02 Find a good number of salient patterns in a matrix

～Infinite plaid models for infinite bi-clustering～

Abstract

Our goal is to find salient bi-clusters from a given relational data matrix automatically. Salient bi-clusters are sub-matrices that have distinct values from other entries of the data matrix. Such bi-clusters often corresponds to informative subsets of the data; e.g. ``good customer groups with best-selling items for them'', and ``specific gene clusters that are reactive for a specific treatment/chemicals''.
Conventionally, bi-clustering requires us to specify the number of bi-clusters to be extracted before the analysis. However it is generally difficult to know the number of bi-clusters before conducting an actual analysis.
Our proposed model enables us to forget about this specification of the number of bi-clusters. The model automatically infer an appropriate number of bi-clusters (up to infinite!) for the given data matrix, and performs effective bi-clustering. This model will help users to conduct ``easy-to-go'' bi-clustering for several situations.

Bi-clustering: extract salient bi-clusters (sub-matrices) with distinct values, compared to the other entries of the data matrix
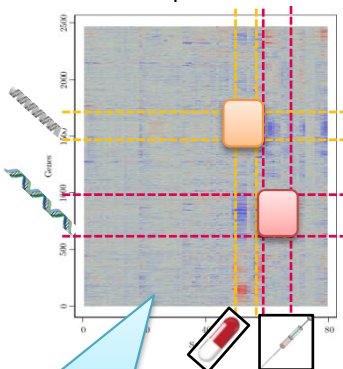
bi-cluster (sub-matrix) = row cluster X column cluster

Relational data matrix
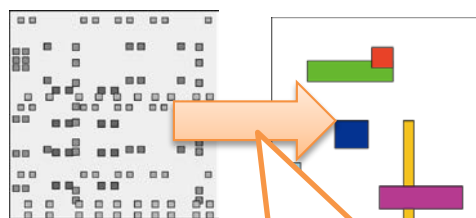
Permutation + Clustering (rows & columns)

Gene expression data

{treatment patterns} x {reactive gene clusters}

User-item purchase data

{good customer patterns} x {best-selling item clusters}

We need to specify the number of bi-clusters, even though they are hidden in the matrix ☹
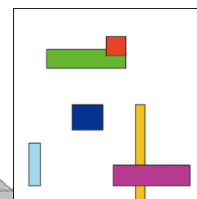
K = 3? K = 6? K = many?

Our proposed model can extract bi-clusters of a unknown number(<∞) from an arbitrary matrix

☺ Automatically infer the appropriate number of bi-clusters → Easy-to-go bi-clustering!!

K = 6!!

【Reference】

[1] K. Ishiguro, I. Sato, M. Nakano, A. Kimura, N. Ueda, "Infinite plaid models for infinite bi-clustering,'' in *Proc. 30th AAAI Conference on Artificial Intelligence (AAAI)*, 2016.

【Contact】
Katsuhiko Ishiguro   Learning and Intelligent Systems Research Group, Innovative Communication Laboratory
E-mail : ishiguro.katsuhiko(at)lab.ntt.co.jp