SoPa: Soft Patterns

Neural Transitions in WFSAs

Extension of a 1-Layer CNN

Takeaways

Text Cat. Experiments

Interpretation

Multiple Patterns in Parallel — an RNN!

A hidden state represents the score of the best path through the pattern.
Pattern match computed efficiently using dynamic programming.
Transition weights explicitly capture a soft notion of words and wildcards.

- What a great movie!
- What a great book!
- What great shoes!
- What a great, funny, magical show!

Pattern Matching using Weighted Finite State Automata (WFSAs)

Multiple Patterns in Parallel — an RNN!

What a great movie!
What a great book!
What great shoes!
What a great, funny, magical show!

Fielding’s funniest and most likeable book in years

<table>
<thead>
<tr>
<th>Pattern1 states</th>
<th>Pattern2 states</th>
<th>word vectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>00</td>
<td>book</td>
</tr>
<tr>
<td>00</td>
<td>09</td>
<td>in</td>
</tr>
<tr>
<td>00</td>
<td>08</td>
<td>years</td>
</tr>
<tr>
<td>00</td>
<td>06</td>
<td>00</td>
</tr>
<tr>
<td>00</td>
<td>04</td>
<td>00</td>
</tr>
<tr>
<td>00</td>
<td>02</td>
<td>00</td>
</tr>
<tr>
<td>00</td>
<td>01</td>
<td>00</td>
</tr>
<tr>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>00</td>
<td>09</td>
<td>book</td>
</tr>
<tr>
<td>00</td>
<td>08</td>
<td>in</td>
</tr>
<tr>
<td>00</td>
<td>06</td>
<td>years</td>
</tr>
<tr>
<td>00</td>
<td>04</td>
<td>00</td>
</tr>
<tr>
<td>00</td>
<td>02</td>
<td>00</td>
</tr>
<tr>
<td>00</td>
<td>01</td>
<td>00</td>
</tr>
<tr>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>00</td>
<td>09</td>
<td>book</td>
</tr>
<tr>
<td>00</td>
<td>08</td>
<td>in</td>
</tr>
<tr>
<td>00</td>
<td>06</td>
<td>years</td>
</tr>
<tr>
<td>00</td>
<td>04</td>
<td>00</td>
</tr>
<tr>
<td>00</td>
<td>02</td>
<td>00</td>
</tr>
<tr>
<td>00</td>
<td>01</td>
<td>00</td>
</tr>
<tr>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
</tbody>
</table>

- A new RNN combining neural learning with WFSAs.
- An extension of a one-layer CNN.
- Models flexible-length spans with insertion and deletion.
- Can be easily customized by swapping in different semirings.
- Paving the bridge between CNNs and RNNs.
- https://github.com/Noahs-ARK/soft_patterns

- With max-plus semiring, without self-loops, ε-transitions and the sigmoid in the main path transition, SoPa is equivalent to a 1-layer CNN with max-pooling.

$S_{\text{span}}(x_{i:i+d}) = \sum_{j=0}^{d-1} w_j \cdot v_{x_{i+j}} + b_j = w_0 \cdot v_{x_{i:i+d}} + \sum_{j=0}^{d-1} b_j$

$S_{\text{doc}}(x) = \max_{1 \leq i \leq n-d+1} S_{\text{span}}(x_{i:i+d})$

- A 1-layer CNN is also learning a restricted form of a WFSA.

- A hidden state represents the score of the best path through the pattern.
- Pattern match computed efficiently using dynamic programming.
- Transition weights explicitly capture a soft notion of words and wildcards.

- What a great movie!
- What a great book!
- What great shoes!
- What a great, funny, magical show!

Takeaways

- A new RNN combining neural learning with WFSAs.
- An extension of a one-layer CNN.
- Models flexible-length spans with insertion and deletion.
- Can be easily customized by swapping in different semirings.
- Paving the bridge between CNNs and RNNs.
- https://github.com/Noahs-ARK/soft_patterns

- With max-plus semiring, without self-loops, ε-transitions and the sigmoid in the main path transition, SoPa is equivalent to a 1-layer CNN with max-pooling.

$S_{\text{span}}(x_{i:i+d}) = \sum_{j=0}^{d-1} w_j \cdot v_{x_{i+j}} + b_j = w_0 \cdot v_{x_{i:i+d}} + \sum_{j=0}^{d-1} b_j$

$S_{\text{doc}}(x) = \max_{1 \leq i \leq n-d+1} S_{\text{span}}(x_{i:i+d})$

- A 1-layer CNN is also learning a restricted form of a WFSA.

- What a great movie!
- What a great book!
- What great shoes!
- What a great, funny, magical show!

Takeaways

- A new RNN combining neural learning with WFSAs.
- An extension of a one-layer CNN.
- Models flexible-length spans with insertion and deletion.
- Can be easily customized by swapping in different semirings.
- Paving the bridge between CNNs and RNNs.
- https://github.com/Noahs-ARK/soft_patterns

- With max-plus semiring, without self-loops, ε-transitions and the sigmoid in the main path transition, SoPa is equivalent to a 1-layer CNN with max-pooling.

$S_{\text{span}}(x_{i:i+d}) = \sum_{j=0}^{d-1} w_j \cdot v_{x_{i+j}} + b_j = w_0 \cdot v_{x_{i:i+d}} + \sum_{j=0}^{d-1} b_j$

$S_{\text{doc}}(x) = \max_{1 \leq i \leq n-d+1} S_{\text{span}}(x_{i:i+d})$

- A 1-layer CNN is also learning a restricted form of a WFSA.

Text Cat. Experiments

Sentiment Analysis

Interpretation

- SoPa is interpretable both at the single pattern level and the document level.

- Highest Scoring Phrases
  - 's ε unsupervised story
  - this ε bad on purpose
  - this ε leaen comedy
  - a ε half-assed story
  - is ε clumsy .st. the writing
  - honest , scathing , and enjoyable
  - soulful , scathing , and joyous
  - unpretentious , charming , quirky
  - forceful , and beautifully energetic
  - surprising

- Analyzed Documents
  - it’s dumb , but more importantly , it’s just not scary
  - though moonlight mile is replete with acclaimed actors and actresses and tackles a subject that ‘s potentially moving , the movie is too predictable and too self-conscious to reach a level of high drama
  - While its careful pace and seemingly opaque story may not satisfy every moviegoer’s appetite , the film’s final scene is so moving , transparently moving.