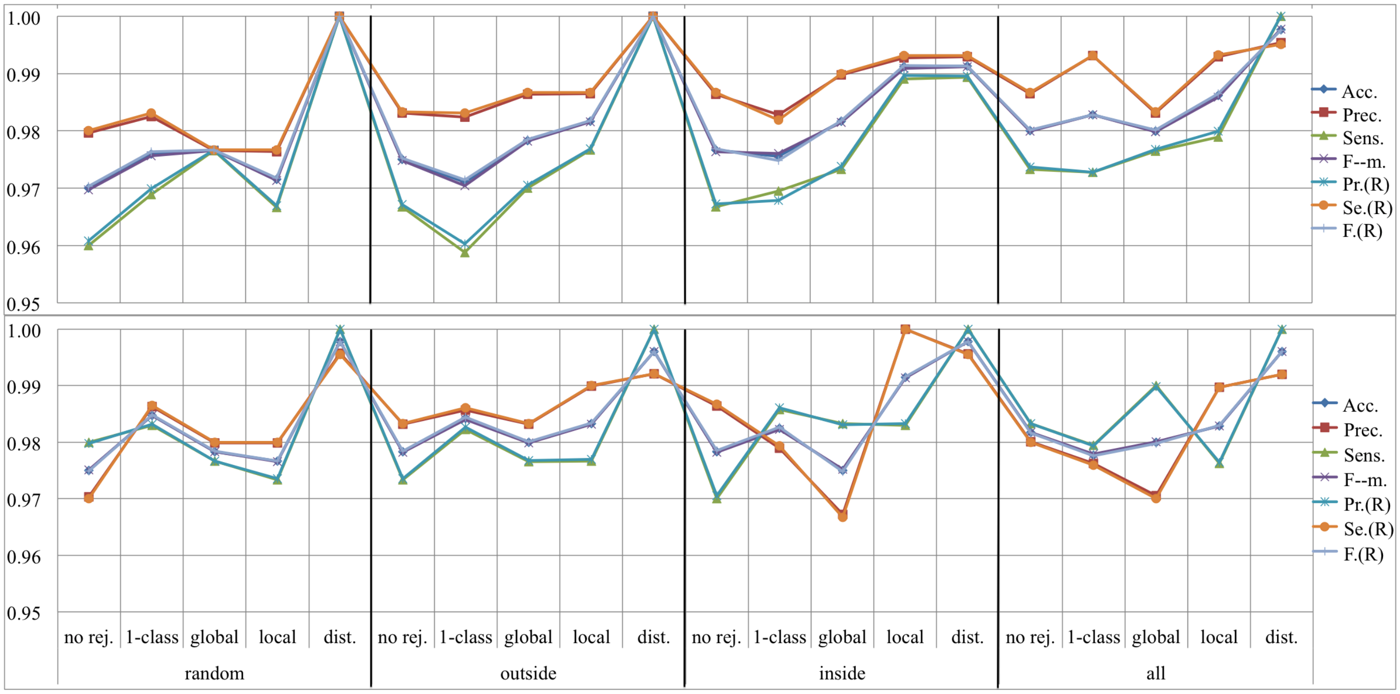

Fig. 9. Classification, no foreign elements, models tested on the learning set (upper chart) on separated training and testing sets, separation ratio: 0.50 (bottom chart)

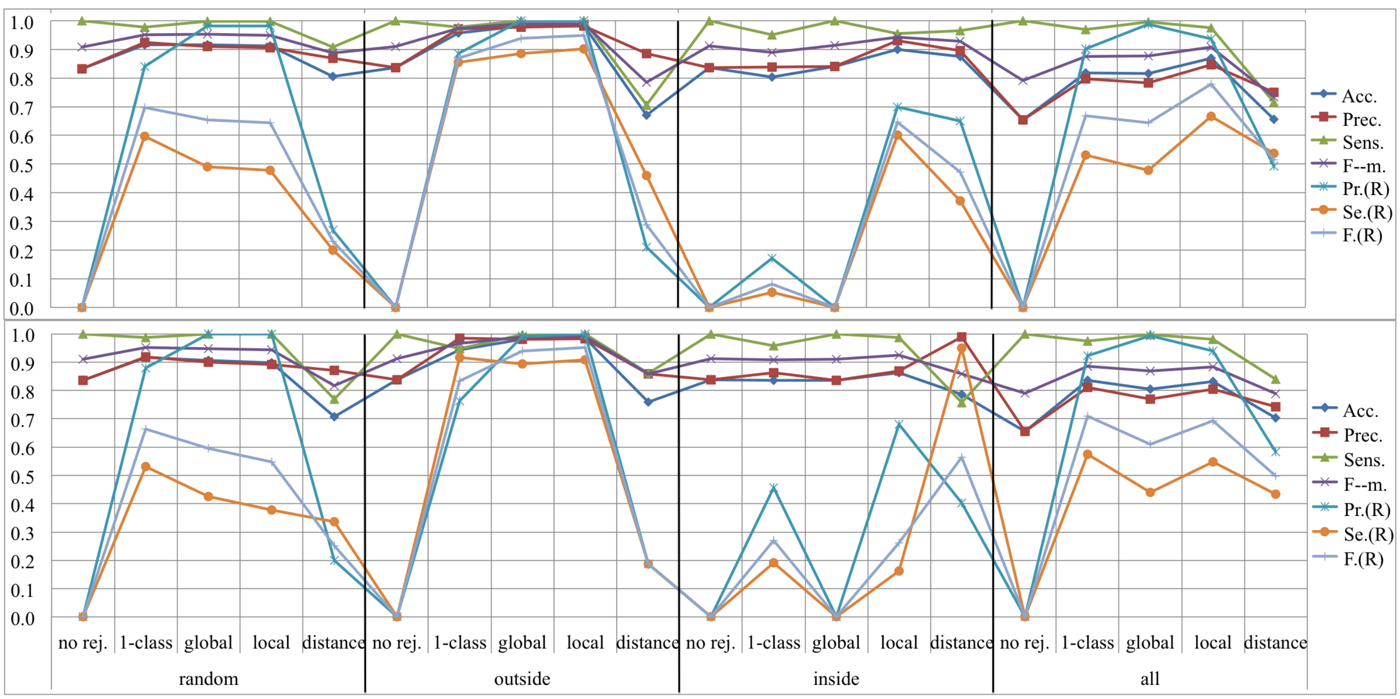

Fig. 10. Classification with rejection, models tested on the learning set (upper chart) and on the learning and testing sets, separation ratio 0.50 (bottom chart)

- foreign elements generated randomly with uniform distribution were added to to native ones,
- outside foreign elements added to to native ones,
- in–between foreign elements added to to native ones,
- all foreign elements added to native ones.

In all tests the factors TP, FN, FP and TN (c.f. Table I for factors' definition) were counted. The factors were used to calculate measures such as the accuracy, the precision, the recall and the F–measure.

In Figure 9 tests done on both classes of native elements are illustrated. Foreign elements were not considered. Figure 10 illustrates results of tests done on all elements: native and

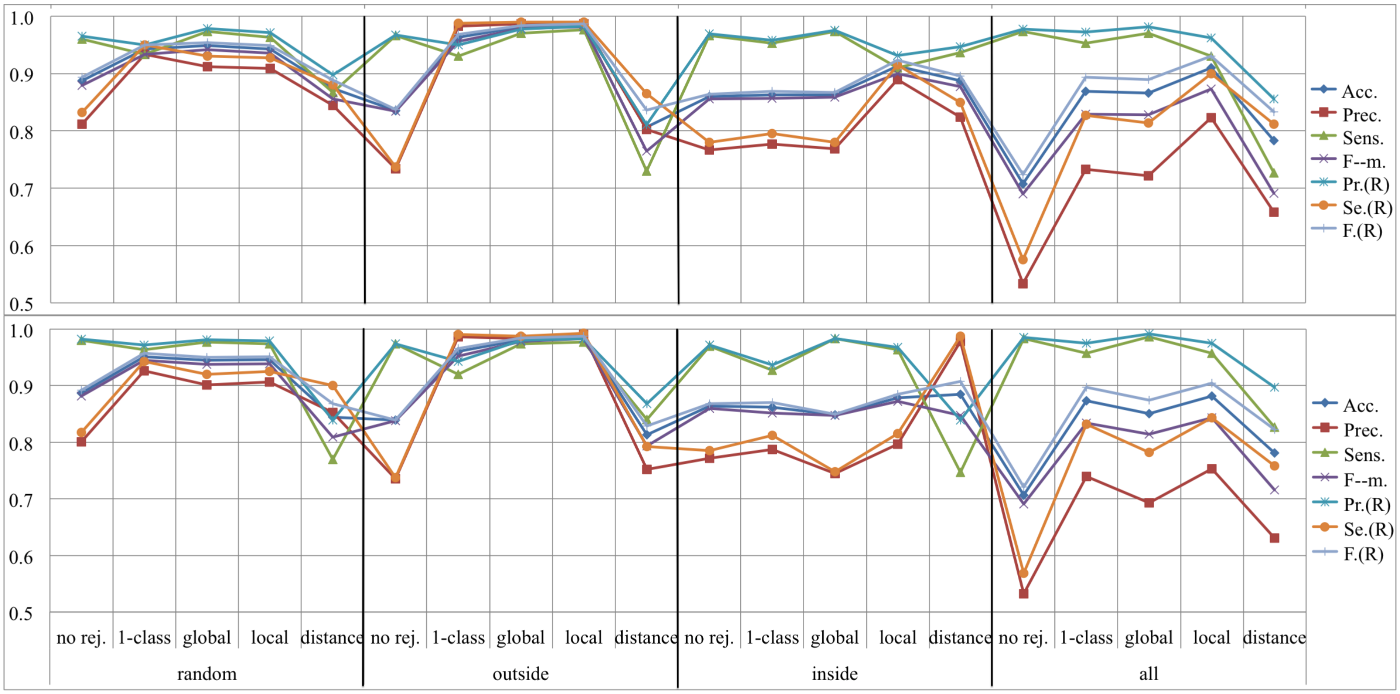

Fig. 11. C1 versus all others, without rejection, models tested on the learning set (upper chart) and testing set with separation ration 0.50 (bottom chart)

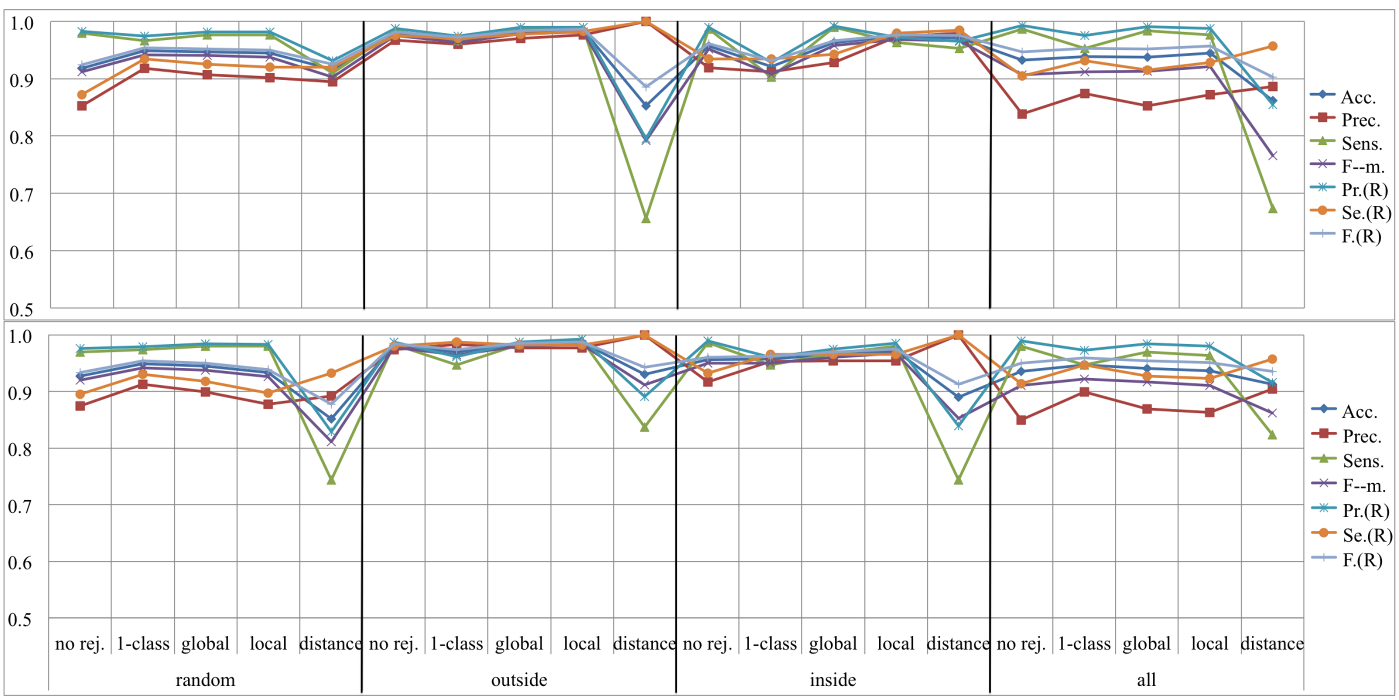

Fig. 12. C2 versus all others, without rejection, models tested on the learning set (upper chart) and testing set with separation ration 0.50 (bottom chart)

foreign ones. Figures 11 and 12 show results of tests performed on configurations native elements of one class contra all other elements (native and foreign). Upper chart of every Figure exposes results for configuration with the testing set included in the learning set while bottom chart depicts results for configuration with separated learning and testing set (c.f. notes on learning and testing sets at the beginning of this section).

D. Discussion

Let us recall that the number of native elements may be reduced by rejection method, c.f. [4]. In tests described here, the number of native elements is equal to 600, they are grouped in two classes. This number stay nearly unchanged in cases of global and local rejections. The rejection based on one–class SVM reduces the number of the native elements by nearly 5 percent. However, the rejection based on SVM distance rejects over 20 percent of the native elements.

Considering classification without foreign elements, all tests reported here show that classification results do not differ significantly between both configuration of datasets, i.e. the configurations with the learning set including the testing one and with the learning and testing sets disjoint. Obtained statistics are from the range 97–100% for the first configuration and 96–100% for the second one, c.f. Figure 9. The best classification results were obtained for rejection based on SVM distance, because all misclassified elements that lied between classes were rejected.

The influence of a learning set is noticeable for the results of the classification with rejection. The difference in the accuracy for the models tested on the testing set included in learning one and on the testing sets separated from the learning one is the biggest for rejection based on SVM distance. The difference is nearly 10% For each group of foreign elements and it is around 5% for all groups of foreign elements collected together. Other statistics show also significant differences. Additionally, big variability can be observed for rejection based on the one–class SVM and the on the local rejection tested on the group of foreign elements placed between classes of native elements.

Most of rejection methods give better results than methods without rejection. Rejection based on SVM distance is the only exception. It brings worse results even for foreign elements placed between both classes of native elements. There is also a problem with global rejection used for foreign elements placed inside both classes of native elements. In such the case, rejection method does not work.

The accuracy is well correlated with the f–measure. Therefore, A rejection method can be evaluated explicitly. Local rejection obtains the best results. The good results are also obtained by the rejection based on one–class SVM and the global rejection, but only for the random and outside foreign elements.

The last series of tests analyzes differences between results for both native classes. When the classification without rejection is used, the difference is equal to 14% in the accuracy calculated for each class separately (elements of the class versus the rest of elements) and for the group of outside foreign elements. This difference is equal to 10% for the inside group of foreign elements. However, for all the rejection method, except the method based on SVM distance, the accuracy for both classes is the same for the outside foreign elements. For the inside foreign elements the difference is smaller than for the model without rejection in all cases except the global rejection.

IV. Conclusions

The task of identifying *native* and *foreign* elements and rejecting *foreign* ones in the pattern recognition problem is discussed in this paper. Such the task is a nonstandard aspect of pattern recognition, which is rarely present in research. In this paper, ensembles of support vector machines solving two–classes and one–class problems are employed as classification tools and as basic tools for rejecting of foreign elements. Evaluation of quality of classification and rejection methods are proposed in the paper and finally some experiments are performed in order to illustrate acquainted terms and methods.

Tests done in the reported experiment affirm importance of the brought up problem of rejection of undesirable elements in pattern recognition problem. Such elements are called foreign ones in contrast to native ones. They may appear at any stage of pattern recognition tasks: measurement, observation and segmentation, feature extraction and selection, classification. There are several sources of foreign elements: errors and mistakes of measurement and observation, distortion and dirtiness of a (paper) document, carelessness of tool's reading and using etc.

The tests proved that the proposed model of local rejection gave the best results in all tested scenarios. Moreover, the method of global rejection and the one–class SVM based rejection can be also used in selected scenarios to reduce a range of learning data.

This study opens interesting directions of real world pattern recognition tasks investigations. In particular, aspects of a paper-to-computer memory technologies is among of high importance tasks in research and practice. Especially, imperfectness of scanned documents' segmentation is of great interest, since this stage produces undesired foreign elements affecting next stages of such technologies.

Acknowledgment

The research is supported by the National Science Center, grant No 2012/07/B/ST6/01501, decision no UMO–2012/07/B/ST6/01501.

References

[1] Bailing Zhang, Breast cancer diagnosis from biopsy images by serial fusion of Random Subspace ensembles, Biomedical Engineering and Informatics (BMEI), 2011 4th International Conference on, pp. 180–186.

[2] Chang C.-C. and Lin C.-J., LIBSVM: A library for support vector machines, ACM Transactions on Intelligent Systems and Technology, Vol 2, pp. 1-27, 2011

[3] Chow, C., On optimum recognition error and reject tradeoff, IEEE Transactions on Information Theory, 16(1), 41–46 (1970)

[4] Homenda W., Luckner M. and Pedrycz W., Classification with rejection: concepts and formal evaluations, Andrzej M.J. Skulimowski (ed.): Proceedings of KICSS'2013, pp. 161-172 Progress & Business Publishers, Krakow 2013

[5] Pillai, I., Fumera, G., Roli, F.: A classification approach with a reject option for multi-label problems. In: Maino, G., Foresti, G.L. (eds.) ICIAP (1). Lecture Notes in Computer Science, vol. 6978, pp. 98–107. Springer 2011

[6] Platt, J.C.: Probabilistic outputs for support vector machines and comparisonsto regularized likelihood methods. In: Smola, A., Bartlett, P., Schlkopf,B., Schurmans, D. (eds.) Advances in Large Margin Classifiers, pp. 61–74.MIT Press (1999)

[7] Schölkopf, B. at al., Estimating the Support of a High-Dimensional Distribution, Neural Comput. Vol. 13, No 7, pp. 1443-1471, 2001

[8] De Stefano, C. and Sansone, C. and Vento, M., To reject or not to reject: that is the question-an answer in case of neural classifiers , Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, Vol. 30, No 1, pp 84–94, 2000

[9] Vapnik V., Statistical Learning Theory, Wiley, New York, 1998

[10] Weber, D.M.; Du Preez, J.A., "A comparison between hidden Markov models and vector quantization for speech independent speaker recognition," Communications and Signal Processing, 1993., Proceedings of the 1993 IEEE South African Symposium on , vol., no., pp.139–144