Bayesian Classifiers for Predicting the Outcome of Breast Cancer Preoperative Chemotherapy

Antônio P. Braga[1], Euler G. Horta[1], René Natowicz[2], Roman Rouzier[3],
Roberto Incitti[4], Thiago S. Rodrigues[5], Marcelo A. Costa[1],
Carmen D.M. Pataro[1], and Arben Çela[2]

[1] Universidade Federal de Minas Gerais, Depto. Engenharia Eletrônica, Brazil
[2] Université Paris-Est, ESIEE-Paris, France
[3] Hôpital Tenon, Service de gynécologie, France
[4] Institut Mondor de Médecine Moléculaire, Plate-forme génomique, France
[5] Universidade Federal de Lavras, Depto. Ciência da Computação, Brazil

Abstract. Efficient predictors of the response to chemotherapy is an important issue because such predictors would make it possible to give the patients the most appropriate chemotherapy regimen. DNA microarrays appear to be of high interest for the design of such predictors. In this article we propose bayesian classifiers taking as input the expression levels of DNA probes, and a 'filtering' method for DNA probes selection.

1 Introduction

We consider the question of predicting the response of the patients to neoadjuvant chemotherapy (treatment given prior to surgery). A pathologic complete response (PCR) at surgery is correlated with an positive outcome while residual disease (NoPCR) is associated with a negative outcome. An accurate prediction of tumor sensitivity to preoperative chemotherapy is important to avoid the prescription of an inefficient treatment to patients with predicted residual disease while allocating the treatment to all the PCR patients

In [4,5] we have proposed a new method for DNA probes selection. In this paper we investigate the performances of bayesian classifiers taking as input the expression levels of the selected probes. We also compare the results with SVMs [2] and Multi-objective neural networks [6].

2 Previous Studies

The data comes from a clinical trail jointly conducted at the MD Anderson Cancer Center (MDACC) in Houston (USA) and at the Institut Gustave Roussy (IGR) in Villejuif (France) and it was used previously in K. Hess & al. [3]. The dataset was composed of 82 cases in the *training set* and 51 cases in the *test set*. A pathologic complete response (PCR) was defined as no histopathologic evidence of any residual invasive cancer cells in the breast, and a non pathologic complete response as any residual cancer cells after histopathologic study. For

each patient case, the data was the outcome of the treatment and the expression levels of 22283 DNA probes of an Affymetrix U133A microarray, measured on tumor cells.

Method of probes selection [5,4]. We assigned two sets of expression levels to any probe s, the sets $E_p(s)$ and $E_n(s)$, computed from the training data as follows [5]. Let $m_p(s)$ et $sd_p(s)$ be the mean and standard deviation of the expression levels of probe s for the PCR training cases, and let $m_n(s)$ and $sd_n(s)$ be that of the NoPCR training cases. The set of expression levels of the PCR training cases was defined as the set difference $E_p(s)$,

$$E_p(s) = [m_p(s) - sd_p(s), m_p(s) + sd_p(s)] \setminus [m_n(s) - sd_n(s), m_n(s) + sd_n(s)]$$

and conversely for the NoPCR training cases,

$$E_n(s) = [m_n(s) - sd_n(s), m_n(s) + sd_n(s)] \setminus [m_p(s) - sd_p(s), m_p(s) + sd_p(s)].$$

For any patient case, the individual prediction of a probe was a discrete value in set $\{pcr, nopcr, unspecified\}$: *pcr* if the expression level of patient p lied within the interval $E_p(s)$ and *nopcr* if it lied within $E_n(s)$. Otherwise, the individual prediction value was *unspecified*.

Let $p(s)$ be the number of PCR training cases correctly predicted *pcr* by probe s, and let $n(s)$ be the number of the NoPCR training cases correctly predicted *nopcr* by the probe. The valuation function of the probes was defined so as to favor probes which correctly predicted high numbers of training cases and moreover, whose sets of correctly predicted training cases were 'good' samplings of the training set. To this end, we have considered the ratios $p(s)/P$ and $n(s)/N$ of correctly predicted training cases. The valuation function $v(s)$, $v(s) \in [0,1]$, was defined as:

$$v(s) = 0.5 \times \left(\frac{p(s)}{P} + \frac{n(s)}{N} \right) \tag{1}$$

In our previous studies ([4,5]) we had defined the k-probes majority decision predictor as the set of the k top ranked probes together with the majority decision criterion: for any patient case, when the majority of 'pcr' and 'nopcr' predictions of the k top ranked probes was 'pcr', the patient was predicted to be 'PCR', and when the majority was 'nopcr' the patient was predicted to be 'NoPCR'. In case of tie the patient was predicted 'UNSPECIFIED'. When computing the performances of the predictor, a false negative was a PCR patient case predicted NoPCR or UNSPECIFIED, and conversely for the false positives.

The best k-probes predictor was for $k = 30$ probes. This number of probes was in accordance with the result of a previous study by K. Hess & al. [3].

3   Bayes Classification and Results

In order to obtain the majority classification rule, the marginal distributions of each one of the 30 probes were estimate separately for the two classes, PCR

Table 1. Comparison between all the classifiers on the test set

	Hess et al	Majority decision	Bayes		Bayes discretized		SVM	MOBJ NN
			equal priors	diff. priors	equal priors	diff. priors		
accuracy	0.76	0.86	0.88	0.86	0.88	0.82	0.88	0.88
sensitivity	0.92	0.92	0.84	0.77	0.92	0.38	0.92	0.92
specificity	0.71	0.84	0.89	0.89	0.87	0.97	0.87	0.87
PPV	0.52	0.67	0.73	0.71	0.71	0.83	0.71	0.71
NPV	0.96	0.97	0.94	0.92	0.97	0.82	0.97	0.97

(class C_{+1}) and NoPCR (class C_{-1}). Therefore, for each probe i the distributions $P(x_i|C_{+1})$ and $P(x_i|C_{-1})$ are known in advance. The majority decision rule can be viewed as:

$$\prod_{i=1}^{30} \frac{P(x_i|C_{+1})}{P(x_i|C_{-1})} = \frac{n_{+1s}}{n_{-1s}} > 1 \tag{2}$$

where n_{+1s} and n_{-1s} are, respectively, the number of +1s and -1s in the final discretized vector of the majority decision rule.

So, we can consider the majority rule as a Bayes rule with equal prior probabilities ($P(C_{+1}) = P(C_{-1})$). The effect of this is to *force a linear separation* in the input space. If the priors were used, this linear surface would *bend* in the direction of the minority class, since the NoPCR class is more likely and has larger variance.

The likelihoods $P(\mathbf{x}|C_{+1})$ and $P(\mathbf{x}|C_{-1})$ were estimated by considering independence between the probes, by $P(\mathbf{x}|C_{+1}) = \prod_{i=1}^{30} P(x_i|C_{+1})$ and $P(\mathbf{x}|C_{-1}) = \prod_{i=1}^{30} P(x_i|C_{-1})$, where $P(x_i|C_{+1})$ and $P(x_i|C_{-1})$ are, respectively, the marginal distributions of probe i for classes C_{+1} and C_{-1}.

The Bayesian rule [1] that assigns a pattern to class C_{+1} is given by Equation 3.

$$\frac{\prod_{i=1}^{30} P(x_i|C_{+1})Np_{+1s}}{\prod_{i=1}^{30} P(x_i|C_{-1})Np_{-1s}} > 1 \tag{3}$$

The results for Equations 2 and 3, with and without estimated prior probabilities for both discretized and non-discretized input vectors are presented in Table 1. SVMs and Multi-Objective neural networks results [4] are also presented in Table 1.

As can be observed in Table 1, the results of SVM and MOBJ-NN were identical. This can be explained by the low complexity models, bearing linear separation, yielded by these models. The resulting norm of the MOBJ neural network weights was very small, what suggests a strong smoothing effect at the network output.This interpretation is also supported by the identical results achieved by the Bayes rule with equal priors that tends to yield a smooth separation. It is surprising, however, that the Bayes rule with estimated prior probability has performed worse than Bayes with equal priors and the other models. This may be due to the small samples effect or to the independence assumption. Sparsity in the data set and small sample sizes may also favor smoother separating surfaces.

4   Conclusion

The trade-off between bias and variance [1] of a model is usually achieved
by complexity control, re-sampling or data-set partition strategies like cross-
validation [1]. SVMs and the MONJ-NN embody complexity control in their
formulations, what makes as believe that the results achieved with these models
point out to limit performance indexes that may be achieved with the current
data set. This assumption is reinforced by the fact that the results, obtained by
different training strategies, are very close to each other. For such a small data
set, representativeness in data set partition may be difficult to achieve, so we
rely on the results presented on Table 1 as reliable estimations of the separating
boundaries of the generator functions. We believe also that the use or not of esti-
mated priors in the Bayes/majority decision rule is a model design decision and
is not against any statistical principle. For our problem, the smoother decision
surface yielded results that are quite close to those obtained with the complexity
control models.

References

1. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, Heidelberg
   (2006)
2. Cortes, C., Vapnik, V.: Support vector networks. Machine Learning 20, 273–279
   (1995)
3. Hess, K.R., Anderson, K., Symmans, W.F., Valero, V., Ibrahim, N., Mejia, J.A.,
   Booser, D., Theriault, R.L., Buzdar, A.U., Dempsey, P.J., Rouzier, R., Sneige, N.,
   Ross, J.S., Vidaurre, T., Gomez, H.L., Hortobagyi, G.N., Pusztai, L.: Pharmacoge-
   nomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and flu-
   orouracil, doxorubicin, and cyclophosphamide in breast cancer. Journal of Clinical
   Oncology 24(26), 4236–4244 (2006)
4. Natowicz, R., Braga, A.P., Incitti, R., Horta, E.G., Rouzier, R., Rodrigues, T.S.,
   Costa, M.A.: A new method of dna probes selection and its use with multi-objective
   neural networks for predicting the outcome of breast cancer preoperative chemother-
   apy. In: European Symposium on Neural Networks (ESANN 2008) (Accepted, 2008)
5. Natowicz, R., Incitti, R., Charles, B., Guinot, P., Horta, E.G., Pusztai, L., Rouzier,
   R.: Prediction of the outcome of preoperative chemotherapy in breast cancer by
   dna probes that convey information on both complete and non complete responses.
   BMC Bioinformatics (Accepted, 2008)
6. Teixeira, R.A., Braga, A.P., Takahashi, R.H.C., Saldanha, R.R.: Improving gener-
   alization of mlps with multi-objective optimization. Neurocomputing (35), 189–194
   (2000)