Self-Adaptive Named Entity Recognition by Retrieving Unstructured Knowledge

○Kosuke Nishida, Naoki Yoshinaga, Kyosuke Nishida (NTT and IIS, Univ. Tokyo)

Summary

Problem: Accurate NER model requires a massive amount of training data in the target domain

Proposal: Our self-adaptive NER retrieves unstructured knowledge for unconfident entities as queries

Results: Proposed model improved the performance on the entities not appeared in the pre-training corpus by retrieving the text mentioning it as knowledge on-the-fly

Proposed Method: Self-Adaptive NER

The model retrieves text chunks as knowledge on-the-fly from the unstructured knowledge base. The retrieval uses embeddings of unconfident entities as queries.

Two-Stage NER Algorithm

Step1 [1st Prediction]: The Model predicts the entity candidates with only the original input

Step2: The model uses the unconfident entities as queries

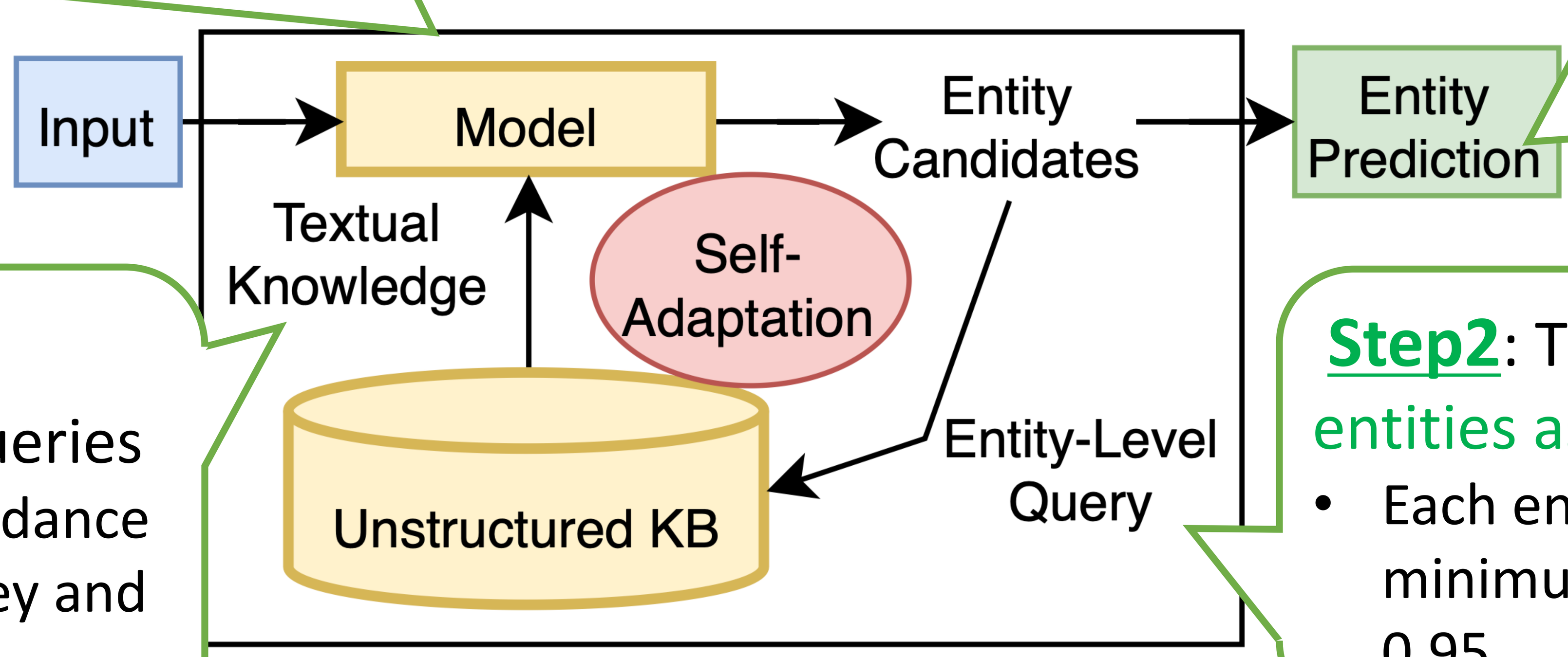
- Each entity candidate is unconfident if the minimum prediction probability is less than 0.95

Step3: The model retrieves text chunks as knowledge with the queries

- Text chunks are retrieved in accordance with the dot products between key and query embeddings

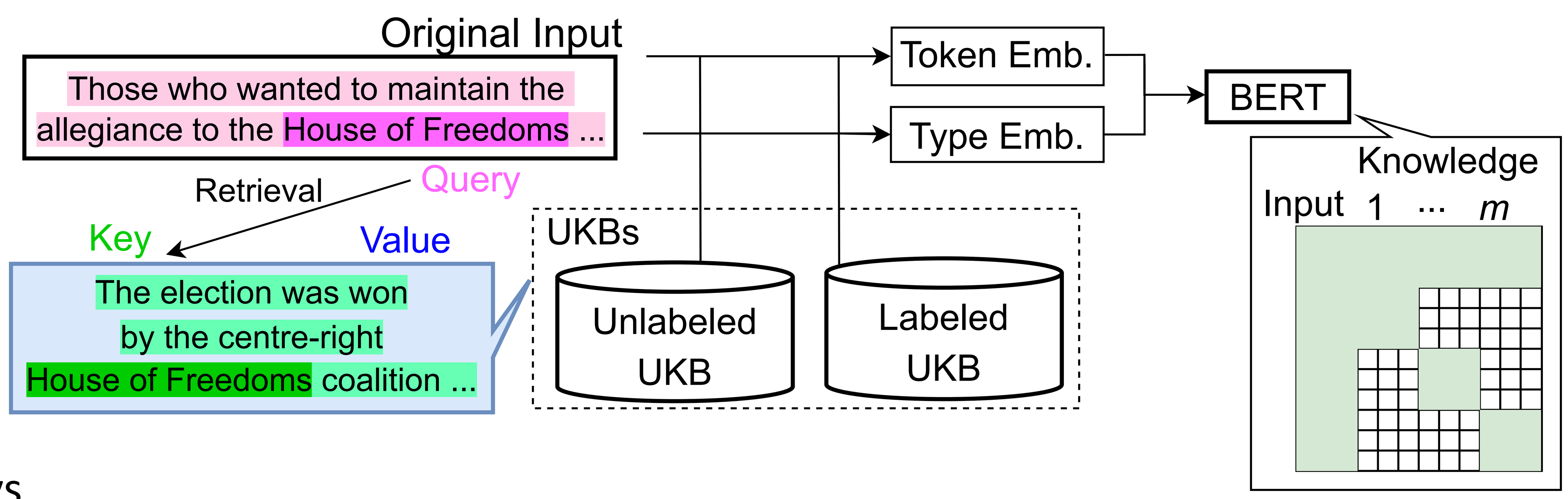
Step4 [2nd Prediction]: The model concatenates the knowledge to the input

- It ignores the intra-knowledge interaction in the self-attention
- The quadratic complexity is reduced to a linear one



Knowledge Base (KB) Construction

- We prepare two KBs
  - For unlabeled KB, we split the textual corpus into chunks
  - For labeled KB, we use training data
  - The labels are represented as token type embeddings
- Keys of KBs are the n-gram and sentence embeddings
  - Each text chunk is copied and retrieved with multiple keys
  - We hold only the n-grams that have a capital letter and no stop-word
  - Embeddings of the unconfident entities/the original input are used for the retrieval with n-gram/sentence embeddings



Evaluation

Data: Cross-NER [Liu+ 2021]

KB: Wikipedia articles in the same domain

Main Results

	AI.	Mus.	Lit.	Sci.	Pol.	Avg.
# Train (# NE types)	100 (14)	100 (13)	100 (12)	200 (17)	200 (9)	—
BERT†	50.37	66.59	59.95	63.73	66.56	61.44
DAPT†	56.36	73.39	64.96	67.59	70.45	66.55
NERBERT‡	60.39	76.23	67.85	71.90	73.69	70.01
BERT on CoNLL03	56.97 (1.05)	69.10 (1.08)	64.37 (0.73)	65.76 (0.58)	70.16 (0.56)	65.27 (0.80)
REALM-NER on CoNLL03	58.05 (1.15)	71.17 (0.63)	64.58 (0.69)	66.33 (0.66)	69.38 (0.36)	66.56 (0.80)
SA-NER on CoNLL03	**60.31** (1.03)	**72.20** (0.79)	**66.23** (1.30)	**68.22** (0.57)	**71.18** (0.57)	**67.62** (0.85)
BERT on NERBERT	62.05 (0.66)	76.45 (0.90)	69.68 (0.26)	72.10 (0.67)	74.38 (0.40)	70.93 (0.58)
REALM-NER on NERBERT	64.32 (0.31)	77.55 (0.69)	70.42 (0.60)	72.52 (0.42)	74.45 (0.38)	71.85 (0.43)
SA-NER on NERBERT	**65.27** (0.95)	**78.71** (0.47)	**71.79** (0.57)	**74.38** (0.19)	**74.63** (0.36)	**72.96** (0.51)

- The proposed model outperformed the previous models across all target domains, and It is further useful for the low-resource setting

Detailed Results

	# Entities	NERBERT	Proposed
All	3472	75.90 (0.22)	77.33 (0.19)
Seen in Training	661	84.05 (1.43)	85.20 (0.21)
Unseen in Training	2811	71.39 (0.29)	73.03 (0.36)
Seen in Pre-Training	3083	77.58 (0.17)	78.83 (0.29)
Unseen in Pre-Training	389	50.90 (1.63)	54.18 (1.85)

- Improvement was largest in the entities that are not appeared in the pre-training corpus
  - The proposed model can retrieve the knowledge which is not learned in the pre-training on-the-fly
  - Baselines were not good in this type because they depend the knowledge learned in pre-training

Qualitative Analysis

- Input has no evidence to confirm that the House of Freedoms is a political party, and the model predicts it as organization
- Knowledge provides the evidence by mentioning it in the context of election, and the model changes its prediction to political party

Input	the Association for the Rose in the Fist of Lanfranco Turci and those who wanted to maintain the allegiance to the House of Freedoms coalition.
Knowledge	The election was won in Sardinia by the centre-right House of Freedoms coalition ... voted party with 30.2% .
Prediction	organization → political party