RNNs are usually trained using Some known problems when training RNNs:

- test loss
- 0/1 flavour
- exposure bias

Contributions:

Can we build on existing approaches?

Learning to Search

Structured prediction

Learning To Search (L2S)

How does it work?

one by one

Links between Learning to Search and RNNs

SEARNN [3]

Overview:

Integrate roll-outs

Leverage

Algorithm

1. Compute costs with roll-in/outs
2. Derive a loss from the costs
3. Use the loss to take a gradient step
4. Rinse and repeat

Why is it better than MLE?

structured information

global

local

SEARNN: Training RNNs with Global-Local Losses

Rémi Leblond*, Jean-Baptiste Alayrac*

* equal contribution

Summary

Some known problems when training RNNs:

- test loss
- 0/1 flavour
- exposure bias

Contributions:

Can we build on existing approaches?

Experiments

Full algorithm:

<table>
<thead>
<tr>
<th>Metric</th>
<th>MLE (ref)</th>
<th>MLE (mixed)</th>
<th>Learned</th>
<th>BSO</th>
<th>LECCAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCR</td>
<td>2.4</td>
<td>2.9</td>
<td>1.9</td>
<td>2.1</td>
<td>1.8</td>
</tr>
<tr>
<td>CoNLL</td>
<td>1.9</td>
<td>2.2</td>
<td>1.6</td>
<td>2.7</td>
<td>1.8</td>
</tr>
<tr>
<td>norm</td>
<td>2.2</td>
<td>2.5</td>
<td>3.7</td>
<td>1.6</td>
<td>1.6</td>
</tr>
<tr>
<td>Running</td>
<td>4.6</td>
<td>4.2</td>
<td>1.2</td>
<td>2.1</td>
<td>1.8</td>
</tr>
<tr>
<td>Spelling</td>
<td>0.4</td>
<td>0.4</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>edit</td>
<td>12.0</td>
<td>18.0</td>
<td>17.5</td>
<td>19.2</td>
<td>17.9</td>
</tr>
<tr>
<td>rouge</td>
<td>25.3</td>
<td>24.6</td>
<td>27.1</td>
<td>28.0</td>
<td>27.5</td>
</tr>
<tr>
<td>bleu</td>
<td>22.5</td>
<td>22.7</td>
<td>27.4</td>
<td>28.0</td>
<td>27.5</td>
</tr>
<tr>
<td>Rename</td>
<td>0.5</td>
<td>0.7</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Scaling approach:

- all 4 tasks
- harder
- bigger
- learned/mixed
- structurally
- close
- warm start
- fraction of the cost

Machine Translation (in progress): ISWL 14 Ger/Eng

References

NIPS

Machine Learning

ArXiv preprint: 1706.04499

ICLR

EMNLP

JMLR

ICML

Check out our project webpage for code/data!