

文書レベル関係抽出における根拠認識の統合

Youmi Ma An Wang 岡崎直観

東京工業大学

{youmi.ma@nlp., an.wang@nlp., okazaki@}c.titech.ac.jp

概要

文書レベル関係抽出 (DocRE) は文書中のすべてのエンティティの組の関係を推定するタスクである。この関係抽出に必要な最小限かつ十分な情報を含む文の集合を根拠と呼ぶ。根拠は関係抽出の性能を改善できるが、既存研究では DocRE と根拠認識を別々のタスクとしてモデル化していた。本稿では、根拠認識を関係抽出のモデルに統合する手法を提案する。さらに、根拠のアノテーションが付与されていないデータに根拠の疑似的な教師信号を付与し、大量の自動ラベル付けデータを活用する方法を提案する。提案手法は DocRED において DocRE と根拠認識の両方で現時点の世界最高性能を達成した¹⁾。

1 はじめに

関係抽出はテキストにおけるエンティティ (実体) の関係を認識するタスクである。関係抽出では、文内に閉じて関係を認識するタスク設定の研究が多く [1, 2, 3, 4, 5], テキスト中で複数の文にまたがって表現される関係は対象外となってしまうため、適用範囲が狭いという課題があった [6]。これに対して、複数の文で言及される関係にも対応したタスク、すなわち**文書レベル関係抽出 (DocRE: Document-level Relation Extraction)**が提案された [6, 7, 8]。DocRE では、複数の文の情報の取捨選択や統合をしながらエンティティ間の関係を推定する必要がある [9, 10, 11]。

DocRE において情報の取捨選択に用いられるのが**根拠 (evidence)**である。根拠は、関係を推定するために必要最小限かつ十分な情報を含む文の集合と定義され、DocRE で広く用いられるデータセット DocRED [6] でもラベル付けされている。図 1 の例では、*Prince Edmund* と *Blackadder* における関係 *present in work* を認識するための必要最小限かつ十

1) 提案手法の実装を <https://github.com/YoumiMa/dream> で公開している。

The Archbishop

[1] "[The Archbishop](#)" is the third episode of the first series of the BBC sitcom [Blackadder](#) (*The Black Adder*). [2] It is set in [England](#) in the late 15th century, and follows the exploits of the fictitious [Prince Edmund](#) as he is invested as [Archbishop of Canterbury](#) amid a [Machiavellian](#) plot by the King to acquire lands from the [Catholic Church](#). [3] ... [5] [Edmund](#), faced with the threat of assassination, attempts to escape to [France](#) into self-imposed exile; and in a later scene, two drunk knights overhear [King Richard IV](#) exclaiming "Who will rid me of this turbulent priest?" [6] The words attributed to [King Henry II](#) which led to [Becket's](#) death in 1170, and embark on a mission to murder [Edmund](#). [7] ...

Subject: *Prince Edmund*
Object: *Blackadder*

Relation: *present in work*
Evidence: 1,2

図 1 DocRED [6] のアノテーションの例。斜体は関係を予測したいエンティティの言及 (メンション) であり、下線はその他のエンティティの言及である。

分な情報は文 1 と 2 であるため、この関係の根拠は文 1 と 2 とラベル付けされる。既存研究では、DocRE のサブタスクとして**根拠認識**に取り組み、エンティティ組の関係を推定する際に必要な情報の取捨選択を行うことが多かった [6, 9, 10, 12]。

既存研究では、DocRE と根拠認識を別々のタスクとしてモデル化しているため [9, 10, 12], 両タスクの関連性を考慮できない。これに対し、本稿では DocRE と根拠認識のモデルを統合した新手法として、Document-level Relation Extraction with Evidence-guided Attention Mechanism (DREAM) を提案する。DREAM では、根拠を単語や文の重要度に関する情報としてテキストのエンコーダに統合する。具体的には、BERT [13] などの事前学習済み言語モデルのエンコーダにおける自己注意機構 [14] への教師信号として根拠を導入し、根拠に高い重みを配分するように誘導しながら DocRE のモデルを学習する。これにより、文書レベル関係抽出から根拠認識モデルを削除し、パラメータ数の削減や推論時のメモリ使用量の低減も実現できる。

なお、文書レベルの関係アノテーションはコストが高いため、学習データが不足しがちな状況にある。表 1 に示すように、現時点で最大規模のデータセットである DocRED でも、人手でラベルが付

表 1 DocRED [6] データセットの統計情報

項目	人手付与	自動付与
文書数	3,053 + 998 + 1,000	101,873
関係ラベル数	97	97
文数/文書数	8.0	8.1
実体数/文書数	19.5	19.3
言及数/実体数	1.3	1.3
関係数/文書数	12.5	14.8
根拠数/関係数	1.6	—

与された文書は 5,051 件しかない。DocRED ではデータ不足を緩和するため、関係ラベルを Distant Supervision [15] を用いて自動付与しているが、根拠ラベルの自動付与は行われていない。本研究では、提案手法である DREEAM を用いて根拠の疑似的な教師信号を自動的に付与し、大量の自動ラベル付与データを活用する。最終的に、提案手法は DocRED において DocRE と根拠認識の両方で現時点の世界最高性能を達成した。

2 予備知識

2.1 タスクの定義

DocRE の目的は、文 $\mathcal{X}_D = \{x_i\}_{i=1}^{|\mathcal{X}_D|}$ からなる文書 D における全てのエンティティ $\mathcal{E}_D = \{e_i\}_{i=1}^{|\mathcal{E}_D|}$ の組が持つ関係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）を $\mathcal{M}_e = \{m_i\}_{i=1}^{|\mathcal{M}_e|}$ とし、エンティティの組 (e_s, e_o) が持つ全ての関係から成る集合を $\mathcal{R}_{s,o} \subseteq \mathcal{R}$ とする。ただし、 \mathcal{R} は関係ラベルの集合である。 (e_s, e_o) の間に関係 $r \in \mathcal{R}$ が存在する場合、 (e_s, r, e_o) に関する根拠を $\mathcal{V}_{s,r,o} \subseteq \mathcal{X}_D$ で表す。根拠認識の目的は、全ての関係抽出における根拠を認識すること、すなわち $(e_s, r, e_o, \mathcal{V}_{s,r,o})$ を得ることである。

2.2 ベースモデル: ATLOP

提案手法のベースとなる手法 ATLOP [16] を説明する。まず、各エンティティの言及の冒頭と末尾に特殊トークン「*」を挿入したトークン列 $\mathcal{T}_D = \{t_i\}_{i=1}^{|\mathcal{T}_D|}$ を BERT [13] などを用いてエンコードする。エンコードの出力から、トークンの分散表現 $H \in \mathbb{R}^{|\mathcal{T}_D| \times d}$ と自己注意機構の重み²⁾ $A \in \mathbb{R}^{|\mathcal{T}_D| \times |\mathcal{T}_D|}$ を記録し、前者をトークン埋め込み、後者をトークン間の依存度と呼ぶ（ d はエンコードの隠れ層の次元数である）。

次に、エンティティ埋め込みを計算する。言及

2) マルチヘッド自己注意の全ヘッドの平均として計算する。

$\mathcal{M}_e = \{m_i\}_{i=1}^{|\mathcal{M}_e|}$ を持つエンティティ e に対して、埋め込み $h_e \in \mathbb{R}^d$ を $h_e = \log \sum_{i=1}^{|\mathcal{M}_e|} \exp(h_{m_i})$ と計算する。ただし、 $h_{m_i} \in \mathbb{R}^d$ は言及 m_i の開始位置に挿入した特殊トークン「*」の埋め込みである。これにより、エンティティ e_s, e_o の埋め込み $h_s, h_o \in \mathbb{R}^d$ を得る。

さらに、エンティティの組 (e_s, e_o) に対する局所文脈埋め込みを計算する。局所文脈埋め込みは文書 D における全トークンの埋め込みの重み付き和であり、重みは e_s, e_o 両方に対する重要度を反映する。トークンの重要度は、トークン間の依存度 A の強さとする。具体的には、エンティティ e_s の各言及 $m_i \in \mathcal{M}_{e_s}$ と他のトークンの依存度から $a_s \in \mathbb{R}^{|\mathcal{T}_D|}$ を求める。ここで a_s は全ての $a_{m_i} \in \mathbb{R}^{|\mathcal{T}_D|}$ の平均であり、 a_{m_i} は m_i の冒頭にある「*」に対応する A のスライスである。同様に $a_o \in \mathbb{R}^{|\mathcal{T}_D|}$ を求め、 (e_s, e_o) に対する各トークンの重要度 $q^{(s,o)} \in \mathbb{R}^{|\mathcal{T}_D|}$ および局所文脈埋め込み $c^{(s,o)} \in \mathbb{R}^d$ を、式 1 と 2 で求める。

$$q^{(s,o)} = \frac{a_s \circ a_o}{a_s^\top a_o} \quad (1)$$

$$c^{(s,o)} = H^\top q^{(s,o)} \quad (2)$$

ただし \circ は要素ごとの積である。局所文脈埋め込みは重要度の高い箇所に注目した分散表現であることから、異なる文の情報を効果的に統合できる。

最後に、 (e_s, e_o) の関係分類を行う。関係分類器の入力はエンティティ埋め込みと局所埋め込みを統合した分散表現で、 $z_s = \tanh(W_s [h_s; c^{(s,o)}] + b_s) \in \mathbb{R}^d$ と計算する。ここで $[\cdot; \cdot]$ はベクトルの連結、 $W_s \in \mathbb{R}^{d \times 2d}$ 、 $b_s \in \mathbb{R}^d$ はパラメータである。 $z_o \in \mathbb{R}^d$ も同様に計算し、全関係ラベルのスコアを $y^{(s,o)} = z_s^\top W_r z_o + b_r \in \mathbb{R}^{|\mathcal{R}|}$ で計算する。ただし $W_r \in \mathbb{R}^{|\mathcal{R}| \times d \times d}$ 、 $b_r \in \mathbb{R}^{|\mathcal{R}|}$ はパラメータである。エンティティ e_s, e_o の間の関係 $r \in \mathcal{R}$ の確率推定は $P(r|s,o) = \text{sigmoid}(y_r^{(s,o)})$ として計算される。

ATLOP の学習は適応閾値損失 (ATL: Adaptive Threshold Loss) に基づく。すなわち、閾値に対応する仮想的な関係ラベル TH を導入し、エンティティ組 (e_s, e_o) に存在する関係ラベルのスコアは TH より高く、存在しない関係ラベルのスコアは TH より低くなるようにモデルを学習する。

3 提案手法

ATLOP では関係推定に重要なトークンに注目して情報を統合するため、局所文脈埋め込みを採用し、BERT の自己注意機構がエンティティ組 (e_s, e_o) に対するトークンの重要度 $q^{(s,o)}$ を暗黙的に学習で

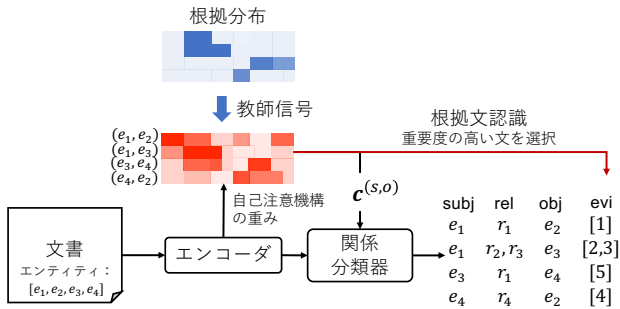


図2 DREAMの概要

きると仮定した。本研究では、根拠を用いて $q^{(s,o)}$ の教師信号を与えることを提案し、根拠認識の関係抽出モデルへの統合、および関係抽出の性能向上を目指す (§ 3.1)。さらに根拠の教師信号を自動で付与し、データ不足の緩和を目指す (§ 3.2)。

3.1 DREAM

提案手法である DREAM の概要を図 2 に示す。自己注意機構から計算されるトークン重要度と根拠から計算される分布が近づくように誘導し、根拠に基づいた局所文脈埋め込みが得られると期待する。

根拠分布の算出 あるエンティティ組 (e_s, e_o) が持つ関係 $r \in \mathcal{R}_{s,o}$ 毎に、根拠ベクトル $\mathbf{v}^{(s,r,o)} \in \{0,1\}^{|\mathcal{X}_D|}$ を定義する。 $\mathbf{v}^{(s,r,o)}$ は二値ベクトルであり、文 $x_i \in \mathcal{X}_D$ が関係 (e_s, r, e_o) の根拠である時に $v_i^{(s,r,o)} = 1$ とする。さらに、全ての関係 $\forall r \in \mathcal{R}_{s,o}$ における根拠ベクトル $\mathbf{v}^{(s,r,o)}$ から、関係ラベルに依存しない根拠の分布 $\mathbf{v}^{(s,o)}$ を次式で求める。

$$\mathbf{v}^{(s,o)} = \frac{\sum_{r \in \mathcal{R}_{s,o}} \mathbf{v}^{(s,r,o)}}{\sum_{r \in \mathcal{R}_{s,o}} \mathbf{1}^T \mathbf{v}^{(s,r,o)}} \quad (3)$$

ただし $\mathbf{1} = (1, 1, \dots, 1)^{|\mathcal{X}_D|}$ である。 $\mathbf{v}^{(s,o)}$ はエンティティ組 (e_s, e_o) に関する文の重要度を全関係ラベル共通に根拠のアノテーションから求めている。

根拠分布を用いた学習 自己注意機構によるトークン重要度 $q^{(s,o)}$ を文の重要度 $\mathbf{p}^{(s,o)} \in \mathbb{R}^{|\mathcal{X}_D|}$ に変換する。トークン列 $t_{\text{START}(x_i)}, \dots, t_{\text{END}(x_i)}$ からなる文 $x_i \in \mathcal{X}_D$ に対し、その重要度を和

$$p_i^{(s,o)} = \sum_{j=\text{START}(x_i)}^{\text{END}(x_i)} q_j^{(s,o)} \quad (4)$$

として計算する。ただし $p_i^{(s,o)}$ は $\mathbf{p}^{(s,o)}$ の i 番目の要素の値である。そして、根拠アノテーションから得られる文の重要度 $\mathbf{v}^{(s,o)}$ を教師信号とし、自己注意機構から計算される文の重要度 $\mathbf{p}^{(s,o)}$ が近くなるように、カルバック・ライブラー距離 (D_{KL}) によ

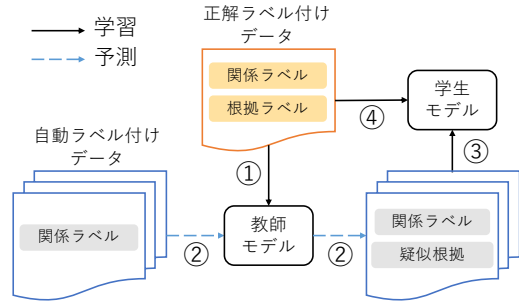


図3 DREAMによる自動ラベル付けデータの活用方法

る損失関数を導入する (式 5)。

$$\mathcal{L}_{\text{ER}} = -D_{\text{KL}}(\mathbf{v}^{(s,o)} \parallel \mathbf{p}^{(s,o)}) \quad (5)$$

モデルのパラメータ推定に用いる損失関数として、根拠抽出の損失関数と式 5 の和を採用することで、関係抽出と根拠認識を同時に学習する。

推論 関係抽出では ATLOP と同様に、適応的閾値を採用する。各エンティティ組 (e_s, e_o) の関係ラベルのスコアを計算し、仮想的な閾値ラベル TH より高いスコアを持つラベルを関係として出力する。根拠認識では静的な閾値 (付録 A 参照) を採用し、 $\mathbf{p}^{(s,o)}$ が閾値より高い文を根拠として出力する。

3.2 根拠の教師信号の自動付与

DREAM ではエンコーダの自己注意機構から計算されるトークン重要度に基づいてエンティティ組の関係の予測に役立つ箇所 (根拠) を推定できる。ところで、1 節で述べたように、DocRED には人手で付与された少量のデータの他に、Distant Supervision で関係ラベルを自動付与した大量のデータがあるが、根拠のラベルが付与されていない。そこで、DREAM から推定される根拠の応用として、DocRE の根拠の教師信号の自動付与を行い、大量の自動ラベル付けデータを活用する方法を提案する。

その概要を図 3 に示す。Tan ら [17] に倣い、教師と学生の二つのモデルを学習する。教師モデルの役割は、根拠の正解ラベル付けデータから根拠認識に関する知識を獲得し、根拠のラベルが無いデータに根拠の疑似的な教師信号を自動付与することである。これは、図 3 のステップ 1 と 2 で実現される。学生モデルは、手動および自動で付与された根拠のラベル付けデータから DocRE と根拠認識のモデルを学習する (図 3 のステップ 3 と 4)。

疑似根拠分布を用いた学習 根拠の正解ラベルで学習した教師モデルを用い、まだ根拠のラベル付けがされていないデータにおけるエンティティ組

(e_s, e_o) のトークンの重要度 $\hat{q}^{(s,o)}$ を予測・記録し、根拠の疑似的な教師信号とする。学生モデルのトークン重要度 $q^{(s,o)}$ を疑似的な根拠 $\hat{q}^{(s,o)}$ に近づけることで、根拠抽出を学習する (式 6)。

$$\mathcal{L}_{ER} = -D_{KL}(\hat{q}^{(s,o)} || q^{(s,o)}) \quad (6)$$

4 実験と考察

4.1 実験結果

DREEAM を DocRED [6] で評価した結果を表 2 に示す。実験設定の詳細は付録 A に示す。

人手のラベル付けデータだけでモデルを学習する場合 DREEAM は EIDER [10] に匹敵する性能を示した。DREEAM と SAIS [12] の間では性能の差があるが、これはマルチタスク学習によるものと考えられる³⁾。なお、DREEAM は根拠認識を DocRE のエンコーダに統合したため、根拠認識モデルが不要となり、学習・予測時のメモリ使用量を大幅に削減できた (詳細な結果は付録 B を参照されたい)。

人手に加えて自動でラベル付けしたデータも用いてモデルを学習する場合 提案手法 (DREEAM*) は関係抽出において SSAN* [18] や KD-DocRE* [17] の性能を上回り、このデータセットにおける最高性能を達成した。これらの既存研究も自動でラベル付けした訓練データを用いているため、提案手法の優位性が示唆される。また、根拠認識においても、提案手法は既存手法の最高性能モデル SAIS [12] を大きく上回った。これにより、提案した根拠の教師信号の自動付与手法が DocRE と根拠認識の両方で有効であることが示された。

4.2 分析

本研究の提案事項の貢献度合いを調べるため、アブレーション実験を行った結果を表 3 に示す⁴⁾。

教師モデル 教師モデルを学習する際に根拠認識の学習 (式 5) を無効化し、性能に与える影響を調べた。表 3 の通り、関係抽出と根拠認識の両方において性能の低下が見られる。ゆえに、DocRE のエンコーダの自己注意機構を根拠で誘導することは、関係抽出にも有益であることが示唆される。

3) SAIS は DocRE と根拠認識以外にも、共参照解析等の多タスクでモデルを学習し、豊富な教師信号を活用している。

4) 表 2 の DREEAM* と表 3 (b) の DREEAM は同じ実験設定であるが、表 2 の実験では RoBERTa_{large} [19] を、表 3 の実験では BERT_{base} [13] をエンコーダとして用いているため、性能に差が現れる。

表 2 DocRED での関係抽出と根拠認識の性能。*が付いている行は自動でラベル付けしたデータを学習に用いた。

モデル	検証データ		評価データ	
	関係 F1	根拠 F1	関係 F1	根拠 F1
SSAN [18]	60.25	—	59.47	—
SSAN [18]*	63.76	—	63.78	—
ATLOP [16]	61.32	—	61.39	—
EIDER [10]	62.34	52.54	62.85	53.01
SAIS [12]	62.23	55.84	63.44	55.67
KD-DocRE [17]*	65.27	—	65.24	—
DREEAM	62.29	54.15	62.12	54.01
DREEAM*	65.52	57.55	65.47	57.34

表 3 DocRED の検証データでのアブレーション実験設定

設定	関係 F1	根拠 F1
(a) 教師モデル (学習データの文書数は 3,053 件)		
DREEAM	59.60	52.08
根拠認識を学習しない	59.21	42.79
(b) 学生モデル (学習データの文書数は 104,926 件)		
DREEAM	63.47	55.68
疑似根拠で根拠認識を学習しない	61.96	53.72
正解根拠で根拠認識を学習しない	63.34	55.27
根拠認識の学習を全く行わない	62.13	47.13

学生モデル 疑似的な教師信号による根拠認識の学習 (式 6)、および手動でラベル付けしたデータによる根拠認識の学習 (式 5) の片方および両方を無効化し、タスクの性能に与える影響を調べる。表 3 から、疑似根拠による根拠認識の学習を無効化すると、DocRE と根拠認識ともに性能が低下することが分かる。一方、人手でラベル付けした根拠データによる根拠認識の学習を無効にしても、性能への影響は限定的であった。これにより、根拠認識に関する知識が教師モデルから学生モデルに期待通り引き継がれていることが示唆される。

5 おわりに

本稿では、文書レベル関係抽出において根拠を活用する手法を提案した。関係抽出器のエンコーダの自己注意機構において、関係予測の根拠に高い重みを配分するように誘導する DREEAM を提案した。続いて、DocRE におけるデータ不足問題を緩和するため、根拠ラベルの持たないデータを活用する手法を提案した。DocRED データセットで評価したところ、提案手法は関係抽出および根拠認識の両方のタスクで現時点での世界最高性能を達成した。

今後は、根拠認識による関係抽出の説明可能性の検討や、長い文書において情報の統合や取捨選択を要する他タスクへの DREEAM の適用、日本語の文書レベル関係抽出の研究に取り組みたい。

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] George Doddington, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel, and Ralph Weischedel. The automatic content extraction (ACE) program – tasks, data, and evaluation. In **Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)**, Lisbon, Portugal, May 2004. European Language Resources Association (ELRA).
- [2] Xu Han, Hao Zhu, Pengfei Yu, Ziyun Wang, Yuan Yao, Zhiyuan Liu, and Maosong Sun. FewRel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing**, pp. 4803–4809, Brussels, Belgium, October–November 2018. Association for Computational Linguistics.
- [3] Iris Hendrickx, Su Nam Kim, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Sebastian Padó, Marco Pennacchiotti, Lorenza Romano, and Stan Szpakowicz. SemEval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. In **Proceedings of the 5th International Workshop on Semantic Evaluation**, pp. 33–38, Uppsala, Sweden, July 2010. Association for Computational Linguistics.
- [4] Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D. Manning. Position-aware attention and supervised data improve slot filling. In **Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing**, pp. 35–45, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.
- [5] Christoph Alt, Aleksandra Gabryszak, and Leonhard Hennig. TACRED revisited: A thorough evaluation of the TACRED relation extraction task. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pp. 1558–1569, Online, July 2020. Association for Computational Linguistics.
- [6] 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, Florence, Italy, July 2019. Association for Computational Linguistics.
- [7] Jiao Li, Yueping Sun, Robin J. Johnson, Daniela Sciaky, Chih-Hsuan Wei, Robert Leaman, Allan Peter Davis, Carolyn J. Mattingly, Thomas C. Wieggers, and Zhiyong Lu. BioCreative V CDR task corpus: a resource for chemical disease relation extraction. **Database**, Vol. 2016, , 05 2016. baw068.
- [8] Patrick Verga, Emma Strubell, and Andrew McCallum. Simultaneously self-attending to all mentions for full-abstract biological relation extraction. In **Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)**, pp. 872–884, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.
- [9] Quzhe Huang, Shengqi Zhu, Yansong Feng, Yuan Ye, Yuxuan Lai, and Dongyan Zhao. Three sentences are all you need: Local path enhanced document relation extraction. In **Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)**, pp. 998–1004, Online, August 2021. Association for Computational Linguistics.
- [10] 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, Dublin, Ireland, May 2022. Association for Computational Linguistics.
- [11] Wang Xu, Kehai Chen, Lili Mou, and Tiejun Zhao. Document-level relation extraction with sentences importance estimation and focusing. In **Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 2920–2929, Seattle, United States, July 2022. Association for Computational Linguistics.
- [12] Yuxin Xiao, Zecheng Zhang, Yuning Mao, Carl Yang, and Jiawei Han. SAIS: Supervising and augmenting intermediate steps for document-level relation extraction. In **Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 2395–2409, Seattle, United States, July 2022. Association for Computational Linguistics.
- [13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [14] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, **Advances in Neural Information Processing Systems**, Vol. 30. Curran Associates, Inc., 2017.
- [15] 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, Suntec, Singapore, August 2009. Association for Computational Linguistics.
- [16] 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.
- [17] 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, Dublin, Ireland, May 2022. Association for Computational Linguistics.
- [18] Benfeng Xu, Quan Wang, Yajuan Lyu, Yong Zhu, and Zhendong Mao. Entity structure within and throughout: Modeling mention dependencies for document-level relation extraction. **Proceedings of the AAAI Conference on Artificial Intelligence**, Vol. 35, No. 16, pp. 14149–14157, May 2021.
- [19] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Ro{bert}a: A robustly optimized {bert} pre-training approach, 2020.
- [20] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In **International Conference on Learning Representations**, 2019.

A 実験設定

4節の実験設定を説明する。DREEAMはATLOP [16]をベースとしているが、トークン埋め込み H とトークン間依存度 A の計算方法が異なる。2.2節で述べたように、ATLOPはエンコーダの出力層のみから H と A を計算するのに対し、DREEAMはエンコーダの上部3層の平均で H と A を求める。

モデルの学習 根拠認識がDocREのモデルに与える影響を調整するため、損失関数 \mathcal{L}_{ER} に係数をかけてから関係抽出の損失との和を計算し、全体の損失関数とした。具体的には、BERT_{base}を用いたモデルでは係数を0.1、RoBERTa_{large}を用いたモデルでは係数を0.05とした。なお、パラメータの更新にはAdam [20]を採用した。その他のハイパー・パラメータの値を表4に示す。

モデルによる関係抽出と根拠認識 関係抽出では、各エンティティの組 (e_s, e_o) における閾値ラベル TH より高い関係ラベルを予測結果とする。根拠認識では、重要度 $p_i^{(s,o)}$ が0.2より高い文を予測とする。さらに、根拠認識の結果を活用するため、予測融合を行う [10]。具体的には、予測結果 $(e_s, r, e_o, \mathcal{V}_{s,r,o})$ 毎に、根拠 $\mathcal{V}_{s,r,o}$ のみからなる部分文書 $\hat{D}_{s,r,o}$ を構成し、関係抽出を行う。最終的な関係抽出結果は、元文書 D と全ての部分文書における予測から算出する。詳細な説明は原著論文 [10] を参照されたい。評価指標はDocREDに従い、関係抽出と根拠認識のF1スコアとする。本稿では、検証・評価データから訓練データとの重複を取り除いたうえで、F1スコアを計測している [6]。なお、全てのスコアは、5つの異なるランダムシードを用いてモデルを初期化し、学習した結果の平均である。

B パラメータ数及びメモリ使用量

従来研究では、根拠認識に特化したモデルを設計し、DocREとは別に根拠認識器を適用することが多い。このとき、各文がエンティティ組の関係判断の根拠となる確率を計算するため、全ての文とエンティティ組を列挙しなければならない。例えばEIDER [10]では、文 x_i がエンティティの組 (e_s, e_o) の関係予測の根拠である確率を式7で求める。

$$P(x_i|e_s, e_o) = \text{sigmoid}(x_i^T Wc^{(s,o)} + b), \quad (7)$$

ただし $x_i \in \mathbb{R}^d$ はトークン埋め込みから得た文 x_i の埋め込み、 $c^{(s,o)} \in \mathbb{R}^d$ は (e_s, e_o) における局所文脈埋

表4 実験に用いたハイパー・パラメータ

	学習 (教師)	学習 (学生)	微調整 (学生)
エンコーダ：BERT_{base}			
エポック数	30	2	10
学習率 (エンコーダ)	5e-5	3e-5	1e-6
学習率 (関係分類器)	1e-4	1e-4	3e-6
最大勾配ノルム	1.0	5.0	2.0
エンコーダ：RoBERTa_{large}			
エポック数	30	5	10
学習率 (エンコーダ)	3e-5	1e-5	1e-6
学習率 (関係分類器)	1e-4	5e-5	3e-6
最大勾配ノルム	1.0	5.0	2.0

表5 パラメータ数及び推論時のメモリ使用量

モデル	パラメータ数 (M)	メモリ使用量 (GiB)
根拠認識の学習なし		
ATLOP [16]	115.4	10.8
SSAN [18]	113.5	6.9
KD-DocRE [17]	200.1	15.2
根拠認識の学習あり		
EIDER [10]	120.2	43.1
SAIS [12]	118.0	46.2
DREEAM	115.4	11.8

め込みであり、 $W \in \mathbb{R}^{d \times d}$ と $b \in \mathbb{R}$ はパラメータである。例えば n 個の文からなる文書に m 個のエンティティがあるとすると、式7の演算を $n \times m \times (m-1)$ 回行うため、特に n や m が大きい場合は、膨大なメモリを消費する。

一方、DREEAMは根拠認識をエンティティの組のエンコーダに組み込み、根拠はエンコーダから計算した文の重要度を用いて直接推定できるため、根拠認識器を外部に用意する必要がない。よって、式7による確率の計算が不要となり、パラメータ W と b を導入せずに済む。

DREEAMと既存手法のパラメータ数およびメモリ使用量を表5に示す。エンコーダはBERT_{base}に揃えた。DREEAMを根拠認識モデルの学習を行う既存手法と比較すると、モデルのパラメータ数は既存研究よりも低く抑えられることが確認できる。また、推論時のメモリ使用量に関しては、DREEAMはEIDER [10]の27.4%、SAISの25.5%まで抑えることが可能で、既存手法に対して大きなアドバンテージとなる。なお、根拠認識モデルを学習しないKD-DocRE [17]と比較しても、DREEAMのパラメータ数およびメモリ使用量は少ないことが分かる。以上のことから、DREEAMは既存手法よりも少ないパラメータ数およびメモリ使用量で、既存手法の同等以上の性能を達成できることが分かる。