Novel Balanced Feature Representation for Wikipedia Vandalism Detection Task

István Hegedűs, Róbert Ormándi, Richárd Farkas, and Márk Jelasity
University of Szeged
Hungary
ihegedus@inf.u-szeged.hu
Our approach

- Supervised learning
- Rich feature set
- Meta-learning scheme
Vector space model (VSM)

- unigrams
- values:
  - $N$ if does not occur in the edit
  - $A$ if in added sequence
  - $D$ if in removed sequence
  - $C$ if in changed sequence
- $\#$features = 47,324
- best 100 by InfoGain
Balanced VSM

- sample is unbalanced
  - 93.9% regular

BVSM:

\[
\text{for } i \text{ in } 1 \text{ to } N \text{ do}
\]

\[
D = \text{vandalism AND random\_regular}
\]

\[
IG += \text{InfoGainScore}(D)
\]

\[
done
\]

\[
VSM = \text{best}(IG,100)
\]
Other features

• CharacterStatistic
  uppercase and lowercase ratio
• RepeatedCharSequences
  – asdasdasdasdasdasd
• ValidWordRatio
  – English/pejorative words
• CommentStatistic
• UserNameOrIP
  – nickname or country from IP
## 10-fold-cross-validation

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC (10-fold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balanced VSM</td>
<td>0.813</td>
</tr>
<tr>
<td>Balanced VSM + stopword</td>
<td>0.843</td>
</tr>
<tr>
<td>Other features</td>
<td>0.883</td>
</tr>
<tr>
<td>Other + unbalanced VSM</td>
<td>0.884</td>
</tr>
<tr>
<td>Other + balanced VSM</td>
<td>0.887</td>
</tr>
</tbody>
</table>
Meta learning

J48=0.3; NaiveBayes=0.09; Logistic=0.61
## Results (eval)

<table>
<thead>
<tr>
<th></th>
<th>AUC (LogReg)</th>
<th>AUC (Voting)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balanced VSM</td>
<td>0.744</td>
<td>0.761</td>
</tr>
<tr>
<td>Other features</td>
<td>0.865</td>
<td>0.876</td>
</tr>
<tr>
<td>Other + balanced</td>
<td>0.854</td>
<td>0.877</td>
</tr>
<tr>
<td>Other + unbalanced</td>
<td>0.864</td>
<td>0.880</td>
</tr>
</tbody>
</table>
Summary

• VSM has no significant added value
• meta-learning (+2%)