A low rank optimization based method for single cell data analysis

Abstract:
Recent advances in single cell technology enable researchers to study heterogeneity of cell populations and dynamics of gene expression in individual cell level. One of the main challenges is to extract the salient features in a manner that reveals the underlying dynamics process.

An optimization method, Single-cell Low Rank Similarity-based Method (ScLRSM), is proposed for identifying cell types associated with cell differentiation and detecting cell lineage from single-cell gene expression data. ScLRSM constructs structured cell-to-cell similarity matrix based on a low rank optimization model and cell types can be derived directly through the similarity matrix using non-negative matrix factorization. The number of cell types is determined automatically via computing the eigenvalue gaps of the constructed consensus matrix while the vast majority of algorithms require the prior knowledge of such a number. In particular, the temporal order of cells is estimated by the non-negative rank one approximation of the cell-to-cell similarity matrix, which captures the global structure of the whole data. Cell lineage is inferred by constructing the minimum spanning tree of the weighted cluster-to-cluster graph.

We applied our method to three different single cell data sets with known lineage and developmental time information from both mouse early embryo and human early embryo. ScLRSM successfully identifies the cell subpopulations within different developmental time stages and reconstructs cell differentiation trajectories which is agreed with the previously experiments. The current results demonstrate the potential and high accuracy of the proposed method in determining cellular differentiation states and reconstructing cell lineages from single cell gene expression data.

Host: Jiawang Nie

Wednesday, May 31, 2017
4:00 PM
AP&M 5402

------------------------------------------