Named Entity Recognition via Machine Reading Comprehension: A Multi-Task Learning Approach

Yibo Wang
Department of Computer Science
University of Illinois Chicago
ywang633@uic.edu

Wenting Zhao
Department of Computer Science
University of Illinois Chicago
wzhao41@uic.edu

Yao Wan
School of Computer Science and Technology
Huazhong University of Science and Technology
wanyao@hust.edu.cn

Zhongfen Deng
Department of Computer Science
University of Illinois Chicago
zdeng21@uic.edu

Philip S. Yu
Department of Computer Science
University of Illinois Chicago
psyu@uic.edu

Abstract

Named Entity Recognition (NER) aims to extract and classify entity mentions in the text into pre-defined types (e.g., organization or person name). Recently, many works have been proposed to shape the NER as a machine reading comprehension problem (also termed MRC-based NER), in which entity recognition is achieved by answering the formulated questions related to pre-defined entity types through MRC, based on the contexts. However, these works ignore the label dependencies among entity types, which are critical for precisely recognizing named entities. In this paper, we propose to incorporate the label dependencies among entity types into a multi-task learning framework for better MRC-based NER. We decompose MRC-based NER into multiple tasks and use a self-attention module to capture label dependencies. Comprehensive experiments on both nested NER and flat NER datasets are conducted to validate the effectiveness of the proposed Multi-NER. Experimental results show that Multi-NER can achieve better performance on all datasets.

1 Introduction

Named Entity Recognition (NER), which aims to locate and classify entity mentions in text into pre-defined types, is a fundamental task in information extraction (Chinchor and Robinson, 1997; Nadeau and Sekine, 2007). Typically, NER is formulated as a sequence labeling task, where each token is classified as one of the pre-defined types. However, the sequence labeling models can only assign one label to a token, resulting in the incapability of handling overlapping entities in nested NER (Finkel

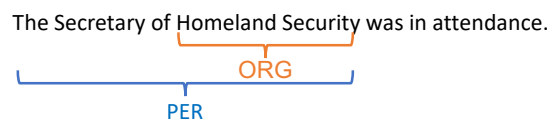The Secretary of Homeland Security was in attendance.
ORG
PER

Figure 1: An example of nested NER.

and Manning, 2009). Figure 1 shows an example of nested NER: *Homeland Security* can be recognized as ORGANIZATION , as well as PERSON.

To mitigate this issue, many works resort to formulating NER as a Machine Reading Comprehension (MRC) question answering task (termed MRC-based NER) (Wang et al., 2020; Li et al., 2020; Wang et al., 2022). For example, to recognize the ORGANIZATION, a natural-language question *"Which ORGANIZATION is mentioned in the text?"* is formulated. Then the goal of NER is transformed to answer the formulated questions through machine reading comprehension, given the contexts. MRC-based NER provides a unified solution for both flat and nested NER tasks since each entity type has its corresponding entity span positions as the answer, and these output answers are independent of each other.

Despite much progress having been made in MRC-based NER, existing approaches tend to ignore the label dependencies among entity types, which are critical for precise NER. Label dependencies indicate that different entities in the text have some relations with each other. For example, in a sentence *"Ousted WeWork founder Adam Neumann lists his Manhattan penthouse for $37.5 million"*, once *Adam Neumann* is recognized as PERSON, it is expected to help with the recognition
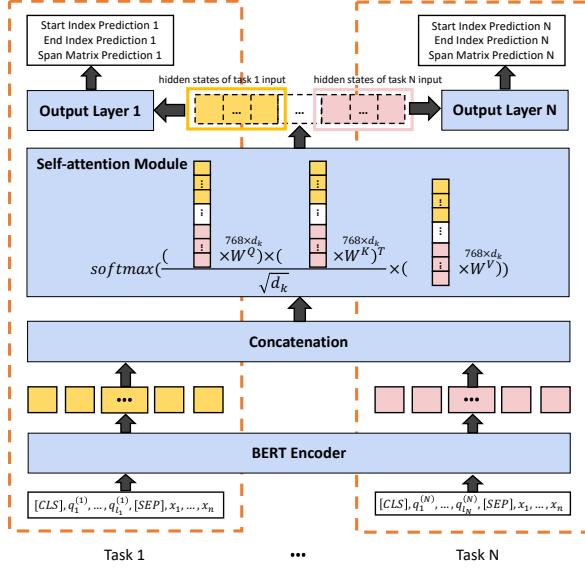
13

Figure 2: The architecture of our proposed Multi-NER. The input of each task is the concatenation of a question associated with task/entity type and context. The outputs are the corresponding start index prediction, end index prediction, and span matrix prediction.

of *WeWork* as ORGANIZATION because the *founder* preceding a person's name implies an organization.

To leverage the label dependencies among entity types, we propose a novel multi-task learning framework (termed Multi-NER) for MRC-based NER. In Multi-NER, MRC-based NER is decomposed into multiple tasks, each task focusing on one entity type. For each task, the corresponding input is the concatenation of an entity-class-related question and the context, and the output is expected to be the corresponding entity spans (i.e., start and end positions). The input is first encoded via a pre-trained BERT (Devlin et al., 2019). The concatenation of embeddings of all tasks are fed into a self-attention module, which can preserve the label dependencies between different entity types. Finally, task-specific output layers are applied to different tasks.

To validate the effectiveness of our proposed Multi-NER, we conduct experiments on the datasets of both flat NER and nested NER. Experimental results show that Multi-NER can benefit the MRC-based NER in both flat NER and nested NER, with the important label dependencies among entities preserved. Additionally, we also visualize the self-attention maps to examine whether the label dependencies have been successfully captured.

Overall, the contributions of this paper are two-fold: 1) We are the first to propose a multi-task

learning framework for MRC-based NER tasks to capture label dependencies between entity types; 2) The introduced self-attention maps are visualized to verify that the self-attention modules can capture label dependencies.

All the source code and datasets are available at https://github.com/YiboWANG214/MultiNER

2 Methodology

2.1 Problem Formulation

Given a sequence $X = \{x_1, x_2, \ldots, x_n\}$, where n denotes length of X, NER aims to find every entity mention in X and assign an entity type $y \in \mathcal{Y}$ to it. BERT-MRC (Li et al., 2020) transforms tagging-style NER to MRC format with a triplet of (QUESTION, ANSWER, CONTEXT). The natural language question $q^{(y)} = \{q_1^{(y)}, \ldots, q_{l_y}^{(y)}\}$, where l_y denotes length of $q^{(y)}$, is related to the entity type y and considered as QUESTION; The positions $P_{start,end}^y$ of entity mentions of y is considered as ANSWER; The input sequence X is considered as CONTEXT. Given X and $q^{(y)}$, the goal of BERT-MRC is to predict $P_{start,end}^y$.

Our Multi-NER applies the same MRC format but further decomposes BERT-MRC into multiple tasks, where each task $i \in \{1, 2, \ldots, |\mathcal{Y}|\}$ only focuses on one entity type. Thus, the task set $\{1, 2, \ldots, |\mathcal{Y}|\}$ and the entity type set $\mathcal{Y}$ are bijection. Instead of processing one QUESTION at a time, Multi-NER processes all QUESTIONs in a multi-task framework. Therefore, in Multi-NER settings, given a CONTEXT X and multiple questions $\{q^{(y)}\}_{y \in \mathcal{Y}}$, the goal is to predict $\{P_{start,end}^y\}_{y \in \mathcal{Y}}$.

2.2 Multi-NER

Figure 2 gives an overview of our proposed Multi-NER, which consists of $|\mathcal{Y}|$ tasks, where each task denotes the recognition of one specific entity type. To share information between tasks, one shared encoder is used across tasks, and a self-attention module is employed to capture label dependencies.

Single Task Learning For every single task $i \in \{1, 2, \ldots, |\mathcal{Y}|\}$, the input sequence is the concatenation of the natural language question $q^{(i)}$ and the context X, as follows:

$$I^{(i)} = [\text{CLS}], q^{(i)}, [\text{SEP}], X . \qquad (1)$$

The output of task i has three components: start index prediction, end index prediction, and span

14

matrix prediction. The start index prediction is the probability of each token being a start position; the end index prediction is the probability of each token being an end position; the span matrix prediction is the probability of each start-end pair being an entity mention position.

Task Interactions Task interactions are two-fold. First, one shared large language model like BERT (Devlin et al., 2019) is used as the encoder for all tasks to make the embedding space consistent. Thus, the embedding of task i is $E^{(i)} = Encoder(I^{(i)})$. Second, a self-attention module (Vaswani et al., 2017) is used across all tasks, accepting the concatenation of the embeddings of every task as input. The self-attention module incorporates information from every task and outputs the concatenation of hidden states

$$E = ATT\left(\left[E^{(1)}, \ldots, E^{(|\mathcal{Y}|)}\right]\right). \quad (2)$$

The self-attention module enables E with the property of capturing the label dependencies between entity types, making E a better representation.

Model Learning At training time, all tasks are trained jointly. The loss function for multi-task learning is defined as follows:

$$\mathcal{L} = \sum_{i=1}^{|\mathcal{Y}|}(\alpha\mathcal{L}_{start}^i + \beta\mathcal{L}_{end}^i + \gamma\mathcal{L}_{span}^i), \quad (3)$$

where α, β, and γ are tunable weights for start index prediction, end index prediction, and span matrix prediction. $\mathcal{L}_{start}^i$, $\mathcal{L}_{end}^i$ and $\mathcal{L}_{span}^i$ are cross entropy loss of task i.

3 Experiments

To evaluate the performance of the proposed Multi-NER, we compare it with a state-of-the-art baseline BERT-MRC (Li et al., 2020), on datasets of flat NER and nested NER. We also perform a case study with attention maps visualized to further analyze the ability of Multi-NER to capture label dependencies.

3.1 Datasets

We adopt three nested NER datasets (i.e., English ACE-2004 (Mitchell et al., 2005), English ACE-2005 (Walker et al., 2006) and GENIA (Ohta et al., 2002)) and one flat NER dataset (i.e., English CoNLL-2003 (Tjong Kim Sang and De Meulder,

2003)) to evaluate the performance of Multi-NER. ACE-2004 and ACE-2005 are two textual datasets from broadcast, newswire, telephone conversations and weblogs. GENIA is a collection of biomedical literature, containing Medline abstracts. CoNLL-2003 is extracted from Reuters news stories between August 1996 and August 1997. We conducted experiments on both nested NER datasets and flat NER dataset since our model is based on BERT-MRC, which can be applied to both flat and nested NER.

3.2 Experimental Settings

For a fair comparison, we select $BERT_{base}$ as the backbone encoder for all models. We adopt a one-layer linear transformation to predict the start index and end index and adopt a two-layer MLP with activation function GELU to predict the span matrix. We set the hidden size as 1,536 and the dropout rate as 0.1. More details of the hyperparameters setting are referred to the Appendix A.1. We follow the same process of question generation in (Li et al., 2020) to use annotation guideline notes, which are the guidelines provided to the annotators when building datasets, as references to construct questions.

3.3 Results and Analysis

Table 1 shows the experimental results on both nested NER datasets and flat NER dataset. From this table, we can observe that our Multi-NER achieves 85.34% on ACE 2004, 84.25% on ACE 2005, 81.13% on GENIA, and 92.33% on CoNLL-2003, achieving +1.3%, +0.4%, +1.24% and 1.25% improvement, respectively, when comparing with BERT-MRC. The performance improvements of Multi-NER on all the datasets indicate that formulating MRC-based NER into a multi-task learning framework to obtain label dependencies between different entity types can indeed bring model performance improvement.

Furthermore, the idea behind our proposed multi-task framework is to use different output layers to disambiguate between entity types and use a self-attention module to obtain label dependencies between entity types. To evaluate the contribution of the different output layers and the self-attention module, we also conduct ablation studies on all datasets. The experimental results in Table 2 show that both different output layers and the self-attention module contribute to Multi-NER.

Table 1: Experimental results on nested NER datasets and flat NER dataset.

Model	Precision	Recall	F1
ACE 2004			
BERT-MRC	84.63	83.46	84.04
Multi-NER	86.01	84.68	85.34 (+1.3)
ACE 2005			
BERT-MRC	83.22	84.48	83.85
Multi-NER	84.47	84.03	84.25 (+0.4)
GENIA			
BERT-MRC	79.47	80.32	79.89
Multi-NER	82.63	79.68	81.13 (+1.24)
CoNLL-2003			
BERT-MRC	90.61	91.55	91.08
Multi-NER	91.96	92.70	92.33 (+1.25)

Table 2: Ablation studies evaluate the contribution of components of Multi-NER.

Model	ACE04	ACE05	GENIA	CoNLL03
Multi-NER	85.34	84.25	81.13	92.33
(w/o att)	84.84	84.01	80.87	91.32
(w/o diff)	84.72	84.05	80.58	91.55

3.4 Case Study

To further study the effect of Multi-NER, we examine some randomly selected examples. As an example, for *"Ahbulaity Ahbudurecy , chairman of the Xinjiang Uigur Autonomous Region , presided at the opening ceremony , and leaders such as Tiemuer Dawamaiti , vice - chairman of the Standing Committee of the National People ' s Congress , Shimayi Aimaiti , member of the State Affairs Committee , etc . cut the ribbon for the meeting ."* from

---

BERT-MRC: [Ahbulaity Ahbudurecy]PER , [chairman of [the Xinjiang Uigur Autonomous Region]GPE]PER , presided at the opening ceremony , and [leaders such as ...... , [[etc]ORG]PER]PER. cut the ribbon for the meeting .

Multi-NER: [Ahbulaity Ahbudurecy]PER , [chairman of [the Xinjiang Uigur Autonomous Region]GPE]PER , presided at the opening ceremony , and [leaders such as ...... , etc]PER. cut the ribbon for the meeting .

Ground Truth: [Ahbulaity Ahbudurecy]PER , [chairman of [the Xinjiang Uigur Autonomous Region]GPE]PER , presided at the opening ceremony , and [leaders such as ...... , etc .]PER cut the ribbon for the meeting .

Figure 3: The ground truth, predictions of BERT-MRC and Multi-NER for a randomly selected example. The middle of the sentence is omitted for better presentation. The complete example is referred to Appendix A.2.
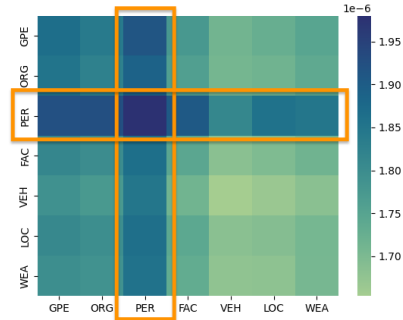

Figure 4: Attention map of the mean scores according to entity types of a randomly selected example from ACE 2004.

ACE 2004, we show the ground-truth entities and predicted results of BERT-MRC and Multi-NER in Figure 3. In BERT-MRC, *etc* is categorized as both ORG and PER, while in Multi-NER and ground truth *etc* is categorized as PER. Ambiguous tokens like *etc* are hard to categorize even with contextual information. However, when applying Multi-NER, different entity types and label dependencies are also considered, which is beneficial to ambiguous tokens.

We also show the attention map of the mean scores according to entity types of this example in Figure 4. We can see that PER has a relatively large impact on other entity types, helping the model improve performance on other entity types with PER information. We attribute it to label dependencies obtained by information sharing between entity types using the self-attention module.

4 Related Work

As language models have advanced (Devlin et al., 2019; Raffel et al., 2020; Zhao et al., 2023; Dong et al., 2023; Zhao et al., 2021; Liu et al., 2021), numerous efforts have emerged to enhance the performance of MRC-based NER. Zhang et al. (2022) incorporated different domain knowledge into MRC-based NER task to improve model generalization ability. Liu et al. (2022) proposed to use graph attention networks to capture label dependencies between entity types when applying MRC-based NER to electronic medical records. However, they only use entity type embeddings to build graph attention networks, ignoring the rich information in context. MRC-based NER is applied to different domains. Du et al. (2022) designed an MRC-based method for medical NER through both sequence

labeling and span boundary detection. Zhang and Zhang (2022) applied MRC-based NER for financial named entity recognition from literature. Wang et al. (2020) proposed MRC-based NER with the help of a distilled masked language model in e-commerce. Jia et al. (2022) applied MRC-based methods for multimodal named entity recognition.

The span-based methods (Eberts and Ulges, 2020) that formulate nested NER as a span classification task are also mentionable. Wan et al. (2022) improved span representation using retrieval-based span-level graphs based on n-gram similarity. Yuan et al. (2022) integrated heterogeneous factors like inside tokens, boundaries, labels, and related spans to improve the performance of span representation and classification. Shen et al. (2021) improved span-based NER using a two-stage entity identifier to filter out low-quality spans to reduce computational costs.

5 Conclusion

In this paper, we propose to incorporate the label dependencies among entity types into a novel multi-task learning framework (termed Multi-NER) for MRC-based NER. A self-attention mechanism is introduced to obtain label dependencies between entity types. Experimental results validate that Multi-NER outperforms BERT-MRC on both nested NER and flat NER. Case study and attention map visualization show that our introduced self-attention module is able to capture label dependencies among entities, contributing to a performance improvement.

Limitations

One limitation of our proposed Multi-NER lies in that the number of tasks depends on the number of entity types because each entity type is considered as a task. Depending on the model structure we use in Multi-NER, the number of parameters will increase by 4M for each additional entity type with $Max\_Length = 128$. One possible solution to solve this problem is using parameter-efficient fine-tuning methods like Hypernetworks (Ha et al., 2016) to effectively generate task-specific output layers. We leave this problem to our future work.

References

Nancy Chinchor and Patricia Robinson. 1997. Muc-7 named entity task definition. In *Proceedings of the 7th Conference on Message Understanding*, volume 29, pages 1–21.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Xiangjue Dong, Jiaying Lu, Jianling Wang, and James Caverlee. 2023. Closed-book question generation via contrastive learning. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 3150–3162, Dubrovnik, Croatia. Association for Computational Linguistics.

Xiaojing Du, Jia Yuxiang, and Zan Hongying. 2022. MRC-based medical NER with multi-task learning and multi-strategies. In *Proceedings of the 21st Chinese National Conference on Computational Linguistics*, pages 836–847, Nanchang, China. Chinese Information Processing Society of China.

Markus Eberts and Adrian Ulges. 2020. Span-based joint entity and relation extraction with transformer pre-training. In *ECAI 2020*, pages 2006–2013. IOS Press.

Jenny Rose Finkel and Christopher D Manning. 2009. Nested named entity recognition. In *Proceedings of the 2009 conference on empirical methods in natural language processing*, pages 141–150.

David Ha, Andrew M. Dai, and Quoc V. Le. 2016. Hypernetworks. *CoRR*, abs/1609.09106.

Meihuizi Jia, Xin Shen, Lei Shen, Jinhui Pang, Lejian Liao, Yang Song, Meng Chen, and Xiaodong He. 2022. Query prior matters: A mrc framework for multimodal named entity recognition. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 3549–3558.

Xiaoya Li, Jingrong Feng, Yuxian Meng, Qinghong Han, Fei Wu, and Jiwei Li. 2020. A unified MRC framework for named entity recognition. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5849–5859, Online. Association for Computational Linguistics.

Shuyue Liu, Junwen Duan, Feng Gong, Hailin Yue, and Jianxin Wang. 2022. Fusing label relations for chinese emr named entity recognition with machine reading comprehension. In *International Symposium on Bioinformatics Research and Applications*, pages 41–51. Springer.

Ye Liu, Yao Wan, Lifang He, Hao Peng, and Philip S. Yu. 2021. Kg-bart: Knowledge graph-augmented bart for generative commonsense reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 6418–6425.

Alexis Mitchell, Stephanie Strassel, Shudong Huang, and Ramez Zakhary. 2005. Ace 2004 multilingual training corpus. *Linguistic Data Consortium, Philadelphia*, 1:1–1.

David Nadeau and Satoshi Sekine. 2007. A survey of named entity recognition and classification. *Lingvisticae Investigationes*, 30(1):3–26.

Tomoko Ohta, Yuka Tateisi, Jin-Dong Kim, Hideki Mima, and Junichi Tsujii. 2002. The genia corpus: An annotated research abstract corpus in molecular biology domain. *Proceedings of the human language technology conference*, pages 73–77.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(140):1–67.

Yongliang Shen, Xinyin Ma, Zeqi Tan, Shuai Zhang, Wen Wang, and Weiming Lu. 2021. Locate and label: A two-stage identifier for nested named entity recognition. *arXiv preprint arXiv:2105.06804*.

Erik F. Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, pages 142–147.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Christopher Walker, Stephanie Strassel, Julie Medero, and Kazuaki Maeda. 2006. Ace 2005 multilingual training corpus. *Linguistic Data Consortium, Philadelphia*, 57:45.

Juncheng Wan, Dongyu Ru, Weinan Zhang, and Yong Yu. 2022. Nested named entity recognition with span-level graphs. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 892–903, Dublin, Ireland. Association for Computational Linguistics.

Qifan Wang, Li Yang, Bhargav Kanagal, Sumit Sanghai, D. Sivakumar, Bin Shu, Zac Yu, and Jon Elsas. 2020. Learning to extract attribute value from product via question answering: A multi-task approach. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20, page 47–55, New York, NY, USA. Association for Computing Machinery.

Yibo Wang, Congying Xia, Guan Wang, and Philip S. Yu. 2022. Continuous prompt tuning based textual entailment model for e-commerce entity typing. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 1383–1388. IEEE Computer Society.

Zheng Yuan, Chuanqi Tan, Songfang Huang, and Fei Huang. 2022. Fusing heterogeneous factors with triaffine mechanism for nested named entity recognition. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 3174–3186, Dublin, Ireland. Association for Computational Linguistics.

Yu Zhang, Jian Deng, Ying Ma, and Jianmin Li. 2022. A multi-task approach for machine reading comprehension form named entity recognition tasks. In *2022 3rd International Conference on Artificial Intelligence and Education (IC-ICAIE 2022)*, pages 480–485. Atlantis Press.

Yuzhe Zhang and Hong Zhang. 2022. Finbert-mrc: financial named entity recognition using bert under the machine reading comprehension paradigm. *arXiv preprint arXiv:2205.15485*.

Wenting Zhao, Ye Liu, Yao Wan, Yibo Wang, Zhongfen Deng, and Philip S. Yu. 2023. Localize, retrieve and fuse: A generalized framework for free-form question answering over tables.

Wenting Zhao, Ye Liu, Yao Wan, and Philip Yu. 2021. Attend, memorize and generate: Towards faithful table-to-text generation in few shots. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 4106–4117, Punta Cana, Dominican Republic. Association for Computational Linguistics.

A Appendix

A.1 Hyperparameters

The hyperparameter details are shown in Table 3.

A.2 Examples

The ground truth, predicted results of Multi-NER and BERT-MRC of a randomly selected example are shown in Table 4 and Figure 5. From the table, we can observe that the true positive, false positive, and false negative positions of BERT-MRC and Multi-NER are 8, 4, 3, and 9, 3, 2. Besides, most errors of Multi-NER are due to the ambiguity of the entity mention boundaries like punctuation.

Table 3: Hyperparameters for all models.

ACE 2004					
Model	Batch Size	Max Length	Learning Rate	Epoch	# Parameters
BERT-MRC	2	128	3e-5	14	112M
Multi-NER	2	128	2e-5	19	133M
ACE 2005					
Model	Batch Size	Max Length	Learning Rate	Epoch	# Parameters
BERT-MRC	2	128	2e-5	11	112M
Multi-NER	2	128	2e-5	16	133M
GENIA					
Model	Batch Size	Max Length	Learning Rate	Epoch	# Parameters
BERT-MRC	2	180	2e-5	9	112M
Multi-NER	2	128	2e-5	15	125M
CoNLL-2003					
Model	Batch Size	Max Length	Learning Rate	Epoch	# Parameters
BERT-MRC	2	200	3e-5	8	112M
Multi-NER	1	200	2e-5	18	123M

Table 4: Results of an example *"Ahbulaity Ahbudurecy , chairman of the Xinjiang Uigur Autonomous Region , presided at the opening ceremony , and leaders such as Tiemuer Dawamaiti , vice - chairman of the Standing Committee of the National People ' s Congress , Shimayi Aimaiti , member of the State Affairs Committee , etc . cut the ribbon for the meeting ."* from ACE 2004.

entity type	BERT-MRC	Multi-NER	Ground Truth
GPE	(5,9)	(5,9)	(5,9)
ORG	(28,37) (32,37) (44,47) (49,49)	(28,37) (32,37) (44,47)	(28,37) (44,47)
PER	(0,1) (3,9) (18,49) (21,22) (24,37) (39,40) (49,49)	(0,1) (3,9) (18,49) (21,22) (24,37) (39,40) (39,47) (42,47)	(0,1) (3,9) (18,50) (21,22) (24,37) (39,40) (42,47) (49,50)
FAC	-	-	-
VEH	-	-	-
LOC	-	-	-
WEA	-	-	-

BERT-MRC: [Ahbulaity Ahbudurecy]PER , [chairman of [the Xinjiang Uigur Autonomous Region]GPE]PER , presided at the opening ceremony , and [leaders such as [Tiemuer Dawamaiti]PER , [vice - chairman of [the Standing Committee of [the National People ' s Congress]ORG]ORG]PER , [Shimayi Aimaiti]PER , **member of** [**the State Affairs Committee**]ORG , [**etc**]ORG]PER]PER. cut the ribbon for the meeting .

Multi-NER: [Ahbulaity Ahbudurecy]PER , [chairman of [the Xinjiang Uigur Autonomous Region]GPE]PER , presided at the opening ceremony , and [leaders such as [Tiemuer Dawamaiti]PER , [vice - chairman of [the Standing Committee of [the National People ' s Congress]ORG]ORG]PER , [[Shimayi Aimaiti]PER , [member of [the State Affairs Committee]ORG]PER]PER , **etc**]PER. cut the ribbon for the meeting .

Ground Truth: [Ahbulaity Ahbudurecy]PER , [chairman of [the Xinjiang Uigur Autonomous Region]GPE]PER , presided at the opening ceremony , and [leaders such as [Tiemuer Dawamaiti]PER , [vice - chairman of [the Standing Committee of the National People ' s Congress]ORG]PER , [Shimayi Aimaiti]PER , [[member of [the State Affairs Committee]PER]ORG]PER , etc .]PER cut the ribbon for the meeting .

Figure 5: The ground truth, results of BERT-MRC and Multi-NER of a randomly selected example. Delete lines and underlines represent false positive and bolded tokens represent false negative. Punctuation errors are not indicated in the figure.