TriG-NER: Triplet Grid Framework for Discontinuous Named Entity Recognition

WWW 2025

Rina Carines Cabral, Soyeon Caren Han, Areej Alhassan, Riza Batista-Navarro, Goran Nenadic, Josiah Poon

Rina Carines Cabral

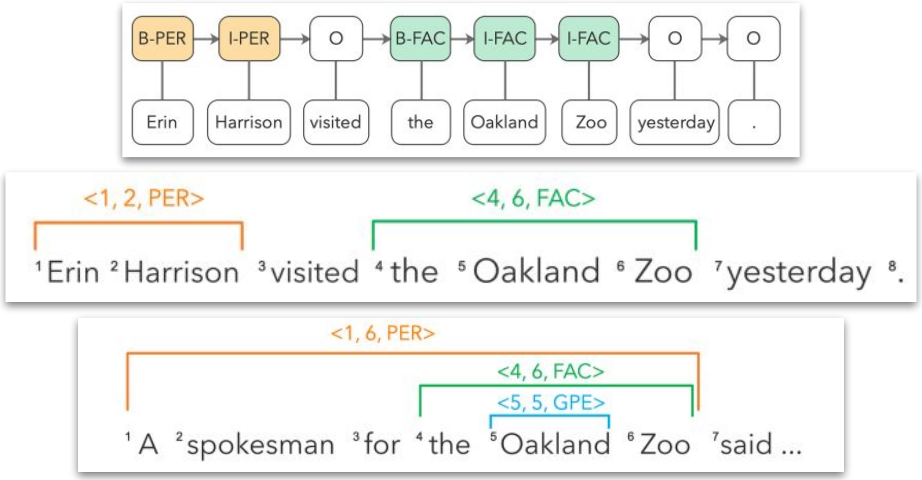
Named Entity Recognition (NER)

Steps:

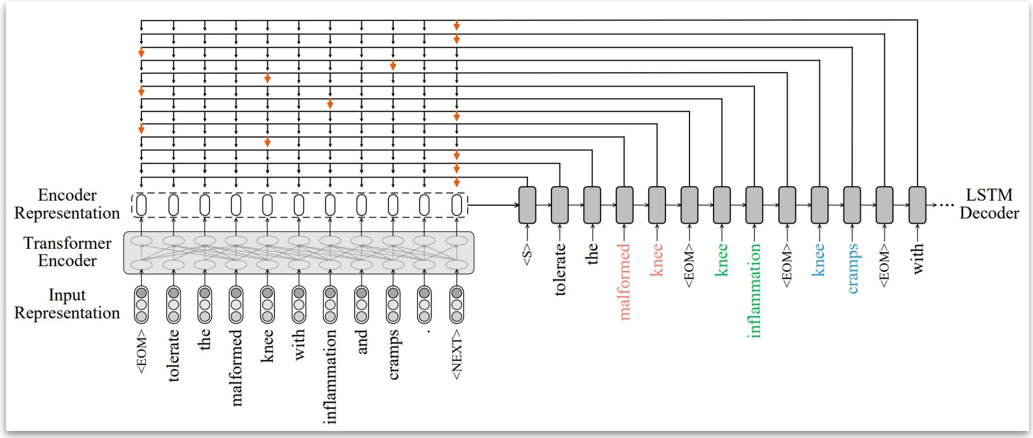
1. *Identifying **entity boundaries**, if any, on a piece of text*
2. *Identifying what the **entity type** is for each entity*

NER Schemes:

- *Sequence Tagging*
- *Span-based/Tuple-based*
- *Sequence-to-sequence*



(Wang et al, 2022)

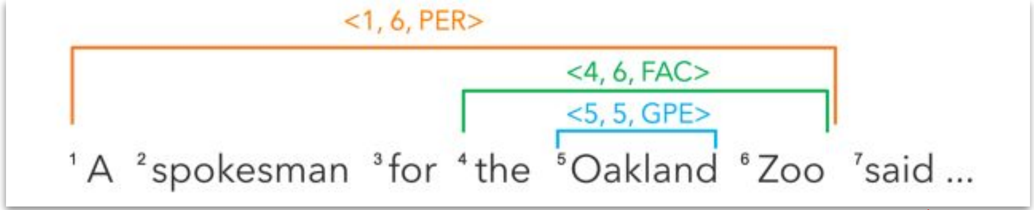


(Fei et al, 2021)

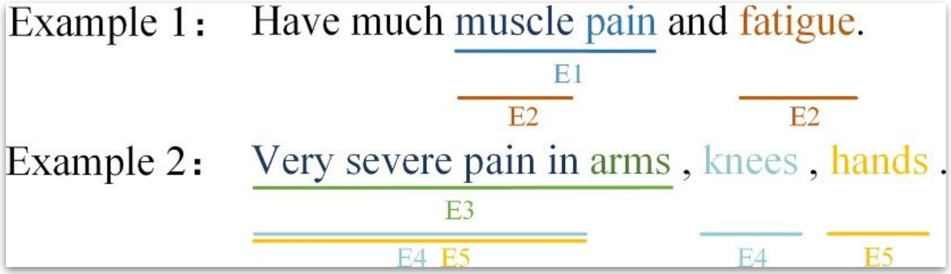
Named Entity Recognition (NER)

Types:

1. *Continuous/Flat Named Entities*
2. *Nested Named Entities*
3. *Overlapping Named Entities*
4. ***Discontinuous Named Entities***
5. *Multi-category Named Entities*



(Wang et al, 2022)



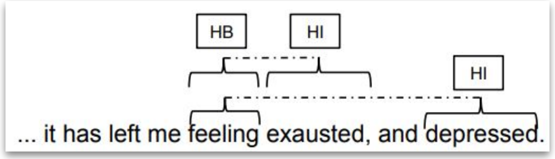
(Zhen et al, 2023)

Discontinuous NER

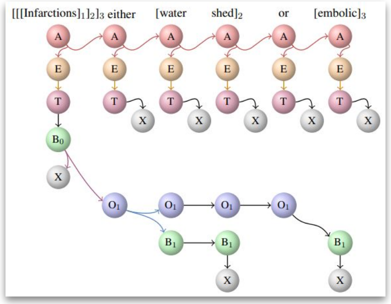
Entity mentions which are interrupted by non-entity tokens within the sequence.

New Schemes:

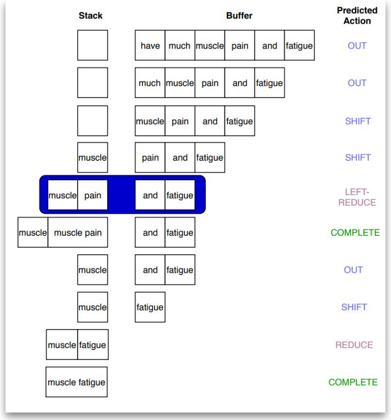
- *Adapted sequence tagging*
- *Adapted span-based*
- *Hypergraphs*
- *Transition-based*
- *Grid tagging*
- *LLM Prompting*



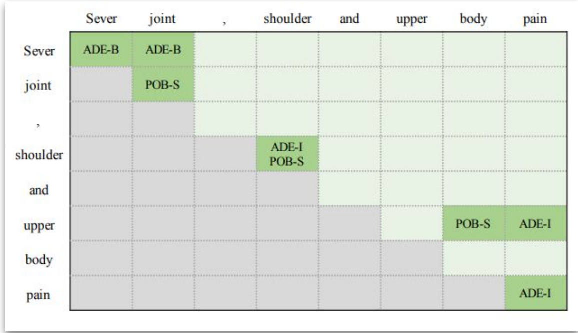
(Metke-Jimenez and Karimi, 2016)



(Muis and Lu, 2016)



(Dai et al, 2020)



(Wang et al, 2021)

Discontinuous NER

Entity mentions which are interrupted by non-entity tokens within the sequence.

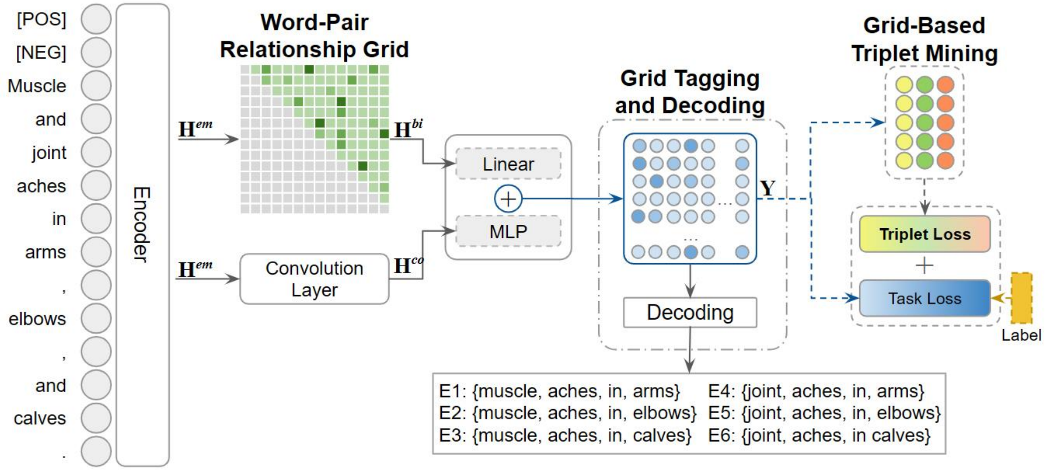
Limitations of new approaches:

- *Heavy reliance on task-specific tagging strategies*
- *Focus on sample-based learning*
- *Existing grid-tagging methods lack mechanisms to differentiate between similar and dissimilar word-pairs*

TriG-NER

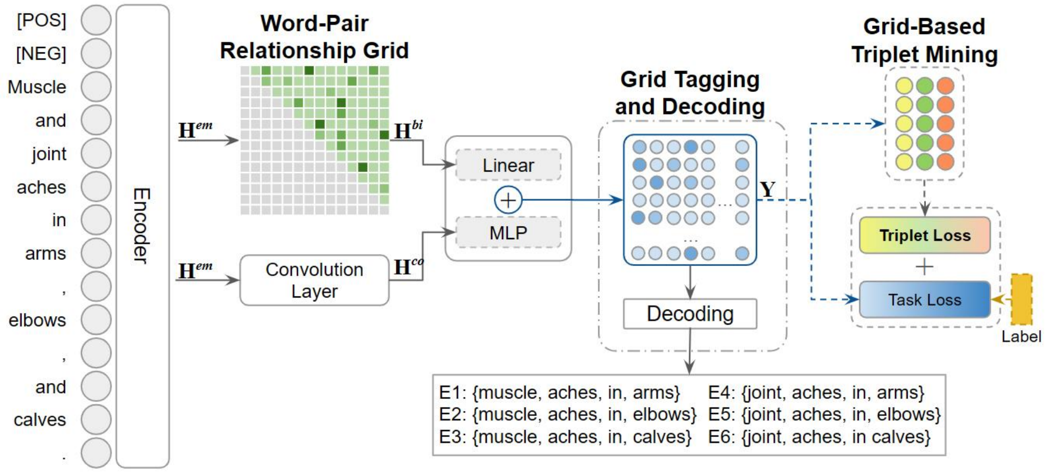
Triplet-Grid Framework that leverages token-based triplet loss to learn fine-grained word-pair relationships for DNER.

1. *Token-based Triplet Loss for NER*
2. *Grid-based Triplet Framework*
3. *Extensive Evaluations and Qualitative Analysis*



Methodology

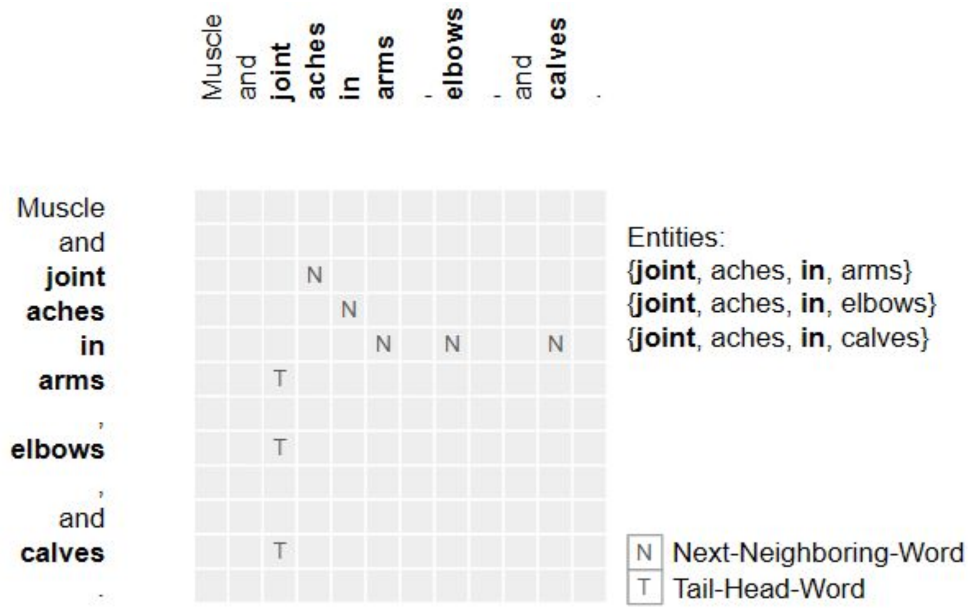
1. *Grid-based NER Models*
2. *Grid Tagging and Decoding*
3. *Word-Pair Relationship Grid*
4. *Grid-Based Triplet Mining*



1. Grid-based NER Models

NER as a word-to-word relation classification problem

- *Input:* $X = \{x_1, x_2, ..., x_n\}$
- *Output:* $Y = \{y_{11}, y_{12}, ..., y_{nn}\} \in R^{n \times n \times c}$



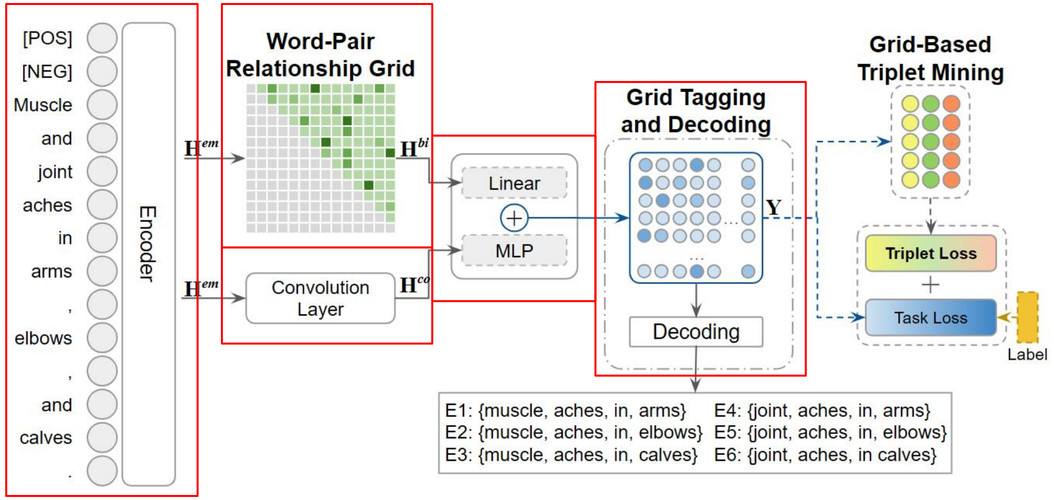
2. Grid Tagging and Decoding

Classes/Tags:

1. *Next-Neighboring-Word (NNW)*
2. *Tail-Head-Word (THW)*
3. *None*

3. Word-Pair Relationship Grid

- *Explicitly models the word-pair relationships using the contextualized representations for each word and applying a biaffine transformation.*
- *Encoder:*
    - *BioBERT*
    - *BioClinicalBERT*
    - *PharmBERT*
    - *PubMedBERT*

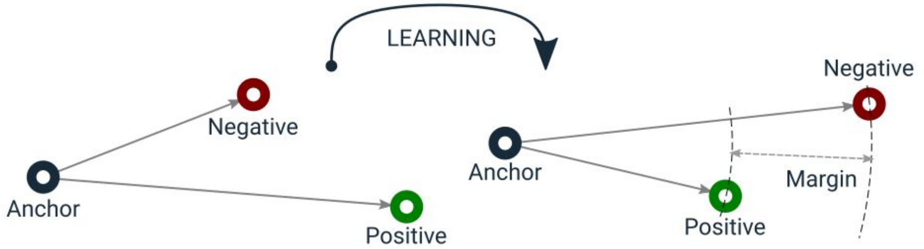


THE UNIVERSITY OF SYDNEY THE UNIVERSITY OF MELBOURNE MANCHESTER 1824 The University of Manchester

Triplet Loss

- *Deep metric learning objective that encourages similar points together and dissimilar points far apart*
- *Components:*
  - *Three points: Anchor (a), Positive (p), Negative (n)*
  - *Distance metric (f)*
  - *Margin (m)*
- *Typically defines similarity by class membership*

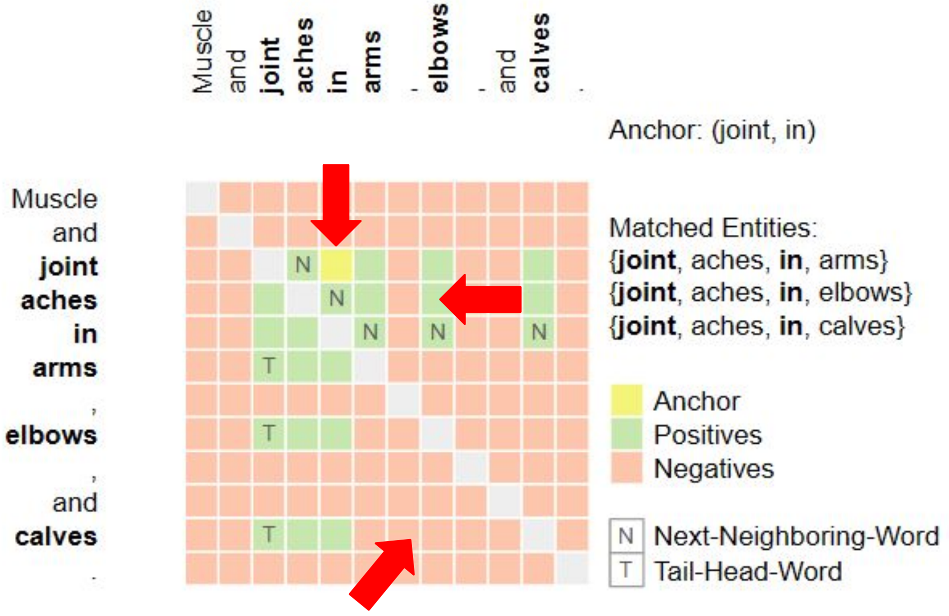
$$L_{triplet} = \sum max(f(a,p) - f(a,n) + m, 0)$$



(Source)

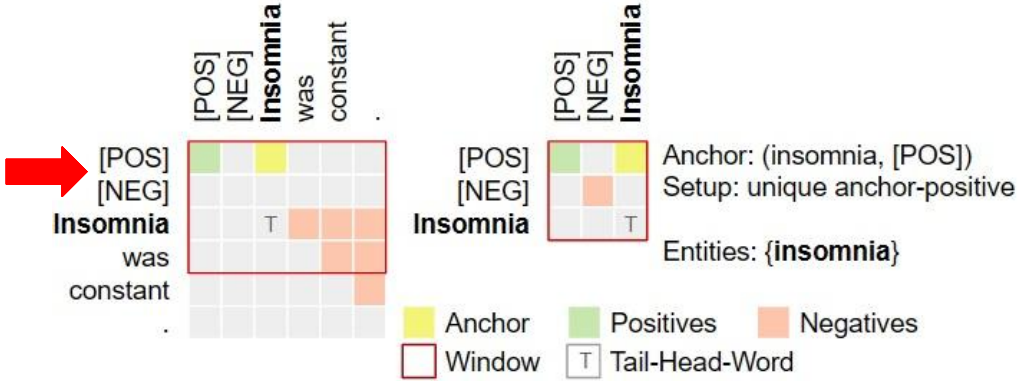
4. Grid-based Token-Level Triplet Mining

- *Token-Level Component Candidates*
  - *Anchor - any word-pair that appears in any entity*
  - *Positive - any word-pair that co-exists with the anchor in any entity*
  - *Negative - any word-pair that does not belong to any entity the anchor is a part of*



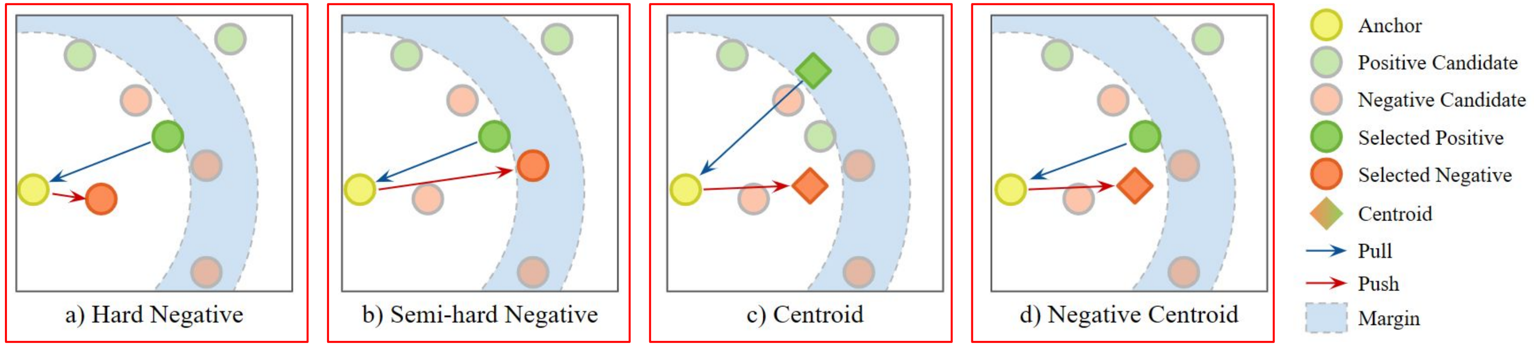
4. Grid-based Token-Level Triplet Mining

- *For special cases such as one-word entities or one-word samples, we include special tokens [POS] and [NEG].*
- *Candidate window*
- *Unique anchor-positive pair setup*



Triplet Selection

Informative valid triplets are where the positive candidate is farther to the anchor than the negative candidate by a certain margin.



Datasets

- *CADEC*
  - *Nature: medication consumer posts from the online forum AskAPatient.com*
  - *Entity: Adverse Drug Reactions (ADR)*
- *ShARe13 and ShARe14*
  - *Nature: clinical notes*
  - *Entity: Disease Disorders*

Table A1: Data statistics

	CADEC	ShARe13	ShARe14
Total Sentences	7,597	18,767	34,618
Total Entities	6,318	11,148	19,073
Continuous Entities	5,639	10,060	17,417
- Percentage	89.25%	90.24%	91.32%
- Number of tokens	1-36	1-9	1-9
Disc. Entities	679	1,088	1,658
- Percentage	10.75%	9.76%	8.68%
- Number of tokens	2-13	2-7	2-7
- Start-End Distance	3-20	3-23	3-23

THE UNIVERSITY OF SYDNEY THE UNIVERSITY OF MELBOURNE MANCHESTER 1824 The University of Manchester

Overall Performance

- Overall improvement of 0.76-1.23% overall
- DiscSent improvement of 0.79-3.19%
- DiscEnt improvement of 3.98-5.13%

Table 2: Comparison of performance from our best-performing models for the overall datasets and for discontinuous elements, including sentences containing at least one discontinuous entity (DiscSent) and discontinuous entities only (DiscEnt). Bold indicates best scores while underline shows next best. † indicates replicated results.

	Overall			DiscSent	DiscEnt
CADEC	**F1**	**P**	**R**	**F1**	**F1**
MAC [36]	71.50	70.50	72.50	69.80	44.40
$W^2NER^\dagger$ [15]	72.67	72.02	73.33	69.25	45.78
$TOE^\dagger$ [16]	72.24	74.28	70.30	67.98	40.00
Corro [3]	71.90	-	-	-	35.90
Ours	**73.43**	**75.35**	71.62	**70.59**	**49.71**
ShARe13	**F1**	**P**	**R**	**F1**	**F1**
MAC [36]	81.20	84.30	78.20	68.10	55.90
$W^2NER^\dagger$ [15]	82.16	84.13	80.29	68.46	57.38
$TOE^\dagger$ [16]	81.92	85.05	79.02	67.82	57.06
Corro [3]	82.00	-	-	-	52.10
Ours	**83.22**	**86.44**	80.24	**69.09**	**60.06**
ShARe14	**F1**	**P**	**R**	**F1**	**F1**
MAC [36]	81.30	78.20	84.70	69.70	54.10
$W^2NER^\dagger$ [15]	81.31	78.93	83.84	63.08	52.70
$TOE^\dagger$ [16]	80.67	78.67	82.78	61.04	49.29
Corro [3]	81.80	-	-	-	49.80
Ours	**82.54**	80.36	84.83	**72.89**	**59.23**

Ablation Studies - Triplet Selection

- *Centroid (CE) strategy consistently shows promising results across the datasets*
- *Semi-hard Negative (SN) and Hard Negative (HN) shows some high scores for the DiscEnt subsets however, it sacrifices the overall performance.*
- *All selection methods, except Hard Negative, generally outperform and is competitive with the best baseline model.*

Table 3: Comparison of different triplet selection methods based on the best-performing setup for each method. Bold indicates best scores while underline shows next best. † indicates replicated results from the baseline. HN: Hard Negative; SN: Semi-hard Negative; CE: Centroid; NC: Negative Centroid

	CADEC		ShARe13		ShARe14	
Method	Overall	DiscEnt	Overall	DiscEnt	Overall	DiscEnt
[15]†	72.67	45.75	82.16	57.38	81.31	52.70
HN	71.61	45.41	81.79	54.45	81.87	57.35
SN	72.21	49.35	82.56	56.30	82.19	53.79
CE	73.43	48.55	83.22	57.14	82.42	56.22
NC	73.33	46.75	82.43	56.22	82.54	54.40

THE UNIVERSITY OF SYDNEY THE UNIVERSITY OF MELBOURNE MANCHESTER 1824 The University of Manchester

Ablation Study - Window Size

- *Implementing a window significantly improves performance compared to not having a window.*
- *Optimal window size varies per dataset*
    - *Larger window for CADEC*
    - *Smaller windows for ShARE13/14*

Table 5: Comparison of different window sizes. Bold indicates best scores while underline shows next best.

Window Size	CADEC	ShARe13	ShARe14
None	71.49	81.74	81.78
1	71.65	81.21	81.91
5	72.77	82.02	**82.54**
10	72.88	**83.22**	81.19
15	70.84	81.26	80.81
20	70.67	81.79	81.33
25	**73.43**	81.83	81.83

Ablation Studies - Encoder Models

- *BioBERT has best results for CADEC*
- *PubMedBERT has best results for ShARe13/14*
- *TriG-NER improves performances by 0.93-1.22% except for BioClinicalBERT.*

Table 6: Comparison of different language models used in the encoder with and without our triplet framework based on the best-performing setup for each dataset. Bold indicates the overall best scores for each dataset while an underline shows the better score regarding the application of our framework.

PLM	TriG-NER	CADEC	ShARe13	ShARe14
BioBERT [13]	×	72.50	80.25	80.75
	✓	73.43	80.72	80.79
BioClinicalBERT [2]	×	71.49	81.78	81.00
	✓	71.42	81.89	80.27
PharmBERT [31]	×	70.78	80.25	80.00
	✓	71.90	80.39	81.11
PubMedBERT [8]	×	70.19	82.00	81.42
	✓	71.39	83.22	82.54

Qualitative Analysis

W^2NER

- *Processes word pairs in isolation limiting its ability to recognize overlapping discontinuous entities*

LLMs

- *Has alignment and indexing problems*
- *Identifies relevant parts of the DNE but does not relate the parts together to return as one entity*
- *Prone to extracting entities that are unrelated to the entity type provided*
- *Prone to simply returning the whole input*

Input
['Pain', 'and', 'cramping', 'in', 'my', 'hands', 'and', 'lower', 'legs', '.']
Gold Standard
{'entity': 'Pain in my hands', 'index': [0, 3, 4, 5], 'type': 'ADR'}, {'entity': 'Pain in my lower legs', 'index': [0, 3, 4, 7, 8], 'type': 'ADR'}, {'entity': 'cramping in my lower legs', 'index': [2, 3, 4, 7, 8], 'type': 'ADR'}, {'entity': 'cramping in my hands', 'index': [2, 3, 4, 5], 'type': 'ADR'}
Ours - 4/4 (100%)
{'entity': 'Pain in my lower legs', 'index': [0, 3, 4, 7, 8], 'type': 'ADR'}, {'entity': 'Pain in my hands', 'index': [0, 3, 4, 5], 'type': 'ADR'}, {'entity': 'Pain in lower legs', 'index': [0, 3, 7, 8], 'type': 'ADR'}, {'entity': 'cramping in my hands', 'index': [2, 3, 4, 5], 'type': 'ADR'}, {'entity': 'cramping in my lower legs', 'index': [2, 3, 4, 7, 8], 'type': 'ADR'}, {'entity': 'cramping in lower legs', 'index': [2, 3, 7, 8], 'type': 'ADR'}
W^2NER - 2/4 (50%)
{'entity': 'Pain in my hands', 'index': [0, 3, 4, 5], 'type': 'ADR'}, {'entity': 'cramping in my hands', 'index': [2, 3, 4, 5], 'type': 'ADR'}
Gemini Zero Shot CoT - 0/4 (0%)
{'entity': 'Pain', 'index': [0], 'type': 'ADR'}, {'entity': 'cramping', 'index': [2], 'type': 'ADR'}, {'entity': 'hands', 'index': [5], 'type': 'ADR'}, {'entity': 'lower legs', 'index': [7, 8], 'type': 'ADR'}
Gemini Few Shot CoT - 0/4 (0%)
{'entity': 'Pain and cramping', 'index': [0, 1, 2], 'type': 'ADR'}, {'entity': 'hands', 'index': [5], 'type': 'ADR'}, {'entity': 'lower legs', 'index': [7, 8], 'type': 'ADR'}
GPT-4o Zero Shot CoT - 0/4 (0%)
{"entity": "Pain", "index": [0], "type": "ADR"}, {"entity": "cramping", "index": [2], "type": "ADR"}
GPT-4o Few Shot CoT - 0/4 (0%)
{"entity": "Pain", "index": [0], "type": "ADR"}, {"entity": "cramping", "index": [2], "type": "ADR"}, {"entity": "Pain and cramping", "index": [0, 1, 2], "type": "ADR"}, {"entity": "Pain and cramping in my hands", "index": [0, 1, 2, 3, 4, 5], "type": "ADR"}, {"entity": "Pain and cramping in my hands and lower legs", "index": [0, 1, 2, 3, 4, 5, 6, 7, 8], "type": "ADR"}

THE UNIVERSITY OF SYDNEY THE UNIVERSITY OF MELBOURNE MANCHESTER 1824 The University of Manchester

Conclusion

- *We present **TriG-NER**, a flexible triplet-grid framework leveraging a token-level, grid-based triplet loss that incorporates and enhances word-pair relationships to improve extraction of discontinuous named entities.*
- *We demonstrate significant improvements over state-of-the-art grid-based architectures.*
- *Future work could explore LLM integration and expanding application to other structured prediction tasks.*

References

- *Hao Fei, Donghong Ji, Bobo Li, Yijiang Liu, Yafeng Ren, and Fei Li. 2021. Rethinking boundaries: End-to-end recognition of discontinuous mentions with pointer networks. In Proceedings of the AAAI conference on artificial intelligence, Vol. 35. 12785–12793.*
- *Xiang Dai, Sarvnaz Karimi, Ben Hachey, and Cecile Paris. 2020. An Effective Transition-based Model for Discontinuous NER. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 5860–5870. doi:10.18653/v1/2020.acl-main.520*
- *Alejandro Metke-Jimenez and Sarvnaz Karimi. 2016. Concept Identification and Normalisation for Adverse Drug Event Discovery in Medical Forums.. In BMDID@ ISWC.*
- *Aldrian Obaja Muis and Wei Lu. 2016. Learning to Recognize Discontiguous Entities. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Jian Su, Kevin Duh, and Xavier Carreras (Eds.). Association for Computational Linguistics, Austin, Texas, 75–84. doi:10.18653/v1/D16-1008*
- *Yu Wang, Hanghang Tong, Ziye Zhu, and Yun Li. 2022. Nested named entity recognition: A survey. ACM Trans. Knowl. Discov. Data 16, 6, Article 108 (Jul. 2022), 29. doi:10.1145/3522593*
- *Yucheng Wang, Bowen Yu, Hongsong Zhu, Tingwen Liu, Nan Yu, and Limin Sun. 2021. Discontinuous Named Entity Recognition as Maximal Clique Discovery. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, Online, 764–774. doi:10.18653/v1/2021.acl-long.63*

Thank you!

Paper *Code*