| **Sequence Labeling as Non-Autoregressive**  **Dual-Query Set Generation** | Sequence labeling is a crucial task in the NLP commu-  nity that aims at identifying and assigning spans within the input  sentence. It has wide applications in various fields such as informa-  tion extraction, dialogue system, and sentiment analysis. However,  previously proposed span-based or sequence-to-sequence models  conduct locating and assigning in order, resulting in problems  of error propagation and unnecessary training loss, respectively.  This paper addresses the problem by reformulating the sequence  labeling as a non-autoregressive set generation to realize locat-  ing and assigning in parallel. Herein, we propose a Dual-Query  Set Generation (DQSetGen) model for unified sequence labeling  tasks. Specifically, the dual-query set, including a prompted type  query and a positional query with anchor span, is fed into the  non-autoregressive decoder to probe the spans which correspond to  the positional query and have similar patterns with the type query.  By avoiding the autoregressive nature of previous approaches,  our method significantly improves efficiency and reduces error  propagation. Experimental results illustrate that our approach can  obtain superior performance on 5 sub-tasks across 11 benchmark  datasets. The non-autoregressive nature of our method allows for  parallel computation, achieving faster inference speed than com-  pared baselines. In conclusion, our proposed non-autoregressive  dual-query set generation method offers a more efficient and accu-  rate approach to sequence labeling tasks in NLP. Its advantages in  terms of performance and efficiency make it a promising solution  for various applications in data mining and other related fields. | Sequence Labeling  Sequence labeling [21], [22], [23], [24] is a critical task in  natural language processing (NLP) that can be used for a vari-  ety of applications, including slot filling, part-of-speech (POS)  tagging and named entity recognition (NER). Recent advances  in sequence labeling methods have leveraged pre-trained lan-  guage models (PLMs) [25], [26], [27] such as RoBERTa [28],  BART [29], and T5 [30] to perform token-level classification  using powerful text encoding capabilities. Span-based classifi-  cation methods [31], [32], [33], [34] have also been proposed to  extend token-level classification to span-level classification. In  addition to these classification-based methods, there have been  efforts to use sequence-to-sequence (Seq2Seq) PLMs to handle  sequence labeling tasks [17], [35], [36]. These models generate  target spans and their tag labels autoregressively through pointer-  based indexing [16] or tagging mechanisms [37]. Researchers  have also focused on decoding strategies. For instance, Zhao  et al. [38] proposes a novel hierarchical decoding model that  dynamically parses act, slot, and value in a structured manner.  In this work, we propose a novel approach to sequence label-  ing that differs from both span-based and Seq2Seq methods.  We model various sequence labeling tasks as dual-query set  generation in a non-autoregressive manner, which eliminates  error propagation and enables us to tag spans in one step.  Specifically, our proposed method leverages a dual-query set,  including a type query (semantic embedding) and a positional  query (anchor spans), to probe the input text and generate sets of  spans with specific semantic types. By modeling these tasks as  the dual-query set generation, we can eliminate error propaga-  tion and tag spans in one step, which can improve performance  while maintaining efficiency.  B. Set Generation  The set generation has been explored in machine learning,  where the output is a set of unordered labels. In the field of  computer vision, researchers have investigated set generation  approaches for Transformer-based object detection models, such  as DETR [39] that uses learnable query sets for task-specific  features. However, the influence of order on the performance  of various NLP tasks has been highlighted in previous studies,  including graph generation [40], NER [41], keyphrase genera-  tion [42], and multi-label classification [43]. Other approaches  have been proposed to explicitly model set properties for these  tasks, such as conducting an exhaustive search for the suitable  order [44] or modifying the optimization of the model [45].  Among the above-described related work about set generation,  Tan et al. [41] introduces an innovative sequence-to-set model  for nested NER that utilizes a set of fixed, trainable vectors  to learn valuable span patterns, which is closely related to our  DQSetGen. However, Tan et al. [41] concatenates the BERT  contextualized embeddings, the GloVE embeddings, part-of-  speech (POS) embeddings and character-level embeddings to-  gether, which runs counter to our simple application of BERT  embedding for sequence labeling tasks. Thus, we don’t compare  with Tan et al. [41] in terms of NER sub-tasks. In contrast,  we propose a dual-query set generation framework in a non-  autoregressive manner that enables the model to identify and  assign spans quickly, which is a fundamental requirement of  sequence labeling. |
| --- | --- | --- |

[16] H. Yan, T. Gui, J. Dai, Q. Guo, Z. Zhang, and X. Qiu, “A unified generative

framework for various NER subtasks,” in Proc. 59th Annu. Meeting Assoc.

Comput. Linguistics 11th Int. Joint Conf. on Natural Lang. Process., 2021,

pp. 5808–5822. [Online]. Available: https://aclanthology.org/2021.acl-

long.451

[17] S. Zhang, Y. Shen, Z. Tan, Y. Wu, and W. Lu, “De-bias for generative ex-

traction in unified NER task,” in Proc. 60th Annu. Meeting Assoc. Comput.

Linguistics, 2022, pp. 808–818, doi: 10.18653/v1/2022.acl-long.59.

[21] H. Chen, Q. Ma, L. Yu, Z. Lin, and J. Yan, “Corpus-aware

graph aggregation network for sequence labeling,” IEEE/ACM

Trans. Audio, Speech Lang. Proc., vol. 29, pp. 2048–2057, 2021,

doi: 10.1109/TASLP.2021.3084105.

[22] P. Zhu et al., “Improving chinese named entity recognition by large-scale

syntactic dependency graph,” IEEE/ACM Trans. Audio, Speech Lang.

Proc., vol. 30, pp. 979–991, 2022, doi: 10.1109/TASLP.2022.3153261.

[23] X. Chen et al., “LightNER: A lightweight tuning paradigm for low-

resource NER via pluggable prompting,” in Proc. 29th Int. Conf.

Comput. Linguistics, 2022, pp. 2374–2387. [Online]. Available: https:

//aclanthology.org/2022.coling-1.209

[24] X. Chen et al., “Continual multimodal knowledge graph construction,”

2023, arXiv.2305.08698.

[25] N. Zhang et al., “A comprehensive study of knowledge editing for large

language models,” 2024, arXiv:2401.01286.

[26] W. X. Zhao et al., “A survey of large language models,” 2023,

doi: 10.48550/arXiv.2303.18223.

[27] Y. Yao et al., “Editing large language models: Problems, methods, and

opportunities,” in Proc. Conf. Empirical Methods Natural Lang. Pro-

cess., 2023, pp. 10222–10240. [Online]. Available: https://aclanthology.

org/2023.emnlp-main.632

[28] Y. Liu et al., “Roberta: A robustly optimized bert pretraining approach,”

2019, arXiv:1907.11692.

[29] M. Lewis et al., “BART: Denoising sequence-to-sequence pre-training for

natural language generation, translation, and comprehension,” in Proc.

58th Annu. Meeting Assoc. Comput. Linguistics, 2020, pp. 7871–7880.

[Online]. Available: https://aclanthology.org/2020.acl-main.703

[30] C. Raffel et al., “Exploring the limits of transfer learning with a unified

text-to-text transformer,” J. Mach. Learn. Res., vol. 21, no. 140, pp. 1–67,

2020. [Online]. Available: http://jmlr.org/papers/v21/20-074.html

[31] H. Lin, Y. Lu, X. Han, and L. Sun, “Sequence-to-nuggets: Nested entity

mention detection via anchor-region networks,” in Proc. 57th Annu. Meet-

ing Assoc. Comput. Linguistics, 2019, pp. 5182–5192. [Online]. Available:

https://aclanthology.org/P19-1511

[32] X. Wang et al., “Automated concatenation of embeddings for struc-

tured prediction,” in Proc. 59th Annu. Meeting Assoc. Comput. Linguis-

tics 11th Int. Joint Conf. Natural Lang. Process., 2021, pp. 2643–2660,

doi: 10.18653/v1/2021.acl-long.206.

[33] S. Bell, H. Yannakoudakis, and M. Rei, “Context is key: Grammatical error

detection with contextual word representations,” in Proc. 14th Workshop

Innov. Use NLP Building Educ. Appl., 2019, pp. 103–115. [Online].

Available: https://aclanthology.org/W19-4410

[34] L. Sun, Y. Sun, F. Ji, and C. Wang, “Joint learning of token

context and span feature for span-based nested NER,” IEEE/ACM

Trans. Audio, Speech Lang. Proc., vol. 28, pp. 2720–2730, 2020,

doi: 10.1109/TASLP.2020.3024944.

[35] B. Athiwaratkun, C. N. D. Santos, J. Krone, and B. Xiang, “Augmented nat-

ural language for generative sequence labeling,” in Proc. Conf. Empirical

Methods Natural Lang. Process., 2020, pp. 375–385. [Online]. Available:

https://aclanthology.org/2020.emnlp-main.27

[36] T. Shibuya and E. H. Hovy, “Nested named entity recognition via second-

best sequence learning and decoding,” Trans. Assoc. Comput. Linguistics,

vol. 8, pp. 605–620, 2020, doi: 10.1162/tacl\_a\_00334.

[37] J. Straková, M. Straka, and J. Hajic, “Neural architectures for nested NER

through linearization,” in Proc. 57th Annu. Meeting Assoc. Comput. Lin-

guistics, 2019, pp. 5326–5331. [Online]. Available: https://aclanthology.

org/P19-1527

[38] Z. Zhao, S. Zhu, and K. Yu, “A hierarchical decoding model for

spoken language understanding from unaligned data,” in Proc. IEEE

Int. Conf. Acoust., Speech Signal Process., 2019, pp. 7305–7309,

doi: 10.1109/ICASSP.2019.8682463.

[39] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov,

and S. Zagoruyko, “End-to-end object detection with transform-

ers,” in Proc. Comput. Vis. 16th Euro. Conf., 2020, pp. 213–229,

doi: 10.1007/978-3-030-58452-8\_13.

[40] X. Chen, X. Han, J. Hu, F. J. Ruiz, and L. Liu, “Order matters:

Probabilistic modeling of node sequence for graph generation,” 2021,

arXiv.:2106.06189.

[41] Z. Tan, Y. Shen, S. Zhang, W. Lu, and Y. Zhuang, “A sequence-to-set

network for nested named entity recognition,” in Proc. 13th Int. Joint

Conf. Artif. Intell., Virtual Event, 2021, pp. 3936–3942, doi: 10.24963/ij-

cai.2021/542.

[42] J. Ye, T. Gui, Y. Luo, Y. Xu, and Q. Zhang, “One2Set: Generating diverse

keyphrases as a set,” in Proc. 59th Annu. Meeting Assoc. Comput. Linguis-

tics 11th Int. Joint Conf. Natural Lang. Process., 2021, pp. 4598–4608.

[Online]. Available: https://aclanthology.org/2021.acl-long.354

[43] P. Yang, F. Luo, S. Ma, J. Lin, and X. Sun, “A deep reinforced sequence-

to-set model for multi-label classification,” in Proc. 57th Annu. Meeting

Assoc. Comput. Linguistics, 2019, pp. 5252–5258. [Online]. Available:

https://aclanthology.org/P19-1518

[44] O. Vinyals, S. Bengio, and M. Kudlur, “Order matters: Sequence to se-

quence for sets,” in Proc. 4th Int. Conf. Learn. Representations, Y. Bengio

and Y. LeCun, Eds., San Juan, Puerto Rico, May 2-4, 2016. [Online].

Available: http://arxiv.org/abs/1511.06391

[45] K. Qin, C. Li, V. Pavlu, and J. Aslam, “Adapting RNN sequence prediction

model to multi-label set prediction,” in Proc. Conf. North Amer. Chapter

Assoc. Comput. Linguistics: Hum. Lang. Technol., 2019, pp. 3181–3190.

[Online]. Available: https://aclanthology.org/N19-1321