Are indirect genomic predictions a good option as the number of genotypes continues to rise?

Shogo Tsuruta, Daniela Lourenco, Yutaka Masuda, Ignacy Misztal - University of Georgia
Tom Lawlor - Holstein Association USA Inc.

Indirect Genomic Prediction

Reducing Computational Cost by GEBV and SNP for randomly selected genotyped animals
Introduction
Computing cost ↑ # genotyped animals ↑

How to minimize the cost?

1. Get a high-speed computer or develop a faster computing program
2. Use APY with a small number of core animals
3. Remove old data
4. Calculate genomic predictions separately and indirectly
   - which animals?
   Genotyped animals with no or less influence on other animals
   - target animals
   i.e., no progeny, no phenotypes
Data set

Phenotypes, pedigrees, and genotypes up to 2018 – full dataset for genomic prediction (benchmark)
10.3M records for 18 linear type traits for 13.6M animals
2.3M genotyped animals with 79K SNPs

Genomic prediction (GEBV)
for up to 13.6M animals using ssGBLUP with APY
except for target genotyped animals
(males: no progeny + females: no phenotypes)

Indirect Genomic Prediction (IGP)
# Target animals for IGP
- no progeny / no phenotypes

<table>
<thead>
<tr>
<th>YOB</th>
<th># genotyped animals for IGP (x 1000)</th>
<th># genoxygened animals for GEBV (x 1000)</th>
<th># IGP in %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Total</td>
</tr>
<tr>
<td>2014 - 18</td>
<td>145</td>
<td>1304</td>
<td>1449</td>
</tr>
<tr>
<td>2015 - 18</td>
<td>118</td>
<td>1171</td>
<td>1289</td>
</tr>
<tr>
<td>2016 - 18</td>
<td>89</td>
<td>1000</td>
<td>1089</td>
</tr>
<tr>
<td>2017 - 18</td>
<td>58</td>
<td>707</td>
<td>765</td>
</tr>
<tr>
<td>2018</td>
<td>23</td>
<td>257</td>
<td>280</td>
</tr>
<tr>
<td>All</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Method

1. Genomic prediction (GEBV) for animals with progeny or phenotypes

\[ \mathbf{u} = \lambda \mathbf{D} \mathbf{Z}' \mathbf{G}^{-1} \hat{\mathbf{a}} \]

(\(\hat{\mathbf{u}}\) =SNP effects, \(\lambda = \frac{\sigma_u^2}{\sigma_a^2}\), \(\mathbf{D}=I\), \(\mathbf{Z}=\text{#animals x #SNP matrix}\), \(\hat{\mathbf{a}}\)=GEBV)

\(\hat{\mathbf{u}}\) from \(\hat{\mathbf{a}}\) for all genotyped animals up to each year group or

\(\hat{\mathbf{u}}\) from \(\hat{\mathbf{a}}\) for randomly sampled genotyped animals

\[ \mathbf{Z} \hat{\mathbf{u}} \]

2. SNP effects

3. Indirect Genomic Prediction (IGP) = \(\mathbf{Z} \hat{\mathbf{u}}\)
GEBV = $b_0 + b_1 \times \text{IGP using 2014 – 2018 IGP group}$

<table>
<thead>
<tr>
<th></th>
<th>$\Delta G$</th>
<th>$b_0$</th>
<th>$b_1$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Mean</td>
<td>1.11</td>
<td>1.29</td>
<td>1.21</td>
<td>0.99</td>
</tr>
<tr>
<td>SD</td>
<td>0.68</td>
<td>0.75</td>
<td>0.72</td>
<td>0.03</td>
</tr>
<tr>
<td>Correlation ($\Delta G, *$)</td>
<td>-</td>
<td>0.98</td>
<td>0.98</td>
<td>0.84</td>
</tr>
</tbody>
</table>

$\Delta G = \text{Genetic progress / year}$
Randomly selected animals to obtain SNPs for 2014 – 2018 IGP

No. randomly selected genotyped animals (x 1000)

- b1 male
- b1 female
- R2 male
- R2 female
<table>
<thead>
<tr>
<th>YOB</th>
<th>Computing time (hrs: wall clock)</th>
<th>No. iterations for GEBV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GEBV</td>
<td>IGP (selected)</td>
</tr>
<tr>
<td>2014 - 18</td>
<td>72</td>
<td>27-5 (1)</td>
</tr>
<tr>
<td>2015 - 18</td>
<td>106</td>
<td>35-8 (1)</td>
</tr>
<tr>
<td>2016 - 18</td>
<td>118</td>
<td>45-10 (1)</td>
</tr>
<tr>
<td>2017 - 18</td>
<td>140</td>
<td>60-12 (1)</td>
</tr>
<tr>
<td>2018</td>
<td>153</td>
<td>70-13 (1)</td>
</tr>
<tr>
<td>All</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Conclusions
• Indirect genomic prediction feasible in large-scale genomic evaluation

• Reducing the computing cost by Indirect Genomic Evaluation

• Significant reduction in computing cost by using randomly selected genotyped animals for IGP

• Indirect Genomic Predictions as accurate as GEBV
Acknowledgements

Holstein Association USA Inc.
and
Council on Dairy Cattle Breeding

Anonymous reviewers in JDS Communications

Thank you
Graphical Summary
No. genotyped animals (x 1000) vs. Computing Cost (hours)

- **No. genotyped animals for IGP**
- **No. genotyped animals for GEBV**
- **Cost with all genotyped animals**
- **Cost with selected genotyped animals**
- **Without constructing G-1 twice**

- **Genotyped animal group**
  - 2014-2018
  - 2015-2018
  - 2016-2018
  - 2017-2018
  - 2018