広島統計談話会

Hiroshima Statistics Study Group

第 295 回談話会を下記のように開催致しますので
御参集下さいますようご案内申し上げます。
You are cordially invited to the 295th meeting as scheduled below.

日　時：　2015 年 12 月 4 日（金）15:00 –
Date　：　December 4th, 2015 (Fri) 15:00 –

場　所：　放射線影響研究所 比治山ホール
Place　：　RERF Hijiyama Hall

演　者：　植木 優夫 博士　（久留米大学 バイオ統計センター）
Speaker　：　Masao Ueki, Ph.D.
Biostatistics Center, Kurume University

演　題：　「不偏なモデル選択による円滑閾値を用いた遺伝的予測」
Title　：　“Smooth-threshold multivariate genetic prediction with unbiased model selection”

要　約：

Summary:

We develop a new genetic prediction method, smooth-threshold multivariate genetic prediction, using single nucleotide polymorphisms (SNPs) data in genome-wide association studies (GWASs). Our method consists of two stages. At the first stage, unlike the usual discontinuous SNP screening as used in the gene score method, our method continuously screens SNPs based on the output from standard univariate analysis for marginal association of each SNP. At the second stage, the predictive model is built by a generalized ridge regression simultaneously using the screened SNPs with SNP weight determined by the strength of marginal association. Continuous SNP screening by the smooth-thresholding not only makes prediction stable but also leads to a closed form expression of generalized degrees of freedom (GDF). The GDF leads to the Stein’s unbiased risk estimation (SURE) which enables data-dependent choice of optimal SNP screening cutoff without using cross-validation. Our method is very rapid because computationally expensive genome-wide scan is required only once in contrast to the penalized regression methods including lasso and elastic net.

Simulation studies which mimic real GWAS data with quantitative and binary traits demonstrate that the proposed method outperforms the gene score method and genomic best linear unbiased prediction (GBLUP), and also shows comparable or sometimes improved per- formance with the lasso and elastic net being known to have good predictive ability but with heavy computational cost. Application to whole-genome sequencing (WGS) data from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) exhibits that the proposed method shows higher predictive power than the gene score and GBLUP methods.