Estimation of Conditional Mixture Weibull Distribution with Right-Censored Data using Neural Network for Time-to-Event Analysis

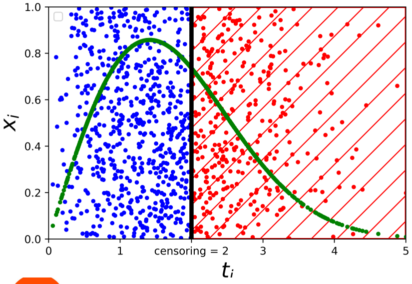
Achraf Bennis   Sandrine Mouysset   Mathieu Serrurier

IRIT - Paul Sabatier University

Video Presentation for PAKDD 2020

Context

Right-Censored Data



- x_i: observation,
- t_i: time recorded,
- censoring: censoring time threshold,
- blue points: $\delta = 1$ and red points: $\delta = 0$ (δ is the event indicator).

Context

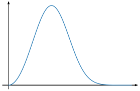
Issue

Classical approach: estimate the risk $P(t > t_h)$, t is time to the event and t_h is the time horizon.

estimate the risk at a given point t_h

Opted approach: estimate the time-to-event distribution.

$$X = \{x_i, t_i, \delta_i\}_{i=1,..,N}$$

$x_i \rightarrow$ 

estimate the risk at any point.

Contribution

Distribution Assumption

- Time-to-event distribution can be modeled by a mixture of Weibull distributions:

$$\sum_{k=1}^{p} \alpha_k W(\beta_k, \eta_k), \quad where : \sum_{k=1}^{p} \alpha_k = 1, \quad \alpha_k \quad 0 \;\; k = 1, .., p$$

=) Analytical expression of the likelihood is known!

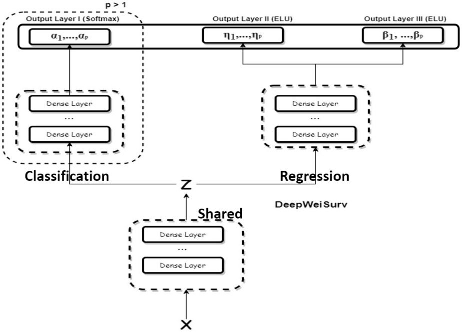
- Goal: Estimate the mixture parameters $\alpha_k, \beta_k, \eta_k$ for each individual x_i.

Contribution

Proposed Model

$$g_p : \mathbb{R}^d \to \mathbb{R}^{p \times 3}$$
$$x_i \mapsto (\alpha, \beta, \eta)$$

g_p models the relationship between the mixture parameters and the features x_i



DeepWeiSurv finds the parameters estimates by resolving the following optimization problem:

$$\min_{\substack{(\alpha,\beta,\eta)\\ \eta>0,\beta\geq 1}} loss = \overbrace{\sum_{i=1}^{n} \delta_i log\left[\sum_{k=1}^{p} \alpha_k f_{W(\beta_k,\eta_k)}(t_i)\right]}^{LL_{\delta=1}}$$

$$\underbrace{+\sum_{i=1}^{n}(1 \quad \delta_i) log\left[\sum_{k=1}^{p} \alpha_k (1 \quad F_{W(\beta_k,\eta_k)}(t_i))\right]}_{LL_{\delta=0}}$$

ELU co-domain =]-1, *1*] =) $\beta \quad \beta + 2 \quad \eta \quad \eta + 1 + \epsilon, \epsilon > 0$

Contribution

Network Configuration

- Network configuration:
  - Shared sub-network (3 dense layers):
    - 128-node dense layer
    - Batch-normalization
    - 64-node dense layer
    - 32-node dense layer
  - Regression sub-network (*reg*, 2 dense layers):
    - 16-node dense layer
    - batch-normalization
    - 8-node dense layer
    - 2 ELU layers (output)
  - Classification sub-network (same as *reg*, except for the output layer: 1 softmax layer instead).

- ReLU activation between the hidden layers. DeepWeiSurv trained using Adam optimizer and learning rate of $10^{\ 4}$

Experiments

Prediction and Evaluation Criteria

The mean lifetime predicted of the i^{th} individual can be calculated as follows:

$$\mu_i = \sum_{k=1}^{p} \alpha_k \eta_{ik} \Gamma(1 + \frac{1}{\beta_{ik}}), \quad \Gamma : \text{Gamma Function}$$

We measure the performance of the models by calculating the *concordance index*:

$$\text{C-index} = \frac{\sum_{i,j} 1_{t_i > t_j} . 1_{\mu_i > \mu_j} . \delta_j}{\sum_{i,j} 1_{t_i > t_j} . \delta_j}$$

Experiments

Real-World Datasets

Some descriptive statistics on three two real-world datasets:

Datasets	No. non-censored	No. censored	No. Features	
			Qualitative	Quantitative
METABRIC	888 (44.8%)	1093 (55.2%)	15	6
SEER BC	9152(42.8%)	12221 (57.2%)	23	11
SEER HD	12014 (49.6%)	12221 (50.4%)		

METABRIC: Molecular Taxonomy of Breast Cancer International Consortium.

SEER : Surveillance, Epidemiology, and End Results.

Experiments

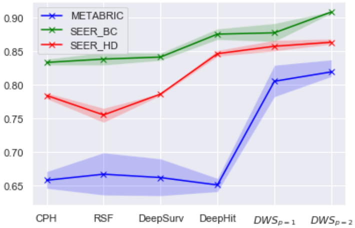
Experiment One: DeepWeiSurv vs. Competing Methods

- Competing methods:
  - CPH: Cox's Proportional Hazard
  - RSF: Random Survival Forest
  - DeepHit (*Lee, C. et al.*)
  - DeepSurv (*Katzman, J.L. et al.*)
- Proposed methods:
  - $\text{DeepWeiSurv}_{p=1}$
  - $\text{DeepWeiSurv}_{p=2}$
- Experimental protocol: 5-fold cross validation. For each fold: calculate μ_is and then C-index.

Experiments

Experiment One: DeepWeiSurv vs. Competing Methods

The results of Experiment One is represented by the figure below:



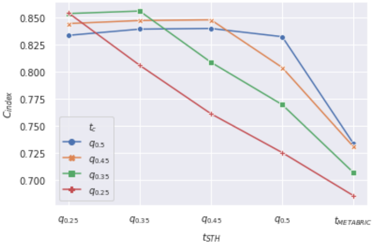
Experiments

Experiment Two: Censoring Threshold Sensitivity Experiment on METABRIC

Distribution of training set's observations (censored/non-censored) for each selected censoring time threshold t_c:

t_c	No. censored	No. non-censored	Added portion
$q_{0.5}$	1285	696	17.6%
$q_{0.45}$	1127	570	29%
$q_{0.35}$	1248	422	42.6%
$q_{0.25}$	1338	311	52.8%

The average of C-index w.r.t survival time horizon t_{STH} for every selected threshold t_c:



Conclusion

DeepWeiSurv:

- Estimates the event-time distribution by learning the mixture parameters.
- Handles highly censored settings.
- Considers any given survival time horizon.
- On real-world dataset experiments, it outperforms the competing methods.