言語横断ラベル射影を用いた 日本語文書レベル関係抽出データセットの構築

Youmi Ma　An Wang　岡崎直観
[1] 東京工業大学大学院
{youmi.ma@nlp.,an.wang@nlp.,okazaki@}c.titech.ac.jp

概要

文書レベル関係抽出（DocRE）は文書中の全てのエンティティ組の関係を推定するタスクである．英語 DocRE の研究は活発に行われてきたが，日本語の DocRE 言語資源はまだ存在しない．本稿では，英語 DocRE 言語資源を活用しつつ，日本語 DocRE 言語資源の構築を目指す．まず翻訳とラベル射影に基づいた日本語 DocRE データセットを自動構築したが，得られたデータセットが実用に耐えないことが分かった．そこで，モデルの予測に人手で修正を加える半自動構築手法を提案した．提案手法はアノテータの負担を軽減しながら，自動構築よりも高品質なデータセットを構築できることを報告する[1)]．

1　はじめに

関係抽出は自然文から知識関係を三つ組 (*subject*, *relation*, *object*) の形で抽出するタスクである．文に閉じた関係に限らず，文の境界をまたがるエンティティ間の関係も認識する設定として，**文書レベル関係抽出**（**DocRE**: **Doc**ument-level **R**elation **E**xtraction）が提案された [1]．DocRE は一般的な関係抽出と同様に知識グラフの補完や質問応答に役立つほか [2, 3]，モデルの長文読解力も反映できる [4]．また，GPT [5, 6] をはじめとする大規模言語モデル（LLM: Large Language Model）の DocRE における性能は低いことが報告されており [7]，この視点からも本タスクは取り組むべき研究課題であると考える．

DocRE の研究は主に英語で行われてきた [1, 8, 9]．しかし，英語テキストで言及されず，特定の言語のテキストでのみ書かれている関係知識が存在する．DocRE の対象言語を英語に限定して研究すると，英語以外の言語の知識グラフの補完が困難となる．本稿では，DocRE を英語以外の言語に拡張することを目指し，英語の言語資源を活用しながら対象言語の DocRE データセットを構築する方法を探求する．本研究では，Web テキストは豊富であるが，DocRE の言語資源が存在しない日本語を対象言語とする．

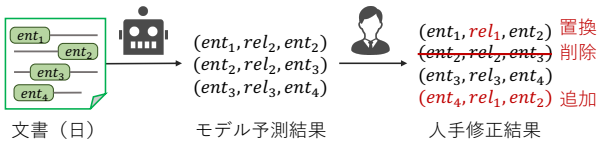

図 1　データセットの半自動構築手法の概要

まず，日本語 DocRE データセットを自動で構築し，その品質を調べる．具体的には，英語の DocRE データセット Re-DocRED [9] を日本語に翻訳すると同時に，**言語横断ラベル射影**（**CLP**: **C**ross-lingual **L**abel **P**rojection）を行い，Re-DocRED^{ja} と呼ぶデータセットを得る．CLP は単語アライメントに基づいてラベルを射影する手法であり，構造化予測タスクの言語横断転移学習での有効性が報告されている [10, 11]．Re-DocRED^{ja} の品質を確認するため，DocRE モデルの学習データとして使用し，得られたモデルの性能を日本語ウィキペディアで測定した．その結果，関係抽出の漏れが多数観察され，Re-DocRED^{ja} が不完全であることが示唆された．

日本語 DocRE データセットの自動構築の問題を解決するため，人手修正を加えた半自動構築手法を採用する．図 1 のように，日本語文書にある関係事例の候補を自動で推薦する．アノテータは文書の内容から候補事例の真偽を判断し，添削を行う．これにより収集したデータセットを JacRED (**Ja**panese **Do**cument-level **R**elation **E**xtraction **D**ataset) と呼び，その統計情報を表 1 に示す．なお，関係の推薦は辞書マッチにより行われてきたが [1]，本稿では Re-DocRED^{ja} で学習したモデルの予測により行う．モデル予測による関係推薦は辞書マッチよりもアノテータの負担を軽減できることを定量的に示した．

実験では，JacRED で学習したモデルが Re-

1)　構築したデータセットを https://github.com/YoumiMa/JacRED で公開する．

表 1　JacRED と既存データセットの比較

データセット	言語	関係ラベル数	文書数	関係三つ組数	文書長（トークン数）	根拠文情報
DocRED [1]	英	96	4,051	50,503	198.4	あり
Re-DocRED [9]	英	96	4,053	120,664	198.4	一部あり
HacRED [12]	中	26	7,731	56,798	122.6	なし
HistRED [13]	韓	20	5,816	9,965	100.6	あり
JacRED	日	35	2,000	42,241	260.1	あり

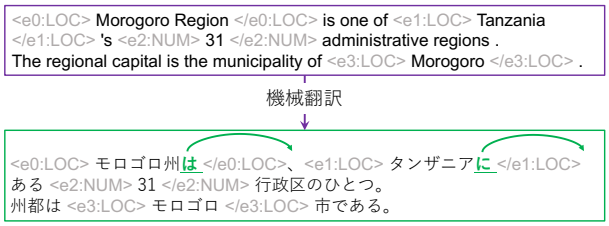

図 2　Re-DocRED [9] を日本語に射影する方法の例

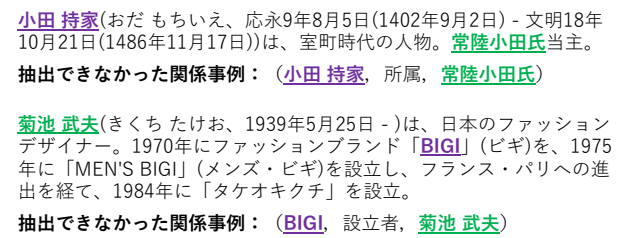

図 3　Re-DocRED^{ja} で学習したモデルが抽出できなかった関係事例の例．簡易のため，文書の一部を省略した．

DocRED^{ja} で学習したモデルより高い関係抽出性能を示した．これにより，高品質で実用に耐えうる日本語 DocRE データセットが収集できた．また，JacRED を日本語 DocRE のベンチマークとし，既存の DocRE モデルの性能評価も行った．

2　データセットの構築

タスク定義　DocRE の目的は，文 $\mathcal{X}_D = \{x_i\}_{i=1}^{|\mathcal{X}_D|}$ からなる文書 D における全エンティティ $\mathcal{E}_D = \{e_i\}_{i=1}^{|\mathcal{E}_D|}$ の組が持つ関係 $r \in \mathcal{R}$ を推定することである．文書 D におけるエンティティ $e \in \mathcal{E}_D$ の言及（メンション）は $\mathcal{M}_e = \{m_i\}_{i=1}^{|\mathcal{M}_e|}$ であり，$\mathcal{R}$ は事前に定義された関係ラベルの集合である．また，エンティティ組 (e_s, e_o) の間に関係 r が成り立つ場合，三つ組 (e_s, r, e_o) の根拠 $\mathcal{V}_{e_s,r,e_o} \subset \mathcal{X}_D$ を文単位で予測する**根拠認識**タスクも考慮する．根拠認識は DocRE のサブタスクとして研究されてきた [1, 14, 15, 16]．

2.1　翻訳に基づいた自動構築

まず言語横断ラベル射影（CLP）を用いて，英語データセット Re-DocRED [9] の日本語版を自動構築する．Re-DocRED は DocRE で最も使われているデータセット DocRED [1] の改良版であり，英語ウィキペディアに基づいている．

翻訳とラベル射影　Re-DocRED の日本語版を mark-then-translate 方式で構築する [10]．この方式では，テキストの翻訳と同時にエンティティのラベル射影を行う．図 2 に示したように，機械翻訳器の入力として，エンティティの言及を XML タグで囲んだ英語文書を渡す．翻訳器は XML タグおよびそれに囲まれた内容の語義を保持したまま翻訳を行い，エンティティ言及の位置情報を含む日本語文書が得られる．翻訳と同時にラベル射影を行うため，単語アラインメントを個別に行う必要はない．追加学習を行わなくても図 2 のように翻訳できる DeepL[2] を翻訳器として採用した．さらに言及スパンの末尾に含まれやすい格助詞を除外する後処理を行った．得られたデータセットを Re-DocRED^{ja} と呼ぶ．

データセットの問題点　Re-DocRED で最良性能を示すモデルである DREEAM [16] を Re-DocRED^{ja} の学習データで学習し，テストデータで評価した結果，72.74 の F 値が得られた．しかし，日本語ウィキペディアから収集した文書にこのモデルを適用した結果，一部の関係事例の抽出が困難であることを観察した[3]．図 3 に抽出できなかった関係事例の例を示した．一番目の例は日本史に関する文書であり，英語ウィキペディアでは対応する記事が存在しない．また，「当主」は現在の日本社会で使われる用語ではないため，英語からの翻訳文で用いられにくく，関係の抽出が難しかったと考えられる．二番目の例では，関係事例（MEN'S BIGI，設立者，菊池武夫）と（タケオキクチ，設立者，菊池武夫）が正常に抽出できたにもかかわらず，（BIGI，設立者，菊池武夫）のみ抽出できなかった．これは抽出できた事例はブランド名が動詞「設立」と隣接しているのに対し，抽出できなかった事例の動詞は省略されたためと考えられる．ここの省略は日本語の母語話者にとって自然であるが，英語由来の翻訳テキストでは省略が起こりにくい．以上により，Re-DocRED^{ja}

2）https://api.deepl.com/v2/translate
3）定量的な評価結果を §4.1 で示す．

は日本語話者が関心を持つような話題を網羅できず，よく用いられる構文を反映できないため，日本語文書に対する関係抽出を十分に行えない．

2.2 人手修正を組み込んだ半自動構築

これを踏まえて，生の日本語文書からデータセットを作成し，母語話者によるアノテーションを行う．集めたデータセットを JacRED と呼ぶ．日本語ウィキペディアの記事に対し，冒頭文で 256 文字より長いものを文書の候補とする．既存研究に倣い [1]，作業をエンティティラベル付与と関係ラベル付与の二段階に分ける．アノテータは自動推薦されたラベルを修正する形でアノテーションを行う．自動推薦の品質が低ければ，人手修正に手間がかかるため，なるべく質の高い自動推薦を提供する必要がある．本稿では自動推薦の品質を改善するため，幾つかの策を講じる（§2.2.2）．

2.2.1 エンティティ言及のラベル付与

ラベル種類　JacRED では IREX (Information Retrieval and Extraction Excise) [17] で使われた 8 種類の固有表現ラベルを採用する．これにより得られたラベルセットは DocRED [1] に類似している．

エンティティ言及の自動抽出　各文書におけるエンティティ言及の抽出は KWJA [18] により行う．KWJA で解析された結果，言及数が 10 個未満の文書は候補から除外される．

文書の選別　文の境界を超えた関係事例数を確保するため，再度文書の選別を行う．まず辞書マッチ [19] により文書に存在しうる関係事例数を概算する．具体的には，自動認識されたエンティティの言及を既存の知識ベース Wikidata [20] にリンクし，エンティティ組ごとに知識関係が存在するかどうかを確認する．関係事例が Wikidata に存在する場合，文書にも当該関係が成り立つとみなす．次に，各文書における関係事例の総数を計算し，文の境界を超えるが 4 つ未満のものを候補から除外する．

人手修正　残された文書からランダムに 2,000 件を選出し，人手修正の対象とする．アノテータは自動抽出されたエンティティ言及を修正し，抽出できなかったエンティティ言及を追加する．

2.2.2 関係のラベル付与

§2.2.1 の結果に基づいて，共参照と関係のラベルを付与する．関係のラベル付与では，ラベル数の削減と自動抽出関係事例の品質改善により，アノテータの負担軽減及びデータセットの品質向上を図る．

ラベル種類　(Re-)DocRED では Wikidata にある関係から頻度の高い 96 種類を選出し，関係ラベルセット $\mathcal{R}$ を構築した [1, 9]．96 種類の関係ラベルを全部理解した上でアノテーションを行うのは困難であり，結果的にはデータセットの品質に影響を与えると考えられる．そのため，関係ラベル数を以下の基準に沿って削減する．(1) rich-ERE [21] で定義された全ての関係種類を網羅する，(2) `HasPart` と `PartOf` のように，Wikidata で逆の関係として定義された関係ラベルは片方だけ残す，(3) (Re-)DocRED で頻度の高い関係ラベルはなるべく採用する．これにより，アノテーションの対象となる関係ラベル種類数を 28 にまで削減できた．人手アノテーション後，(2) に該当する逆の関係を自動補完し，得られた関係ラベルセット $\mathcal{R}'$ の種類数は 35 である．関数 $f(\cdot) : \mathcal{R} \rightarrow \mathcal{R}'$ を用いて Re-DocRED の関係事例を JacRED の関係ラベルセットに射影した結果，88% 以上の関係事例を温存できることを確認した．

関係事例の自動抽出　既存研究では §2.2.1 で紹介した辞書マッチの結果を自動推薦として用いたが [1]，本稿では Re-DocRED$^{\text{ja}}$ で学習したモデルで関係抽出を行った結果を自動推薦とする．この方法で推薦された関係事例は，以下の点において辞書マッチより優れると期待できる．(1) Wikidata を含め，知識ベースの網羅性は限定的であり，全ての関係事例を網羅できない，(2) 辞書マッチは文脈を考慮できないが，モデル予測は周囲の文脈に基づいた柔軟な関係抽出ができる．辞書マッチによる推薦とモデル予測による推薦の定量的な比較を §3 で行う．

人手修正　アノテータは自動抽出された関係事例 (e_s, r, e_o) の真偽を判断し，成立しない関係を削除する．関係が成立する場合，文書からその関係を裏付ける根拠 $\mathcal{V}_{e_s,r,e_o}$ を特定し，文単位で提示する．最後に，自動抽出できなかった関係事例を追加する．

3 モデル予測による推薦の優位性

§2.2.2 では，モデル予測は辞書マッチより品質の高い推薦ができる理由を定性的に述べた．本節では，その根拠を定量的に示す．そのため，収集した JacRED から 400 件の文書をランダムでサンプリングし，自動推薦から人手修正結果に至るまでに要する編集ステップ数を計算する．モデル予測と辞書マッチの 2 通りの自動推薦から，最終的なアノテー

表 2 関係事例の推薦をモデル予測で行う場合と辞書マッチで行う場合に要する人手修正ステップ数の比較．

	推薦	削除	置換	追加
モデル予測	6,500	1,266	224	2,740
辞書マッチ	3,200	1,459	113	6,233

表 3 JacRED のテストデータにおける DREEAM 関係抽出結果の適合率，再現率と F 値．データセットの文書数をカッコの中に示す．

訓練データ	適合率	再現率	F 値
JacRED (1,400)	**64.76**	**73.29**	**68.73**
Re-DocRED^{ja} (3,053)	56.14	53.67	54.87

ション結果に到達するまでの削除・置換・追加[4)]のステップ数を表 2 に示す．

表 2 から，モデル予測は辞書マッチより約 2 倍の関係事例を推薦できることが分かった．また，モデル予測を起点とした場合，推薦事例の 25%（6,500 件中 1,366+234=1600 件）を編集し，さらに 42%（6,500 件に対して 2,740 件）を追加することで最終のアノテーション結果に到達できるが，辞書マッチを起点とした場合，推薦事例の 50%を編集し，194%を追加することになる．この事実は，提案したモデル予測を起点とするアノテーション方法がアノテータの負担軽減につながることを示唆している．

4 実験

目的 以下の問いに答えるため実験を行う．(1) 人手修正を加えることにより，データセットの品質は改善できたか？ (2) 既存の DocRE モデルは日本語 DocRE をどれほど解けるのか？

設定 JacRED の文書全 2,000 件を，学習・検証・テストデータとして 1,400/300/300 に分割する．モデルは全て一枚の Tesla V100 16GB GPU で学習した．ランダムに初期化されたモデルを 5 つ学習し，予測結果の F 値の平均を報告する．

4.1 人手修正によるデータセットの品質

ここでは，自動構築データセット Re-DocRED^{ja} と JacRED の品質を定量的に評価する．具体的には，それぞれのデータセットで学習したモデルを JacRED のテストデータで評価し，その性能を比較する．表 3 に示すように，JacRED で学習したモデルは少ない学習事例数で Re-DocRED^{ja} で学習したモデルを上回る性能を達成できた．特に再現率に乖離があることは，§2.1 で例示した自動構築データセットの問題点の裏付けである．

4) 削除・置換・追加の例は図 1 を参照されたい．

表 4 モデルの性能（F 値）．英語で関係抽出は Re-DocRED [9]，根拠認識は DocRED [1] の結果である．

	JacRED		(Re-)DocRED	
	関係	根拠	関係	根拠
ATLOP [8]	68.04	–	77.56	–
DocuNet [22]	67.66	–	77.87	–
KD-DocRE [23]	68.29	–	78.28	–
EIDER [14]	68.61	57.16	73.80	51.27
DREEAM [16]	**68.73**	**62.11**	**80.73**	**51.71**
GPT-3.5 [5]	13.17	–	6.55	–
GPT-4 [6]	27.45	–	14.29	–

4.2 既存モデルの性能

JacRED をベンチマークとして用い，既存モデルの日本語 DocRE における性能を測る．ここでは，DocRE の教師ありモデル [8, 22, 23, 16] と汎用的な大規模言語モデル [5, 6] を対象とする．大規模言語モデルによる DocRE は関係ラベルごとに抽出例を 7 個提示し，in-context learning を行う．この方法は一回の予測に必要な入力プロンプトが長く，コストがかさむため，テストデータからランダムに選んだ 10 個の文書での評価結果を報告する．

表 4 に実験結果を示す．まず，教師ありモデルに着目する．関係抽出の性能では，ATLOP が最下位で DREEAM が最上位であり，Re-DocRED と同様の傾向を示した．一方，教師ありモデルの Re-DocRED での F 値が 70 台であるが，JacRED での F 値は 60 台であった．これにより，日本語 DocRE の精度はまだ英語 DocRE に及ばないことが分かった．ゆえに，日本語 DocRE にはまだ課題が残されていると考える．次に，大規模言語モデルの性能に着目する．教師ありモデルと比べて，GPT-3.5 と GPT-4 は JacRED と Re-DocRED の両方で性能が低かった．この結果は先行研究と一致しており [7]，大規模言語モデルの DocRE での有効性が限定的であることを示唆する．

5 おわりに

本稿では，日本語 DocRE の言語資源構築に英語 DocRE 言語資源を活用する方法を探求した．翻訳とラベル射影による自動構築の問題点を洗い出した後，人手修正とモデル予測を組み合わせたデータセットの半自動構築手法を提案した．提案手法を用いて構築したデータセット JacRED は，初の日本語 DocRE データセットとして公開される．

今後は，忠実性チェックや質問応答などのタスクにおける JacRED の活用法について調べていきたい．

謝辞

この成果は, 国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）の委託業務（JPNP18002）の結果得られたものです.

参考文献

[1] Yuan Yao, Deming Ye, Peng Li, Xu Han, Yankai Lin, Zhenghao Liu, Zhiyuan Liu, Lixin Huang, Jie Zhou, and Maosong Sun. DocRED: A large-scale document-level relation extraction dataset. In **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**, pp. 764–777, July 2019.

[2] Mo Yu, Wenpeng Yin, Kazi Saidul Hasan, Cicero dos Santos, Bing Xiang, and Bowen Zhou. Improved neural relation detection for knowledge base question answering. In **Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics**, pp. 571–581, July 2017.

[3] Bayu Distiawan Trisedya, Gerhard Weikum, Jianzhong Qi, and Rui Zhang. Neural relation extraction for knowledge base enrichment. In **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**, pp. 229–240, July 2019.

[4] Haotian Chen, Bingsheng Chen, and Xiangdong Zhou. Did the models understand documents? benchmarking models for language understanding in document-level relation extraction. In **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics**, pp. 6418–6435, July 2023.

[5] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback, 2022.

[6] OpenAI. Gpt-4 technical report, 2023.

[7] Somin Wadhwa, Silvio Amir, and Byron Wallace. Revisiting relation extraction in the era of large language models. In **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics**, pp. 15566–15589, July 2023.

[8] Wenxuan Zhou, Kevin Huang, Tengyu Ma, and Jing Huang. Document-level relation extraction with adaptive thresholding and localized context pooling. In **Proceedings of the AAAI Conference on Artificial Intelligence**, 2021.

[9] Qingyu Tan, Lu Xu, Lidong Bing, Hwee Tou Ng, and Sharifah Mahani Aljunied. Revisiting DocRED - addressing the false negative problem in relation extraction. In **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing**, pp. 8472–8487, December 2022.

[10] Yang Chen, Chao Jiang, Alan Ritter, and Wei Xu. Frustratingly easy label projection for cross-lingual transfer. In **Findings of the Association for Computational Linguistics: ACL 2023**, pp. 5775–5796, July 2023.

[11] Leonhard Hennig, Philippe Thomas, and Sebastian Möller. MultiTACRED: A multilingual version of the TAC relation extraction dataset. In **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics**, pp. 3785–3801, July 2023.

[12] Qiao Cheng, Juntao Liu, Xiaoye Qu, Jin Zhao, Jiaqing Liang, Zhefeng Wang, Baoxing Huai, Nicholas Jing Yuan, and Yanghua Xiao. HacRED: A large-scale relation extraction dataset toward hard cases in practical applications. In **Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021**, pp. 2819–2831, August 2021.

[13] Soyoung Yang, Minseok Choi, Youngwoo Cho, and Jaegul Choo. HistRED: A historical document-level relation extraction dataset. In **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics**, pp. 3207–3224, July 2023.

[14] Yiqing Xie, Jiaming Shen, Sha Li, Yuning Mao, and Jiawei Han. Eider: Empowering document-level relation extraction with efficient evidence extraction and inference-stage fusion. In **Findings of the Association for Computational Linguistics: ACL 2022**, pp. 257–268, May 2022.

[15] Yuxin Xiao, Zecheng Zhang, Yuning Mao, Carl Yang, and Jiawei Han. SAIS: Supervising and augmenting intermediate steps for document-level relation extraction. In Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz, editors, **Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 2395–2409, July 2022.

[16] Youmi Ma, An Wang, and Naoaki Okazaki. DREEAM: Guiding attention with evidence for improving document-level relation extraction. In **Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics**, pp. 1971–1983, May 2023.

[17] Satoshi Sekine and Hitoshi Isahara. IREX: IR & IE evaluation project in Japanese. In **Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00)**, May 2000.

[18] Nobuhiro Ueda, Kazumasa Omura, Takashi Kodama, Hirokazu Kiyomaru, Yugo Murawaki, Daisuke Kawahara, and Sadao Kurohashi. KWJA: A unified Japanese analyzer based on foundation models. In **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics**, pp. 538–548, July 2023.

[19] Mike Mintz, Steven Bills, Rion Snow, and Daniel Jurafsky. Distant supervision for relation extraction without labeled data. In **Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP**, pp. 1003–1011, August 2009.

[20] Denny Vrandečić and Markus Krötzsch. Wikidata: A free collaborative knowledgebase. **Commun. ACM**, Vol. 57, No. 10, p. 78–85, sep 2014.

[21] Zhiyi Song, Ann Bies, Stephanie Strassel, Tom Riese, Justin Mott, Joe Ellis, Jonathan Wright, Seth Kulick, Neville Ryant, and Xiaoyi Ma. From light to rich ERE: Annotation of entities, relations, and events. In **Proceedings of the The 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation**, pp. 89–98, June 2015.

[22] Ningyu Zhang, Xiang Chen, Xin Xie, Shumin Deng, Chuanqi Tan, Mosha Chen, Fei Huang, Luo Si, and Huajun Chen. Document-level relation extraction as semantic segmentation. In Zhi-Hua Zhou, editor, **Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21**, pp. 3999–4006, 8 2021. Main Track.

[23] Qingyu Tan, Ruidan He, Lidong Bing, and Hwee Tou Ng. Document-level relation extraction with adaptive focal loss and knowledge distillation. In **Findings of the Association for Computational Linguistics: ACL 2022**, pp. 1672–1681, May 2022.

[24] Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. brat: a web-based tool for NLP-assisted text annotation. In **Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics**, pp. 102–107, April 2012.

[25] Junpeng Li, Zixia Jia, and Zilong Zheng. Semi-automatic data enhancement for document-level relation extraction with distant supervision from large language models. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, **Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing**, pp. 5495–5505, December 2023.

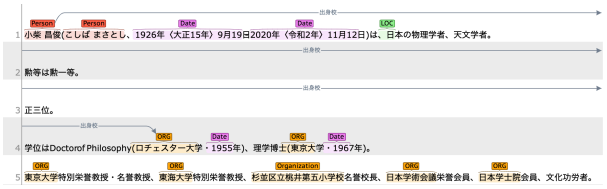

図 4　アノテーション画面の例

A　アノテーション作業について

作業プロセス　タスクの複雑性から，アノテーション作業はクラウドソーシングを用いず，専門家に依頼した．アノテータは事前に作成されたガイドラインに基づき，自動抽出されたエンティティ言及および関係事例の修正を行う．作業中は必要に応じて議論とガイドラインの修正を行う．

作業画面　エンティティ言及付与と関係付与両方とも BRAT [24] を用いる．エンティティ言及付与では，文書単位で各エンティティ言及のスパンおよびラベルの正誤を行う．関係付与では，図 4 に示すように，関係は事例単位で提示される．提示された関係が正しい場合，それをサポートする根拠を文 ID としてコメント欄に記入する．

B　既存データセットとの比較

英語データセット　DocRE のベンチマークとして一般的に使われてきたのは DocRED [1] である．しかし DocRED におけるアノテーション漏れが問題視されており，それを軽減するデータセットとして Re-DocRED が収集された [9]．Re-DocRED は DocRED の欠損関係事例を補完したが，補完対象である関係の根拠文の追加は行われていない．JacRED は (Re-)DocRED と同じくウィキペディアに基づいたデータセットであるため，3 者をより詳しく比較する．表 5 に示したように，JacRED における文書の長さおよび文書ごとのエンティティ数は (Re-)DocRED と匹敵する．一方，JacRED の文書ごとの関係事例数は DocRED より多く，アノテーション漏れの問題はある程度回避できたと考えられる．また，根拠文数から見て，JacRED は Re-DocRED より多く，DocRED に匹敵する．これにより，JacRED は根拠文と関係両方のアノテーションに配慮したバランスの良いデータセットであると考えられる．

英語以外のデータセット　(Re-)DocRED の他，中国語データセットとして HacRED [12]，韓国語データセットとして HistRED [13] がある．両データセットとも独立で収集され，(Re-)DocRED と異なるドメインとラベルセットを持っている．それに対し，JacRED は (Re-)DocRED を起点とするため，(Re-)DocRED と性質が近い．また，JacRED の収集過程で，翻訳および言語横断ラベル射影による他言語 DocRE データセットの自動構築だけでは不十分であるが，自動構築データセットを用いた半自動構築はアノテータの負担軽減に繋がることを示した．

表 5　(Re-)DocRED と JacRED の比較

	DocRED	Re-DocRED	JacRED
文数/文書数	7.98	7.98	8.39
実体数/文書数	19.51	19.45	17.87
関係数/文書数	12.45	29.77	21.12
根拠文数/関係数	1.60	0.88	1.67

C　プロンプト

§4.2 で報告された大規模言語モデルの in-context learning に用いるプロンプトの詳細を図 5 に示す．既存研究では全ての関係を一斉に出力するようにプロンプトを設計してきた [7, 25]．しかし文書における全関係の一斉出力は困難であるため，本稿では関係ラベルごとにプロンプトを設計し，一回の API 呼び出しに関係を一種類だけ抽出する．これにより GPT-3.5 は既存研究を上回る関係抽出性能を示した．

```
Perform Document-level Relation Extraction task. Given a context and an entity list, identify all entity pairs with relation type {located in the administrative territorial entity} in the context. Note that only a few entity pairs hold relations. Please return entity pairs as {head, tail} and make sure they follow the relation definition:

located in the administrative territorial entity: {head} is located in the administrative territorial entity {tail}.

###

Context: 東京・板橋出身。

Entity List: 東京||板橋

Extracted Entity Pairs: {板橋, 東京}

###

Context: 南都六宗(なんとろくしゅう、なんとりくしゅう)とは、奈良時代、平城京を中心に栄えた日本仏教の6つの宗派の総称。三論宗(さんろんしゅう、中論・十二門論・百論)-華厳宗や真言宗に影響を与えた成実宗(じょうじつしゅう、成実論)-三論宗の付宗(寓宗)法相宗(ほっそうしゅう、唯識)倶舎宗(くしゃしゅう、説一切有部)-法相宗の付宗(寓宗)華厳宗(けごんしゅう、華厳経)律宗(りっしゅう、四分律)-真言律宗等が生まれたなお、奈良時代当時から「南都六宗」と呼ばれていたわけではなく、平安時代以降平安京を中心に栄えた「平安二宗」(天台宗・真言宗)に対する呼び名である。

Entity List: 奈良時代||平安時代||平城京||日本||平安京||平安

Extracted Entity Pairs: {平安京, 日本}

###

(examples)

###

Context: アンソニー世界を駆ける(アンソニーせかいをかける)は、アメリカ合衆国のCNNで放送されているテレビ番組。2013年4月から放送を開始した。エミー賞を4回受賞、また、脚本賞、音響賞、編集賞、撮影賞に11回ノミネートされている。また2013年にはアメリカのテレビ・ラジオ・ウェブサイトの優れた放送作品に贈られるピーボディ賞を受賞した。自ら料理人であり、ノンフィクション「キッチン・コンフィデンシャル」の著者でもあるアンソニー・ボーディンが世界の津々浦々を旅し、あまり知られていない地域の景観、風俗、食材、料理などを紹介する。

Entity List: アメリカ合衆国||アンソニー世界を駆ける||CNN||2013年4月||エミー賞||2013年||ピーボディ賞||キッチン・コンフィデンシャル||アンソニー・ボーディン

Extracted Entity Pairs:
```

図 5　DocRE に用いるプロンプトの例