Fine-Grained Spatial Information Extraction in Radiology as Two-turn Question Answering

UTHealth The University of Texas Health Science Center at Houston School of Biomedical Informatics

Surabhi Datta, MS[1], Kirk Roberts, PhD[1]

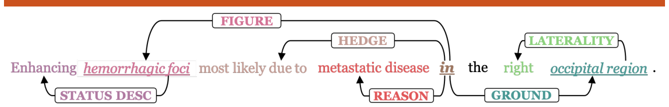
[1]The School of Biomedical Informatics | The University of Texas Health Science Center at Houston

For additional information, please contact the author via email: Surabhi.Datta@uth.tmc.edu

INTRODUCTION

- Radiology reports contain important clinical information about spatially-grounded radiographic findings and medical devices
- NLP-generated fine-grained spatial labels can be used for various applications such as automated image labeling and deep phenotyping
- Recent research[1-4] shows advantages of framing IE tasks as QA over traditional sequence labeling approaches
- Multi-turn QA[5-7] captures hierarchical dependency of entities
- We frame spatial IE as two-turn QA (BERT[8]-based)

REPRESENTATION SCHEMA



- Information organized according to frame semantics – Rad-SpatialNet[9]
- Lexical units of a frame
  - spatial trigger (*in*), radiological entities (*hemorrhagic foci*)
- Spatial Frame Element - Figure (w.r.t. trigger), Size (w.r.t. entity)
- Descriptive Frame Element - Status, Temporality (w.r.t. entity)

RESULTS

Target entities	Uncased	Cased	M+Cased
Spatial trigger	89.99	89.50	90.07
Finding	76.89	78.26	76.11
Anatomy	87.56	87.40	87.46
Device	91.87	92.68	93.12
Tip	99.18	98.41	99.32
Location descriptor	81.50	81.21	80.89
Other descriptor	84.19	84.24	84.09
Assertion	78.48	80.85	79.40
Position	69.68	71.41	72.81
Quantity	85.54	85.37	83.23
Process	60.93	59.26	60.19

Table 1. Average F1 measures over two 10-fold CVs for target entity extraction. M+Cased: MIMIC+Cased.

Table 2. Average F1 measures over 10-fold CV for frame element extraction. Desc: Descriptive. Q_f: Query$_{find}$. Q_{f+d}: Query$_{find + desc}$. M+Cased: MIMIC+Cased. Count: # of annotations.

	Frame Elements	Proposed approach Uncased		Cased		M+Cased		Baseline M+Cased	Count
		Q_f	Q_{f+d}	Q_f	Q_{f+d}	Q_f	Q_{f+d}		
Spatial	Figure	**78.13**	77.29	76.72	77.57	76.44	77.40	65.12	1491
	Ground	83.76	83.40	83.31	82.27	83.17	**83.77**	71.51	1537
	Hedge	75.47	76.44	**77.18**	76.42	75.90	74.97	57.82	249
	Diagnosis	69.32	73.32	**73.94**	72.67	65.47	67.92	50.76	190
	Position Status	68.72	68.75	66.98	67.12	68.43	**70.37**	60.37	167
	Relative Position	77.19	76.42	**77.53**	76.71	75.78	76.15	66.33	398
	Distance	84.65	86.54	85.36	85.20	87.94	**90.09**	88.05	45
	Reason	39.51	32.34	39.51	**49.81**	17.71	44.89	0	33
	Associated Process	48.52	**54.63**	43.15	42.29	38.95	41.36	0	21
	Morphologic	52.48	58.14	49.92	**60.52**	48.04	45.53	–	69
	Size Desc	76.16	73.80	78.16	78.94	78.56	**78.98**	–	93
	Distribution Pattern	57.45	63.62	59.74	64.01	59.22	**66.03**	–	65
	Composition	41.46	33.63	41.67	**46.88**	26.49	20.48	–	17
	Laterality	88.43	88.51	**89.35**	87.49	87.78	87.32	–	612
	Size/Measurement	45.43	**48.44**	41.51	43.59	34.46	32.06	–	23
Desc	Status	**64.67**	62.60	63.38	61.67	59.17	59.09	–	452
	Quantity	72.56	72.32	72.82	71.61	72.47	**73.11**	–	130
	Temporal	70.87	70.63	70.5	71.47	67.31	**71.78**	–	113
	Negation	58.08	61.06	**67.75**	65.04	60.95	61.83	–	103

MATERIALS & METHODS

Data

- 400 MIMIC-III[10] radiology reports annotated as per Rad-SpatialNet
- Chest X-rays, Brain MRIs, Babygrams

Problem formulation

- Information extracted from report text by answering queries
- Two-turn QA (Target entity & FE extraction)
  - 1st turn: Spatial triggers and main radiological entities
  - 2nd turn: Spatial and descriptive FEs

Query Construction Example

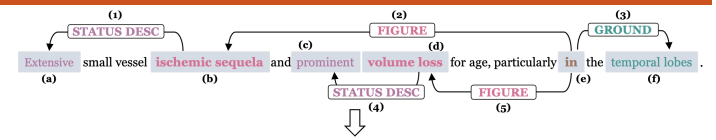
- **Target entity extraction**
  - **Entity type -** Spatial trigger
  - Q_f **-** find all spatial trigger entities in the context.
- **Spatial and descriptive FE extraction**
  - **FE type -** Figure • **Target entity type -** Spatial trigger
  - **Target entity span** from turn 1 – in
  - Q_f **-** find all clinical finding entities in the context that have a figure relationship with spatial trigger entity **in**.
  - Q_{f+d} **-** Figure refers to finding or device or tip entities that are described with respect to an anatomical structure. + Q_f

QA framework

- Input to BERT - merge query q & context c (report text)
  - `[[CLS]` q `[SEP]` c `[SEP]]`
- Handles queries that have multiple answers
- Models for target (turn 1) and FE (turn 2) extraction trained jointly

Baseline - Sequence Labeling

- BERT model pre-trained on MIMIC-III and fine-tuned to extract spatial triggers & spatial FEs
- Input to BERT - `[[CLS]` *sentence* `[SEP]]`
- Encoder output → linear classification layer to predict labels per token



		Queries	Answer spans
Turn 1 Entities	b, d	find all clinical finding entities in the context	*ischemic sequela, volume loss*
	e	find all spatial trigger entities in the context	*in*
	f	find all anatomical structure entities in the context	*temporal lobes*
	a, c	find all descriptor entities in the context	*extensive, prominent*
	-	find all medical device entities in the context	NONE
Turn 2 Frame Elements	2, 5	find all clinical finding entities that have a figure relationship with spatial trigger entity ***in***	*ischemic sequela, volume loss*
	3	find all anatomical structure entities that have a ground relationship with spatial trigger entity ***in***	*temporal lobes*
	1	find all descriptor entities that have a status descriptor relationship with clinical finding entity ***ischemic sequela***	*extensive*
	4	find all descriptor entities that have a status descriptor relationship with clinical finding entity ***volume loss***	*prominent*
	-	find all medical device entities that have a figure relationship with spatial trigger entity ***in***	NONE

Figure. Two-turn QA overview.

DISCUSSION

- Performance (F1) improvement of QA approach compared to sequence labeling
- Casting IE as QA is still under-explored on clinical domain datasets
- Covers more detailed radiological information
- First work to use QA for spatial & radiology IE
- Incorporating more information about a frame element in the query results in better results
- Moderate results for infrequent frame elements
- Future work - evaluate on larger dataset, evaluate generalizability on multi-institutional datasets and other imaging modalities

CONCLUSIONS

- Our two-turn QA approach outperforms traditional transformer-based sequence labeling in extracting spatial triggers and spatial FEs
- Extracting radiology findings with contextual information facilitates downstream clinical applications

ACKNOWLEDGEMENTS

This work was supported by NIBIB: R21EB029575 and PCORI: ME-2018C1-10963.

REFERENCES

1. X. Li, J. Feng, Y. Meng, Q. Han, F. Wu, J. Li, A Unified MRC Framework for Named Entity Recognition, *in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 5849–5859.
2. J. Liu, Y. Chen, K. Liu, W. Bi, X. Liu, Event Extraction as Machine Reading Comprehension, *in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 1641–1651.
3. C. Sun, Z. Yang, L. Wang, Y. Zhang, H. Lin, J. Wang, Biomedical named entity recognition using BERT in the machine reading comprehension framework, arXiv:2009.01560 [cs]arXiv:2009.01560.
4. P. Banerjee, K. K. Pal, M. Devarakonda, C. Baral, Knowledge Guided Named Entity Recognition for BioMedical Text, arXiv:1911.03869 [cs]arXiv:1911.03869.
5. X. Li, F. Yin, Z. Sun, X. Li, A. Yuan, D. Chai, M. Zhou, J. Li, Entity-Relation Extraction as Multi-Turn Question Answering, *in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 1340–1350.
6. F. Li, W. Peng, Y. Chen, Q. Wang, L. Pan, Y. Lyu, Y. Zhu, Event Extraction as Multi- turn Question Answering, *in: Findings of the Association for Computational Linguistics: EMNLP 2020*, 2020, pp. 829–838.
7. X. D. Wang, L. Weber, U. Leser, Biomedical Event Extraction as Multi-turn Question Answering, *in: Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis*, 2020, pp. 88–96.
8. 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.
9. Surabhi Datta, Morgan Ulinski, Jordan Godfrey- Stovall, Shekhar Khanpara, Roy F. Riascos- Castaneda, and Kirk Roberts. 2020b. Rad-SpatialNet: A Frame-based Resource for Fine-Grained Spatial Relations in Radiology Reports. *In Proceedings of The 12th Language Resources and Evaluation Conference*, pages 2251–2260.
10. Alistair E.W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data, 3*(1), pages1-9.

NLP - Natural Language Processing, IE – Information Extraction, QA - Question Answering, BERT - Bidirectional Encoder Representations from Transformers, FE - Frame Element