Statistical Method for Increasing the Accuracy of Genomic Prediction

Tom Flemming*

Department of Nano Science Technology, University of Waterloo, Waterloo, Canada

INTRODUCTION

The Genome-Wide Association Study (GWAS) has significant success in pinpointing the genetic regions responsible for a variety of complex traits and disorders. Important genetic insights into the pathophysiology of diseases are provided by the findings. We provide an overview of commonly employed methods and tactics for GWAS analysis, along with a basic approach to handling GWAS data.

Other advanced topics like the problem of missing heritability, meta-analysis, set-based association analysis, copy number variation analysis, and GWAS cohort analysis are also briefly introduced. The issues regarding data quality control, population structure, association analysis, multiple comparison, and visual presentation of GWAS results are also discussed. In the presence of outliers, it is anticipated that the predictive ability of genetic prediction algorithms may suffer.

An outlier in agriculture science might appear as a result of incorrect data imputation, an outlying response, or a sequence of trials conducted at different times or locations.

Even though there are numerous statistical techniques for outlier detection in the literature, finding a genuine outlier is still difficult, especially when dealing with high dimensional genomic data.

Here, we have outlined an effective strategy for identifying outliers in high-dimensional genomic data. Our method uses a variety of p-value based methods to provide a single p-value for outlier detection.

DESCRIPTION

Through the use of evaluation metrics including precision, recall, and others, the robustness of our technique has been examined using simulated data.

By identifying outliers and managing them appropriately utilising our suggested approach and real data, it has been seen that the performance of genomic prediction has improved significantly.

Advances in genetics, statistical genetics can be seen as a traditional topic of applied probability and statistics that has recently experienced significant growth.

The association between gene features, such as diseases and genetic susceptibilities, must be studied in an unprecedented way given the availability of current technology, new approaches, and much more data information.

In applied statistics, applied mathematics, biological/medical studies, and other allied sciences, this area is one of the most popular research areas.

In the field of plant and animal science, Genomic Selection (GS) is a common method for choosing suitable candidates for breeding. Numerous studies have recently been conducted. In GS, a cutting-edge breeding technique, the genetic merit of each individual in a breeding programme is predicted using data from genome wide density markers. GS is a promising strategy for enhancing the genetic gain of study subjects in the current environment. Meuwissen, et al., first to introduce genomic selection? In this method, the individual effects of each marker are calculated, and the combined effects of all the markers are used to calculate each individual's genotypic value, or Genome Estimated Breeding Value (GEBV).

The first step in the GS process is to create a statistical model using individuals who have both genotypic and phenotypic data (also known as the training set). This model is then used to estimate GEBVs for individuals who only have genotypic data. Next, people are sorted according to their GEBVs, and then the best people are chosen. Plants and animals have successfully used genomic selection techniques. However, the quality of the data necessary for using the various statistical models determines whether genomic selection is successful. However, in real world situations, the quality of genomic data frequently falls short of the ideal standard and is frequently constrained by factors like the presence of significant observations, missing data, noise, etc.

CONCLUSION

The Genome-Wide Association Study (GWAS) has been extremely effective at identifying the genetic loci underlying a wide range of complex traits and illnesses. Additionally covered

Correspondence to: Tom Flemming, Department of Nano Science Technology, University of Waterloo, Waterloo, Canada; E-mail: flemmington@edu.ca

are the problems with data quality assurance, population structure, association analysis, multiple comparison, and the visual presentation of GWAS results. The genetic quality of each individual in a breeding programme is predicted using information from genome-wide density markers in the cutting-edge breeding technique known as genomic selection. The genotypic value of each individual is determined using this method by combining the individual effects of each marker with those of the other markers.