

DICE: Data-Efficient Clinical Event Extraction with Generative Models

Mingyu Derek Ma, Alexander K. Taylor, Wei Wang, Nanyun Peng

University of California, Los Angeles

読み手: 三輪誠 (豊田工大)

図表は論文・著者スライド, [1]より

[1] Caufield et al., 2019. A comprehensive typing system for information extraction from clinical narratives. medRxiv.

論文のまとめ

• 医療分野のイベント抽出

- 世界初の医療分野のイベントコーパスMACCROBAT-EE
 - イベント: **トリガ**と**項**の述語項関係
- 生成モデルを利用したイベント抽出 (DICE) を提案
- MACCROBAT-EEにおけるDICEの評価・エラー分析を通して、その有用性・課題を確認
- ニュースのイベント抽出 (ACE05) においても評価し、DICEの汎用性を実証

A man presented with an **abnormal nodule** measuring 0.8 x 1.5 cm in the **left upper lung lobe** imaged through **chest computed tomography** scanning.

| | |
|----------------------|----------------------|
| Event trigger | nodule |
| Event type | Sign_symptom |
| Detailed description | abnormal |
| Area | 0.8 x 1.5 cm |
| Biological structure | left upper lung lobe |

Diagnostic_procedure

| | |
|----------------------|----------------------|
| Event trigger | computed tomography |
| Event type | Diagnostic_procedure |
| Biological structure | chest |

医療分野のイベント抽出における問題と提案の概要

1. 項を考慮したイベント抽出のコーパスがない
→ イベント抽出用コーパスの構築
2. 一般分野のイベントよりもトリガやエンティティの言及 (mention) が長く，項の種類も多い
→ 言及を重視した項ごとのモデル

Sign_symptom

A man presented with an abnormal nodule measuring 0.8 x 1.5 cm in the left upper lung lobe imaged through chest computed tomography scanning.

| | |
|----------------------|----------------------|
| Event trigger | nodule |
| Event type | Sign_symptom |
| Detailed description | abnormal |
| Area | 0.8 x 1.5 cm |
| Biological structure | left upper lung lobe |

Diagnostic_procedure

| | |
|----------------------|----------------------|
| Event trigger | computed tomography |
| Event type | Diagnostic_procedure |
| Biological structure | chest |

イベント抽出用コーパスの構築

医療分野のイベント抽出
コーパスの作成

- 高いアノテーションコスト
- プライバシーの問題

➔ 既存のタグ付きコーパス
MACCROBAT [1] を
イベント抽出用に再構成

| | MACCROBAT | MACCROBAT-EE |
|--|-----------|--------------|
| Entities | ✓ | ✓ |
| Event triggers | ✓ | ✓ |
| Event types | ✓ | ✓ |
| Arguments | ✗ | ✓ |
| Argument roles | ✗ | ✓ |
| Definition sentences of event types and argument roles | ✓ | ✓ |

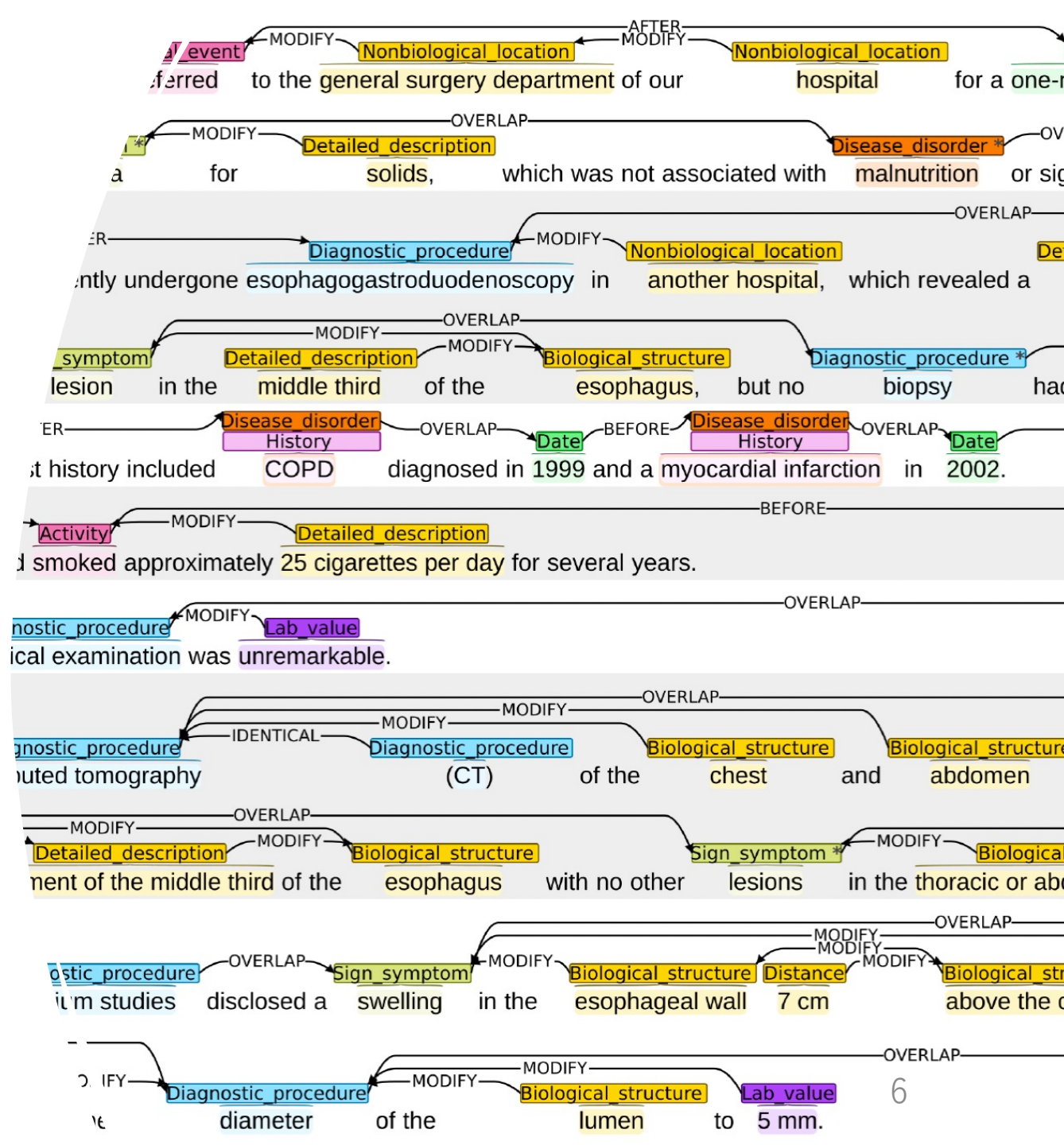
[1] Caufield et al., 2019. A comprehensive typing system for information extraction from clinical narratives. medRxiv.

MACCROBAT コーパス [1]

- MACCROBATの名前の由来
 - MACCR (Metadata Acquired from Clinical Case Reports) にACROBAT (Annotation for Case Reports using Open Biomedical Annotation Terms) というスキームで、タグづけ
- 200例の症例報告に**イベント (トリガ)** と関連する**エンティティ**, その間の**関係**, イベント間の**時間関係**を6人のアナテータでタグづけ

[1] Caufield et al., 2019. A comprehensive typing system for information extraction from clinical narratives. medRxiv.

2023/9/28



MACCROBAT-EEコーパス

- MACCROBATコーパスの**イベント**と**エンティティ**，その間の**Modify** **関係**を取り出し，述語（トリガ）と項の形に変換
 - Modify: あるエンティティ/イベントが別のエンティティ/イベントを変更
 - MAACROBATのエンティティのタイプを項のタイプとして利用
- サンプルしたイベントについて，2人の評価者が**100個中96個を** **包括的で適切である**と判断

| Event Type | Role |
|-----------------------|--|
| Sign_symptom | Biological_structure, Detailed_description, Severity, Lab_value, Distance, Shape, Area, Color, Texture, Frequency, Volume, Quantitative_concept, Qualitative_concept, Biological_attribute, Subject, Other_entity, History, Mass |
| Diagnostic_procedure | Lab_value, Biological_structure, Detailed_description, Qualitative_concept, Nonbiological_location, Frequency, Distance, Subject, Shape, Quantitative_concept, Texture, Severity, Age, Color, Area, Volume, Administration, Mass |
| Therapeutic_procedure | Detailed_description, Biological_structure, Lab_value, Dosage, Nonbiological_location, Frequency, Distance, Qualitative_concept, Subject, Quantitative_concept, Area, Administration, Other_entity |
| Disease_disorder | Detailed_description, Biological_structure, Severity, Lab_value, Quantitative_concept, Distance, Nonbiological_location, Shape, Volume, Qualitative_concept, Area, Subject, Biological_attribute |
| Medication | Dosage, Administration, Detailed_description, Frequency, Lab_value, Nonbiological_location, Quantitative_concept, Biological_structure, Volume |
| Clinical_event | Nonbiological_location, Detailed_description, Frequency, Biological_structure, Subject, Lab_value, Quantitative_concept, Volume |
| Lab_value | Biological_structure, Detailed_description, Color, Severity, Frequency |
| Activity | Detailed_description, Nonbiological_location, Biological_structure, Other_entity, Frequency, Lab_value, Quantitative_concept |
| Other_event | Biological_structure, Quantitative_concept, Nonbiological_location, Severity, Detailed_description |
| Outcome | Nonbiological_location, Subject, Detailed_description, Age |
| Date | - |
| Time | - |
| Duration | - |

既存のイベント抽出用コーパスとの比較

- 項の種類が多い
- 文あたりのイベントが多い
- イベント辺りの項の数は少ない
- 言及が長い

| Metric | ACE05 | ERE | MACCROBAT-EE |
|----------------------------------|--------|--------|--------------|
| Unique event types | 33 | 38 | 13 |
| Unique argument roles | 22 | 21 | 22 |
| Unique arg. roles per event type | 4.73 | 2.87 | 10 |
| Documents # | 599 | 459 | 200 |
| Sentences # | 20,862 | 17,114 | 4,539 |
| Entities # | 54,820 | 46,185 | 23,898 |
| Trigger mentions # | 5,348 | 7,287 | 13,128 |
| Argument mentions # | 8,102 | 10,479 | 8,599 |
| Avg entities # per sentence | 3.18 | 3.20 | 5.43 |
| Avg events # per sentence | 1.34 | 1.47 | 3.21 |
| Avg args # per sentence | 2.39 | 2.24 | 2.67 |
| Avg args per event # | 1.48 | 1.42 | 0.81 |
| Avg entity word count | 1.12 | 1.10 | 1.89 |
| Avg trigger word count | 1.05 | 1.06 | 1.61 |
| Avg argument word count | 1.14 | 1.14 | 1.72 |

医療分野のイベント抽出における問題と提案の概要

1. 項を考慮したイベント抽出のコーパスがない
→ イベント抽出用コーパスの構築
2. 一般分野のイベントよりもトリガやエンティティの言及 (mention) が長く， 項の種類も多い
→ 言及を重視した項ごとのモデル

Sign_symptom

A man presented with an abnormal nodule measuring 0.8 x 1.5 cm in the left upper lung lobe imaged through chest computed tomography scanning.

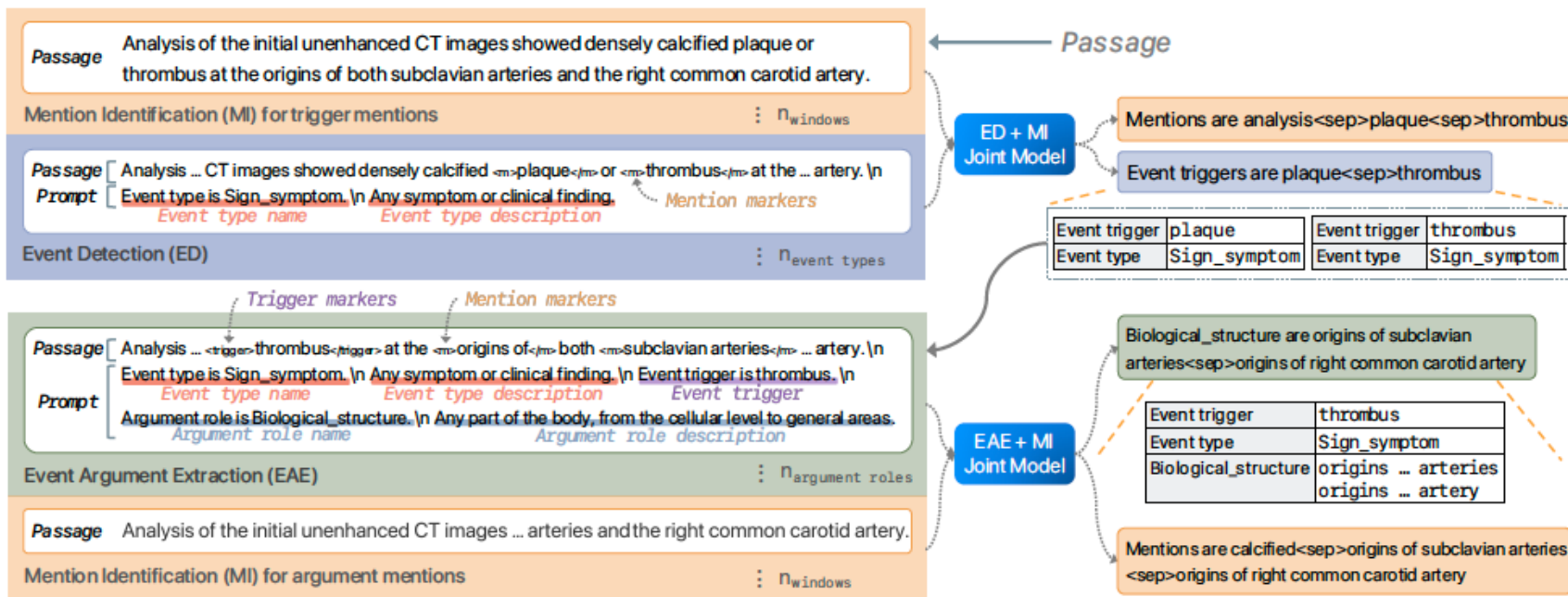
| | |
|----------------------|----------------------|
| Event trigger | nodule |
| Event type | Sign_symptom |
| Detailed description | abnormal |
| Area | 0.8 x 1.5 cm |
| Biological structure | left upper lung lobe |

Diagnostic_procedure

| | |
|----------------------|----------------------|
| Event trigger | computed tomography |
| Event type | Diagnostic_procedure |
| Biological structure | chest |

イベント抽出手法DICE

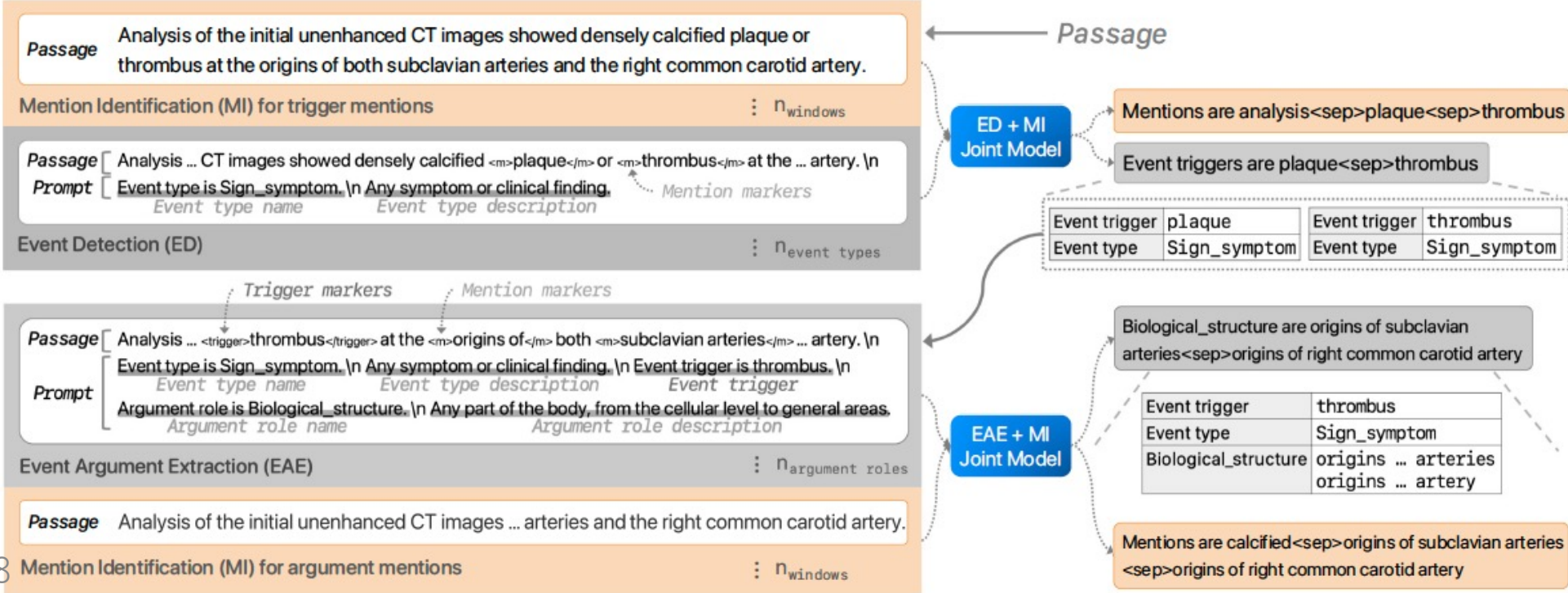
- 著者らのDEGREE (Data-Efficient GeneRation-based Event Extraction) [2] を拡張した**Sequence-to-sequence (S2S)** モデル
- **言及抽出で強化したイベント抽出**
 - 言及抽出, トリガ抽出, イベント項抽出の3つのモジュール



[2] Hsu et al., 2022. DEGREE: A Data-Efficient Generation-Based Event Extraction Model. NAACL-HLT 2022.

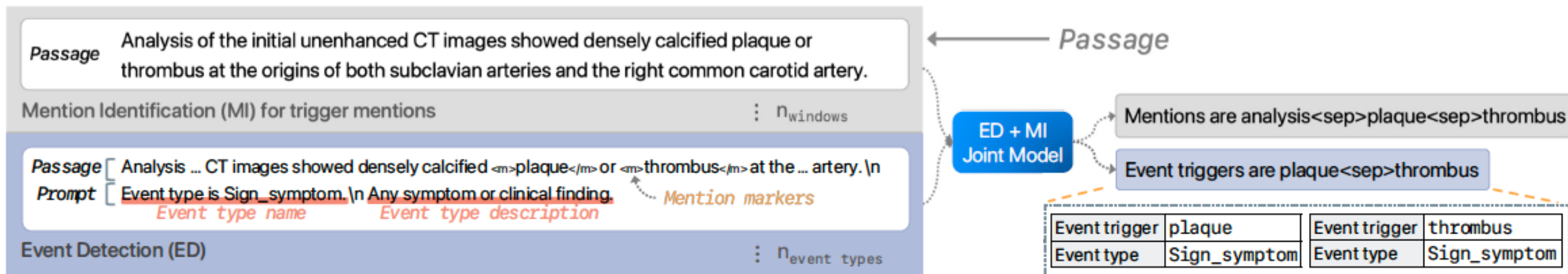
言及抽出 (Mention Identification; MI)

- **トリガ・項の候補**を別々に抽出 → 結果をそれぞれの抽出に利用
 - 抽出対象である Passage全体からの生成とスライディングウィンドウによる Passageの一部からの生成を組み合わせ、カバレッジを改善
 - 実験では10語のウィンドウを4語ずつスライド



トリガ抽出 (Event Detection; ED)

- イベントの種類ごとに**トリガ**を抽出
 - 言及抽出の結果をマーカとして追加
 - イベント名とその説明をプロンプトとして利用
- **言及抽出を補助タスク (auxiliary task)** としてjointで学習




イベント項抽出 (Event Argument Extraction; EAE)

- 特定のトリガについて項の種類ごとに**項**を抽出
 - 言及抽出・トリガ抽出の結果をマーカとして追加
 - トリガ抽出のプロンプトに，項の名前とその説明を加えたものをプロンプトとして利用
- **言及抽出を補助タスク**としてjointで学習




学習上の工夫: 境界の摂動による 対照学習

- 通常の変分エントロピー損失に加え, ランダムに境界を変えた k 個 (実験では10個) の負例によるInfoNCE損失を利用



Positive target

Negative targets



Mentions are calcified<sep>origins of subclavian arteries

Mentions are densely calcified<sep>origins of subclavian arteries

Mentions are calcified<sep>~~origins~~ of subclavian arteries

Mentions are calcified<sep>origins of subclavian ~~arteries~~

Mentions are calcified plaque<sep>~~origins of~~ subclavian arteries and

$$\mathcal{L}_N = \frac{1}{|\mathcal{X}|} \sum_{\mathbf{x}_i \in \mathcal{X}} \left[\log \frac{f(\text{out}_i, \text{in}_i)}{\sum_{(n_i^j, \text{in}_i) \in \mathcal{X}_i} f(n_i^j, \text{in}_i)} \right]$$

InfoNCE loss

学習設定

- Scheduled sampling
 - 後半のモジュール (ED, EAE) は一定の確率で正解を利用.
 - 確率はエポック毎に線形に減らし, 前のモジュールの結果を利用するように.
- ビーム探索 (ビーム幅=2)
- オプティマイザはAdamW

| Hyperparameter | Search Range | Best |
|--|------------------------------|-------------|
| Negative instance # for ED | 1, 2, 3, 4, 5, 8, 10, all | 10 |
| Negative instance # for EAE | 1, 2, 3, 4, 5, 8, 10, all | 10 |
| MI module sliding window size | 4, 6, 8, 10, 12 | 10 |
| MI module sliding window step | 2, 4, 6, 8, 10 | 4 |
| MI module sliding window retains original long sequence during training | True, False | True |
| MI module sliding window retains original long sequence during inference | True, False | False |
| Batch size | 1, 2, 3, 4 | 4 |
| Learning rate | 1e-4, 5e-5, 1e-5, 5e-6, 1e-6 | 1e-5 |
| Decoding method | beam search, greedy | beam search |
| Max epochs | | 70 |

結果

MACCROBAT-EE

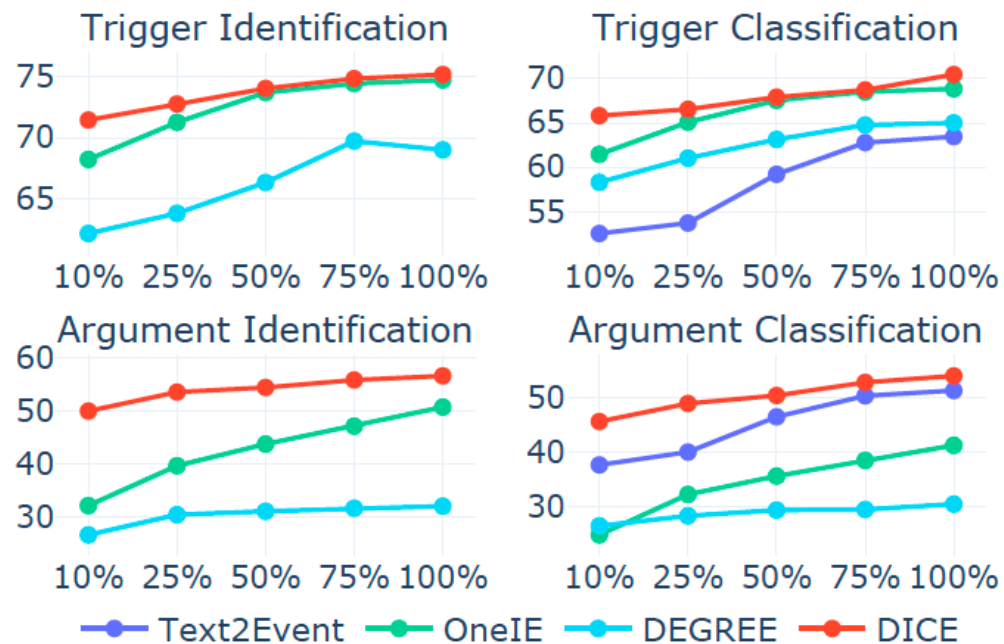
| # Model | Trigger | | | | | | Argument | | | | | |
|----------------|----------------|--------------|--------------|----------------|--------------|--------------|----------------|--------------|--------------|----------------|--------------|--------------|
| | Identification | | | Classification | | | Identification | | | Classification | | |
| | Prec. | Recall | F1 | Prec. | Recall | F1 | Prec. | Recall | F1 | Prec. | Recall | F1 |
| 1 Text2Event | – | – | – | 66.64 | 60.57 | 63.46 | – | – | – | 55.29 | 47.89 | 51.33 |
| 2 OneIE | 74.60 | 74.93 | 74.77 | 68.74 | 68.96 | 68.85 | 48.99 | 52.59 | 50.72 | 39.82 | 42.95 | 41.32 |
| 3 DEGREE | 71.91 | 66.33 | 69.01 | 67.59 | 62.59 | 65.00 | 46.84 | 24.31 | 32.02 | 44.75 | 23.23 | 30.58 |
| 4 Vanilla DICE | 65.03 | 74.08 | 69.26 | 60.51 | 70.28 | 65.03 | 49.10 | 53.60 | 51.25 | 45.95 | 50.76 | 48.24 |
| 5 DICE | 73.53 | 76.98 | 75.22 | 68.12 | 72.97 | 70.46 | 55.41 | 57.87 | 56.61 | 53.02 | 55.03 | 54.01 |

ACE05 (一般ドメイン)

| Model | 100% | |
|------------|-------------------|-------------------|
| | Tri-C | Arg-C |
| Text2Event | 71.9 [†] | 53.8 [†] |
| OneIE | 74.7 [†] | 56.8 [†] |
| DEGREE | 72.2 [†] | 55.8 [†] |
| DICE | 75.5 | 57.6 |

- トリガー・項について，スパン同定・分類で評価
- 既存のSOTAよりも（一般分野も含めて）高性能
 - Text2Event (T5 large), OneIE (BERT large), DEGREE (BART large), DICE (T5 large)

学習データを減らした時の影響



| Model | 10% | |
|------------|-------------------------|-------------------------|
| | Tri-C | Arg-C |
| Text2Event | 47.0 [‡] | 24.9 [‡] |
| OneIE | 61.5 [‡] | 26.8 [‡] |
| DEGREE | <u>66.1[†]</u> | <u>42.1[†]</u> |
| DICE | 68.9 | 44.7 |

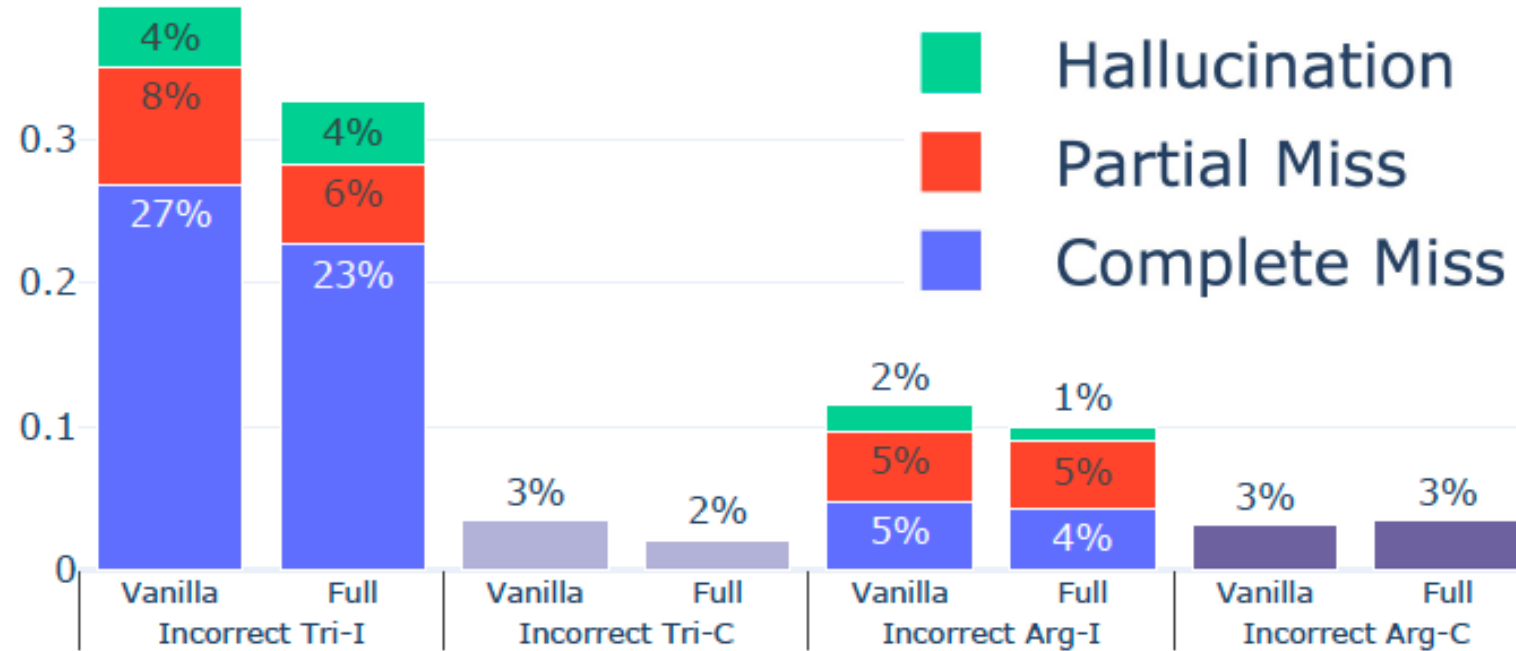
- データが少ないときでもDICEは頑健に動作
 - DICEのDIは**D**ata-**E**fficient

言及についての工夫の効果

| # Mention-enhancing techniques | Trigger | | | | | | Argument | | | | | |
|--|----------------|--------|--------------|----------------|--------|--------------|----------------|--------|--------------|----------------|--------|--------------|
| | Identification | | | Classification | | | Identification | | | Classification | | |
| | Prec. | Recall | F1 | Prec. | Recall | F1 | Prec. | Recall | F1 | Prec. | Recall | F1 |
| 1 Vanilla DICE | 65.03 | 74.08 | 69.26 | 60.51 | 70.28 | 65.03 | 70.76 | 76.48 | 73.51 | 66.47 | 72.71 | 69.45 |
| 2 Vanilla w/ aux. task | 69.54 | 74.59 | 71.98 | 65.02 | 71.00 | 67.88 | 73.24 | 76.48 | 74.83 | 68.31 | 73.03 | 70.59 |
| 3 Vanilla w/ marker | 72.91 | 70.71 | 71.79 | 68.58 | 67.70 | 68.14 | 74.27 | 76.91 | 75.57 | 69.66 | 72.82 | 71.20 |
| 4 Vanilla w/ contrastive | 70.02 | 75.12 | 72.48 | 66.93 | 72.04 | 69.39 | 73.86 | 77.41 | 75.59 | 69.92 | 72.89 | 71.37 |
| 5 Vanilla w/ all three (Full DICE) | 73.53 | 76.98 | 75.22 | 68.12 | 72.97 | 70.46 | 75.73 | 77.62 | 76.66 | 71.14 | 73.91 | 72.50 |
| 6 Vanilla w/ perfect marker [†] | 97.04 | 94.11 | 95.55 | 85.23 | 88.66 | 86.91 | 91.91 | 90.72 | 91.31 | 81.71 | 86.73 | 84.14 |

- 他の影響を省くため，項については正解のトリガを利用
- 言及抽出を補助タスクとしたjointでの学習 (2)，マーカの追加 (3)，対照学習 (4) いずれも有効.

エラー分析



- Complete miss (正解との重複なし), Partial Miss (正解の部分列), Hallucination (正解と重複しているが余計な範囲を出力)
- 言及の工夫によるエラー削減ができています

エラーの例

-
- 1 **Task:** ED **Passage:** An {audiology evaluation} showed {severe} {bilateral} {sensorineural} {high-frequency} {hearing loss} ({-70 dB}).
Ground-truth: (bilateral sensorineural high-frequency hearing loss, Sign_symptom) **Pred. of DICE:** (hearing loss, Sign_symptom)
-
- 2 **Task:** EAE **Passage:** The patient underwent a {resection} of the { 15 cm segment IVb } mass [SIGN_SYMPTOM] in {June 2010} .
Ground-truth: (15 cm, Distance), (segment IVb, Biological_structure) **Pred. of DICE:** (15 cm segment IVb, Biological_structure)
-
- 3 **Task:** ED **Passage:** Core biopsies from the {breast lump} showed {ductal carcinoma} in situ (sample labelled P1.1).
Ground-truth: (biopsies, Diagnostic_procedure), (ductal carcinoma, Disease_disorder) **Pred. of DICE:** None
-
- 4 **Task:** EAE **Passage:** Serum total bilirubin and {tumor markers} , carcinoembryonic antigen [DIAGNOSTIC_PROCEDURE] ({CEA}) and carbohydrate antigen 19-9 [DIAGNOSTIC_PROCEDURE] ({CA19-9}), were all {within normal ranges} .
Ground-truth: None **Pred. of DICE:** (within normal ranges, Lab_value) was predicted as the argument for both events.
-

- 言及抽出は正解とずれていることもある
 - 他のモジュールが分割したり，まとめたりして間違っているものが散見される

まとめ

- 医療分野のイベント抽出の提案
 - 既存コーパスから変換したイベント構造を持ったコーパス
 - 既存のS2Sによるモデルを拡張し，言及抽出を補助とした，項ごとの抽出を行うモデル
 - 一般分野でも高い性能を発揮
 - 様々な結果の解析