

A Geometrical Approach to Rejecting Option in Pattern Recognition Problem

Jakub Ciecierski¹, Bartomiej Dybisz¹, Agnieszka Jastrzebska¹(✉),
and Witold Pedrycz^{2,3}

¹ Faculty of Mathematics and Information Science,
Warsaw University of Technology, ul. Koszykowa 75, 00-662 Warsaw, Poland
a.jastrzebska@mini.pw.edu.pl

² Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6,
01-447 Warsaw, Poland

³ Department of Electrical & Computer Engineering, University of Alberta,
Edmonton, AB T6R 2G7, Canada

Abstract. Frequently it happens that during symbols recognition, not all of them are the proper ones. This may cause deterioration of a classifying process. In this paper we present a way to “separate the wheat from the chaff”, by constructing a rejector, based on geometrical figures enclosing “wheat” and excluding “chaff”. We assume that entities of wheat, called native elements, are structured in some way and that there is no a priori knowledge about chaff, named foreign symbols. For the purpose of this study we present simple geometrical figures to generalize the distribution of symbols and to govern the rejection process.

Keywords: Pattern recognition · Rejecting option · Native and foreign elements

1 Introduction

In a standard attempt to pattern recognition an object is classified into one of given classes. However, in practice, this assumption is often too optimistic. Unfortunately, in important practical applications we do not only have elements belonging to one of proper classes, but also elements that do not belong to any of these classes, cf. [2, 5]. Such elements, which we shall not classify into any of proper classes, could appear for example due to an error in input signal segmentation.

Surprisingly, the issue of rejecting foreign symbols is rarely considered in otherwise very impressive research volume on pattern recognition, despite its obvious importance. We may refer to [2] and [3], where the issue of contaminated datasets is recalled. Among few recent papers, dealing with the issue of contaminated datasets we can list [1] and [4]. The literature study reveals that dominating technique for classification with rejection proposes to issue “flexible” decisions regarding class belongingness. This does not necessarily entail application of fuzzy sets, though in [11] authors point out that this is a viable method. Among studies oriented on rejection in pattern recognition problems we may refer to [8, 9].

Since elements not belonging to any of proper classes are usually not known a priori, i.e. at the stage of recognizer construction. Moreover, we cannot assume that they form their own class(es) and we cannot use them at this stage. To distinguish between these two types of elements the following terms are used:

- *native* elements for elements of recognized classes and
- *foreign* elements for ones not belonging to any given class.

Objectives of this study are aimed on inventing and exploring geometrical methods that could be employed for rejecting foreign elements, to test proposed methods on synthetic data and validate them on real data. In addition, the study focuses on comparison of different geometrical figures used for symbol rejection. In particular, two figures are considered: n -dimensional ellipsoid and hyperrectangle. As a working hypothesis we assume that the latter one will perform worse.

Let us emphasise that it is not our goal to elaborate and work on native symbols classification, i.e. determining their class belongingness. Such techniques are already well wrought. Nonetheless, when the need occurs we allow ourselves to use some well-known methods of this area.

The paper is structured as follows. In Section 2 mathematical rudiments of presented work are presented. Section 3 contains information about conducted experiments for both synthetic (3.1) as well real data and with their transformations (3.1). Eventually, Section 4 concludes the work.

2 Preliminaries

2.1 Basic Concepts

Conventionally, a pattern recognition problem is an action of dividing a set of objects $\mathbf{S} = \{s_1, s_2, \dots, s_{|\mathbf{S}|}\}$ into subsets, which include similar objects. Let us assume that $\mathbf{S} = \{S_1, S_2, \dots, S_{|\mathbf{C}|}\}$ such that $(\forall i \neq j)(S_i \cap S_j = \emptyset)$. Mapping $\sigma : \mathbf{S} \rightarrow \mathbf{C}$, where $\mathbf{C} = \{1, 2, \dots, |\mathbf{C}|\}$, constitutes a basic model for a task of splitting.

In order to compare and assign class labels to each object we identify them with a vector of some measurable characteristics, which are called features, and then perform a classification. Such approach can be split into two mappings: $\phi : \mathbf{S} \rightarrow \mathbf{X}$ and $\omega : \mathbf{X} \rightarrow \mathbf{C}$, where the first one goes from the space of object to the space of features and the second one from the space of features to the space of classes. It can be easily seen that $\sigma = \omega \circ \phi$ and from now on σ can be referred to as a classifier. In addition, throughout this paper we assume that $\mathbf{X} = X_1 \times X_2 \times \dots \times X_n = \mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R}$ i.e. space of features is the real coordinate space of n -dimensions - \mathbb{R}^n and we use lower-case and upper-case letters to distinguish vectors and sets, respectively.

Since set \mathbf{S} is commonly not available as a whole, we will construct the rejector only on some part of considered space. Let $\mathbf{S} \supset \mathbf{L} = L_1 \cup L_2 \cup \dots \cup L_{|\mathbf{C}|}$ such that $(\forall i \in \langle 1, |\mathbf{C}| \rangle)(L_i \subset S_i)$. We call \mathbf{L} a learning set. Furthermore we split the learning set into training set(\mathbf{Tr}) and testing set(\mathbf{Ts}) in the following

fashion: $\mathbf{L} = (Tr_1 \cup Ts_1) \cup (Tr_2 \cup Ts_2) \cup \dots \cup (Tr_{|C|} \cup Ts_{|C|})$ such that every class of the learning set is split into a training set and a testing set, namely: $(\forall i, j \in \langle 1, |C| \rangle)(Tr_i \cup Ts_i = L_i \wedge Tr_i \cap Ts_j = \emptyset)$ and $\mathbf{Tr} = Tr_1 \cup Tr_2 \cup \dots \cup Tr_{|C|}$ along with $\mathbf{Ts} = Ts_1 \cup Ts_2 \cup \dots \cup Ts_{|C|}$.

It is also worthy to note that the word *element* is used all over this paper to indicate examined object, as well as its features vector. Usually difference can be distinguished from the context, but wherever it may cause any doubts we explicitly state which is which.

2.2 Rejector Construction

Since we are interested only in evaluating whether given element belongs or not to any of the proper classes (rather than specifying to which one exactly), each element will be classified either as a native or a foreign one. In this case we assume that the rejector is represented by following mapping: $\rho : \mathbf{S} \rightarrow \mathbf{X} \rightarrow \bar{\mathbf{C}} = \{native, foreign\}$.

Let $\mathbf{Z} = Z_1 \cup Z_2 \cup \dots \cup Z_{|C|}$ represent a geometrical region in \mathbb{R}^n being a union of some figures Z_i . This region is assumed to enclose all points in training sets, such that Z_i is created based on training set T_i , for all $i \in \langle 1, |C| \rangle$. For convenience we assume that figures Z_i are convex and compact. Such a region will be used to determine if a feature vector from \mathbf{X} represents native or foreign element. We will write $x \in \mathbf{Z}$ to indicate the fact that some x from \mathbb{R}^n belongs to one or more figures from \mathbf{Z} . Since precise information about geometrical objects are defined in Section (2.3) - we are not covering any further details about \mathbf{Z} .

Clearly, we are interested in a mapping:

$$\lambda : \mathbf{X} \rightarrow \bar{\mathbf{C}} \quad \text{such that} \quad (\forall x \in \mathbf{X}) \quad \lambda(x) = \begin{cases} native & \text{if } x \in \mathbf{Z} \\ foreign & \text{otherwise} \end{cases}$$

In other words if a vector of features belongs to the region \mathbf{Z} , we classify it as native. Otherwise it will be treated as a foreign element. To sum up, we complete ρ with a new mapping $\lambda: \mathbf{S} \xrightarrow{\phi} \mathbf{X} \xrightarrow{\lambda} \bar{\mathbf{C}}$, where λ is the object of our study.

2.3 Geometry

As mentioned in the previous Section, there is still a need to precise, which geometrical figures we will use to construct the rejector, i.e. the region \mathbf{Z} . We have chosen n -dimensional ellipsoids and hyperrectangles and in what follows we define how to establish a membership to each of them. Choice of figures has been motivated by their simplicity (hyperrectangles are really straight-forward to compute) and intuitiveness (using ellipsoidal figures rely on the expectation that features vectors will be normally distributed among each class).

Ellipsoids. An n -dimensional ellipsoid E can be defined as follows:

$$E = \{x \in \mathbb{R}^n : (x - x_0)^T A(x - x_0) \leq 1\}, \tag{1}$$

where $x_0 \in \mathbb{R}^n$ is the center of the ellipsoid E and A is a positive definite matrix, cf. [10, 13]. Having this in mind, it is obvious that any ellipsoid is uniquely represented by its center x_0 and matrix A .

Given volume of the unit n -dimensional hypersphere V_0 , we can compute volume of the ellipsoid E in the following way:

$$Vol(E) = \frac{V_0}{\sqrt{\det(A)}}, \quad (2)$$

Referring to Section 2.2, we define a region \mathbf{Z} as union of ellipsoids:

$$\mathbf{Z} = E_1 \cup E_2 \cup \dots \cup E_{|\mathbf{C}|}$$

such that the ellipsoid E_i encloses all elements of the training set Tr_i and has minimal volume, for all classes in \mathbf{C} . Let us notice that according to Formula (2) volume minimization is equivalent to maximization of the determinant. The task of finding minimal volume ellipsoids enclosing given sets of elements was solved by implementation of methods given in [13].

Finally, let us state that for any $x \in \mathbf{X}$, $x \in \mathbf{Z}$ if and only if it is enclosed by at least one ellipsoid from \mathbf{Z} .

Hyperrectangles. Alike in the case of ellipsoids, we now construct a rejection region \mathbf{Z} as union of n -dimensional hyperrectangles enclosing training sets of native elements:

$$\mathbf{Z} = H_1 \cup H_2 \cup \dots \cup H_{|\mathbf{C}|}$$

such that the hyperrectangle H_i encloses all elements of the training set Tr_i , its edges are parallel to axes and it has minimal volume, for all classes in \mathbf{C} .

More precisely, the hyperrectangle H_i enclosing the training set Tr_i is a Cartesian product of the following finite intervals:

$$H_i = I_i^1 \times I_i^2 \times \dots \times I_i^n$$

where, for all classes $i \in \mathbf{C}$, the interval I_i^k is the minimal interval containing values of k -th parameter for all points of the training set Tr_i . Assuming that $f_k(x)$ is the value of k -th feature of an elements x , we get:

$$I_i^k = \langle \min \{f_k(x) : x \in Tr_i\}, \max \{f_k(x) : x \in Tr_i\} \rangle$$

Again, let us state that for any $x \in \mathbf{X}$, $x \in \mathbf{Z}$ if and only if it is enclosed by at least one hyperrectangle from \mathbf{Z} .

2.4 Evaluation

For a better understanding of 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 their original sources here. The following parameters were used in defining several measures outlining classification's quality. These parameters create so called confusion matrix, which is given in Table 1. The parameters given in the matrix are numbers of elements of a testing set, which have the following meaning:

- TP - the number of elements of the considered class correctly classified to this class,
- FN - the number of elements of the considered class incorrectly classified to other classes,
- FP - the number of elements of other classes incorrectly classified to the considered class,
- TN - the number of elements of other classes correctly classified to other classes (no matter, if correctly, or not).

Table 1. Confusion matrix for rejecting in pattern recognition problem

| | Classification to the class | Classification to other classes |
|---------------|-----------------------------|---------------------------------|
| The class | True Positives (TP) | False Negatives (FN) |
| Other classes | False Positives (FP) | True Negatives (TN) |

In this study, original pattern recognition problem is a multiclass problem. However, from the point of view of rejection task, it is important whether a native element is classified to any native class, even to an incorrect native class. Therefore, the question if a native element is classified to the correct or any other native class is not considered here. Hence, a multiclass pattern recognition with rejection is turned to a two-class problem. Finally, the following measures were used assess the quality of the classifier:

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN}{TP+FN+FP+TN} \\
 \text{Sensitivity} &= \frac{TP}{TP+FN} \\
 \text{Precision} &= \frac{TP}{TP+FP} \\
 \text{F-measure} &= 2 \cdot \frac{\text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{3}
 \end{aligned}$$

The following comments help to understand these characteristics:

- Accuracy is a measure of a rejecting performance. This measure describes the ability to distinguish between native and foreign elements. Of course, the higher the value of this measure, the better the identification.
- Precision is the ratio of the number of not rejected native elements to the number of all elements (native and foreign) not rejected. Precision evaluates the rejection ability to separate foreign elements from native ones. The higher the value of this measure, the better ability to distinguish foreign elements from native ones.

- Sensitivity is the ratio of the number of not rejected native elements to the number of all native ones. This measure evaluates the rejection ability to identify native elements. The higher the value of Sensitivity, the more effective identification of native elements. Unlike the Precision, this measure does not evaluate the effectiveness of separation between native and foreign elements. The higher the value of this measure, the better ability to identify native elements.
- Precision and Sensitivity are complementary: increasing sensitivity can cause a drop in precision since, along with increasing the number of not rejected native elements, there might be more not rejected foreign ones. It is there to express the balance between precision and sensitivity since, in practice, these two affect each other. There exists yet another characteristic that combines them: the F-measure. The higher the value of this measure, the better balance between identification of native elements along with separation native elements from foreign ones.

3 Experiment

Experiments are performed for two kinds of data: synthetic and real, both described in Section 3.1. Elements of both kinds are characterized by 24 features. In both, ellipsoids and hyperrectangles were built on chosen set of elements, cf. Sections 2.2 and 2.3, and effectiveness of rejectors has been checked. Both kinds of datasets have been split into learning and testing set.

In addition, we observe how rejection behaves when we start to remove some of points from training classes and reconstruct geometric figures according to shrunken training sets. We shrink number of elements 4 times; at each stage we subtract 5% of points from the current cardinality of training set.

Figure 1 depicts the process of ellipsoids and hyperrectangles shrinking. In this Figure, elements of two classes are presented in the 2-dimensional space of features in order to keep explanation as lucid as possible.

3.1 Datasets

Synthetic Data. For this kind of data randomizing symbols demands several stages. First of all we are generating 24 intervals (using uniform distribution), where each interval can take values from 0 to 20 inclusively. These intervals serve as a scope of possible values for each of features. Next, we are creating 10 different classes of symbols by assigning to them random vectors of features, as well drawn from a uniform distribution. At this stage we have 10 points in our features space. Next, we distort them by creating clouds of points around each symbol. Process of spreading points is done as follows: take features vector of some symbol and for every value of this vector add a number drawn from $N(0, 1)$. This is done 1500 times per each previously generated symbol. It gives us 10 clouds of points, each having 1500 elements. We treat clouds as learning sets, where 1000 points belong to training set and remaining 500 to testing set.

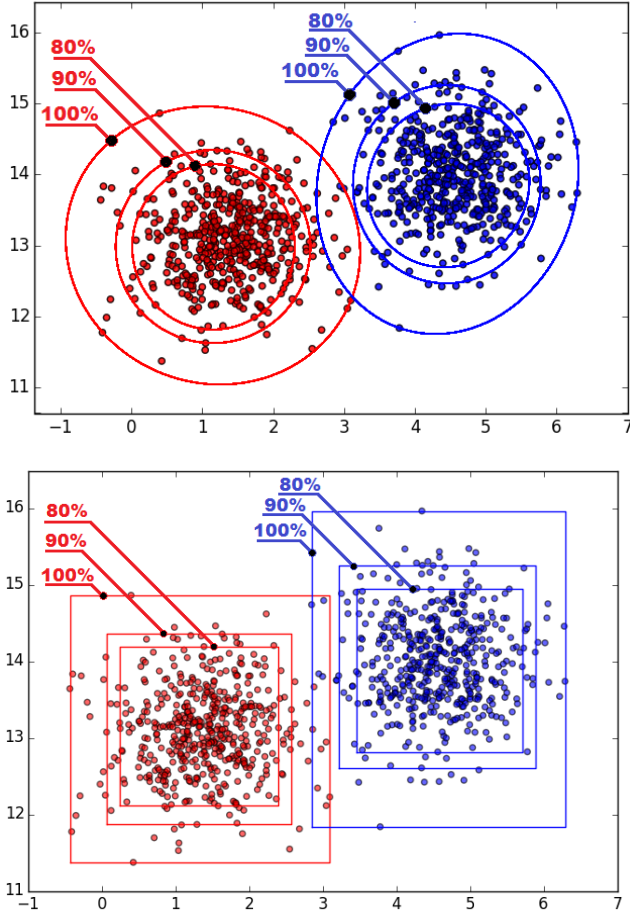


Fig. 1. Consecutive shrinking process applied for: ellipsoids (upper part) and hyper-rectangles (bottom part).

As for foreign symbols, two approaches were applied:

- homogeneous distribution in the space of features, which spreads elements uniformly around the Cartesian product of features intervals. Elements accidentally overlapping with natives were redrawn.
- non-homogeneous distribution - created by applying following process: for each segment between a pair of native clouds centers find its middle point. Treat the points as a centre of new cloud of 1000 foreign points, created using the normal distribution.

In this way, in both manners, we get 10000 foreign points.

Real and Semi-synthetic Data. The datasets were constructed based on the MNIST database of handwritten digits [12]. The following figure (2), represents a sample of natives elements used in the experiments:

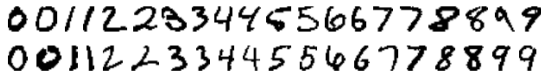


Fig. 2. Samples of elements of native classes.

Then, based on this database, two sets of foreign elements were constructed. Sample of such set is represented by Figure 3.

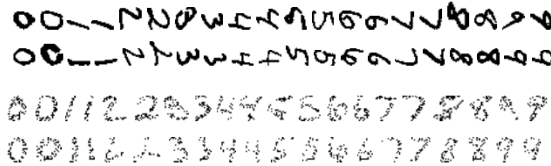


Fig. 3. Transformed native elements: rotated by 90 degrees (upper part) and randomly distorted (bottom part).

On the other hand, we refer to collection of foreign symbols as ‘semi-synthetic’, because they are neither pure real, nor pure synthetic objects. They have been created by modifying in some specific fashion (therefore synthetic) the native set of symbols (therefore real). In this paper we incorporate two such sets: A and B. First one comprises of native symbols rotated by 90 degrees, while the second one of uniformly distorted native elements, cf. Figure 3.

3.2 Results

This Section is built mostly upon tables, which present results of the experiments. Although we separate synthetic data measurements from real/semi-synthetic ones, at the end there is an array concluding both of them.

Synthetic Data. First of all we present Tables 2 and 3. Each of them is divided into considered geometric figure (upon which rejector has been based on), type of foreign symbols distribution and consecutive steps of shrinking figures numbered as 1000, 950, 903, 858, 816 (number of points in each figure at each stage). The only difference between them is that both training and testing sets are used in the first case, while test set in the second case. One can observe natural behaviour of hyperrectangles, which in the end of the volume minimizing, enclose more native points than ellipsoids. What is more, ellipsoids have better foreign recognition properties, as mostly they are smaller (in terms of volumes) than hyperrectangles. In addition, both figures work better with homogeneous distributed foreign, which is caused by an assumption that feature’s vectors of foreign symbols are equally likely to appear within the considered space.

Table 2. Confusion matrices for consecutive shrinking of hyperrectangle (h-r) and ellipsoid (ell) at training set (in %).

| training set | Rejector built for synthetic data sets | | | | | | | |
|------------------|--|-------|--------|-------|-----------------|-------|--------|-------|
| data set | homogeneous | | | | non-homogeneous | | | |
| figure | h-r | | ell | | h-r | | ell | |
| confusion matrix | TP | FN | TP | FN | TP | FN | TP | FN |
| | FP | TN | FP | TN | FP | TN | FP | TN |
| -0% | 100.00 | 0.00 | 100.00 | 0.00 | 100.00 | 0.00 | 100.00 | 0.00 |
| | 0.67 | 99.33 | 0.14 | 99.86 | 25.39 | 74.61 | 14.96 | 85.04 |
| -5% | 98.46 | 1.54 | 94.84 | 5.16 | 98.46 | 1.54 | 94.84 | 5.16 |
| | 0.48 | 99.52 | 0.07 | 99.93 | 21.67 | 78.33 | 10.54 | 89.46 |
| -10% | 96.98 | 3.02 | 90.10 | 9.90 | 96.98 | 3.02 | 90.10 | 9.90 |
| | 0.38 | 99.62 | 0.03 | 99.97 | 17.62 | 82.38 | 7.91 | 92.09 |
| -15% | 95.53 | 4.47 | 85.64 | 14.36 | 95.53 | 4.47 | 85.64 | 14.36 |
| | 0.31 | 99.96 | 0.02 | 99.98 | 17.13 | 82.87 | 5.38 | 94.62 |
| -20% | 94.23 | 5.77 | 81.56 | 18.44 | 94.23 | 5.77 | 81.56 | 18.44 |
| | 0.29 | 99.71 | 0.01 | 99.99 | 16.99 | 83.01 | 4.83 | 95.17 |

Table 3. Confusion matrices for consecutive shrinking of hyperrectangle (h-r) and ellipsoid (ell) at testing set (in %).

| testing set | Rejector built for synthetic data sets | | | | | | | |
|------------------|--|-------|-------|-------|-----------------|-------|-------|-------|
| data set | homogeneous | | | | non-homogeneous | | | |
| figure | h-r | | ell | | h-r | | ell | |
| confusion matrix | TP | FN | TP | FN | TP | FN | TP | FN |
| | FP | TN | FP | TN | FP | TN | FP | TN |
| -0% | 95.10 | 4.90 | 87.96 | 12.04 | 95.10 | 4.90 | 87.97 | 12.03 |
| | 0.71 | 99.29 | 0.03 | 99.97 | 31.03 | 68.97 | 8.33 | 91.67 |
| -5% | 93.36 | 6.64 | 80.66 | 19.34 | 93.36 | 6.64 | 80.66 | 19.34 |
| | 0.47 | 99.53 | 0.02 | 99.98 | 27.64 | 72.36 | 6.31 | 93.69 |
| -10% | 92.12 | 7.88 | 76.16 | 23.84 | 92.12 | 7.88 | 76.16 | 23.84 |
| | 0.39 | 99.61 | 0.00 | 100.0 | 24.32 | 75.68 | 5.63 | 94.37 |
| -15% | 91.02 | 8.98 | 71.68 | 28.32 | 91.02 | 8.98 | 71.68 | 28.32 |
| | 0.33 | 99.67 | 0.00 | 100.0 | 22.55 | 77.45 | 4.55 | 95.45 |
| -20% | 89.52 | 10.48 | 66.74 | 33.26 | 89.52 | 10.48 | 66.74 | 33.26 |
| | 0.26 | 99.74 | 0.00 | 100.0 | 21.43 | 78.57 | 3.84 | 96.16 |

Real and Semi-synthetic Data. As for this type, we provided only one Table (4), which concerns both sets: A and B. ‘h-r’ is an abbreviation for hyperrectangle, while ‘ell’ means ellipsoid. In this case each stage of shrinking is marked as ‘-0%’, ‘-5%’, ‘-10%’, ‘-15%’ or ‘-20%’, because our samples contain approximately 1000 features’ vectors per symbol class, which causes slight

Table 4. Confusion matrices for consecutive shrinking in ellipsoid (ell) and hyperrectangle (h-r) approach. Considered are sets A and B of foreign elements.

| testing set | Rejector built for real data set | | | | | | | |
|------------------|----------------------------------|-------|-------|-------|-------|-------|-------|-------|
| data set | A | | | | B | | | |
| geometry | h-r | | ell | | h-r | | ell | |
| confusion matrix | TP | FN | TP | FN | TP | FN | TP | FN |
| | FP | TN | FP | TN | FP | TN | FP | TN |
| -0% | 98.42 | 1.58 | 85.06 | 14.93 | 98.45 | 1.55 | 85.98 | 14.02 |
| | 62.41 | 37.59 | 12.10 | 87.90 | 3.18 | 96.82 | 0.11 | 99.89 |
| -5% | 96.45 | 3.55 | 80.82 | 19.18 | 96.74 | 3.26 | 81.83 | 18.17 |
| | 54.16 | 45.84 | 7.53 | 92.47 | 2.57 | 97.43 | 0.10 | 99.90 |
| -10% | 95.60 | 4.40 | 77.25 | 22.75 | 95.45 | 4.55 | 76.93 | 23.07 |
| | 49.96 | 50.04 | 5.51 | 94.49 | 1.95 | 98.05 | 0.05 | 99.95 |
| -15% | 93.92 | 6.08 | 73.61 | 26.39 | 93.88 | 6.12 | 73.37 | 26.63 |
| | 46.41 | 53.59 | 4.27 | 95.73 | 1.87 | 98.13 | 0.02 | 99.98 |
| -20% | 92.77 | 7.23 | 69.53 | 30.47 | 66.74 | 33.26 | 93.15 | 6.85 |
| | 43.20 | 56.80 | 3.18 | 96.82 | 0.00 | 100.0 | 1.80 | 98.20 |

differences in number of points enclosed by each of geometrical figure, hence it can not be generalized. Of course ‘-0%’ symbolizes situation, in which geometric figures are based on all points in training sets. As in the previous Section, data are presented in confusion matrices built with testing set.

First remark is that in both foreign sets hyperrectangle approach is not always worse than the ellipsoid one - we have got 100% rejected foreign symbols in case of hyperrectangles in B, while ellipsoids rejected 98.20%. Second observation is that for two sets ellipsoid-based rejectors works well, while hyperrectangles acquire rather average results in A. Hyperrectangle approach in sets A and B manifests difference in points classified as true positive. In the first one, at stage ‘-20%’, it is 92.77% while in the second 66.74%. This reveals difference in native elements distribution i.e. in set B they are more ‘compressed’ than in A, hence shrinking process has created hyperrectangles of a bigger volume in the case of set A than in B. In addition in set A, ellipsoid technique, we achieved high foreign symbols rejection rate, but at the same time the rate of native elements’ classification decreased.

Summary. As stated in Section 1, one of the main aims of this study is to measure differences between approach based on ellipsoids and hyperrectangles. In this Section we present Table 5, which is the resume of performed tests, based on measuring sensitivity, accuracy, precision and F-measure features, as described in 2.4. This array helps us observing behaviour of both considered geometrical figures.

Table 5. Measuring rejecting results for synthetic and real data for shrunken figures ('-0%', '-5%', '-10%', '-15%' or '-20%'). h-r -hyperrectangle, ell - ellipsoid.

| testing sets | synthetic | | | | real/semi-synthetic | | | |
|--------------|-------------|-------|--------------|-------|---------------------|-------|-------|-------|
| | homogeneous | | non-homogen. | | 90ccl | | AND | |
| figure | h-r | ell | h-r | ell | h-r | ell | h-r | ell |
| Sensitivity | 95.10 | 87.96 | 95.10 | 87.96 | 98.41 | 85.06 | 98.45 | 85.98 |
| Accuracy | 97.89 | 95.97 | 77.68 | 90.43 | 57.99 | 86.95 | 97.37 | 95.23 |
| Precision | 98.52 | 99.93 | 60.51 | 84.07 | 44.32 | 78.01 | 93.97 | 99.75 |
| F-measure | 96.78 | 93.56 | 73.96 | 85.97 | 61.11 | 81.38 | 96.16 | 92.35 |
| Sensitivity | 93.36 | 80.66 | 93.36 | 80.66 | 96.45 | 80.82 | 96.74 | 81.83 |
| Accuracy | 97.47 | 93.54 | 79.36 | 89.35 | 62.82 | 88.56 | 97.20 | 93.85 |
| Precision | 99.00 | 99.95 | 62.81 | 86.47 | 47.33 | 84.42 | 94.99 | 99.76 |
| F-measure | 96.10 | 89.27 | 75.10 | 83.46 | 63.50 | 82.58 | 95.86 | 89.90 |
| Sensitivity | 92.12 | 76.16 | 92.12 | 76.16 | 95.60 | 77.25 | 95.45 | 76.93 |
| Accuracy | 97.11 | 92.05 | 81.16 | 88.30 | 65.32 | 88.70 | 97.18 | 92.24 |
| Precision | 99.16 | 100.0 | 65.44 | 87.12 | 49.13 | 87.62 | 96.10 | 99.87 |
| F-measure | 95.51 | 86.47 | 76.52 | 81.27 | 64.90 | 82.11 | 95.78 | 86.91 |
| Sensitivity | 91.02 | 71.68 | 91.02 | 71.68 | 99.91 | 73.61 | 93.88 | 73.37 |
| Accuracy | 96.79 | 90.56 | 81.97 | 87.53 | 67.12 | 88.31 | 96.71 | 91.07 |
| Precision | 99.28 | 100.0 | 66.87 | 88.73 | 50.53 | 89.69 | 96.19 | 99.94 |
| F-measure | 94.97 | 83.50 | 77.10 | 79.30 | 65.70 | 80.86 | 95.02 | 84.62 |
| Sensitivity | 89.52 | 66.74 | 89.52 | 66.74 | 92.77 | 69.52 | 93.15 | 69.53 |
| Accuracy | 96.33 | 88.91 | 82.22 | 86.35 | 68.86 | 87.66 | 96.50 | 89.79 |
| Precision | 99.42 | 100.0 | 67.62 | 89.68 | 52.01 | 91.69 | 96.30 | 99.97 |
| F-measure | 94.21 | 80.05 | 77.05 | 76.53 | 66.65 | 79.08 | 94.70 | 82.02 |

We have divided the array into datasets types, foreign types (homogeneous, non-homogeneous; set A, set B), rejector kind which has been used ('h-r' means hyperrectangle; 'ell' describes ellipsoid) and results for consecutive shrinking of geometrical figures.

First, we observe that in ellipsoid approach sensitivity decreases quicker than in hyperrectangle one. Also, interesting is fact that for non-homogeneous foreign symbols, F-measure of hyperrectangle technique actually elevates, while in all others is getting smaller (especially in the ellipsoid type of rejectors). In all cases during shrinking process precision rises, but some of the accuracy is lost.

4 Conclusion and Future Research Directions

We have proposed an approach to native elements recognition/foreign elements rejection. Native elements belong to proper classes, which we plan to process, while foreign elements may appear in a data set due to an unexpected error. Because classification itself has not been in the scope of this study, presented

methods may play crucial role for example in data preprocessing, where one can firstly dispose of foreign elements and then focus on proper classification. Foreign elements removed early from a dataset would not be subjected to a classification process and therefore the proposed technique could improve the overall pattern recognition process.

Two methods have been analysed: based on hyperrectangles and on ellipsoids. Despite the fact that ellipsoid-based rejector exhibits high precision in detecting foreign elements, it forfeits significant amount of native elements (especially in shrinking process). On the other hand, hyperrectangular rejector detects native elements on acceptable level but its ability to discern foreign ones is closely related to the learning set layout. As for real life application, it is beneficial to use ellipsoid-based rejector, without shrinking technique, but one must keep in mind that with this method approximately 14% of native points may be wrongly rejected.

As to future directions, we plan to employ models of bipolar uncertainty to features and their aggregation [6,7]. Since final decision is crisp (a pattern is either native, or foreign) preferred are univariate bipolar models based on the uncertainty interval $[-1,1]$ rather than bivariate models based on two unit intervals representing negative/positive uncertainties.

Acknowledgement. The research is supported by the National Science Center, grant No 2012/07/B/ST6/01501, decision no UMO?1?72012/07/B/ST6/01501.

References

1. Bertolami, R., Zimmermann, M., Bunke, H.: Rejection strategies for offline handwritten text line recognition. *Pattern Recognition Letters* **27**(16), 2005–2012 (2006)
2. Bromley, J., Denker, J.S.: Improving rejection performance on handwritten digits by training with rubbish. *Neural Computing* **5**, 367–370 (1993)
3. Chow, C.: On optimum recognition error and reject tradeoff. *IEEE Transactions on Information Theory* **16**(1), 41–46 (1970)
4. Elad, M., Hel-Or, Y., Keshet, R.: Pattern detection using a maximal rejection classifier. In: Arcelli, C., Cordella, L.P., Sanniti di Baja, G. (eds.) *IWVF 2001*. LNCS, vol. 2059, pp. 514–524. Springer, Heidelberg (2001)
5. Homenda, W.: Optical music recognition: the case study of pattern recognition. In: Kurzynski, M., Puchala, E., Wozniak, M., et al. (eds.) *Computer Recognition Systems. Advances in Soft Computing*, pp. 835–842 (2005)
6. Homenda, W.: Balanced Fuzzy Sets. *Information Sciences* **176**, 2467–2506 (2006)
7. Homenda, W., Pedrycz, W.: Processing of uncertain information in linear space of fuzzy sets. *Fuzzy Sets & Systems* **44**, 187–198 (1991)
8. Homenda, W., Lesinski, W.: Optical music recognition: a case of pattern recognition with undesirables and garbage symbols. In: Choras, R., et al. (eds.) *Image Processing and Communications Challenges*, pp. 120–127 (2009)
9. Koerich, A.: Rejection strategies for handwritten word recognition. In: *Proc. of the 9th International Workshop on Frontiers in Handwriting Recognition* (2004)
10. Kumar, P., Yildirim, E.A.: Minimum-Volume Enclosing Ellipsoids and Core Sets. *Journal of Optimization Theory and Applications* **126**(1), 1–21 (2005)

11. Meel, A., Venkat, A.N., Gudi, R.D.: Disturbance classification and rejection using pattern recognition methods. *Ind. Eng. Chem. Res.* **42**(14), 3321–3333 (2003)
12. LeCun, Y., Cortes, C., Burges, C.: The MNIST database of handwritten digits (1996). <http://yann.lecun.com/exdb/mnist/>
13. Todda, M.J., Yildirim, E.A.: On Khachiyan's algorithm for the computation of minimum-volume enclosing ellipsoids. *Discrete Applied Mathematics* **155**(13), 1731–1744 (2007)