

項の表現に着目した質問応答による関係分類

山田 晃士 三輪 誠 佐々木 裕
豊田工業大学

{sd18086,makoto-miwa,yutaka.sasaki}@toyota-ti.ac.jp

概要

関係分類の新しい手法として、質問応答による関係分類が提案されている。これらの手法では、既存の質問応答モデルを用いているため、関係分類タスク特有の情報を利用できていない。また、回答の開始・終了位置のみを予測するため、その間の表現を利用していない。本研究では、関係に関するマーカーと、スパン表現を用いた質問応答モデルを用いて、タスク特有の情報と間の表現を利用可能な質問応答による関係分類手法を提案する。SemEval-2010 Task 8 データセットを用いた評価の結果、マーカーを用いない場合と比べて、性能は1.9%ポイント向上した。また、解析では、マーカーの追加により、正しく用語を回答できるようになることを確認した。

1 はじめに

構造化されていない文献中から情報を抽出するタスクの一つとして、文献中に含まれる用語間に成り立つ関係を候補となる関係ラベルに分類する関係分類が研究されている。

関係分類手法のうち、高い精度を達成しているものとして、質問応答を用いた関係分類手法 [1, 2, 3] がある。例として、「フランスの首都はパリである」という対象文中で与えられた用語であるフランスとパリの間に首都という関係があるかどうかを判定する問題を、「フランスの首都はどこですか?」という質問文をこの対象文に対して適用して回答を探す問題に置き換える。パリが回答であると判定されればフランスとパリの間に首都の関係があると判定する。具体的には、まず、候補となる関係ラベルすべてに対して用語をスロットとする質問文のテンプレートを作成して、そこにデータセットで与えられる用語ペアの片方の用語を当てはめることで質問文を作成する。作成した質問文と元の文を質問応答モデルに入力し、得られた回答がもう一つの用語のときテンプレートがもつ関係があるとして関係分類を

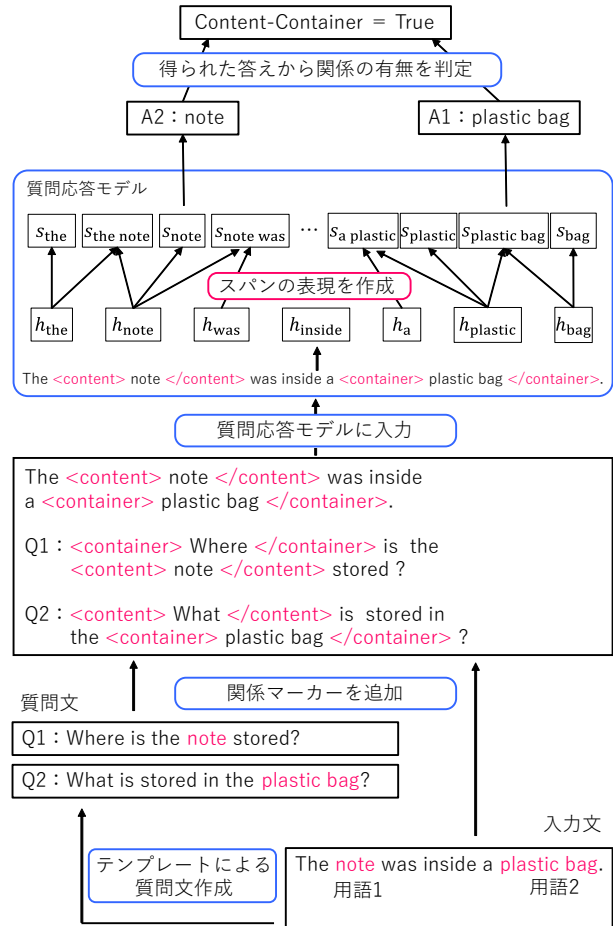


図1 提案手法の概要

する。質問応答を用いた関係分類手法には、関係ごとに異なる質問文を用意することで、それぞれの関係に注目した処理を行えるという利点がある。一方で、既存の質問応答を用いた関係分類手法では、多くの関係分類のデータセット [4, 5] が持つ関係分類特有の情報を利用できていないという問題点がある。関係分類特有の情報とは、タスクにおいて既知である用語のタイプや、関係の項それぞれのタイプなどを指す。

そこで、本研究では、関係分類タスク特有の情報の一つである関係の項のタイプの有効活用による、

質問応答を用いた関係分類手法の性能向上を目的とする。データセットに含まれる項のタイプを、元の文と質問文に対してマーカーとして追加することで、質問応答型の関係分類手法の利点である各関係に注目した入力を保ったまま、モデルの大幅な変更をすることなく、関係分類に特化させる。

2 関連研究

Cohen らは図 2 のように関係の始点と終点をそれぞれ予測する 2 つの質問を用いて双方向の質問応答を行い、得られた回答を基に関係を分類する手法 [1] を提案している。関係ごとに 2 つ用意したテンプレートを用いて、用語と関係からなる質問文を用語ペアそれぞれについて作成して質問応答を行う。2 つのテンプレートは関係の始点を当てはめて終点を予測するものと、終点を当てはめて始点を回答するものになっている。2 つの質問文に対して片方でも質問応答の回答がもう片方の用語であった場合に関係があると判定して分類する。質問応答は回答の開始位置と終了位置をそれぞれ予測するモデルで行う。BERT[6] に文と質問文を入力し、得られた文の各トークンの表現から、そのトークンの開始位置のスコアと終了位置のスコアを個別に計算し、それぞれのスコアが最も高い開始位置と終了位置を持つスパンが質問の回答となる。この方法では、回答の終了位置が開始位置より前になってしまうという実際にはありえないスパンが回答になってしまう可能性がある。また、回答の開始位置と終了位置のみを予測するため、その間の表現を利用できない。さらに、この手法は既存の質問応答をそのまま用いているため、タスクで与えられている情報を利用していないという問題がある。

3 提案手法

本研究では関係分類タスクに質問応答モデルを用いる際に、関係分類タスク特有の情報を活用する手法を提案する。Cohen らの手法 [1] を拡張し、質問文のテンプレートで与えられる関係の情報に加えて、タスクに与えられる関係の項それぞれのタイプを示すマーカーを加えることで、関係分類において事前に与えられる用語の情報を質問応答モデルに与える。また、質問応答モデルをスパンそのものを回答するようにすることで、回答がありえないスパンになることを防ぐとともに、開始位置と終了位置以外の間のトークンの表現も含む回答全体の情報を利

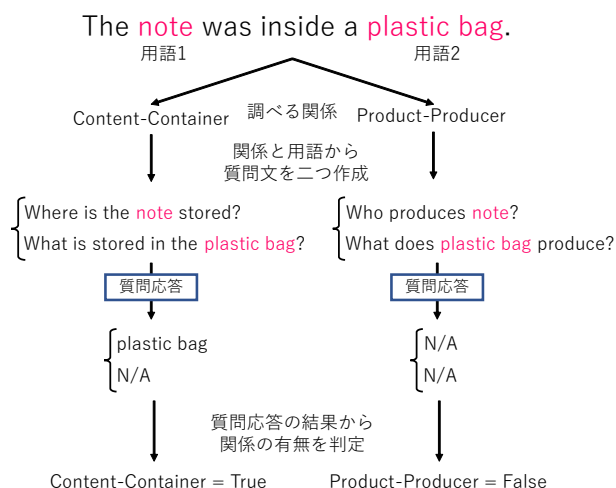


図 2 双方向の質問応答による関係分類 [3]

用する。図 1 に提案手法の概要を示す。

3.1 入力文の整形

質問応答モデルへの入力の前準備として、関係ごとの質問文の作成と関係マーカーの付与を行う。図 1 のように、関係ごとに、関係分類を行う文とテンプレートを用いて作成した質問文をそれぞれ関係情報のマーカーを加えたものに整形する。質問文のテンプレートは Cohen らのもの [1] を利用し、双方向に質問を行う。関係マーカーは本稿で対象としている SemEval 2010 Task 8 データセット [4] で与えられる関係の項のタイプの情報を利用する。使用したテンプレートとマーカーは付録 A に示した。関係分類を行う文に対しては用語を挟むようにマーカーを、質問文には用語の前後と回答に対応する疑問詞に対してマーカーを追加する。

3.2 スパン表現を用いた質問応答

図 3 に質問応答モデルの概要を示す。質問応答モデルは BERT[6] を利用したモデルとなっており、文と質問文を入力とし、文中に質問の回答があればその回答となるトークンを、なければ回答なしを表す BERT の CLS トークンを出力する。以降、モデルの詳細について説明する。

まず、BERT の CLS トークンと SEP トークンを利用して、質問文と文を連結して BERT に入力し、トークンごとの表現を得る。

[CLS] 質問文 [SEP] 文 [SEP]

BERT への入力を x とした時、出力である表現の系列 H を以下のように表す。ここで、 N は入力トークン数、 h_i は i 番目のトークンの表現に相当し、 h_1

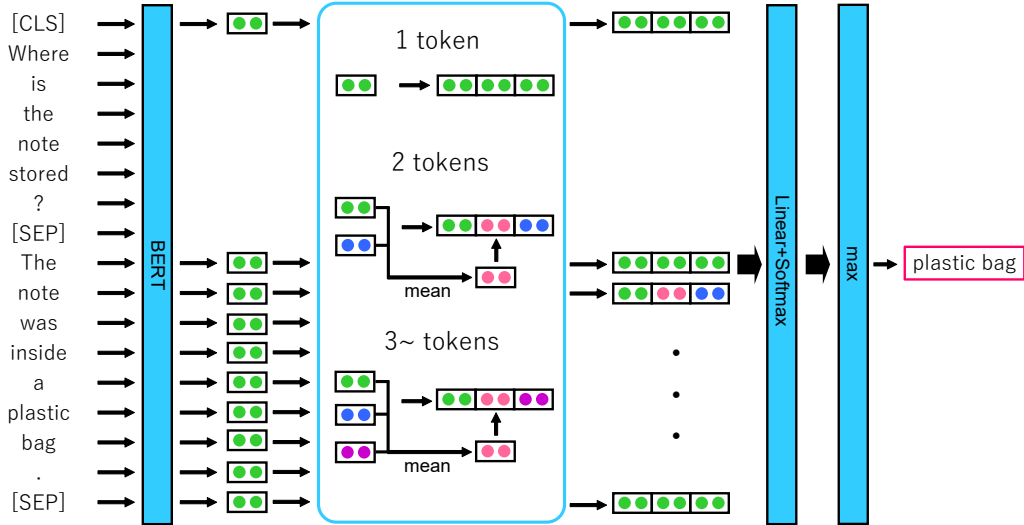


図3 スパン表現を用いた質問応答モデル

は文頭である CLS トークンの表現となる。

$$\mathbf{H} = [h_1, h_2, \dots, h_N] = \text{BERT}(x) \quad (1)$$

次に得られた表現 \mathbf{H} のうち CLS トークンと文の表現から、質問文の回答の候補となるスパンの表現を作成する。CLS トークンの表現は回答なしの候補として用いる。あるトークン h_i から始まり、 h_j で終わるスパンの表現 s は以下のように求める。

$$s = \text{Concat}(h_i, \text{mean}(h_i, \dots, h_j), h_j) \quad (2)$$

以上により得られたスパン表現のベクトル s をまとめ、以下のようにスパン表現の系列 \mathbf{S} を作成する。 N' は質問の回答となる可能性のあるスパン候補の数であり、 s_1 は CLS トークンの表現である。

$$\mathbf{S} = [s_1, s_2, \dots, s_{N'}] \quad (3)$$

生成したスパン表現系列 \mathbf{S} に全結合層と Softmax 関数を適用することで確率の形に変換し、最も確率の高いスパンを質問の回答とする。CLS トークンが回答として選ばれた場合は、回答なしとする。

4 実験と考察

4.1 データセット

提案手法の評価を SemEval-2010 Task 8 のデータセット [4] を用いて行う。このデータセットは一文中の用語間の関係分類を行うデータセットで、訓練データ 8,000 件とテストデータ 2,717 件で構成される。訓練データを関係ラベルの割合が均一になるように 4:1 に分割し、6,400 件を訓練データ、1,600 件を開発データとして用いる。分類する関係ラベルは

表1 提案手法と各種ベースラインの F 値

手法	開発	テスト
提案手法	87.7	87.9
マーカーなし	86.3	86.0
ベースライン	86.7	86.9

表2 回答の表現とマーカーによる比較

回答の表現	マーカー	マクロ F
開始・終了位置	なし	85.6
	関係	87.3
	用語	87.1
スパン	関係+用語	85.6
	なし	86.3
	関係	87.7
関係ラベル	用語	87.2
	関係+用語	85.9

正例 9 種類と負例の Other の 10 種類であるが、関係には方向があるため、方向を含めると正例 18 種類と負例 1 種類の計 19 種類の分類となる。評価指標には負例を除いた 9 種類の関係ラベルを持つデータに対するマクロ F 値を用いる。

4.2 実験環境

実装にはプログラミング言語 Python 3.7.11 を用いた。また、深層学習モデルを実装するために Pytorch 1.10.0[7]、事前学習済みモデルを利用するために Transformers 4.11.3[8] を使用した。計算機は、CPU に Intel(R) Xeon(R) W-3225、GPU に NVIDIA RTX A6000 をそれぞれ実験に使用した。

表3 マーカーの有無によって回答が変化した事例（太字が正しい回答）

質問文	Where is the box cutters stored?
文と回答	The scissors were in a small plastic soap container, next to a bar of soap, and the box cutters were in a box next to a bottle of after-shave lotion.
マーカーなし	soap container
マーカーあり	box
質問文	What does factory produce?
文と回答	The factory products have included flower pots, Finnish rooster-whistles, pans, trays , tea pots, ash trays and air moisturisers.
マーカーなし	flower pots
マーカーあり	trays
質問文	From what Calluses caused?
文と回答	Calluses are caused by improperly fitting shoes or by a skin abnormality .
マーカーなし	abnormality
マーカーあり	skin abnormality

4.3 実験設定

提案手法を評価するため、関係マーカーを追加しなかった場合のモデルと BERT の CLS トークンを用いて文分類として関係分類を行うベースラインモデルとの F 値の比較を行う。また、関係マーカーの代わりにデータセットに含まれる用語を表す用語マーカー<e1>、<e2>を用いた場合と、関係マーカーと用語マーカー両方を用いた場合について開発データで実験を行った。さらに、回答の開始位置と終了位置を回答の表現とする質問応答モデルを用いた場合についても開発データで実験を行った。事前学習済みモデルとして、BERT-base-uncased[6]を使用した。現状の SotA な手法である Cohen らのモデル [1] では BERT-large を用いているため、直接の比較はできず、この比較は今後の課題である。

4.4 実験結果

表 1 の結果から、関係マーカーを入力文に追加することで性能が向上していることがわかる。また、表 2 の結果から開始・終了位置を予測するモデルでもマーカーによる精度向上が確認できる。さらに、質問応答にスパン表現を用いることでの性能向上や、関係ごとに固有のマーカーを用いることの有効性も確認できる。

4.5 考察

関係マーカーの有効性を、マーカーを加えたことによって改善した事例を用いて確認する。マーカー

の有無によって変化した事例のいくつかを表 3 に示した。マーカーを加えたことで改善した回答として、用語以外を回答としていたものや、用語の一部を回答としていたものがあった。また、マーカーありとなしの場合における質問応答の正誤について McNemar 検定を行ったところ、有意に差が見られた ($p < 0.001$)。これらの結果は関係マーカーを追加することの有効性を示している。

5 おわりに

本研究では、関係分類タスク特有の情報を有効活用し、質問応答を用いた関係分類の性能向上を目的に研究を行った。質問応答型の関係分類に対して、マーカーを用いて関係の項のタイプの情報を有効利用する方法を提案し、SemEval-2010 Task 8 のデータセットを用いて、提案手法を評価したところ、マーカーを用いないベースラインに対して、1.9%ポイントの性能向上が得られ、提案手法の有効性を示すことができた。また、解析の結果、マーカーによって用語を正しく回答できていることがわかった。今後は、1つの用語が複数の関係を持つ場合や、1組の用語ペアが複数の関係をもつ場合への拡張を行う。また、用語のタイプなどのさらなる追加情報を有効活用する方法を検討する。さらに、現状の SotA なモデルとの比較を行い、本手法の有効性を調査する。

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] Amir DN Cohen, Shachar Rosenman, and Yoav Goldberg. Relation classification as two-way span-prediction. **arXiv preprint arXiv:2010.04829v2**, 2021.
- [2] Omer Levy, Minjoon Seo, Eunsol Choi, and Luke Zettlemoyer. Zero-shot relation extraction via reading comprehension. In **Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)**, pp. 333–342, Vancouver, Canada, August 2017. Association for Computational Linguistics.
- [3] Xiaoya Li, Fan Yin, Zijun Sun, Xiayu Li, Arianna Yuan, Duo Chai, Mingxin Zhou, and Jiwei Li. Entity-relation extraction as multi-turn question answering. In **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**, pp. 1340–1350, Florence, Italy, July 2019. Association for Computational Linguistics.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. **Advances in neural information processing systems**, Vol. 32, pp. 8026–8037, 2019.
- [8] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. Huggingface’s transformers: State-of-the-art natural language processing. **arXiv preprint arXiv:1910.03771**, 2019.

A 質問テンプレートとマーカー

本研究で用いた質問文テンプレートと関係マーカーを表 4 に示す。質問文テンプレートは Cohen らが用いたもの [1] を利用し、関係マーカーは SemEval-2010 Task 8 のデータセット [4] の関係ラベル名から作成した。

表 4 SemEval-2010 Task 8 の質問文テンプレートと関係マーカー [1]

関係名	質問文テンプレート	関係マーカー
Content-Container	Q1: Where is the e_1 stored? Q2: What is stored in the e_2 ?	<content>, </content> <container>, </container>
Component-Whole	Q1: What whole is the e_1 component of? Q2: What is the component of the e_2 ?	<component>, </component> <whole>, </whole>
Product-Producer	Q1: Who produces e_1 ? Q2: What does e_2 produce?	<product>, </product> <producer>, </producer>
Instrument-Agency	Q1: Who uses a e_1 ? Q2: What does e_2 use?	<instrument>, </instrument> <agency>, </agency>
Member-Collection	Q1: What collection e_1 is part of? Q2: What is a fraction of e_2 ?	<member>, </member> <collection>, </collection>
Entity-Origin	Q1: Where does e_1 come from? Q2: What comes from e_2 ?	<entity1>, </entity1> <origin>, </origin>
Entity-Destination	Q1: What is the e_1 's destination? Q2: Who does e_2 serve as a destination?	<entity2>, </entity2> <destination>, </destination>
Message-Topic	Q1: What is the topic of the e_1 ? Q2: Who does e_2 serve as a topic?	<message>, </message> <topic>, </topic>
Cause-Effect	Q1: What caused the e_1 ? Q2: From what e_2 caused?	<cause>, </cause> <effect>, </effect>