Prediction of Probability of Survival in Critically Ill Patients Optimizing the Area Under the ROC Curve

By Oscar Luaces, José R.Quevedo, Francisco Taboada, Guillermo M.Albaiceta, Antonio Bahamonde

Lina YE

Abstract: This article presents the method of Support Vectors Machines (SVM) for predicting probability of survival in critically ill patients by using Platt's method to fit a sigmoid[1]. Moreover, the comparison of two interesting exploitations is provided: the maximization of the AUC and the minimization of the error rate in binary classification tasks.

Introduction: This paper insists on using the method of SVM for learning probabilities that is very important in the assessment of the efficiency of intensive care units (ICU) treatments. In addition, AUC is used here to measure the performance of prediction by estimating the degree of coherence between a continuous output and a binary classification, which is quite fruitful, especially when classes are greatly unbalanced where the accuracy is usually inadequate.

Objective: The crucial step in learning a probability distribution by using SVM is to transform the continuous outputs into probabilities with Platt's method to fit a sigmoid. For this, the supposition in this paper is that to compute the sigmoid, the optimization of AUC first is preferable to the minimization of the error rate. So the objective of this article is to propose a new method for learning probabilities that optimize AUC and to then support it by experimental evidence.

Methods: Four approaches for learning probabilities are compared by using the collection of data sets of survival probabilities in critically ill patients: the accuracy optimizer, which is the standard method based on SVM, called SVM(Accu); the AUC optimizer, which is the proposal of this article, presented by SVM(AUC); the regression approach that is the straightforward approach using regression, used as a baseline for measuring the merits of the other options; and the fourth predictor, the commercial system APACHE III[2]. It is important to note that SVM (AUC) sets its parameters to optimize AUC, which is opposition to SVM (Accu) and SVR whose parameters were set to optimize the Brier score. For the comparison, two measures of the performance of probability estimations used here are AUC and Brier score in respect that AUC is considered as the probability of a correct ranking and that Brier score is the average of quadratic deviations of true and predicted probabilities.

Results: The results obtained by the three support vector approaches(except APACHE III) show that in general, the multivariate SVM(AUC) has the best performance concerning both AUC measure and Brier scores measure. The fact that SVM(AUC) outperforms both SVM(Accu) and SVR in terms of the Brier score is surprising, considering that two latter approaches set their parameters to optimize this score, while SVM(AUC) is to optimize the AUC measure. Another interesting point is found in the results is that if more training cases are used, the performance of SVM and SVR could be both improved. So the explication of the great difference between the three support vector approaches and the commercial system APACHE III could be that the latter was trained with data sets that were several times bigger.

[1] John Platt : Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, (2000)
[2] Rivera-Fernandez et al: The APACHE III prognostic system : customized mortality predictions for Spanish ICU patients, (1998)

Conclusion: From the methods and the results, it is reasonable to conclude that looking for an optimum AUC first is better than minimizing the error rate with a classification SVM. It is necessary to point out that the method proposed in this article outperforms the standard SVM, especially in the case that the available data is scarce.

Comments

This article is well written, especially as regards the organization of its contents. Although the proposal of the new learning method for estimating probabilities is to resolve one real world problem that is the prediction of survival in critically ill patients, it is easy to generalize this approach to other learning tasks. Moreover, it seems that SVM (AUC) can be seen as a direct generalization of the conventional classification SVM. In others words, when using error rate as the loss function, SVM (Accu) arises as a special case of SVM (AUC).

However, it is important to notice some defects of this article when reading it. First of all, in regard to the experimentation, the comparison is unfair, since the system APACHE III was trained with data sets that were several times bigger than the other three support vector approaches. In fact, the size of the available data sets has important impact on the performance of the approach for estimating probabilities.

secondly, it is doubtful that the hypothesis space used by SVR is adequate enough to induce probability distributions from a reduced set of training data, considering the fact that although the optimization problem posed to SVR is precisely the minimization of the distance between true and predicted probabilities, a large amount of data is required to tie the scores of SVM (AUC) in the Brier score.

Finally, it seems that the predominance of the new method SVM (AUC) is restricted to the cases when classes are very unbalanced, the misclassification rate is usually inadequate. It is regretted that there is no evidence in this article of superiority of SVM (AUC) in the case of the unbalance of the classes.