Estimating Survival Days of Transplanted People. A Linear Regression Approach.

ZUYAO LI AND JUAN GABRIEL MARTÍNEZ

Introduction

- What is transplantation?
- Which criteria is applied?
  - Medical Urgency
  - Waiting Time
- 2010, Germany: 25 % less heart donors, waiting list doubled.
- Problem: Organs transplanted to very sick people.
- Result: Low expected outcome and non-optimal use of organs.

LUND
UNIVERSITY

Outline

- Method
  - Working with RDF and SPARQL
  - Parsing the Files
  - Prediction Method
- Results
- Demo

Method

- Set of data provided by Lund's hospital.
  - 95,639 individuals
  - 74 features
- Data provided in RDF format in three files:
  - Properties.ttl
  - Skos_ISHTL.ttl
  - ISHTL.ttl

Working with RDF and SPARQL

- Sesame framework to work with RDF data.
- SPARQL queries to obtain data.
- Script in Python using sparql-client.
- Problem: Missing values.
  - SPARQL does not work with NULL values.
  - No information could be retrieved for 74 features.
- Fastest solution: Parse the files.

LUND UNIVERSITY

Parsing the Files

- Parse Properties.ttl to obtain the name of the 74 features.
- Parse ISHTL.ttl to obtain information from all the individuals.
- What happens with the missing information?
  - Delete all individuals with a missing attribute.
  - Select value from a subset with similar values.
  - Select most frequent value for that feature.

Prediction Model

- The target value is a number
- We choose linear regression model
- Use Stepwise model selection method to get the linear regression model formula.
- The model starts from intercept model
- Each step it add predictors until the p-value of predictors reach the threshold

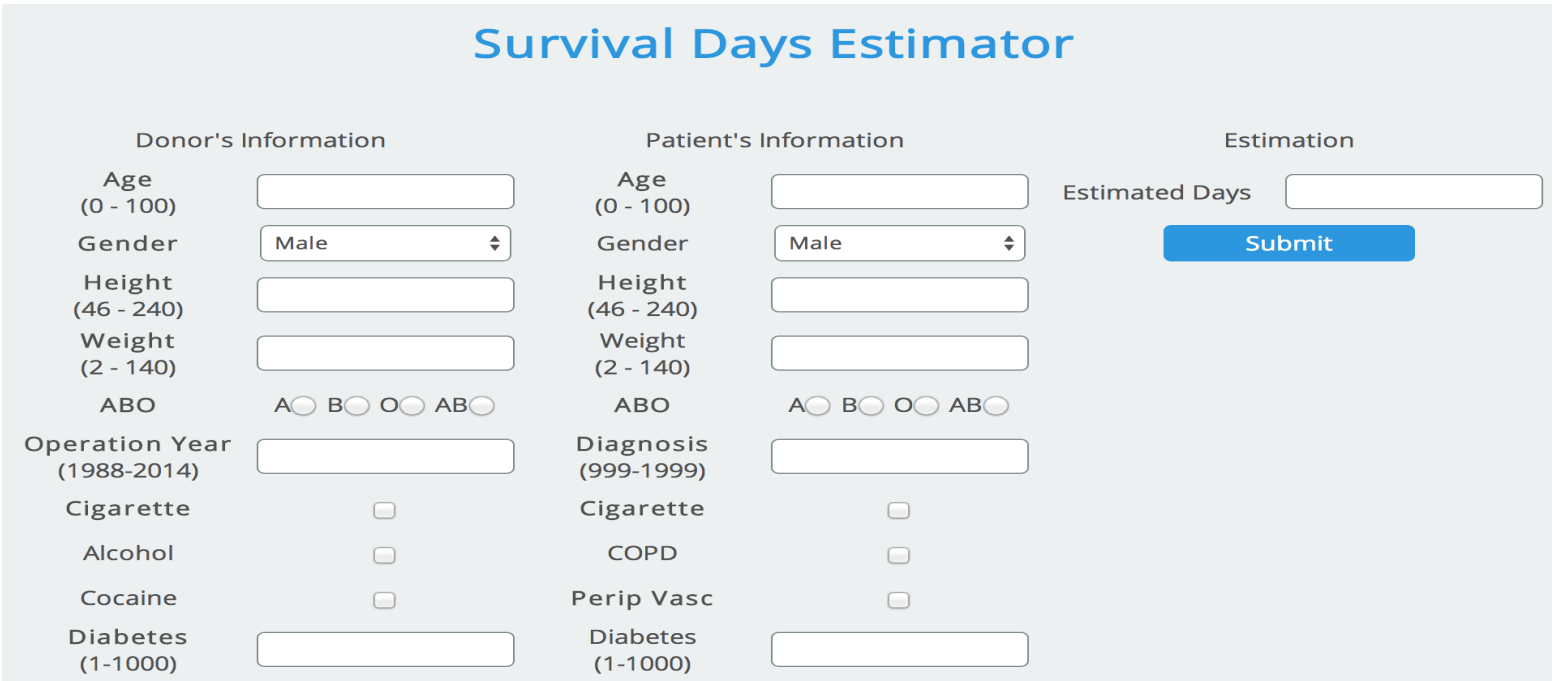
Linear regression model

Features	Weight	P - Value
age	-11.673	0.0031788
diagnosis	-1.3251	0.00039242
history_cigarette_Y	1.0395e+05	1.06e-09
age_donor	-999.21	8.9229e-15
gender_donor_F	-24812	1.1612e-08
age*age_donor	-0.20325	1.5689e-15
diagnosis*age	0. 0061983	0.0198
age*gender_donar_F	3.1292	4.5424e-05

Results

- Obtained 32 predictors from 74 features
- Number of observations: 95,639
- Root Mean Squared Error: 1.58e+03
- R-squared: 0.385
- Results obtained from the set of data

Demo



http://130.235.17.116:8000/surgery-form/

LUND UNIVERSITY

LUND
UNIVERSITY