

# Conditional Hazard Estimating Neural Networks

**Antonio Eleuteri**

*Royal Liverpool University Hospital, UK*

**Azzam Taktak**

*Royal Liverpool University Hospital, UK*

**Bertil Damato**

*Royal Liverpool University Hospital, UK*

**Angela Douglas**

*Liverpool Women's Hospital, UK*

**Sarah Coupland**

*Royal Liverpool University Hospital, UK*

## INTRODUCTION

Survival analysis is used when we wish to study the occurrence of some event in a population of subjects and the time until the event of interest. This time is called *survival time* or *failure time*. Survival analysis is often used in industrial life-testing experiments and in clinical follow-up studies. Examples of application include: time until failure of a light bulb, time until occurrence of an anomaly in an electronic circuit, time until relapse of cancer, time until pregnancy.

In the 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 (Cox & Oakes, 1984). In contrast, semi-parametric models do not make assumptions on the distributions of failures, but instead make assumptions on how the system features influence the survival time (the usual assumption is the proportionality of hazards); furthermore, these models do not usually allow for direct estimation of survival times. Finally, non-parametric models usually only allow for a qualitative description of the data on the population level.

Neural networks have recently been used for survival analysis; for a survey on the current use of neural networks, and some previous attempts at neural network

survival modeling we refer to (Bakker & Heskes, 1999), (Biganzoli et al., 1998), (Eleuteri et al., 2003), (Lisboa et al., 2003), (Neal, 2001), (Ripley & Ripley, 1998), (Schwarzer et al. 2000).

Neural networks provide efficient parametric estimates of survival functions, and, in principle, the capability to give personalised survival predictions. In a medical context, such information is valuable both to clinicians and patients. It helps clinicians to choose appropriate treatment and plan follow-up efficiently. Patients at high risk could be followed up more frequently than those at lower risk in order to channel valuable resources to those who need them most. For patients, obtaining information about their prognosis is also extremely valuable in terms of planning their lives and providing care for their dependents.

In this article we describe a novel neural network model aimed at solving the survival analysis problem in a continuous time setting; we provide details about the Bayesian approach to modeling, and a sample application on real data is shown.

## BACKGROUND

Let  $T$  denote an absolutely continuous positive random variable, with distribution function  $P$ , representing the time of occurrence of an event. The survival function,  $S(t)$ , is defined as:

$$S(t) = \Pr(T > t),$$

that is, the probability of surviving beyond time  $t$ . We shall generally assume that the survival function also depends on a set of covariates, represented by the vector  $x$  (which can itself be assumed to be a random variable). An important function related to the survival function is the *hazard rate* (Cox & Oakes, 1984), defined as:

$$h_r(t) = P'(t)/S(t)$$

where  $P'$  is the density associated to  $P$ . The hazard rate can be interpreted as the instantaneous force of mortality.

In many survival analysis applications we do not directly observe realisations of the random variable  $T$ ; therefore we must deal with a missing data problem. The most common form of missingness is *right censoring*, i.e., we observe realisations of the random variable:

$$Z = \min(T, C),$$

where  $C$  is a random variable whose distribution is usually unknown. We shall use a censoring indicator  $d$  to denote whether we have observed an event ( $d=1$ ) or not ( $d=0$ ). It can be shown that inference does not depend on the distribution of  $C$  (Cox & Oakes, 1984).

With the above definitions in mind we can now formulate the log-likelihood function necessary for statistical inference. We shall omit the details, and only report the analytical form:

$$L = \sum_i d_i \log h_r(t_i, x_i) - \int_0^{t_i} h_r(u, x_i) du.$$

For further details, we refer the reader to (Cox & Oakes, 1984).

## CONDITIONAL HAZARD ESTIMATING NEURAL NETWORKS

### Neural Network Model

The neural network model we used is the Multi-Layer Perceptron (MLP) (Bishop, 1995):

$$a(t, x; w) = b_0 + \sum_k v_k g(u_k^T x + u_0 t + b_k)$$

where  $g()$  is a sigmoid function, and  $w = \{b_0, v, u, u_0, b\}$  is the set of network parameters. The MLP output defines an analytical model for the logarithm of the hazard rate function:

$$a(t, x; w) \triangleq \log h_r(t, x)$$

We refer to this continuous time model as Conditional Hazard Estimating Neural Network (CHENN).

### Bayesian Learning of the Network Parameters

The Bayesian learning framework offers several advantages over maximum likelihood methods commonly used in neural network learning (Bishop, 1995), (MacKay, 1992), among which the most important are automatic regularization and estimation of error bars on predictions.

In the conventional maximum likelihood approach to training, a single weight vector is found, which minimizes the error function; in contrast, the Bayesian scheme considers a probability distribution over weights  $w$ . This is described by a prior distribution  $p(w)$  which is modified when we observe a dataset  $D$ . This process can be expressed by Bayes' theorem:

$$p(w | D) = \frac{p(D | w) p(w)}{p(D)}.$$

To evaluate the posterior distribution, we need expressions for the likelihood  $p(D | w)$  (which we have already shown) and for the prior  $p(w)$ .

The prior over weights should reflect the knowledge, if any, we have about the mapping we want to build. In our case, we expect the function to be very smooth, so an appropriate prior might be:

$$p(w) \propto \exp\left(-\frac{1}{2} \sum_k \alpha_k w_k^T w_k\right)$$

which is a multivariate normal density with zero mean and diagonal covariance matrix with elements  $1/\alpha_k$ . In this way, weights centered on zero have higher probability, a fact which encourages very smooth functions.

4 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/chapter/conditional-hazard-estimating-neural-networks/10277](http://www.igi-global.com/chapter/conditional-hazard-estimating-neural-networks/10277)

## Related Content

---

### Identifying Influencers in Online Social Networks: The Role of Tie Strength

Yifeng Zhang, Xiaoqing Li and Te-Wei Wang (2013). *International Journal of Intelligent Information Technologies* (pp. 1-20).

[www.irma-international.org/article/identifying-influencers-online-social-networks/75543](http://www.irma-international.org/article/identifying-influencers-online-social-networks/75543)

### Memes and Mutation: Societal Implications of Evolutionary Agents in Push Technologies

Kenneth E. Kendall and Julie E. Kendall (2005). *International Journal of Intelligent Information Technologies* (pp. 17-29).

[www.irma-international.org/article/memes-mutation-societal-implications-evolutionary/2377](http://www.irma-international.org/article/memes-mutation-societal-implications-evolutionary/2377)

### Fusion of Improved Retinex Enhancement and Graph Neural Networks for Retinal Image Quality Classification

Wei Lin (2025). *International Journal of Intelligent Information Technologies* (pp. 1-20).

[www.irma-international.org/article/fusion-of-improved-retinex-enhancement-and-graph-neural-networks-for-retinal-image-quality-classification/383644](http://www.irma-international.org/article/fusion-of-improved-retinex-enhancement-and-graph-neural-networks-for-retinal-image-quality-classification/383644)

### An Intelligent and Dynamic Decision Support System for Nonlinear Environments

S. Uma and J. Suganthi (2012). *International Journal of Intelligent Information Technologies* (pp. 43-61).

[www.irma-international.org/article/intelligent-dynamic-decision-support-system/74829](http://www.irma-international.org/article/intelligent-dynamic-decision-support-system/74829)

### A New Two-Level Clustering Approach for Situations Management in Distributed Smart Environments

Achouri Mounir, Alti Adel, Dourdour Makhoul, Laborie Sébastien and Roose Philippe (2019). *International Journal of Ambient Computing and Intelligence* (pp. 91-111).

[www.irma-international.org/article/a-new-two-level-clustering-approach-for-situations-management-in-distributed-smart-environments/225773](http://www.irma-international.org/article/a-new-two-level-clustering-approach-for-situations-management-in-distributed-smart-environments/225773)