### Introduction
- Text encoded by paths in RNN output
- Characters can be repeated: "ab" → "aab"
- And can be followed by blanks: "aab" → "aaa-b"
- Word Beam Search (WBS) decodes RNN output

![CTC-trained neural network](image)

### Proposed Method
- **Beam state**
  - Inside word: constrain beam by prefix tree
  - Between words: allow arbitrary many non-word characters

![Beam states](image)

- **Word-level LM with 4 possible scoring-modes**
  - Only constrain words
  - N-gram score whenever beam finishes a word
  - N-gram lookahead: possible words given a prefix
  - N-gram sampled lookahead: subset of words

![Extending a beam in word-state](image)

### Related Algorithms
- Best path decoding
  - Collapse best-scoring path
- Token passing
  - Sequence of dictionary words, word-level LM
- Vanilla beam search (VBS)
  - Tree of beams, character-level LM

![Tree of beams](image)

### Evaluation on IAM and Bentham HTR datasets
- Two different texts to train LM
  - Perfect: text of test-set (OOV-rate 0%)
  - Rudimentary: text of training-set + 370k word list

<table>
<thead>
<tr>
<th></th>
<th>Perfect LM</th>
<th>Rudimentary LM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best Path</td>
<td>8.77 / 29.07</td>
<td>8.77 / 29.07</td>
</tr>
<tr>
<td>Token Passing</td>
<td>8.57 / 12.23</td>
<td></td>
</tr>
<tr>
<td>VBS W</td>
<td>8.27 / 27.34</td>
<td>8.48 / 28.24</td>
</tr>
<tr>
<td>VBS N</td>
<td>5.62 / 13.04</td>
<td>8.95 / 24.19</td>
</tr>
<tr>
<td>WBS N-F</td>
<td>5.33 / 13.90</td>
<td>10.00 / 27.65</td>
</tr>
<tr>
<td>WBS N+F</td>
<td>5.21 / 9.82</td>
<td>8.61 / 22.86</td>
</tr>
<tr>
<td>WBS N+F+S</td>
<td>5.21 / 9.78</td>
<td>8.62 / 22.91</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Perfect LM</th>
<th>Rudimentary LM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best Path</td>
<td>8.58 / 17.06</td>
<td>8.58 / 17.06</td>
</tr>
<tr>
<td>Token Passing</td>
<td>8.19 / 12.24</td>
<td></td>
</tr>
<tr>
<td>VBS W</td>
<td>5.35 / 16.02</td>
<td>5.55 / 16.39</td>
</tr>
<tr>
<td>VBS N</td>
<td>4.22 / 13.02</td>
<td>5.47 / 14.03</td>
</tr>
<tr>
<td>WBS N-F</td>
<td>4.07 / 13.90</td>
<td>6.15 / 23.80</td>
</tr>
<tr>
<td>WBS N+F</td>
<td>4.06 / 7.36</td>
<td>6.76 / 18.06</td>
</tr>
<tr>
<td>WBS N+F+S</td>
<td>4.06 / 7.39</td>
<td>6.76 / 18.06</td>
</tr>
</tbody>
</table>

### Results
- Dependence of running time on dictionary size
  - Token passing: only feasible for 2k words
  - Small dependence: VBS and WBS modes W and N
  - Large dependence: WBS modes N+F and N+F+S

![Running time [ms] for Bentham with different dictionary sizes](image)

### Conclusion
- Decodes output of CTC-trained neural network
- Words constrained by dictionary
- Allows arbitrary number of non-word characters between words
- Optional word-level LM
- Faster than token passing

**Code:**
github.com/githubharald/CTCWordBeamSearch

**Contact:**
harald_scheidl@hotmail.com