RA-NER: RETRIEVAL AUGMENTED NER FOR KNOWLEDGE INTENSIVE NAMED ENTITY RECOGNITION

Zhenwei Dai
Amazon
zwdai@amazon.com

Chen Luo
Amazon
cheluo@amazon.com

Zhen Li
Amazon
amzzhn@amazon.com

Xianfeng Tang
Amazon
xianft@amazon.com

Hanqing Lu
Amazon
luhanqin@amazon.com

Rahul Goutam
Amazon
rgoutam@amazon.com

Haiyang Zhang
Amazon
hhaiz@amazon.com

ABSTRACT

NER (named entity recognition) model aims to recognize the named entities in the keywords. However, when the entities are extremely knowledge intensive, traditional NER model cannot encode all the knowledge in its parameters, thus fails to recognize those entities with high accuracy. In this paper, we propose retrieval-augmented NER model (RA-NER) to address this issue. RA-NER retrieves the most relevant information from an exhaustive external knowledge database to assist the entity recognition. We implement RA-NER for media related entity recognition task on an E-commerce search dataset, and achieve significant performance boost over the traditional deep-learning based NER model.

1 INTRODUCTION

Named Entity Recognition (NER) plays a crucial role in extracting and classifying entities within a given text, as demonstrated by Erdogan (2010). Current deep learning based NER model (Collobert & Weston, 2008) learns an encoder to embed each token of the query into an embedding vector and classifies the embedding vector into different named entity categories. However, this conventional approach may fall short in accurately identifying knowledge-intensive entities, as these entities might lack clear semantic distinctions from others. For instance, if we attempt to use an NER model to annotate the movie title in the query "Yosemite 2015", the term "Yosemite" is more likely to be identified as a national park name rather than a movie title. To correctly recognize it as a movie title, the NER model must have memorized the specific title "Yosemite" during training. Yet, given the vast number of movie titles and their constant updates, it becomes challenging for the NER model to memorize them all and keep up with the evolving landscape of movie titles.

To address the above issue, we propose retrieval-augmented NER (RA-NER) model which leverages an external knowledge database to help the NER model "memorize" the movie titles. Still using "Yosemite 2015" as the example, RA-NER could find the movie "Yosemite (2015)" from the movie title database. With this information, the NER model can easy recognize "Yosemite 2015" refers to a movie title instead of a national park. In our experiment, we implement RA-NER to recognize the media related entities (media titles and media-related person names) on an E-commerce search dataset and achieves over 10% F_1-score boost comparing to the deep learning based NER model.

2 METHOD

RA-NER includes two parts: (1) a retriever to retrieve the most relevant information from an external knowledge database and (2) a NER model that classifies tokens into different entities. Figure 1 shows

the design of RA-NER. Given a query, the retriever will first retrieve the most related information from the external knowledge database. Then, RA-NER will parse the input query along with the retrieved information to the NER model. Finally, the NER model encodes and annotate each token of the input query with a predetermined entity.

Retriever: The retriever aims to retrieve the most relevant information from the external knowledge dataset, where the "relevance" of the information can measured in using semantic or string-level similarity. For example, we can use HNSW (Hierarchical Navigable Small World) to search information with similar semantics (Malkov & Yashunin, 2018), or LSH (locality sensitive hashing) to retrieve information with high string-level similarity (Jaccard similarity).

NER Model: The NER model includes an encoder to encode the input tokens into embeddings, and a classification head to classify the token into different entities. Note that classification head only applies to the input query tokens. To improve the classification accuracy, NER model usually uses conditional random field (CRF) to better model the sequence labeling and label annotations (Patil et al., 2020).

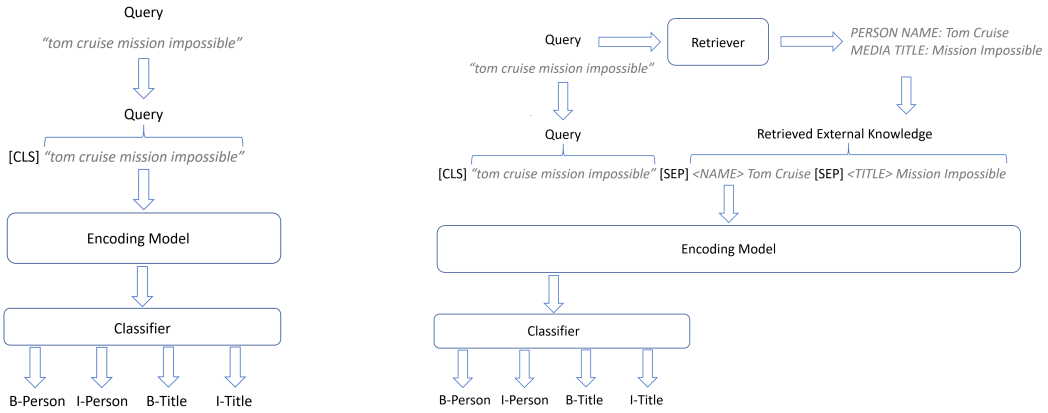

Figure 1: (a) model structure of vanilla deep learning based NER model (b) model structure of RA-NER.

3 EXPERIMENT

We conducted a performance comparison between the RA-NER model and the conventional deep learning-based NER model, utilizing an E-commerce search dataset. The annotation of entities, including product type, color, brand, etc., helps comprehend customers' search intent, facilitating the identification of products aligned with their preferences. In addition to general search queries, media-related searches necessitate the extraction of media titles such as book or movie titles, as well as media related individuals like authors, actors, and characters. Recognizing these media-related entities accurately poses a challenge due to their knowledge-intensive nature, making the conventional approach less effective for precise identification.

Dataset: The E-commerce search dataset comprises over 500K training samples and 100K testing samples. Each sample includes a genuine search keyword and its token-level entity annotation. The dataset annotates 15 distinct entity categories, encompassing media titles and media-related individuals, and spans over 10 different languages.

Model Training: We collect a media knowledge database including most of the popular media titles and media related individuals. RA-NER uses LSH as the retriever to retrieve the related information from the media knowledge database. For both the RA-NER and traditional deep learning based NER model, we take the same pre-trained encoder as the initial encoding model and fine-tune the whole NER model over the training samples.

Experiment Results: The performance of NER models are evaluated using the span-level (each named entity is a span) annotation F_1 score. Table 1 shows that RA-NER consistently outperforms the baseline by a great margin, demonstrating the strength of augmenting external media knowledge. Since the test samples do not just include the media related search keywords, only retrieving the media related knowledge may mistakenly augment the media knowledge to the non-media search

keywords. However, our experiment suggests that it does not hurt the non-media entity annotation accuracy, where RA-NER and baseline show comparable performance. Therefore, RA-NER is robust to the noise during the retrieval stage.

4 FUTURE WORKS

In this paper, we showcase the efficacy of retrieving an external knowledge database to enhance the accuracy of identifying knowledge-intensive entities. Moving forward, our aim is to develop a more intelligent retriever and seamlessly integrate the training of the retriever with the NER model. This synergistic approach holds the promise of retrieving more valuable information, thereby consistently enhancing the overall performance of the NER model.

5 URM STATEMENT

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2024 Tiny Papers Track.

REFERENCES

Ronan Collobert and Jason Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*, pp. 160–167, 2008.

Hakan Erdogan. Sequence labeling: Generative and discriminative approaches. In *Proc. 9th Int. Conf. Mach. Learn. Appl.*, pp. 1–132, 2010.

Yu A Malkov and Dmitry A Yashunin. Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. *IEEE transactions on pattern analysis and machine intelligence*, 42(4):824–836, 2018.

Nita Patil, Ajay Patil, and BV Pawar. Named entity recognition using conditional random fields. *Procedia Computer Science*, 167:1181–1188, 2020.

A APPENDIX

The experiment results comparing the performance of RA-NER versus the traditional deep learning based NER model.

Language	Media Title F_1	Media Related Individuals F_1	Other Entities F_1
English	+6.7	+2.6	-0.1
German	+6	+4.1	0.4
Spanish	+3.8	+2.2	-0.1
French	+6.7	+5.8	0.1
Italian	+8	+1.4	0.5
Japanese	+9.3	+0.1	0.2
Dutch	+9.5	+8.3	-0.7
Polish	+10.1	+5.6	-0.6
Portuguese	+6.6	+4.4	1.3
Turkish	+5	+8.6	-0.6
Arabic	+17.8	+12.2	-1.5

Table 1: Span-level annotation performance boost of RA-NER over the traditional deep learning-bsaed NER model.