

# 複数の形式・表現の質問を利用した多角的な関係抽出

山田 晃士 三輪 誠 佐々木 裕

豊田工業大学

{sd22439,makoto-miwa,yutaka.sasaki}@toyota-ti.ac.jp

## 概要

質問応答を用いた関係抽出手法は、各関係タイプごとにその有無を確認する質問を利用し、それぞれの関係タイプに特化した処理を可能にしている。さらに、これらの手法では、質問を変えると結果が変わる柔軟性をもつが、既存手法はそれを十分に利用できていない。本研究では、質問の柔軟性を利用し、関係そのものを複数の方法で捉える、多角的な捉え方が可能な関係抽出の実現を目指し、複数の形式・表現の質問を利用する関係抽出手法を提案する。DrugProt データセットを用いた評価によって、複数の質問形式の利用が有効であること、質問の表現による関係抽出性能への影響を確認した。

## 1 はじめに

文献中から構造化された情報を取り出す情報抽出タスクの1つとして、文献中に存在する用語間の関係を抽出する関係抽出が研究されている [1, 2, 3]。既存の関係抽出手法の多く [2, 4, 5] は、用語ペア間の関係を事前に定義された候補の関係タイプから選択し、用語ペアを分類する手法である。これらの手法では、用語ペア間の関係を候補となる関係タイプから一度の処理で選択し分類するため、それぞれの関係タイプに特化した処理ができず、各関係タイプに着目できていない。

それぞれの関係タイプに着目した関係抽出手法として、質問応答を用いた関係抽出手法 [6, 7, 8] がある。これらの手法では、まず、候補となる関係タイプすべてに対して用語をスロットとして持つ質問文テンプレートを作成し、用語ペアの片方を当てはめることで質問文を作成する。作成した質問文と用語ペアを含む元の文を用いて質問応答を行い、回答が質問文に当てはめていない方の用語であれば、テンプレートに対応する関係タイプであると判定し、関係抽出を行う。質問応答を用いる手法では、関係タイプごとの質問を用いて質問応答を行うことで、そ

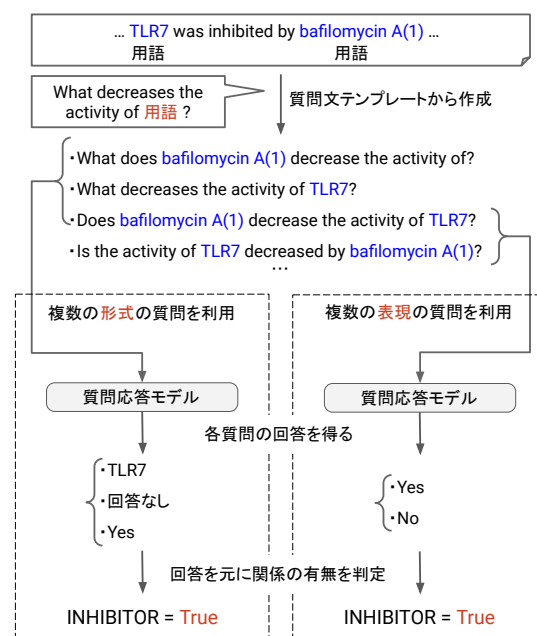


図1 複数の形式・表現の質問を用いる関係抽出

れぞれの関係タイプに特化した処理が可能になるという利点がある。さらに、質問応答を用いた関係抽出手法は、質問を変更することで関係抽出の結果が変化するという、質問の柔軟性もある。しかし、既存手法はこの柔軟性を十分に利用できていない。

そこで本研究は、質問応答を用いた関係抽出における質問の柔軟性を利用して、関係そのものを複数の方法で捉える多角的な捉え方が可能な関係抽出手法の実現を目的として、形式・表現の異なる複数の質問応答を用いる関係抽出手法を提案する。本稿で対象とする薬物タンパク質問関係抽出データセット DrugProt [9] の各関係タイプに対して、回答対象の異なる複数の形式の質問を作成し、関係抽出に利用する。また、各形式について、質問表現の異なる複数の表現の質問も作成・利用する。本研究の貢献は以下の通りである。

- 質問の柔軟性を利用して関係を多角的に捉える手法として、複数の形式・表現の質問を利用す

る手法を提案した。

- 複数の質問形式の利用が関係抽出に有効であることを示した。
- 質問の表現が関係抽出性能に影響を与えることを確認した。

## 2 関連研究

### 2.1 複数の質問を用いた関係抽出

Cohen らは、用語ペアの始点と終点の用語をそれぞれ含む2つの形式の質問を用いて、双方向の質問応答を行うことで関係抽出を行う手法 [8] を提案している。まず、候補となる各関係ラベルに対して、用語ペアの始点の用語を当てはめて終点の用語を回答する質問文のテンプレートと、終点の用語を当てはめて始点の用語を回答するテンプレートの2種類を作成する。2つの質問のうち片方でも質問文に当てはめていない用語ペア内のもう1つの用語が回答であるときに、その用語ペアはテンプレートの持つ関係タイプであると判定する。著者らは、Cohen らの手法を用いて、質問文の表現を一部変更した2つの質問文を比較し、質問文の表現が関係抽出の性能に影響を与えることを報告した [10]。

### 2.2 関係抽出に適した質問応答モデル

著者らは、関係抽出に適した質問応答モデルとして、タスク情報を利用する2値分類型の手法 [11] を提案した。まず、関係抽出タスクで事前に与えられる情報として、質問文と入力文中の用語の前後には各用語タイプのマーカーを、疑問詞の前後には回答となる用語のタイプのマーカーを付与した。また、質問応答を用いた関係抽出において必要な、回答がテンプレートに当てはめていないもう1つの用語であるかを2値分類によって直接判定するモデルを提案した。BERT (Bidirectional Encoder Representations from Transformers) [12] を用いて、回答の候補となる用語のスパン表現を作成し、2値分類を行うモデルを提案した。

## 3 提案手法

本研究では、質問の柔軟性を利用して、関係そのものを複数の質問で捉える、多角的な捉え方が可能な関係抽出手法の実現を目的とする。提案手法の概要を図1に示す。1つの関係に対する複数の質問として、複数の形式・表現の質問を利用するためのテ

ンプレートを作成した。3.1 節では回答の対象を変更した複数の形式の質問文テンプレートについて、3.2 節では、各形式の質問に対して表現を変えた複数の表現の質問文テンプレートについて、それぞれ説明する。なお、いずれのテンプレートも本稿で対象とする薬物タンパク質問関係抽出データセットの DrugProt [9] のアノテーションガイドライン [13] を参考に作成した。

### 3.1 複数の形式の質問テンプレート

1つの関係に対する複数の質問として、本稿で対象とする薬物タンパク質問関係抽出データセットの DrugProt のすべての関係タイプに対して3つの形式の質問文テンプレートを作成した。ここで、形式とは、質問応答における回答とする対象を指す。1つ目の形式は薬物とタンパク質の用語ペアに対して、薬物の用語を当てはめて、回答がタンパク質の用語であれば関係があると判定するもの (D→P) で、2つ目の形式はタンパク質の用語を当てはめて、薬物の用語であれば関係があると判定するもの (P→D) である。これらの2つの形式の質問文テンプレートは、Cohen らの双方向の質問を用いた手法 [8] を基に著者らが作成したもの [10] である。形式名の D と P はそれぞれ薬物 (Drug) とタンパク質 (Protein) を指す。3つ目の形式の質問文テンプレートは、用語ペアの両方を当てはめ、用語ペア間に関係があるかどうかを yes か no で回答する形式の質問文テンプレート (yes/no) である。

### 3.2 複数の表現の質問テンプレート

1つの関係に対して複数の質問を利用するもう1つの方法として、複数の表現の質問文テンプレートの作成を行った。ここで、表現とは質問表現を指し、3.1 節で作成した3つの形式の質問文テンプレートセットの一部に対して表現を変更することで、異なる2つの表現の質問文テンプレートセットを作成した。2つのテンプレートのうち、1つは薬物とタンパク質の用語ペアに対して、薬物の用語または対応する疑問詞を主語にしたもの (D→P-D, P→D-D, yes/no-D) で、もう1つはタンパク質の用語または対応する疑問詞を主語としたもの (D→P-P, P→D-2P, yes/no-P) である。ただし、P→D-D と yes/no-D については、3.1 節で作成した P→D と yes/no と同一のものである。質問セットの名前は、ハイフンの前が形式、後が主語となる用語タイプを指す。

表 1 複数の形式の質問を用いた結果 (%)

手法	P	R	F
文分類	71.4	77.1	74.1
D→P	76.3	76.1	76.2
P→D	77.6	76.0	76.8
yes/no	74.9	77.3	76.1
vote-1/3	69.4	83.3	75.7
vote-2/3	78.0	77.3	77.6
vote-3/3	84.0	68.8	75.6
mean	82.2	71.7	76.6

表 2 複数の表現の質問を用いた結果 (%)

手法	P	R	F
D→P-D	76.2	76.2	76.2
P→D-D	77.6	76.0	76.8
yes/no-D	74.9	77.3	76.1
D→P-P	75.0	75.9	75.3
P→D-P	76.3	76.0	76.1
yes/no-P	74.1	77.9	76.0
D→P-vote	72.1	80.4	76.0
P→D-vote	73.1	80.5	76.6
yes/no-vote	70.7	81.5	75.7
all-vote-3/6	76.2	78.7	77.4
all-vote-4/6	79.3	75.4	77.3

## 4 実験設定

### 4.1 データセット

関係抽出データセットとして、DrugProt データセット [9] を用いて、学習および開発データでの評価を行った。DrugProt データセットは薬物とタンパク質の用語を含む薬学文献のアブストラクトで構成され、薬物とタンパク質の用語間に 13 種類の関係タイプが設定されているデータセットである。また、関係が存在しない薬物とタンパク質のペアも存在する。評価指標にはマイクロ F 値を用いる。

### 4.2 複数形式の質問の利用

複数の形式の質問を利用する手法の評価のため、比較手法として、BERT の CLS トークンを用いて文分類を行う質問を用いない手法と、3.1 節で作成したそれぞれの質問形式のうち 1 つのみを用いて学習を行う手法 (D→P, P→D, yes/no) との比較を行う。また、複数の質問形式を利用する手法として、3 つの形式で学習したモデルのうち 1 つ以上で予測された関係を予測する手法 (vote-1/3)、同様に、3 つの形式のうち、2 つ以上、3 つ以上で予測された関係を採用する手法 (vote-2/3, vote-3/3)、3 つの形式の出力を平均して用いる手法 (mean) で比較を行った。付録 C に実際に使用したモデルの設定を示した。

### 4.3 複数表現の質問の利用

複数の形式の質問を利用に対する評価のため、比較手法として、それぞれの 3.2 節で作成した質問セットを用いる手法 (D→P-D/P, P→D-D/P, yes/no-D/P) との比較を行う。複数の表現の質問を利用する手法として、各形式の異なる 2 つの表現でそれぞれ学習したモデルのうち、片方でも予測された関係を採用す

るもの (D→P-vote, P→D-vote, yes/no-vote)、複数の形式・表現を利用する手法として、異なる 6 種類の質問のうち、3 つ以上または 4 つの以上で予測された関係を採用するもの (all-vote-3/6, all-vote-4/6) で比較を行う。

## 5 結果と考察

### 5.1 実験結果

複数の質問形式を用いる手法と各比較手法の結果を表 1 に示した。文分類による関係抽出手法と比較すると、質問応答を用いるどの手法もより高い F 値となっている。この結果は、関係抽出に対して質問応答を利用する有効性を示唆している。また、複数の質問形式を用いる手法は、1 つの質問形式を用いる手法と比較すると、3 つの形式のうち 2 つ以上の形式で予測された関係を採用する手法 (vote-2/3) はより高い F 値が、3 つのモデルの出力を平均して用いる手法 (mean) は同程度の F 値が、他の 2 つの手法はより低い F 値がそれぞれ得られた。この結果から、複数の質問形式の利用は、vote-2/3 のように利用する方法によっては有効であると考えられる。

複数の表現の質問を用いる手法と各比較手法の結果を表 2 に示した。1 つの表現の質問のみを用いた手法で比較すると、薬物を主語にした表現を用いた場合は、タンパク質を主語とした表現を用いた場合よりも高い F 値になる傾向が見られた。これは、DrugProt データセットではそれぞれの関係タイプが薬物の用語を始点・タンパク質を終点として定義されているため、薬物の用語が主語の質問文の方が適しているという結果になったと考えられる。また、



**表 3** 複数の質問形式の利用によって変化した事例（斜字は薬物、太字はタンパク質の用語。どちらも関係 INHIBITOR に対する事例）

入力文	… <b>TLR7</b> ligands was inhibited by <i>bafilomycin A(1)</i> …
正解した質問文	What does <i>bafilomycin A(1)</i> decrease the activity of? (D→P) Does <i>bafilomycin A(1)</i> decrease the activity of <b>TLR7</b> ? (yes/no)
誤答した質問文	What decreases the activity of <b>TLR7</b> ? (P→D)
入力文	<b>MAO-A</b> inhibition, as judged by the decrease of plasma DHPG concentration, was significantly different from placebo with moclobemide but not with <i>toloxatone</i> .
正解した質問文	Does <i>toloxatone</i> decrease the activity of <b>MAO-A</b> ? (yes/no)
誤答した質問文	What does <i>toloxatone</i> decrease the activity of? (D→P) What decreases the activity of <b>MAO-A</b> ? (P→D)

2つの表現の質問を利用した手法は、1つの表現のみを用いた場合の結果に比べて、性能の向上が見られなかった。使用した質問の表現が2つであり、そのうちのどちらか一方でも予測されれば、その関係タイプを採用しているため、どちらかの予測での誤りを排除できないことや、タンパク質を主語とする表現の質問（D→P-P など）を用いた際の性能が薬物を主語とする質問（D→P-D など）を用いた場合と比較して低いことなどが原因と考えられる。また、複数の形式・表現を同時に用いる手法（all-vote-3/6, 4/6）は複数の表現のみを用いる手法（D→P-vote, P→D-vote, yes/no-vote）と比較した場合は性能の向上が見られるが、表 2 の複数の形式のみを用いる手法（vote-2/3）と比較しても性能の向上が見られないのも同様の原因であると考えられる。

以上の結果より、複数の質問形式の利用は、関係抽出に対して有効であると考えられる。一方で、今回作成した2つの表現の質問の利用による関係抽出の性能向上は得られなかった。しかし、それぞれの表現ごと抽出結果から、質問の表現の違いによる関係抽出結果への影響が見られたため、表現の選択によって、複数の表現の質問が関係抽出に有効に作用する可能性もあると考えられる。

## 5.2 解析

複数の形式の質問文を利用することによる関係抽出結果への影響を、複数の形式を用いた手法（vote-2/3）によって結果が変化した事例を用いて確認する。表 3 に、複数の質問形式によって、結果が改善した事例と改悪した事例をそれぞれ示した。上の事例では、関係 INHIBITOR に対する3つ形式の質問文のうち、タンパク質を当てはめて薬物を回答する形式の質問は誤った回答によって関係が正

しく予測できていないが、他の2つの形式は正しく予測できているため、3つの形式を合わせて利用することで、正しく関係 INHIBITOR を予測することができている。一方で、下の事例も同じく、関係 INHIBITOR に対する事例だが、yes か no で回答する形式のみ正しく予測できているため、他の2つの形式では誤った予測をしているため、3つの形式の利用した手法では、関係 INHIBITOR がないと誤った予測をしてしまっている。

## 6 おわりに

本研究では、質問応答を用いた関係抽出における質問の柔軟性を利用して、関係そのものを複数の方法で質問することで捉えるという多角的な捉え方が可能な関係抽出手法の実現を目的に、複数の形式・表現の質問文を利用することを提案し、そのための質問文テンプレートの作成を行った。DrugProt データセットで評価した結果、複数の形式の質問によって、1つの形式のみを用いた場合よりも高い性能が得られることを示した。一方で、複数の表現の質問を用いた手法では、性能の向上は見られなかったが、各表現の質問ごとに性能に差異が見られ、質問の表現が関係抽出の結果に影響を与えることを示した。

今後は、関係そのものを複数の方法で質問するのではなく、関係が持つ複数の属性や性質といった側面に対する質問の利用など、質問の柔軟性の異なる利用方法を検討する。また、どのような質問文をいくつ利用するのが有効かは明確でなく、すべて人手で設計するのは困難であるため、質問もしくはそれに相当する入力を自動で作成する手法も検討する。

## 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] 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.
- [2] 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.
- [3] 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 Anna Korhonen, David Traum, and Lluís Màrquez, editors, **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**, pp. 764–777, Florence, Italy, July 2019. Association for Computational Linguistics.
- [4] Christoph Alt, Marc Hübner, and Leonhard Hennig. Improving relation extraction by pre-trained language representations. In **Automated Knowledge Base Construction (AKBC)**, 2019.
- [5] Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski. Matching the blanks: Distributional similarity for relation learning. In Anna Korhonen, David Traum, and Lluís Màrquez, editors, **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**, pp. 2895–2905, Florence, Italy, July 2019. Association for Computational Linguistics.
- [6] 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.
- [7] 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.
- [8] Amir DN Cohen, Shachar Rosenman, and Yoav Goldberg. Supervised relation classification as two-way span-prediction. In **4th Conference on Automated Knowledge Base Construction**, 2022.
- [9] Martin Krallinger, Obdulia Rabal, Antonio Miranda-Escalada, and Alfonso Valencia. DrugProt corpus: Biocreative VII Track 1 - Text mining drug and chemical-protein interactions, June 2021.
- [10] 山田晃士, 三輪誠, 佐々木裕. 複数の質問形式を利用した分類型の質問応答による薬物タンパク質間関係抽出. 言語処理学会第 29 回年次大会, 2023.
- [11] Koshi Yamada, Makoto Miwa, and Yutaka Sasaki. Biomedical relation extraction with entity type markers and relation-specific question answering. In Dina Demner-fushman, Sophia Ananiadou, and Kevin Cohen, editors, **The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks**, pp. 377–384, Toronto, Canada, July 2023. Association for Computational Linguistics.
- [12] 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.
- [13] Obdulia Rabal, Jose Antonio López, Astrid Lagreid, and Martin Krallinger. DrugProt corpus relation annotation guidelines [ChemProt - Biocreative VI], June 2021.
- [14] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, **3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings**, 2015.
- [15] 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 Transactions on Computing for Healthcare**, Vol. 3, No. 1, p. 1–23, Jan 2022.

## A 作成した質問文テンプレート

3.1 節の複数の質問形式のテンプレートの例を表 4 に、3.2 節の複数表現の質問テンプレートの例を表 5 にそれぞれ示す。

表 4 関係 INHIBITOR に対する複数形式の質問テンプレート (DRUG は薬物, PROTEIN はタンパク質の用語)

形式	質問文テンプレート
D→P	What does DRUG decrease the activity of?
P→D	What decreases the activity of PROTEIN?
yes/no	Does DRUG decrease the activity of PROTEIN?

表 5 関係 INHIBITOR に対する複数表現の質問テンプレート (DRUG は薬物, PROTEIN はタンパク質の用語)

表現	質問文テンプレート
D→P-D	What does DRUG decrease the activity of?
P→D-D	What decreases the activity of PROTEIN?
yes/no-D	Does DRUG decrease the activity of PROTEIN?
D→P-P	What activity is decreased by DRUG?
P→D-P	What is the activity of PROTEIN decreased by?
yes/no-P	Is the activity of PROTEIN decreased by DRUG?

## B データセットの統計

実験に使用した DrugProt データセット [9] の統計を表 6 に示した。ただし、DrugProt データセットはテストデータのラベルが公開されていないため、開発データを約 1:1 となるように分割してそれぞれ開発、テストデータとしている。

表 6 訓練、開発、テストセットの各関係タイプの事例数

関係	訓練	開発	テスト
INDIRECT-DOWNREGULATOR	1,329	168	164
INDIRECT-UPREGULATOR	1,378	153	149
DIRECT-REGULATOR	2,247	230	228
ACTIVATOR	1,428	121	125
INHIBITOR	5,388	575	575
AGONIST	658	66	65
AGONIST-ACTIVATOR	29	4	6
AGONIST-INHIBITOR	13	1	1
ANTAGONIST	972	111	107
PRODUCT-OF	920	81	77
SUBSTRATE	2,003	245	249
SUBSTRATE.PRODUCT-OF	24	1	2
PART-OF	885	130	127
Total	17,274	1,886	1,875

## C モデル設定

実験において用いるモデルの設定について述べる。モデルには、2.2 節のタスク情報を利用する 2 値分類型の質問応答モデル [11] を利用した。モデルは、質問文に含まれない用語が回答のときに関係があると判定する質問形式に対して、回答の

候補となる用語のスパン表現から、その用語が回答かどうかを 2 値分類するモデルである。yes か no で回答する質問形式に対しては、回答の候補が入力文中にないため、疑問符のスパン表現から回答が yes かどうかで 2 値分類を行った。最適化手法には Adam [14] を、事前学習モデルには PubMedBERT-base-uncased-abstract-fulltext [15] を用いた。表 7 に実験で用いたパラメータを示した。各実験はシード値を 40-42 に設定して 3 回学習・評価を行い、その平均をとった値を最終的な評価値とした。

表 7 実験で利用するモデルの設定

パラメータ	値
学習率	3e-6
ミニバッチサイズ	32
エポック数	10
2 値分類の閾値	0.7