A Review and Outlook of the Latest Results on Document-level Information Extraction

Huanyuan Li[1,3], Lai Wei[2,4,*], Ziheng Wang[1,5]

[1]School of Computer Information and Engineering, School of Henan University of Economics and Law, Zhengzhou, China
[2]College of Computer Science and Technology, Hainan University, Haikou, China

[3]L18839191123@icloud.com
[4]maoxiali@ldy.edu.rs
[5]2830631926@qq.com
*corresponding author

Abstracts. Relationship extraction is a key step in information extraction and aims to identify entities from text data and identify semantic links between them, which is the foundation of knowledge graph construction. Improving relationship extraction efficiency can effectively enhance the quality of information extraction, which has an impact on the construction of a knowledge graph and subsequent downstream tasks. The paper improves the efficiency and quality of relationship extraction by optimizing the model, choosing the way of relationship extraction and proposing new evaluation parameter criteria. The paper summarizes the current state-of-the-art research results, and briefly describes SIEF, an evaluation mechanism that focuses on important sentences, Coreference Aid, which captures the internal structure of entities as well as the external fine-grained information, and FILR, a framework that extracts global document information and performs multi-granular logical reasoning. Aggregate semi-automated data augmentation by integrating large model languages, remote supervision and construction of domain graphs integrating domain information of entities, cross-article entity-centered relationship extraction capable of capturing connections between different articles, and the introduction of mean-accuracy averaging capable of evaluating the comprehensibility of the model are presented. This paper presents a vision of new and better optimization algorithms by describing the latest approaches to relationship extraction optimization and finding that they all have their own limitations.

Keywords: Document-Level Relationship Extraction, Framework Optimization, Natural Language Processing, Model Evaluation, Cross-document Relation Extraction.

1. Introduction

DocRE is a significant research topic in NLP, where the main task is to automatically identify the relationship between two individuals in a document containing numerous sentences or even multiple paragraphs. This is different from the traditional sentence-level relationship extraction, which usually only needs to extract the relationship between entities within a single sentence. In contrast, at the document level, relationships between entities often span multiple sentences and even require

understanding the context of an entire paragraph or document. This complexity poses an even greater challenge to researchers. In practical applications, many relationships between entities cannot be expressed in a single sentence. For example, in long documents such as news reports, academic papers, legal documents, etc., multiple details describing an event are usually scattered in multiple sentences or paragraphs. Therefore, DocRE is widely used in information extraction tasks, such as event correlation, character relationship, legal analysis and other scenarios. In a document, the relationship between two entities may not be directly described in the same sentence. For example, in a report, a certain character A may be mentioned in the first paragraph, and another character B and the relationship between them are mentioned only a few paragraphs later. Such cross-sentence or even cross-paragraph relational reasoning requires the model to have stronger contextual understanding. Sometimes, document-level relationship extraction also involves multi-hop reasoning, where the model not only extracts direct relationships between entities, but also reasons through multiple intermediate entities. For example, the relationship between A and B is not directly stated in the document, but the model can deduce the relationship between A and B by referring to the third-party entity C.

Due to the long length of documents, the number of entities and relationships involved is often large, which requires the model to be able to efficiently handle long texts and accurately model the connections between entities in complex contexts. This is an important reason why document-level relationship extraction is more challenging than sentence-level relationship extraction. In terms of optimization of the framework, previous studies usually make the whole document as a hint to predict the relationships of all entity pairs, but the models may not be robust and will make errors when non-evidence sentences are removed. These models fall short in learning cross-sentence entity interactions and handling back references in documents and most existing Bio-DocuRE research approaches do not consider relational reasoning with limited effectiveness. Recently, a Sentence Importance Estimation and Focusing (SIEF) framework has been proposed to boost models to concentrate on testimony sentences by designing sentence importance scores and sentence focusing losses. The results of experiments show that SIEF not only enhances overall performance but also enhances the robustness of the DocRE model. A new framework has been proposed through experiments to explicitly and jointly model co-referential and back-referential information to capture fine-grained interactions between entities. The graph structure is dynamically learned by introducing new document graphs and attention mechanisms to fully model the data transfer on the graph and extract expressive entity representations for final classification. In addition, evidence retrieval is introduced as a supplementary task to help the model filter irrelevant information. The experimental results show that the proposed back-finger-assisted model outperforms previous approaches on both datasets. Lastly, experiments are presenting a new Bio-DocuRE model FILR that utilizes multidimensional fusion information and multi-granularity logical reasoning, including a multidimensional information fusion module MDIF and a multi-granularity reasoning module MGLR. Experiments on two biomedical datasets show that FILR achieves state-of-the-art performance.

In terms of optimization methods for relation extraction, the existing dataset DocRED suffers from the problem of false negatives and imperfect data distribution, and there are limitations in the existing methods, such as under-utilizing the rich information in the knowledge base, non-customization of inputs and under-utilizing the text-path connections. A study proposes a method that combines the Large Language Model (LLM) and Natural Language Inference (NLI) modules to generate relational triples to augment the DocRE dataset, creates the DocGNRE dataset, and experimentally verifies the effectiveness of the method in zero-sample document-level RE and training using remote triples. It has also been proposed to establish a neighborhood knowledge graph containing the neighbors of the entity graph in the input documents and incorporate it into a biomedical DSRE model by combining it with an entity representation using a remotely supervised corpus, which has been experimentally demonstrated to improve the micro-averaged F1 scores on the ChemDisGene dataset. There are also studies proposing entity-based document context filters and cross-path entity relationship attention models, which are experimentally validated on the CodRED dataset, significantly driving the performance of CodRE to the state-of-the-art.

In terms of the parameters of the evaluation criteria, the decision rules of the existing models in this task are yet to be investigated, and it is not clear whether they are capable of language comprehension and reasoning. A study proposed the DocREDHWE dataset by annotating DocRED with human-annotated word-level evidence to analyze DocRE more comprehensively.A representative model among the document-level RE models was selected, and by observing the key words that the model considered in the reasoning process through the feature attribution method, it was found that there are differences in the model's decision rules from those of humans and that the model relies on irrelevant information (e.g., entity names, certain fixed locations, and irrelevant words) with spurious associations with the final predictions. In addition, the unreliability of the model's decision rules and their vulnerability to real-world applications were revealed by designing RE-specific attacks.

Finally, the article introduces Mean Average Precision (MAP) to assess the comprehension and reasoning ability of the model, and considers both performance and comprehension when evaluating the model. In this paper, the paper first explore the progress of current optimization of relational extraction techniques at the framework level, including the SIEF, the Fingerback Assisted Model, and the FILR framework, etc.; then, the paper introduces the optimization of relational extraction methods, covering data augmentation, remote supervision, and the construction of the domain knowledge graph, etc.; and finally, the paper discusses the new evaluation parameter --Mean Average Precision (MAP) in assessing model understanding and reasoning capabilities, and look forward to future research directions.

2. Up-to-date information extraction methods

2.1. Impact of different frameworks on relational extraction

Wang Xu has proposed a new Sentence Importance Estimation and Focusing (SIEF) framework to motivate the model to concentrate on evidence sentences and predict entity relationships between sentences. Wang Xu determines the significance of each sentence for a particular entity pair. Sentences with lower scores are deemed non-evidence and can be removed without altering the DocRE prediction. Wang Xu posits a sentence importance score based on the DocRE prediction that either includes or excludes the sentence in question. The process of extracting relationships for evidence is usually monotonous, which means more relationships will be predicted for more sentences. A sentence that is removed and there is a decrease in the predicted probability of a relation is more likely to be evidence. The sentence is likely to be unsubstantiated if the predicted probability does not change. When sentences are deleted, the predicted probability may increase, causing the DocRE model to be unreliable and violating monotonicity [1].

Converter-based models take as input only the word sequences of documents and utilize converters to implicitly capture remote contextual dependencies between entities. However, these models are limited in their ability to handle complex instances that require reasoning because they only consider word sequences as input without considering the internal structure of entities and cannot explicitly learn external interactions between entities. Recent studies suggest improving the reasoning abilities of converter-based models by transforming a document into a graph, where entity mentions serve as nodes. Graph-based models emphasize the creation of a document graph and the explicit learning of interactions between entities derived from this graph [2].

The current inference models are based solely on one type of elemental information, disregarding the fact that different granularities of inference information are complementary. In addition, logical reasoning requires access to rich document information, but many previous models fail to fully utilize document information, which limits their inference capability. Lishuang Li proposes a new Bio-DocuRE model FILR that combines information from multiple dimensions and logical reasoning with multi-granularity. Specifically, FILR proposes a multidimensional information fusion module MDIF to extract sufficient global document information. Then, FILR proposes a multi-granularity reasoning module MGLR to obtain rich inference information by reasoning on entity pairs and mention pairs [3].

2.2. Introduction to the optimization of the relation extraction method

Context learning based on large language models has potential in relation extraction, but the direct use of conventional context learning methods is not feasible in document-level relation extraction due to the need to deal with a large number of predefined fine-grained relation types and the uncontrolled generation of LLMs. To solve this problem, Junpeng Li proposes a method that combines LLM and NLI modules to generate relation triples [4], thus expanding the document-level relation dataset. In addition, the authors demonstrate the effectiveness of their approach by introducing the DocGNRE dataset and point out that the approach has the potential for wider application in the definition of domain-specific relation types and has practical benefits in advancing the semantic understanding of generalized languages. The concept of the Document Level Relationship Extraction (DocRE) task is introduced. The DocRE task focuses on the extraction of fine-grained relationships between pairs of entities in long textual contexts. The complexity of DocRE, in contrast to sentence-level relation extraction (RE), is highlighted due to the significant number of entity pairs and relation types present within a document.

The main role of this method. 1. GPT performs poorly in the zero-learning document-level relationship extraction task, but performance is improved by combining the NLI module with relationship descriptions. 2. The framework can automatically generate distantly labeled data for document-level relationships, which, after manual validation, resulted in a new test set DocGNRE with more relationship triples than Re-DocRED.3. In the experiments, it was found that most of the relations generated by GPT are natural language representations, which do not exactly match the predefined relation types, but the NLI module is able to efficiently map GPT answers to the predefined relation types.4. The method achieved high recall on the DocGNRE test set.5. This research provides new ideas and methods for document-level relationship extraction, which is expected to promote the development of this field.

Fengqi Wang proposed new solutions to advance the state-of-the-art in cross-document relational extraction [5]. The article focuses on building inputs for the relationship extraction model and proposes an entity-based document context filter. This filter utilizes bridging entities in the text path to retain useful information in a given document. The article proposes a cross-document relational extraction model based on cross-path entity relational attention. This model allows entity relationships across text paths to interact with each other. To validate the effectiveness of this cross-document relationship extraction method, the article compares it with the current state-of-the-art method on the CodRED dataset. The results show that the method outperforms the other methods by at least 10% on the F1 score, which proves its effectiveness.

Fengqi Wang proposed a method called Entity-based Cross-path Relation Inference Method (ECRIM).

The core ideas and advantages of the ECRIM method [5]:

CORE IDEA: The focus of optimizer3 is to address the above issues and improve the performance of cross-document relationship extraction by proposing an entity-based cross-path relationship inference method (ECRIM).

Input Optimization: ECRIM first proposes an entity-based document context filter to carefully construct the inputs to the cross-document relational extraction model. This consists of two steps: first, filtering out a number of sentences based on the scores of the bridge entities; second, using heuristics to characterize the importance scores of the bridge entities and assigning these scores to the sentence filter. Utilizing bridge entities: in this way, ECRIM is able to capture more accurately the bridge entities and textual paths that are used to reason about the relationships between the target entities.

Utilizing connections between text paths: ECRIM not only utilizes the information of each text path individually, but also fully considers the global connections between all text paths. This enables the model to better utilize the connections between different text paths for relational reasoning. The importance of cross-document relationship extraction and the shortcomings of existing approaches are presented, as well as the proposal of the ECRIM approach and its core advantages. ECRIM aims to improve the accuracy of cross-document relationship extraction by optimizing the inputs, and by exploiting the connections between bridge entities and text paths.

In conclusion this method can capture the global dependencies between text paths more accurately, which helps to improve the accuracy and efficiency of relationship extraction.

Zhepei Wei proposes a new idea: adopting a brand new perspective instead of the previous categorical perspective [6], modeling the relationship as a mapping function from S to 0. A new framework is proposed: Compositional Activity Service-Oriented RELational Model (CASREL), i.e., Compositional Activity Service-Oriented Relational Model, is a service-oriented system architecture design methodology. The CASREL model is mainly used to guide the informatization construction of enterprises or organizations, helping them to CASREL model is mainly used to guide the informatization construction of enterprises or organizations, helping them to better realize the integration of business processes, organizational structures and systems.

The three main components in the model and their roles are:

1.BERT-based encoder module can be replaced with different encoding frameworks that mainly encode the words in the sentence, the paper ends up being BERT-based with strong results.

2. subject tagging module: the purpose is to identify the subject in a sentence.

3. relation-specific object tagging module: find possible relations and objects according to subject.

The CASREL framework effect uses two publicly available datasets, NYT and WebNLG.

The specific implementation results are where the CASREL model uses the BERT coding side with random initialization parameters, the LSTM coding side, and the pre-trained BERT coding side, respectively.

The CASREL framework is indeed effective in improving the efficiency of relational extraction, and all three coding structures are far more effective, and the CASREL model performance can be made more efficient by using BERT pre-training.

2.3. Introduction of new parameters to evaluate the model

Haotian Chen chooses one from each class of document-level RE models (DocuNet for graph-based approach and ATLOP for transformer-based approach) and generates attributes by feature attribution (FA) method. Integral Gradient (IG) is chosen as the attribution method in this paper because of its proven simplicity and reliability [7], which makes IG more applicable to other text-related tasks [8-11]. Integral Gradient is a reference-based method that calculates both the model output on the input and the model output at the reference point. The difference between the outputs is assigned as an importance score for each token. The DocREDScratch dataset and their own proposed DocREDHWE dataset are used as test sets.

In the experiments, this paper analyzes and describes the comprehensibility of the SOTA model in DocRE, exposes the bottlenecks of the model, and then introduces a new evaluation metric to select plausible and robust models from those that perform well. Calculation: MAP is a widely used metric for evaluating the performance of models, considering "human-annotated evidence words" as "relevant items for the user" and "most important words considered by the model" as "recommended words". The "human-annotated evidence words" are considered as "relevant items for the user", and the "most important words considered by the model" are considered as "items recommended by the recommender system". For a given K words with the highest attribution value, MAP is formulated as $MAP(k) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{k}AP(K) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{k}\sum_{i=1}^{K}p_t(i) * l_t(i)$, where $l_t(i)$ denotes the indicator function that predicts the i-th important word for the t-th relational fact, and outputs a value of 1 if the word is in the human-annotated word-level evidence, and 0 otherwise. The choice of K depends on the needs of RE practitioners, and is usually set to 1, 10, 50, and 100, and all possible values can be chosen to form the MAP curves and measure the AUC to fully assess the comprehension of the model. Assessment Role: Based on MAP, the degree to which a model's decision rules differ from those of humans can be measured to assess the model's vulnerability and robustness. Models with lower values of MAP are vulnerable in the face of attacks, while those with higher values of MAP are more trustworthy and robust, and are able to be better deployed in real-world applications. The experimental results show that the MAP values of all existing models are much lower than the average of common recommender systems, which explains their vulnerability to attacks.

3. Theoretical basis and concrete realization of the latest results

3.1. Specific implementations of different frameworks

Wang Xu proposed using loss-based sentence analysis to make it easier for the model to produce identical output distributions by using the whole document as input and removing non-evidence sentences. Perfectly, if any combination of sentences in kij (the set of generated non-evidence sentences) is deleted, the predictive probability should remain unchanged, and therefore, a penalty is applied to the extent to which the predictive probability changes. Wang Xu presents a versatile framework that can be readily adapted to various DocRE models without adding extra parameters. Experimental findings indicate that SIEF consistently enhances the performance of the foundational models across different domains while also increasing the robustness of DocRE models [1].

Chonggang Lu uses off-the-shelf NLP tools to recognize back references. Specifically, lexical annotation is performed using the annotation function of the space, which treats all pronouns (PRON) as potential analogies. In addition, Chonggang Lu utilizes the dependency parser of the space to help identify analogies. In particular, when the dependency of a token is "det" (indicative qualifier) and its text is "the", it involves identifying all the tokens that lie between this particular "the" token Chonggang Lu proposes a method for explicitly and jointly modeling co-referential and referential information in order to effectively capture the internal structure and external interactions of entities. Moreover, a dynamic algorithm is utilized for optimizing the graph structure and pruning, necessitating very few extra annotations. Chonggang Lu also introduces evidence retrieval as an auxiliary task to enhance the encoder. Empirical studies on well-established benchmarks confirm the effectiveness of the model proposed in this paper [2].

FILR begins by acquiring the global context and structural information of the document. Subsequently, the MDIF module effectively combines these two types of document information. Next, the MGLR module conducts entity-pair and mention-pair reasoning simultaneously. Finally, Lishuang Li assesses the FILR model in the paper using two popular biomedical datasets, CDR and GDA, with experimental results demonstrating that the FILR model achieves state-of-the-art performance on both datasets[3].

3.2. Application of relational optimization methods

Junpeng Li in In order to align GPT-generated relations with predefined relation types, the authors first combine a natural language inference (NLI) model with GPT to solve zero-sample DocRE. results show that although GPT generation only hits a part of Re-DocRED's truth, it detects a large number of externally valid relation triples. Therefore, the authors designed a pipeline framework to further complement Re-DocRED's test set and automatically generate remote training annotations by combining GPT and NLI modules. GPT-3.5 (gpt-3.5-turbo) was chosen as the LLM module, considering the balance between cost and performance. Given that the original DocRED dataset provides a list of entities for each document, the authors restricted the response of GPT to use only the entities in the provided list. In order to realize the mapping of GPT-generated relations to predefined types, the authors used an NLI model, which has shown effectiveness in evaluating factual consistency. With the above steps, each document in the Re-DocRED training set can be processed to generate additional remote relation triples [4].

The latest RE research has shifted to cross-document RE (CodRE), where the target entities are located in different documents. A CodRE model needs to first retrieve relevant documents and then identify critical textual paths in these documents for relational reasoning.

An entity-based document context filtering approach is proposed in Fengqi Wang to carefully construct the inputs for our cross-document RE model [5], which consists of two steps: 1) The authors filter out some sentences based on the scores of bridging entities. Three heuristics are used to characterize the importance scores of bridging entities, and these scores are then assigned to sentences for filtering.2) After filtering out the lower scoring sentences, the authors use a semantic-based sentence filter to rearrange the remaining sentences so that they become a relatively coherent document. In the

work on cross-document level relationship extraction authors consider modeling global dependencies between multiple textual paths (i.e., cross-paths) based on bridging entities, which ensures more reliable reasoning in CodRE. Two documents in each path may share multiple entities, call such entities bridging entities. The authors' main layering for the cross-document relationship extraction model: first, an entity-based document context filter receives as input textual paths, each consisting of two documents. The filter removes less relevant sentences from the text paths and reorganizes the remaining sentences into more compact inputs for subsequent layers. Afterwards, the BERT encoder generates representations for tokens and entities. The cross-path entity-relationship attention module then constructs a package-level entity-relationship matrix to capture global dependencies between entities and relationships in a package and outputs entity-relationship representations for all text paths. Finally, the paper uses a classifier to aggregate these representations and predict the relationships between head and tail entities.

Implementation framework: firstly, the representations of all entity mentions in the package are collected, and then the relationship representations are generated for the entity pairs; after that, in order to model the interaction of the relationships between the paths, the authors constructed a relationship matrix; Entity-based sentence filtering aspect: the filtering procedure is to select those informative sentences with a priori distributional knowledge of the bridging entities. For this purpose, the authors use three steps:

Step 1: Calculate the co-occurrence score for each bridged entity.

Step 2: Calculate the importance score for each sentence s by summarizing the scores of all bridging entities contained in the sentences

Step 3: Sort the sentences in descending order of importance scores and select the top K sentences as the candidate set S = {s1, s2, ..., sK}, where K is a hyperparameter. In the authors' implementation, the candidate set size K is set to 16 based on experiments on the development set. if there are several sentences with the same score, they are prioritized based on their distance from the highest scoring sentence. The encoder module aspect uses unused tokens from the BERT vocabulary to mark the beginning and end of each entity. BERT is then utilized as an encoder to generate the tagged representations; in terms of cross-path entity relation attention first collects the representations of all entity mentions in a package and then generates a relation table for the entity pairs; and in terms of training details the authors introduce an additional threshold to control which category should be output.

In the CasRel model mentioned in Zhepei Weil [6], relation extraction is considered as a joint task, i.e., a model extracts subject, predicate and object simultaneously. This is different from the traditional pipeline-based distributional extraction, which predicts the predicate when the two entities subject and object are known.CasRel encodes the text through a BERT encoder, and then the Casrel model, by identifying the subject in the sentence, and based on the identified subject, the recognizes all possible relations and corresponding objects. This process includes key parts such as head entity recognition layer, relation, and tail entity joint recognition layer. Specifically, the CasRel model uses the BERT pre-trained model as the encoder, and the decoder mainly consists of SubjectTaggera, Relation-Specific Object Taggersa, which can extract Subject entities and tag all possible relations and corresponding Object entities for each Subject entity, so as to identify all possible relations and corresponding Object entities based on the identified subject. corresponding Object entities, thus obtaining the entity-relationship ternary (Subject, Relation, Object). This joint extraction approach can effectively deal with the overlapping relationship situation that exists in real scenarios, i.e., an entity is part of multiple triples at the same time. The experimental results show that the CasRel model achieves high precision, recall, and F1 values in the named entity recognition task proving the effectiveness of the model in handling complex natural language processing tasks. In addition, the implementation of the CasRel model demonstrates how to use PyTorch9 for data processing, model definition, training, and validation, which provides a concrete implementation framework and reference for researchers.

3.3. Introducing the application of new parameters

MAP can be used to evaluate the performance of a model. For example, in a text categorization task, different categories of text can be considered as related items, and the results of the model classification

can be considered as candidates, and the classification accuracy of the model can be evaluated by calculating the MAP.KAMAL NIGAM has conducted an experiment using the MAP criterion [12], in which he compared the performance of the traditional Naive Bayes classifiers, which were trained using only labeled data, and the classifiers trained using EM algorithms combining labeled and unlabeled data. Bayes) and the performance of classifiers trained using an EM algorithm combining labeled and unlabeled data. MAP values were used to measure the classification accuracy of these classifiers on different datasets (e.g., 20 Newsgroups, WebKB, Reuters).Haotian Chen has also analyzed and described the comprehension of the SOTA model in DocRE while using MAP, which measures the extent to which the model's decision rules differ from those of humans, thereby assessing the model's vulnerability and robustness.

4. Challenges and prospects

4.1. Challenges
This paper argues that information extraction still faces a number of challenges today.

On different model choices: Accurately identifying which sentences are evidence sentences remains challenging, especially in complex documents where there may be multiple related sentences. And although some frameworks can be adapted to multiple DocRE models, how to effectively implement the framework in different models still needs to be explored. For example, although the SIEF framework can be adapted to multiple DocRE models, it is not effectively used in all models. In graph-based models, it is still a challenge to effectively capture the internal structure and external interactions between entities, especially when dealing with complex co-referencing and back-referencing relationships.

In Relation Extraction Optimization Methods; The latest research on relation extraction method adopts a new perspective instead of the previous classification perspective, modeling the relation as a mapping function from S to 0. A brand new framework is proposed: CASREL. [optimizer3] The results of CASRE framework extracting ternary (subject, relation, object) mentioned in optimizer3 show that CASREL framework is indeed effective, and the effect of the three coding structures is much higher than the performance of other models at this stage. In this paper, the paper is going to refer to its framework to solve the problems of insufficiently efficient model performance and fast processing speed at this stage.

on new parameters for the assessment of model quality: This paper identifies the following problems: Data Distribution Issues: All data in the DocRED dataset used in the article were taken from Wikipedia and Wikidata, which may result in the training and test data being independently and identically distributed (i.i.d. assumption). This assumption may hinder the intuition that "a model with a higher MAP will get a higher F1 score on the test set", as the model may get a higher F1 score by greedily absorbing all the correlations in the training data (including spurious correlations), rather than truly understanding the document.

Limitations of the dataset: although the authors present the DocREDHWE dataset and annotate DocRED, there may be some problems with DocRED itself, such as labeling noise. Furthermore, extending the study to cleaner Re - DocRED and analyzing the role of unobservable mislabels are important and interesting ideas, but have not been fully investigated.

4.2. Outlook
It is argued that this paper can enhance the quality and efficiency of information extraction to some extent in the following three ways.

In terms of models, firstly, it is an important challenge to comprehensively extract and utilize the global information in documents to support complex logical reasoning. Second, although the models are all optimized, how to optimize the computational complexity and efficiency issues to improve the speed and accuracy still needs to be investigated. The paper hopes that the models can be further optimized in future research to improve the efficiency and accuracy of the framework and capture the internal

structure and external interactions between entities more effectively. And it can extract evidence sentences more accurately.

In terms of relationship extraction, today's results show that cross-document relationship extraction centered on entities produces too much redundant information, is too complex to implement and the actual meaning of the same entity in the article may change leading to errors. This problem can be studied by changing the idea of event-centered relationship extraction to reduce redundant information and improve the accuracy of extraction.

In terms of model assessment parameters, this paper argues that the average precision mean can also be optimized in the following ways. In addition to the word level, the evaluation parameters can be designed to examine the comprehension and reasoning ability at the phrase or sentence level to achieve the effect of multi-scale evaluation. It is also possible to develop a comprehensive evaluation framework that can be used not only for DocRE, but also for other NLP tasks, such as Named Entity Recognition (NER), Natural Language Reasoning (NLI), and so on. Not only that, feedback from human experts is introduced to validate the model's decision-making process when necessary to ensure that the model's behavior is not only statistically valid but also logically sound. Finally, a dynamic evaluation system is established, which can adjust the evaluation criteria according to the model's performance in different scenarios, so as to better adapt to the ever-changing data distribution.

5. Conclusion

This paper reviews recent research results in the field of document-level relationship extraction (DocRE), focusing on model optimization, improvements in relationship extraction methods, and new parameters for model evaluation. By analyzing several current mainstream frameworks such as SIEF, Fingerback Assisted Modeling, FILR and their applications on biomedical datasets, the contributions and limitations of these approaches in improving the performance of DocRE are revealed. At the same time, this paper points out the remaining challenges in model selection and implementation, relation extraction optimization methods, and model evaluation, such as how to identify evidence sentences more efficiently, deal with complex co-referential and back-referential relations, and address the noise present in the dataset.

Based on the current research trends, the paper expects that future research will focus on the following aspects: first, further optimization of the model to support more complex logical reasoning, while reducing computational complexity and improving processing speed; second, shifting from entity-centered to event-centered in the relationship extraction approach to reduce redundant information and improve the extraction accuracy; third, developing a multi-scale assessment framework, which is not only limited to the word level evaluation, but also to cover phrase- and sentence-level comprehension and reasoning ability, and even to introduce feedback from human experts to ensure the reasonableness of modeling decisions, and to establish a dynamic evaluation system to adapt to changes in data distribution. Despite the challenges of document-level relational extraction, with the development and application of new technologies, the paper has reason to believe that research in this area will make greater breakthroughs and provide more powerful tools for natural language processing.

References

[1] Xu, W., Chen, K. H., Mou, L. L., Zhao, T. J. (2022). Document-Level Relation Extraction with Sentences Importance Estimation and FocusingProceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 2920 - 2929

[2] Lu, C., Zhang, R., Sun, K., Kim, J., Zhang, C., Mao, Y. (2023). Anaphor Assisted Document-Level Relation Extraction. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, pages 15453-15464

[3] Li, L., Lian, R., Lu, H., Tang, J. (2023). Document-level Biomedical Relation Extraction Based on Multi-Dimensional Fusion Information and Multi-Granularity Logical Reasoning.

Proceedings of the 29th International Conference on Computational Linguistics, pages 2098-2107

[4]   Li, J., Jia, Z., Zheng, Z. (2024). Semi-automatic Data Enhancement for Document-Level Relation Extraction with Distant Supervision from Large Language Models. Technical Report. National Key Laboratory of General Artificial Intelligence, BIGAI. https://github.com/bigai-nlco/DocGNR

[5]   Wang, F., Li, F., Fei, H., Li, J., Wu, S., Su, F., Shi, W., Ji, D., & Cai, B. (2022). Entity-centered Cross-document Relation Extraction. Title of Periodical, volume number (issue number).

[6]   Wei, Z., Su, J., Wang, Y., Tian, Y., & Chang, Y. (2020). A Novel Cascade Binary Tagging Framework for Relational Triple Extraction. Technical Report. School of Artificial Intelligence, Jilin University; Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education; International Center of Future Science, Jilin University; Shenzhen Zhuiyi Technology Co.; University of North Carolina at Chapel Hill.

[7]   Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic Attribution for Deep Networks. Proceedings of the 34th International Conference on Machine Learning, Volume 70 of Proceedings of Machine Learning Research, pages 3319-3328. PMLR.

[8]   Mudrakarta, P. K., Taly, A., Sundararajan, M., & Dhamdhere, K. (2018). Did the Model Understand the Question? In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1896-1906, Melbourne, Australia. Association for Computational Linguistics.

[9]   Liu, F., & Avci, B. (2019). Incorporating Priors with Feature Attribution on Text Classification. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 6274-6283, Florence, Italy. Association for Computational Linguistics

[10]   Bastings, J., & Filippova, K. (2020). The Elephant in the Interpretability Room: Why Use Attention as Explanation When We Have Saliency Methods? In Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP, pages 149-155, Online. Association for Computational Linguistics.

[11]   Hao, Y., Dong, L., Wei, F., & Xu, K. (2021). Self Attention Attribution: Interpreting Information Interactions Inside Transformer. In Proceedings of the AAAI Conference on Artificial Intelligence, Volume 35, pages 12963-12971.

[12]   Nigam, K., McCallum, A., Kachites, C., Thrun, S., & Mitchell, T. (2000). Text Classification from Labeled and Unlabeled Documents Using EM. Technical Report. School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA.