An Ordinal Data Method for the Classification with Reject Option

Ricardo Sousa, Jaime S. Cardoso
INESC Porto, Faculdade Engenharia, Universidade Porto
Campus da FEUP, Rua Dr. Roberto Frias, n. 378
4200-465 Porto, Portugal
{rsousa,jaime.cardoso}@inescporto.pt

Abstract

In this work we consider the problem of binary classification where the classifier may abstain instead of classifying each observation, leaving the critical items for human evaluation. This article motivates and presents a method to learn the reject region on complex data. Observations are replicated and then a single binary classifier determines the decision plane. Our method is compared with standard techniques on synthetic and real datasets.

1. Introduction

Decision support systems are becoming ubiquitous in many human activities, most notably in finance and medicine. Automatic models are being developed to imitate, as closely as possible, the usual human decision [1]. One of the problems with classifying complex items is that many items from distinct classes have similar structures in a feature space, resulting in a setting with overlapping classes. The automation of decisions in this region leads invariably to many wrong predictions. On the other hand, although using labels $only$ as 'good' or 'bad', the deployment of a decision support system in many environments has the opportunity to label critical items for manual revision. Therefore, the development of tripartite classifiers, with a third output class, the reject class, in-between the good and bad classes, is attractive.

2. Problem Statement and Standard Solutions

The design of classifiers with reject option can be systematised in three different approaches: 1) the design of two, *independent*, classifiers. A first classifier is trained to output $\mathcal{C}_{-1}$ only when the probability of $\mathcal{C}_{-1}$ is high and a second classifier trained to output $\mathcal{C}_{+1}$ only when the probability of $\mathcal{C}_{+1}$ is high. 2) the design of a single, standard binary classifier (SBC). If the classifier provides some approximation to the a posterior class probabilities, then a pattern is rejected if the maximum of the two posterior probabilities is lower than a given threshold. If the classifier does not provide probabilistic outputs, then a rejection threshold targeted to the particular classifier is used. 3) the design of a single classifier with embedded reject option. This approach has consisted in the design of algorithms specifically adapted for this kind of problems [3].

3. An Ordinal Data Approach for Detecting Reject Regions

The rejection method to be presented—rejoSVM—is an extension of a method already proposed in the literature but for the classification of ordinal data. For more detail about this method, the reader should consult [2].

3.1. The Data Replication Method for Detecting Reject Regions

The scenario of designing a classifier with reject option shares many characteristics with the classification of ordinal data. It is also reasonable to assume for the reject option scenario that the three output classes are naturally ordered as $\mathcal{C}_1, \mathcal{C}_{reject}, \mathcal{C}_2$. In the scenario of designing a classifier with reject option, we are interested on finding two boundaries: a boundary discriminating $\mathcal{C}_1$ from $\{\mathcal{C}_{reject}, \mathcal{C}_2\}$ and a boundary discriminating $\{\mathcal{C}_1, \mathcal{C}_{reject}\}$ from $\mathcal{C}_2$.

We proceed exactly as in the data replication method for ordinal data. We start by transforming the data from the initial space to an extended space, replicating the data, according to the rule:

$$\mathbf{x} \in \mathbb{R}^d \begin{array}{l} \nearrow \left[\begin{smallmatrix} \mathbf{x} \\ h \end{smallmatrix}\right] \in \mathbb{R}^{d+1} \\ \searrow \left[\begin{smallmatrix} \mathbf{x} \\ 0 \end{smallmatrix}\right] \in \mathbb{R}^{d+1} \end{array}, \text{ where } h = \text{const} \in \mathbb{R}^+$$

If we design a binary classifier on the extended training data, without further considerations, one would obtain the

same classification boundary in both data replicas. Therefore, we modify the misclassification cost of the observations according to the data replica they belong to. In the first replica (the extension feature assumes the value zero), we will discriminate $\mathcal{C}_1$ from $\{\mathcal{C}_{reject}, \mathcal{C}_2\}$; therefore we give higher costs to observations belonging to class $\mathcal{C}_2$ than to observations belonging to class $\mathcal{C}_1$. This will bias the boundary towards the minimisation of errors in $\mathcal{C}_2$. Similar approach is conducted for the second replica.

3.2. Selecting the Misclassification Costs

The typical adoption of the same cost for erring and rejecting on the two classes is as follows: assign C_{low} cost when classifying a class as reject and assign C_{high} cost when misclassifying. Therefore, $C_{reject} = \frac{C_{low}}{C_{high}} = w_r$ is the cost of rejecting (normalised by the cost of erring). The data replication method with reject option tries to minimizes the empirical risk $w_r R + E$, where R accounts for the rejection rate and E for the misclassification rate.

4. Experimental Results

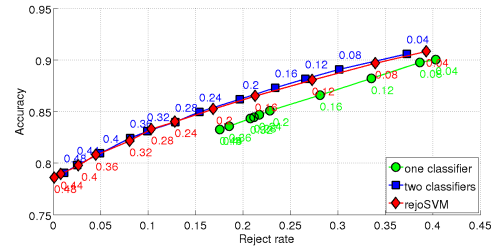
The performance of the classification methods were assessed over two datasets. As in [2], for the synthetic dataset, we began by generating 400 example points $\mathbf{x} = [x_1\ x_2]^t$ in the unit square $[0,1] \times [0,1] \subset \mathbb{R}^2$ according to a uniform distribution. Then, we assigned to each example $\mathbf{x}$ a class $y \in \{-1, +1\}$ corresponding to

$$\begin{aligned} (b_{-2}, b_{-1}, b_0, b_1) &= (-\infty; -0.5; 0.25; +\infty) \\ \varepsilon_1 \sim N(0, 0.125^2),\ \alpha &= 10(x_1 - 0.5)(x_2 - 0.5) \\ t = \min_{r \in \{-1,0,+1\}} &\{r : b_{r-1} < \alpha + \varepsilon_1 < b_r\} \\ \varepsilon_2 \sim Uniform(b_{-1}, b_0),\ y &= \begin{cases} t & t \neq 0 \\ +1 & t = 0 \wedge \varepsilon_2 < \alpha \\ -1 & t = 0 \wedge \varepsilon_2 > \alpha \end{cases} \end{aligned} \tag{1}$$

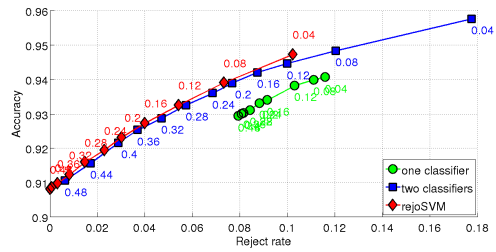
The second dataset encompasses 960 observations [1] expressing the aesthetic evaluation of Breast Cancer Conservative Treatment (BCCT). Here, we aggregated Excellent and Good into one class, and the Fair and Poor cases into another.

We randomly split each dataset into training and test sets, with 5% and 95% of the data, respectively. The splitting of the data into training and test sets was repeated 100 times in order to obtain more stable results for accuracy by averaging and also to assess the variability of this measure. A 5-fold cross validation was performed. The performance was represented by an Accuracy-Reject curve where each different point of the A-R curve corresponds to different values of w_r.

Figure 1 summarises the results obtained for all three methods on the datasets. A first main assertion is that rejoSVM performs better than the simpler solution based on a single classifier. Comparing the rejoSVM with the two independent classifiers approach, neither of the two techniques outperformed the other, although rejoSVM benefits of simplicity and interpretability.



(a) Synthetic dataset.



(b) BCCT dataset.

Figure 1. The A-R curves for the three datasets.

5. Conclusion

Here we expose an extension of the data replication method [2] that directly embeds reject option taking a new perspective by viewing the three output classes as naturally ordered. Also, this method has the advantages of using a SBC and embedding the design of the reject region during the training process.

References

[1] J. S. Cardoso and M. J. Cardoso. Towards an intelligent medical system for the aesthetic evaluation of breast cancer conservative treatment. *Artificial Intelligence in Medicine*, 40:115–126, 2007.

[2] J. S. Cardoso and J. F. P. da Costa. Learning to classify ordinal data: the data replication method. *Journal of Machine Learning Research*, 8:1393–1429, 2007.

[3] G. Fumera and F. Roli. Support Vector Machines with Embedded Reject Option. In *SVM '02: Proceedings of the First International Workshop on Pattern Recognition with Support Vector Machines*, pages 68–82, London, UK, 2002. Springer-Verlag.