A performance evaluation of AdaBoost and SVM forFace and Cancer Classification

Conference Paper · January 2017

CITATIONS
0

READS
210

4 authors, including:

Mohammed Ngadi
Université Ibn Tofail
20 PUBLICATIONS **6** CITATIONS
SEE PROFILE

Aouatif Amine
Université Ibn Tofail
89 PUBLICATIONS **246** CITATIONS
SEE PROFILE

Hanaa Hachimi
Université Ibn Tofail
73 PUBLICATIONS **32** CITATIONS
SEE PROFILE

A performance evaluation of AdaBoost and SVM for Face and Cancer Classification

Mohammed NGADI[1], $Aouatif\ AMINE^2$, $Bouchra\ NASSIH^3$, $Hanaa\ HACHIMI^4$
Systems Engineering Laboratory
National School of Applied Sciences
Ibn Tofail University Kenitra Morocco
$^1Ngadi.mohammed@univ-ibntofail.ac.ma$
$^2amine\_aouatif@univ-ibntofail.ac.ma$
$^3Nassih.bouchra@univ-ibntofail.ac.ma$
$^4hanaa.hachimi@univ-ibntofail.ac.ma$

***Abstract*—In this paper, we present a comparative study of Support Vector Machine (SVM) and Adaboost, as being two decision based classification tools in the field of shape recognition. The aim of our work is to study their theoretical foundations, their learning algorithms and to see their performance in classification capacity. In order to evaluate the contributions of our approach, we carried out a series of experiments on two famous training datasets widely used by the community, namely the MIT-CBCL face and Wisconsin diagnosis breast cancer (WDBC). The quality of decision of each classifier depends on the choice of its parameters and its implementation. The results of these experiments are promising and confirm the advantage of using SVM and Ababoost.**

***Keywords*—supervised learning, unsupervised learning, classification, shape recognition, SVM, Adaboost.**

I. Introduction

The field of pattern recognition knows a revolution since the mid-90s with the statistical learning theory and the advent of the Support Vector Machines (SVM) [1, 2, 3] for the resolution of detection problems, classification and regression. In recent years, appeared a set of highly interdependent disciplines, concerning the information treatment, decision theory and methods of pattern recognition, Boosting methods [4]. Their field of applications is much expanded and extended to several areas, in particular in: shapes recognition, the approximation of functions, image processing, speech recognition, classification...
Let this principle, we have found other methods of classifying and superficial overview on the principles of the first major approach inferred from a sample of examples classified a procedure (decision function) classification of new examples unlabelled. Supervised methods can be based on probabilistic assumptions (Naive Bayes) [5,6] or on concepts of proximity (nearest neighbors) [7,8] or even on research in areas of assumptions (decision trees [9], neural networks [10]).
This work is part of the development of medical systems to aid in diagnosis and face detection which has many direct applications in video surveillance, biometrics, robotics. It has been devoted to the study of techniques Of supervised classification. After presenting the theoretical of SVM and Adaboost, we spoke of the utilities in classification problems. The algorithms have been implemented under the Matlab and python environment. Finally, the simulation results obtained were presented and an evaluation of the designed system was made with a comparison of such algorithms. The results obtained were generally satisfactory, allowing us to achieve a classification rate close to 99%.

The rest of the paper is organized as follows: A brief description of Adaboost and SVM is given in Section 2 and 3. The experimental results are presented in Section 4, while Section 5 concludes the paper.

II. Adaboost

The basic idea of Adaboost [4] is to combine simple "rules" (assumptions) to create an ensemble for which the performance of each member is amplified. The ensemble composed by hypotheses is defined as follows:
Let $h_1, h_2, ..., h_t$ be a simple ensemble hypothesis, and consider the set of composite hypothesis:$f(x) = \sum_{t=1}^{T} \alpha_t h_t(X)$
Where α_t is the weight ascribed to the hypothesis h_t of the ensemble. The weights α_t and the assumptions h_t must be trained for the Boosting procedure.
In general, there are several possible approaches to select α_t coefficients and h_t assumptions. When Boosting, α_t and h_t are selected iteratively with weighted training examples. At each iteration, the weights of training examples are recomputed so as to attribute a high weight to misclassified learning examples and a low weight to the others . This technique helps focus the learning process on the examples that are hard to classify. The final classifier (most powerful) is the combination of all the weak classifiers.

III. Support Vector Machines (SVM)

The method of Large Margins separators [11] is a very general classification technique for determining a boundary between two classes. This boundary is defined by the principle of structural risk minimization made by Vapnik.

Let $D = x_1, ..., x_N$ a set of N samples examples labeled by $y_i = \{-1, 1\}$ according to the corresponding class of the sample x_i. The purpose of this method is to find a separating hyperplane that maximizes the margin between the closest samples of each class.
To describe the technique of construction of the optimal hyperplan separating data belonging to two different classes, suppose we have the empirical data:

$$(x_1, y_1), ..., (x_i, y_i) \in R^N \times \{\pm 1\} \quad (1)$$

Let $f(x) = w.x + b$ be the hyperplan that satisfy the following conditions:

$$y_i(w.x_i + b) \geq 1, for i = 1, ..., n \quad (2)$$

The optimal hyperplane is obtained by resolving the following quadratic optimization problem:

$$min \frac{1}{2} \parallel w^2 \parallel$$
$$\forall i, y_i(w.x_i + b) \geq 1 \quad (3)$$

Using the method of Lagrange multiplier allows us to formulate the problem as the maximization of the Lagrangian:

$$L(w, b, \alpha) = min \frac{1}{2} \parallel w^2 \parallel - \sum_{i=1}^{n} [\alpha_i (y_i (w.x_i + b) - 1)] \quad (4)$$

By substituting in the Lagrangian, we eliminate the variable w and b and we obtain the Lagrangian dual:

$$max[\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{j,i=1}^{n} \alpha_i \alpha_j y_i y_j < x_i, x_j >]$$
$$\forall i, \alpha_i \geq 0, \sum_{i=1}^{n} \alpha_i y_i = 0 \quad (5)$$

By resolution of the dual optimization problem, the coefficients α_i are obtained, necessary for the expression of the vector w and one can, therefore, establish the following decision function:

$$f(x) = \sum_{i=1}^{n} \alpha_i^* y_i x_i . x + b^* = 0 \quad (6)$$

Points x_i with $\alpha_i > 0$ are called support vectors (SV). The classification rule for a new observation x based on the maximum margin hyperplane is given by[12]:

$$Sign(\sum_{i=1}^{n} y_i \alpha_i^* < x_i, x > + b^*) \quad (7)$$

IV. Experimental results

After deriving the mathematics behind each classifier, we will now test and compare the mentioned classifiers on real datasets. We have used two famous datasets very used by the community, namely the MIT-CBCL [1] face database and the Wisconsin diagnostic breast cancer WDBC [2] database.
The original MIT-CBCL face database (Fig.1) has 6,977 training images (with 2,429 faces and 4,548 nonfaces) and 24,045 test images (472 faces and 23,573 nonfaces).

Fig. 1. A subset of MIT-CBCL face dataset used for classification.

Breast cancer detection experiments has been carried out using the Wisconsin Diagnosis Breast Cancer WDBC dataset created by Dr. William H. Wolberg at the University of Wisconsin. This dataset consists of 569 observations of patients with breast cancer among which 357 are benign and 212 are malignant status (Fig.2). Each instance has 32 features including id number and the class label that correspond to the type of breast cancer being benign or malignant. These features are computed from digital image of ne needle of aspirates (FNA) of breast masses that describes the characteristics of the cell nuclei in the image (Fig.3).

Breast Cancer Wisconsin (Diagnostic) data set			
Type	Classification	Origin	Real world
Features	30	(Real / Integer / Nominal)	(30 / 0 / 0)
Instances	569	Classes	2
Missing values?			No

Fig. 2. WDBC data set.

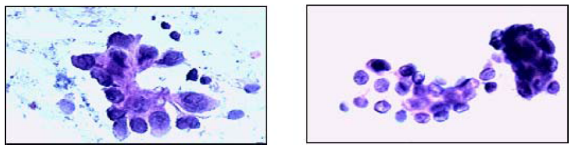

Fig. 3. Fine Needle Biopsies of Breast. Malignant (left) and benign (right).

A. Adaboost

The method of Adaboost is particularly interesting because we can choose the number of classifiers in order to achieve the desired error rates on samples examples. Moreover, we observe that the error rate decreases exponentially with the number of used weak classifiers, allowing us to achieve a classification rate close to 99% (Fig.4 and Fig.5).

[1] http://faculty.ucmerced.edu/mhyang/face-detection-survey.html

[2] https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)

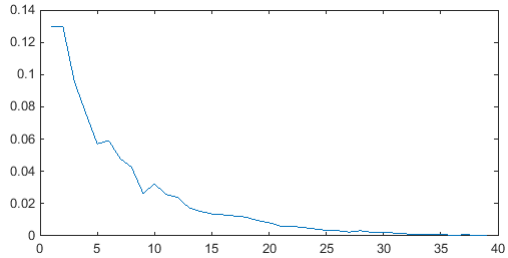

Fig. 4. Classification error versus number of weak classifiers"MIT-CBCL"

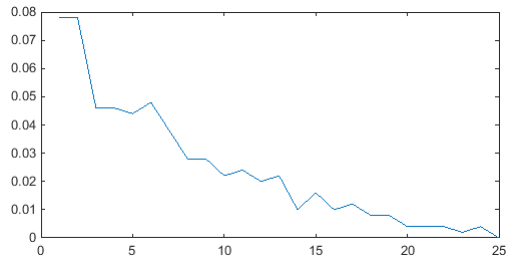

Fig. 5. Classification error versus number of weak classifiers "WDBC"

B. SVM

We have used LIBSVM [13] to measure the classication accuracy using SVM. A typical use of LIBSVM involves two steps: first, training a dataset to obtain a model and second using the model to predict information of a testing dataset. As kernel function, we have used a Gaussian radial basis function (RBF).

TABLE I. SHOWS THAT THE BEST CLASSIFICATION ACCURACY OF THE LIBSVM.

Dataset	Kernel	Accuracy (%)
MIT-CBCL	RBF	99.95
WDBC	RBF	96.95

C. Discusion

Classifiers used for our experiments under Python are K Nearest Neighbours (KNN), Linear SVM, decision tree, Random Forest, Adaboost and Naive Bayes.
The different Classifiers have been constructed on two data sets. ROC curves of MIT-CBCL and WDBC have been illustrated in the Fig.6 and Fig.7.

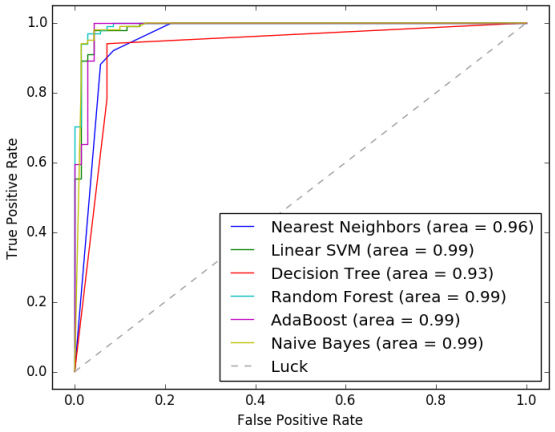

Fig. 6. Receiver operating characteristic on the MIT-CBCL dataset

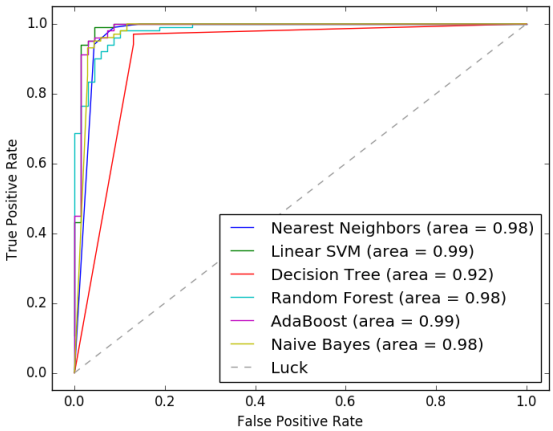

Fig. 7. Receiver operating characteristic on the WDBC dataset

In this study, the WDBC and MIT-CBCL datasets were used with several classification algorithms. It follows that most of them gives good precision results. However, we can see clearly that the SVM and Adaboost outperforms other classifiers regardless of the given base (99%).Not to mention the ability to efficiently manage practical applications where learning data can come from different environments.

V. CONCLUSION

In this article, we have presented a detailed comparative study of two classification algorithms, Adaboost and SVM. First, we started with the historical development of each classifier. Secondly, we applied them to two famous training datasets, widely used by the community, namely the MIT-CBCL face and the Wisconsin diagnosis breast cancer datasets. The main criteria that we used for comparison is the accuracy of the classification.
The experimental results under MATLAB and Python show that Adaboost and SVM perform better than other learning

algorithms on all the data that we have used.
Our goal in the near future is to continue the study of SVM and Adaboost in order to test the relationships that exist between them. Then we will try to establish the relationship between the dimensionality reduction and find the best compromise between precision and execution time.

REFERENCES

[1] Kim, S.Pang,M.Je. Constructing support vector machine ensemble. Pattern Recognition, (36), pp.2757-2767, 2005.

[2] I.Buciu. Demonstrating the stability of support vector machines for classification. Signal Processing. (86), pp,2364-2380, 2006.

[3] X.Li, L.Wang, E.Sung. A Study of AdaBoost with SVM Based Weak Learners. Proceedings of International Conference on Neural Networks, pp.196-200, 2005.

[4] Cao Ying, Miao Qi-Guang, Liu Jia-Chen, Gao Lin. Advance and prospects of AdaBoost algorithm. Acta Automatica Sinica, 39(6), pp.745-758, 2013.

[5] Guozhong Feng, Jianhua Guo, Bing-Yi Jing, Tieli Sun. Feature Subset Selection Using Naive Bayes for Text Classification, Pattern Recognition Letters, 2015.

[6] Kharya Shweta, Agrawal Shika, Soni Sunita. Naive Bayes Classifiers: A Probabilistic Detection Model for Breast Cancer.International Journal of Computer Applications. 92(10), 2014.

[7] Zhun-ga Liu, Quan Pan, Jean Dezert. A new belief-based K- nearest neighbor classification method. Pattern Recognition, 46, pp.834-844, 2013.

[8] Medjahed Seyyid Ahmed, Saadi Tamazouzt Ait, Benyettou Abdelkader. Breast Cancer Diagnosis by using k-Nearest Neighbor with Different Distances and Classification Rules. International Journal of Computer Applications 62(1), 2013.

[9] D.Lavanya, K.Usha Rani. ENSEMBLE DECISION TREE CLASSIFIER FOR BREAST CANCER DATA.International Journal of Information Technology Convergence and Services, pp.17-24, 2012.

[10] Rodriguez-Galiano, et al. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and..., Ore Geol. Rev. 2015.

[11] Abedi, M., Norouzi, G.H., Bahroudi, A. Support vector machine for multi-classification of mineral prospectivity areas. Comput. Geosci. 46, pp.272-283, 2012.

[12] Zuo, R., Carranza, E.J.M. Support vector machine: A tool for mapping mineral prospectivity. Computers Geosciences, 37, pp.1967-1975, 2011.

[13] Chang, Chih-Chung, and Chih-Jen Lin. LIBSVM: A library for support vector machines.ACM Transactions on Intelligent Systems and Technology (TIST), 2(3), 2011.