Noname manuscript No.
(will be inserted by the editor)

Developing a parsimonius predictor for binary traits in sugar beet (*Beta vulgaris*)

Filson Nazzarini · Simone Marini · Piergiorgio Stevanato · Nelppo Biscicari

Received: 05 August 2014 / Accepted:

Abstract Insert your abstract here. Include keywords, PACS and mathematical subject classification numbers as needed.

Keywords binary traits · genomic predictions · parsimonious predictor · sugar beet

1 Introduction

binary traits in plants and animals: why they are important in practice Sugar beets in particular: we work on root vigor [1]. The concept of parsimony: when many possible predictors are available, it is useful to select a subset to limit analysis cost and time. Moreover: use the minimun necessary information set, occam razor, and so forth.

As the technology advances, and available predictors grow, not only the prediction precision becomes important, but also the actual cost must be considered.

Filippo Biscarini and Nelson Nazzicari contributed equally to the work.

F. Biscarini
Fondazione Parco Tecnologico Padano
Tel.: +123-45-678910
E-mail: filippo.biscarini@tecnoparco.org

S. Marini
second address

N. Nazzicari
Fondazione Parco Tecnologico Padano
Tel.: +123-45-678910
E-mail: nelson.nazzicari@tecnoparco.org

P. Stevanato
address

In this paper we propose statistical methods to highlight and select the most useful predictors given a set. We started on real world data and validated our approach on a XXX dataset. We found that it is possible to strongly reduce the dimension of the predictors set and still achieve high performance.

2 Material and methods

Text with citations [3] and [2].

2.1 Plant material and SNP genotypes

Root vigor. Available data. SNP technology used, imputation.
Copypaste from other articles. Dataset description.

2.2 Development of a parsimonious predictor

we use a two steps approach: - a ranker to rank the various available predictors (SNPs in our case). We used the BOSS algorithm - this is an iterative step. we progressively reduced the predictors set, taking away the laest useful predictor and applying to the resulting subset a ridge logistic regression apprach. Thus, we obtained as many performances estimation as the number of original predictors.

Rank of predictors This explain the BOSS algorithm

Selection of predictors and classification method We take one predictor out at each iteration You put the model formula for ridge logistic regression

Classification error Cross validation: how many times, what fractions. Explanation of error rate and other parameters (ROC?)

2.3 Comparison with another method to rank predictors

Another ranker: why use one, and its description. P value and SNP effect (as it is done in GWAS)

2.4 Software

R, weka, perl.

3 Results

Possible charts: - Precision as a function of the number of predictors. - Breakdown of two types of error.

If possible: probability of assignment as a function of predictors, maybe with ROC curve? Maybe in discussion?

4 Discussion

General overview why error rates are not evenly distributed? Reminder: it works very well because of LD and H2

4.1 SNP effects

Manhattan plot with BOSS weights and weights from the other articles, somehow compared (same chart? two charts? only ten best?).

Do the peaks make sense from the biological perspective?

4.2 Relative performance of rankers

why using Pvalues and not other standard rankers (e.g. backward stepwise selection)? Because of the specific nature of the problem

4.3 Genotyping strategies and applications to breeding

genotyping strategies: Costs, possible technologies (gbs, snp chip, macroarrays), implications

applications to breeding: why is it important root vigor early detection. Other binomial traits (e.g. disease resistance)

5 Conclusions

Concluding remarks

Acknowledgements Do we need to ack somebody? (projects?)

References

1. Biscarini, F., Stevanato, P., Broccanello, C., Stella, A., Saccomani, M.: Genome-enabled predictions for binomial traits in sugar beet populations. BMC Genetics **18**(5), 1–9 (2014)

2. Saccomani, M., Stevanato, P., Trebbi, D., McGrath, J.M., Biancardi, E.: Molecular and morpho-physiological characterization of sea, ruderal and cultivated beets. Euphytica **169**(1), 19–29 (2009)
3. Stevanato, P., Broccanello, C., Biscarini, F., Del Corvo, M., Sablok, G., Panella, L., Stella, A., Concheri, G.: High-throughput rad-snp genotyping for characterization of sugar beet genotypes. Plant Molecular Biology Reporter pp. 1–6 (2013)