

訓練可能な k 近傍 Retriever で関係抽出事例を導入したニューラルプロンプティング

牧野 晃平 三輪 誠 佐々木 裕
豊田工業大学

{sd21505,makoto-miwa,yutaka.sasaki}@toyota-ti.ac.jp

概要

近傍法による関係抽出では、関係の事例を推論に利用すると、少ない訓練データでも高い性能で抽出が可能になるという報告がある。LLM による抽出では、Retriever によって得た事例を使う方法があるものの、事例の選択が微分できないため、Retriever を end-to-end に訓練できない。そこで本研究では、微分可能な k 近傍事例のソフトな選択を利用して、事例をソフトプロンプトとして導入するニューラルプロンプティングを行うことで、関係抽出で end-to-end に Retriever を訓練できるようにする。標準的なベンチマーク用データセットの TACRED に対する抽出性能は、訓練事例数を 10% と 5% とした場合に、F 値でそれぞれ 71.0% と 70.9% となり、訓練事例を削減した条件では最先端の性能が得られた。

1 はじめに

文章中の用語ペア間の関係性を抽出する関係抽出は、自然言語で記述された情報の構造化において重要な役割を果たしている [1]。関係抽出器の作成では、人手で作成された教師を利用して深層学習モデルを訓練するのが一般的だが [2]、タグ付けのコストが高いことから、少ない訓練データで高い抽出性能が得られるモデルが求められている。このような中で、近傍法で関係抽出の事例を推論時に利用して関係抽出をする抽出器 [3] は、関係ラベル同士の関係性をモデル化することで、少ない訓練データで訓練できるようになってきている。

加えて、近年では、それらのモデルのベースに BERT [4] のような事前学習言語モデル (Pretrained Language Model; PLM) の利用が標準的になってきている。最近では、事前学習モデルとして大規模言語モデル (Large Language Model; LLM) を利用した研究が注目されている [5]。特に、指示やデモンスト

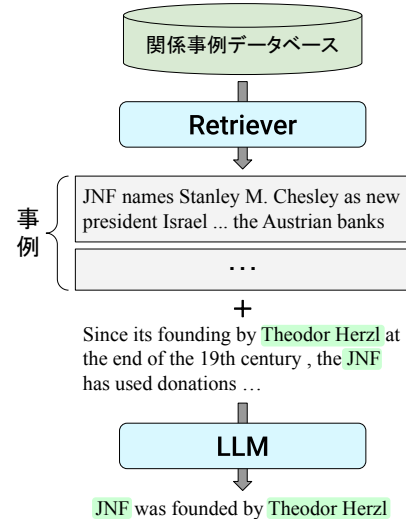


図 1 事例を利用した LLM による関係抽出

レーションとなる事例を文脈に与えて少ない事例で訓練を実現する文脈内学習 (In-context learning; ICL) が可能になり、関係抽出に事例を利用する研究が進んでいる [6]。ICL に利用される技術の一つである Retrieval-Augmented Generation (RAG) では、図 1 のように、入力に合わせて事例を選択する Retriever によって得られた事例のテキストをプロンプトに導入することで、ICL で不足する知識や ICL の方向性を調整している [7]。

ただ、RAG では、Retriever による事例の選択は微分不可能な操作であるため、事前に Retrieval に合わせて学習することはできても [8]、対象タスクに合わせて end-to-end に Retriever を最適化するのは難しい。これを解決するために、事例からソフトプロンプトを作成するニューラルプロンプティングが提案されたものの [9]、依然として Retriever による事例の選択は微分不可能な操作であるため、end-to-end に訓練できないという問題点がある。

そこで本研究では、関係抽出において、Retriever まで含めて end-to-end に最適化可能にし、関係抽出に特化した事例の選択を可能にすることを目的とす

る。Retriever を最適化可能にするためには、 k 近傍事例の選択を微分可能な操作にする必要がある。そこで、近傍ニューラルネットワーク [10] に着想を得て、実際にサンプリングせずにソフトに選択することで微分可能にし、その事例をソフトプロンプトとして導入するニューラルプロンプティングをすることで、Retriever を含めて関係抽出を end-to-end に訓練した時の効果を確認する。

本研究の貢献は以下の点である。

- LLM による関係抽出において、 k 近傍事例の選択をソフトにして微分可能にし、ニューラルプロンプティングに利用する手法を提案し、Retriever まで end-to-end に訓練可能な方法を示した。
- 関係抽出の標準的なベンチマーク TACRED で、訓練事例数を 10%, 5% に減らす設定において最先端の性能を示した。

2 関連研究

深層学習による関係抽出器は、文脈を考慮したモデル [2] や、文章構造のような構造をグラフ構造として利用するモデル [11] など、様々なモデル化がされている。関係事例を利用するモデルである近傍法による関係抽出 [3] は、他の関係事例との関係性を明確にすることで、事例数が少ない関係ラベルでも高い性能での予測が可能になった。

既存の LLM による関係抽出では、LLM を QA モデルとして利用 [12] したり、関係ラベル名を生成するようにプロンプトを設計 [5] したりと、単純な方法が取られており、抽出性能も LLM を利用しない関係抽出と大きく変わらなかった。BERT のような PLM を利用した関係抽出手法の中では、関係を様々な観点から自然言語に書き下し、自然言語推論によって含意する関係を見つけることでこの問題を解決した方法 [13] が提案された。このような中で、出力の形式を自然言語に合わせて言語モデルに適合させるために、文章を関係事例へ要約するタスクとみなして関係抽出を行うモデル SuRE (Summarization as Relation Extraction) [14] が提案された。

3 LLM による関係抽出

本節では LLM による関係抽出器、および Retriever を用いたその拡張を定義する。入力の記事を x 、文章中の関係の始点と終点の用語をそれぞれ h, t 、そ

Algorithm 1: 微分可能な k 近傍事例の選択

Input: K : 選択する数

$E = [E_1, E_2, \dots, E_{|D|}]$: 事例の埋め込み

E_{in} : 入力記事の埋め込み

Output: $X \in \mathbb{R}^{K \times L \times D_{emb}}$: k 近傍事例の埋め込み

$s_i = \text{Dist}(E_{in}, E_i)$

for $k \in [1, K]$ **do**

$w_i = \frac{\exp(-s_i)}{\sum_i \exp(-s_i)}$

$x_k = \sum_i w_i E_i$

$s_i = s_i + \log(1 - w_i)$

end for

の間の関係を r とすると、分類モデルによる関係抽出器の予測は $\hat{r} = \text{argmax}_r P(r|x, h, t)$ となる。一般的な LLM を利用した関係抽出では、入力の記事をテンプレートに当てはめてテキストプロンプトに変換し、それを LLM に入力して得られるテキストを関係ラベルにマッピングして関係を抽出する。

$$\hat{r} = \text{Map}(\text{LLM}(y|\text{Template}(x, h, t))) \quad (1)$$

次に、LLM による関係抽出器を、図 1 のような Retriever によって獲得した事例を利用する関係抽出に拡張する。そのために、式 (3) のように、Retriever によって事例集合や外部知識のデータベース D から事例を選択し、その事例からプロンプト z を作成して LLM の入力に与える。

$$z = \text{Retriever}(x, h, t, D) \quad (2)$$

$$\hat{r} = \text{Map}(\text{LLM}(y|z, \text{Template}(x, h, t))) \quad (3)$$

Retriever は、データベースから事例を選択する機能と、選択した事例を LLM に入力可能なプロンプトに変換する機能の 2 つの機能から構成される。既存の RAG [7] の Retriever では、入力とデータベース D 中の事例を埋め込み、近傍の事例を選択し、その事例のテキストを直接プロンプトに利用する。ニューラルプロンプティング [9] では、文字列一致で事例を選択し、事例をニューラルネットワークで埋め込み、ソフトプロンプトとして利用する。

4 提案手法

k 近傍事例の選択をソフトな操作として微分可能にし、それによって得られた埋め込みをソフトプロンプトとして利用するニューラルプロンプティングによって k 近傍 Retriever を訓練可能にする。

まず、 k 近傍 Retriever を定義する。 k 近傍 Retriever では、訓練データの関係事例の集合などのデー

データベース D 中の要素 d_i を埋め込みのリスト $E_i = [e_i^1, e_i^2, \dots, e_i^L] \in \mathbb{R}^{L \times D_{\text{emb}}}$ に変換し、入力埋め込みと距離が近い事例を選択する。

$$E_i = \text{Embed}(d_i) \quad (4)$$

$$E_{\text{in}} = \text{Embed}(x, h, t) \quad (5)$$

$$s_i = \text{Dist}(E_{\text{in}}, E_i) \quad (6)$$

$$z = \text{KNNRetriever}(x, h, t, D) = [z_1, z_2, \dots, z_K] \quad (7)$$

$$= \parallel_{k \in \text{arg top-}k_i s_i} \text{Template}(d_k) \quad (8)$$

通常の k 近傍 Retriever では、事例の選択部分が微分できず、プロンプトもテンプレートを利用したテキストになっているため、Retriever の訓練ができない。

そこで、 k 近傍事例の選択を微分可能にし、埋め込み E を利用するソフトプロンプトにすることで、 k 近傍 Retriever 全体を訓練可能な状態にする。具体的には、既存の微分可能な k 近傍事例の選択 [10] を参考に、アルゴリズム 1 で得られる k 近傍事例の埋め込みをプロンプトにして、訓練可能な k 近傍 Retriever を実現する。この Retriever では、距離を利用して計算した Softmax の値を重みとして、埋め込みの重み付け和をとることで、 k 近傍事例の埋め込みを仮想的に計算する。これによって、事例の選択が微分可能になり、プロンプトもソフトプロンプトとすることで、Retriever を含めて LLM を end-to-end に訓練できるようになる。

ただ、この微分可能な k 近傍 Retriever では、 D 内の事例全体に対して距離を計算しなければならない。これを訓練時にも行った場合、すべてのデータベース中の事例の埋め込みを計算して、すべての埋め込みに対する更新をしなければならない。そこで、これを解決するために、ランダムに事例の部分集合をサンプリングし、その事例の中で Retriever を適用して訓練をする。

5 実験

5.1 実験設定

関係抽出の抽出性能の評価をするために、標準的に利用されている TACRED データセット [15] を使用する。評価指標は、それぞれのタスクで一般的に利用されているものとして、関係なしのクラス no_relation を負例として計算した micro 平均 F 値、適合率 (P)、再現率 (R) の百分率を使用する。それぞれの結果としては、計算コストのため、一度実行し

たときの値を利用する。また、付録 B に示したように、訓練事例を 10%、5%、1% に減らした場合についても評価を行い、訓練事例数に対する性能の変化を確認する。

評価ではベースのモデルに Flan-T5 の large サイズを利用して [16]、関係抽出を要約タスクとして定式化した SuRE [14] をもとに、関係事例をプロンプトに追加する実験をした。提案手法の有効性の確認のため、訓練可能な k 近傍 Retriever を利用して事例を導入した場合を評価した。Retriever で利用する表現は、SuRE で利用されているテンプレートを当てはめた文章を RoBERTa-base [17] でエンコードし、関係の始点と終点の用語の表現と関係を述べている部分の表現を、トークンの表現を平均して作成した。また、訓練時に Retriever でサンプリングする事例数は 32 として、10 事例をプロンプトに導入する条件で実験した。Retriever の事前の訓練のための Warmup として、最初の 300 ステップは Retriever のパラメタのみを更新し、その後 Flan-T5 を更新した。Retrieval に利用するデータベースは訓練データ中の関係事例として、計算機の都合上、5000 件以上訓練事例がある場合には 5000 件をランダムにサンプリングしてデータベース D とした。詳細なモデルの設定は、付録 A に記述した。

既存の関係抽出の最先端のモデルとして、Pegasus-large をベースにした SuRE に加えて、構造化のための事前学習モデルである DeepStruct [18]、言語化した関係が入力文章に含まれるかどうかを NLI で解いて関係抽出をする NLI.DeBERTa [13] と比較する。

5.2 結果

表 1 に示した結果より、訓練データ数が少ない 10% と 5% のときに、既存の訓練事例数を削減した場合の SOTA である SuRE および NLI.DeBERTa と比べ、提案手法が高い性能を示した。ベースとした SuRE と比較すると、訓練事例数を 10% 以下にした場合に、提案手法によって再現率が向上し、F 値も向上した。一方で、訓練事例が十分な領域では、抽出性能が低下することが確認できた。このことから、訓練可能な Retriever から得られた事例を利用するのは、限定的だが有効に作用すると確認できた。

6 解析

本章では、5 節で行った抽出性能の評価が得られた要因やモデルの挙動を解析するために、アブレー

表 1 関係抽出性能の比較 [%]

| | 100% | | | 10% | | | 5% | | | 1% | | |
|---------------------|------|------|------|------|------|-------------|------|------|-------------|------|------|-------------|
| | P | R | F | P | R | F | P | R | F | P | R | F |
| DeepStruct [18] | - | - | 76.8 | - | - | - | - | - | - | - | - | - |
| SuRE (Pegasus) [14] | - | - | 75.1 | - | - | 70.7 | - | - | 64.9 | - | - | 52.0 |
| NLI DeBERTa [13] | 72.5 | 75.3 | 73.9 | 62.4 | 74.4 | 67.9 | 64.1 | 74.8 | 69.0 | 59.5 | 68.5 | 63.0 |
| SuRE (Flan-T5) | 73.5 | 73.2 | 73.3 | 74.1 | 62.5 | 67.8 | 68.0 | 61.5 | 64.6 | 60.6 | 50.7 | 55.2 |
| + k 近傍 Retriever | 70.5 | 74.4 | 72.4 | 67.5 | 75.0 | 71.0 | 67.8 | 74.3 | 70.9 | 51.9 | 63.2 | 57.0 |

表 2 アブレーション研究 [%]

| | P | R | F |
|-----------|------|------|------|
| 提案手法 | 67.8 | 74.3 | 70.9 |
| 訓練なし | 70.0 | 70.7 | 70.4 |
| Warmup なし | 68.1 | 72.7 | 70.3 |
| ランダム | 67.2 | 65.6 | 66.4 |
| CLS | 69.6 | 60.2 | 64.5 |

シオン研究と、事例数に対する性能変化を確認する。以降の実験は 5 節で使用したモデルをベースにして、性能の向上が確認できた設定である 5% の訓練事例を利用した際に、条件を変化させた場合の性能を確認して解析を行った。

6.1 アブレーション研究

提案手法のうち、性能の向上に寄与した要因を確認するために、事例に関する部分と Retriever モデルを対象としてアブレーション研究をした。具体的には、ランダムな事例を利用した場合（ランダム）、Retriever を訓練せずに k 近傍事例を利用した場合（訓練なし）、利用する表現を用語ではなく CLS トークンにした場合（CLS）、Retriever の訓練時の Warmup をしない場合（Warmup なし）、のそれぞれの設定での性能を比較した。

表 2 に示した通り、すべての場合について F 値が低下しており、提案手法の各部分が性能向上に寄与していることが確認できた。まず、Retriever の訓練をしない場合や Warmup をしない場合のように、Retriever の訓練能力を弱めると性能が低下することが確認できた。また、プロンプトに利用する事例をランダムにした場合には、性能が大幅に低下したことから、距離を用いた近傍事例の選択は重要な役割を果たしていることが確認できた。プロンプトや Retriever に利用する表現を CLS トークンの表現にした場合では、性能がさらに大幅に低下した。このことから、利用する表現を関係抽出に特化させて設計

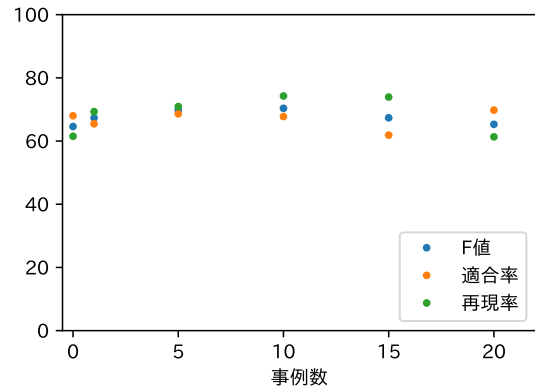


図 2 プロンプトに利用する事例数に対する抽出性能の変化

することも重要な要素であることが確認できた。

6.2 事例数に対する性能の変化

5.2 節では事例を利用すると再現率及び F 値が向上したため、その影響を確認するためにプロンプトに利用する事例数を変化させたときの性能変化を確認した。図 2 に示した結果では、10 を境に、事例数を増加した場合も減少させた場合も F 値と再現率が徐々に低下した。一方で、適合率には明確な傾向が確認できなかった。このことから、適切な事例数の選択は再現率に有利に働くことが確認できた。

7 おわりに

本研究では、LLM による関係抽出において、プロンプトに利用する事例を選択する Retriever を訓練可能にするために、微分可能な k 近傍サンプリングを利用したニューラルプロンプティングを提案した。実験では、訓練事例数が少ない場合に再現率および F 値の向上が確認でき、事例の利用を訓練可能にする効果が確認できた。ただ、訓練事例数が多い場合や極端に少ない場合には性能の向上が限定的であったため、今後はその改善が必要である。

謝辞

本研究は JSPS 特別研究員奨励費 22KJ2983 の助成を受けたものです。

参考文献

- [1] Linlin Wang, Zhu Cao, Gerard de Melo, and Zhiyuan Liu. Relation classification via multi-level attention CNNs. In **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1298–1307, Berlin, Germany, August 2016. Association for Computational Linguistics.
- [2] Makoto Miwa and Mohit Bansal. End-to-end relation extraction using LSTMs on sequences and tree structures. In **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1105–1116, Berlin, Germany, August 2016. Association for Computational Linguistics.
- [3] Zhen Wan, Qianying Liu, Zhuoyuan Mao, Fei Cheng, Sadao Kurohashi, and Jiwei Li. Rescue implicit and long-tail cases: Nearest neighbor relation extraction. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing**, pp. 1731–1738, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.
- [4] 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.
- [5] Somnath Wadhwa, Silvio Amir, and Byron Wallace. Revisiting relation extraction in the era of large language models. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 15566–15589, Toronto, Canada, July 2023. Association for Computational Linguistics.
- [6] Kai Zhang, Bernal Jimenez Gutierrez, and Yu Su. Aligning instruction tasks unlocks large language models as zero-shot relation extractors. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, **Findings of the Association for Computational Linguistics: ACL 2023**, pp. 794–812, Toronto, Canada, July 2023. Association for Computational Linguistics.
- [7] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, **Advances in Neural Information Processing Systems**, Vol. 33, pp. 9459–9474. Curran Associates, Inc., 2020.
- [8] Xinze Li, Zhenghao Liu, Chenyan Xiong, Shi Yu, Yu Gu, Zhiyuan Liu, and Ge Yu. Structure-aware language model pretraining improves dense retrieval on structured data. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, **Findings of the Association for Computational Linguistics: ACL 2023**, pp. 11560–11574, Toronto, Canada, July 2023. Association for Computational Linguistics.
- [9] Yijun Tian, Huan Song, Zichen Wang, Haozhu Wang, Ziqing Hu, Fang Wang, Nitesh V. Chawla, and Panpan Xu. Graph neural prompting with large language models, 2023.
- [10] Tobias Plötz and Stefan Roth. Neural nearest neighbors networks. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, **Advances in Neural Information Processing Systems**, Vol. 31. Curran Associates, Inc., 2018.
- [11] Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. Connecting the dots: Document-level neural relation extraction with edge-oriented graphs. In **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, pp. 4925–4936, Hong Kong, China, November 2019. Association for Computational Linguistics.
- [12] Amir DN Cohen, Shachar Rosenman, and Yoav Goldberg. Supervised relation classification as twoway span-prediction. In **4th Conference on Automated Knowledge Base Construction**, 2022.
- [13] Oscar Sainz, Oier Lopez de Lacalle, Gorka Labaka, Ander Barrena, and Eneko Agirre. Label verbalization and entailment for effective zero and few-shot relation extraction. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 1199–1212, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.
- [14] Keming Lu, I-Hung Hsu, Wenxuan Zhou, Mingyu Derek Ma, and Muhao Chen. Summarization as indirect supervision for relation extraction. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, **Findings of the Association for Computational Linguistics: EMNLP 2022**, pp. 6575–6594, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.
- [15] 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.
- [16] Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Yunxuan Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, Albert Webson, Shixiang Shane Gu, Zhuyun Dai, Mirac Suzgun, Xinyun Chen, Aakanksha Chowdhery, Alex Castro-Ros, Marie Pellat, Kevin Robinson, Dasha Valter, Sharan Narang, Gaurav Mishra, Adams Yu, Vincent Zhao, Yanping Huang, Andrew Dai, Hongkun Yu, Slav Petrov, Ed H. Chi, Jeff Dean, Jacob Devlin, Adam Roberts, Denny Zhou, Quoc V. Le, and Jason Wei. Scaling instruction-finetuned language models, 2022.
- [17] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019.
- [18] Chenguang Wang, Xiao Liu, Zui Chen, Haoyun Hong, Jie Tang, and Dawn Song. DeepStruct: Pretraining of language models for structure prediction. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, **Findings of the Association for Computational Linguistics: ACL 2022**, pp. 803–823, Dublin, Ireland, May 2022. Association for Computational Linguistics.

A 詳細な実験設定

実験の設定は、基本的に SuRE[14] に準じて行っており、テンプレートは元論文と同様のものを利用した。SuRE の入力用のテンプレートは図 3 の形式で、ここから関係ラベルごとに定義された要約用のテンプレートを予測する。要約用のテンプレートは、例えば、関係ラベル“org:founded_by”に対しては、“\${ 始点の用語 } was founded by \${ 終点の用語 }”というようなテンプレートが SuRE で提案されていた。Retriever の埋め込みに利用した表現は、PLM に対して入力用のテンプレートと要約用のテンプレートを結合して入力し、図 3 中の下線部の埋め込みと、要約テンプレート部分の埋め込みを利用した。予測時の要約用のテンプレート部分の埋め込みについては、EOS トークンの表現を利用した。

SuRE による分類時のパラメタである探索ビーム幅は、SuRE と同じ値である 4 とした。Flan-T5 モデルおよび、Retriever 部分の RoBERTa モデルは、LoRA を適用して最適化をした。LoRA のパラメタは、両者ともに、 $r = 32, \alpha = 16$ 、ドロップアウト率を 0.1 とした。モデルの最適化には AdamW を利用して、Flan-T5 は 5×10^{-4} 、それ以外のパラメタは 1×10^{-3} の学習率で、バイアス部分と Layer Normalization のパラメタを除く部分に対しては重みを 5×10^{-6} として重み減衰を適用した。バッチサイズは 64 とした。評価に利用するパラメタは、patience を 5 とした Early Stopping によって訓練が完了したときのものを利用した。それぞれの実験では、産総研の AI 橋渡しクラウド (ABCI) の NVIDIA A100 SXM4 を一枚利用した。

B データセット

訓練事例数が少ない場合についても性能調査を行うために、TACRED データセットを 100% 利用した場合に加えて、10%、5%、1% に減らした場合についての実験を行った。この時のデータの分割や利用するデータは、[13] の設定に従って決定した。そのときのデータセットの統計量は、表 3 に示した。

The head entity is \${ 始点の用語} . The tail entity is \${ 終点の用語} . The type of \${ 始点の用語} is \${ 始点の用語ラベル} . The type of \${ 終点の用語} is \${ 終点の用語ラベル} . \${ 入力文} .

図 3 SuRE の入力テンプレート

表 3 TACRED データセットの統計量

| | 訓練 | 開発 | 評価 |
|------|--------|--------|--------|
| 100% | 68,125 | 22,631 | |
| 10% | 6,815 | 2,265 | 15,509 |
| 5% | 3,407 | 1,133 | |
| 1% | 682 | 227 | |