

# Personalized treatment recommender system

## Problem Statement

Medical practitioners use survival models to explore and understand the relationships between patient's covariates (e.g. clinical and genetic features) and the effectiveness of various treatment options. Standard survival models like the linear Cox proportional hazards model require extensive feature engineering or prior medical knowledge to model treatment interaction at an individual level. While nonlinear survival methods, such as neural networks and survival forests, can inherently model these high-level interaction terms, they have yet to be shown as effective treatment recommender systems.

Personalized medicine, or more effective treatment based on individual health data paired with predictive analytics, is also closely related to better disease assessment. The domain is presently ruled by supervised learning, which allows physicians to select from more limited sets of diagnoses, for example, or estimate patient risk based on symptoms and genetic information. Personalized medicine (PM) aims to tailor disease prevention, diagnosis, and treatment to individuals based on their genes, lifestyle and environments. Patients and interested organizations may potentially play an important role in the realization of PM.

## Background

PM relies on building statistical models able to predict personalized treatments, tailored to the characteristics of specific patients. In turn, developing such statistical models call for large amounts of patient data. However, in practice, patients often prefer not to divulge certain kinds of data due to privacy concerns.

Medical researchers use survival models to evaluate the significance of prognostic variables in outcomes such as death or cancer recurrence and subsequently inform patients of their treatment options. One standard full list of author information is available or related to the article survival model is the Cox Proportional Hazards model (CPH). The CPH is a semi parametric model that calculates the effects of observed covariates on the risk of an event occurring (e.g. 'death'). The model assumes that a patient's log-risk of failure is a linear combination of the patient's covariates. This assumption is referred to as the linear proportional hazards condition. However, in many applications, such as providing personalized treatment recommendations, it may be too simplistic to assume that the log-risk function is linear. As such, a richer family of survival models is needed to better fit survival data with nonlinear log-risk functions.

## Methodology

To model nonlinear survival data, researchers have applied three main types of neural networks to the problem of survival analysis. These include variants of: (i) classification methods, (ii) time encoded methods (iii) and risk predicting methods This third type is a feed-forward neural network (NN) that estimates an individual's risk of failure. In fact, Faraggi-Simon's network is a nonlinear extension of the Cox proportional hazards model.

In this section, the main methodology for providing personalized treatment recommendations using DeepSurv is explained. First, the architecture and training details of DeepSurv is explained which is an open source Python module that applies recent deep learning techniques to a nonlinear CPH network. Second, DeepSurv is defined as a prognostic model and how to use the networks predicted log-risk function to provide personalized treatment recommendations is shown.

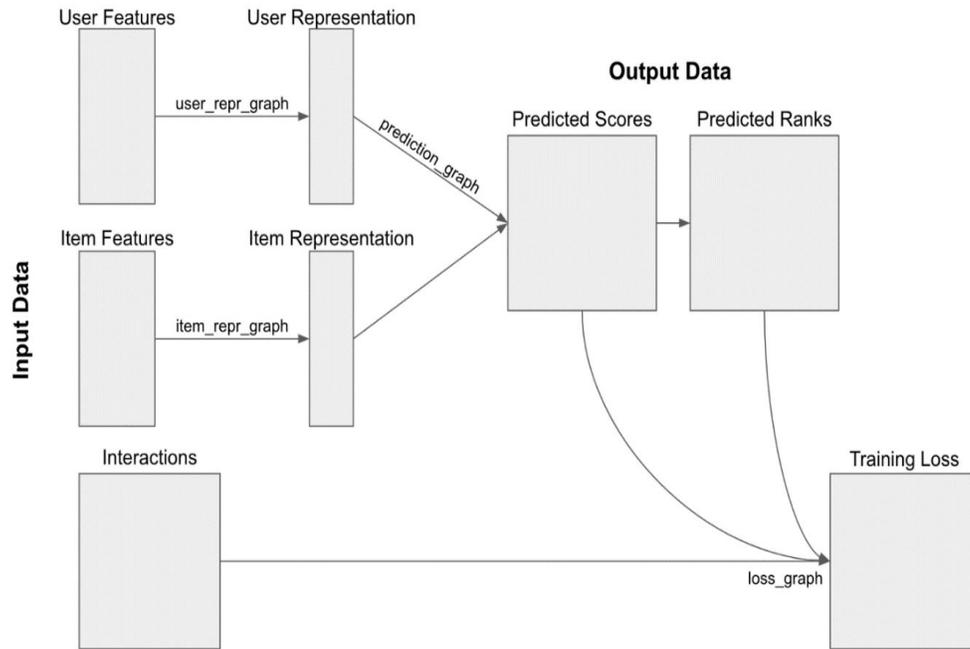


Fig:1 Recommendation system using Deep learning

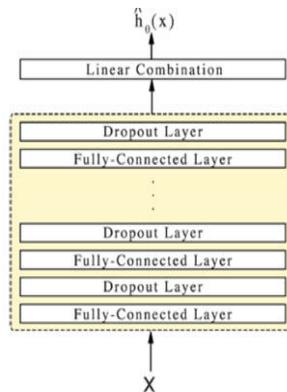


Fig:2 Diagram of the DeepSurv

## **Experimental Design**

### *Dataset*

Dataset such as ICCR will be used for experimentation and evaluation.

### *Evaluation Measures*

Measures such as accuracy of prediction of hazards will be computed by comparing the proposed treatment and hazards with the ones prescribed by the doctors. To evaluate the model's predictive accuracy on the survival data, the concordance-index (C-index) will be measured. The C-index is the most common metric used in survival analysis and reflects a measure of how well a model predicts the ordering of patients' death times.

## **Methodology**

Architecture of the DeepSurv is shown in figure 2.

### *Step 1: Data collection and dataset preparation*

Survival data will be taken from dataset ICCR and also collected from local hospitals.

### *Step 2: Developing a DeepSurv based treatment recommendation system*

A recommended treatment for each patient in the test set will be built using DeepSurv and the Random Survival Forest (RSF). Without preselected treatment-interaction terms, the CPH model will compute a constant recommender function and recommend the same treatment option for all patients. This would effectively be comparing the survival rates between the control and experimental groups. DeepSurv and the RSF can predict an individual's hazard per treatment because each compute relevant interaction terms.

### *Step 3: Training and experimentation on datasets*

The DeepSurv model will be trained on the large scale datasets.

### *Step 4: Deployment and analysis on real life scenario*

The trained and tested recommender system will be developed in real-life scenario where historical medical records of cancer patients will be collected from local hospitals.