大規模ニュースを対象とした政治イベント情報抽出における LLM 活用の試行と課題

Exploring the Use of LLMs for Political Event Extraction from Large-Scale News

朝倉卓人 Takuto Asakura　東海林拓人 Takuto Tohkairin　韓南琦 Han Namgi　阪本拓人 Takuto Sakamoto　宮尾祐介 Yusuke Miyao

東京大学
The University of Tokyo

大規模ニュースコーパスを対象とした政治イベント情報抽出は、複雑なイベント定義や属性設計を伴い、従来は高コストな人手アノテーションを要してきた。本研究では、ニュース記事群を対象に、PLOVER 系の政治イベントスキーマに基づく情報抽出を LLM で実行し、その実用可能性と課題を検討する。特に、イベント種別を先行抽出し、その結果に応じて属性抽出を分岐させる二段階ワークフローを採用した。実験の結果、イベント種別間および属性間で抽出成功率に大きな偏りが見られ、特に第 2 段階の属性抽出ではスキーマの複雑さに起因する高い失敗率が確認された。一方で、LLM の長文・表構造理解能力自体は高水準であることも示唆された。

1. はじめに

政治イベント情報抽出において、PLOVER [PLOVER 18, Halterman 23a] のように専門的に定義され、国際関係研究において事実上の標準となっているスキーマに合致するデータを大規模に整備することは、理論検証や政策分析の基盤として重要である [Olsen 24]。しかし同スキーマは、イベント種別の数が多く、それぞれに詳細な定義が与えられており、境界的事例の判定には政治学的知識と文脈理解を要する。また各イベントには多数のイベント属性が規定され、さらに付与すべき属性の種類はイベント種別ごとに異なるため、単純な抽出問題には還元できない構造的複雑性を有する。

従来、このような仕様を満たすデータ構築には、政治学の専門的素養を持ち、かつアノテーションスキーマの詳細を理解した専門家による高コストの人手作業が不可欠であった [Olsen 24]。一方で近年の大規模言語モデル（LLM）は、各種ベンチマークにおいて博士課程学生相当以上の専門知識を示すことが報告されており [OpenAI 23, Chowdhery 23, Hendrycks 20]、政治学向けデータをより効率的かつ大規模に生成できる可能性が期待される。

しかし、LLM が PLOVER のような複雑な仕様を厳密に運用し、定義に忠実なデータを安定的に生成できるかは未検証である。本研究では、イベント種別の定義に基づく抽出と、抽出結果に対するイベント属性付与を二段階に分離する枠組みを採用し、ニュース記事からスキーマ仕様を満たすデータ構築を試みた。専門家による定性評価の結果、種別および属性ごとに妥当性のばらつきが確認されたものの、一部については実用に資する水準でのデータ生成が可能であることが示唆された。

なお、本研究で使用したプロンプトや抽出結果データは我々のプロジェクト・ウェブサイト [*1] で公開する。

2. 背景

政治学においては、20 世紀後半以降、ニュース記事を基に現実世界の出来事を構造化したイベントデータベースが構築され、仮説検証に広く用いられてきた [McClelland 78]。初期の WEIS や COPDAB [Azar 80] は人手によるコーディングに依拠し、その後も UCDP GED や ACLED など、専門家がニュースや複数情報源を参照して記録する手動データベースが発展している [Sundberg 13, Raleigh 10]。多くは Factiva[*2] や LexisNexis[*3] といった商用ニュースアグリゲータを情報源として利用する。一方、KEDS や TABARI に基づく ICEWS や GDELT などの機械生成型データベースも登場し、CAMEO オントロジーに基づく大規模抽出が行われてきた [Schrodt 94, Leetaru 13]。近年は CAMEO の後継として PLOVER が提案され、POLECAT データセットが NGEC モデルにより構築されている [Halterman 23a, Halterman 23b]。これらは自動化と精度の両立を目指すが、依然として手動コーディングとの質的差異が指摘されている [Olsen 24]。

なお既存の政治イベントデータベースは、紛争や抗議行動など特定の現象に特化したものが多いのに対し、PLOVER は多様な政治的相互作用を包括的に扱うことを目的としたオントロジーである。この包括性は、特定領域に依存しない汎用的な政治イベント抽出の検証という本研究の目的に適合している。

3. 問題設定と方法

本研究の目的は、PLOVER スキーマに合致する政治イベントデータを、ニュース記事から大規模言語モデルにより自動生成できるかを検証することである。実験対象として、Factiva に収録された 2024 年 1 月の記事 1,000 件を用いた。PLOVER で定義される 16 種類のイベントのうち、本稿では敵対的（Conflict）イベント 9 種類を抽出対象とした（詳細は [PLOVER 18] 参照）。

- REQUEST: 要求・命令・要請などの言語的行為。デモや集会など物理的行動を伴うものは含まれない。
- ACCUSE: 非難・告発・申し立て・訴訟・調査など。政策や行為に対する公的な批判や司法的措置を含む。
- REJECT: 提案や要求に対する拒否・拒絶の表明。関係の縮減を伴う場合は SANCTION に区別される。
- THREATEN: 深刻な帰結を伴う強い警告や威嚇。通常は言語的行為であり、実際の武力行使は含まない。
- PROTEST: 市民による集団的・公開的な抗議行動。物理

連絡先: 朝倉卓人，東京大学，takuto@is.s.u-tokyo.ac.jp

*1 https://mynlp.is.s.u-tokyo.ac.jp/digital_observatory/

*2 http://www.factiva.com

*3 https://www.lexisnexis.com

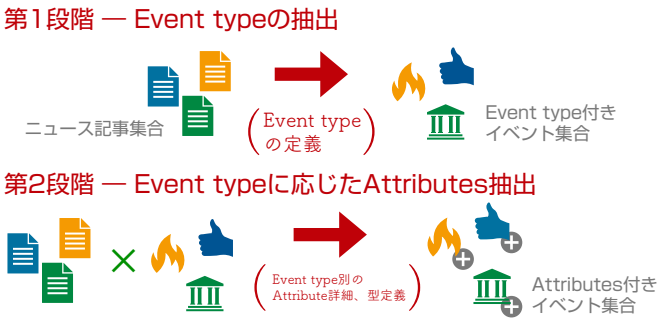

図 1　二段階の抽出ワークフロー

的集会やデモを含む。

- SANCTION: 既存の協力関係の縮減や制裁措置。形式的な経済制裁に限らず、関係の低減全般を含む。
- MOBILIZE: 実際の武力行使に至らない軍事・警察的動員。武力の示威や準備行動を指す。
- COERCE: 暴力に至らない強制的措置や権利制限。弾圧や抑圧的行為を含む。
- ASSAULT: 実際の物理的危害を伴う意図的行為。武器の有無を問わず身体的攻撃を含む。

各イベントについては、PLOVER において当該種別に定義されたすべてのイベント属性に加え、政治学の専門家から要望のあった若干数の独自属性も抽出対象とした。モデルには、全実験を通じて OpenAI の GPT-4o-mini を用いた。

スキーマ仕様を精査した結果、イベント種別が全体構造において中核的役割を果たすことが明らかになった。第一に、いずれかのイベント種別定義に該当するか否かが抽出対象の可否を決定する。第二に、抽出すべきイベント属性の種類は種別ごとに異なる。例えば、MOBILIZE や RETREAT のような主体の行為を中心とするイベントでは客体に関する属性は必須ではないが、AID や SANCTION では客体の明示が不可欠となる。このように、種別判定は後続の属性付与条件を規定する枠組みとなっている。

以上を踏まえ、本研究では二段階のワークフローを採用した（図 1）。第一段階では、ニュース記事本文と全イベント種別の定義をプロンプトに与え、記事内に定義に合致するイベントがあればすべて列挙させた。該当イベントが存在しない場合には 0 件とすることも許容した。第二段階では、第一段階で抽出された各イベントについて、当該記事本文および抽出済みのイベント概要に加え、当該種別に対して要求される属性仕様を提示し、必要なイベント属性を抽出させた。この段階でも、イベント属性に関する記述が記事中に存在しない場合には空のままとすることを許容した。

各段階のプロンプトは、PLOVER のスキーマ仕様 PDF を参照させた上で ChatGPT に草案を生成させ、著者が仕様との整合性を確認しながら調整を行った。この設計により、LLM がスキーマ定義をどの程度内在化し、種別依存の制約を運用できるかを検証した。

4.　定量的な抽出の兆候

1,000 件の記事を対象に抽出を行った結果、合計 1,104 件のイベントが得られた。図 2 に示すとおり、ニュースソース以外のフィルタリングを行っていないため政治イベントと無関係な記事も一定数含まれ、抽出件数が 0 件の記事も存在する。一方

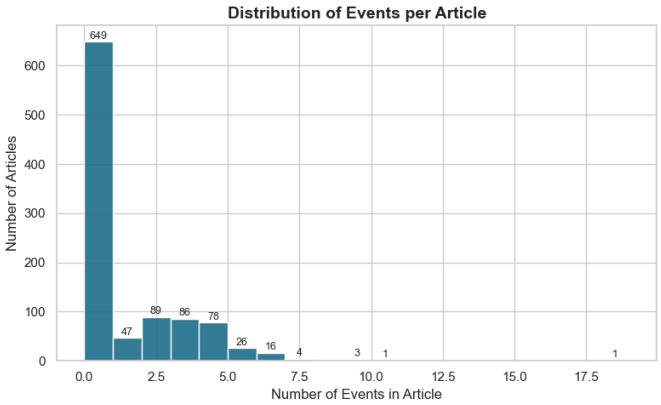

図 2　記事ごとの抽出イベント数の分布

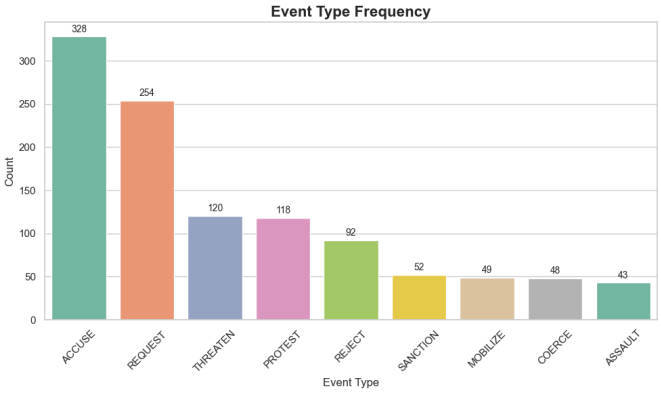

図 3　抽出結果のイベント種別頻度

で、複数のイベントが抽出された記事もあり、概要的に複数事象を扱う報道では分解的な抽出が行われている。これは LLM が一記事一イベントを前提とせず処理している兆候といえるが、同時に過抽出の可能性も否定できない。図 3 に示すイベント種別頻度には明確な偏りがあり、ACCUSE や REQUEST は比較的多いのに対し、COERCE、MOBILIZE、ASSAULT は少数かつ不安定であった。これがニュース分布の反映か、モデルの判定傾向によるものかは現段階では切り分けられていない。

属性抽出の傾向を図 4 に示す。location は比較的安定して抽出される一方、actor や recipient はばらつきが大きく、イベント種別によって成功状況が大きく変動した。また第 2 段階では、要求したデータ構造自体が返却されないケースを含む抽出失敗が 44.1% に達した。種別ごとに必須属性が異なるというスキーマの複雑性が影響し、定義に整合する形式での出力が安定しない可能性が示唆される。

5.　専門家による定性評価

抽出結果に対する専門家の定性評価では、全体として事実記述を超えた推論に基づく分類が散見された。特に、将来的な展開や情勢判断を読み込み、実際には発生していない出来事をイベントとして同定する傾向が指摘された。ACCUSE や REQUEST では定義が広く解釈され、単なる政治的批判や拒否的ニュアンスを含む言説まで含める例が見られた。PROTEST では国民の不満表明を実際の抗議行動とみなす推論的分類があり、COERCE、MOBILIZE、SANCTION でも定義との乖離が確認された。一方、ASSAULT は比較的妥当に抽出されてい

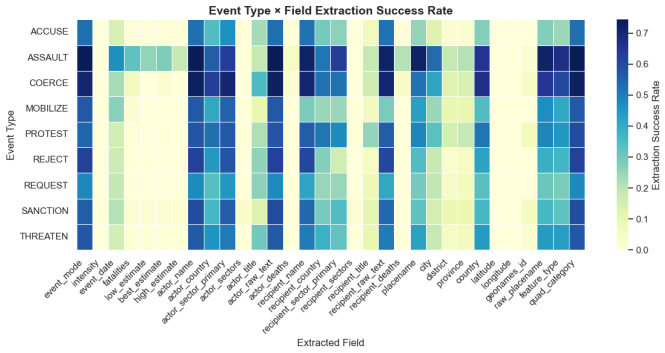

図 4　イベント種別ごとの属性抽出成功率

る事例が多かった。

イベント属性については、location は概ね適切に抽出される傾向があったが、actor と recipient の逆転や、主体として法令・措置が出力される事例が複数確認された。また、スポーツ記事が一部混入するなど、イベント種別自体の誤判定が属性抽出にも影響している。個別事例を見ると、種別と属性の双方が整合する例も存在するが、種別の誤同定や定義解釈の揺らぎが後続の属性付与に連鎖する構造が示唆された。

6.　結論

本研究は、PLOVER スキーマに基づく政治イベント情報抽出を LLM で自動化できるかを検証した。定量的傾向および専門家による定性評価の結果、イベント種別や属性ごとに妥当性のばらつきが見られ、特にスキーマ定義の厳密な運用には課題が残ることが明らかとなった。一方で、種別や属性によっては実用に資する水準でのデータ生成も可能であり、設計の工夫次第で大規模データ構築への応用可能性が示唆された。

謝辞

本研究は東京大学デジタルオブザーバトリ研究推進機構において実施されたものです。

参考文献

[OpenAI 23] Achiam, J., et al.: GPT-4 Technical Report. arXiv:2303.08774, (2023).

[Azar 80] Azar, E.: The Conflict and Peace Data Bank (COPDAB) Project. *Journal of Conflict Resolution*, 24(1):143–152, (1980).

[Chowdhery 23] Chowdhery, A., et al.: PaLM: Scaling Language Modeling with Pathways. *Journal of Machine Learning Research*, 24.240:1–113, (2023).

[Halterman 23a] Halterman, A., et al.: PLOVER and POLECAT: A New Political Event Ontology and Dataset. International Studies Association Conference Paper, (2023).

[Halterman 23b] Halterman, A., et al.: Creating Custom Event Data without Dictionaries: A Bag-of-Tricks, arXiv:2304.01331, (2023).

[Hendrycks 20] Hendrycks, D., et al.: Measuring Massive Multitask Language Understanding. arXiv:2009.03300, (2020).

[Leetaru 13] Leetaru, K. and Schrodt, A.: GDELT: Global Data on Events, Location, and Tone, 1979–2012. In ISA Annual Convention, (2013).

[McClelland 78] McClelland, C.: World Event/Interaction Survey, 1966-1978. WEIS Codebook ICPSR, 5211(640):49, (1978).

[Olsen 24] Olsen, H., et al.: Socio-political Events of Conflict and Unrest: A Survey of Available Datasets, In Proceedings of CASE 2024.

[PLOVER 18] Open Event Data Alliance: Political Language Ontology for Verifiable Event Records. `https://github.com/openeventdata/PLOVER`, (2018). Accessed: 2026-02-17.

[Raleigh 10] Raleigh, C., et al.: Introducing ACLED: An Armed Conflict Location and Event Dataset. *Journal of Peace Research*, 47(5):651–660, (2010).

[Schrodt 94] Schrodt, P.: Automated Coding of International Event Data Using Sparse Parsing Techniques. In Annual Meeting of the International Studies Association, Chicago, (1994).

[Sundberg 13] Sundberg, R. and Melander, E.: Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research*, 50(4):523 – 532, (2013).