Can artificial intelligence be used on historical cow data to improve data quality and standardization of records

Where Will New Technologies Take Milk Recording?
Miel Hostens - m.m.hostens@uu.nl
MY HISTORY

CURRENT
This is how you can describe me now

PhD Vet Met 2013

MS Data Sci 2016

MS Vet Med 2006

Ass. Prof at department of Farm Animal Health, 2019
Can artificial intelligence be used on historical cow data to improve data quality and standardization of records

Where Will New Technologies Take Milk Recording?
Miel Hostens - m.m.hostens@uu.nl
Need for milk in 50 years

- FAO - 2017: 912,224,857,490
- FAO estimate for 2057: 1,240,361,000,000
- SAT current estimate for 2057: 938,175,000,000

- Global milk consumption

100% 191% 147%
Efficiency through genetic and management
Efficiency through genetic and management
Efficiency through genetic and management

1
TRUCK
of MILK
Efficiency through genetic and management

1 TRUCK of MILK
2 TRUCKs of MANURE
Efficiency through genetic and management

1. TRUCK of MILK
2. TRUCKs of MANURE
3. TRUCKs of WATER
Efficiency through genetic and management

1. TRUCK of MILK
2. TRUCKs of MANURE
3. TRUCKs of WATER
4. Gb of data
Can artificial intelligence be used on historical cow data to improve data quality and standardization of records

Where Will New Technologies Take Milk Recording?
Miel Hostens - m.m.hostens@uu.nl
Precision livestock farming
Precision livestock farming

One day we will be able to predict disease
Precision livestock farming
Precision livestock farming

"We keep adding more sensors without making sense of the sensors we already have"
Genomic selection
Genomic selection

One day we will be able to breed for healthy animals
Hypothesis - partim 1
Hypothesis - partim 1

Highest occurrence of data quality issues in disease recording
Hypothesis - partim 2

- Days to first breeding
- Number of inseminations
- Days to last breeding
- Interval to first luteal activity
- Age at first calving
- First calving interval
- Mastitis
- Ketosis
- Retained placenta
- Metritis

Heritability

Donagh Berry, 2014
Hypothesis - partim 2

"Is heritability of disease really low?"

or an artifact of low data quality in disease recording

Donagh Berry, 2014
Hypothesis - partim 4

Any prediction, from SENSOR-DATA or MILK-BIOMARKERS or GENOMICS will fail if our reference data has low quality
Hypothesis - partim 4

Any prediction, from SENSOR-DATA or MILK-BIOMARKERS or GENOMICS will fail if our reference data has low quality

Novel approaches to assess the quality of fertility data stored in dairy herd management software

K. Hermans,*1 W. Waegeman,† G. Opsomer,* B. Van Ranst,* J. De Koster,* M. Van Eevelde,* and M. Hostens*
*Faculty of Veterinary Medicine, Department of Reproduction, Obstetrics and Herd Health, Ghent University, Salisburylaan 133, 9820 Merelbeke, Belgium
†Faculty of Bioscience Engineering, Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Coupure Links 653, 9000 Ghent, Belgium
Disease classification
Standards
Disease classification

Standards

> 100 disease ????????????????????????
Disease classification
Standards

> 100 disease ?????
Disease classification
Producer records
Disease classification
Producer records

The old way of creating herd-management-software

An impossible flow of 'Data Entry' input screens leading to low quality and under reporting
Disease classification
Producer records

The old way of creating herd-management-software

An impossible flow of 'Data Entry' input screens leading to low quality and under reporting
Ask the people with boots in the manure
Ask the people with boots in the manure

"10 events is already too much ..."
We urgently need to

LEARN

UNLEARN

RELEARN

Alvin Toffler
We urgently need to

**LEARN**

**UNLEARN**

**RELEARN**

**DEEPLEARN**

when talking about producer-recorded events in dairy cows

*Miel Hostens*
ADSA in 2019

Can artificial intelligence be used on historical cow data to improve data quality and standardization of records

Where Will New Technologies Take Milk Recording?  
Miel Hostens - m.m.hostens@uu.nl
What did we LEARN

Rutten et al. JDS
44.
How can we use the information we ALREADY have, to improve event/disease classification
Given animal information (display and印象) in the application.
Animal history

Birth  Insemination  Calving  ...

50.
Animal history

Birth  Insemination  Calving  ...

Basic model
Animal history

Birth  Insemination  Calving  ...

Picture model
Animal history

Birth  Insemination  Calving  ...

Picture model
Animal history

Birth       Insemination       Calving       ...

Augment the history with features from the pictures

Picture model
MmmooOgle data platform
MmmooOgle data platform

- Real-time access for data scientists to entire DATAHUB
- Anonymous data
- Short data-to-model DevOps
- Event driven data structure implemented
- Swagger documentation of the entire domain
- Models/data available via REST-API
- ...
MmmooOgle data platform
MmmooOgle data platform

- In total we collected
  - 222 756 recordings
  - 22 unique events
  - 49 671 different cows
  - 6 different herds

- For 26 903 of these events, one or more pictures were available

- Model building
  - Markov for discrimination models
  - Neural network (deep learning)
    - Long Short Term-Memory networks
    - Convolutional Neural Network
Data pipeline
Data pipeline
Sliding windows
Sliding windows length + class balancing

Figure 5: Average evaluation scores by sequence length
Data pipeline using deep learning
Data pipeline using deep learning
Tensorflow deep learning
Tensorflow deep learning
Can artificial intelligence be used on historical cow data to improve data quality and standardization of records

Where Will New Technologies Take Milk Recording?
Miel Hostens - m.m.hostens@uu.nl
Model performance

Evaluation
- Percentage Correctly Classified (PCC)
- weighted Percentage Correctly Classified (wPCC)
- Top-3 accuracy (Top-3 PCC)
## Model Performance

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2</th>
<th>Dataset</th>
<th>PCC</th>
<th>wPCC</th>
<th>Top-3 PCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFD-1</td>
<td>MFD-2</td>
<td>Unbalanced</td>
<td>0.67</td>
<td>0.39</td>
<td>0.87</td>
</tr>
<tr>
<td>LSTM</td>
<td>LSTM</td>
<td>Unbalanced</td>
<td>0.75</td>
<td>0.44</td>
<td>0.93</td>
</tr>
<tr>
<td>LSTM</td>
<td>Balanced</td>
<td>Balanced</td>
<td>0.71</td>
<td>0.49</td>
<td>0.92</td>
</tr>
<tr>
<td>CNN</td>
<td>CNN</td>
<td>Unbalanced</td>
<td>0.67</td>
<td>0.38</td>
<td>0.84</td>
</tr>
<tr>
<td>CNN</td>
<td>Balanced</td>
<td>Balanced</td>
<td>0.67</td>
<td>0.39</td>
<td>0.85</td>
</tr>
<tr>
<td>MFD-2</td>
<td>LSTM</td>
<td>Unbalanced</td>
<td>0.68</td>
<td>0.38</td>
<td>0.88</td>
</tr>
<tr>
<td>LSTM</td>
<td>LSTM</td>
<td>Balanced</td>
<td>0.75</td>
<td>0.43</td>
<td>0.92</td>
</tr>
<tr>
<td>LSTM</td>
<td>Balanced</td>
<td>Balanced</td>
<td>0.73</td>
<td>0.50</td>
<td>0.92</td>
</tr>
<tr>
<td>CNN</td>
<td>CNN</td>
<td>Unbalanced</td>
<td>0.68</td>
<td>0.36</td>
<td>0.85</td>
</tr>
<tr>
<td>CNN</td>
<td>Balanced</td>
<td>Balanced</td>
<td>0.68</td>
<td>0.40</td>
<td>0.87</td>
</tr>
<tr>
<td>LSTM</td>
<td>Unbalanced</td>
<td>Unbalanced</td>
<td>0.76</td>
<td>0.42</td>
<td>0.95</td>
</tr>
<tr>
<td>LSTM</td>
<td>Balanced</td>
<td>Balanced</td>
<td>0.73</td>
<td>0.51</td>
<td>0.94</td>
</tr>
<tr>
<td>CNN</td>
<td>Unbalanced</td>
<td>0.76</td>
<td>0.40</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>CNN</td>
<td>Balanced</td>
<td>Balanced</td>
<td>0.75</td>
<td>0.44</td>
<td>0.94</td>
</tr>
<tr>
<td>LSTM</td>
<td>CNN</td>
<td>Balanced</td>
<td>0.60</td>
<td>0.58</td>
<td>0.90</td>
</tr>
<tr>
<td>CNN</td>
<td>Unbalanced</td>
<td>0.69</td>
<td>0.53</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>CNN</td>
<td>Balanced</td>
<td>Balanced</td>
<td>0.60</td>
<td>0.59</td>
<td>0.90</td>
</tr>
<tr>
<td>CNN</td>
<td>Unbalanced</td>
<td>0.42</td>
<td>0.15</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>CNN</td>
<td>Balanced</td>
<td>Balanced</td>
<td>0.41</td>
<td>0.17</td>
<td>0.76</td>
</tr>
<tr>
<td>CNN</td>
<td>Balanced</td>
<td></td>
<td>0.38</td>
<td>0.21</td>
<td>0.49</td>
</tr>
</tbody>
</table>
## Model Performance

Ensemble of historical data and picture information is able to have **high Top-3 PCC**

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2</th>
<th>Dataset</th>
<th>PCC</th>
<th>wPCC</th>
<th>Top-3 PCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFD-1</td>
<td>MFD-2</td>
<td>Unbalanced</td>
<td>0.67</td>
<td>0.39</td>
<td>0.87</td>
</tr>
<tr>
<td>LSTM</td>
<td>Unbalanced</td>
<td>0.67</td>
<td>0.39</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>LSTM</td>
<td>Balanced</td>
<td>0.75</td>
<td>0.44</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>LSTM</td>
<td>Balanced</td>
<td>0.71</td>
<td>0.49</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>CNN</td>
<td>Unbalanced</td>
<td>0.67</td>
<td>0.38</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>CNN</td>
<td>Balanced</td>
<td>0.67</td>
<td>0.39</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>MFD-2</td>
<td>LSTM</td>
<td>Unbalanced</td>
<td>0.68</td>
<td>0.38</td>
<td>0.88</td>
</tr>
<tr>
<td>LSTM</td>
<td>Unbalanced</td>
<td>0.75</td>
<td>0.43</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>LSTM</td>
<td>Balanced</td>
<td>0.73</td>
<td>0.50</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>CNN</td>
<td>Unbalanced</td>
<td>0.68</td>
<td>0.36</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>CNN</td>
<td>Balanced</td>
<td>0.68</td>
<td>0.40</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>LSTM</td>
<td>LSTM</td>
<td>Unbalanced</td>
<td>0.76</td>
<td>0.42</td>
<td>0.95</td>
</tr>
<tr>
<td>LSTM</td>
<td>Balanced</td>
<td>0.76</td>
<td>0.51</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>CNN</td>
<td>Unbalanced</td>
<td>0.75</td>
<td>0.40</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>CNN</td>
<td>Balanced</td>
<td>0.75</td>
<td>0.44</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>LSTM</td>
<td>CNN</td>
<td>Unbalanced</td>
<td>0.60</td>
<td>0.58</td>
<td>0.90</td>
</tr>
<tr>
<td>CNN</td>
<td>Unbalanced</td>
<td>0.60</td>
<td>0.59</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>CNN</td>
<td>Balanced</td>
<td>0.60</td>
<td>0.59</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>CNN</td>
<td>Balanced</td>
<td>0.42</td>
<td>0.15</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>CNN</td>
<td>Balanced</td>
<td>0.41</td>
<td>0.17</td>
<td>0.76</td>
<td></td>
</tr>
</tbody>
</table>
Individual class performance
Individual class performance

- Mastitis
- Abortion
- Culling/Dead

High PCC
Can artificial intelligence be used on historical cow data to improve data quality and standardization of records

Where Will New Technologies Take Milk Recording?
Miel Hostens - m.m.hostens@uu.nl

Introduction
Hypothesis
M&M
Results
Discussion
Who am I
Why are these results relevant
Why are these results relevant
Why are these results relevant

1. Pneumonia?
2. Diarrhea?
3. ...
Why are these results relevant

Guide low-trained farmers/personnel to the right diagnosis

1. Pneumonia?
2. Diarrhea?
3. ...
MmmooOgle data platform
MmooOgle data platform
What other information is out there
17/06/2019 12:37
Melkvee

BE 614134595
6415 - Br Alida 6415

17/06/2019 12:37
brought from the back. sick.. treated with 30cc vecoxan
What other information is out there
What other information is out there

Using Natural Language Processing, we could label the animal as
- Sick
- Treated
- Dose 30cc
- Product Vexocan
Can artificial intelligence be used on historical cow data to improve data quality and standardization of records

Where Will New Technologies Take Milk Recording?
Miel Hostens - m.m.hostens@uu.nl