Biomedical Relation Extraction with Entity Type Markers and Relation-specific Question Answering

Koshi Yamada, Makoto Miwa, and **Yutaka Sasaki**
Computational Intelligence Laboratory
Toyota Technological Institute
2-12-1 Hisakata, Tempaku-ku, Nagoya, Aichi, 468-8511, Japan
{sd22439, makoto-miwa, yutaka.sasaki}@toyota-ti.ac.jp

Abstract

Recently, several methods have tackled the relation extraction task with QA and have shown successful results. However, the effectiveness of existing methods in specific domains, such as the biomedical domain, is yet to be verified. When there are multiple entity pairs that share an entity in a sentence, a QA-based relation extraction model that outputs only one single answer to a given question may not extract desired relations. In addition, these methods employ QA models that are not tuned for relation extraction. To address these issues, we first extend and apply a span QA-based relation extraction method to the drug-protein relation extraction by creating question templates and incorporating entity type markers. We further propose a binary QA-based method that directly uses the entity information available in the relation extraction task. The experimental results on the DrugProt dataset show that our QA-based methods, especially the proposed binary QA method, are effective for drug-protein relation extraction. Our source code is available at https://github.com/tticoin/BioRE-ETM-QA

1 Introduction

In recent years, several methods that tackle the relation extraction (RE) task with question answering (QA) have shown impressive results (Levy et al., 2017; Li et al., 2019; Cohen et al., 2021). Instead of directly finding a relation between an entity pair in a sentence, in these methods, questions relating to the relation between the entity pair are answered by a QA model to determine whether the entity pair has the relation.

For example, suppose that we have the following sentence "Tom was born in 1999". To determine whether the relation *born_in* exists or not between the entity pair Tom and 1999, Cohen et al. (2021) builds two questions, each containing one of the two entities, e.g., "When was Tom born?" for an entity Tom and "Who was born in 1999?" for an entity 1999. If the answer from a span QA model is the other entity that is not in the question for one of these questions, the method determines that there is a relation "born_in" between the pair. Such a QA-based RE method can focus on each relation by preparing different questions for different relations.

However, existing methods have been applied only to general domain datasets, and their effectiveness has not been verified in specific domains, such as the biomedical domain. When there are multiple entity pairs that share an entity in a sentence, a QA model that can output only one answer to a question cannot extract relations since the questions become identical for the same entity. For example, when using a question "Who was born in 1999?" to determine if there is a relationship born_in between the entity pair ("Tom", "1999") and ("John", "1999") in the sentence "John was born in 1999, like Tom.," the answer will be either Tom or John because the input to the model is the same. In this case, it is determined that there is no relation "born_in" for one of the entity pairs. Moreover, the employed QA models are not tuned for RE.

In this study, we extend and apply a span QA-based RE method (Cohen et al., 2021) for drug-protein RE to check the effectiveness of the method in the biomedical domain task. For this purpose, we manually create question templates for a drug-protein RE dataset DrugProt (Krallinger et al., 2021). To deal with each entity pair, even if multiple entity pairs share the same entity, we introduce entity type markers that incorporate entity types available in the RE task into the question and input sentence. Furthermore, we propose a binary QA-based method that directly classifies whether the answer to a question, which includes one entity of the target entity pair, is the other entity of the pair or not.

Our contributions are summarized as follows:

- We extend and apply a span QA-based RE method to the DrugProt drug-protein RE

The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks, pages 377–384

dataset by creating question templates and incorporating entity type markers.

- We propose a binary QA-based method that directly uses the entity information available in the RE task.
- We show that the binary QA-based method is effective for drug-protein RE.

2 Related Work

Levy et al. (2017) have proposed the method of zero-shot RE using QA. Using an existing QA model, their method could make predictions even for unseen relation types by preparing questions corresponding to the relation types. Li et al. (2019) proposed the method for entity-relation extraction by multi-turn QA. The hierarchical structure of the relation types can be captured by extracting entities with QA at each turn and using them to construct questions at the next turn.

Cohen et al. (2021) proposed the RE method that uses QA with two-way questions, each containing a head or tail entity. First, they created two question templates for each relation type. One is a template in which the tail entity is answered using the head entity, and the other is a template in which the head entity is answered using the tail entity. If at least one of the answers to the two questions corresponds to the entity not used in the questions from the templates, it is determined that the entity pair has a relation type for which the template is created. They extract all the relation types by repeating this process for all the relation types.

QA-based methods have also been applied to other tasks, such as named entity recognition (Li et al., 2020b), event extraction (Li et al., 2020a; Liu et al., 2020), and coreference resolution (Wu et al., 2020). In the biomedical domain, the effectiveness of QA-based methods has been verified in event extraction (Wang et al., 2020) and spatial information extraction (Datta and Roberts, 2022).

3 Method

This section describes our RE models that employ the span QA-based RE following (Cohen et al., 2021) to drug-protein RE and proposes a binary QA-based method. We first introduce the question templates to apply QA-based RE to drug-protein RE in Section 3.1. We next introduce entity type markers, which incorporate the information about the entity types into the input sentence and question in Section 3.2. We then explain a span QA-based method that predicts the start and end of the answer span in Section 3.3. We finally propose a binary QA-based method in Section 3.4.

3.1 Question templates

We created question templates for all relation types in the DrugProt dataset (Krallinger et al., 2021) to apply the RE method using QA to drug-protein RE. For each relation, we created two types of question templates: one is a template with a slot for the drug and the other for the protein, following the two-way QA format in (Cohen et al., 2021). DrugProt has 13 relation types, so the total number of question templates is 26. These question templates were created based on the annotation guidelines (Rabal et al., 2021) for the dataset. For example, the template for the relation "INDIRECT-DOWNREGULATOR" was set to "What does DRUG downregulate via other targets?" and "What downregulates PROTEIN via other targets?" using the part of the definition in the annotation guideline "CEM downregulates GPRO via other target" (CEM stands for drug and GPRO stands for protein). The templates are shown in Table 3 of Appendix A.

3.2 Entity type marker

We insert entity type markers into the input sentence and question to make the input for each entity pair unique and to use the entity types available in the RE task as input to the QA models. For each target entity pair, we assign the markers to the drug and protein entities in the input sentence and the question. The interrogatives are also assigned markers corresponding to the answer entity type to link the question and the answer.

Table 1 shows examples of inserting markers into input sentences and questions. First, the drug-protein pair (TT-B, kinase) and (TT-B, EGFR) are created from the entities in the input sentence "The kinase activity of EGFR was little inhibited by TT-B in a cell-free system". Next, drug markers (<D> and </D>) and protein markers (<P> and </P>) are assigned to the target entity pairs in the input and question sentences. In addition, the markers corresponding to the answer are assigned to the interrogative "What" in the question.

3.3 Span QA-based RE

To evaluate the RE method using QA, we build a span QA model that predicts the start and end indices of the answer span, as shown in Figure 1

Input sentences	Questions
The <P>*kinase*</P> activity of *EGFR* was little inhibited by <D>*TT-B*</D> in a cell-free system.	<P>What</P> does <D>*TT-B*</D> decrease the activity of? <D>What<D> decreases the activity of <P>*kinase*</P>?
The *kinase* activity of <P>*EGFR*</P> was little inhibited by <D>*TT-B*</D> in a cell-free system.	<P>What</P> does <D>*TT-B*</D> decrease the activity of? <D>What<D> decreases the activity of <P>*EGFR*</P>?

Table 1: Example of the questions and the input sentences with entity type markers for a relation type INHIBITOR. The italic text represents entities, and the markers P and D denote protein and drug, respectively.

following Cohen et al. (2021). First, entity type markers are assigned to the question and the input sentence, as shown in Section 3.2. Next, a sequence $S = \{w_1, w_2, \cdots w_n\}$, which is the concatenation of the question and the input sentence with the [SEP] token between them, is input to Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) to obtain the representation of each token $\mathbf{h}_i$.

$$\{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n\} = \mathrm{BERT}(w_1, w_2, \ldots w_n) \quad (1)$$

From the representation of each token, the probabilities p_i^s and p_i^e that indicate whether the i-th token is the start and end tokens of the answer span are obtained by applying the fully connected layer and the softmax function.

$$\mathbf{z}_i^s = \mathbf{W}^s \mathbf{h}_i + \mathbf{b}^s \quad (2)$$

$$\mathbf{z}_i^e = \mathbf{W}^e \mathbf{h}_i + \mathbf{b}^e \quad (3)$$

$$\{p_1^s, p_2^s, \cdots\} = \mathrm{Softmax}(\mathbf{z}_1^s, \mathbf{z}_2^s, \ldots) \quad (4)$$

$$\{p_1^e, p_2^e, \ldots\} = \mathrm{Softmax}(\mathbf{z}_1^e, \mathbf{z}_2^e, \ldots) \quad (5)$$

If at least one answer span of two-way questions is the entity, the target entity pair is determined to have the relation type corresponding to the question. When the answer span is the entity for multiple relation types, the target entity pair is assumed to have multiple relation types.

3.4 Binary QA-based relation extraction

In the method of Section 3.3, the agreement between the answer span of the question and the entity span is checked. We consider this too much since it is sufficient to determine whether the answer is the target entity or not for RE. Therefore, we propose a binary QA-based method that uses a binary classification to determine whether the entity is the answer to the question based on the span representation of the entity that is not included in the question text, as shown in Figure 2.

In this method, the representations $\boldsymbol{h}_i$ in Equation (1) of Section 3.3 are used to create the span representation of the candidate answer entity. The span representation is created by concatenating the representations of the first and last tokens of the span and the mean of the representations of all tokens in the span, as shown below.

$$\mathbf{h}^{mean} = \mathrm{mean}(\mathbf{h}_s, \cdots, \mathbf{h}_e) \quad (6)$$

$$\mathbf{h}^{span} = \mathrm{Concat}(\mathbf{h}_s, \mathbf{h}^{mean}, \mathbf{h}_e) \quad (7)$$

Here, s and e denote the start and end positions of the entity span, and $\mathbf{h}^{span}$ denotes the span representation. The probability that the entity is the answer p^{span} is obtained by applying a fully connected layer and a sigmoid function to this span representation.

$$p^{span} = \mathrm{Sigmoid}(\mathbf{W}^{span}\mathbf{h}^{span} + \mathbf{b}^{span}) \quad (8)$$

Similarly to Section 3.3, classification is performed in two ways for all relation types, and an entity pair can have multiple relation types.

4 Experimental Settings

4.1 Dataset

We used the DrugProt drug-protein RE dataset for the evaluation. The dataset consists of abstracts from the pharmaceutical literature that contain drug and protein mentions and 13 types of drug-protein relations. DrugProt originally consisted of training, development, and test data, but the labels for the test data were not publicly available, so the documents in the original development set were split into 1:1 and used for our development and test data. Table 4 in Appendix B shows the statistics of the data. As a preprocessing, each document was split into sentences using SciSpacy (Neumann et al., 2019). We used the micro-averaged F-score for all relations for evaluation, which is calculated using evaluation scripts provided by the task organizer.

4.2 Compared methods

To verify the effectiveness of the QA-based methods, we compare the span QA-based method (SpanQA) in Section 3.3 and the binary QA-based method (BinQA) in Section 3.4. In order to evaluate the QA-based RE with the templates created

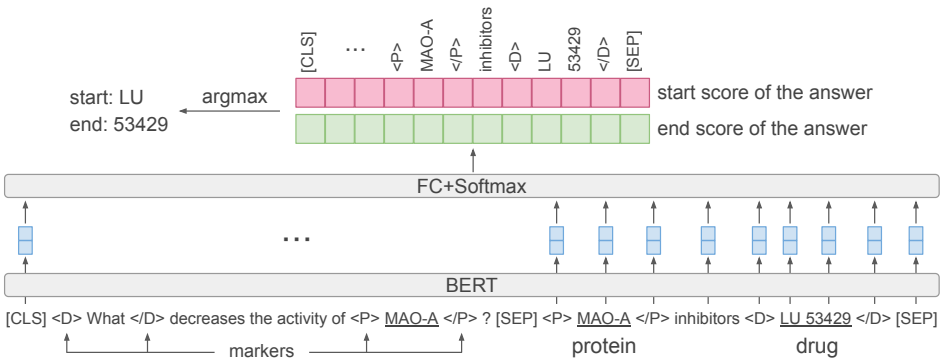

Figure 1: Span QA-based RE

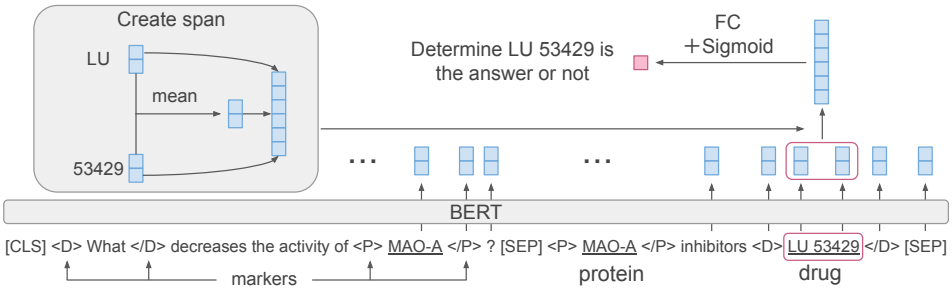

Figure 2: Binary QA-based RE

in Section 3.1, we experiment with the method using special tokens (STQA) specific to the relation types instead of the question templates to ensure that the natural language questions are effective or not and the method of RE by sentence classification using BERT's CLS tokens (CLS). In STQA, for the relation "INHIBITOR", instead of the natural language templates "What does DRUG decrease the activity of?" and "What decreases the activity of PROTEIN?", the templates with special tokens "<INHIBITOR> DRUG?" and "</INHIBITOR> PROTEIN?" are used. Furthermore, to confirm the effectiveness of the entity type markers, we evaluated BinQA without the markers.

All methods used PubMedBERT-base-uncased-abstract-fulltext (Gu et al., 2022), which has been pre-trained using a large biomedical literature, as an encoder. Adam (Kingma and Ba, 2015) was used as the optimization method, and the learning rate was set to 3e-6. The threshold for the binary QA-based method described in Section 3.4 was set to 0.7. In addition, we trained and evaluated the methods five times with different seeds and calculated the average score.

5 Results

Table 2 shows the micro-averaged F-scores for the span QA-based methods (SpanQA), the proposed method (BinQA), the method using special tokens (STQA) instead of question templates, and the RE method using sentence classification (CLS)[1]. BinQa improved the micro-averaged F-score by 1.6 percentage points on the test set compared to the SpanQA. These results indicate that the binary QA-based method is effective for RE. In a comparison of BinQA and STQA, the proposed method has a 1.0 percentage point higher micro-averaged F-score than STQA on the test set, indicating the effectiveness of using natural language for the questions. In addition, a comparison between the QA-based methods (SpanQA, BinQA, and STQA) and the sentence classification method (CLS) shows that the QA-based methods are also effective in drug-

	dev	test
CLS	74.2	74.9
SpanQA	75.5	75.8
STQA	76.2	76.4
BinQA	**76.3**	**77.4**
BinQA w/o markers	74.7	75.3

Table 2: Comparison of QA-based and classification-based methods in the Micro-averaged F-scores (%) on the development and test sets.

[1]We summarize the RE performance for each relation type in Appendix C

protein RE. When the markers were removed from BinQA, the micro-averaged F-score decreased by 2.1 percentage points on the test set. These results show that the entity type markers are effective.

6 Conclusions

In this paper, we applied and extended span QA-based RE method to the drug-protein RE by creating question templates and incorporating entity type markers. We also proposed the binary QA-based method that directly uses the entity information available in the RE task. We showed the QA-based methods, especially the proposed binary QA method, are effective for drug-protein RE. For future work, we will analyze the advantages and problems of the proposed QA-based methods by comparing them with other RE methods in the biomedical domain. In addition, we will analyze difficulties in the biomedical domain by applying and comparing QA-based methods in the general domain.

Limitations

In this study, we used entity type markers to enable a QA model to output multiple answers for RE for multiple entity pairs that share an entity in a sentence. However, we have not evaluated other QA models that can output multiple answers. In addition, since entity type markers are incorporated before and after entities, it is difficult to apply the method to nested entity pairs.

QA-based RE methods are not computationally efficient since they require multiple QA processes to extract the relation for an entity pair. Furthermore, the question templates in QA-based RE affect the performance of RE, but in this study, only one template set was used for evaluation, and there is room for improvement.

Direct comparison with existing SotA (Weber et al., 2022) is not performed because of the unavailability of the original DrugProt test data set. In addition, the SotA model is an ensemble of 10 pretrained language models, and we cannot directly compare the performance in a fair setting.

Acknowledgements

This work was supported by JSPS Grant-in-Aid for Scientific Research JP20K11962.

References

Amir DN Cohen, Shachar Rosenman, and Yoav Goldberg. 2021. Relation classification as two-way span-prediction. *arXiv preprint arXiv:2010.04829v2*.

Surabhi Datta and Kirk Roberts. 2022. Fine-grained spatial information extraction in radiology as two-turn question answering. *International Journal of Medical Informatics*, 158:104628.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. 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)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. 2022. Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare*, 3(1):1–23.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Martin Krallinger, Obdulia Rabal, Antonio Miranda-Escalada, and Alfonso Valencia. 2021. DrugProt corpus: Biocreative VII Track 1 - Text mining drug and chemical-protein interactions.

Omer Levy, Minjoon Seo, Eunsol Choi, and Luke Zettlemoyer. 2017. Zero-shot relation extraction via reading comprehension. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 333–342, Vancouver, Canada. Association for Computational Linguistics.

Fayuan Li, Weihua Peng, Yuguang Chen, Quan Wang, Lu Pan, Yajuan Lyu, and Yong Zhu. 2020a. Event extraction as multi-turn question answering. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 829–838, Online. Association for Computational Linguistics.

Xiaoya Li, Jingrong Feng, Yuxian Meng, Qinghong Han, Fei Wu, and Jiwei Li. 2020b. A unified MRC framework for named entity recognition. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5849–5859, Online. Association for Computational Linguistics.

Xiaoya Li, Fan Yin, Zijun Sun, Xiayu Li, Arianna Yuan, Duo Chai, Mingxin Zhou, and Jiwei Li. 2019. Entity-relation extraction as multi-turn question answering. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1340–1350, Florence, Italy. Association for Computational Linguistics.

Jian Liu, Yubo Chen, Kang Liu, Wei Bi, and Xiaojiang Liu. 2020. Event extraction as machine reading comprehension. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1641–1651, Online. Association for Computational Linguistics.

Mark Neumann, Daniel King, Iz Beltagy, and Waleed Ammar. 2019. ScispaCy: Fast and robust models for biomedical natural language processing. In *Proceedings of the 18th BioNLP Workshop and Shared Task*, pages 319–327, Florence, Italy. Association for Computational Linguistics.

Obdulia Rabal, Jose Antonio López, Astrid Lagreid, and Martin Krallinger. 2021. DrugProt corpus relation annotation guidelines [ChemProt - Biocreative VI].

Xing David Wang, Leon Weber, and Ulf Leser. 2020. Biomedical event extraction as multi-turn question answering. In *Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis*, pages 88–96, Online. Association for Computational Linguistics.

Leon Weber, Mario Sänger, Samuele Garda, Fabio Barth, Christoph Alt, and Ulf Leser. 2022. Chemical–protein relation extraction with ensembles of carefully tuned pretrained language models. *Database*, 2022. Baac098.

Wei Wu, Fei Wang, Arianna Yuan, Fei Wu, and Jiwei Li. 2020. CorefQA: Coreference resolution as query-based span prediction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6953–6963, Online. Association for Computational Linguistics.

A Question templates

Table A shows the question templates for the DrugProt dataset. Two templates are defined for each relation type, and each template consists of a question with a slot for a drug or a protein. Before inputting the model, the DRUG or PROTEIN of the template is replaced with the target entity.

B Dataset statistics

Table 4 shows dataset statistics for the DrugProt dataset. As explained in Section 4.1, the documents in the original development data were split and used for our development and test data. The training data consist of 3,500 documents, and the development and test data consist of 375 documents each.

C RE performance for each relation type

Tables 5 and 6 show the F-scores for each relation type and micro-averaged F-scores for the methods described in Section 4.2. We trained and evaluated the methods five times with different seeds and calculated the average scores.

Relation	Question templates
INDIRECT-DOWNREGULATOR	What does DRUG downregulate via other targets? What downregulates PROTEIN via other targets?
INDIRECT-UPREGULATOR	What does DRUG upregulate via other targets? What upregulates PROTEIN via other targets?
DIRECT-REGULATOR	What does DRUG directly regulate? What directly regulates PROTEIN?
ACTIVATOR	What does DRUG increase the activity of? What increases the activity of PROTEIN?
INHIBITOR	What does DRUG decrease the activity of? What decreases the activity of PROTEIN?
AGONIST	What does DRUG act as an agonist to? What acts as an agonist to PROTEIN?
AGONIST-ACTIVATOR	What does DRUG increase activity by acting as an agonist to? What acts as an agonist on PROTEIN to increase its activity?
AGONIST-INHIBITOR	What does DRUG decrease activity by acting as an agonist to? What acts as an agonist on PROTEIN to decrease its activity?
ANTAGONIST	What does DRUG act as an antagonist to? What acts as an antagonist to PROTEIN?
PRODUCT-OF	What produces DRUG by the enzymatic reaction? What is the product of the enzymatic reaction of PROTEIN?
SUBSTRATE	What does DRUG act as a substrate for? What acts as a substrate for PROTEIN?
SUBSTRATE_PRODUCT-OF	What causes the enzymatic reaction with DRUG as substrate and product? What is the substrate and product of the enzymatic reaction of PROTEIN?
PART-OF	What has a structural relationship to DRUG? What is structurally related to PROTEIN?

Table 3: Question templates for DrugProt

Relation	train	dev	test
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

Table 4: Data statistics in our settings for the DrugProt dataset. We split the original development data into our development and test data.

Relation	CLS	SpanQA	STQA	BinQA	BinQA w/o markers
INDIRECT-DOWNREGULATOR	72.2	75.6	76.5	75.9	76.1
INDIRECT-UPREGULATOR	79.6	77.0	79.5	78.6	78.4
DIRECT-REGULATOR	59.8	60.9	59.4	62.1	57.9
ACTIVATOR	75.2	78.7	78.9	78.6	75.0
INHIBITOR	81.4	83.0	84.4	83.4	82.0
AGONIST	82.5	84.1	81.8	84.9	81.8
AGONIST-ACTIVATOR	0	0	0	0	0
AGONIST-INHIBITOR	0	20.0	20.0	20.0	60.0
ANTAGONIST	91.6	91.9	91.9	93.2	89.4
PRODUCT-OF	60.2	61.5	67.1	62.2	63.6
SUBSTRATE	64.7	65.4	66.9	69.7	66.0
SUBSTRATE_PRODUCT-OF	0	0	0	0	0
PART-OF	73.7	73.2	72.7	72.8	75.1
Micro-averaged F-score	74.2	75.5	76.2	76.3	74.7

Table 5: RE performance (%) of the QA-based and classification-based methods on the development set.

Relation	CLS	SpanQA	STQA	BinQA	BinQA w/o markers
INDIRECT-DOWNREGULATOR	71.9	73.0	77.1	77.2	71.8
INDIRECT-UPREGULATOR	68.7	71.8	72.6	74.5	73.3
DIRECT-REGULATOR	70.6	71.0	70.9	73.9	69.7
ACTIVATOR	71.2	73.5	73.0	74.0	72.8
INHIBITOR	85.4	85.8	85.4	86.7	85.4
AGONIST	66.4	68.6	68.3	68.7	64.3
AGONIST-ACTIVATOR	0	23.9	0	0	44.3
AGONIST-INHIBITOR	0	13.3	0	0	0
ANTAGONIST	84.6	90.9	89.2	90.6	87.6
PRODUCT-OF	60.9	58.5	59.7	63.9	60.9
SUBSTRATE	67.3	67.8	69.3	69.8	68.0
SUBSTRATE_PRODUCT-OF	0	0	0	0	0
PART-OF	71.4	70.3	70.7	68.9	68.6
Micro-averaged F-score	74.9	75.8	76.4	77.4	75.3

Table 6: RE performance (%) of the QA-based and classification-based methods on the test set.