Genomic prediction in Nordic Holstein population using a single-step approach

Hongding Gao, Ole F. Christensen, Per Madsen
Ulrik S. Nielsen, Yuan Zhang, Mogens S. Lund, Guosheng Su

Dept. of Molecular Biology and Genetics
Aarhus University, Denmark

August 29, 2011
EAAP Stavanger, Norway

AARHUS
UNIVERSITY
DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

Background

- Not all individuals can be genotyped in practice
- Expectation: reliability enhanced by blending genomic information with traditional EBV
- Single-step blending is theoretically superior over multi-step blending approaches

Single-step blending method

GEBV: Phenotypes + Pedigree info. + Marker info.

Superiority: All information is used to predict GEBV simultaneously

Objective

Compare genomic prediction models in Nordic Holstein data

- GBLUP
- Selection index blending
- Single-step blending

Data

- De-regressed proof (DRP)
- Genotyped bulls (5,214)
- Non-genotyped bulls (9,374)
- Pedigree animals (42,144)
- SNP markers (48,073)
- 16 traits (sub-indices) in the Nordic Total Merit (NTM) index

Reference and validation datasets

- Partition the DRPs into reference and validation data
- Reference data: $<$ 2001-10-01
- Validation data: $>$ 2001-10-01
- Reference data of single-step blending method include all the animals have DRPs with and without genotyped before the cut-off date

GBLUP model (VanRaden, 2008; Hayes et al., 2009)

- Model

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{g} + \mathbf{e}$$

where $\mathbf{g} \sim N(\mathbf{0}, \mathbf{G}\sigma_g^2)$, $\mathbf{e} \sim N(\mathbf{0}, \mathbf{D}\sigma_e^2)$

- $\mathbf{D}$ is a diagonal matrix with $d_{ii} = 1/w_i$
- w_i is a weighting factor with $w_i = r_{DRP}^2/(1 - r_{DRP}^2)$
- Account for heterogeneous residual variances due to difference in reliabilities of DRP

Single-step blending

Based on Christensen and Lund, 2010

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

$\mathbf{G}_p$: Modified genetic relationship matrix by combining with pedigree

$$\mathbf{G}_p = \begin{bmatrix} \mathbf{G}_\alpha & \mathbf{G}_\alpha \mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{G}_\alpha & \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{G}_\alpha\mathbf{A}_{11}^{-1}\mathbf{A}_{12} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{bmatrix}$$

Single-step blending

Based on Christensen and Lund, 2010

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

$\mathbf{G}_p$: Modified genetic relationship matrix by combining with pedigree

$$\mathbf{G}_p = \begin{bmatrix} \mathbf{G}_\alpha & \mathbf{G}_\alpha \mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{G}_\alpha & \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{G}_\alpha\mathbf{A}_{11}^{-1}\mathbf{A}_{12} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{bmatrix}$$

$\mathbf{A}_{11}$: a sub-matrix of $\mathbf{A}$ for genotyped animals

AARHUS UNIVERSITY
DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

Single-step blending

Based on Christensen and Lund, 2010

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

$\mathbf{G}_p$: Modified genetic relationship matrix by combining with pedigree

$$\mathbf{G}_p = \begin{bmatrix} \mathbf{G}_\alpha & \mathbf{G}_\alpha \mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{G}_\alpha & \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{G}_\alpha\mathbf{A}_{11}^{-1}\mathbf{A}_{12} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{bmatrix}$$

$\mathbf{A}_{22}$: a sub-matrix of $\mathbf{A}$ for non-genotyped animals

Single-step blending

Based on Christensen and Lund, 2010

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

$\mathbf{G}_p$: Modified genetic relationship matrix by combining with pedigree

$$\mathbf{G}_p = \begin{bmatrix} \mathbf{G}_\alpha & \mathbf{G}_\alpha \mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{G}_\alpha & \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{G}_\alpha\mathbf{A}_{11}^{-1}\mathbf{A}_{12} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{bmatrix}$$

$\mathbf{A}_{12} = \mathbf{A}'_{21}$: sub-matrices of $\mathbf{A}$ describing the relationship between genotyped and non-genotyped animals

AARHUS UNIVERSITY
DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

Single-step blending

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Za} + \mathbf{e}$$

$\mathbf{G}_p$: Modified genetic relationship matrix by combining with pedigree

$$\mathbf{G}_p = \begin{bmatrix} \mathbf{G}_\alpha & \mathbf{G}_\alpha \mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{G}_\alpha & \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{G}_\alpha\mathbf{A}_{11}^{-1}\mathbf{A}_{12} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{bmatrix}$$

$$\mathbf{G}_\alpha = (1-\omega)\mathbf{G} + \omega\mathbf{A}_{11}$$

Use 8 ω to weight polygenic effect (0.05-0.40)

Single-step blending

- Inverse of $\mathbf{G}_p$ is

$$\mathbf{G}_p^{-1} = \begin{bmatrix} \mathbf{G}_\alpha^{-1} - \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \mathbf{A}^{-1}$$

Selection index blending (VanRaden et al., 2009)

The GEBV was obtained from a selection index including three items:

- Direct genomic value (DGV) from the GBLUP model
- Pedigree index($\mathrm{PI}_{\mathrm{sub}}$) calculated from a subset of data consisting of the genotyped bulls and using the **A** matrix
- $\mathrm{PI}_{\mathrm{full}}$ from the full dataset using the **A** matrix

A scale factor (0.85, 0.90, 0.95) was used on the DGV and the reliability of DGV to reduce the inflation of GEBV

Selection index blending (VanRaden et al., 2009)

The GEBV was obtained from a selection index including three items:

- Direct genomic value (DGV) from the GBLUP model
- Pedigree index($\mathrm{PI}_{\mathrm{sub}}$) calculated from a subset of data consisting of the genotyped bulls and using the **A** matrix
- $\mathrm{PI}_{\mathrm{full}}$ from the full dataset using the **A** matrix

A scale factor (0.85, 0.90, 0.95) was used on the DGV and the reliability of DGV to reduce the inflation of GEBV

Validation

- Reliabilities of GEBV were measured as squared correlations between predicted breeding value and DRP for bulls in the validation data and divided by reliability of DRP
- Unbiasedness of genomic predictions was measured as the regression of DRP on the genomic prediction
- The analyses of the GBLUP and single-step blending were performed using DMU package (Madsen et al., 2010)

Impact of different weights

The reliability and regression coefficients using single-step blending method

ω	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
Mean Rel.	0.377	0.379	0.379	0.379	0.378	0.377	0.375	0.372
Mean Dev.[1]	0.10	0.10	0.09	0.08	0.08	0.08	0.08	0.07

[1] Mean of absolute deviation from 1 for regression coefficient

Impact of different scale factors

Reliability and regression coefficients using selection index blending

	scale = 0.85	scale = 0.90	scale = 0.95
Mean Rel.	0.373	0.374	0.375
Mean Dev.[1]	0.082	0.084	0.090

[1] Mean of absolute deviation from 1 for regression coefficient

Reliabilities of genomic models

Trait	GBLUP	Index – blending[1]	Single – step[2]
Milk	0.431	0.447	0.452
Fat	0.455	0.453	0.459
Protein	0.429	0.425	0.436
⋮	⋮	⋮	⋮
Fertility	0.411	0.431	0.425
Mastitis	0.362	0.380	0.386
Mean	0.360	0.375	0.379

[1] Scale factor = 0.95

[2] weighting factor $\omega = 0.15$

AARHUS UNIVERSITY
DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

Reliabilities of genomic models

Trait	GBLUP	Index – blending[1]	Single – step[2]
Milk	0.431	0.447	0.452
Fat	0.455	0.453	0.459
Protein	0.429	0.425	0.436
⋮	⋮	⋮	⋮
Fertility	0.411	0.431	0.425
Mastitis	0.362	0.380	0.386
Mean	0.360	0.375	0.379

[1] Scale factor = 0.95

[2] weighting factor $\omega = 0.15$

Reliabilities of genomic models

Trait	GBLUP	Index – blending[1]	Single – step[2]
Milk	0.431	0.447	0.452
Fat	0.455	0.453	0.459
Protein	0.429	0.425	0.436
⋮	⋮	⋮	⋮
Fertility	0.411	0.431	0.425
Mastitis	0.362	0.380	0.386
Mean	0.360	0.375	0.379

[1] Scale factor = 0.95

[2] weighting factor $\omega = 0.15$

Reliabilities of genomic models

Trait	GBLUP	Index – blending[1]	Single – step[2]
Milk	0.431	0.447	0.452
Fat	0.455	0.453	0.459
Protein	0.429	0.425	0.436
⋮	⋮	⋮	⋮
Fertility	0.411	0.431	0.425
Mastitis	0.362	0.380	0.386
Mean	0.360	0.375	0.379

[1] Scale factor = 0.95

[2] weighting factor $\omega = 0.15$

Regression coefficients of DRP on genomic predictions

Trait	GBLUP	Index − blending[1]	Single − step[2]
Milk	0.92	0.95	0.95
Fat	0.88	0.92	0.89
Protein	0.85	0.87	0.86
⋮	⋮	⋮	⋮
Fertility	0.98	1.03	1.01
Mastitis	0.94	0.97	0.95
Mean Dev.[3]	0.11	0.09	0.09

[1] Scale factor = 0.95
[2] weighting factor $\omega = 0.15$
[3] Mean of absolute deviation from 1 for regression coefficient

AARHUS UNIVERSITY
DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

Summary

- Both single-step blending and selection index blending approach was more accurate than the GBLUP model
- Reliability of genomic predictions from the single-step blending was on average 0.4% higher than selection index blending model (used the same info. sources)
- The inflation of genomic predictions can be reduced through appropriate weighting factors in single-step blending and scale factors in selection index blending

Summary

- Both single-step blending and selection index blending approach was more accurate than the GBLUP model (1.9% and 1.5% higher on average) and less biased (0.02 reduction in mean absolute deviation)
- Reliability of genomic predictions from the single-step blending was on average 0.4% higher than selection index blending model (used the same info. sources)
- The inflation of genomic predictions can be reduced through appropriate weighting factors in single-step blending and scale factors in selection index blending

Summary

- Both single-step blending and selection index blending approach was more accurate than the GBLUP model
- Reliability of genomic predictions from the single-step blending was on average 0.4% higher than selection index blending model (used the same info. sources)
- The inflation of genomic predictions can be reduced through appropriate weighting factors in single-step blending and scale factors in selection index blending

Summary

- Both single-step blending and selection index blending approach was more accurate than the GBLUP model
- Reliability of genomic predictions from the single-step blending was on average 0.4% higher than selection index blending model (used the same info. sources)
- The inflation of genomic predictions can be reduced through appropriate weighting factors in single-step blending and scale factors in selection index blending

Conclusion

The single-step blending could be a feasible approach for genomic prediction in practical breeding programs