Application of Deep Learning in Genomic Selection

Yang Liu

Informatics Institute,
University of Missouri, Columbia
Columbia, MO USA
Ylmk2@mail.missouri.edu

Duolin Wang

Christopher S. Bond Life Sciences
University of Missouri, Columbia
Columbia, MO USA
wangdu@missouri.edu

Abstract— **Genomic selection (GS) is a marker-assisted selection approach to enhance quantitative traits in breeding population in which whole genome single-nucleotide polymorphisms (SNPs) markers can be used to predict breeding values (BV). GS has been proved to increase breeding efficiency in both plant and animal breeding, such as dairy cattle, pig, rice, soybean and loblolly pine. Here, we propose a deep-learning model using convolutional neural network (CNN) to predict genomic estimated breeding value (GEBV) and also to investigate neighboring SNP effects within linkage disequilibrium. We have applied our models on two datasets: 1) grain yield (YLD) trait on *Glycine Max (*soybean) nested association mapping (NAM) dataset and 2) stem height (HT) trait on a *Pinus taeda* (loblolly pine) dataset. The SoyNAM population contains 4,313 SNPs from 5,139 individuals with a trait heritability of 0.345 and the Loblolly Pine population contains 4,853 SNPs from 861 individuals with a trait heritability of 0.31. Our deep-learning model was tested with a 10-fold cross-validation, run in parallel on graphic processing units (GPUs). Our model prediction accuracy, which was calculated by Pearson Correlation between GEBV and observed values, outperforms traditional statistical RR-BLUP, Bayesian LASSO and BayesA models. The results indicate that deep-learning model is efficient in accurately computing breeding values and simultaneously studying nearby SNP effects from CNNs. It also indicates powerful potential in interpreting phenotype-genotype associations over the entire genome.**

Keyword: Genomic selection, soybean, deep learning, breeding