Multistage Naive Bayes Classifier with Reject Option for Multiresolution Signal Representation

Urszula Libal
Institute of Computer Engineering, Control and Robotics, Wroclaw University of Technology, Wroclaw, Poland
urszula.libal@pwr.wroc.pl

Keywords: Multistage Classifier, Naive Bayes, Reject Option, Pattern Recognition, Wavelet, Resolution.

Abstract: In the article, two approaches to pattern recognition of signals are compared: a direct and a multistage. It is assumed that there are two generic patterns of signals, i.e. a two-class problem is considered. The direct method classifies signal in one step. The multistage method uses a multiresolution representation of signal in wavelet bases, starting from a coarse resolution at the first stage to a more detailed resolutions at the next stages. After a signal is assigned to a class, the posterior probability for this class is counted and compared with a fixed level. If the probability is higher than this level, the algorithm stops. Otherwise the signal is rejected and on the next stage the classification procedure is repeated for a higher resolution of signal. The posterior probability is calculated again. The algorithm stops when the probability is higher than a fixed level and a signal is finally assigned to a class. The wavelet filtration of signal is used for feature selection and acts as a magnifier. If the posterior probability of recognition is low on some stage, the number of features on the next stage is increased by taking a better resolution. The experiments are performed for three local decision rules: naive Bayes, linear and quadratic discriminant analysis.

1 INTRODUCTION

Sometimes the direct approach to classification does not give the desired results. Then a classifier with reject option (Devroye et al., 1996) may be used. The object rejection is a cancellation of the object assignment to one of the classes, if the decision is not certain on a reasonable level. This approach can reduce the risk of misclassification (Pudil et al., 1992).

In opposition to a multistage classifier based on decision trees (Burduk and Kurzyński, 2006); (Kurzyński, 1988); (Libal, 2010), the new multistage approach to classification is presented in this article. There is proposed a multistage classifier, which is a sequence of Bayes decision rules with the reject option. The new classifier is dedicated to signal recognition and uses wavelet representation of signals. There are assumed only two classes of signals. In case of inability to identify the class at some stage (i.e. signal rejection), it will try to classify the signal to one of the two classes at the next stage. It should be noted that at each stage there are still the same two classes considered, and the number of steps of the algorithm is not determined arbitrarily.

To avoid the curse of dimensionality (the empty space phenomenon), the signal is represented by the wavelet approximation coefficients in the following way: at an early stage classifier uses signal representation in a low resolution. And if it is not enough (i.e. rejection case), then classifier will use signal representation in an increased resolution at the next stage. The method of obtaining wavelet coefficient vectors $cA_1, cA_2, \dots, cA_m$ by the wavelet decomposition of signals with the use of the Mallat algorithm is described in the section 2.1.

2 MULTISTAGE CLASSIFIER

The considered problem is to classify a noised signal to one of two classes. There is shown a multistage algorithm with reject option, i.e. on every stage a local classifier assigns an analysed signal to a class from $\{1, 2\}$ or rejects it. If the signal was assigned to class 1 or 2, then algorithm stops. On the other hand, after the rejection signal stays unclassified and waits for the classification on the next stage. The difference between stages is a representation of the signal:

1) cA_m - at the first stage a coarse signal representation in wavelet bases is considered (with a small number of wavelet approximation coefficients for a low resolution),
2) cA_{m-1} - at the second stage and so on,

⋮

m) cA_1 - till a detailed representation in a high resolution (with a large number of coefficients).

2.1 Signal Representation

A multiresolution signal representation, in the form of the sequence $(cA_1, cA_2, \dots, cA_m)$, can be obtained by a wavelet transform of the signal (Mallat, 1989). The wavelet decomposition of signal (Daubechies, 1992) allows to approximate signal for various resolutions, lower than the initial resolution of a signal before the transformation. Fast wavelet decomposition can be performed by the Mallat algorithm, shown in figure 1.

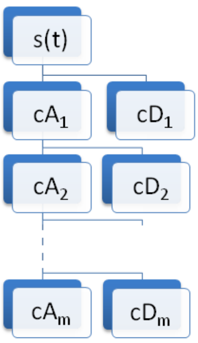

Figure 1: Wavelet decomposition by the Mallat algorithm.

At the first decomposition level it filtrates a signal $s(t)$ with low- and high-pass filters and then decimates the filtration results. This procedure computations are very fast and give two coefficient vectors: approximation cA_1 and detail cD_1 coefficients:

$$s = cA_1 + cD_1. \tag{1}$$

At every following level the algorithm filtrates the approximation of signal cA_i (received at the previous step) with low- and high-pass filters and the results are also down-sampled, i.e.:

$$cA_i = cA_{i+1} + cD_{i+1}. \tag{2}$$

As noticed by Mallat, the multiresolution representation of a signal or an image is a very effective method of extracting information (Mallat, 1989). In figure 2 are presented two patterns (in the first row) generating two classes of noised signals (second row). The approximation cA_4 of signal at 4^{th} decomposition level (third row) reveals a noticeable division between classes, especially for 9^{th} coefficient.

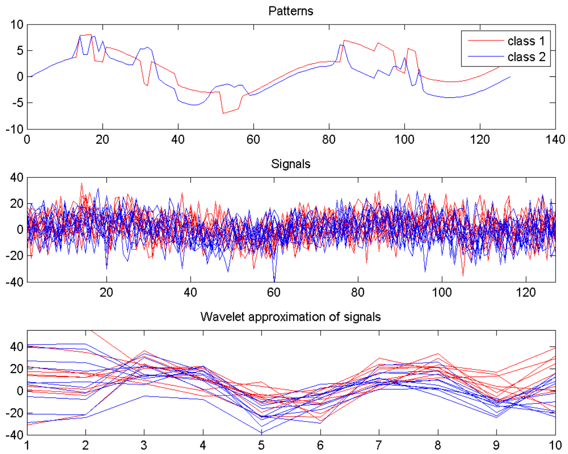

Figure 2: Two class patterns, noisy signals for $\sigma = 10$ and wavelet approximation coefficients cA_4 of 4^{th} level.

2.2 Direct Classifier with Reject Option

The local Bayesian classifier Ψ_R with reject option assigns a signal to one class from the set $M = \{1, 2\}$ or rejects it what is denoted by zero. The signal is represented by a sequence of wavelet approximation coefficients cA_i. The classification rule Ψ_R is given by formula

$$\Psi_R = \begin{cases} 1, & \text{if } 1 - \eta(cA_i) \leq \lambda \\ 0, & \text{if } \min\{\eta(cA_i), 1 - \eta(cA_i)\} > \lambda \\ 2, & \text{if } \eta(cA_i) \leq \lambda \end{cases} \tag{3}$$

where the posterior probability of assigning a signal to class 1, after observing values of wavelet coefficient vector x, is

$$\eta(x) = \frac{p_1 f_1(x)}{p_1 f_1(x) + p_2 f_2(x)}. \tag{4}$$

The posterior probability of assigning a signal to class 2 is equal to $1 - \eta(x)$. The prior probabilities of class occurrences are p_1 and p_2. The probability density functions in both classes are given by $f_1(x)$ and $f_2(x)$.

The classification rule (3) can be transformed into the following form

$$\Psi_R = \begin{cases} 1, & \text{if } \eta(cA_i) \in [1 - \lambda, 1] \\ 0, & \text{if } \eta(cA_i) \in (\lambda, 1 - \lambda) \\ 2, & \text{if } \eta(cA_i) \in [0, \lambda] \end{cases} \tag{5}$$

what leads to the one-stage local decision rule in the multistage algorithm, defined in the next section 2.3.

2.3 Multistage Algorithm

The multistage algorithm has the following form:

Algorithm 1. *Multistage Classifier*

```
1: for i from m to 1
   {
2:     if η(cA_i) ∈ [1 − λ, 1]
         then class=1
         END
3:     if η(cA_i) ∈ (λ, 1 − λ)
         then class=0
         GO TO STEP 1
4:     if η(cA_i) ∈ [0, λ]
         then class=2
         END
   }
5: return class
```

The maximum number of stages is m. There will be less stages if the posterior probability η achieves an appropriate level, dependent on the rejection threshold λ.

2.4 Theoretical Risk

The difference between risk values for Bayes classifiers with and without reject option (Libal, 2012), denoted by Ψ_R and Ψ respectively, is

$$R[\Psi] - R[\Psi_R] = \tag{6}$$

$$= \int_{D_R^0} (\min\{\eta(x), 1 - \eta(x)\} - \lambda) \cdot$$

$$\cdot (p_1 f_1(x) + p_2 f_2(x)) \, dx \geq 0,$$

where D_R^0 is the rejection area if the signal is represented by a vector x and the rejection threshold is fixed to λ. The area $D_R^0 \neq \emptyset$ only if there is available the reject option, i.e. $\lambda \in [0, 0.5)$. If $\lambda = 0.5$, there is no reject option and the area $D_R^0 = \emptyset$ is an empty subspace of a feature space. Examples of rejection areas are presented in figures 3 and 4.

3 EXPERIMENTS

There were 3 multistage classifiers tested with the basic decision rules as:

1) naive Bayes,
2) linear discriminant analysis (LDA),
3) quadratic discriminant analysis (QDA).

Every one-stage rule was performed under assumption that all features (i.e. wavelet approximation coefficients on a fixed decomposition level) are pairwise uncorrelated. The algorithm has 4 stages and the signals of length 128 were decomposed for 4 levels by the Daubechies wavelet family of order 2 (Daubechies, 1992). The signal was represented (if it was necessary):

- at 1st stage by $x = cA_4$, (10 coefficients),
- at 2nd stage by $x = cA_3$, (18 coefficients),
- at 3rd stage by $x = cA_2$, (34 coefficients),
- at 4th stage by $x = cA_1$, (65 coefficients).

The figures 3 and 4 shows in 2D the partition of a feature space into decision areas of class 1 and 2 and rejection area, but the dimensionality of feature space in performed experiments was always higher, i.e. depending on the stage there were 10, 18, 34 or 65-dimensions.

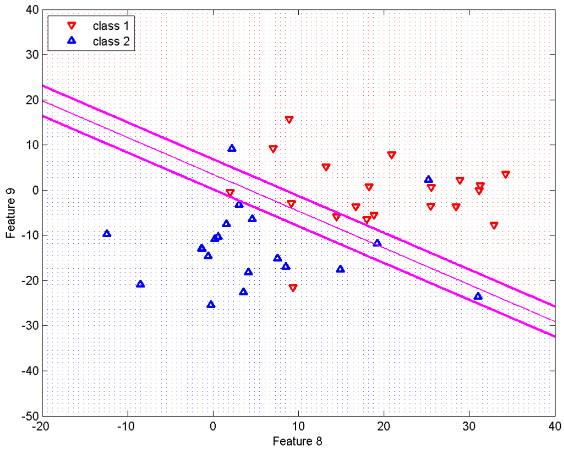

Figure 3: Decision areas for classes and rejection area (between lines) for LDA.

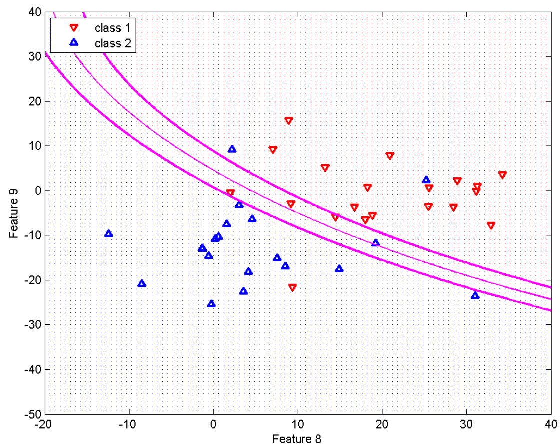

Figure 4: Decision areas for classes and rejection area (between curves) for QDA.

4 CONCLUSIONS

According to the inequality (6) the theoretical risk of one-stage Bayes classifier Ψ without reject option is higher than or equal to the risk of classifier Ψ_R with reject option, i.e.

$$R[\Psi] \geq R[\Psi_R]. \tag{7}$$

This fact has been confirmed by the experimental results. In the figures 5, 6 and 7 is shown the risk of incorrect classification of signals, containing class patterns and a Gaussian white noise at the level of σ (see figure 2). For the multistage classifier (i.e. for $\lambda = 0.1, 0.2, 0.3$ and 0.4) the risk values from all four stages were summarized. At the last stage classifier has to choose a class from $\{1, 2\}$, so there were no unclassified signals at the end. For $\lambda = 0.5$ the classifier has no reject option and only one-stage.

The rejection threshold λ is a suffered loss after rejecting a signal (choosing class 0). The loss after choosing a wrong class from the set $\{1, 2\}$ is 1, and there is no loss after choosing a correct class, what means that the loss is equal to 0 then.

The lower the $\boldsymbol{\lambda}$, the lower the risk of misclassification. The application of the presented classifier (given by algorithm 1) for wavelet representation of signals improves the classifier efficiency (see figures 5, 6 and 7). The lowest values of experimental risk among the three methods were obtained for linear discriminant analysis (LDA).

REFERENCES

Burduk, R. and Kurzyński, M., (2006). Two-stage binary classifier with fuzzy-valued loss function. *Pattern Analysis & Applications*, 9(4), pp.353-358.

Daubechies, I., (1992). *Ten Lectures on Wavelets*. Lecture Notes Vol. 61, SIAM, Philadelphia.

Devroye, L. Györfi, L. and Lugosi, G., (1996). *A probabilistic theory of pattern recognition*. Springer-Verlag, New York.

Kurzyński, M., (1988). On the multistage Bayes classifier, *Pattern Recognition*, 21(4), pp.355-365.

Libal, U., (2010). Multistage classification of signals with the use of multiscale wavelet representation. In *MMAR'10, 15th IEEE Int. Conference on Methods and Models in Automation and Robotics*, pp.154-159.

Libal, U., (2012). Multistage pattern recognition of signals represented in wavelet bases with reject option. In *MMAR'12, 17th IEEE Int. Conference on Methods and Models in Automation and Robotics*, pp.79-84.

Mallat, S.G., (1989). A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7), pp.674-693.

Pudil, P., Novovicova, J., Blaha, S., Kittler, J., (1992). Multistage pattern recognition with reject option. In *Pattern Recognition, Vol.II. Conference B: 11th IAPR Int. Conference on Pattern Recognition Methodology and Systems*, pp.92-95.

APPENDIX

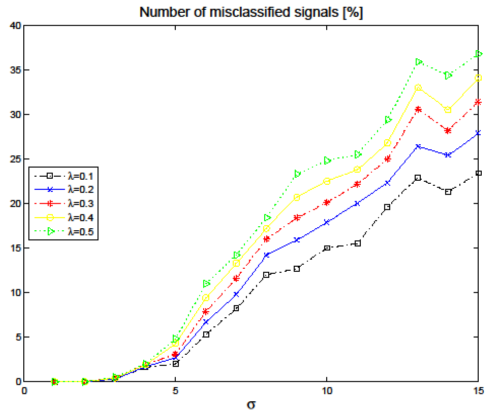

Figure 5: Risk for naive Bayes classifier.

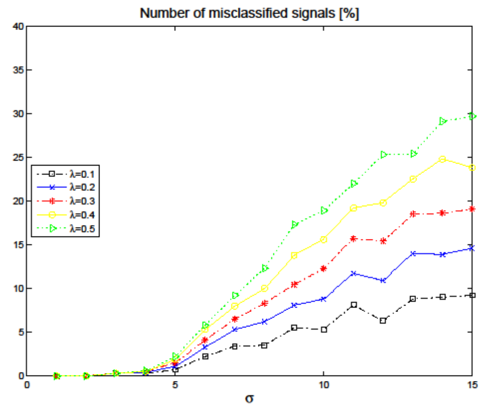

Figure 6: Risk for linear discriminant analysis (LDA).

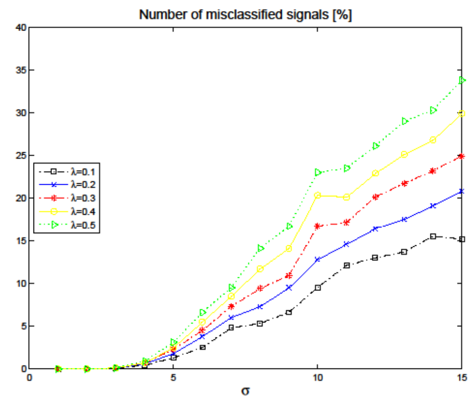

Figure 7: Risk for quadratic discriminant analysis (QDA).