

Compressed Sensing in Disease Classification

Anna Ramisch and Martin Vingron

When classifying diseases based on gene expression data it is often desirable to find a small number of disease related genes rather than a huge set of genes. We model the situation as a linear system. Since the number of patients is typically much smaller than the number of genes, this system will not have a unique solution but infinitely many. However, the sparsity assumption on the gene set makes it possible to fit the disease classification task into the mathematical theory of Compressed Sensing. With its many theoretical results, Compressed Sensing is a new technique for efficient reconstruction of solutions to under-determined linear systems under the requirements of sparsity. Following this approach, we propose a new classification model with which we classify Non-Small Cell Lung Cancer data. We use two recovery methods with different approaches. The Orthogonal Matching Pursuit is an iterative Greedy algorithm to determine the support of sparse coefficients and the LASSO is one of the standard convex optimization techniques for ℓ_1 minimization and sparse regressions. In comparison to state-of-the-art methods in classification, our algorithms show a fast performance and comparably good accuracy results. In the end, we study the resulting gene sets and their biological relevance which is very convincing.