Information Geometry for Survival Analysis and Feature Selection

A. Eleuteri (1,2), R. Tagliaferri (3,4), F. Acernese (1,2), L. Milano (1,2) and M. De Laurentiis (5)

(1) DSF Università di Napoli "Federico II", Complesso Univ. di M.S. Angelo, via Cintia, 80126, Napoli, Italia
eleuteri@na.infn.it
(2) INFN Sez. Napoli, Complesso Univ. di M.S. Angelo, via Cintia, 80126, Napoli, Italia
(3) DMI Università di Salerno, Salerno, Italia
robtag@unisa.it
(4) INFM Unità di Salerno, Salerno, Italia
(5) Dipartimento di Oncologia Molecolare e Clinica, Università di Napoli "Federico II", Napoli, Italia

Keywords. Information Geometry, Survival analysis, Feature selection, Neural networks

Introduction

In this paper an information geometric approach to survival analysis is described. A neural network is designed to model the probability of failure of a system, and it is trained by minimising a suitable divergence functional in a Bayesian framework. By using the trained network, minimisation of the same divergence functional allows for *fast, efficient* and *exact* feature selection.

Survival Analysis and Feature Selection

In literature we find many different modeling approaches to survival analysis. Conventional parametric models may involve too strict assumptions on the distributions of failure times and on the form of the influence of the system features on the survival time, assumptions which usually extremely simplify the experimental evidence, particularly in the case of medical data. In contrast, semi-parametric models do not make assumptions on the distributions of failures, but make instead assumptions on how the system features influence the survival time; furthermore, usually these models do not allow for direct estimation of survival times. Finally, non-parametric models only allow for a qualitative description of the data.

Neural networks have been recently used for survival analysis; for a survey on the current use of neural networks we refer to [8] and [10]. The only neural network architectures aimed at survival analysis and trained in a Bayesian framework are described by [2], [3], [4] and [7].

In many survival analysis applications, the issue of feature selection has a central role. Often the phenomenon we are trying to model is very complex, and there is not a priori knowledge which can be used to select the input variables which are relevant to modeling the survival function. So, the usual approach is to use all the available inputs. This is true, for example, in the biostatistics and bioinformatics fields, in which researchers are not willing to exclude possibly relevant variables. The inclusion of many inputs however has many drawbacks: interpreting the model is difficult, irrelevant variables act as noise worsening the generalization capability of the model, data gathering can be much more costly and small data sets become less appealing because of the curse of dimensionality.

Variable selection methods are a main subject of research in statistics [5], with application to commonly used modeling techniques. However, it turns out that much of the developed theories and algorithms are not usable in the case of complex layered models (like neural networks) [9].

Many of the existing methodologies tackle the more general problem of architecture selection, that is the elimination of parameters to get simpler models; variable selection criteria are then based on these pruning strategies. Other selection criteria define some direct "measures" of relevance of parameters, or evaluate their relevance by some functionals of the model outputs (e.g. functions of first or second derivatives w.r.t. the inputs). Almost all methods require retraining of the model, and have therefore a very high computational cost. Furthermore, these methods are typically coupled with suboptimal search criteria, like forward and backward substitution [5], which cannot take into account relations between input variables. For a review of some variable selection algorithms applied to layered models the reader can refer to [9]. The fundamental relevance of survival analysis and feature selection for bioinformatics research is reported in the best paper of the CAMDA 03 conference [6], where the authors however use common modeling approaches. Therefore, we see the need for improvements in modeling and feature selection techniques.

A Novel Information Geometric Approach

In this paper we describe a neural network architecture [3], a generalisation of the model proposed in [4], which overcomes all the limitations of currently available neural network models without making assumptions on the underlying survival distribution and without using discretizations or piecewise approximations as other neural network approaches to survival analysis do.

The neural network architecture is defined and trained according to information geometric principles [1] by minimising a suitable divergence functional in a Bayesian framework. The same geometric concepts are applied to exploit the geometric structure of the network and an algorithm is formulated which efficiently solves the feature selection problem by evaluating "projections" between the manifolds of probability densities implied by models with different input dimensions [3]. The feature selection algorithm makes full use of the information obtained in the training phase, and *does not need retraining* of the network, *nor the full exploration of the feature space*. Instead, the monotonicity of the proposed selection criterion allows for a fast and "clever" exploration of the feature space by using efficient "branch and bound" [5] search algorithms. Our method takes into account possible dependencies between input variables

References

[1] S. Amari. Methods of Information Geometry. TMM 191, Oxford University Press, Oxford, 2000.

[2] B. Bakker, T. Heskes. A neural-Bayesian approach to survival analysis. Proceedings IEE Artificial Neural Networks, 832-837, 1999.

[3] A. Eleuteri. On a novel neural computation approach to conditional probability estimation and variable selection. Ph.D. thesis, DMA ``R. Caccioppoli", Università degli Studi di Napoli ``Federico II", 2003. Available at: http://people.na.infn.it/~eleuteri

[4] A. Eleuteri, R. Tagliaferri, L. Milano, S. De Placido, M. De Laurentiis. A novel neural network-based survival analysis model. Neural Networks 16, 855-864, 2003.

[5] K. Fukunaga. Introduction to Statistical Pattern Recognition, second edition. Academic Press, 1990.

[6] J. Morris, G. Yin, K. Baggerly, C. Wu, L. Zhang. Identification of Prognostic Genes, Combining Information Across Different Institutions and Oligonucleotide Arrays. CAMDA 2003, Duke University, 2003.

[7] R. M. Neal. Survival Analysis Using a Bayesian Neural Network. JSM report, Atlanta, 2001.

[8] B. D. Ripley, R. M. Ripley. Neural Networks as Statistical Methods in Survival Analysis. Artificial Neural Networks: Prospects for Medicine (R. Dybowsky and V. Gant eds.), Landes Biosciences Publishers, 1998.

[9] W. S. Sarle. How to measure the importance of inputs?. Technical Report, SAS Institute Inc., Cary, NC, USA, 2000.

[10] G. Schwarzer, W. Vach, M. Schumacher. On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology. Statistics in medicine 19, 541-561, 2000.