Rheumatoid Arthritis Responder Challenge: Team Lucia

Victor Bellon1,2,3, Chloé Agathe Azencott1,2,3, Véronique Steven1,2,3, Olivier Collier1,2,3, Azadeh Khaleghi1,2,3, Valentina Boeva1,2,3 and Jean Philippe Vert1,2,3

1. INRS-ENSM, Varennes, Quebec, Canada
2. UMR 1928, INRA, Nancy, France
3. INRAE, Nancy, France

Introduction

Rheumatoid arthritis is an autoimmune disease that results in a chronic inflammatory disorder that principally attacks flexible joints and can lead to substantial loss of functioning and mobility. The disease is usually controlled by treatment with drugs that try to suppress the activity of TNF-α, which plays a major role in the immunological response and in the inflammation pathway. However, up to 30 per cent of the cases there is no response to the treatment. Hence, it is important predict which will be the response of a certain patient, to a certain treatment.

Data

- **Genotype Data**: 2.5 million SNPs for 3766 individuals
- **Clinical Data**: 6 treatments
  - Age
  - Sex
  - DAS before treatment
- **Outcome Data**: ADAS, Responder or not

SNPs selection

**Biological dataset**

- SNPs
- Thalassemia related genes
- Drug activity modulators
- Drug targets
- 12,737 SNPs
- 3,840 SNPs

**Statistical dataset**

- SNPs
- 2.5 million
- 3,000 SNPs per treatment
- Statistically selected SNPs

We have created two sets of selected features.

The first one contains SNPs extracted from the literature. Those SNPs are located in the exons of drug targets, drug action regulators and TNF-α related genes.

The second one contains SNPs selected using their Mutual Information Score. The number of used SNPs is obtained through 10-fold cross-validation.

Prediction with Support Vector Regression

We used a stratified model, i.e. we trained a different model for each treatment.

To deal with the large amount of variables we used Support Vector Regression and the kernel trick.

We calculated a different kernel for each one of the variables, and we combined them using their mean.

**Variable**

- SNP
- Sex
- Age

**Kernel**

- MinMax
- Dirac
- Linear

Final results

We obtained the second best performance.

We observe no difference between using MinMax or Linear Kernel and using Dirac Kernel, which can be considered as using only the clinical variables.

Correlation

10 Fold Cross Validation

- Human genetic and clinical model
- Only clinical model
- Linear genetic and clinical model

References


Rheumatoid Arthritis Responder Challenge https://www.synapse.org/#/Synapse/syn1734172