Problem & Summary of Work

Patents certify and protect new ideas. They are an essential part of modern innovation and are a primary driver of economic growth.

However, current patent processes are inefficient and imprecise:
- 650k annual applications to USPTO, beginning to overwhelm the office
- Only 10% of patent inventors are women; micro-entities and women are meaningfully less likely to have their patents approved

Better acceptance prediction and model understanding can reduce strain on USPTO, save governments money, and reduce bias in granting applications

Our work improves acceptance accuracy and model understanding:

- Accuracy Improvement in Patent Acceptance Prediction
- For the first time, analysis saliency of Prediction Model
- Confirms that BERT models cannot yet outperform Naive Bayes

Technical Methods

Naive Bayes (top performing baseline)
- Assumes independence of tokens when classifying

\[ p(Accepted | x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i | Accepted) \]

DistILBERT (second-best performing baseline) by Sarh et al. @ HuggingFace
- Lighter, Cheaper, Faster: 40% decrease in model size and 60% faster pretraining, retains 97% of performance of BERT model
- Uses knowledge distillation to minimize loss with “teacher” BERT model

Prior work acknowledges limits of only training on one section

Our Custom Ensemble Architectures
- Ensemble is better than training one large model:
  - Abstract and Claims average 143 tokens combined (too big for BERT)
  - Each section is semantically different, preserve nuance
  - Can achieve full advantage of different data representations

Experiments & Results

Dataset
- Using largest, richest patent dataset: Harvard USPTO Patent Dataset
- Trained on 2011-2013 subset for efficiency — 664509 patent applications
- Validation set is balanced between rejected/accepted, i.e., true baseline of 50%

Model Evaluation: Overall Decision Accuracy

<table>
<thead>
<tr>
<th>Bernoulli Naive Bayes</th>
<th>DISTILBERT</th>
<th>Ensemble #1</th>
<th>Ensemble #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract — 0.63</td>
<td>Abstract — 0.63</td>
<td>61.76</td>
<td>62.91</td>
</tr>
<tr>
<td>Claims — 0.67</td>
<td>Claims — 0.59</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Analyzed saliency to understand how BERT models think
- Use integrated gradient saliency analysis for DISTILBERT baseline
- We look at words that have strong impact on the overall decision prediction
- Technical words like circuit, semiconductor, device and adjectives that stress novelty like first have strong positive impact in the classification decision
- Action words like introduced, and, method, controls, commented are penalized as they are generic and do not stress the concept's novelty

Baseline for acceptance accuracy set by Naive Bayes (not BERT models):
- Prior work trained on only one section at a time (only Abstract; only Claims)
- State-of-the-art baseline set with Bernoulli Naive Bayes

We train and test on #1 International Patent Classification (IPC) subclass, G06F: Electric Digital Data Processing, which constitutes 10.4% of applications

References & Acknowledgements

3. Oscar O’Rahilly, Alex Lerner, and Oscar O’Rahilly. (2019). Thank you to all great mentors, Michelle Singer and Katherine Goldman for your invaluable guidance on our project.