

反復的な抽出により抽出結果を精錬する生物医学イベント抽出

金兎一矢¹ 井上純大¹ 三輪誠^{1,2}

¹ 豊田工業大学大学院 ² 産業技術総合研究所人工知能研究センター
 {sd24416,sd24410,makoto-miwa}@toyota-ti.ac.jp

概要

生物医学イベント抽出では、イベントを引数を持つ階層構造などを含む複雑な構造を持ったイベントを対象とするため、複数のイベント間の関係を考慮する必要がある。本研究では、表形式によるイベント表現と離散拡散モデルを統合し、抽出したイベント情報を考慮して抽出結果を段階的に改善する反復的精錬手法を提案する。実験では、データセットに GENIA11 を用いて、性能を評価した。

1 はじめに

生物医学イベントはテキストに記述された遺伝子やタンパク質などのエンティティ間の相互作用や状態変化などの生物学的反応であり、文献からの生物医学イベントの抽出はゲノム解析や新薬の開発の補助などへの幅広い応用が期待される [1]。

生物医学イベントには3つの複雑な構造が存在する [2]。1つ目は、1つのトリガーに複数の引数に対応する1対多の関係の構造である。2つ目は同一のトリガーが複数の異なるイベントタイプに対応する重複構造である。例えば、図1では「過剰発現」というトリガーから *Gene_expression* タイプのイベントと *Positive_regulation* タイプのイベントが同時に対応する。3つ目はイベントが別のイベントの引数となる階層構造である。例えば、図1では「過剰発現」をトリガーとするイベントが「必要」をトリガーとするイベントの引数となる。これらの複雑な構造を持ったイベントを正確に抽出するためには、複数のイベント間の関係を考慮する必要がある。

生物医学イベント抽出における既存手法ではイベントを順に抽出する手法が主流である。例えば、Trieu ら [3] は深層学習モデルを用いてエンティティとトリガーを検出した後、再帰的にイベントを構築する。また、Wang ら [4] は質問応答の枠組みで「このイベントの引数は何か」など質問を繰り返し、イベントの要素をあらかじめ決められた順番で抽出

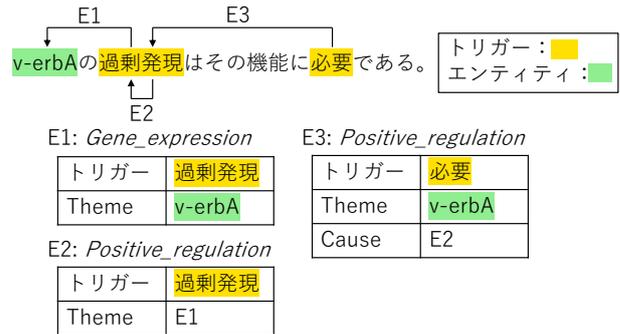


図1 生物医学イベントの例

し、イベントを構築する。これらの手法では、既に抽出された他のイベント情報を活用して、抽出の誤りを修正する機構を持たない。この結果、抽出の誤りが修正されないまま最終的なイベント構造が構築され、複数のイベント間で矛盾が生じる可能性がある。

本研究では、この問題を解決するため、抽出したイベント情報を活用して予測を段階的に改善するイベント抽出手法を提案する。具体的には、エンティティ・イベント間の関係を捉えることができる表形式 [5] で表現した抽出結果を、反復的な精錬能力を持ち画像生成などの分野で成功を収めている拡散モデル [6] を用いて反復的に精錬する。本手法では、各ステップにおける抽出結果を条件として、新たな抽出結果を「生成」する拡散モデルの逆生成過程を通して、抽出の誤りが修正されることを期待する。

本研究の貢献は次の通りである。

1. 拡散モデルの反復的精錬能力を利用し、表形式で全体を表現した抽出結果を段階的に改善しながらイベント抽出を行うフレームワークを提案した。
2. GENIA11 データセットにおける包括的な実験と分析を通じて、拡散モデルによる反復的精錬の効果が限定的であることを示した。

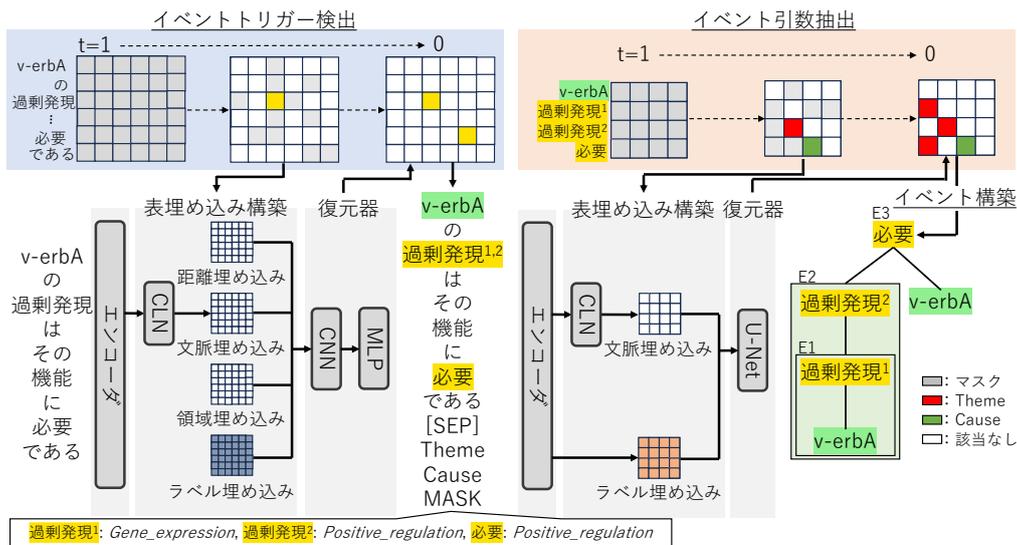


図2 提案手法の概要図。

2 関連研究

2.1 表形式による情報抽出

表形式による情報抽出では、単語対の関係を二次元の表形式で表現し、各セルにラベルを割り当てることで、複数のエンティティや関係を一度に表現し、抽出する [5]。この手法は、固有表現抽出 (NER) や関係抽出といったタスクにおいて発展し、近年では生物学イベント抽出にも適用されている [7]。

Li ら [8] は NER を単語対間の関係分類問題として定式化する W2NER を提案した。W2NER は、連続した通常のエンティティ・互いに重なり合って重複したエンティティ・離れた複数のスパンから構成される不連続なエンティティを表現する表形式を用いることで、文中に現れる全てのエンティティの一括抽出を可能としている。具体的には表を上三角領域と下三角領域に分割し、上三角にはエンティティ内部の連結を表す関係を、下三角には境界及びエンティティタイプを割り当てる。モデルは最初に入力文を BERT, BiLSTM により単語の文脈情報を得る。次に表の埋め込みを各単語対に対して条件付き層正規化により構築する。この埋め込みを強化するため、CLN により構築された文脈埋め込みに、単語間距離埋め込み、表内の上三角・下三角を区別する領域埋め込みを連結することで表埋め込みを得る。最後に構築した表埋め込みを CNN, MLP から構成される予測器に入力することで最終的な予測を得る。

2.2 生物学イベント抽出

Trieu ら [3] はエンティティ検出から階層構造構築までをエンドツーエンドで実行する DeepEventMine を提案した。DeepEventMine は、SciBERT[9] をベースとし、BERT 層、エンティティ・イベントトリガー検出層、関係抽出層、イベント構築層を持ち、各層を同時に訓練することで知識共有を実現する。イベント構築層では、検出されたトリガーと引数からイベント候補を列挙し、ボトムアップ方式でイベントを引数に持たないイベントから高次の階層的なイベントへと再帰的に構築し、イベントを抽出する。

3 提案手法

本研究では、文中のイベントを構成する情報を表形式で表現し、マスク吸収型の離散拡散モデルにより表全体を反復的に精練する手法を提案する。提案手法の概要図を図2に示す。本手法では、拡散モデルの逆過程において、全セルをマスクで初期化した表から開始し、複数ステップにわたって各セルのイベントタイプまたは引数役割を段階的に確定する。次ステップ以降は、各ステップで確定したセルを入力とし、残りのマスクされたセルを予測することで、抽出結果を反復的に精練する。

本手法は既存研究 [10] と同様、イベントをイベントトリガー検出とイベント引数抽出の2段階で抽出する。イベントトリガー検出ではイベントの発生を最も表す表現であるトリガーとイベントの種類を表すイベントタイプを特定し、イベント引数抽出はイ

イベントの要素である引数となるエンティティもしくはイベントとその役割を特定する。

3.1 イベントトリガー検出

本節では、入力文からトリガーとそのイベントタイプを段階的に抽出するイベントトリガー検出手法について述べる。提案手法は、W2NER [8] を基盤とし、表をマスクから反復的に復元する離散拡散モデルとして定式化することで、予測を反復的に更新しながら抽出結果の精錬を実現する。ここで用いるエンコーダおよび表の予測器は W2NER と同一のモデルを用い、本研究では予測器を拡散過程の逆過程における復元器として利用する。

W2NER は入力文を単語列として処理するが、本研究ではアノテーションがスパン単位で与えられる設定に対応するため入力文をトークン列として扱い、2.1 節で説明した W2NER の表形式と同様の手法で、表埋め込みを得る。また、本研究では拡散過程におけるセルの状態（マスク、ラベル、該当なし）に対応する学習可能な埋め込みからなるラベル埋め込みを新たに構築し、表埋め込みに連結する。これにより、各ステップで確定した情報を次ステップ以降の予測に利用できるようにする。

結合した表埋め込みを拡散モデルに入力し、逆過程を複数ステップ反復して表全体を復元する。重複構造に対応するため、各セルの分類をマルチラベル分類として扱い、同一のスパン（同一セル）が複数のイベントタイプを同時に持つ場合に対応する。

3.2 イベント引数抽出とイベントの構築

本節では、3.1 節で抽出したトリガーと正解のエンティティを用いてイベント引数抽出を行う方法、および最終的に階層的なイベント構造を構築する方法について述べる。

表を構築するために、入力文をエンコーダ (BERT + BiLSTM) に入力し、各トークンの文脈表現を得る。次に、3.1 節で得られたトリガーと文中のエンティティを統合し、文中の出現順に整列した要素列を作成する。この要素（トリガー、エンティティ）を表の行・列の単位として扱う。各要素について、対応するスパンの文脈表現を、Trieu ら [3] に倣い、集約し、先頭トークン・末尾トークン・スパン内平均の表現にタイプ埋め込みを付加して要素埋め込みを構築する（トリガーはイベントタイプ、エンティティはエンティティタイプを用いる）。この要素埋

め込みを W2NER の CLN に入力することで $N \times N$ の表埋め込みを構築する (N は要素数)。表の各セルは、要素間の引数役割（例：Theme, Cause, AtLoc など）または「該当なし」を表す。

さらに、引数役割の意味情報をモデルに明示的に与えるため、役割名の自然言語表現（例：AtLoc → “At Location”）から役割埋め込みを構築し、セル特徴に連結する。役割埋め込みは、入力文と役割名（自然言語表現）を [SEP] で連結して同じエンコーダに入力し、役割名に対応するトークン部分の出力を平均プーリングして得る。

CLN により構築した表埋め込みと前段で構築した役割ラベル埋め込みを連結し、拡散モデルに入力し、3.1 節と同様に、逆過程を複数ステップ反復して表全体を復元する。復元器には表全体の大域的な文脈を考慮した予測を行うため、U-Net を利用する。

イベント引数抽出により得られた表から最終的な階層的イベント構造を構築する。各トリガー t に対し、表から役割付き引数の集合 $\{(r_1, a_1), (r_2, a_2), \dots, (r_m, a_m)\}$ を抽出する。 t はトリガー、 r は引数役割、 a は引数を表す。

同一トリガーに対して同一役割を持つ複数の引数が存在する場合、提案手法では各役割に対する引数候補の数が等しいときは同じ位置にある引数同士を組み合わせる。例えば、トリガー t に対して Theme 役割を持つ引数が a_1, a_2 の 2 つ、Cause 役割を持つ引数が c_1, c_2 の 2 つ存在する場合、1 番目同士を組み合わせた $(t, \text{Theme} : a_1, \text{Cause} : c_1)$ と、2 番目同士を組み合わせた $(t, \text{Theme} : a_2, \text{Cause} : c_2)$ の 2 つのイベントを生成する。役割ごとの引数候補数が異なる場合は、すべての組み合わせを列挙する。

引数がエンティティのみのイベントから構築し、引数がトリガーの場合はそのトリガーが構成するイベントを引数としてイベントを構築する。

3.3 拡散モデルの学習と推論

学習では、時刻 t を 0 から 1 の範囲でランダムにサンプリングし、時刻 t に対応するマスク（ノイズ）を正解の表 \mathbf{x}_0 に適用し、時刻 t の表 \mathbf{x}_t から表 \mathbf{x}_0 を予測する。損失関数には、イベントトリガー検出はバイナリ交差エントロピー損失、イベント引数抽出には交差エントロピー損失を用いる。イベントトリガー検出の推論では、すべてのセルをマスクで初期化した表 \mathbf{x}_1 から開始し、逆拡散過程を反復的に実行

表1 イベントトリガー検出性能の比較(%). 各評価指標の最高スコアは太字で示す.

モデル	Steps	R	P	F1
DeepEventMine [3]	-	73.32	68.72	70.95
シングルラベル	-	70.81	70.93	70.87
提案手法	1	73.32	69.59	71.41
	2	69.18	72.20	70.66
	3	73.74	68.12	70.82
	4	69.60	71.06	70.32
	5	73.99	68.04	70.89
	10	69.56	71.04	70.29

表2 イベント抽出性能の比較(%). 各評価指標の最高スコアは太字で示す.

モデル	Steps	R	P	F1
DeepEventMine [3]	-	56.52	70.52	62.75
提案手法	1	57.59	58.01	57.80
	2	57.61	58.01	57.81
	3	57.63	58.00	57.82
	4	57.64	57.89	57.76
	5	57.49	57.78	57.64

して表 \tilde{x}_0 を生成する. イベント引数抽出の推論では各ステップでイベントのスキーマ制約を検証し, 制約違反が検出された場合は該当するトリガーの行をマスクする. 制約は (1) 引数を持つ行は Theme 役割が必須, (2) 番号付き役割 (例: Site2) は対応する ThemeN (例: Theme2) が必須, (3) イベントタイプごとに許容される役割の制限である.

4 実験

4.1 実験設定

生物医学イベント抽出の標準的なデータセットである GENIA Shared Task 2011 (GENIA11) [11] を用いる. このデータセットは, 複雑な階層構造を含むイベントを PubMed 抄録にアノテーションしたものである. GENIA11 は2つのエンティティタイプと9つのイベントタイプ, 6つの引数役割から構成される. イベント抽出は正解のエンティティとイベントトリガー検出で予測されたトリガーが与えられた状態で評価する. 実験は開発データを用いて行い, 評価指標は再現率, 適合率, F1 スコアを用いる. 提案手法は異なる乱数シードで5回実験を行い, その平均値を報告する. なお, GENIA11 には文を跨ぐイベントが付与されているが, 本手法では文単位でのイベント抽出に焦点を当てているため, 文単位のイベントについてのみ抽出を行った. そのため, 文を跨ぐイベントは偽陰性となり, 理論上の F1 値の最大値は 96.48% となる. イベントトリガー検出とイベント引

数抽出の基盤モデルには BioBERT [12] を採用する. 付録 C に利用したハイパーパラメタを示す.

4.2 結果

イベントトリガー検出の結果を表 1 に示す. 反復的精練なしのシングルラベル分類を行うベースラインモデルからマルチラベル分類へ拡張することでシングルラベルの設定から F1 スコアが 0.54% 上昇し, DeepEventMine と比較しても 0.46% 上昇した. しかし, ステップ数を増やし, 反復的精練を行なっても一貫した傾向は見られなかった.

イベント抽出の結果を表 2 に示す. 反復的精練を行わないベースラインは DeepEventMine と比較して F1 スコアが 4.95% 低下した. これはイベント引数抽出モデルの学習時に, イベントトリガー検出の偽陽性が混入する状況を想定した学習が十分ではなく, U-Net による畳み込みで誤った情報が伝播したことが考えられる. サンプルングステップを増やすことで, ステップ数 1 から 4 までは再現率が上昇し, F1 スコアはステップ数 3 が最も高い数値となったが効果は限定的であった. サンプルングステップ 1 と 3 を比較すると, GENIA11 の制約違反が平均 42 件から 37.4 件に減少した. 制約違反により評価から除外されるイベントが減少したことが, 再現率の向上に寄与したと考えられる.

イベントトリガー検出, イベント抽出において反復的精練の効果が限定的であった理由として, BERT から構成された表埋め込みが分類に十分な情報を持っていたことが挙げられる. 各ステップで予測されたラベル埋め込みを条件として次の予測を行うが, BERT の事前学習により獲得された文脈表現に対して, ラベル埋め込みによる追加情報は限定的であったと考えられる.

5 おわりに

本研究では, 抽出したイベント情報を活用して予測を段階的に改善するイベント抽出手法の実現を目指し, 表形式で表現した抽出結果を反復的な精練能力を持つ拡散モデルを用いて反復的に精練する手法を提案した.

実験結果から提案手法の反復的精練による効果が限定的であることがわかった. 今後の課題としては, 抽出した結果の活用方法や学習と推論の乖離を解消する学習手法などが挙げられる.

謝辞

この成果の一部は、NEDO（国立研究開発法人新エネルギー・産業技術総合開発機構）の委託業務（JPNP25006）の結果得られたものです。また、産総研及び AIST Solutions が提供する ABCI 3.0 を「ABCI 3.0 開発加速利用」の支援を受けて利用しました。

参考文献

- [1] Giacomo Frisoni, Gianluca Moro, and Antonella Carbonaro. A survey on event extraction for natural language understanding: Riding the biomedical literature wave. **IEEE Access**, Vol. 9, pp. 160721–160757, 2021.
- [2] Ioannis Korkontzelos Sophia Ananiadou Makoto Miwa, Paul Thompson. Comparable study of event extraction in newswire and biomedical domains. In **COLING**, 2014.
- [3] Khoa N. A Duong Anh Nguyen Makoto Miwa Hai-Long Trieu¹, Thy Thy Tran and Sophia Ananiadou. Deep-eventmine: end-to-end neural nested event extraction from biomedical texts. In **Bioinformatics**, 2020.
- [4] U.Leser X.Wang, L.Weber. Biomedical event extraction as multi-turn question answering. In **ACL**, 2020.
- [5] Makoto Miwa and Yutaka Sasaki. Modeling joint entity and relation extraction with table representation. In **In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, 2014.
- [6] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In **Advances in Neural Information Processing Systems**, Vol. 33, pp. 6840–6851, 2020.
- [7] Liangyu Gao, Zhihao Yang, Jinzhong Ning, Kai Ma, Lei Wang, Wei Liu, Yin Zhang, Ling Luo, Bo Xu, Jian Wang, Yumeng Yang, Zhehuan Zhao, Yuanyuan Sun, and Hongfei Lin. Biomedical event extraction as semantic segmentation. In **2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)**, pp. 3218–3221, 2024.
- [8] Jingye Li, Hao Fei, Jiang Liu, Shengqiong Wu, Meishan Zhang, Chong Teng, Donghong Ji, and Fei Li. Unified named entity recognition as word-word relation classification. **Proceedings of the AACL Conference on Artificial Intelligence**, Vol. 36, No. 10, pp. 10965–10973, Jun. 2022.
- [9] Iz Beltagy, Kyle Lo, and Arman Cohan. SciBERT: A pretrained language model for scientific text. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, pp. 3615–3620, Hong Kong, China, November 2019. Association for Computational Linguistics.
- [10] I-Hung Hsu, Kuan-Hao Huang, Elizabeth Boschee, Scott Miller, Premkumar Natarajan, Kai-Wei Chang, and Nanyun Peng. Degree: A data-efficient generation-based event extraction model. In **ACL**, 2022.
- [11] Jin-Dong Kim, Yue Wang, Toshihisa Takagi, and Akinori Yonezawa. Overview of Genia event task in BioNLP shared task 2011. In Jun’ichi Tsujii, Jin-Dong Kim, and Sampo Pyysalo, editors, **Proceedings of BioNLP Shared Task 2011 Workshop**, pp. 7–15, Portland, Oregon, USA, June 2011. Association for Computational Linguistics.
- [12] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. Biobert: a pre-trained biomedical language representation model for biomedical text mining. **Bioinformatics**, Vol. 36, No. 4, pp. 1234–1240, 2020.

A 役割ラベルの自然言語マッピング

本節では、イベント引数抽出において用いる役割ラベルと自然言語テキストの対応を示す。GENIA データセットにおける引数役割ラベルは、以下の表 3 のように自然言語テキストに変換される。

表 3 GENIA データセットにおける役割ラベルと自然言語テキストの対応

役割ラベル	自然言語テキスト
Theme	Theme
Theme2	Second Theme
Theme3	Third Theme
Theme4	Fourth Theme
Cause	Cause
Site	Site
Site2	Second Site
CSite	Cause Site
AtLoc	At Location
ToLoc	To Location
MASK	MASK

B データセットの統計

GENIA11 データセットの詳細な統計情報を表 4 に示す。

表 4 イベント抽出データセットの統計情報。

統計量	GENIA11		
	Train	Dev	Test
文書数	908	259	347
文数	8,691	2,900	3,371
エンティティ数	12,105	4,871	5,301
トリガー数	7,882	2,388	-
関係数	11,142	3,698	-
イベント数	10,310	3,250	-
エンティティタイプ数	2	2	1
イベントタイプ数	9	9	-
引数役割数	6	6	-

C ハイパーパラメータ

ハイパーパラメータの探索範囲を表 5 に示す。

表 5 ハイパーパラメータの探索範囲。

ハイパーパラメータ	探索範囲
BERT の学習率	{1e-5, 3e-5, 5e-5, 1e-4}
学習率	{1e-5, 3e-5, 5e-5, 1e-4}
バッチサイズ	{8, 16, 32}
warmup ステップ数	{0, 50, 100}
weight_decay	{0, 1e-5, 1e-4}