

利用生醫大數據及演算來預測漸凍人病程

Predicting Disease Progression of ALS using Biomedical Big Data

指導教授：楊士德

專題成員：楊震

開發工具：R, Python 2.7

測試環境：Linux on IBM Cloud

一、簡介:

肌萎縮性脊髓側索硬化症（Amyotrophic lateral sclerosis，ALS），或稱漸凍人，是一種病程發展快速的神經退化性疾病，從發病到死亡平均 3 至 5 年的時間，病人之間卻也存在很大的差異性（物理學家史蒂芬·霍金，距他被診斷為漸凍人至今，已逾 50 年）。希望藉由大數據的分析，縮短藥物開發的時程，為治癒 ALS 找到一線曙光。

我們使用漸凍人功能性量表（ALS Functional Rating Scale，ALSFRS）的變化率作為病程進展的指標，並且用確診後 0-3 個月的分數，預測 3-12 個月的情形。

我們使用超過 7000 位病人的資料當成訓練集，用以預測約 650 位病人的病程，這些病患的資料來自於 PRO-ACT（Pooled Resource Open-Access ALS Clinical Trials Database）。

特徵選取的部分，我們計算了特徵的熵，並依此得出特徵的候選名單。候選名單中的特徵，相對於原始資料中的其他特徵，與預測目標有更高的相關性。而結果的預測，我們選用 Gradient Boosted Regression Trees (GBRT) 作為預測的方法。

二、結果:

我們選出的最佳特徵為”onset-delta”(診斷至今的天數)，“trunk”(ALSFRS-軀幹)，“Phosphorus”(磷)，“Q1-Speech”(ALSFRS-Q1)，“Q5-Cutting”(ALSFRS-Q5)，“leg”(ALSFRS-腿)

預測結果：



