

# Anchoring を行う生成的關係抽出

広田航<sup>1</sup> 高橋洸丞<sup>1</sup> Benjamin Heinzerling<sup>2,3</sup> Qin Dai<sup>3</sup> 近江崇宏<sup>1</sup> 乾健太郎<sup>4,3,2</sup>

<sup>1</sup> ストックマーク株式会社 <sup>2</sup> 理化学研究所 <sup>3</sup> 東北大学 <sup>4</sup> MBZUAI  
wataru.hirota@stockmark.co.jp kosuke.takahashi@stockmark.co.jp  
benjamin.heinzerling@riken.jp qin.dai.b8@tohoku.ac.jp  
takahiro.omi@stockmark.co.jp kentaro.inui@mbzuai.ac.ae

## 概要

生成的言語モデルを用いた關係抽出(生成的関係抽出)によって關係抽出の精度は大きく向上したが、再現率が依然低いという問題がある。この問題に対し、本論文では任意の生成的関係抽出に適用できる新手法 Anchoring を提案する。Anchoring はあらかじめ3つ組 (*head*, *relation*, *tail*) の *head* の候補となるスパン (anchor) を抽出し、各 anchor に対して個別に生成的関係抽出を行う手法である。評価実験では複数の關係抽出データセットにおいて Anchoring が精度および再現率が向上させることを示した上で、特に *head* および *tail* の抽出性能の向上に寄与すること、および文の主題でない *head*, *tail* に対しても頑健性が高まることを示す。

## 1 はじめに

生成的言語モデルの登場により關係抽出の性能は大きく向上している。従来の關係抽出はスパン特定モデルおよび關係分類モデルを用いて段階的に解くパイプライン方式が一般的であったが、近年は生成的言語モデルを使用して關係抽出を解く生成的関係抽出が提案されている(図1上段, 中段)。

生成的関係抽出は高い精度を示すものの、特定のスパンに関する3つ組のみを抽出する傾向があり再現率が低いという問題がある[1]。この問題に対し、本論文では任意の生成的関係抽出に適応可能な新手法 **Anchoring** を提案する。Anchoring は事前に3つ組の *head* の候補となるスパン (anchor) を抽出し、それぞれの anchor に対して個別に生成的関係抽出を解く方法である。(図1下段)。Anchoring によって生成的関係抽出の長所である高い精度を維持しつつ、より抜け漏れなく3つ組を抽出できることが期待される。

評価実験では複数の關係抽出データセットで精度および再現率を比較し、Anchoring の有効性を検証す

る。また3つ組の *head*, *tail* の抽出性能も個別に評価し、提案手法は特に多様な *head* に関する3つ組を抽出する効果があることを示す。さらに3つ組に含まれるスパンが文の主題かどうかによって分けて性能を比較し、主題性が關係抽出に及ぼす影響を論じる。

## 2 関連研究

關係抽出は古典的に(1)スパン特定(2)關係分類の2段階で解くパイプライン方式が提案されてきた[2, 3]。近年ではこの2つを同時に解く End-to-End 形式の手法も提案されている[4, 5]。

Cabot ら[6]は關係抽出を End-to-End の seq2seq 問題として解く生成的関係抽出を提唱した。近年では、關係抽出や固有表現抽出、イベント抽出などの情報抽出タスクを統一したテンプレートせ生成的に解く手法が高い抽出性能を示している[7, 8, 9]。

生成的関係抽出は大規模言語モデル(LLM)と組み合わせることで高い精度を示す[7, 9, 10]一方で、再現率にあまり向上が見られないという問題がある[7, 1]。Ding ら[1]は特に3つ組数が多い關係抽出データで生成的関係抽出の再現率が低下することを示している。

## 3 前提

### 3.1 問題設定

關係ラベルの集合を  $R$ , 文  $D$  に存在するスパンの集合を  $S_D$  とするとき、關係抽出は  $D$  中の全ての3つ組の集合

$$\{(h, r, t) | h \in S_D, r \in R, t \in S_D\} \quad (1)$$

を得る問題である。



図1 パイプライン方式, 生成的関係抽出および生成的関係抽出 w/ Anchoring (提案手法) の概略図. 入力テキストおよび各手法の出力は Re-DocRED データセットにおける実際のデータからサンプルした. 本例では, 提案手法はパイプライン方式で見られた偽陽性 (赤色) を防ぎつつ Anchoring を行わない生成的関係抽出で抽出されなかった3つ組 (青色) を正しく抽出している. 簡単のため, 入力文および両手法の予測の一部を省略した.

### 3.2 ベースライン手法

本論文ではベースライン手法としてパイプライン方式と Anchoring を行わない生成的関係抽出を採用する. パイプライン方式は関係抽出を (1) スパン特定 (2) 関係分類の2段階に分けて解く方法である. パイプライン手法の概略図を図1上段に示す. 生成的関係抽出は seq2seq 問題として関係抽出を解く手法である (図1中段). 本論文では生成的関係抽出のテンプレートとして CodeIE [9] と UIE [7] を採用する. CodeIE と UIE のテンプレートを付録Aに示す.

## 4 提案手法: Anchoring

本論文では生成的関係抽出における Anchoring を提案する. Anchoring はあらかじめ3つ組の head の候補となるスパン (anchor) を抽出し, 次にテキストと個々の anchor の2つを入力して生成的関係抽出を解く方法である (図1下段). Anchoring は任意の生成的関係抽出に対して適用できる.

Anchoring の核心は, モデルに anchor に対して注目を向けさせるよう明示的に指示することである. モデルが各 anchor に対して注意深く関係抽出を行うよう仕向けることで, より3つ組の抽出漏れが少なくなることが期待される.

## 5 評価実験

### 5.1 実験設定

**データセット** 本実験では関係抽出のデータセットとして Re-DocRED [11], SciERC [12] および

BioRED [13] を用いる. Re-DocRED, SciERC, BioRED はそれぞれ Wikipedia 文書, 人工知能分野の学術文献, 生物医学分野の学術文献に対して関係ラベルを付与したデータセットである. 各データセットの詳細な情報を表1に示す.

**ベースライン** 生成的関係抽出のテンプレートには CodeIE [9] と UIE [7] を用いる. CodeIE (w/ Anchoring) と UIE (w/ Anchoring) はそれぞれのテンプレートに対し Anchoring を適用した手法である. これらのテンプレートを付録Aに示す. ベースライン手法との公平な比較のため, CodeIE, UIE の Anchoring におけるスパン特定はそれぞれ CodeIE, UIE を用いる. パイプライン方式におけるスパン特定は, より精度の高い UIE テンプレートを使用する. パイプライン方式における関係分類の学習では, 偽陽性を抑制するため, 正例の80%の量の負例を自動生成して教師データに追加した. これは Ni ら [10] が提案している値である.

**モデルと学習** 本実験では全ての手法で Llama 3.1 8B Instruct [14] に LoRA [15] チューニング ( $r = 16, \alpha = 32$ ) を行った. 全ての学習で学習率を  $1.0 \times 10^{-4}$ , 最適化手法を AdamW [16], バッチサイズを 16, エポック数を 3 に定めた.

### 5.2 ベースラインとの比較

**目的** 提案手法によって関係抽出の精度・再現率が向上するか検証する.

**結果と考察** 表2に關係抽出の精度, 再現率および F1 スコアを示す. Re-DocRED と SciERC において提案手法はベースライン手法と比べ全体的に高

表1 データセットの詳細な情報.

データセット	文書の種類	関係数	関係の対称性	平均トークン数	平均3つ組数	平均スパン数
Re-DocRED	Wikipedia	96	非対称	235.3	31.9	63.7
SciERC	学術論文	8	非対称	151.6	9.2	18.5
BioRED	学術論文	9	対象	69.7	2.7	6.0

表2 関係抽出の精度 (P), 再現率 (R) および F1 スコア (F).

手法	Re-DocRED			SciERC			BioRED		
	P	R	F	P	R	F	P	R	F
パイプライン方式	0.594	0.341	0.433	0.188	<b>0.379</b>	0.251	<b>0.518</b>	0.527	<b>0.522</b>
CodeIE	0.527	0.044	0.081	0.230	0.194	0.210	0.489	0.432	0.459
(w/ Anchoring)	0.540	0.522	0.531	0.177	0.261	0.211	0.302	<b>0.578</b>	0.397
UIE	0.632	0.281	0.389	0.256	0.242	0.249	0.503	0.442	0.470
(w/ Anchoring)	<b>0.635</b>	<b>0.527</b>	<b>0.576</b>	<b>0.313</b>	0.373	<b>0.340</b>	0.498	0.516	0.507

表3 head のみ, tail のみ, ペアのみの精度と再現率.

	Head		Tail		Pair	
	P	R	P	R	P	R
<i>Re-DocRED</i>						
パイプライン方式	<b>0.845</b>	0.679	0.755	0.596	0.640	0.367
CodeIE	0.796	0.069	0.739	0.670	0.589	0.048
(w/ Anchoring)	0.743	<b>0.853</b>	0.764	0.684	0.601	<b>0.583</b>
UIE	0.821	0.413	0.830	0.732	<b>0.699</b>	0.311
(w/ Anchoring)	0.748	0.766	<b>0.855</b>	<b>0.751</b>	0.693	0.575
<i>SciERC</i>						
パイプライン方式	0.451	0.600	0.491	0.673	0.211	<b>0.432</b>
CodeIE	<b>0.508</b>	0.428	0.535	0.544	0.284	0.252
(w/ Anchoring)	0.375	<b>0.617</b>	0.516	0.509	0.207	0.320
UIE	0.464	0.487	0.639	0.591	0.307	0.283
(w/ Anchoring)	0.451	0.614	<b>0.746</b>	<b>0.716</b>	<b>0.353</b>	0.429
<i>BioRED</i>						
パイプライン方式	<b>0.804</b>	0.778	-	-	0.631	0.646
CodeIE	0.801	0.737	-	-	<b>0.638</b>	0.568
(w/ Anchoring)	0.567	<b>0.930</b>	-	-	0.379	<b>0.723</b>
UIE	<b>0.804</b>	0.731	-	-	0.632	0.560
(w/ Anchoring)	0.775	0.802	-	-	0.594	0.627

い精度・再現率を示した。このことから、Anchoring は3つ組をより正確に、網羅的に行う効果があることがわかる。一方で BioRED においては Anchoring によって精度が低下した。これは BioRED データセットは関係が対照的であることに起因すると考えられる。すなわち、3つ組 (*head, relation, tail*) と (*tail, relation, head*) は同義であるが、Anchoring を行った場合は別々に推論が行われるため、偽陽性が増え精度が低下したと考えられる。対称性のある関係に対する精度の向上は提案手法の今後の課題である。

### 5.3 head および tail の抽出性能

**目的** 提案手法の性能が向上した理由として考えられる以下の仮説を検証する。

- head の抽出性能が上がった
- head に対する tail の抽出性能が上がった。
- 関係ラベルの分類性能が上がった。

**実験設定** 各手法の予測結果を以下の指標で比較する。

- head の精度と再現率 (Head).
- 各 head に対する tail の精度と再現率 (Tail).
- (head, tail) のペアに対する精度と再現率 (Pair). 関係ラベルの正誤は問わない。

なお BioRED は関係が対称的であることから Head と Tail は同一の評価になるため、Head のみ評価する。

**結果と考察** 結果を表3に示す。

Head においては、提案手法の精度はベースライン手法と比べ同等か少し下がる一方、再現率は大きく向上した。Tail においては提案手法が精度と tail のいづれも高い値を示した。Pair においても提案手法が高い精度、再現率を示した。Pair の精度、表2で示した3つ組の精度、再現率の差は手法によらず同程度であった。両者の違いは関係ラベルの正誤を考慮するかどうかにあることから、関係の分類性能は手法間の差異がないことがわかった。

以上を総合し、提案手法は (1) head の網羅性を向上させ (2) head が与えられた上での tail の精度、再現率を高く維持したことで関係抽出の性能が向上していることがわかった。

### 5.4 スパンの主題性

**目的** 「ベースライン手法は文の主題ではないスパンの3つ組を取りのながすことが多く、提案手法は主題ではないスパンに対しても頑健である」という仮説を立て検証した。ここではスパンが主題である

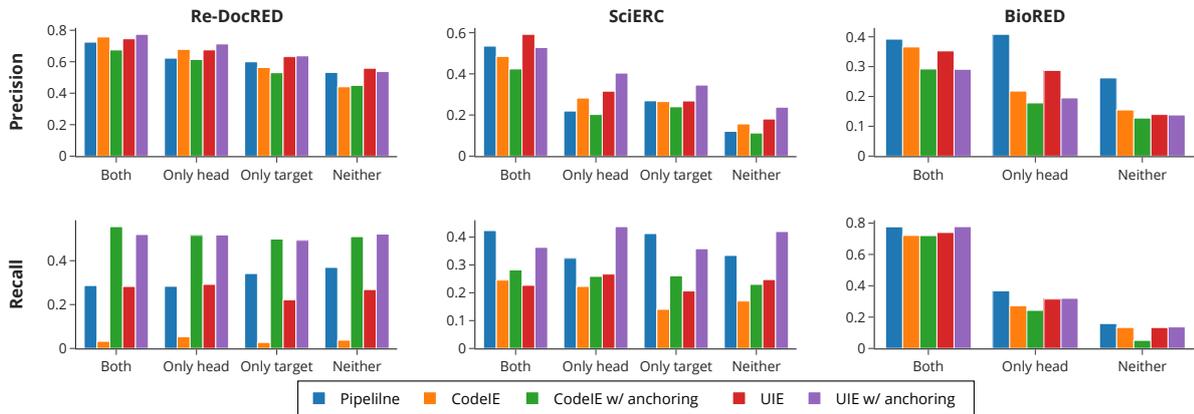


図2 スパンを主題・非主題に分けた時の関係抽出の性能. 上段の3つのグラフの縦軸は精度, 下段の3つのグラフの縦軸は再現率を表す. グラフの横軸はテストデータのカテゴリを表し, 異なる色の縦棒でそれぞれの手法を表す.

表4 トピックの主題・非主題で分類した4つのデータセットの3つ組数.

データセット	Both	Only head	Only tail	Neither
Re-DocRED	5,039	7,442	3,349	2,309
SciERC	450	158	155	207
BioRED	888	547	-	161

ことを「文がそのスパンについて主に述べている」ことと定める. 例えば「テスラは自動車メーカーであり, アメリカに本社がある」という文では, テスラが主題スパン, 自動車・アメリカは非主題スパンに分類される.

**実験設定** GPT-4o<sup>1)</sup>を用いて正解の3つ組に含まれる *head* および *tail* が文の主題であるかどうかを分類し, 3つ組を「*head* も *tail* も主題である (Both)」「*head* のみが主題である (Only head)」「*tail* のみが主題である (Only tail)」「どちらも主題ではない (Neither)」の4つのカテゴリに分けた. 4つのカテゴリの3つ組数を表4に示す. 分類に使用したプロンプトは付録Bに示す. なお BioRED は関係に対称性があることから Only *tail* の評価は行わない.

**結果と考察** 図2に4つのカテゴリでの精度と再現率を示す. 結果より, Re-DocRED と SciERC においてベースライン手法は非主題のスパンに対する再現率が低下するのに対し, 提案手法では非主題のスパンに対しても高い再現率を示すことがわかった. このことから, Anchoring を行わない生成的関係抽出はより主題のスパンに偏って着目して3つ組を抽出するが, Anchoring によってそれが緩和されていること

がわかる. 一方で非主題のスパンに対する精度は全ての手法・データセットにおいて低く, 提案手法でも改善が見られない. 非主題のスパンに対する精度の向上は今後の課題である.

また興味深いことに, 生成的関係抽出では Only *tail* の性能は Only *head* に比べ低いという現象がみられた. この現象はパイプライン方式では見られないことから, 生成的関係抽出のモデルには3つ組の *head* と *tail* の順番に対するバイアスがあることが推察される.

## 6 おわりに

本論文では, 生成的関係抽出の再現率が低いという問題に対する解決策として Anchoring を提案し, 評価実験によりその効果を確認した. また評価実験では (1) 提案手法は多様な *head* に対する3つ組の抽出に効果があること (2) 生成的関係抽出は主題ではないスパンの3つ組に対して性能が劣化するが, 提案手法ではそれが緩和されることをそれぞれ示した.

本研究の今後の課題として, 5.2章で論じた対称性のある関係抽出における精度の向上, および 5.4章で論じた非主題スパンの精度の向上が挙げられる. これらの問題は関係ラベルなど *head* 以外の anchor を採用することで解決できる可能性がある.

本論文で提案した Anchoring はあらゆる生成的関係抽出に簡単に適用できる上に, 追加の教師データも必要としない. この利点を踏まえ, 関係抽出において幅広く本手法が使われることを期待する.

1) モデルは gpt-4o-2024-08-06 を用いた.

## 参考文献

- [1] Zepeng Ding, Wenhao Huang, Jiaqing Liang, Yanghua Xiao, and Deqing Yang. Improving recall of large language models: A model collaboration approach for relational triple extraction. In Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue, editors, **Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation**, pp. 8890–8901, Torino, Italia, May 2024. ELRA and ICCL.
- [2] Nanda Kambhatla. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations. In **Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions**, p. 22–es, Barcelona, Spain, 2004. Association for Computational Linguistics.
- [3] Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao. Relation classification via convolutional deep neural network. In Junichi Tsujii and Jan Hajic, editors, **Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers**, pp. 2335–2344, Dublin, Ireland, August 2014. Dublin City University and Association for Computational Linguistics.
- [4] Makoto Miwa and Yutaka Sasaki. Modeling joint entity and relation extraction with table representation. In Alessandro Moschitti, Bo Pang, and Walter Daelemans, editors, **Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing**, pp. 1858–1869, Doha, Qatar, October 2014. Association for Computational Linguistics.
- [5] Arzoo Katiyar and Claire Cardie. Going out on a limb: Joint extraction of entity mentions and relations without dependency trees. In Regina Barzilay and Min-Yen Kan, editors, **Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 917–928, Vancouver, Canada, July 2017. Association for Computational Linguistics.
- [6] Pere-Lluís Hugué Cabot and Roberto Navigli. REBEL: Relation extraction by end-to-end language generation. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, **Findings of the Association for Computational Linguistics: EMNLP 2021**, pp. 2370–2381, Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.
- [7] Yaojie Lu, Qing Liu, Dai Dai, Xinyan Xiao, Hongyu Lin, Xianpei Han, Le Sun, and Hua Wu. Unified structure generation for universal information extraction. In **Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 5755–5772, Dublin, Ireland, May 2022. Association for Computational Linguistics.
- [8] Giovanni Paolini, Ben Athiwaratkun, Jason Krone, Jie Ma, Alessandro Achille, RISHITA ANUBHAI, Cicero Nogueira dos Santos, Bing Xiang, and Stefano Soatto. Structured prediction as translation between augmented natural languages. In **International Conference on Learning Representations**, Virtual Event, May 2021. PMLR.
- [9] Peng Li, Tianxiang Sun, Qiong Tang, Hang Yan, Yuanbin Wu, Xuanjing Huang, and Xipeng Qiu. CodeIE: Large code generation models are better few-shot information extractors. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics**, pp. 15339–15353, Toronto, Canada, July 2023. Association for Computational Linguistics.
- [10] Jian Ni, Gaetano Rossiello, Alfio Gliozzo, and Radu Florian. A generative model for relation extraction and classification. **arXiv [cs.CL]**, February 2022.
- [11] Qingyu Tan, Lu Xu, Lidong Bing, Hwee Tou Ng, and Sharifah Mahani Aljunied. Revisiting DocRED - addressing the false negative problem in relation extraction. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing**, pp. 8472–8487, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.
- [12] Yi Luan, Luheng He, Mari Ostendorf, and Hannaneh Hajishirzi. Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun’ichi Tsujii, editors, **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing**, pp. 3219–3232, Brussels, Belgium, October–November 2018. Association for Computational Linguistics.
- [13] Ling Luo, Po-Ting Lai, Chih-Hsuan Wei, Cecilia N Arighi, and Zhiyong Lu. Biored: a rich biomedical relation extraction dataset. **Briefings in Bioinformatics**, Vol. 23, No. 5, p. bbac282, July 2022.
- [14] Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, and Angela Fan et al. The llama 3 herd of models. **arXiv [cs.AI]**, July 2024.
- [15] Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In **International Conference on Learning Representations**, Virtual Event, may 2022. PMLR.
- [16] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In **International Conference on Learning Representations**, Louisiana, U.S., May 2019. PMLR.

## A 生成的関係抽出のテンプレート

### A.1 CodeIE

CodeIE は以下のようなソースコードに模したテンプレートで関係抽出を解く。

```
def relation_extraction(input_text):
    """ extract the relations of named entities
    from the input text. """
    input_text = "Steve became CEO of Apple."
    entity_relation_list = []
    # extracted relations
    entity_relation_list.append(
        {
            "rel_type" "work_for",
            "ent1_name": "Steve",
            "ent2_name": "Apple"
        }
    )
```

Anchoring を行った CodeIE は input\_text 以降の部分を変換するように変更する。

```
input_text = "Steve became CEO of Apple."
anchor = "Steve"
entity_relation_list = []
# extracted relations
entity_relation_list.append(
    {
        "rel_type" "work_for",
        "ent2_type": "Apple"
    }
)
```

### A.2 UIE

UIE は以下のようなテンプレート (Structured Extraction Language; SEL) で関係抽出を解く。

```
入力: Steve became CEO of Apple.
出力:
(
  (person: Steve
    (work for: Apple)
  )
)
```

Anchoring を行った UIE は以下のようなテンプレートになる。

```
入力: Steve became CEO of Apple.
anchor: Steve
出力:
(
  (person: Steve
```

```
(work for: Apple)
)
)
```

## B 主題スパン・非主題スパンの分類

主題スパン・非主題スパンの判定は以下のプロンプトを用いた。

Task:

You are an expert in identifying topics and categorizing entities from text. Your goal is to analyze the given text and a list of entities to classify them into two categories:

1. Main Topic Entities: Entities that represent the central topic(s) or main focus of the text.
2. Other Entities: Entities that are mentioned in the text but are not central to its main topic(s).

Input:

- `text`: A paragraph or document.
- `entities`: A list of entities extracted from the text.

Output:

A JSON object with two fields:

- "main\_topic\_entities": A list of entities that represent the main topic(s).
- "other\_entities": A list of entities that are not central but still mentioned in the text.

Example Input:

Text: "Electric vehicles are gaining popularity as a sustainable alternative to traditional cars. Tesla, a leading manufacturer of EVs, has announced plans to expand its charging network. The government is also offering subsidies to promote electric vehicle adoption."

Entities:

["Electric vehicles", "Tesla", "charging network", "government", "subsidies", "traditional cars"]

Example Output:

```
---
{
  "main_topic_entities": ["Electric vehicles",
    "Tesla"],
  "other_entities": ["charging network",
    "government", "subsidies", "traditional cars"]
}
---
```