

# BertSRC: BERT-based Semantic Relation Classification

Yeawon Lee

Yonsei University

Jinseok Son

Yonsei University

Min Song (✉ [min.song@yonsei.ac.kr](mailto:min.song@yonsei.ac.kr))

Yonsei University

---

## Research Article

**Keywords:** Relation extraction, Semantic relation classification, Corpus construction, Annotation method, Deep learning, BERT, Fine-tuning

**Posted Date:** March 16th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1425623/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

The relationship between biomedical entities is complex, and many relationships have not yet been identified. By applying semantic relation classification techniques to the accumulated biomedical literature, existing relationships between biomedical entities that can help to infer previously unknown relationships are efficiently grasped. To develop semantic relation classification models, constructing a training dataset that is manually annotated by biomedical experts with semantic relationships among biomedical entities, such as genes, diseases, and compounds, is essential. In this paper, we aim to build such a dataset. To validate the usability of the training dataset and to improve the performance of the relation extraction task, we built a relation classification model based on bidirectional encoder representations from transformers (BERT) trained on our dataset, applying our newly proposed fine-tuning methodology. In experiments comparing performance among several models based on different deep learning algorithms, our model with the proposed fine-tuning methodology showed the best performance. The experimental results show that the constructed training dataset is an important information resource for the development and evaluation of semantic relationship extraction models. Furthermore, relationship extraction performance can be improved by integrating our proposed fine-tuning methodology into semantic relation algorithms. Therefore, this can lead to the promotion of future text mining research in the biomedical field.

## 1. Introduction

Biomedical literature is rapidly accumulating, and a large amount of this information is in the form of raw text, making it difficult to easily gather details on topics of interest. Text mining techniques can be effectively applied to vast literature for key information extraction, thus facilitating biomedical research processes [1].

Among the various biomedical text mining techniques evolving as a result of research achievements in the natural language processing (NLP) field, relation extraction is critical. Relation extraction is defined as extracting connections between entities in literature. There are several types of relation extraction, including semantic relations, grammatical relations, negations, and coreferences, depending on the focus and aim of the task [2].

Specifically, researchers primarily focus on semantic relations in bio-text mining to identify various relationships between bio-entities and to infer undiscovered knowledge. This perspective of information extraction motivates limiting our scope to semantic relation extraction. Semantic relation classification in the biomedical field enables the automatic extraction of relationships between biomedical entities such as diseases, medications, chemicals, genes/proteins, or medical tests from a particular work. Therefore, new relationships can be inferred, allowing scientific hypotheses or new knowledge to be discovered or confirmed by identifying mechanisms of interaction between these entities or pathways to target materials; this in turn facilitates biological or new drug development research, biological database curation, drug repositioning, and clinical decision making [3, 4].

In the machine learning field, relation extraction is a classification task that predicts whether there is any semantic interaction between two entities (binary-class classification) or what type of relation belongs among multiple predefined relation types (multi-class classification) from a sequence or features set containing these entities. This usually involves annotating unstructured natural language text with named entities and relationships between them. This training dataset manually annotated by experts is essential in relation extraction but expensive. Constructing a dataset with semantic relation information is more complex and difficult than constructing a corpus annotated with named entities. Entity annotation is a task of simply recognizing bio-instances and categorizing them into a proper type, whereas relation annotation takes entity annotation as a prerequisite and determines the semantic interaction between two entities, which further relies on human judgment. This requires considerable time and resources to build a relation-annotated dataset.

Therefore, intending to promote relation classification research in the biomedical field, after reviewing the available benchmark relational corpora (usually for protein-protein interactions (PPIs)), this study presented our newly built corpus annotated with unique relation types on a meaningful scale. Then, to demonstrate the feasibility of our corpus, we built a bidirectional encoder representations from transformers (BERT)-based relation classification model, called BertSRC, trained on our dataset. Moreover, we proposed a new fine-tuning methodology regarding formatting input tokens for BERT, which is the second contribution of our study. The corpus for semantic relation classification and the BertSRC code are publicly available at <https://github.com/tsmmbio/BertSRC>.

## 1.1. Construction of a Training Dataset Annotated with Semantic Relations

Semantic relation classification in the biomedical field has been studied primarily as part of shared tasks aimed at evaluating and advancing natural language processing (NLP) techniques. Currently, most prestigious datasets tagged with semantic relations are from these tasks, such as BioNLP shared tasks on the recognition of biological events, which introduced the BioNLP-09, 11, 13 event corpus, and BioCreative shared tasks on PPI extraction, which generated the BioCreative-II relation corpus [5]. In most of these corpora, the entities that mainly receive attention are genes/proteins, and much of the focus is centered around the relationship between them [6, 7]. Examples of such PPI corpora include LLL, BioInfer, IEPA, and HPRD50 [1]. Adding semantic relation annotations to existing popular corpora to use them as training data for relation extraction has also been attempted. Huang et al. [8] proposed the Revised JNLPBA by expanding the original JNLPBA corpus, an NER corpus, in order to utilize it for biomedical event extraction (PPIE, BEE).

Among these various semantic relation corpora, only those annotated with binary relations are included as the main scope of our study. A binary relation is the type of relationship where a pair of two entities come as arguments. Since this type of relation is easy to understand and make use of, it is the corpus annotated with this type of relation that most current information extraction (IE) systems require. For this reason, we deliberately excluded complex relations or events where a different level of relations could be

nested and become an argument for the other relation along with the named entities from the scope of our study. The table comparing several representative benchmark corpora annotated with binary relations is provided in Appendix A. We gathered information about these corpora by reviewing each corpus or through the literature presenting or explaining the corpus [9].

According to Appendix A, some corpora annotate only genes/proteins as entities, while others annotate other participating entity types exhaustively along with genes/proteins. With regard to relation type, there are some corpora that do not define separate relation types at all; in contrast, they only determine the presence or absence of a connection within the scope they have determined. Other corpora explicitly define relation types and express dynamic hierarchical relationships among them, forming complex ontologies. The former case has a limitation in that there is only limited utility for bio-researchers because various aspects of the relation between bio-entities and semantic relationships between those relations cannot be expressed. On the other hand, the latter has the advantage of being able to reflect detailed characteristics and meaning of relations. However, introducing many classes, such as BIOINFER, which has more than 60 relationship types, requires a larger dataset, a more thoughtfully designed structure, and a complicated parameter tuning process in order to make a model generalize well. This could limit the applicability of the corpus compared with the dataset working with the model without much configuration.

Therefore, we present a new corpus annotated with semantic relation types differentiated from those of the existing corpora to provide researchers with useful and easily implemented resources for bio-text mining. Regarding entity, our corpus exhaustively covers many types including genes/proteins as long as they form meaningful relations. The relation type is divided into two according to the presence or absence of causality, and if there is no causality, it is considered an undirect link, which is a negative example. On the other hand, a direct link, which is judged to be causal, can be classified as a more subdivided type, such as a positive cause and negative cause, if the direction of causality is clearly revealed in the sentence. This hierarchical structure of relations has no biological meaning and is intended to introduce certainty about the interaction. It allows researchers to determine the reliability of the relation revealed in the text on their own because the relation type itself defines the level of its specificity and certainty.

## **1.2. Deep Learning-Based Semantic Relation Classification Model**

We evaluated the feasibility of the dataset that we built by constructing a semantic relation classification model and training it on the dataset. Furthermore, we sought to improve the performance of relation extraction by suggesting a new fine-tuning method. Since relation extraction has significant applications in various NLP tasks, improving the performance of relation extraction models can also improve the quality of various application tasks such as information extraction and knowledge graph construction.

Recently, deep learning-based pre-training language models such as BERT [19] have recorded state-of-art (SOTA) performance in many NLP tasks, including relation classification [20, 21]. An NLP model that exceeds the performance of conventional deep learning models, such as convolutional neural network

(CNN), recurrent neural network (RNN), and long short-term memory (LSTM) or rule-based models, is expected to accurately extract meaningful information from text data [22, 23].

In summary, we built a training dataset for semantic relation classification that annotates various bio-entity types and their semantic relational information. In addition, we proposed a new fine-tuning method to improve the performance of relation classification tasks. By comparing the relationship extraction performance of models with various methodologies trained on the constructed dataset including the proposed methodology, we evaluated the usability of our constructed dataset and the performance of our proposed methodology. The creation of these novel datasets and fine-tuning methodology for classifying relations provides a meaningful contribution to this field and is expected to advance future semantic relation classification research.

## 2. Material And Methods

### 2.1. Construction of Semantic Relation Corpus

#### *Data Preparation*

First, PubMed data that were published between 2004 and 2019, a total of 15 years, were collected regardless of topic within the boundaries of the biomedical field. A total of 1,500 candidate abstracts were obtained. After a brief inspection of the abstracts, we excluded documents that were too short or lacked sufficient entities for a relation to be assigned. The number of screened papers was 154, and a total of 1,346 filtered papers in our dataset were subject to annotations. The collected data were then separated into sentences, which became the unit of annotation in this work. Sentences were distinguished from one another based on the combination of the PMID that they belonged to and the sentence identity (ID) that represented the sequential order of the sentence in the text. The title of each work was treated as the sentence with the sentence ID of 0.

#### *Annotation*

We extracted different types of entities from the titles and abstracts of papers provided by PubMed and then built a semantic relationship corpus that assigned semantic relation types to those entities based on the contextual information surrounding them and an expert's judgment. Recognizing different types of entities outside the scope of proteins/genes and granting those entities semantic relationships were highly dependent on manual expert curation. Such studies have not been extensively performed in the past.

Various bio-entities in each sentence were manually identified and annotated with the corresponding entity type. When annotating them or resolving any ambiguity after the annotation process, annotators referred to three databases, namely, Pubtator, Uniprot, and GeneCards, that contain information about bio-entities. Entity types that were annotated, with their definitions given later in the paper, included biological

processes, cells, compounds, DNA, enzymes, genes, hormones, molecular functions, phenotypes, proteins, RNA, and viruses.

After the process of annotating entities was complete, the process of assigning the eight types of semantic relations between those annotated entities followed. The relation was decided based on the context of the sentence or the full abstract. The eight types of semantic relations included the following: undirected link, directed link, positive cause, positive increase, negative decrease, negative cause, positive decrease, and negative increase.

For the sake of convenience, the annotation work was assisted by an online annotation tool developed by our research team (Figure 1). The software tool is based on an open-source project TextAE (Text Annotation Editor)<sup>[1]</sup>, a visual annotation tool, that can support not only creating and saving annotation results but also conveniently retrieving and editing existing ones. This tool saved us significant time whenever we needed to revise the annotation guideline or amend previously accumulated annotations. The tool can be accessed at <http://165.132.151.153/annotator/>.

## **Annotation Guidelines**

### *Entity Annotations*

Recognition of named entities is a prerequisite in the relationship extraction process. Discovering the relationship between two entities is possible after they are accurately recognized. To ensure accurate and consistent annotation, multiple annotators worked independently at first, and in the event of a discrepancy, a consensus was reached after multiple agreements to resolve it. Brief explanations of each entity type are shown below (Table 1). Please see Appendix B for details.

### **<Table 1 > Twelve entity types**

Entity Type	Explanation
<b>Genes, DNA, RNA, and Proteins</b>	A gene is the functional unit of heredity and the nucleotide sequence of DNA or RNA that holds instructions for synthesizing either RNA or protein.
<b>Enzymes</b>	Proteins that act as biological catalysts.
<b>Hormones</b>	Signaling molecules that act distant from its site of production.
<b>Compounds</b>	Additives such as drugs or chemicals.
<b>Molecular functions</b>	Proteins with the role of binding such as hormones and antigen-antibodies.
<b>Phenotypes</b>	Limited to human diseases.
<b>Biological processes</b>	Processes/activities that occur within cells.
<b>Cells</b>	Smallest functional units of an organism with which biological experiments are conducted to observe their mechanisms or to grow targeted compounds.
<b>Viruses</b>	Used to indicate when experiments are conducted on a virus.

If more than one entity was recognized in a sentence, the sentence moved to the step of assigning the type of semantic relationship between the entities.

### *Semantic Relation Annotations*

Relationship annotation is performed on two entities that appear within a sentence. In principle, the assignment of relations is based on the sentence in which the entities appear, but when ambiguous, the entire abstract of a piece can be read, and the annotator can decide based on the full context in the final stage. To define semantic relationships between entities, we extracted relational verbs and contextual information between two entities. When determining semantic relationships between entities, contextual information must be considered along with the meaning of the verb. For example, if “miR-194,” “basal and insulin-stimulated glucose uptake,” and “glycogen synthesis” were recognized as named entities in the sentence “Knockdown of miR-194 in L6 skeletal muscle cells induced an increase in basal and insulin-stimulated glucose uptake and glycogen synthesis” [24], the verb associated with “miR-194,” “basal and insulin-stimulated glucose uptake,” and “glycogen synthesis” would be “induce.” Although the verb “induce” is usually classified as positive, the word “knockdown of” followed by “miR-194” means inhibition; therefore, it should be classified as negative rather than positive. Thus, contextual information, which plays an important role in correctly classifying relationships, should be annotated together.

The semantic relationships defined in this study were classified into a total of eight types: undirected link, directed link, positive cause, positive increase, negative decrease, negative cause, positive decrease, and negative increase (Fig. 2). They were structured into three layers, and each level represents the extent of

granularity. Initially, a relation between two entities in a sentence was classified into an undirected link and a directed link at the top level based on whether the causal relationship between the two entities was clearly revealed. When a relation was identified as a directed link where there was a causal relationship based on the sentence, if it could be decided whether the two entities were positively or negatively correlated, the relation proceeded down one level down and was matched with finer types (positive cause/negative cause) or stopped at the first level (directed link). Similarly, at the second level (positive cause/negative cause), if the sentence captured causality according to an increase or decrease in the amount of each entity or its strength with explicit expressions of quantity, the relation proceeded down to the lowest level (positive increase/negative decrease, positive decrease/negative increase) or stopped at the second level (positive cause/negative cause). With regard to the relation of an undirected link, which is a correlation without causation revealed, since this type of relation is rarely a subject of attention for researchers, it does not need to be broken down to a more granular level, and this type of relation was treated as a negative example. In summary, with respect to the relationships that researchers might be interested in, each relation was classified into the most detailed and specific type possible. Examples of sentences corresponding to each relation type and detailed descriptions are provided in Appendix C.

## **Annotation Procedure**

We finished collecting data in February 2019 and performed trial annotations until April 2019. During the trial period, entities and relations were annotated for a small amount of collected data referring to previous related works for exploratory purposes. At this time, entities were pre-annotated through an automated biomedical NER/RE system called PKDE4J [25], and annotators had to modify incorrect entity annotations or add missed annotations. Relations between these readily annotated entities were classified manually by four to six annotators. However, as the decision to include gene-related entities in our annotation scope in a more exhaustive way was made, the method of entity annotation also shifted from utilizing PKDE4J to manual annotation to cover entity types that PKDE4J is not aimed at extracting, such as enzymes and viruses. Throughout the manual annotation process, the lead annotator, a biology expert with rich hands-on experience in bio-corpus construction projects, refined the guidelines and detailed the workflows.

The full-scale annotation process based on the final guidelines and workflows began in May 2019. The annotators included the lead annotator and eight researchers working on text mining in the biomedical field. More specifically, two of the eight researchers were selected as senior annotators to coordinate the entire annotation process among the multiple annotators and played important roles in settling any ambiguities, such as mediating disagreements that failed to be resolved in the previous stage. The lead annotator controlled the final verification and resolved any remaining ambiguous cases.

As we decided to annotate both entities and relations manually from the beginning, we developed a web-based annotation tool to streamline the annotation process and introduced this tool to our task in earnest starting in June. The annotation process was completed at the end of June 2020. The entire process, including data preparation, took approximately one and a half years to complete.

The final annotation workflow consisted of three stages: *annotation*, *error review*, and *final verification* (Fig. 3).

### *Annotation*

Annotators who had conducted research in the field of text mining and had experience building corpora in the biomedical field read the abstracts and manually annotated bio-entities by referring to the Uniprot, GeneCards, and Pubtator databases. They selected sentences in which two or more entities appeared and annotated the verb between the two entities and other contextual information that could help resolve any ambiguities. Based on this information and the annotator's judgment after reading the sentence, the relationship between two entities was mapped into one of the eight semantic relation types we defined in this work.

### *Error Review*

Except for the lead and senior annotators, six annotators formed two teams of three to assess for simple errors within their own teams. If there was a disagreement within the teams, they attempted to reach an agreement in a reasonable direction through discussion and provision of evidence. The discussion usually resolved the disagreement. If the issue persisted, however, the teams convened to discuss the issue and moved forward with an agreement or passed the disagreement on to the senior annotators.

The annotation result, which had undergone the first error review process, was delivered to the remaining two senior annotators who did not belong to either of the teams, and they conducted a second review process. The delivered annotation modifications and unresolved discrepancies were reviewed once more by these senior annotators, and agreement was attempted. Any inconsistencies that were not settled were finally resolved in the next verification step.

### *Final Verification*

To increase the quality of the dataset, the lead annotator verified the annotation result one last time in consideration of the context of the original text based on his or her biological knowledge. The lead annotator then corrected the result if necessary and decided whether to include it in the final dataset. The lead annotator not only adjudicated disagreements that could not be settled in the previous steps but also reviewed the entire dataset, including data that were already agreed upon by the teams for final verification. In practice, only a few simple errors were found after the multi-step review process, and these were quickly corrected.

Annotations that remained ambiguous even in the final verification stage were excluded from the final corpus.

## **2.2. Semantic Relation Classification Model Training**

To verify the credibility and usefulness of the semantic relation corpus we established, we attempted to build a semantic relationship classification model that utilized the corpus. We compared several deep

learning-based pre-training models and suggested fine-tuning techniques for relation classification that produced optimal performance.

### *Pre-trained Language Model for Relation Classification*

Many studies have been conducted to improve the performance of models for relation classification by applying deep learning algorithms such as CNN, RNN, and LSTM [22, 26–30]. Recently, BERT has shown the best performance in NLP tasks. BERT is a pre-trained language model that leverages the structure of the transformer encoder [31]. Language models learn and utilize universal text embeddings rich in grammatical and semantic features from pre-training on vast amount of textual data text, and only a simple additional layer is needed for the aiming task. The basic BERT model was trained on the Book corpus (800M words) and Wikipedia (2.5B), achieving SOTA in most common NLP tasks [19]. Since then, its variant models, which were pre-trained on domain-specific data, such as BioBERT [32] and SciBERT [33], as well as advanced versions of BERT with fine-tuned methods or a structure of layers, such as ALBERT [34] and RoBERTa [35], have been announced.

Since the focus of this study was to implement use of the corpus we constructed, we chose BERT and its several variants as our models due to their ease of use and capability of making predictions immediately without much configuration and compared their performance.

### *Pre-training and Fine-tuning Stage of BERT*

In the pre-training stage of BERT, a masked language model (MLM) and next sentence prediction (NSP) were utilized to learn various characteristics of natural languages [19]. The MLM method is a method of randomly replacing 15% of tokens with [MASK] tokens in the input sentence, expecting the BERT model to predict the original word of the [MASK] token. The pre-training of BERT involves two different sentences divided by the [SEP] token as input. At this time, 50% of the sentence pairs are in order, with the next sentence being the actual sentence that follows the prior in the original text, and the rest are not, with the first sentence being followed by a random sentence. NSP involves models learning to determine whether these two statements are in order. To develop a BERT model trained with these methods in the pre-training stage to perform downstream NLP tasks, such as relationship extraction, additional layers for detailed tasks are built after the transformer encoder layer having learned weights, and further fine-tuning is performed using relevant data for the desired tasks. This is how fine-tuning can provide a model for handling detailed downstream tasks.

The simplest approach to fine-tuning the relation classification task is using a single sentence containing the relationship between entities as input without any pre-processing treatment. However, to achieve optimal performance, introducing a slightly more complex input data processing method is necessary. One study that compared several fine-tuning methodologies for relationship extraction [36] found that the entity marker–entity start structure performed the best among all structures. The structure has the advantage of being able to learn the span of entities in a sentence by introducing marker tokens ([E1],

[E2]). Additional layers for relation classification are also added after the output layer and correspond to the location of marker tokens, enhancing the model performance (Fig. 4).

### *Masked Input*

Yang et al. [37] demonstrated that there is a mismatch in the BERT model. The [MASK] tokens are used in the pre-training process but not in the fine-tuning stage. To compensate for this limitation, in this work, we use a methodology that utilizes [MASK] tokens for input data in both the pre-training and fine-tuning stages for relation classification.

In the pre-training of BERT, input sentences containing [MASK] tokens are received as input data, and the original tokens for the [MASK] tokens within each sentence are predicted using the output of the final layer corresponding to the location of [MASK] tokens. Namely, the output of the final layer at [MASK] token contains contextual information needed to predict the original token replaced with [MASK]. Likewise, during the fine-tuning process, if the token corresponding to the entities of interest within the input sentence is replaced with a [MASK] token, the final output layer at the location of the replaced tokens can be considered to output a semantic and contextual vector for the token.

There are several objectives to using this approach. The first is to maintain consistency between the pre-training and fine-tuning training of the model. As mentioned above, the pre-training process for BERT employs [MASK] tokens that are not introduced in the fine-tuning stage, resulting in the disadvantage of inconsistent models. In this work, the [MASK] tokens are also utilized as input data in the fine-tuning process to increase the consistency of the model. The second objective is to effectively convey to the model the information of entities that span over multiple tokens. In relation classification, each entity often consists of multiple words. For these cases, one of the suggested methodologies is “entity marker–entity start” from [36], which employs additional marker tokens, such as [E1], [/E1] and [E2], [/E2], before and after the entity to convey information about where the entity is located in the sentence, has the disadvantage of inconsistency because the tokens are not used for pre-training and the entity itself is not replaced (Fig. 4). Although this method enables the model to learn the span of entities in a given sentence, the model is limited in accurately recognizing additional marker tokens that were unseen in pre-training. Furthermore, it fails to directly convey information of entities as a whole to the model. On the other hand, when replacing an entity itself with a [MASK] token, the entity exists as a token, and the output vector corresponding to the token contains contextual information that helps to effectively predict the meaning of the original word for the token (i.e., an entity) (Fig. 5).

Finally, it can alleviate the out-of-vocabulary (OOV) problem that occurs when a token that the model did not learn during training is introduced as input data. This problem can be improved using word piece tokenizing with BERT [19] but is not fully resolved. Tasks including named entity recognition and relation classification are likely to cause OOV problems because entities often contain proper nouns that have many variations of case or abbreviation (e.g., BERT, Bert, bert, bert algorithm). If an OOV problem occurs and the model cannot recognize entity tokens in a sentence correctly, it may struggle to predict the relation type between the entities. Replacing the entity with the [MASK] token more effectively prevents

OOV problems. However, the masked input method has a fatal drawback in that it loses the original token information of the entity. Therefore, we propose the two masked sentence input method to overcome this weakness.

### *Two-Sentence Input*

Two-sentence input is a method of utilizing two identical sentences that are linked with [SEP] tokens as input data in fine-tuning. In this paper, specifically, we propose the two masked sentence input method, which masks one of the two entities in each of the sentences (Fig. 6). The masked entities are different from each other to keep one of the original entities unmasked, which is more beneficial than simply linking the duplicate of the sentence.

The first advantage of this method is that this method maintains consistency in the pre-training and fine-tuning stages. In the original paper, the pre-training stage of BERT exploited two sentences linked with the [SEP] token. During fine-tuning, however, only one sentence was used, which is disadvantageous due to the model learning process is inconsistent. The second advantage is that this method prevents information loss associated with the masked input methodology. Within sentences that have multiple entities, we can preserve token information by replacing only one entity at a time with the [MASK] token. This method completely prevents the loss of token information that can be caused by using masked input. Finally, this method conveys sequential information about the entity to the model. In relation extraction tasks, the relationship might be decided differently if the order of the first entity and the second entity are reversed. Therefore, it is critical that the model accurately recognizes the order of entities. Traditional methods attempt to use additional marker tokens to carry sequence information to the model, but there is a limit to model recognition of additional tokens unseen during pre-training as sequence information. In the two masked sentence input method, the first entity is replaced with the [MASK] token in the first sentence, and the second entity is replaced with the [MASK] token in the second sentence, effectively passing the semantic and contextual information corresponding to the first and second entities to the model in order.

For comparison, we replaced two entities in a sentence differently by experimenting with the two-sentence input method combined with additional entity tokens, such as [E1] and [E2], instead of [MASK] tokens (Fig. 7). In this method—named the two-sentence entity token input method—the two tokens for each entity are differentiated from each other.

### *Downstream Layer Structure*

We built a SciBERT-based model for classifying semantic relations by adding a linear layer to the output layer corresponding to the location of the entities, masked by the [MASK] token, in the first and second sentences of the input data with the two masked sentence input methodology applied (Fig. 8). Two output vectors of BERT produce vectors containing semantic and contextual information for each entity. With these vectors as input, the added layer optimizes the weights for classifying relations between entities. This method has the advantage of more effective learning for relation prediction by directly

utilizing vectors corresponding to information of entities as input than CLS downstream layers, which are from the original BERT paper [19] and use the special token [CLS] for classification.

For comparison, we also experimented with the conventional CLS token layer method and a three-token layer method that utilizes both the outputs corresponding to [MASK] tokens for entities and the output from the [CLS] token, which is used as the starting input token for the downstream layer.

In the CLS downstream layer, which receives one BERT layer output as input, the input size of the classification output layer is set to  $h$  (512), which is the output size of each layer of the pre-trained language model. In the two-mask token layer proposed here, which receives two BERT layer outputs as input, the input size is set to  $2 \cdot h$  (1024), which is the size of two outputs combined. In the three-token layer, which receives three BERT layer outputs as input, the input size is set to  $3 \cdot h$  (1536), which is the size of three outputs combined. The output size for all the methods is set to  $c$  (10), which is the number of relation types or the number of classes to be predicted. The loss function of the model is cross-entropy loss.

<sup>[1]</sup> Text annotation editor. TextAE, 2020. Accessed 2021 Aug 30. Available from: <https://textae.pubannotation.org/>

## 3. Results

### 3.1. Dataset Overview

We constructed a semantic relation corpus consisting of 1,346 abstracts annotated with 5,045 relations classified into eight types. A total of 5,014 distinct bio-entities of 12 types were identified including 2,346 in the left portion of the sentence and 2,668 in the right portion of the sentence. The general statistics for the corpus are shown in the figure below (Fig. 9).

The verb in the sentence is one point of reference when classifying the relation type; however, it does not exclusively determine the relation type. A relation type is determined by comprehensively considering the verb and verb-related information in the sentence, contextual information around entities, and human interpretation. For example, common context words such as “inhibition” and “decreased” can reverse the meaning of a verb. Thus, the semantic relation type assigned to the sentence might be the opposite of the original meaning of the verb.

### 3.2. Semantic Relation Classification Model

*Performance Comparison between Pre-trained Language Models (Table 2)*

While applying an effective masking methodology for input sentences and downstream layers to the model is crucial, it is also important to select a pre-trained language model that best fits our tasks and data as the base model. Therefore, we compared the performance of various existing pre-trained BERT

models for our dataset with the same hyperparameters specified in each original paper unless otherwise noted, setting the input masking method to two masked sentence input for all models.

**<Table 2 > Performance comparison of pre-trained language models**

<b>Model</b>	<b>F1</b>
BERT [19]	78.1
ALBERT [34]	80.4
RoBERTa [35]	81.3
SpanBERT [20]	80.5
XLNET [37]	80.2
<b>SciBERT [33]</b>	<b>81.7</b>

Comparisons have shown that SciBERT models pre-trained on scientific publication datasets using the BERT model performed better than ALBERT, RoBERT, SpanBERT, and the autoregressive transformer (XLNET), which improve the model layers and pre-training methodologies of BERT. This confirms that when building a downstream model using a pre-trained language model, the data used for pre-training should have the same domain as those used for fine-tuning. SciBERT, which performed the best on our PubMed datasets, will be used as the base model in later experiments.

*Performance Comparison between Methods of Masking Input (Table 3)*

For this experiment, we used SciBERT, the language model with the best performance in the abovementioned comparative experiment, as the base model and the CLS token layer as the downstream output layer. We trained the model with 10 epochs and used a batch size of 8 during training, a learning rate of  $5e-5$  with 1,100 warm-up steps, and a weight decay of 0.01 to prevent overfitting. We evaluated each method of masking input sentences to determine which one performed the best.

We specifically compared the performance of the following methods: masked input, two masked sentence input, entity token input, masking with additional tokens other than [MASK], and entity marker–entity start, marking a span of entities.

**<Table 3 > Performance comparison of masking input methods**

Method	F1
Entity Marker–Entity Start	79.9
Masked Input	79.6
<b>Two Masked Sentence Input</b>	<b>81.4</b>
Two-Sentence Entity Token Input	80.5

The comparison of input methods showed that the two-sentence entity token input and the two masked sentence input methods, which used two combined sentences as input data, performed better than the entity marker–entity start method or the original masked input method, which is the way BERT is pre-trained. Two masked sentence input, which replaces entities with [MASK] tokens, was used rather than using additional tokens, [E1] and [E2] tokens, to replace entities and also showed better performance.

These findings show that leveraging [MASK] tokens is better than introducing additional tokens such as [E1] and [E2] to replace entities. They also confirm that maintaining consistency between the pre-training and fine-tuning stages can lead to improved performance. Furthermore, [MASK] tokens, which have only been used in pre-training phases, can be appropriately utilized in downstream tasks. In addition, the two masked sentence input methodology performed better than the methodologies where only one sentence is entered as input; this finding suggests that in such a relation classification task, using two identical sentences, each of which contains a masked entity with [MASK], can lead to improved performance.

*Performance Comparison of Downstream Layers (Table 4)*

Additional downstream layer construction is essential to train specific NLP tasks using pre-trained language models. Relation extraction models require the addition of the classification output layer for relation prediction after the output of the transformer encoder. For performance comparisons between different layer structures, we equally apply the two masked sentence input methodology to the same SciBERT model with the same hyperparameters as the previous experiment. In this experiment, we compare the performance of the CLS token layer, using the output of the [CLS] token location, two-mask token layer, using the output values of the two [MASK] token locations, and three-token layer, using the output values of the [CLS] and [MASK] token locations.

**<Table 4 > Performance comparison of downstream layers**

Layer	F1
CLS token layer	81.4
<b>Two-mask token layer</b>	<b>81.7</b>
Three-token layer	81.5

The experiments showed no significant difference, but the two-mask token layer showed slightly better performance than the other two techniques. The trivial difference between downstream layers seems to be due to the structure of the transformer encoder, where multiple layers are stacked and interconnected in both directions. The structural characteristics of BERT are exemplified by layers corresponding to all tokens in the input sequence, and these layers are interconnected and pass through multiple transformer encoders. This distinctive structure enables a well-trained model for predicting the relationship between entities regardless of the number or location of the final output vectors in the downstream layer.

*Final Comparison of Model Performance (Table 5)*

The experiments comparing different base models, input methodologies, and output layer structures using our datasets showed that utilizing SciBERT as a base model with the two masked sentence input methodology and two-mask token layer applied performs best, resulting in an F1 score of 81.7. Finally, we compared this model with existing models presented in related works using our dataset. In addition to BERT-based models that have shown SOTA performance in relation extraction tasks, such as [36], we also included models based on other deep learning algorithms such as the CNN [38] and entity attention Bi-LSTM, which is a semantic relation classification model using bidirectional LSTM networks with entity-aware attention using latent entity typing [22].

**<Table 5 > Overall performance comparison of the models**

<b>Model</b>	<b>F1</b>
Word2vec + CNN [38]	70.8
Entity Attention Bi-LSTM [22]	78.7
Matching the Blanks [36]	79.9
<b>Our Model (Two Masked Sentences)</b>	<b>81.7</b>

The experimental results confirm that the final proposed SciBERT-based model with the two masked sentence input methodology and two-mask token layer performed best.

To determine how well our model predicts on each class and examine situations where our model has limitations, we further analyzed the per-class performance for each of the eight types of relations (Table 6).

**<Table 6 > Per-class performance**

Class	Precision	Recall	F1-score	Support
Directed Link	0.928	0.850	0.888	167
Negative Cause	0.845	0.893	0.868	140
Negative Decrease	0.817	0.918	0.865	73
Negative Increase	0.788	0.820	0.804	50
Positive Cause	0.882	0.922	0.901	218
Positive Decrease	0.720	0.667	0.692	27
Positive Increase	0.738	0.818	0.776	55
Undirected Link	0.900	0.839	0.868	279

Support: Number of instances in the test data represents 20% of the full data set, which is proportional to each class.

In general, per-class performance relied on the number of data instances under the class. Classes designated as negative increase, positive decrease, and positive increase, which had the fewest data instances of 50, 27 and 55, respectively, obtained the lowest f1 scores among the different relation types. Other than the issue, the model showed generally even scores over the classes.

We closely examined the data points where the value predicted by the model differed from the annotated target value to objectively assess the limitations of our model or corpus and obtain insights for future reinforcement. As a result of the observation, we were able to identify two interesting patterns in false cases.

First, our model revealed its weakness when the verb between entities did not directly convey the meaning of increase/decrease or a cause-and-effect relationship, such as “improve,” “exacerbate,” and “aggravate,” making it difficult to accurately infer the relationship through context words surrounding entities. In this case, to correctly determine the direction in the quantity of the right entity, knowing whether the entity instance itself held a positive or negative meaning was necessary, such as in the sentence below:

*Moreover, hepatic knockdown of HFREP1 improved insulin resistance in both mice fed a high-fat diet and ob/ob mice.*

The target relation type associating “HFREP1” and “insulin resistance” belongs to the negative decrease class, but the model incorrectly predicted it as the negative increase class. To accurately classify their relationship, in this case, the model needs to know whether insulin resistance itself has a positive or negative meaning. This type of error could be alleviated with a language model pre-trained on richer literature in the biomedical field, resulting in more comprehensive coverage of semantic meaning for bio-vocabulary.

Second, we found several cases of errors due to the conflict between the annotators' contextual considerations of the entirety of the literature findings and the model predictions that exploit contextual words limited to each sentence in classifying the relationship between entities. An example of this is as follows:

*Our data suggest that titanium particles may cause less leukocyte activation and inflammatory tissue responses than other particulate biomaterials used in total joint arthroplasty.*

For this sentence, the annotator classified the relationship between titanium particles and inflammatory tissue response as the negative cause class, and the model predicted the positive cause class. The annotator compared the relationship between these two entities to other entities in the sentence and focused on the intention of the sentence. However, if we simply considered the directional association between the two entities of interest, we could assign the positive cause class, which was the model prediction.

To avoid this controversial gray area, the data that required abstract and complex consideration of context were excluded as much as possible from the corpus construction stage; as a result of this, few of these cases were found. However, we specifically paid attention to this example because it provided insights on how the model prediction works in these special circumstances and which direction to move forward in future research to overcome this limitation. In the example sentence, the model prediction cannot be regarded as wrong, but the main finding conveyed in the sentence must have been that titanium particles cause "less" inflammatory reactions, not the fact that they do. Therefore, this case demonstrates that meaningful relation types, which better reflect the intentions of the text and provide benefit to researchers, require elaborately reflecting not only the causality between entities and its direction but also the relative extent of the increased or decreased amount of a particular entity. This is possible by pushing beyond the limits of the current relation classification based on binary entities and addressing subtle and complicated interactions among multiple bio-entities appearing in a sentence.

## 4. Conclusion

Machine learning-based relational classification tasks can be successfully performed based on good quality training data and well-designed algorithms. However, if the training data are unreliable due to poor quality or small quantities, the performance of machine learning algorithms cannot be accurately identified and compared. This poses significant obstacles to studying how to improve the algorithms. In particular, while algorithms are domain-independent, corpora are not; thus, constructing training datasets annotated with bio-entities and the relationships between them is an urgent task to promote text mining research in biomedical fields [39].

In this paper, we developed a corpus with a wide range of bio-entities, such as biological processes, cells, compounds, DNA, enzymes, genes, hormones, molecular functions, phenotypes, proteins, RNA, and viruses, along with their annotated semantic relations. The construction of a corpus with multiple types of bio-entities and their rich relationships is essential to extract complex and significant biological

information from a wide range of bio-entity types that the biomedical literature contains. Considering this need, our newly constructed corpus, built by manually tagging a wide range of bio-entities and their relationships from a rich amount of biomedical literature, represents a significant contribution. We comprehensively annotated verbs situated between entities, contextual information, such as positive/negative and active/passive information that affects their meanings, and other meaningful information in the sentence as features to consider in assigning semantic relation types, opening the possibility of further research on semantic relationships. This corpus could be used as a reliable reference standard in the development of text mining systems.

Another contribution of this paper is that we demonstrated the utilization of the dataset that we built by training and evaluating deep learning-based semantic relation classification models leveraging the data and further presented a way to improve the performance of the relationship classification task. Tweaking existing BERT-based models that are already known to show good performance for the classification task, we devised a new technique that can achieve better performance by alleviating the limitations of existing models for classification. By introducing a [MASK] token respectively on two identical input sentences, we effectively improved problems such as OOV words and inconsistency between pre-training and fine-tuning that afflict existing relation extraction models. In the overall comparison experiment between our model with all of its methods and the existing models suggested in the related works, our proposed model proposed showed the best performance, confirming that this methodology is effective in improving the performance of fine-tuning the language model for relation classification tasks.

In summary, the developed training dataset for semantic relation classification was successfully applied to train the classification model. Therefore, this could be used as a valuable resource for similar text mining research. We also made significant improvements to the algorithms of the relation classification model. We expect that the biological information extracted with high accuracy through our proposed training dataset and relation extraction technique will be used as a trusted source of information in the development of a biomedical text mining system. We also believe that the annotation processes we elaborated here, the explanations of possible problems in the current deep learning algorithms or in the training dataset, and the ideas that can improve them will be of significant help to fellow researchers performing similar work.

However, there remains room for improvement. As already mentioned in the per-class performance analysis, our semantic relation based on the causality of binary entities and its direction showed limitations at sufficiently describing complex semantic associations among bio-entities in a sentence. For example, in cases where the relative intensity of the association needs to be revealed for meaningful knowledge discovery, the current relation type might be insufficient unless introducing a complex semantic relation of which degree is higher than two. Moreover, the comparative experimental results between downstream layers indicating minor differences between the three different layer structures warrants more in-depth future research. This suggests that a novel effective downstream layer providing significantly improved performance compared with the existing layer structure needs to be proposed. If follow-up studies are conducted to address these listed limitations based on the realizations obtained

through the experiments and analysis in this paper, we can expect further improvement in constructing a dataset and deep learning model for effective semantic relation classification to be achieved in the near future.

## Declarations

### *Ethics approval and consent to participate*

Not applicable.

### *Consent for publication*

Not applicable.

### *Availability of data and materials*

The datasets generated during the current study are available at <https://github.com/tsmmbio/BertSRC>.

### *Competing interests*

The authors declare that they have no competing interests.

### *Funding*

This work was supported by the Bio-Synergy Research Project (NRF-2013M3A9C4078138) of the Ministry of Science, ICT, and Future Planning through the National Research Foundation.

### *Authors' contributions*

YL wrote the main manuscript text. JS conducted model experiments and prepared Figure 8. MS confirmed the contents of the manuscript. All authors reviewed the manuscript.

### *Acknowledgements*

Not applicable.

## References

1. S.C. Onye, A. Akkeles, N. Dimililer, Review of biomedical relation extraction, European International Journal of Science and Technology. 6 (2017) 1–14.
2. Zhou D, Zhong D, He Y. Biomedical relation extraction: from binary to complex. Comput Math Methods Med. 2014;2014:298473. doi: 10.1155/2014/298473. Epub 2014 Aug 19. PMID: 25214883; PMCID: PMC4156999.

3. W.W. Chapman, K.B. Cohen, Guest editorial: Current issues in biomedical text mining and natural language processing. *J. of Biomedical Informatics*. 42(5) (2009) 757–759.  
<https://doi.org/10.1016/j.jbi.2009.09.001>.
4. H. Kilicoglu, G. Roseblat, M. Fiszman, D. Shin, Broad-coverage biomedical relation extraction with SemRep. *BMC Bioinformatics*. 21(1) (2020) 1–28. <https://doi.org/10.1186/s12859-020-3517-7>.
5. Y. Luo, Ö. Uzuner, P. Szolovits, Bridging semantics and syntax with graph algorithms—state-of-the-art of extracting biomedical relations. *Briefings in Bioinformatics*. 18(1) (2017) 160–178.
6. H. Gurulingappa, A.M. Rajput, A. Roberts, J. Fluck, M. Hofmann-Apitius, L. Toldo, Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. *Journal of Biomedical Informatics*. 45(5) (2012) 885–892.  
<https://doi.org/10.1016/j.jbi.2012.04.008>.
7. E.M. van Mulligen, A. Fourrier-Reglat, D. Gurwitz, M. Molokhia, A. Nieto, G. Trifiro, J.A. Kors, L.I. Furlong, The EU-ADR corpus: Annotated drugs, diseases, targets, and their relationships. *Journal of Biomedical Informatics*. 45(5) (2012) 879–884. <https://doi.org/10.1016/j.jbi.2012.04.004>.
8. M.-S. Huang, P.-T. Lai, R.T.-H. Tsai, W.-L. Hsu, Revised JNLPBA corpus: A revised version of biomedical NER corpus for relation extraction task. *ArXiv Preprint ArXiv:1901.10219*, 2019.
9. Pyysalo, S., Airola, A., Heimonen, J. et al. Comparative analysis of five protein-protein interaction corpora. *BMC Bioinformatics* 9, S6 (2008). <https://doi.org/10.1186/1471-2105-9-S3-S6>
10. GENIA corpus, <http://www.geniaproject.org/genia-corpus/relation-corpus>, 2022 (Accessed February 2022).
11. BioNLP Shared Task, <https://sites.google.com/site/bionlpst/bionlp-shared-task-2011/entity-relations-supporting-task-rel>, 2022 (Accessed February 2022).
12. Nédellec C, Learning language in logic - genic interaction extraction challenge. *Proceedings of LLL'05* 2005:31–37.
13. LLL corpus, <http://genome.jouy.inra.fr/texte/LLLchallenge/>, 2022 (Accessed February 2022).
14. BioCreative-ii corpus, <https://biocreative.bioinformatics.udel.edu/resources/corpora/biocreative-ii-corpus/>, 2022 (Accessed February 2022).
15. Bunescu R, Ge R, Kate RJ, Marcotte EM, Mooney RJ, Ramani AK, Wong YW: Comparative experiments on learning information extractors for proteins and their interactions. *Artif Intell Med, Summarization and Information Extraction from Medical Documents* 2005, 33(2):139–155.
16. Pyysalo S, Ginter F, Heimonen J, Björne J, Boberg J, Järvinen J, Salakoski T: BioInfer: A corpus for information extraction in the biomedical domain. *BMC Bioinformatics* 2007, 8(50)
17. Fundel K, Kuffner R, Zimmer R: RelEx—Relation extraction using dependency parse trees. *Bioinformatics* 2007, 23(3):365–371.
18. Ding J, Berleant D, Nettleton D, Wurtele E: Mining MEDLINE: Abstracts, sentences, or phrases? *Proceedings of PSB'02* 2002:326–337

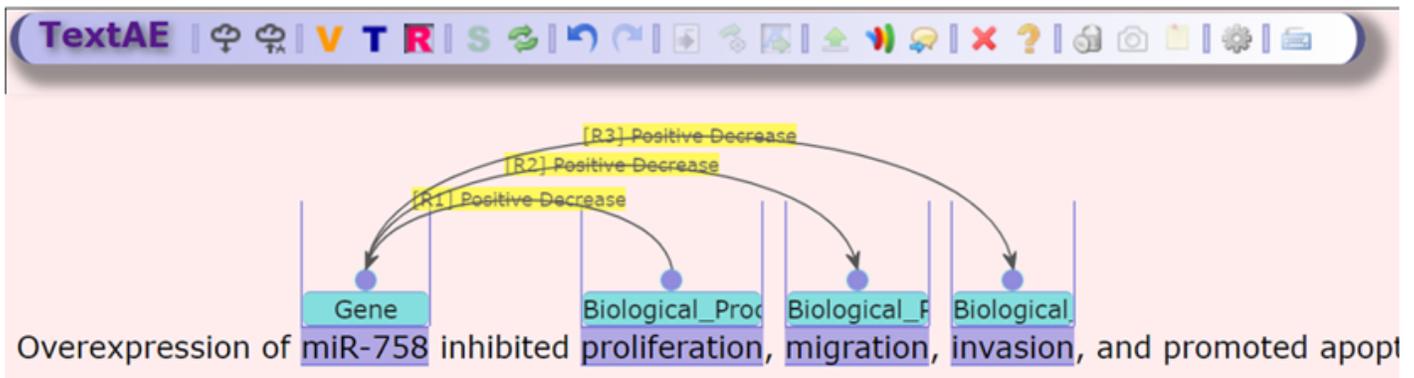
19. J. Devlin, M.W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL HLT 2019–2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 1(Mlm), 2019, 4171–4186.
20. M. Joshi, D. Chen, Y. Liu, D.S. Weld, L. Zettlemoyer, O. Levy, SpanBERT: Improving pre-training by representing and predicting spans. CoRR, abs/1907.1, 2019. <http://arxiv.org/abs/1907.10529>.
21. P. Shi, J. Lin, Simple {BERT} models for relation extraction and semantic role labeling. CoRR, abs/1904.0, 2019. <http://arxiv.org/abs/1904.05255>.
22. J. Lee, S. Seo, Y.S. Choi, Semantic relation classification via bidirectional LSTM networks with entity-aware attention using latent entity typing. Symmetry. 11(6) (2019) 785.
23. Y. Shen, X. Huang, Attention-based convolutional neural network for semantic relation extraction. COLING, 2016.
24. C. Latouche, A. Natoli, M. Reddy-Luthmoodoo, S.E. Heywood, J.A. Armitage, B.A. Kingwell, MicroRNA-194 modulates glucose metabolism and its skeletal muscle expression is reduced in diabetes. PLoS One. 11(5) (2016) e0155108–e0155108. <https://doi.org/10.1371/journal.pone.0155108>.
25. M. Song, W.C. Kim, D. Lee, G.E. Heo, K.Y. Kang, PKDE4J: Entity and relation extraction for public knowledge discovery. Journal of Biomedical Informatics. 57 (2015) 320–332. <https://doi.org/10.1016/j.jbi.2015.08.008>.
26. Z.Q. Geng, G.F. Chen, Y.M. Han, G. Lu, F. Li, Semantic relation extraction using sequential and tree-structured LSTM with attention. Information Sciences. 509 (2020) 183–192. <https://doi.org/10.1016/j.ins.2019.09.006>.
27. M. Xiao, C. Liu, Semantic relation classification via hierarchical recurrent neural network with attention. COLING 2016–26th International Conference on Computational Linguistics, Proceedings of COLING 2016: Technical Papers, 1254–1263.
28. K. Xu, Y. Feng, S. Huang, D. Zhao, Semantic relation classification via convolutional neural networks with simple negative sampling. Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing, 2015, 536–540. <https://doi.org/10.18653/v1/d15-1062>.
29. X.H. Yatian Shen, Attention-based convolutional neural network for semantic relation extraction. Anaesthesia Critical Care and Pain Medicine. 36(6) (2017) 411–418. <https://doi.org/10.1016/j.accpm.2017.08.001>.
30. D. Zeng, K. Liu, S. Lai, G. Zhou, J. Zhao, Relation classification via convolutional deep neural network. COLING, 2014.
31. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need. Advances in Neural Information Processing Systems, 2017, 5999–6009.
32. J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C.H. So, J. Kang, BioBERT: A pre-trained biomedical language representation model for biomedical text mining. Bioinformatics. 36(4) (2020) 1234–1240. <https://doi.org/10.1093/bioinformatics/btz682>.

33. I. Beltagy, K. Lo, A. Cohan, Scibert: A pretrained language model for scientific text. ArXiv, 2019, 3615–3620.
34. Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, R. Soricut, Albert: A lite BERT for self-supervised learning of language representations. ArXiv, 2019, 1–17.
35. Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, RoBERTa: A robustly optimized BERT pretraining approach. ArXiv, 1, 2019.
36. L.B. Soares, N. FitzGerald, J. Ling, T. Kwiatkowski, Matching the blanks: Distributional similarity for relation learning. ACL 2019–57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, 2019, 2895–2905. <https://doi.org/10.18653/v1/p19-1279>.
37. Z. Yang, Z. Dai, Y. Yang, J.G. Carbonell, R. Salakhutdinov, Q.V. Le, XLNet: Generalized autoregressive pretraining for language understanding. CoRR, abs/1906.0, 2019. <http://arxiv.org/abs/1906.08237>.
38. T.H. Nguyen, R. Grishman, Relation extraction: Perspective from convolutional neural networks. Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing, 2015, 39–48. <https://doi.org/10.3115/v1/W15-1506>.
39. J.D. Kim, T. Ohta, Y. Tateisi, J. Tsujii, GENIA corpus - A semantically annotated corpus for bio-textmining. *Bioinformatics*. 19 (SUPPL. 1), 2003, 180–182. <https://doi.org/10.1093/bioinformatics/btg1023>.
40. MedlinePlus, What is a gene? <https://medlineplus.gov/genetics/understanding/basics/gene/>, 2021
41. E.D. Hirsch, *The New Dictionary of Cultural Literacy: What Every American Needs to Know*, Houghton Mifflin, Boston, 2002.
42. National Institute of General Medical Sciences, What is genetics? <https://www.nigms.nih.gov/education/fact-sheets/Pages/genetics.aspx>, 2021.
43. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Molecular Biology of the Cell*, Fourth Ed., Garland Science, New York, 2002.
44. M. Shuster, *Biology For a Changing World, With Physiology*, Second Ed., New York, 2014.
45. N. Neave, *Hormones and Behaviour: A Psychological Approach*, Cambridge Univ. Press, Cambridge, 2008.
46. MedlinePlus, Hormones. <https://medlineplus.gov/hormones.html>, 2021 (Accessed 29 August 2021).
47. Encyclopedia Britannica, Hormones. <https://www.britannica.com/summary/hormone>, 2021 (Accessed 21 May 2021).

## Figures

PMID	Sent_ID	Left Context	Left Entity	Left Entity Type	Right Context	Left Context	Right Entity	Right Entity Type	Right Context
30446524	13	over-expression	miR-758	GENE			proliferation	BIOLOGICAL_PROCESS	
Infinitive Verb	Relation Verb	Sentence				Relation Type_Y	Relation Type_Y		
inhibit	inhibits	Our results support that over-expression of miR-758 inhibits proliferation, migration and invasion, and promotes apoptosis of NSCLC cells by negative regulating HMGB2.				Directed Link	Positive Decrease		

<Figure 1-1> Sentence within the corpus that was subject to annotation



<Figure 1-2> Sentence annotation result visualized with TextAE

Figure 1

**Annotation tool based on TextAE:** TextAE is a text annotation tool that can annotate named entity and relation information in the text. Each term (entity) can be dragged or double-clicked in *Term Edit Mode*, and the corresponding type can be selected to annotate them. If the corresponding entity type does not exist at the time, a new type can be defined. Likewise, in *Relation Edit Mode*, a type of relation can be selected or created and visualized. It is also possible to annotate multiple relations where one entity is associated with multiple other entities at once. The results of the annotation can be downloaded locally in the form of a .json file.

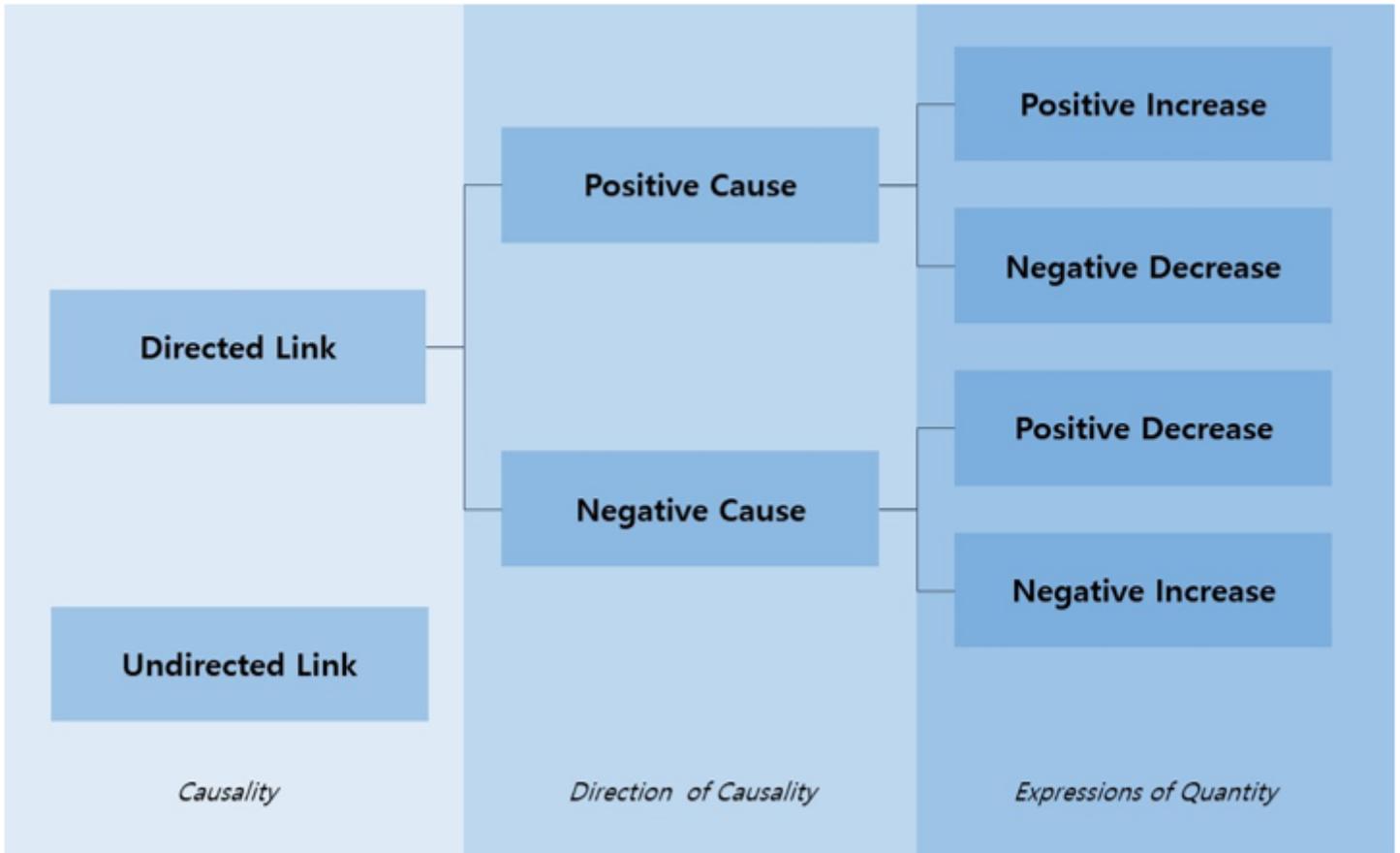


Figure 2

Eight types of relations



## Figure 5

**Masked input:** Input sentence masked as [MASK] for entities

The overexpression of miR-758 inhibits proliferation.

Entity 1

Entity 2

The overexpression of [MASK] inhibits proliferation. [SEP] The overexpression of miR-758 inhibits [MASK].

## Figure 6

**Two masked sentence input**

The overexpression of miR-758 inhibits proliferation.

Entity 1

Entity 2

The overexpression of [E1] inhibits proliferation. [SEP] The overexpression of miR-758 inhibits [E2].

## Figure 7

**Two-sentence entity token input**

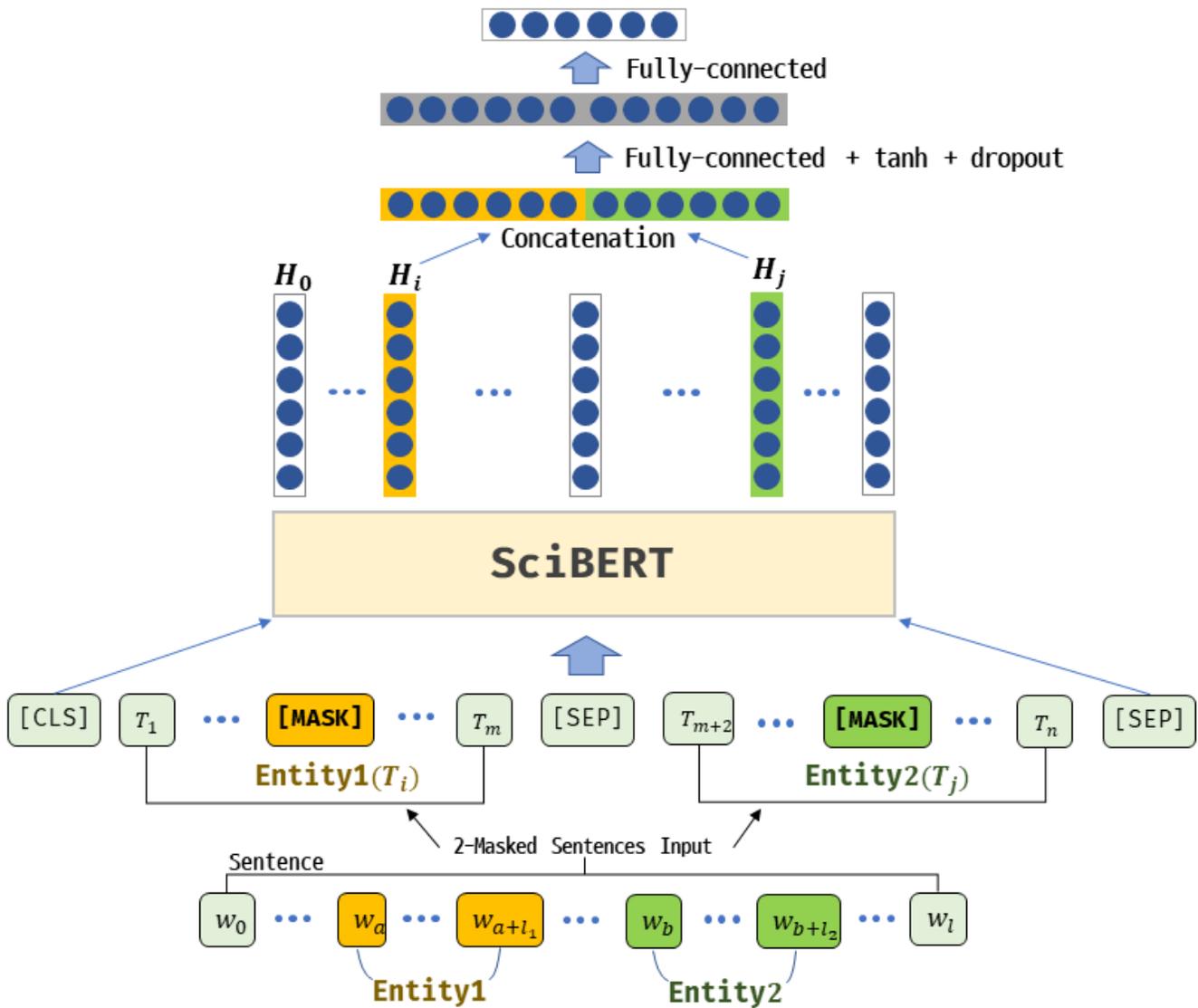


Figure 8

SciBERT-based model with the two masked sentence input method applied

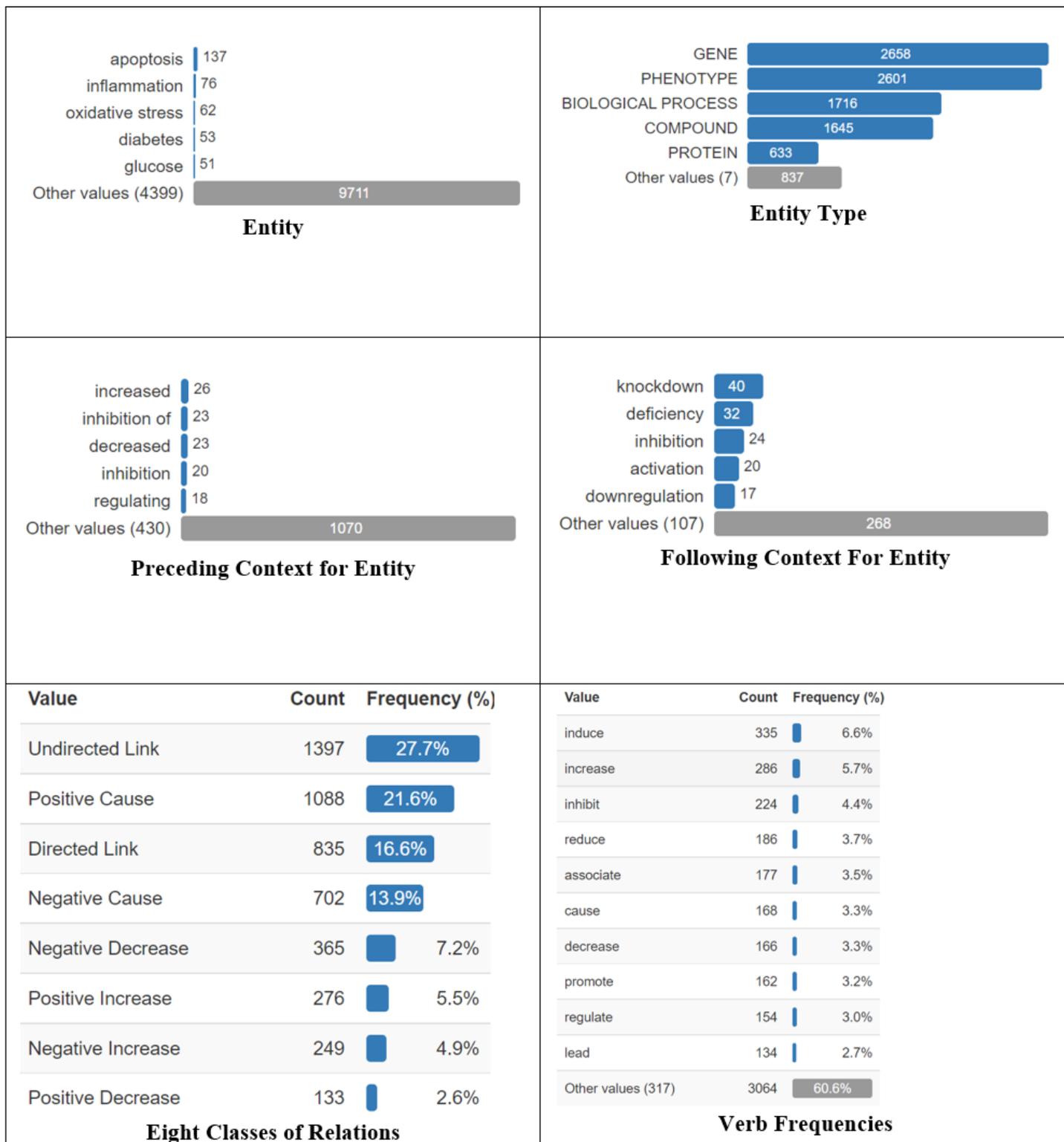


Figure 9

Dataset statistics

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- BertSRCAppendix.docx