Predicting Conference Paper Acceptance
William Jen, Shichang Zhang, and Muyun Chen
{wjen, shichang, muyunc}@stanford.edu

Background and Motivation
The explosion of scientific research in machine learning has led to the rapid growth of paper submissions to top conferences. Can we predict whether a machine learning paper will be accepted using machine learning methods?

Kang, et. al. published initial work on this topic in April 2018 with the public release of PeerRead, a dataset that collects research papers from AI/ML conferences.

Models
We reproduced the models from Kang’s paper together with some other models.
- Logistic regression with L2/L1 regularization.
- Random Forest
- SVM with RBF kernel
  \[
  \min_{w,b} \frac{1}{2}||w||^2 \quad s.t. \quad y^{(i)}(w^T x^{(i)} + b) \geq 1, \quad i = 1, \ldots, m \\
  K(x, \tilde{x}) = \exp \left( \frac{||x - \tilde{x}||^2}{\sigma^2} \right)
  \]
- AdaBoost
  - We used 50 weak classifiers.
- Neural Network
  - We used ReLU activation function and 20 different network structures.

Datasets and Features
We took all the 427 papers submitted to ICLR 2017 as our dataset. There are 172 accepted and 255 rejected papers. For each paper, we extracted 18 features of numerical and Boolean values.

Important Features
- Whether abstract contains deep, neural, embedding, outperform, outperform, novel, or state_of_the_art
- Number of figures, tables, sections, equations, theorems
- Number of references
- Bag-of-words in abstract
- Average of GloVe word embeddings in abstract

Result
The first row shows the accuracy of the baseline model of predicting by majority, which in this case is reject all the papers. Our best models outperform Kang’s best model, which has 65.3% test accuracy on average.

<table>
<thead>
<tr>
<th>Model</th>
<th>Train Accuracy (%)</th>
<th>Test Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority</td>
<td>60.17</td>
<td>60.53</td>
</tr>
<tr>
<td>Logistic L2</td>
<td>42.41</td>
<td>42.10</td>
</tr>
<tr>
<td>Logistic L1</td>
<td>68.48</td>
<td>68.42</td>
</tr>
<tr>
<td>SVM RBF</td>
<td>72.49</td>
<td>71.05</td>
</tr>
<tr>
<td>Random Forest</td>
<td>99.43</td>
<td>63.16</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>96.56</td>
<td>50.00</td>
</tr>
<tr>
<td>Neural Network</td>
<td>63.04</td>
<td>60.53</td>
</tr>
</tbody>
</table>

Future Work
Our work focused on the ICLR dataset, which has limited examples. Similar studies can be done on other conferences with more submissions, like NIPS, or for the same conference but with submissions across years. The challenging part is parsing and featurizing the large number of papers, which is computationally expensive.

Reference