ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

Interval coded scoring systems for survival analysis

V. Van Belle[12*], S. Van Huffel[12], J.A.K. Suykens[12] and S. Boyd[3]

1- Katholieke Universiteit Leuven - Department of Electrical Engineering (ESAT-SCD)
Kasteelpark Arenberg 10/2446, 3001 Leuven - Belgium
2- IBBT-K.U.Leuven Future Health Department
Kasteelpark Arenberg 10/2446, 3001 Leuven - Belgium
3- Stanford University - Department of Electrical Engineering
Stanford, CA 94305-9510 USA

Abstract. Black-box mathematical models are powerful tools in classification and regression problems. Thanks to the use of (unknown) transformations of the inputs, the outcome can be estimated, improving performance in comparison to standard statistical models. A disadvantage of these complex models however, is their lack of interpretability. This work illustrates how advanced methods can be made interpretable. Using constant B-spline kernel functions and sparsity constraints, interval coded scoring models for survival analysis are presented.

1 Introduction

Clinical decision support systems are often based on standard statistical models with linear effects of the inputs. The machine learning techniques are ideal to model non-linearities present in clinical data and to incorporate interactions between inputs in an automatic way. However, these techniques are seldomly used in clinical practice due to the lack of interpretability of the resulting models.

Popular decision support systems are too often based on a rough approximation of (logistic) regression models. A study of the clinical literature on decision support [1, 2, 3] illustrates that clinicians are interested in decision support supplied without interfering with the clinical work flow, in an automatic way and providing recommendations. A commonly used decision support tool is a scoring chart. Such a chart consists of the effects of several inputs, which are represented by consecutive intervals, within which the effects are assumed to be constant. Although these tools have nice properties concerning interpretation and applicability in a clinical setting, they have major drawbacks: (i) they are a rough approximation of a previously built model [4] and thus (ii) do not include a control mechanism for the possible loss of information by creating input

*We would like to thank P. Neven and V. Harvey for the collection of the data. This research was supported by Research Council KUL: GOA Ambiorics, GOA MaNet, CoE EF/05/006 Optimization in Engineering (OPTEC), projects: FWO G.0302.07 (SVM/Kernel), G0226.06 (cooperative systems and optimization); IBBT, Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, 'Dynamical systems, control and optimization', 2007-2011). V. Van Belle is supported by a grant of the IWT Vlaanderen (Agentschap voor Innovatie door Wetenschap en Technologie) and a post-doctoral grant (BOF-PDM) from the Katholieke Universiteit Leuven.

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

intervals, (iii) the generated intervals are depending on the model builder, and (iv) the performance highly depends on the chosen number of intervals.

In order to accomplish the wishes of the end user while overcoming the drawbacks of the existing tools, a support vector machine for the analysis of survival data [5, 6] is adapted such that the obtained model automatically results in a score chart. The intervals of the inputs, as well as the number of intervals, are defined within the optimization problem. The resulting models are represented by means of color bars for improved visual interpretation.

2 Interval coded scoring systems

In order to obtain scoring systems, the survival problem is tackled by means of transformation models [5]. These models combine a ranking step, in which a score as concordant as possible to the failure time is searched, and a reconstruction step, linking the score from the previous step with a survival estimate. Interpretable survival systems can be obtained by adapting the first step of these models. A discussion on the use of interval coded scoring systems (ICS) for classification and medical applications is found in [7].

2.1 Step 1: Interval coded scoring systems for survival analysis

To develop an interval coded score system (ICS) for prognostic problems, we start from a support vector machine for the analysis of survival data, combining ranking and regression constraints in order to deal with the incomplete information of censored observations [6]. The SVM survival model is then adapted at three points: (i) the model is constrained to be a generalized additive model [8], (ii) with an explicit feature map with functional forms closely related to constant B-splines [9], and (iii) sparsity constraints (minimizing the total variation of the coefficients vector [10]) are added in order to reduce the number of intervals to a minimum and perform feature selection. Let $\mathcal{D} = \{(x_i, y_i, \delta_i)\}_{i=1}^n$ be a dataset with x_i, y_i and δ_i the inputs, survival time and censoring indicator for the i^{th} observation, respectively. Let x_i^p be the p^{th} input out of d and $w_{p,l}$ the weight corresponding to the l^{th} interval and $k_p + 1$ the number of intervals (thresholds $\tau_{p,l}$) of the p^{th} input. The model is then written as a convex optimization problem [11]:

$$
\begin{aligned}
&\min_{w,b,\epsilon,\xi,\xi^*} \boxed{\textstyle\sum_{p=1}^{d} \sum_{l=1}^{k_p+1} |w_{p,l} - w_{p,l-1}|}^{\text{iii}} + \gamma \sum_{i=1}^{n} \epsilon_i + \mu \sum_{i=1}^{n} (\xi + \xi^*) \\
&\text{s.t.} \begin{cases}
\hat{y}_i = \boxed{\textstyle\sum_{p=1}^{d}}^{\text{i}} \left(\boxed{\textstyle\sum_{l=1}^{k_p+1} w_{p,l} \mathcal{I}(\tau_{p,l-1} \leq x_i^p < \tau_{p,l})}^{\text{ii}} \right) + b, & \forall\, i = 1, \ldots, n\,, \\
\hat{y}_i - \hat{y}_{i-1} + \epsilon_i \geq y_i - y_{i-1}, & \forall\, i = 2, \ldots, n\,, \\
\hat{y}_i \geq y_i - \xi_i, & \forall\, i = 1, \ldots, n\,, \\
-\delta_i \hat{y}_i \geq -\delta_i y_i - \xi_i^*, & \forall\, i = 1, \ldots, n\,, \\
\xi_i, \xi_i^*, \epsilon_i \geq 0, & \forall\, i = 1, \ldots, n\,.
\end{cases}
\end{aligned}
$$

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

To reduce the number of steps, an iteratively reweighted L_1 minimization is performed. The difference between the weights of two consecutive intervals is weighted with $\chi_{p,l} = \frac{1}{\varepsilon + a|w_{pl} - w_{pl-1}|}$, $\forall\ p = 1, \ldots, d,\ \forall\ l = 1, \ldots, k_p + 1$, where ε is a small positive value (e.g. 0.0005) and the value of a is optimized for the problem at hand.

Although the result is easy to interpret, it is not yet easy to use. We therefore propose to normalize the coefficients $w_{p,l}$ such that the smallest non-zero absolute value of the coefficients becomes 1. All other normalized coefficients are rounded to the nearest integer ($\tilde{w}_{p,l}$). The final score for a new observation $x_\star$ is then found as $\sum_{p=1}^{d} \left(\sum_{l=1}^{k_p+1} \tilde{w}_{p,l} \mathcal{I}(\tau_{p,l-1} \leq x_\star^p < \tau_{p,l}) \right)$.

2.2 Step 2: Estimation of the survival function

Once the scores are calculated, a survival function needs to be estimated. Preferably, one survival curve is estimated for each possible score. However, estimation of this function will only be reliable when enough observations (with events) have the same score. The ICS survival model is therefore used a second time, using the scores as input and the failure times as output. The obtained step function will now denote which scores correspond to the same survival and can therefore be taken together when estimating the survival curves $\hat{S}$.

The cumulative distribution function (CDF), equal to $1 - \hat{S}$, is estimated by means of monotone least-squares support vector regression [12]. To include censored observations, the data are preprocessed. An augmented data set $\mathcal{D}_{\text{aug}} = \{\mathcal{D}_i\}_{i=1}^{n}$ is created. Each data set $\mathcal{D}_i = \{(x_i, y_{i,k})\}_{k=1}^{n_t}$ represents a replication of observation i within n_t consecutive time intervals $k = 1, \ldots, n_t$. The outcome $y_{i,k}$ is zero when the event did not occur before the end of the k^{th} time interval. For events, $y_{i,k} = 1$ for all intervals ending after the observed failure time. For censored data, the observations are only replicated within the intervals in which they are known to be at risk. The model then becomes

$$\min_{w,b,\epsilon} \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{i=1}^{n} \sum_{k=1}^{n_t} \epsilon_{i,k}^2$$

$$\text{s.t.} \begin{cases} w^T \varphi(x_{i,k}) + b = y_{i,k} + \epsilon_{i,k}, & \forall\ i = 1, \ldots, n; \forall\ k = 1, \ldots, n_t\,, \\ w^T (\varphi(x_{i,k}) - \varphi(x_{i,k-1})) \geq 0, & \forall\ i = 1, \ldots, n;\ \forall\ k = 2, \ldots, n_t\,, \\ w^T \varphi(x_{i,k}) + b \geq 0, & \forall\ i = 1, \ldots, n;\ \forall\ k = 1, \ldots, n_t\,, \\ -w^T \varphi(x_{i,k}) - b \geq -1, & \forall\ i = 1, \ldots, n;\ \forall\ k = 1, \ldots, n_t\,, \end{cases}$$

with $x_{i,k}$ the augmented input $x_{i,k} = [x_i, k]^T$. In the above formulation x_i is the score obtained from the first step of the transformation model and k the k^{th} time interval.

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

3 Illustrative example

The ICS for survival analysis is illustrated on the prognosis of primary operable breast cancer patients. The model is trained on a set of 1923 patients with complete information, treated at the University Hospitals Leuven between January 2000 and June 2005. The scoring system is then validated on an external set containing complete information on 1192 patients treated in New Zealand (Auckland Breast Cancer Registry) between January 2000 and December 2005. Only patients with complete information for age, tumor size, number of positive lymph nodes, expression of the progesterone (PR) and human epidermal growth factor receptor 2 (HER2) and tumor grade were considered in the analysis. The model was trained using 10-fold cross-validation to tune the hyperparameter. In order to find the optimal weight parameter a, 5-fold cross-validation was used. The obtained ICS model is illustrated in Figure 1. The ICS model is used a second time with the ICS scores as a single input in order to obtain the best cut-off values to define risk groups. Five different risk groups are recognized, with predicted survival curves closely aligning with the observed curves, in training as well as in test set (results not shown).

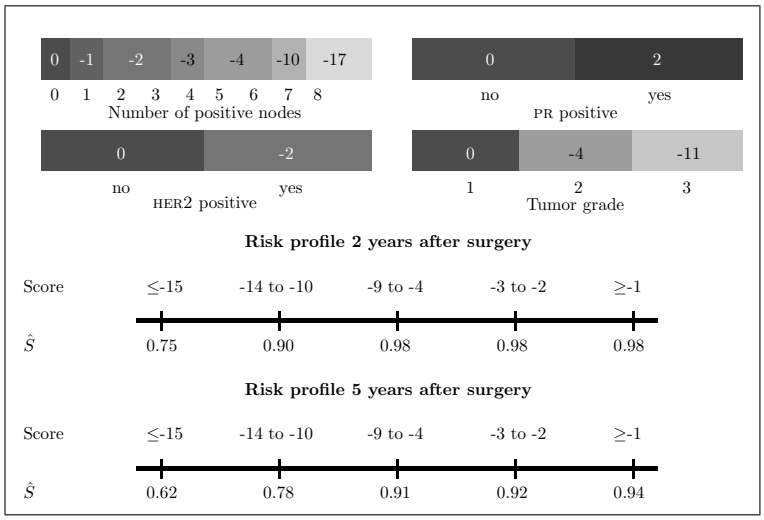

Fig. 1: ICS model to predict the prognosis of primary operable breast cancer patients. Given this chart, the clinician knows which variables need to be collected in order to obtain an estimate of the patient's prognosis. For each of the represented bars, the corresponding points need to be calculated. The total score is obtained by addition of all these points.

The ICS model is compared with the standard model for defining breast cancer risk groups, the Notthingham prognostic index (NPI) [13] and an improved version (iNPI) [14]. Both models use three risk groups. However, this number is chosen arbitrarily. The ICS method is used a second time with the continuous

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

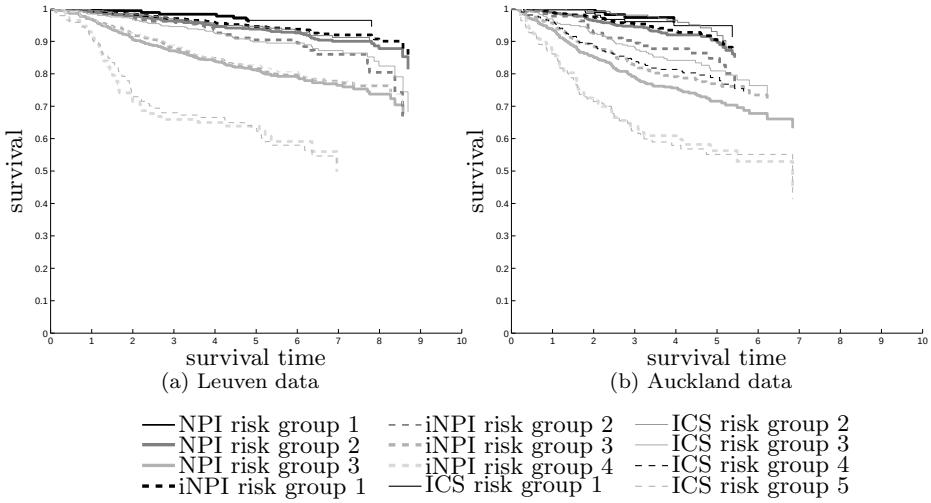

Fig. 2: Kaplan-Meier survival curves according to the risk groups defined by NPI, iNPI and ICS. The best separation is found using ICS.

outcome of the NPI or iNPI as input to obtain the number of risk groups with different survival curves. Application of this approach on the NPI yields 3 risk groups with cut-offs at 2.6 and 4.4. The resulting groups on the iNPI are defined by the cut-offs 3.5, 4.3 and 6.3. The Kaplan-Meier curves of the different risk groups are represented in Figure 2. The NPI can only divide patients into three risk groups, for which the predicted survival at two and five years differ less than for both other models. The iNPI obtains the largest estimated survival difference between the most extreme risk groups. The NPI is able to define a very good prognostic group, but does not find a risk group with a very poor prognosis. The iNPI on the contrary does find a very poor prognostic group, but fails to find a very good prognostic group. The ICS model finds both a very good and a very bad prognostic group.

4 Conclusions

This paper presents an attractive way to visualize a survival model. A study of the properties needed to lower the threshold to use clinical decision supports systems in clinical practice, learned that clinicians appreciate the representation of a model by means of intervals. A SVM survival model was therefore adapted such that the resulting models automatically lead to clinical yes/no questions. Depending on the answers, a point is added to the score. The final score is then used to attach a patient-specific estimate of the risk on the event over time.

The model was illustrated on the prognosis of primary operable breast cancers and validated on an independent test set. The results are promising: the model

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

is able to identify which variables are important to predict relapse, but it is also able to identify how many different survival groups can be noted. A comparison with currently used methods for the classification of patients within risk groups indicates that the ICS method is able to define more risk groups than both reference models and the survival curves have a wider spread.

In the future, it will be necessary to adapt the model structure in order to allow for interactions between input variables. Additionally, further research is necessary to estimate survival curves for large datasets and more time points.

References

[1] G P Percell. What makes a good clinical decision support system. *British Medical Journal*, 330:740–741, 2005.

[2] K Kawamoto, C A Houlihan, E A Balas, and D F Lobach. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *British Medical Journal*, 330:765–773, 2005.

[3] J A Osheroff. *Improving medication use and outcomes with clinical decision support: a step-by-step guide*. Healthcare Information and Management Systems Society, Chicago, IL, 2009.

[4] L M Sullivan, J M Massaro, and R B D'Agostino. Presentation of multivariate data for clinical use: The framingham study risk score functions. *Statistics in Medicine*, 23(10):1631–1660, 2004.

[5] V Van Belle, K Pelckmans, J A K Suykens, and S Van Huffel. Learning Transformation Models for Ranking and Survival Analysis. *Journal of Machine Learning Research*, 12:819–862, 2011.

[6] V Van Belle, K Pelckmans, S Van Huffel, and J A K Suykens. Support vector methods for survival analysis: a comparison between ranking and regression approaches. *Artificial Intelligence in Medicine*, 53(2):107–118, 2011.

[7] V Van Belle, B Van Calster, D Timmerman, T Bourne, C Bottomley, L Valentin, P Neven, S Van Huffel, J A K Suykens, and S Boyd. A Mathematical Model for Interpretable Clinical Decision Support with Applications in Gynecology. Technical report, 10-170, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), 2010. Submitted for publication.

[8] T Hastie and R Tibshirani. *Generalized additive models*. Chapman and Hall, 1990.

[9] C de Boor. *A Practical Guide to Splines*. Springer, Berlin, 1978.

[10] L I Rudin, S Osher, and E Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 1992.

[11] S Boyd and L Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004.

[12] K Pelckmans, M Espinoza, J De Brabanter, J A K Suykens, and B De Moor. Primal-dual monotone kernel regression. *Neural Processing Letters*, 22(2):171–182, 2005.

[13] M Galea, R Blamey, C Elston, and I Ellis. The Nottingham Prognostic Index in primary breast cancer. *Breast Cancer Research and Treatment*, 22(3):207–219, 1992.

[14] V Van Belle, B Van Calster, O Brouckaert, I Vanden Bempt, S Pintens, R Paridaens, F Amant, K Leunen, A Smeets, R Drijkoningen, H Wildiers, M R Christiaens, I Vergote, S Van Huffel, and P Neven. Qualitative assessment of the progesterone receptor and HER-2 improve the Nottingham Prognostic Index for short term breast cancer prognosis. *Journal of Clinical Oncology*, 28(27):4129–4134, 2010.