Lightweight Contextual Logical Structure Recovery

Po-Wei Huang, Abhinav Ramesh Kashyap, Yanxia Qin, Yajing Yang, Min-Yen Kan

National University of Singapore

SDP@COLING 2022
Problem Statement

Classify each individual line into 23 predefined classes that indicate the hierarchy of the document structure.
Can we obtain (near-)SOTA performance on logical structure recovery without relying on feature-rich information, but on context only?
Dataset

**Data Source**

- **Original SectLabel Dataset** *(Luong et al., IJDLS 2010)*:
  - 20 ACL 2009 Papers
  - 20 CHI 2008 Papers
- **Extended Testing Dataset**:
  - 20 ACL 2020 Papers
- **Unlabelled Dataset**:
  - 570 ACL 2021 Long Papers
  - 1895 NeurIPS 2021 Papers

8:1:1 Document Split on SectLabel Dataset for Training, Validation and Testing
Contextual Model Construction

*Sliding Window Attention*

<table>
<thead>
<tr>
<th>Baseline</th>
<th>Sliding Window 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>author</em></td>
<td>reference</td>
</tr>
<tr>
<td><em>reference</em></td>
<td>reference</td>
</tr>
<tr>
<td><em>bodyText</em></td>
<td>reference</td>
</tr>
<tr>
<td><em>reference</em></td>
<td>reference</td>
</tr>
</tbody>
</table>

Pooling Methods for Sentence Embeddings

[CLS] Token
Pooling Methods for Sentence Embeddings

Mean Pooling

BERT (or any other pretrained transformer model)

\[
\text{Mean Pooling} = \frac{1}{n} \sum_{i=1}^{n} \text{emb}_i
\]
Pooling Methods for Sentence Embeddings

Attention Pooling

Attention Pooling

Self Attention Layer

BERT (or any other pretrained transformer model)
### Semi-Supervised Learning

**Data Augmentation Techniques**

<table>
<thead>
<tr>
<th>Original</th>
<th>Once upon a midnight dreary, while I pondered, weak and weary,</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synonym Replacement (EDA)</td>
<td><strong>Erstwhile</strong> upon a midnight dreary, while I pondered, weak and weary,</td>
</tr>
<tr>
<td>Random Insertion (EDA)</td>
<td>Once upon a midnight dreary, while I pondered, weak and <strong>once</strong> weary,</td>
</tr>
<tr>
<td>Random Swap (EDA)</td>
<td>Once upon <strong>I</strong> midnight dreary, while <strong>a</strong> pondered, weak and weary,</td>
</tr>
<tr>
<td>Random Delete (EDA)</td>
<td>Once upon a _ dreary, while I pondered, _ and weary,</td>
</tr>
<tr>
<td>Back Translation</td>
<td>Once at midnight it was bleak while I was thinking, weak and tired,</td>
</tr>
</tbody>
</table>
Semi-Supervised Learning

Unsupervised Data Augmentation

Diagram showing labeled text, unlabeled text, augmented text, and the process of using these with a backbone model and TSA to achieve cross-entropy and consistency loss.
Semi-Supervised Learning

FixMatch
Loss Engineering

Training Signal Annealing

Sentence Embedding → Prediction → Training Signal Annealing

Select
Max Prediction
Prob < f(t)

Cross Entropy Loss

f(t)
Loss Engineering
Supervised Data Augmentation
## Results

<table>
<thead>
<tr>
<th>Model</th>
<th>SectLabel</th>
<th></th>
<th></th>
<th>Extended</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Macro F1</td>
<td>Micro F1</td>
<td>Macro F1</td>
<td>Micro F1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>SciWING (Ramesh Kashyap and Kan, 2020)</em></td>
<td>0.732</td>
<td>0.900</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RoBERTa-Attn Model (OURS)</td>
<td>0.806</td>
<td>0.904</td>
<td>0.596</td>
<td>0.870</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RoBERTa-Attn Model + UDA_{log} †</td>
<td>0.784</td>
<td>0.906</td>
<td><strong>0.669</strong></td>
<td><strong>0.887</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RoBERTa-Attn Model + SDA_{log} †</td>
<td><strong>0.832</strong></td>
<td><strong>0.929</strong></td>
<td>0.623</td>
<td>0.886</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>SectLabel (Luong et al., 2010)</em> ‡</td>
<td><strong>0.847</strong></td>
<td><strong>0.934</strong></td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Scan to read the full paper!

Connect with the first author!
✉️ huangpowei@comp.nus.edu.sg