What patient information allows us to make accurate predictions of outcome?

D. R. Lovell, C. R. Dance, M. Niranjan, R. W. Prager and K. J. Dalton‡
Cambridge University Engineering Department, Trumpington St., Cambridge CB2 1PZ, UK
Tel: +44 1223 332754, Fax: +44 1223 332 662, Email: {drl,crd,niranjan,rwp}@eng.cam.ac.uk

‡Cambridge University Department of Obstetrics and Gynaecology,
Rosie Maternity Hospital, Robinson Way, Cambridge, CB2 2SW, UK.

Abstract—**Only some of the information contained in a medical record will be useful to the prediction of patient outcome. We describe a novel method for selecting those outcome predictors which allow us to reliably discriminate between adverse and benign end results. Using the area under the receiver operating characteristic as a non-parametric measure of discrimination, we show how to calculate the maximum discrimination attainable with a given set of discrete valued features. This upper limit forms the basis of our feature selection algorithm. We use the algorithm to select features (from maternity records) relevant to the prediction of failure to progress in labour. The results of this analysis motivate investigation of those predictors of failure to progress relevant to parous and nulliparous sub-populations.**

Keywords— **Feature selection, risk prediction, pregnancy**

I. Introduction

TODAY'S health information systems allow medical professionals access to an immense amount of data. Prediction of patient outcome can be made on the basis of detailed information from many other similar cases. We tackle two questions: (1) What information is essential to make accurate predictions of outcome? (2) Given specific items of information about a number of patients, how accurately does that information allow us to discriminate between patients who have adverse or benign outcomes?

From the standpoint of pattern recognition, the first question describes the problem of *feature selection* [1, 2]. Many feature selection techniques rely upon statistical measures of the between-class separation provided by different feature sets [1, 3]. One drawback with this is that, while certain features may provide a high degree of class separation, there is no guarantee they form a useful representation for making predictions about *new* data. In a medical context, for example, one could achieve perfect separation between cases with adverse or benign outcome by using each patient's social security number as a descriptive feature. Obviously, this feature will be of no use in predicting outcomes for new patients. In this paper, we review a practical method of selecting those discrete valued features which allow us to reliably discriminate between two classes [4].

II. Maximum attainable discrimination

We represent retrospective patient information by a vector of discrete valued features (describing the patient), and a class label (to denote either an adverse or a benign outcome for that patient). This representation places patient records into what we term *bins*, *i.e.*, groups of records with the same attributes, but not necessarily the same class. This multinomial form makes no assumptions about the true distribution of data and also allows all interactions between features to be retained.

We use the *area under the receiver operating characteristic* (ROC) *curve* as a non-parametric measure of the discrimination achieved when cases are represented by a given set of discrete features [5]. When bins contain a mixture of benign *and* adverse outcome cases, it is impossible for any classification system to achieve perfect discrimination. However, the area under the ROC will be maximized by a prediction system which ranks bins in ascending order of likelihood of adverse outcome [5, 4, 6]. This is the means by which we calculate the *maximum attainable discrimination* for a specific data set, described by discrete valued features.

III. Estimating average expected discrimination

The area under the ROC curve can also be used as a measure of the discrimination attainable with *new* data. A classic way to estimate the expected performance of a system is to reserve, or "hold out" data for testing. If we employ this principle to estimate the maximum discrimination expected with a given set of discrete features, we must use two sets of data, D^1 and D^2, drawn from the population we wish to classify. We then

1. Maximize attainable discrimination with D^1 by placing its bins in ascending order of likelihood ratio.
2. Apply that ordering to the corresponding bins in D^2 then measure the discrimination obtained on D^2.

The area under the ROC curve for D^2 indicates how well the features representing the data allow us to predict the class of new data.

In practice, this estimate of expected discrimination is formed by randomly partitioning a data set, D, into two disjoint subsets, D^1, D^2. Estimates from a series of such random partitions can be averaged to form a *rotation error* estimate [2, p. 216] which mitigates the effects of sampling variability. The rotation error estimate can be used to assess the discriminatory power of subsets of discrete valued features via this *backwards sequential selection* algorithm:

$$W_{n+1} \quad = \quad W_n - \operatorname*{argmax}_{x_i \in W_n} d_{\text{avg}}(W_n - \{x_i\}), \qquad (1)$$

where the working set W_n initially contains all discrete features used to describe the data (*i.e.*, the *saturated* model of data). The $d_{\text{avg}}(W_n)$ function is our rotation error estimate of the discrimination expected when the features in W_n represent a given data set. Thus, features will be removed from the working set in order of increasing importance to attainable discrimination.

IV. Prediction of failure to progress

We used the backwards selection algorithm to choose features predictive of mothers who have a failure to progress in labour. Data describing 771571 singleton births was obtained from the Scottish Maternity Register (SMR). Since the SMR uses the International Classification of Diseases [7] to encode maternal conditions, we had to use univariate methods to pre-select a subset of 49 predictors — strongly associated with failure to progress — from over 1900 candidate features.

Table I shows the order in which some of the 49 predictors were removed from the saturated model in five different runs of the backwards selection algorithm. Unimportant features tend to be removed first whilst features that reliably discriminate between adverse and benign outcomes are retained to the last. Figure 1 plots the average expected discrimination ($d_{\text{avg}}(W_n)$) obtained at each step of the selection algorithm. Clearly, the most discriminating information is contained in a few features: parity, maternal age, Caesarian sections. This finding prompted further analysis of nulliparous and parous sub-groups, the results of which are also plotted in Figure 1. With parity "removed" in this way, different predictors (including height and anaemia in pregnancy) were seen to be more important in the nulliparous population.

V. Conclusions

We have given a brief overview of a new method for selecting discriminative discrete valued features. Tests have shown how this method can successfully determine features relevant to the prediction of failure to progress. We hope that other researchers will find this feature selection technique useful in different problem domains.

Acknowledgements

We are pleased to acknowledge the assistance of Dr Susan Cole of the Scottish Home and Health Department for collecting and providing us with access to Scottish Maternity Register data. This work was funded by the Quality Assurance in Maternity Care project, EC Biomed Grant Ref. BIOMED 1 CT93 1113. The first author gratefully acknowledges The British Council for providing travel funding for him to join the QAMC project as a research associate.

References

[1] J. Kittler, "Mathematical methods of feature selection in pattern recognition," *International Journal of Man-Machine Studies*, vol. 7, pp. 609–637, 1975.

[2] W. Siedlecki and J. Sklansky, "On automatic feature selection," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 2, no. 2, pp. 197–220, 1988.

[3] D. E. Boekee and J. C. A. Van der Lubbe, "Some aspects of error bounds in feature selection," *Pattern Recognition*, vol. 11, pp. 353–360, 1979.

[4] D. R. Lovell, C. R. Dance, M. Niranjan, R. W. Prager, and K. J. Dalton, "Limits on the discrimination possible with discrete valued data, with application to medical risk prediction," Tech. Rep. CUED/F-INFENG/TR243, Cambridge University Engineering Department, January 1996.

[5] D. Green and J. Swets, *Signal Detection Theory and Psychophysics*. New York: John Wiley, 1966.

[6] J. Hanley and B. McNeil, "Maximum attainable discrimination and the utilization of radiologic examinations," *Journal of Chronic Disease*, vol. 35, pp. 601–611, 1982.

[7] World Health Organization, *International Classification of Diseases, Ninth Revision*. Geneva: World Health Organization, 1975.

Feature and Range		Order of removal				
Social class	0-5	1	1	1	1	1
Blood group	0-4	2	2	2	2	2
Spontaneous abortions	0,1	3	3	3	3	3
Support	0,1	4	5	4	4	4
Theraputic abortions	0,1	6	4	5	5	5
GU TRACT INFECT IN PREG	0,1	7	8	6	6	7
EDEMA IN PREGNANCY	0,1	10	7	7	12	8
⋮		⋮	⋮	⋮	⋮	⋮
UNWANTED PREGNANCY NEC	0,1	38	40	38	21	42
ANEMIA IN PREGNANCY	0,1	44	44	45	45	44
OTH CURRENT COND OF PREG	0,1	42	43	16	41	43
OTH VENOUS COMPL IN PREG	0,1	43	39	37	37	41
Previous neonatal deaths	0,1	45	45	44	44	45
Height > 155cm	0,1	46	46	46	46	46
Caesarian sections	0,1	47	47	47	47	47
Maternal age	0-5	48	48	48	48	48
Parity	0-5	49	49	49	49	49

Table I. An abridged list of the 49 discrete variables used to predict failure to progress, and the order in which they were removed from the saturated model across 5 separate runs of the backwards elimination algorithm. Capitalized entries indicate variables abstracted from ICD-9 codes describing maternal conditions. These results are based on the data set containing all mothers.

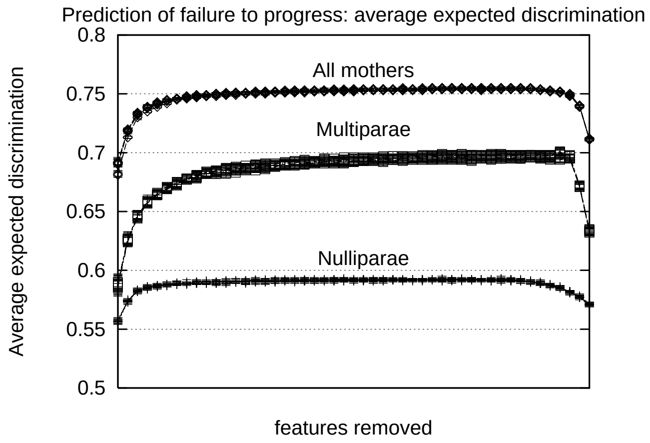

Fig. 1. This graph shows the average expected discrimination for prediction of failure to progress, obtained over 10 runs of the backwards elimination algorithm. Three different data sets are considered: all mothers, multiparae and nulliparae. Features are removed, one at a time, from left to right, starting with the saturated model.