MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

Towards Structuring Clinical Texts: Joint Entity and Relation Extraction from Japanese Case Report Corpus

Yoshimasa KAWAZOE[1]

Associate Professor

[1] *Graduate School of Medicine, The University of Tokyo*

Co-authors: Daisaku SHIBATA[1], Emiko SHINOHARA[1], Kiminori SHIMAMOTO[1]

AIDH IMIA

Background: Challenges with clinical text usage

- Clinical text contains valuable information for research purposes
- it is still challenging to extract specific or relevant information by NLP

Spelling inconsistency / variants

血圧110-120 mmHg　　収縮期120mmHg　　転倒した　ベッドから落ちた

Blood pressure　　**Systolic BP**　　**Fall**　　**Fall from bed**

Factuality（positive or Negative）

関節炎を認めた　　筋力低下や錐体路徴候を伴わない　　胸水あるも腹水なく

Arthritis present　　**Muscles weakness and pyramidal signs absent**　　**Pleural effusion present, ascites absent**

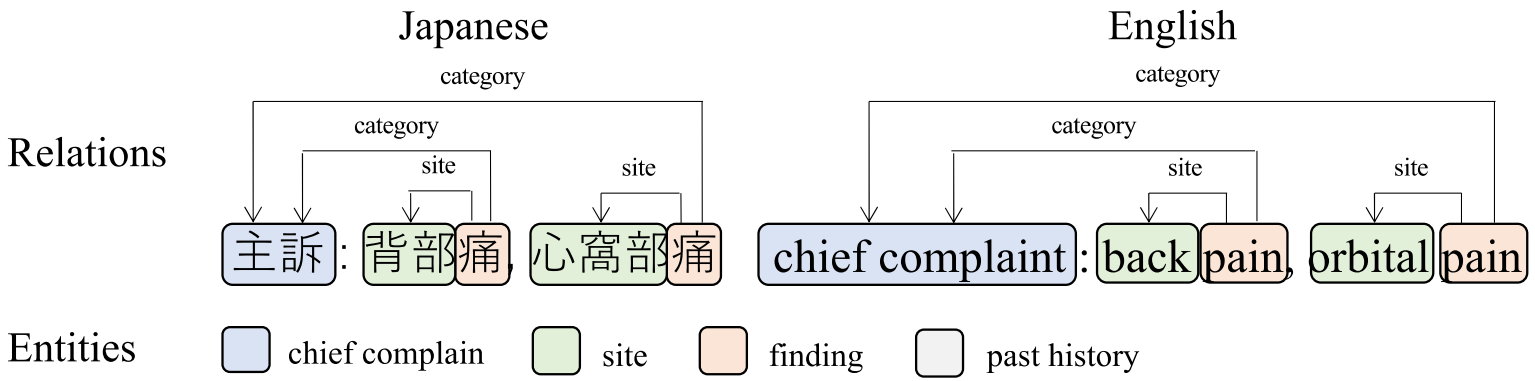
Factuality (General knowledge .. It did not occur in the patient）

術後に出血が予想される　　網膜症が合併しやすい

Bleeding is expected after surgery.　　**Retinopathy tends to occur as a complication**

Background: Information retrieval requires a text corpus

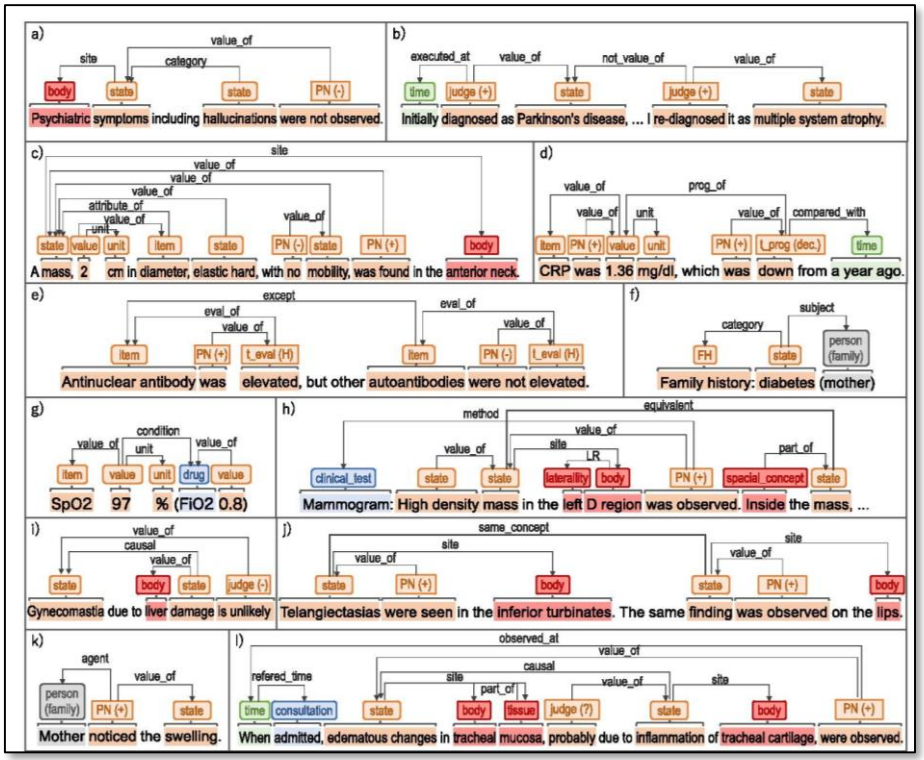
- Datasets (text + annotation) for developing NLP technology to extract information from text.
- Entity tags represent the type of entity, while relationship tags represent the type of relationship between entities.



Background: iCorpus[1]

- A corpus consists cases reports of rare and intractable disease.
- Based on annotation criteria of versatility, comprehensiveness, and consistency.
- Formulated as NLP tasks (Named Entity Recognition and Relation Extraction)
  - 183 Documents, 56 Entity types, 35 Relationships
- Publicly available for research use

https://ai-health.m.u-tokyo.ac.jp/home/research/corpus



1. Shinohara E et al. J Biomed Inform. 2022 Oct;134:104200.

Objective

1. Showing the baseline performance of Named Entity Recognition (NER) and Relation Extraction (RE) using iCorpus.
2. Also, showing the difference in performance between two types of BERT[1]
   - **Clinical-BERT:** Pretrained on clinical text
   - **Wikipedia-BERT:** Pretrained on Wikipedia (JP) text

1. Jacob Devlin et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805

Method: Compare the two types of BERTs

BERT is pre-trained on a large corpus of text, then fine-tuned with just one additional layer for specific tasks, such as NER and RE.

Clinical-BERT [1]

- BERT-base: 12 (L), 12 (A), 768 (E)
- Clinical text (120 million lines)
- 25,000 vocabularies (BPE)
- **Domain: Clinical**

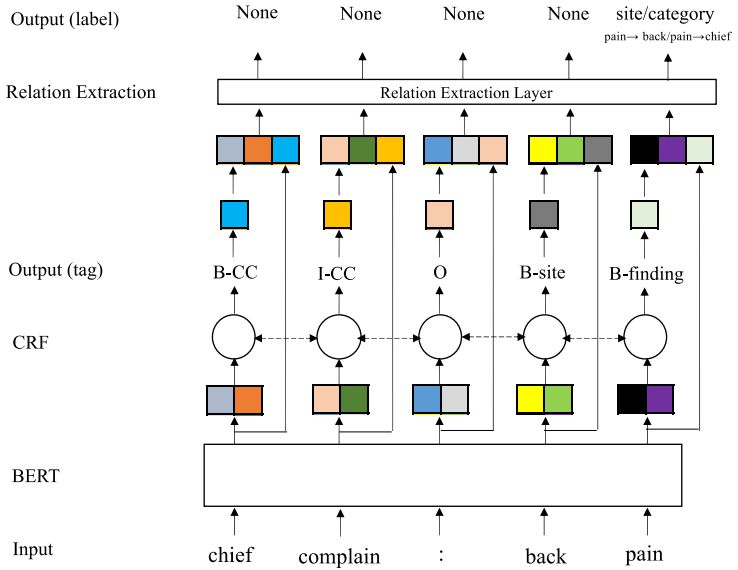
Wikipedia-BERT [2]

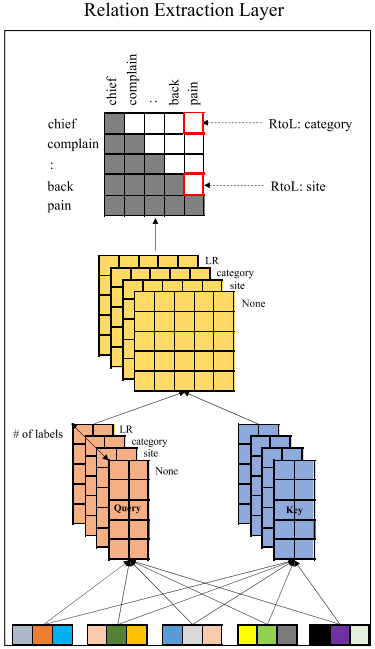
- BERT-base: 12 (L), 12 (A), 768 (E)
- Wikipedia (ja) (17 million lines)
- 32,000 vocabularies (BPE)
- **Domain: General**

1. Kawazoe Y, et al. PLoS One. 2021 Nov 9;16(11):e0259763.

2. https://alaginrc.nict.go.jp/nict-bert/index.html

Method: Joint NER-RE Model[1]





- NER assigns a suitably named entity tag to every token
- RE classifies the relation labels between tokens
- Joint NER-RE model simultaneously conducts NER and RE and optimizes the model weight.

1. Ma Y et al. Named Entity Recognition and Relation Extraction Using Enhanced Table Filling by Contextualized Representations. https://doi.org/10.5715/jnlp.29.187

Method: Experiment settings

Table 1. Statistics of iCorpus

Documents		183
Average	Characters (S.D)	1,692 (593)
	Entities (S.D)	394 (129)
	Relations (S.D)	387 (127)
Entity tags		113 (B-, I- + O)
Relation tags		36 (35 + None)

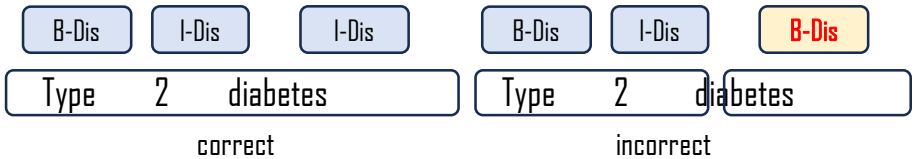
Evaluate the two Joint NER-RE models

Metrics: The average of Macro-F1 and Micro-F1 through 5-fold cross-validation

- Train 64%, Val 16%, Test 20%

Macro-F1: Calculated by the total true positives, false positives, and false negatives.

Micro-F1: The average of F1 scores across multiple classes



Results

	NER		RE	
	Micro-F1 (95%CI)	Macro-F1 (95%CI)	Micro-F1 (95%CI)	Macro-F1 (95%CI)
Clinical-BERT	0.912 * (0.907-0.916)	0.601 (0.580-0.621)	0.759 * (0.757-0.761)	0.611 (0.585-0.637)
Wiki-BERT	0.892 (0.884-0.899)	0.586 (0.567-0.606)	0.741 (0.734-0.747)	0.591 (0.571-0.611)

* significantly difference

Discussions: Comparison with related research

1. i2b2/VA dataset [1]
   - Discharge summary (en)
   - Entity: **3**, Relationships: **8**
   - NER **0.859**, RE: **0.737** (Micro F1)

2. JaMIE [2]
   - Radiology reports (jp)
   - Entity: **9**, Relationships: **10**
   - NER **0.956**, RE: **0.865** (Micro F1)

3. JaMIE [2]
   - Medical history Reports (jp)
   - Entity types: **9**, Relationships: **10**
   - NER **0.855**, RE: **0.710** (Micro F1)

4. iCorpus (this study)
   - Case Reports (jp)
   - Entity types: **56**, Relationships: **35**
   - NER **0.912**, RE: **0.759** (Micro F1)

1. Uzuner Ö et al. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. J Am Med Inform Assoc. 2011 Sep-Oct;18(5):552-6.
2. Fei Cheng et al. JaMIE: A Pipeline Japanese Medical Information Extraction System. arXiv:2111.04261.

Conclusions

1. We showed the baseline performances of NER and RE on iCorpus.
   - iCorpus: Manually annotated 56 types of entities and 35 types of relationships.
   - NER: Micro-F1 of 0.913, Macro-F1 of 0.601 (Clinical-BERT)
   - RE: Micro-F1 of 0.759 and Macro-F1 of 0.611 (Clinical-BERT)
2. We also showed the domain-specific clinical BERT tended to show better performance than the Wikipedia BERT.