Joint Entity Linking with BERT

Amund Faller Råheim

Master of Science in Computer Science
Albert-Ludwigs-Universität Freiburg

1st of June 2022
Joint Entity Linking with BERT
Amund Faller Råheim

Outline

1 Problem
2 Solution
3 Evaluation
**Problem**

Entity Linking: Connecting an Input Document to a Knowledge Base

**Solution**

Towards a Joint Entity Linking Pipeline

BERT

Joint Entity Linking with BERT

**Evaluation**

AIDA-CoNLL Dataset

Evaluation Criteria

Evaluation Details
Problem

Entity Linking: Connecting an Input Document to a Knowledge Base

Example

- **Paris Hilton** is visiting **Paris** this weekend.
Problem

Entity Linking: Connecting an Input Document to a Knowledge Base

Example

- Paris Hilton is visiting Paris this weekend.
Problem

Entity Linking: Connecting an Input Document to a Knowledge Base

Example

- **Paris Hilton** is visiting **Paris** this weekend.

Entities in the **Knowledge Base**
Joint Entity Linking with BERT

Problem
The Classic Pipeline
Artificial Neural Networks
Problem Summary

Solution
Towards a Joint Entity Linking Pipeline
BERT
Joint Entity Linking with BERT

Evaluation
AIDA-CoNLL Dataset
Evaluation Criteria
Evaluation Details

Entity Linking = Mention Detection + Entity Disambiguation
The Classic Pipeline

1. Mention Detection ← Document
2. Candidate Generation ← Mention Texts
3. Entity Disambiguation ← Candidates and Context
The Classic Pipeline

1. Mention Detection ← Document
2. Candidate Generation ← Mention Texts
3. Entity Disambiguation ← Candidates and Context

... three disjoint methods! 😞
Artificial Neural Networks

SOTA for Entity Linking: Deep Learning!
Artificial Neural Networks

SOTA for Entity Linking: Deep Learning!

1. ANN for Mention Detection
2. Mapping for Candidate Generation
3. ANN for Entity Disambiguation
Artificial Neural Networks

SOTA for Entity Linking: Deep Learning!

1. ANN for Mention Detection
2. Mapping for Candidate Generation
3. ANN for Entity Disambiguation

No end-to-end deep learning with a mapping in the middle!
Problem Summary

Questions

ANN for Mention Detection, Mapping for Candidate Generation, ANN for Entity Disambiguation

No benefits of end-to-end learning!
Problem Summary

Questions

Summary

ANN for Mention Detection,
Mapping for Candidate Generation,
ANN for Entity Disambiguation

No benefits of end-to-end learning!
Towards a Joint Entity Linking Pipeline

Joint Mention Detection and Entity Disambiguation
Towards a Joint Entity Linking Pipeline

Joint Mention Detection and Entity Disambiguation

Q: How to solve both tasks with one model?
Towards a Joint Entity Linking Pipeline

Joint Mention Detection and Entity Disambiguation

Q: How to solve both tasks with one model?

A: Mention Embeddings and Entity Embeddings from shared Word Embeddings
A Transformer Network
Gives a contextualized embedding of words
A Transformer Network

Gives a contextualized embedding of words using "self-attention":

![Diagram of self-attention connections between words like "Paris", "is", "the", "capital", "and", "most", "populous", "city", "of", "France".]
A Transformer Network
Gives a contextualized embedding of words using "self-attention":

Pre-trained for general language tasks,
Fine-tuned for specific tasks
Joint Entity Linking with BERT

Amund Faller Råheim

Problem
The Classic Pipeline
Artificial Neural Networks
Problem Summary

Solution
Towards a Joint Entity Linking Pipeline
BERT
Joint Entity Linking with BERT

Evaluation
AIDA-CoNLL Dataset
Evaluation Criteria
Evaluation Details

Mention Detection
Entity Disambiguation

BERT

Tokenizer

'paris', 'hilton', 'is', 'visiting', 'paris', 'this', 'weekend', '.

Paris Hilton is visiting Paris this weekend.
Joint Entity Linking with BERT

Amund Faller Råheim

Problem
The Classic Pipeline
Artificial Neural Networks

Solution
Towards a Joint Entity Linking Pipeline
BERT
Joint Entity Linking with BERT

Evaluation
AIDA-CoNLL Dataset
Evaluation Criteria
Evaluation Details

Knowledge Base
Candidate Generation
Wikipedia2vec

Mention Detection
Entity Embeddings
EL-BERT

'paris', 'hilton', 'is', 'visiting', 'paris', 'this', 'weekend', '.'
Joint Entity Linking with BERT

Amund Faller Råheim

Problem
The Classic Pipeline
Artificial Neural Networks
Problem Summary

Solution
Towards a Joint Entity Linking Pipeline
BERT
Joint Entity Linking with BERT

Evaluation
AIDA-CoNLL Dataset
Evaluation Criteria
Evaluation Details

Joint Entity Linking with BERT

Knowledge Base
Candidate Generation
Wikipedia2vec

B, I, O, O, B, O, O, O

Mention Detection
Entity Embeddings

EL-BERT

'paris’, 'hilton’, 'is’, 'visiting’, 'paris’, 'this’, 'weekend’, ’'

1. Mention Detection & Entity Disambiguation
2. (Candidate Generation)
Solution Summary

Questions?

Joint representation from BERT, with a Mention Detection prediction and an Entity Embedding prediction. Entity Embeddings compared with Wikipedia2vec target entity embedding.
Questions?

Solution Summary

Joint representation from BERT, with a Mention Detection prediction and an Entity Embedding prediction.
Solution Summary

Questions?

Solution Summary

Joint representation from BERT, with a Mention Detection prediction and an Entity Embedding prediction.

Entity Embeddings compared with Wikipedia2vec target entity embedding.
AIDA-CoNLL Dataset

Demo
AIDA-CoNLL Dataset

Demo

HAVEL PRAISES CZECH NATIVE ALBRIGHT AS FRIEND. Klara Gajduskova PRAGUE 1996-12-06 Czech President Vaclav Havel on Friday welcomed the appointment of Madeleine Albright, who is of Czech extraction, as the United States' first woman Secretary of State. In a statement Havel, who is recovering from cancer surgery, said: "Madeleine Albright is a distinguished friend, a tested diplomat, and a true American of fine origins. " I look forward to continuing our good relations ... with the United States and with the first woman ever to hold the position of Secretary of State. I wish her well," Havel said in a statement to Reuters. Havel, who helped lead the "velvet revolution" that ousted the Communist regime in Prague in 1989, invited Albright, then working for a private foreign policy think tank, to advise his new democratic government in 1990. Havel had a small malignant tumour removed from his lung on Monday and is recovering in hospital. Albright, born Marie Korblova to a Czechoslovak diplomat in 1937, fled with her family to the United States after the Communists came to power in a coup in 1948.

As an academic, Albright studied and lectured on Europe's 20th century problems before becoming U.S. ambassador to the United Nations. Czech diplomats, seeking to have their country included in the expected expansion of NATO, praised the selection of Albright, known to be a strong supporter of alliance's integration of former Soviet-bloc countries. "The nomination ... is a clear signal that one key of the lines of foreign policy will be the strengthening of the trans-Atlantic cooperation, a creation of strategic partnership between Europe and the US," Foreign Minister Josef Zieleniec told Reuters. "(Albright) is a convinced advocate of NATO enlargement and of stabilisation of security structures." Czech ambassador to the United Nations, Karel Kovanda, told the daily Mlada Fronta Dnes that Albright "is a little light in our diplomatic heaven, but warned against expecting her to exert any influence in favour of the Czechs.
Evaluation Criteria

- **Precision:**
  \[
  \text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}
  \]

- **Recall:**
  \[
  \text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}
  \]

- **F1 score:**
  \[
  \text{F1 score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}
  \]
Evaluation Criteria

- **Precision:**

\[
\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}
\]

- **Recall:**

\[
\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}
\]

- **F1 score:**

\[
\text{F1 score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}
\]

**In-KB F1 Score:** Ignores unknown entities
## Evaluation Results

Results **without** Candidate Generation:

<table>
<thead>
<tr>
<th></th>
<th>AIDA-CoNLL Test (Micro F1 Score)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mention Detection</td>
</tr>
<tr>
<td><strong>Our Model</strong></td>
<td>95.1</td>
</tr>
<tr>
<td>Chen et al. (2019)</td>
<td>69.4</td>
</tr>
<tr>
<td>Broscheit (2019)</td>
<td></td>
</tr>
</tbody>
</table>
## Evaluation Results

### Results with Candidate Generation:

<table>
<thead>
<tr>
<th></th>
<th>AIDA-CoNLL Test (Micro F1 Score)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mention Detection</td>
</tr>
<tr>
<td>Broscheit (2019)</td>
<td></td>
</tr>
<tr>
<td>Martins et al. (2019)</td>
<td>92.5</td>
</tr>
<tr>
<td>Kolitsas et al. (2018)</td>
<td></td>
</tr>
<tr>
<td><strong>Our Model</strong></td>
<td>95.1</td>
</tr>
<tr>
<td>Poerner et al. (2020)</td>
<td></td>
</tr>
<tr>
<td>Chen et al. (2019)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Performance on entities *seen* in the training set vs. *unseen* in the training set:

<table>
<thead>
<tr>
<th></th>
<th>All Mentions</th>
<th>Seen</th>
<th>Unseen</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Our Model - No CG</strong></td>
<td>59.4</td>
<td>93.1</td>
<td>7.6</td>
</tr>
<tr>
<td><strong>Our Model - CG</strong></td>
<td>88.1</td>
<td>97.2</td>
<td>74.6</td>
</tr>
</tbody>
</table>
Evaluation Summary

• Promising results on seen entities without Candidate Generation,
Evaluation Summary

- Promising results on *seen* entities without Candidate Generation,
- Large performance boost with Candidate Generation.
Evaluation Summary

• Promising results on *seen* entities without Candidate Generation,

• Large performance boost with Candidate Generation.
Evaluation Summary

- Promising results on *seen* entities without Candidate Generation,
- Large performance boost with Candidate Generation.

Questions?
Joint Entity Linking with BERT

Amund Faller Råheim

Problem

The Classic Pipeline
Artificial Neural Networks
Problem Summary

Solution

Towards a Joint Entity Linking Pipeline
BERT
Joint Entity Linking with BERT

Evaluation

AIDA-CoNLL Dataset
Evaluation Criteria
Evaluation Details
AIDA-CoNLL Stats (Table 2)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Docs</th>
<th>Ment</th>
<th>Ment Annot</th>
<th>Unique Ent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>946</td>
<td>23,396</td>
<td>18,541</td>
<td>4,084</td>
</tr>
<tr>
<td>Validation</td>
<td>216</td>
<td>5,917</td>
<td>4,791</td>
<td>1,644</td>
</tr>
<tr>
<td>Test</td>
<td>231</td>
<td>5,616</td>
<td>4,485</td>
<td>1,536</td>
</tr>
<tr>
<td>Total</td>
<td>1,393</td>
<td>34,929</td>
<td>27,817</td>
<td>5,593</td>
</tr>
</tbody>
</table>

**Table:** Number of documents, mentions, mentions annotated with Wikipedia entities and unique mentioned entities in the AIDA-CoNLL datasets.
## Model Comparisons (Table 12)

<table>
<thead>
<tr>
<th></th>
<th>Our Model</th>
<th>Chen et al. (2019)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BERT</strong></td>
<td>Uncased</td>
<td>Cased</td>
</tr>
<tr>
<td><strong>Hidden Layers</strong></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Training Epochs</strong></td>
<td>180</td>
<td>190</td>
</tr>
<tr>
<td><strong>Training Time</strong></td>
<td>4 hrs 50 mins</td>
<td>5 hrs 6 mins</td>
</tr>
<tr>
<td><strong>Loss Function λ</strong></td>
<td>0.01</td>
<td>0.1</td>
</tr>
<tr>
<td><strong>Dropout</strong></td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Table:** Characteristics of our best-performing model and our implementation with the settings of Chen et al. (2019).