ICR The Institute of Cancer Research

An Efficient Training Algorithm for Kernel Survival Support Vector Machines

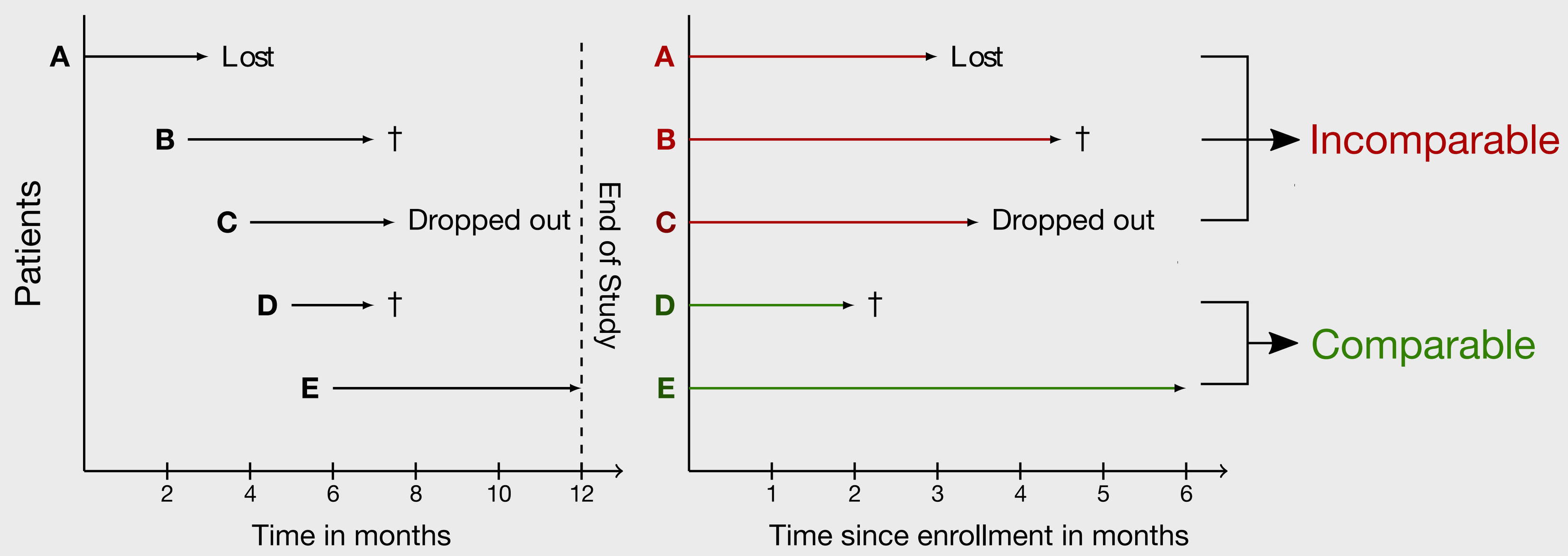
Sebastian Pölsterl[1] / Nassir Navab[2,3] / Amin Katouzian[4]

Overview

- **Survival analysis** tries to establish a connection between a set of features and the time until an event of interest.
- **Problem**: For a dataset with *n* samples and *p* features, previous training algorithms for the kernel survival SVM (SSVM) require $O(n^4)$ space and $O(pn^6)$ time.[1,2]
- Recently, an **efficient training algorithm for linear SSVM** with much lower time complexity and linear space complexity has been proposed.[3]
- We **extend this optimisation scheme to the non-linear case** and show that it allows analysing large-scale data with no loss in prediction performance.

Survival Analysis

- In survival analysis, parts of the training and test data can only be partially observed.
- Patients that remain event-free during the study-period are **right censored**, because it is unknown whether an event has or has not occurred after the study ended. Only **partial information** about their survival is available.



Kernel Survival Support Vector Machine

- The SSVM[1,2] is an extension of the Rank SVM[4] to right censored survival data: *patients with a lower survival time should be ranked before patients with longer survival time*.
- Survival data consists of feature vectors $x_i \in \mathbb{R}^p$, the time of an event/censoring $y_i > 0$, and an event indicator $\delta_i > 0$.

Objective function: $\mathcal{P} = \{(i,j) \mid y_i > y_j \land \delta_j = 1\}_{i,j=1}^n$

$$\min_{\boldsymbol{w}} \quad \frac{1}{2}\|\boldsymbol{w}\|_2^2 + \gamma \sum_{(i,j)\in P} \max(0, 1 - \boldsymbol{w}^\top(\phi(\boldsymbol{x}_i) - \phi(\boldsymbol{x}_j)))$$

Lagrange dual problem with $\boldsymbol{K}_{i,j} = \phi(\boldsymbol{x}_i)^\top \phi(\boldsymbol{x}_j)$:

$$\max_{\boldsymbol{\alpha}} \quad \boldsymbol{\alpha}^\top \boldsymbol{1}_m - \frac{1}{2}\boldsymbol{\alpha}^\top \boldsymbol{A}\boldsymbol{K}\boldsymbol{A}^\top \boldsymbol{\alpha}$$

$$\text{subject to} \quad 0 \leq \alpha_{ij} \leq \gamma, \quad \forall (i,j) \in P,$$

where $\boldsymbol{A}_{k,i} = 1$ and $\boldsymbol{A}_{k,j} = -1$ if $(i,j) \in \mathcal{P}$ and 0 otherwise.

Proposed Optimisation Scheme

- Find a function $f \colon \mathcal{X} \to \mathbb{R}$ from a reproducing Kernel Hilbert space $\mathcal{H}_k$ with $k \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ (usually $\mathcal{X} \subset \mathbb{R}^p$):

$$\min_{f \in H_k} \quad \frac{1}{2}\|f\|_{H_k}^2 + \frac{\gamma}{2} \sum_{(i,j)\in P} \max(0, 1 - (f(\boldsymbol{x}_i) - f(\boldsymbol{x}_j)))^2$$

- Directly apply representer theorem[5] $(f(\boldsymbol{z}) = \sum_{i=1}^n \beta_i k(\boldsymbol{x}_i, \boldsymbol{z}))$ and minimise *squared hinge loss* with respect to $\boldsymbol{\beta} \in \mathbb{R}^n$:

$$\begin{aligned} R(\boldsymbol{\beta}) =& \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n \beta_i \beta_j k(\boldsymbol{x}_i, \boldsymbol{x}_j) \\ &+ \frac{\gamma}{2} \sum_{(i,j)\in P} \max\left(0, 1 - \sum_{l=1}^n \beta_l (k(\boldsymbol{x}_l, \boldsymbol{x}_i) - k(\boldsymbol{x}_l, \boldsymbol{x}_j))\right)^2 \\ =& \frac{1}{2}\boldsymbol{\beta}^\top \boldsymbol{K}\boldsymbol{\beta} + \frac{\gamma}{2}(\boldsymbol{1}_m - \boldsymbol{A}\boldsymbol{K}\boldsymbol{\beta})^\top \boldsymbol{D}_{\boldsymbol{\beta}} (\boldsymbol{1}_m - \boldsymbol{A}\boldsymbol{K}\boldsymbol{\beta}) \end{aligned}$$

$$(\boldsymbol{D}_{\boldsymbol{\beta}})_{k,k} = \begin{cases} 1 & \text{if } f(\boldsymbol{x}_j) > f(\boldsymbol{x}_i) - 1 \Leftrightarrow \boldsymbol{K}_j \boldsymbol{\beta} > \boldsymbol{K}_i \boldsymbol{\beta} - 1, \\ 0 & \text{else.} \end{cases}$$

- The form of the optimisation problem is very similar to the one of linear SSVM, which allows using truncated Newton optimisation and order statistic trees to avoid storing $\boldsymbol{A}$.[3,5]

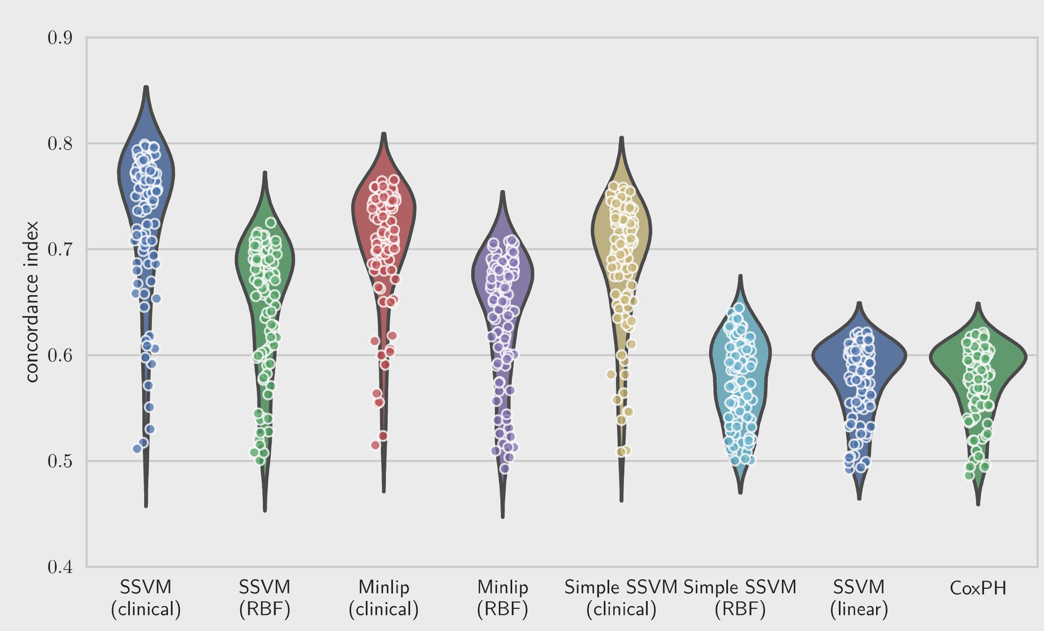
Complexity

- **Space**: $O(n)$ or $O(n^2)$ if $\boldsymbol{K}$ is to be stored in memory.
- **Time** (assuming evaluating kernel function costs $O(p)$):
  - $O(n^3 p)$ to compute $\boldsymbol{K}_i \boldsymbol{v}$ for all $i = 1, \ldots, n$
  - $O(n \log n)$ to sort samples according to $\boldsymbol{K}_i \boldsymbol{v}$
  - $O(n^2 + n + n \log n)$ to compute Hessian-vector product
- **Overall** (if kernel matrix is stored in memory):

$$O(n^2 p) + \left[O(n \log n) + O(n^2 + n + n \log n)\right] \cdot \bar{N}_{\text{CG}} \cdot N_{\text{Newton}}$$

Experiments

Synthetic:



Real world:

		SSVM (ours)	SSVM (simple)[6]	Minlip[7]	SSVM (linear)[3]	Cox
AIDS study	Uno's c	0.711	0.621	0.560	0.659	0.663
	iAUC	0.759	0.685	0.724	0.766	0.771
Coronary artery disease	Uno's c	0.780	0.751	0.745	0.730	0.732
	iAUC	0.753	0.641	0.703	0.716	0.777
Framingham offspring	Uno's c	0.732	0.674	0.724	0.699	0.742
	iAUC	0.827	0.742	0.837	0.829	0.832
Lung cancer	Uno's c	0.664	0.605	0.716	0.709	0.712
	iAUC	0.740	0.630	0.790	0.783	0.780

References

1. Evers et al.: Sparse kernel methods for high-dimensional survival data. Bioinformatics 24(14). pp. 1632–38. 2008
2. Van Belle et al.: Support vector machines for survival Analysis. 3rd Int. Conf. Comput. Intell. Med. Healthc. 2007
3. Pölsterl et al.: Fast training of support vector machines for survival analysis. ECML PKDD 2015
4. Herbrich et al.: Large margin rank boundaries for ordinal regression. Advances in Large Margin Classifiers. 2000
5. Kuo et al.: Large-scale kernel RankSVM. SIAM International Conference on Data Mining. 2014
6. Van Belle et al.: Survival SVM: a practical scalable algorithm. 16th Euro. Symb. Artif. Neural Netw. 2008
7. Van Belle et al.: Learning transformation models for ranking and survival analysis. JMLR. 12, pp. 819–62. 2011

Contact

1 Institute of Cancer Research, London UK; 2 Technische Universität München, Munich, Germany; 3 Johns Hopkins University, Baltimore, MD, USA; 4 IBM Almaden Research Center, San Jose, CA, USA

Acknowledgments

This work has been supported by the CRUK Centre at the Institute of Cancer Research and Royal Marsden (Grant No. C309/A18077), The Heather Beckwith Charitable Settlement, The John L Beckwith Charitable Trust, and the Leibniz Supercomputing Centre. We would like to thank Bissan Al-Lazikani and Carmen Rodriguez-Gonzalvez.

CANCER RESEARCH UK   The ROYAL MARSDEN NHS Foundation Trust   NHS National Institute for Health Research