

Flow cytometry analysis based on projection pursuit automated gating of data nuggets

Mahan Dastgiri, Rebecca Mary Peters, PoChung Chou, Ge Cheng, Yajie Duan, Chun-Pang Lin, Jocelyn Sendeki, Helena Geys, Kanaka Tatikola, Davit Sargsyan, Ah-ng Kong, Javier Cabrera

Abstract

Projection pursuit creates low dimensional (1- or 2-D) projections that display the main structure of high dimensional data. We proposed an application of this approach for flow cytometry data analysis. Flow cytometry is a powerful, well-established tool in drug discovery that helps detect and identify cell populations by measuring their physical and biochemical characteristics. When applied to flow cytometry data, projection pursuit produces a small number of projections that visualize the cluster structure of the data. These projections involve a few variables in each dimension, and the structure is generally much clearer than that found in ordinary 2-D projection plots. If the number of cells in the samples is very large, the data can be reduced by summarizing it with a few thousand data nuggets making it possible to find the optimal projections at the highly reduced computational cost.

Our new methodology is a combination of these two techniques, projection pursuit and data nuggets, that results in a more efficient clustering that can be automated and applied to flow cytometry data to find regions differentially populated across experimental conditions such as genotype or treatment. This is an alternative to the traditional gating strategies employed in the flow cytometry data analysis.