Measure of dissimilarity

Aleksandr Savenkov
ols2010@med.cornell.edu

02/07/2017

Intro

- Goal: to define a measure of similarity between time series.

- Clustering

- ANOVA type problem

Definition

Consider two time series:

- X_t, $t = 1, ..., n$
- Y_t, $t = 1, ..., m$.

Assume ARMA model M_1 estimated based on x_t and ARMA model M_2 estimated based on y_t.

- Apply M_1 to time series X_t to obtain prediction error σ_{11}^2
- Apply M_2 to time series X_t to obtain prediction error σ_{12}^2
- Apply M_1 to time series Y_t to obtain prediction error σ_{21}^2
- Apply M_2 to time series Y_t to obtain prediction error σ_{22}^2

Measure of dissimilarity

Define the measure of dissimilarity by

$$d(X_t, Y_t) = \frac{\sigma_{12}^2 - \sigma_{11}^2}{\sigma_{11}^2} + \frac{\sigma_{21}^2 - \sigma_{22}^2}{\sigma_{22}^2}$$

The proposed measure satisfies following properties:

1. $d(X_t, Y_t) > 0$
2. $d(X_t, Y_t) = 0$ iff $X_t = Y_t$
3. $d(X_t, Y_t) = d(Y_t, X_t)$

Test statistic

The test statistic for H_0 that there is no group difference is

$$T = \frac{\text{average between group differences}}{\text{average within group differences}}$$

- ► BD - all between groups differences
- ► WD - all within groups differences

Permutation test

1. Calculate statistics T
2. Evaluate value T_i for each permutation, $i = 1, ..., N$
3. Approximate p-value as follow

$$\hat{p} = \frac{\sum_{i=1}^{N} I(T_i > T)}{N}$$

Next steps

- Literature search

- Simulation study