

ID レベル関係抽出における不要な文の自動選択

辻村 有輝 三輪 誠 佐々木 裕
豊田工業大学

{sd18602,makoto-miwa,yutaka.sasaki}@toyota-ti.ac.jp

概要

言及を超えた文書に現れるエンティティ間の関係を、エンティティに対応するデータベース上の ID 間の関係として抽出する ID レベル関係抽出が取り組まれ始めている。ID により文書を超えた利用性の高い関係情報を抽出できる一方で、入力ノイズの増大が問題となる。本研究では、関係抽出対象のエンティティ ID ペアに対する入力文書中の各文の重要性を判断し、不要な文をマスクする ID レベル関係抽出モデルを提案する。さらに固有表現抽出とエンティティリンキングを組み合わせることでテキストのみから ID レベルの関係を抽出するパイプラインシステムを実現し、その性能を明らかにする。

1 はじめに

関係抽出は文書中に出現するエンティティ同士の意味的な関係を抽出するタスクであり、データベース拡張や下流タスクでの利用性から広く研究されている [1, 2, 3]。既存研究では文書中に出現するエンティティの特定の言及スパン同士の間で成立する関係を抽出する言及レベルの関係抽出が広く取り組まれてきた [1, 2]。一方で、文書中には同一エンティティを表す言及が複数出現することが多く、文書の意味をよく捉えた関係抽出のためには、文書中の同一エンティティについて情報の統合が必要である。

そこで、近年、エンティティ間の関係性を取り扱う研究が取り組まれ始めている [3]。このような研究では、各エンティティに対する言及を専門家が整備する用語データベース上の ID にリンクすることによって概念的な同一性の粒度を取り扱う、ID レベル関係抽出のタスク設定が一般的である。ID レベル関係抽出における抽出結果は $(ID_{Arg1}, relation, ID_{Arg2})$ のトリプルの形で表現される。エンティティ各項がすでにデータベース上の ID とリンクされているため、言及レベルの関係抽出

に比べて表記ゆれの問題がなく、文書を超えた利用が可能であり、データベースの参照や拡張に直接利用できたり、他の下流タスクでの利用時に ID による参照できたりするなど、応用上のメリットも大きい。

このようなメリットの一方で、ID レベル関係抽出は言及レベルの関係抽出に比べて、入力として利用する文脈が必然的に長くなり、その結果、入力中に含まれる不要な情報も増えるという問題点がある。関係抽出の性能向上には、関係の判断に必要な情報の取捨選択が重要であることが知られており、例えば単文上の言及レベル関係抽出では、ルールベースでの文の構文木上の最短経路を用いた重要単語の選択手法の有効性が知られている [1]。この手法は、観察によって得られたヒューリスティックに基づいており、重要な単語が選択できないケースが存在するため、著者らはルールの代わりに学習による重要単語の選択手法を提案した [2]。

本研究ではこの学習ベースの重要単語の選択手法を ID レベル関係抽出に拡張し、入力文中の不要な文を自動選択する ID レベル関係抽出モデルを提案する。提案モデルは関係抽出対象のエンティティ ID ペアと各文の表現ベクトルから、その文の重要度スコアを計算し、この重要度スコアに基づいて不要な文を離散的にマスクする。マスクの基準はタスク損失から end-to-end に学習する。

実験では生化学分野の論文概要上での ID レベル関係抽出を取り扱う BioRED データセット [3] を用いて、提案モデルの評価を行った。実験結果から、提案モデルの性能は recall に比べて precision が高くなる傾向が確認された。また本研究では、提案モデルを用いたパイプライン情報抽出システムの構築と評価も行った。このパイプラインシステムは、関係抽出に加えて固有表現抽出とエンティティリンキングを実施することで、テキストのみを入力とした ID レベルの情報抽出を実現した。

本研究の主な貢献は以下の2点である。

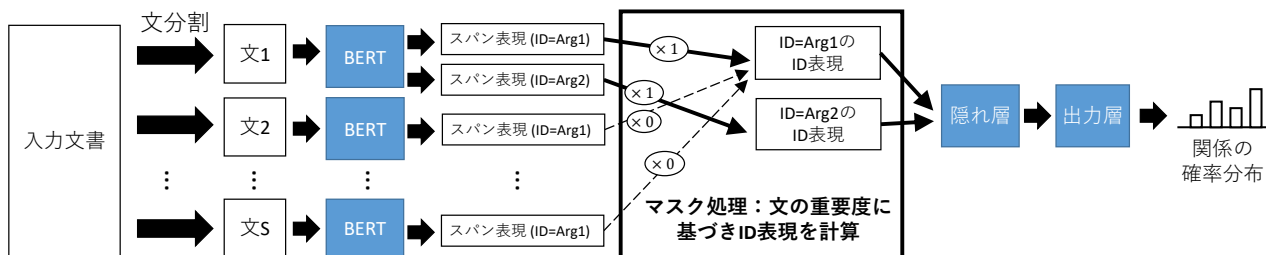


図 1 関係抽出モデルの外観. 本モデルは文分割された文書と対象のエンティティ ID ペアに属する全言及スパン位置を入力とし, 対象のペアを関係クラスに分類する.

- 不要な文の判断をタスク損失から end-to-end に学習する, 入力中の不要な文をマスクする ID レベル関係抽出モデルを提案し, BioRED データセットで評価した. 提案モデルは不要な文をマスクしない場合に比べて, precision が高くなることを確認した.
- テキストのみの入力から ID レベルの関係を抽出するパイプライン情報抽出システムを構築し, BioRED データセットで評価した. 最終的な関係抽出の性能として, システムは 32.2% のテスト F 値を記録した.

2 関連研究

2.1 情報抽出

情報抽出はテキストから構造化された情報を抽出する自然言語処理技術である [1]. 情報抽出によって抽出される情報の多くは文書中に出現するエンティティとその間に成立する関係の形で構成される. 情報抽出のサブタスクとしては, テキストからエンティティの言及スパンを取り出す固有表現抽出, 同一概念のクラスタリングを行う共参照解析, 各言及スパンについて, それが意味するエンティティに対応する外部 ID を識別するエンティティリンキング, エンティティ間の関係を抽出する関係抽出などがある.

2.2 BioRED

BioRED は文書に対する ID レベル関係抽出データセットである [3]. 各文書は PubMed¹⁾ に登録されている生化学分野の論文概要であり, 6 クラスのエンティティと 8 クラスの関係から構成される. エンティティのラベルは各言及スパンの位置とそのクラス, およびデータベース上の ID である. リンク先となるデータベースはエンティティクラスごとに異

1) <https://pubmed.ncbi.nlm.nih.gov/>

なる. 関係は各 ID ペアごとに高々 1 つのラベルがつく, シングルラベル設定であり, 負例ペアも存在する. データセットの分割として公式の訓練 (400 文書), 開発 (100 文書), テストセット (100 文書) が提供されている. 各正例関係にはその関係がその論文中で新たに見つかったものを表す新規性ラベルが付与されているが, 本研究では新規性ラベルは使用しない.

2.3 文レベル関係抽出における重要箇所

著者らは, 単文からの言及レベル関係抽出において, 学習を通じて重要単語を自動で選択する手法を提案した [2]. 選択の基準は Gumbel-Softmax サンプルリング [4] を利用することで, 離散的な振る舞いを表現し, タスク損失から end-to-end に学習した.

3 提案手法

3.1 ID レベル関係抽出モデル

文書からの関係抽出の入力中には, 抽出対象のエンティティ ID ペアに無関係な記述が多く存在する. また, 対象 ID ペアに関連する記述ではあるが, 関係については言及していない記述も存在し, そのような不要な情報は関係抽出の性能に悪影響を及ぼすと考えられる. そこで, 各文について, その文の対象エンティティ ID ペアに対する重要度を判断し, 不要な文をマスクするマスク処理を備えた ID レベルの関係抽出モデルを提案する. 不要な文の判定はタスク損失から end-to-end に学習され, Gumbel-Softmax サンプルリングによって離散的なマスク処理の学習を実現する. 提案モデルの外観を図 1 に示す.

提案モデルは, 文分割された文書と対象のエンティティ ID ペアに属する全言及スパン位置を入力とし, そのペア間の関係クラスを出力する分類モデルである. 初めに各文を BERT [5] によってエ

ンコードし、BERT 最終層の隠れベクトルとして各トークンの表現ベクトルを得る。続いてエンティティの言及スパンごとに、そのスパンに含まれるトークンの表現ベクトルに対して平均プーリングを行うことで、スパン表現ベクトルを計算する。

$$\mathbf{v}_{i,j}^{\text{span}} = \frac{1}{|\mathbf{M}_{ij}|} \sum_{t \in \mathbf{M}_{ij}} \mathbf{v}_{i,t}^{\text{token}}, \quad (1)$$

ここで $\mathbf{v}_{i,j}^{\text{span}}$ は文書中の i 番目の文において j 番目に出現する言及のスパン表現であり、 \mathbf{M}_{ij} はスパンに含まれるトークンインデックスの集合、 $\mathbf{v}_{i,t}^{\text{token}}$ は i 番目の文の t 番目のトークンに対する BERT の最終層の隠れベクトルである。

次に、対象 ID ペアに対する i 番目の文の重要度スコア c_i^{mask} を計算する。重要度スコアは、文中のトークンの表現ベクトルの平均から計算される文表現と、対象 ID ペアそれぞれのエンティティの言及スパン表現の平均から計算されるエンティティの表現から計算される。

$$\begin{aligned} \mathbf{v}_i^{\text{sent}} &= \frac{1}{T_i} \sum_{t=1}^{T_i} (\mathbf{v}_{i,t}^{\text{token}}) \\ \{\mathbf{v}_1^{\text{lstm}}, \mathbf{v}_2^{\text{lstm}}, \dots, \mathbf{v}_S^{\text{lstm}}\} &= \text{LSTM}(\mathbf{v}_1^{\text{sent}}, \mathbf{v}_2^{\text{sent}}, \dots, \mathbf{v}_S^{\text{sent}}) \\ \mathbf{v}_i^{\text{csent}} &= (\mathbf{W}^{\text{csent}} \mathbf{v}_i^{\text{lstm}} + \mathbf{b}^{\text{csent}}) + \mathbf{v}_i^{\text{sent}} \\ \mathbf{v}_i^{\text{key}} &= \mathbf{W}^{\text{key}} \mathbf{v}_i^{\text{csent}} \\ \mathbf{v}_k^{\text{ID}} &= \frac{1}{|\mathbf{E}_k|} \sum_{(i,j) \in \mathbf{E}_k} (\mathbf{v}_{i,j}^{\text{mention}}) \\ \mathbf{v}^{\text{query}} &= \mathbf{W}^{\text{query}} \left[\mathbf{v}_{\text{Arg1}}^{\text{ID}}, \mathbf{v}_{\text{Arg2}}^{\text{ID}} \right] \\ c_i^{\text{mask}} &= \sigma \left(\frac{\mathbf{v}^{\text{query}} \mathbf{T} \mathbf{v}_i^{\text{key}}}{\sqrt{\dim(\mathbf{v}^{\text{query}})}} \right), \quad (2) \end{aligned}$$

ここで $\mathbf{v}_i^{\text{sent}}$ が i 番目の文に対する文表現であり、 T_i は i 番目の文の文長を、 S は文書中の文数を表す。LSTM は双方向の出力を結合する双方向 LSTM、 $[\mathbf{v}_1; \mathbf{v}_2]$ はベクトル \mathbf{v}_1 と \mathbf{v}_2 の結合、 σ はシグモイド関数である。 $\dim(\mathbf{v})$ はベクトル \mathbf{v} の次元数を表す。 \mathbf{E}_k はエンティティ ID が k であるエンティティに対応する言及スパンが属する文インデックスとスパンインデックスのペア集合、 \mathbf{v}_k^{ID} はエンティティ ID が k であるエンティティの表現、Arg1 および Arg2 が関係抽出対象のエンティティ ID ペアのそれぞれを表す。 $\mathbf{W}^{\text{csent}}$, \mathbf{W}^{key} , $\mathbf{W}^{\text{query}}$ と $\mathbf{b}^{\text{csent}}$ はそれぞれ重み行列とバイアスである。

そして、計算された重要度スコア c_i^{mask} に基づき、 i 番目の文のマスク m_i^{mask} を決定する。学習時のマ

スクは、Gumbel-Softmax サンプリングを利用することで、離散的な値に近いマスクを得る。

$$m_i^{\text{mask}} = \text{Gumbel}_2 \left(c_i^{\text{mask}} \right), \quad (3)$$

ここで Gumbel_2 は 2 クラスの Gumbel-Softmax サンプリングである。このサンプリングは微分可能であり、勾配法によって重要度スコアに影響するパラメータを学習する。テスト時のマスクは重要度スコアが 0.5 を超える場合 1、超えない場合 0 として決定的に計算する。

最後に、計算されたマスクに基づいてエンティティの表現を計算し、関係を予測する。

$$\begin{aligned} \mathbf{v}_k^{\text{mID}} &= \frac{\sum_{(i,j) \in \mathbf{E}_k} (m_i^{\text{mask}} \mathbf{v}_{i,j}^{\text{span}})}{\sum_{(i,j) \in \mathbf{E}_k} m_i^{\text{mask}}} \\ \mathbf{v}^{\text{h}} &= \tanh \left(\mathbf{W}^{\text{h}} \left(\left[\mathbf{v}_{\text{Arg1}}^{\text{mID}}, \mathbf{v}_{\text{Arg2}}^{\text{mID}} \right] + \mathbf{v}^{\text{noise}} \right) + \mathbf{b}^{\text{h}} \right) \\ \mathbf{p} &= \text{softmax} \left(\mathbf{W}^{\text{o}} \mathbf{v}^{\text{h}} + \mathbf{b}^{\text{o}} \right), \quad (4) \end{aligned}$$

$\mathbf{v}_k^{\text{mID}}$ は ID が k であるエンティティの言及のスパン表現をその言及の属する文の重要度を考慮して統合したエンティティの表現であり、関係予測では対象エンティティ ID ペアの各エンティティ表現 $\mathbf{v}_{\text{Arg1}}^{\text{mID}}$ と $\mathbf{v}_{\text{Arg2}}^{\text{mID}}$ を入力として分類を行う。 \tanh と softmax はそれぞれ \tanh 関数とソフトマックス関数を表す。 \mathbf{W}^{h} , \mathbf{W}^{o} と \mathbf{b}^{h} , \mathbf{b}^{o} はそれぞれ重み行列とバイアスである。出力 \mathbf{p} は対象エンティティペアに対する関係ラベルの確率分布である。

最終的な予測としては、 \mathbf{p} において最大の確率となった関係クラスを選ぶ。 $\mathbf{v}^{\text{noise}}$ はマスクされなかった文数が多いほど分散が大きくなるノイズであり、不要な情報を多く利用するほど入力がノイズに晒されることをモデルしている。この設定により、ノイズの予測への影響を低減させるために使用する文数を減らすよう訓練されることを期待している。ノイズの計算式は付録 A に示す。テスト時にはノイズは使用しない。

3.2 パイプライン情報抽出システム

本研究では、提案する関係抽出に、固有表現抽出とエンティティリンキングを組み合わせた、パイプライン ID レベル情報抽出システムも構築した。本システムは、初めに文書に対して固有表現抽出を実行し、言及スパンを抽出する。続いて各スパンにエンティティリンキングを行い ID を識別する。最後に、抽出されたスパンと ID から提案モデルによって ID レベルの関係を抽出する。

表 1 “Chemical” および “Disease” エンティティのみからなる ID ペアを対象とした関係抽出の評価。結果はそれぞれ micro 平均スコアの 5 回の評価の平均値である。Luo らは全エンティティクラスのペアに対して実験を行っており、本研究で対象とする 2 種のみ microF 値は報告されていないため、論文中の ID ペア種別ごとの F 値のうち対応するものの macro 平均値を参考として記した。また、各スコアにおいて最大の値を太字で示した。

手法	関係抽出のみ (シングルタスク設定)						パイプライン					
	開発セット			テストセット			開発セット			テストセット		
	Precision	Recall	F 値	Precision	Recall	F 値	Precision	Recall	F 値	Precision	Recall	F 値
ベースライン	0.653	0.606	0.629	0.648	0.633	0.640	0.321	0.358	0.339	0.286	0.364	0.320
提案手法	0.690	0.597	0.640	0.668	0.586	0.624	0.348	0.353	0.351	0.302	0.346	0.322
共起のみ	0.671	0.604	0.635	0.628	0.641	0.634	0.346	0.373	0.359	0.275	0.366	0.314
Luo ら [3]	-	-	-	-	-	~0.601	-	-	-	-	-	-

4 実験と考察

4.1 実験設定

実験では、BioRED データセット [3] を用いて学習と評価を行った。パイプラインシステムの固有表現抽出には著者らのニューラル固有表現抽出モデル [6] を使用し、BioRED で学習と評価を行った。エンティティリンキングには著者らのニューラルリンキングモデル [7] の、NLM-CHEM データセット [8] で学習された学習済みモデルを用いた。このリンキングモデルは MeSH ID²⁾ へのリンキングを行うが、BioRED データセット上のエンティティでは、“Chemical” と “Disease” クラスが MeSH ID とリンクされているため、固有表現抽出モデルがこの 2 クラスへ分類したエンティティに限ってリンキングを行った。関係抽出においてもこの 2 クラスのエンティティのみを対象とした。ベースラインとして、学習によるマスクを用いず常に全文を利用する設定での実験も行った。また、比較として学習によるマスクの代わりに、対象エンティティペアが共起する文のみを利用する設定での実験も行った。学習によるマスクを用いない場合は、学習時も式 (4) 中のノイズは利用しない。各設定についてそれぞれ 5 回ずつ乱数シードを変更して学習と評価を行った。データセットの分割は公式の訓練・開発・テストセットを採用し、100 エポックの学習において最大の開発 F 値を計測したエポック時のパラメータでテスト評価を行った。より詳細な実験設定は付録 B に示す。

4.2 実験結果と考察

関係抽出の評価を表 1 に示す。提案手法は開発セットではシングルタスク設定においてベースラインよりも 1%ポイント以上高い F 値であったが、テ

ストセットにおいては優劣が逆転した。共起のみを用いた場合でも同様の傾向が確認された。なお、エンティティ ID ペア 2 種のみでの設定ではあるが、いずれの手法も先行研究 [3] よりも高い F 値となった。パイプラインシステムに用いた固有表現抽出とエンティティリンキングの性能は付録 C に示す。

提案手法は比較手法と比べて常に最高の precision となったが、逆に recall は常に最低となった。また、パイプライン設定では、シングルタスク設定同様に提案手法がベースラインよりも 1%ポイント以上高い開発 F 値となった一方、テストセットではほとんど差がない結果となっている。パイプライン設定では言及スパンや ID がシングルタスク設定より信頼しにくいいため、この点で precision の高い提案モデルが有利であった可能性が示唆されるが、より詳細な分析は今後の課題である。

シングルタスク設定において、学習済みの提案手法が開発セット上でマスクしなかった文数を計測したところ、1 文書当たりの平均で約 3.0 文であった。これは文書中の文数の約 25%にあたる。もし各対象エンティティ ID のスパンが別々の文に出現する場合、最低 2 文は入力として利用しなければならないが、提案手法はこれに近い数まで利用する文をマスクしていることが確認された。

5 おわりに

本研究では、ID レベル関係抽出における不要な情報の増大に対処するため、不要な文をマスク処理によって自動で選択する ID レベル関係抽出モデルを提案した。さらに提案モデルに固有表現抽出とエンティティリンキングを組み合わせることで、テキストのみから ID レベルの関係を抽出するパイプライン ID レベル情報抽出システムを実現した。今後の課題として、マスク処理が利用する文についてのより詳細な分析が挙げられる。

2) <https://www.nlm.nih.gov/mesh/meshhome.html>

謝辞

本研究は JSPS 科研費 JP20K11962 の助成を受けたものです。

参考文献

- [1] Razvan Bunescu and Raymond Mooney. A shortest path dependency kernel for relation extraction. In **Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing**, 2005.
- [2] 辻村有輝, 三輪誠, 佐々木裕. 関係分類における依存木上の重要トークンの自動判別. *自然言語処理*, Vol. 27, No. 2, pp. 211–235, 2020.
- [3] 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, 2022.
- [4] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. **Computing Research Repository**, Vol. arXiv:1611.01144, , 2017. version 5.
- [5] 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, 2019. Association for Computational Linguistics.
- [6] Tomoki Tsujimura, Koshi Yamada, Ryuki Ida, Makoto Miwa, and Yutaka Sasaki. Contextualized medication event extraction with striding ner and multi-turn qa. **Journal of Biomedical Informatics**, p. 104416, 2023.
- [7] Tomoki Tsujimura, Makoto Miwa, and Yutaka Sasaki. Large-scale neural biomedical entity linking with layer overwriting. **Journal of Biomedical Informatics**, p. 104433, 2023.
- [8] Rezarta Islamaj, Robert Leaman, David Cissel, Meng Cheng, Cathleen Coss, Joseph Denicola, Carol Fisher, Rob Guzman, Preeti Kochar, Nicholas Miliaras, Zoe Punske, Keiko Sekiya, Dorothy Trinh, Deborah Whitman, Susan Schmidt, and Zhiyong Lu. The chemical corpus of the nlm-chem biocreative vii track: Full-text chemical identification and indexing in pubmed articles. In **Proceedings of the seventh BioCreative challenge evaluation workshop**, 2021.
- [9] Mark Neumann, Daniel King, Iz Beltagy, and Waleed Ammar. ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing. In **Proceedings of the 18th BioNLP Workshop and Shared Task**, pp. 319–327, Florence, Italy, 2019. Association for Computational Linguistics.
- [10] Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. Domain-specific language model pre-training for biomedical natural language processing. **ACM Trans. Comput. Healthcare**, Vol. 3, No. 1, 2021.
- [11] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. **Journal of Machine Learning Research**, Vol. 15, No. 56, pp. 1929–1958, 2014.

A 提案モデルにおけるノイズの計算

学習時の式 (4) 中のノイズ $\mathbf{v}^{\text{noise}}$ は以下の式によって計算する.

$$\begin{aligned} r^{\text{num}} &= \sum_S m_i^{\text{mask}} \\ r^{\text{penalty}} &= \max(0, r^{\text{num}} - 1) \\ \mathbf{p} &\sim \mathcal{N}(\mathbf{0}, \alpha^2 \mathbf{I}) \\ \mathbf{v}^{\text{noise}} &= r^{\text{penalty}} \mathbf{p}, \end{aligned} \quad (5)$$

ここで α はハイパーパラメータであり, 実験では 0.5 とした.

B 実験設定の詳細

ここではより詳細な実験設定について説明する.

関係抽出では, 前処理として BioRED データセットの各入力文書を scispaCy [9] によって文分割した. scispaCy のモデルには “en_core_sci_scibert” を用いた. 提案モデルの BERT エンコーダーの初期値には PubMedBERT [10] を用いた. PubMedBERT の最終層隠れベクトルの次元数は 768 次元である. 式 (2) 中のパラメータの各次元数は, $\mathbf{W}^{\text{csent}}$ と $\mathbf{b}^{\text{csent}}$ はそれぞれ 768×1536 次元と 768 次元, \mathbf{W}^{key} と $\mathbf{W}^{\text{query}}$ はそれぞれ 768×768 次元と 768×1536 次元とし, 双方向 LSTM の各方向の次元数も 768 次元とした. 式 (4) 中のパラメータの各次元数は, \mathbf{W}^{h} と \mathbf{b}^{h} はそれぞれ 1536×1536 次元と 1536 次元, \mathbf{W}° と \mathbf{b}° はそれぞれ 9×1536 次元と 9 次元とした. 式 (5) 中の \mathbf{p} の次元数は 1536 次元とした. 汎化性能を向上させるため, 隠れ層の出力に対してドロップアウト [11] を用いた. 各入力では入力文書中の 1 つのエンティティ ID ペアのための関係抽出を行うこととし, そのエンティティ ID ペアに対応する全ての言及スパンの始まりと終わりの位置にマーカーとなる特殊トークンを挿入した. 対象エンティティが共起する文のみを利用する設定では, 入力文書中に利用できる文が存在しなくなるエンティティ ID ペアも存在するが, その場合は常に関係なしと予測することとした.

C 固有表現抽出とエンティティリンキングの性能

表 2 固有表現抽出の性能. それぞれのスコアは micro 平均によって計算した.

手法	対象クラス	開発セット			テストセット		
		Precision	Recall	F 値	Precision	Recall	F 値
辻村ら [6]	Chemical & Disease	0.891	0.908	0.899	0.876	0.914	0.895
	全 6 クラス	0.908	0.931	0.919	0.918	0.932	0.925
Luo ら [3]	全 6 クラス	-	-	-	-	-	0.893

表 3 “Chemical” および “Disease” クラスのエンティティに対するエンティティリンキングの性能. 文書レベルのスコアは micro 平均によって計算した.

手法	データ	スパンレベル	文書レベル (シングルタスク)			文書レベル (パイプライン)		
		Accuracy	Precision	Recall	F 値	Precision	Recall	F 値
辻村ら [7]	開発セット	0.7062	0.7258	0.8411	0.7792	0.6726	0.7876	0.7255
	テストセット	0.6876	0.7331	0.8490	0.7868	0.6521	0.7957	0.7168

パイプラインシステムに用いた固有表現抽出の性能を表 2 に, エンティティリンキングの性能を表 3 に示す.