

Sequential Support Vector Regression Model with Embedded Entropy for Disease Prediction and SNP Selection

Yulan Liang^{*}, University of Maryland, Baltimore
Arpad Kelemen, University of Maryland, Baltimore

Abstract

Comprehensive evaluation of common genetic variations through association of SNP structure with common diseases on the genome-wide scale is currently a hot area in human genome research. For less costly and faster diagnostics, advanced computational approaches are needed to select the minimum SNPs with the highest prediction accuracy for common complex diseases. In this paper, we present a sequential support vector regression model with embedded entropy algorithm to deal with the redundancy for the selection of the SNPs that have best prediction performance of diseases. We implemented our proposed method for both SNP selection and disease classification, and applied it to simulation data sets and two real disease data sets. Results show that on the average, our proposed method outperforms the well known methods of Support Vector Machine Recursive Feature Elimination, logistic regression, CART, and logic regression based SNP selections for disease classification and prediction.

Keywords: Support vector regression; Sequential algorithm; Entropy measures; Embedded methods; Sliding window; Single nucleotide polymorphism; Common complex disease.

^{*} Presenting author