Universidad Carlos III de Madrid

Escuela Politécnica Superior

Grado en Ingeniería de Sistemas de Comunicaciones

Abstract

Neural classifiers with unbalanced binary problems

Autora: Eva Pérez Íñigo
Tutor: Marcelino Lázaro Teja

Julio 2016

Capítulo 1

Abstract

1.1. Motivation

This Degree's Final Project is developed in the general framework of Artificial Intelligence (AI), and particularly, within the scope of machine learning. AI intends to equip machines with the ability to solve problems through human intelligence paradigm. To do so, different computational and statistical tools are used. These have advanced and developed new techniques for decades. However, years ago big amounts of data weren't available. That fact has changed thanks to Internet and present data storage capacity. Due to this, to improve machine learning is a possibility.

Many companies already use machine learning to improve their decisions. Some like BlackRock (Williams, 2015) use data from Google and Twitter to take decisions on investments. World's most powerful companies, such as Google, Apple, Facebook, Microsoft, etc. (Rodríguez, 2016) have invested for years in investigation for the implementation of applications based on AI. Some examples are Siri from Apple or Cortana from Microsoft, which work as virtual assistants.

Out of the economic sphere, these techniques also help in critical aspects such as saving lives through medical diagnosis (Kononenko, 2001), with applications such as the analysis of electrocardiogram signals.

The amount of money involved in these applications lead the decisions of both private and public companies to invest on them.

One of the recurring problem of AI and machine learning is pattern classification (Duda et al., 2001), that allows machines to decide when facing different stimuli. Within classification, there are numerous occasions in which data are unbalanced, that is, the number of samples of each class might be sensibly different. In addition, in these cases, the class with the least number of samples usually is the most important to be correctly detected. An example of this kind of problems might me fraud detection, in which, a greater number of legal use samples are available compared to the ones of fraud crimes.

About regulatory framework, this will depend fundamentally on the concrete application on which the application is used. The main aspects to take into consideration are: data confidentiality, patents for used methods, and software licenses.

Ultimately, AI, and therefore, classification, is an instrument of great utility and social impact, and it is expected to be omnipresent in a not so distant future, since its applications only have a limit in imagination.

1.2. Objetives

The present project is developed in the framework of pattern classification when data are unbalanced, that is, when the number of samples of each class is significantly different.

Regarding this kind of problem, it is intended to be solved using diferent classification techniques, analyzing which offers better results.

Three main objectives have been proposed:

- To implement a neural classifier, and specifically, a multi-layer perceptron.
- To evaluate supervised learning methods with modifications to confront unbalanced problems. A method based on minimization of weighted mean square error, and a method based on Bayesian formulation will be tested.
- To evaluate the combination of individual classifiers as a method of classification, intending to increase the performance of this kind of problems.

In addition to the previous objectives, real unbalance databases will have to be obtained, and after the evaluation, a comparison between the methods will be carried out.

1.3. Problem Statement

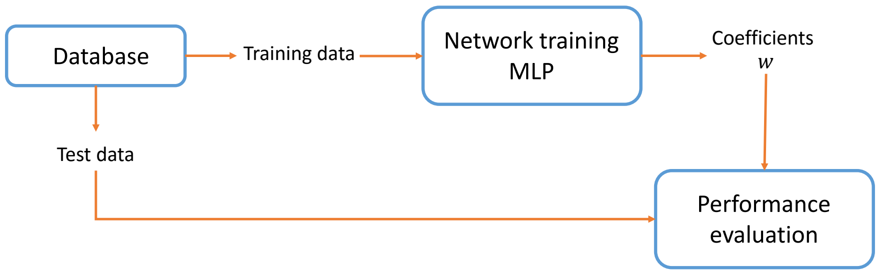

Figura 1.1: Basic scheme of the methodology for the evaluation of different sorting methods.

A set of 8 databases with unbalanced data obtained from KEEL repository (Alcalá-Fdez et al., 2011) will be used. Subsequently, the network is trained with the set of training data to obtain the weights $\mathbf{w}$ that minimizes the cost function through the adaptive algorithm of gradient descent.

Once the neural networks are trained, and parameters $\mathbf{w}$ are obtained, the evaluation of each method is carried out by calculating the performance over the test set. For evaluation purposes, different merit figures have been used: ROC curve, and the area under it.

The methods for network training that will be evaluated are:

- Method based in the Weighted Mean Square Error, which cost function to be minimized is

$$J^{WMSE}(\mathbf{w}) = \frac{\alpha}{N_0} \sum_{k \in S_0} (y_k - o_k)^2 + \frac{1-\alpha}{N_1} \sum_{k \in S_1} (y_k - o_k)^2, \qquad (1.1)$$

  being S_0 and S_1 the sets of indexes for the patterns of each class, and where N_0 and N_1 are the number of patterns of class 0 and 1, respectively. In addition, α defines the relation of importance over the adjustment of the labels of the patterns for each class.

- Recently proposed method (Lazaro et al., 2015) based on Bayesian formulation, which cost function can be expressed by

$$J^{Bayes}(\mathbf{w}) = \alpha\, \hat{p}_{FA} + (1-\alpha)\, \hat{p}_{M}. \quad (1.2)$$

  in which $\hat{p}_{FA}$ and $\hat{p}_{M}$ are the estimation of the probabilities of false alarm and missing, respectively. The parameter α ponders the importance of failure over each of the two classes.

- Two combiners of individual classifiers. The individuals are trained with the previous Bayesian formulation. The first of them takes the decision according to the decisions of the majority of individual classifiers. The second one is constructed from a Bayesian decision rule exposed in (Lazaro et al., 2015).

1.4. Results and conclusions

After training the network with the different methodologies and evaluate them, they will be compared to extract the most relevant results.

Firstly, WMSE and Bayesian method are assessed when parameter α for both cost functions equals 0.5.

Database	1	2	3	4	5	6	7	8
Bayes	**0.9163**	**0.8924**	**0.8372**	**0.9737**	0.9119	**0.8774**	**0.9872**	**0.9411**
WMSE	0.812	0.8385	0.7211	0.8906	**0.9832**	0.7164	0.7847	0.9006

Tabla 1.1: AUC results of WMSE and Bayesian method for the 8 databases with $\alpha = 0.5$.

According to the results obtained, Bayesian method offers better performance than WMSE in all databases excepting database number 5. This one is the one with the least number of samples, and WMSE offers better results compared to the others.

Following, both methods have been trained for different working points, with $\alpha \in \{0.1, 0.2, \ldots, 0.9\}$

Database	1	2	3	4	5	6	7	8
Bayes	**0.8068**	**0.8715**	**0.8194**	**0.9728**	**0.9355**	**0.8666**	**0.9877**	**0.9216**
WMSE	0.7078	0.7296	0.6652	0.7502	0.8274	0.6658	0.7441	0.7085

Tabla 1.2: AUC results of WMSE and Bayesian method for the 8 databases with different values of α.

With the results obtained in table 1.2, the Bayesian method proves to obtain better results for every database. The variance of AUC for both methods across the databases is similar.

After this, Bayesian method is singlely compared using ROC curve for $\alpha = 0.5$, and other different values of α.

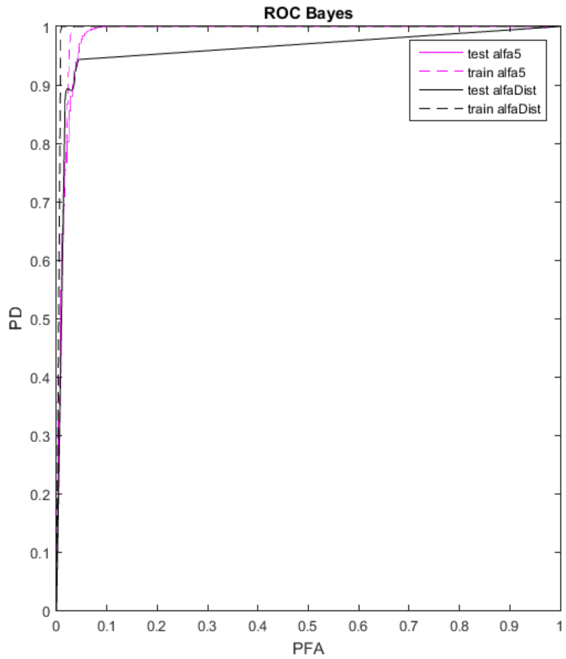

Figura 1.2: ROC curves for database 7.

If both ROCs are observed (see figure 1.2) for training data, the ROC obtained for different values of α is greater than the one obtained for a single $\alpha = 0.5$. However, when analyzing the test curves, an intersection between both curves is found in $\alpha = 0.5$. For values of α over 0.5, the performance in the curve with different values of α is better than the one of $\alpha = 0.5$. On the contrary, if the intention is to obtain a better adjustment for the labels of the minority class, the curve for $\alpha = 0.5$ overcomes the performance than the one with different values of α. This is due to a problem of overfitting.

Lastly, Bayes'AUC from table 1.2 is compared with the results of the two types of classifiers combiners.

Database	Individual	Combination by majority	Combination with Bayes
1	0.8068	0.8268	0.9199
2	0.8715	0.8725	0.8775
3	0.8194	0.8284	0.8499
4	0.9728	0.9782	0.9863
5	0.9355	0.9379	0.9543
6	0.8666	0.8778	0.8991
7	0.9877	0.9885	0.9912
8	0.9216	0.9294	0.9311

Tabla 1.3: AUC comparison of individual classifier and combinations of individual classifiers.

According to table 1.3, the performance of the combination of classifiers improves when compared to the individual classifier. Among the two possibilities of combination of classifier, the second one offers better results. In some databases the difference between both of them is very significant, as it is the case of database 1. This begs the question on whether is it worth to use more complex classifiers aiming for a slight improvement in performance. In the end, depending on the use given to it a classifier of greater complexity will be justified, according to the application's restrictions.

Bibliografía

A. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, y F. Herrera. "Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework.". págs. 255–287, 2011.

R. O. Duda, P. E. Hart, y D. G. Stork. *Pattern classification*. John Wiley & Sons, segunda edición. 2001.

I. Kononenko. "Machine learning for medical diagnosis: hystory, state of the art and perspective". *Artificial Intelligence in Medicine*, (23):89–109, 2001.

M. Lazaro, F. Herrera, y A. R. Figueiras-Vidal. "Classification of binary imbalanced data using a bayesian ensemble of bayesian neural networks". 2015.

M. Rodríguez. "Cómo están utilizando la inteligencia artificial las grandes empresas". software disponible en https://www.euroresidentes.com/tecnologia/avances-tecnologicos/como-estan-utilizando-la-inteligencia-artificial-las-grandes-empresas. 2016.

A. Williams. "Blackrock y google negocian una posible 'joint venture"'. software disponible en http://www.expansion.com/empresas/2015/11/09/5641228322601d960f8b4570.html. 2015.