

大規模言語モデルを用いた出来事ネットワークの構築

澤畑 尚希[†] 町澤 大輔[†] 池尻 良平[‡] 澄川 靖信[†]
拓殖大学[†] 広島大学[‡]

1. はじめに

「愚者は体験によって学ぶという。私は他人の経験によって利益を得ることを好む」という一文が示すように、現代の問題によく似た過去の因果関係を分析することは現代の問題解決の土台となりうるということが知られている。このような分析を行うためには、過去の出来事から因果関係を適切に抽出して記録することが重要である。実際、新聞記事で報道された出来事から因果関係を抽出してタイムライン[1]や木構造[2]、ネットワーク[3]を生成することで構造化する研究がこれまでに行われている。しかし、このような先行研究は解析できる文章が長文であることを仮定しているが、必ずしも全ての出来事が長文で記述されているとは限らない。長文を仮定した因果関係を生成する手法[4]を短文の出来事に対して適用したところ、F値が約12%しか得られなかった。この結果から、短文でも適切に因果関係を構築できる手法を提案することが重要であると考えられる。

本研究では、短文で書かれた出来事でも因果関係を自動的に判定できるようにすることを目指し、WikipediaのPortal:Current eventsから抽出した短文の出来事で構成したネットワークを含むデータセットの構築と出来事ネットワークを自動で構築する手法を提案する。短文では文章内の情報だけで出来事間の関係性を解析するのは難しいので、本手法は大規模言語モデル（以降、LLMと呼ぶ）による推論を用いる。

2. データセット作成

本研究では、出来事ネットワークの教師データをWikipediaの2016年と2017年の出来事として記録されている5,204件から作成した。

データセットで表現したいことは、連鎖的に発生している出来事を並べたものであるタイムライン間における収束や分岐である。例えば、

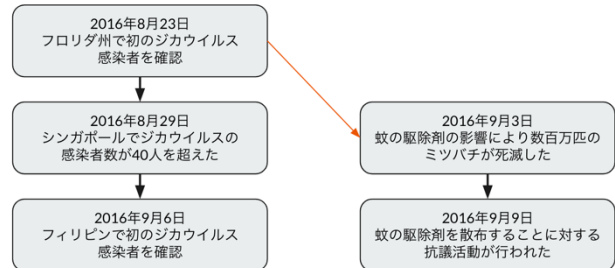


図1 ジカウイルスに関するタイムライン

図1の左側に示したジカウイルスの感染拡大のタイムラインには、「フロリダ州で初のジカウイルス感染者を確認」、「シンガポールでジカウイルスの感染者数が40人を超えた」、「フィリピンで初のジカウイルス感染者を確認」の3つの出来事が並んでいる。一方で右側のタイムラインには、「蚊の駆除剤の影響により数百万匹のミツバチが死滅した」、「蚊の駆除剤を散布することに対する抗議活動が行われた」の2つが存在する。このとき、2つのタイムラインは、ジカウイルスは蚊が媒介するという因果関係によって、結びつける必要があるといえる。これらの構造を表現するためには、1つの出来事をノードとした、ネットワーク構造で表す必要がある。

データセット作成は、タイムライン作成とタイムライン接続の2段階で実施した。タイムライン作成では、日付順に出来事を取り出し、既に取り出してある出来事と新しい出来事の間に関係性が定義できるかどうかを確認し、関係性を見出せたときにタイムラインとして日付順で並べた。戦争や政治のような複数の国や地域が関係する出来事のように、タイムラインとしては別のグループに属していても関係性が見出せる出来事同士が確認できたとき、それらの出来事同士に辺を定義し、それらの出来事を含むタイムライン同士を接続した。

データセットは2名の作業者がそれぞれ2016年と2017年の1年ずつ作成し、3人目の作業者が全ての結果の妥当性を検証した。このとき、意見の相異があれば2名が議論し、両者の意見が一致するようにタイムラインとネットワークを修

The Event Evolutionary Network for Tracking Causal Relationships using LLM

[†] Naoki Sawahata, Takushoku University

[†] Daisuke Machizawa, Takushoku University

[‡] Ryohei Ikejiri, Hiroshima University

[†] Yasunobu Sumikawa, Takushoku University

正した。この修正結果の妥当性を評価する評価者を4人目の作業員として上記の3人とは別のの人に依頼し、全作業員の意見が一致したものだけをデータセットに含めた。その結果、タイムライン作成では、2,879件のタイムラインを定義し、186件のタイムラインを接続した。

3. ネットワークの自動生成

本研究では、LLMを用いて出来事ネットワークを構築する。ここで使用する出来事は、短い文書で記述されている。

LLMは、入力となる文章を与えてその文章に対する答えを得るものである。今回は、1回の処理で2つの出来事を与え、これらに因果関係があるかどうかを答えさせる。もし、因果関係があると判定された場合、この2つの出来事間に辺を張る。これを出来事の組み合わせの数だけ繰り返すことによって、出来事をノードとしたネットワークの作成を行う。

全ての組み合わせを試すことは不要な時間とコストを使用するため、本手法はLLMに処理を行わせる出来事の組み合わせを減らすために以下の前処理を行った。まず、2つの出来事をOpenAI社がAPIとして提供している「text-embedding-3-large」を用いて特徴ベクトルに変換し、それらのコサイン類似度を計算した。この値が閾値を超えたものだけを使用し、類似度が大きい順に並び替えた。

4. 実験

本手法の有効性を評価するために、第2章のデータセットを用いた。本稿で使用したプロンプトエンジニアリング手法は以下の通りである。

Few-Shot prompting (FS)

FSは、プロンプト内に具体的な入力文に対する適切な解答例を1つ以上明示することによって、解答形式や行わせたい処理を明確化する手法である。この例示を用いないものをZero-Shot (ZS)として記載する。

Chain-of-Thought prompting (CoT)

CoTは、直接答えを出力させるのではなく、与えられた入力を考察させたり、推論途中の流れを前置きとして記述させたりすることによって、回答の精度を向上させる手法である。

上記の手法をそれぞれ使用した場合としなかった場合で、4種類のプロンプトを作成した。モデルは、OpenAI社が提供するGPT-4oを使用した。

表1 実験結果

	適合率	再現率	F値
EEG	29.8%	26.2%	27.9%
ZS	31.0%	46.9%	37.4%
FS	35.2%	35.0%	35.1%
ZS-CoT	29.1%	43.7%	34.9%
FS-CoT	36.1%	36.2%	36.1%

ベースラインとして、Yangらが提案した手法であるEEG [3]を元に本データセットに合わせて変更を行ったものを使用した。

評価手法は、先行研究[3]と同様に、手動で作成したネットワークと自動的に作成されたネットワークの辺の張り方の違いを評価した。すなわち、各手法が、データセットに存在する出来事間の辺を同じように張ることが出来ていれば正解とした。指標としては、適合率 (Precision) と再現率 (Recall)、そしてこれらの調和平均であるF値の3つを用いた。

表1にすべての手法の結果を示す。この結果から、適合率、再現率、F値のすべてにおいて、LLMを用いた手法がベースラインを上回っていることがわかる。LLMによる手法同士を比較すると、ZS-CoTよりZSの精度が高く、ZSからFSにすることで適合率が上がり再現率が下がることが確認できる。

5. まとめ

本稿では、WikipediaのPortal:Current eventsから収集した短文出来事を用いて、データセットの構築とネットワーク構造を生成する手法を提案した。本稿の実験では、既存の長文を仮定した手法を短文で書かれた出来事間の因果関係抽出に使用することが難しいこと、LLMが一定の文脈を補完できることを示した。本稿で作成したデータセットは2016年と2017年の出来事のみを含むので、今後の課題としては、データセットを拡張することが考えられる。

参考文献

- [1]: D. Gholipour Ghalandari and G. Ifrim, "Examining the state-of-the-art in news timeline summarization," ACL'20, pp. 1322–1334, 2020.
- [2]: C. Zhang, J. Lyu, and K. Xu, "A storytree-based model for inter-document causal relation extraction from news articles," Knowl. Inf. Syst., 65 (2), pp. 827–853, 2022.
- [3]: C. C. Yang, X. Shi and C. -P. Wei, "Discovering Event Evolution Graphs From News Corpora," IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 39 (4), pp. 850–863, 2009.
- [4]: K. Radinsky and E. Horvitz, "Mining the web to predict future events," WSDM'13, pp. 255–264, 2013.