A Few-Shot Learning-Focused Survey on Recent Named Entity Recognition and Relation Classification Models

Sakher Khalil Alqaaidi, Elika Bozorgi, Afsaneh Shams and Krzysztof J. Kochut
School of Computing, University of Georgia, Athens, Georgia, U.S.A.

Keywords: Named Entity Recognition, Relation Classification, Natural Language Processing, Information Extraction.

Abstract: Named Entity Recognition (NER) and Relation Classification (RC) are important tasks for extracting information from unstructured text and transforming it into a machine-readable format. Recently, the field of few-shot learning has gained increased interest due to its ability to enable models to generalize across multiple domains using minimal labeled data. However, no studies have addressed the recent achievements in the NER and RC fields within the few-shot learning paradigm. In this work, we aim to fill this gap by presenting a survey on recent few-shot learning models in the fields of NER and RC. Our survey provides a thorough introduction to these tasks, along with a summary of the latest approaches and achievements. We conclude with our observations on the current state of research in these domains.

1 INTRODUCTION

Named Entity Recognition (NER) and Relation Classification (RC) are crucial tasks for extracting information from unstructured text and converting it into a machine-readable format. Several Natural Language Processing (NLP) applications employ the two steps, either separately or simultaneously, such as information retrieval and information extraction, knowledge graph construction (Zou, 2020), question answering, and other domain-specific applications, such as biomedical data mining (Quirk and Poon, 2016).

The NER task targets labeling subsets of words in a text that designate entities and assigning a type to each entity, such as Person or Location, etc. The RC task aims to identify all the valid semantic relations between two given entities. Figure 1 shows a general diagram for the two tasks. A variety of methods have been proposed to train NER and RC models. Early ones used rule-based algorithms, such as text pattern mining (Huffman, 1995), feature-based methods (Kambhatla, 2004) or graphical methods (McClosky et al., 2011). Followed that models that used text representation as input in neural networks (Luo et al., 2020; Wang et al., 2020a). Neural network models have achieved state-of-the-art performance. Thanks to the advancements in deep-learning techniques. Models also varied by the learning approaches. Supervised learning is a widely used approach. However, supervised learning requires large amounts of labeled data and heavy resources for training. Furthermore, it cannot generalize to multiple domains. Primitively, these issues were handled by weak or distant supervision models (Hoffmann et al., 2011). However, noisy labels have always put obstacles to reaching good results in the weak or distant supervision models. On the other hand, few-shot learning has shown its capabilities in achieving reasonable scores using minimal labeled data and can easily adapt to new domains (Bragg et al., 2021).

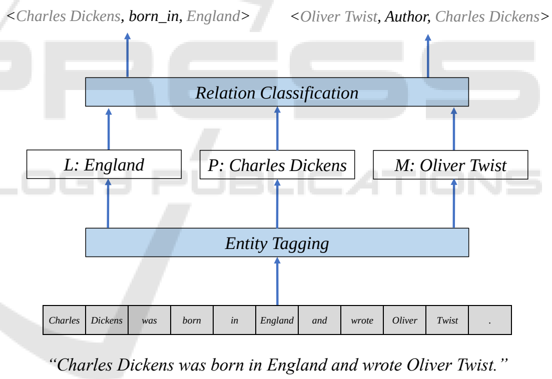

Figure 1: The diagram of the NER and RC tasks.

Recent NER and RC surveys focused on deep learning models (Yadav and Bethard, 2019; Li et al., 2020a), and a few considered surveying both fields in a single study. Accordingly, A survey that focuses on recent few-shot learning models in both fields would

Alqaaidi, S., Bozorgi, E., Shams, A. and Kochut, K.
A Few-Shot Learning-Focused Survey on Recent Named Entity Recognition and Relation Classification Models.
DOI: 10.5220/0012791600003756

In *Proceedings of the 13th International Conference on Data Science, Technology and Applications (DATA 2024)*, pages 450-459
ISBN: 978-989-758-707-8; ISSN: 2184-285X

enhance current research efforts. Consequently, we address this gap by presenting this survey that studies the recent few-shot learning models in the NER and RC fields with a quick review of other models that adopted different learning approaches. This contribution entails presenting the most recent advancements and exploring the latest methodologies. Our survey categorizes the recent works based on the model input and follows careful criteria in selecting the papers.

Our survey is divided as follows: Section 2 describes previous and related surveys. Section 3 explains our methodology in selecting the models for this survey. Section 4 sets the foundational terms and definitions. Section 5 describes the datasets that we found commonly used in both the NER and RC tasks. Section 6 shows the models that have been found handling both the NER and RC tasks. Section 7 shows the models that solely addressed the NER task. Section 8 shows the models that addressed the relation classification task only. Finally, we conclude our observations in Section 9.

2 PREVIOUS WORKS

Our work is the first work that considers the two tasks with a focus on few-shot learning methods. The surveys in (Yadav and Bethard, 2019; Li et al., 2020a) considered the NER task methods only, they showed early approaches and focused on deep learning models. The survey in (Nasar et al., 2021) considered works from the NER and RE tasks with a focus also on deep learning models. The survey in (Han et al., 2020) reviewed the works in the Relation Extraction (RE) task and categorized them based on their approaches, then discussed more paths in the RE task to be explored.

3 METHODOLOGY

Given that early works in the NER and RC fields complied with pattern-based or feature-based approaches, followed by machine learning-based approaches, we only considered machine learning-based models in this survey for the following reasons. First, pattern-based or feature-based models have significantly lower scores on several benchmarks compared to deep learning models (Nasar et al., 2021; Yadav and Bethard, 2019). Accordingly, a few models adopted pattern-based or feature-based methods solely during the last few years. Second, several surveys have addressed non-machine learning models sufficiently while we target surveying recent accomplishments.

With hundreds of works in the NER and RC tasks available in the literature and to present a survey that focuses on deep learning-based models for the reasons mentioned earlier, we chose the models that were published in 2019 and later. We selected the 2019 year because it witnessed the beginning of using some revolutionary Pre-trained Language Models (PLMs), such as BERT (Devlin et al., 2018) and GPT (Brown et al., 2020); these PLMs were employed to score new state-of-the-art performances in most of the NLP tasks. With the adoption of the English language for many NLP benchmarks and evaluations, we excluded works that pursued other languages solely from our search results. Furthermore, we excluded domain-specific works to survey general-use models that can be adapted for other domains. We searched Google Scholar for the terms: "*Relation Extraction*", "*Named Entity Recognition*", "*Relation Classification*", and "*Triple Extraction*". We selected the papers that have any of the terms in the title or the content that appeared in the first 100 search results, and then we gave a rank based on the following factors:

- Number of citations.
- The model presents a few-shot learning results.
- The model handles both NER and RC tasks together.
- Publication year.

The last factor is considered for fairness with papers that were published in the same year of writing this survey and did not receive an adequate number of citations.

4 PRELIMINARIES

The NER task targets labeling subsets of words in the text that designate entities. Formally $f(W) = E$, where f is a trainable function, W is a sequence of words of size n, $W = \{w_1, w_2 ... w_n\}$, E is a set of labels, and $E = \{e_1, e_2 ... e_n\}$. $e \in L$, where L is a set of entity types, such as Person, Location, etc. An entity e may contain multiple w words. It is not necessary that all words within an entity are adjacent, this type of entity is called a discontinuous entity. For example, the term *"The teams of France and Italy"* incurs two entities *"The team of France"* and *"The team of Italy"*. An entity of multiple words may contain instances of sub-entities. For example, *"The governor of Bryxton"* is an entity, and *"Bryxton"* is a sub-entity; this type of entity is called a nested entity.

Table 1: Statistics of popular NER, RC, and RE Benchmarks.

Benchmark	Train	Validation	Test	Total
CoNLL2003	14,041	3,250	3,453	20,744
OntoNotes5.0	59,924	8,528	8,262	76,714
FEW-NERD (INTRA)	99,519	19,358	44,059	162,936
FEW-NERD (INTER)	130,112	18,817	14,007	162,936
TacRed	68,124	22,631	15,509	106,264
Re-TacRed	58,465	19,584	13,418	91,467
FewRel	44,800	11,200	14,000	70,000
NYT	56,196	5,000	5,000	66,196
WebNLG	5,019	500	703	6,222

The RC task, in its simplest implementation, aims to identify if a relation between two entities exists within a text sequence. However, the task is more practical when identifying the relation type. Formally, $f(W,e1,e2) = r$, where f is a trainable function, W is a text sequence, $e1$ and $e2$ are two tagged entities in W, and r is the relation. In the open-world scenario, the relation can be either predicted without the need to be seen during training. Controversially, a relations set needs to be pre-defined in the closed-world scenario.

Several models employed the NER and RC tasks jointly to extract relational triples (Li et al., 2021; Sui et al., 2020; Cabot and Navigli, 2021; Tang et al., 2022). A relational triple consists of the following three items respectively, a head or subject entity, a semantic relation, and a tail or object entity. Different names were used for this simultaneous task, such as Relation Extraction (RE) and Triple Extraction (TE). In a sentence, multiple triples may share a single entity in a case named *Single Entity Overlap*, Figure 1 shows an example of the entity *"Charles Dickens"* that is found in two triples because it is a part of two input items in the RC task. A more complicated scenario is when multiple relations connect the same entities, this case is called *Entity Pair Overlap*. For instance, the entities *"Bern"* and *"Switzerland"* can have the two relations *"capital_of"* and *"city_in"* in the sentence *"Bern is not only a city in Switzerland but also the capital"*.

Few-shot learning has been found practical in classification problems when access to labeled data is limited (Bragg et al., 2021). Different approaches can be used to design a few-shot learning model, including data generation to provide sufficient labeled data (Mishra et al., 2018), or transfer learning when pre-trained models are employed (Sun et al., 2019). There are different model architectures for few-shot learning; the most popular and the easiest to grasp is the N-way-K-shot framework, where a trainable function is exposed to a support set of few instances of size K in each label. The size of the labels is N. After training the function, a query set is used for evaluation. A loss function measures how far the predictions are from the ground truth labels.

5 BENCHMARKS

In this section, we describe widely used benchmarks in the NER and RC tasks. Statistics of the datasets are shown in Table 1.

5.1 Named Entity Recognition Benchmarks

CoNLL2003 (Sang and De Meulder, 2003) is a dataset that was built using the Reuters news corpus. It has four entity types: Persons, Locations, Organizations, and Miscellaneous. **OntoNotes5.0** (Weischedel et al., 2013) is an annotated text dataset that has part of speech (POS) and NER tags. The dataset was built using a corpus of various types of text content, such as news, conversational telephone speech, weblogs, newsgroups, broadcasts, and talk shows. OntoNotes has different language variants including English. **FEW-NERD** (Ding et al., 2021) is the first released dataset for few-shot NER evaluation. It has two variants. In FEW-NERD (INTRA), the evaluation entity types are not seen during training, which makes it harder compared to FEW-NERD (INTRA), where splits share the same types. The dataset sentences were retrieved from Wikipedia articles. The dataset has 491.7k annotated entities of 8 coarse-grained types and 66 fine-grained types.

5.2 Relation Classification Benchmarks

TacRed (Zhang et al., 2017) is a dataset that has been used in both the RC and RE tasks. It is derived from news articles and web content. The original release had 41 relation types. The dataset was designed for supervised learning evaluation. The work in (Sabo et al., 2021) showed some drawbacks in popular few-shot learning datasets and proposed an approach to customize the supervised learning ones for few-shot evaluation, such as TacRed. Later on, Re-TacRed was released as an improved version of the original one (Stoica et al., 2021). **FewRel** (Gao et al., 2019b) is a Few-Shot relation classification dataset of 100 relations in sentences derived from Wikipedia and labeled by crowdsourcing. The training part has 64 relations, the validation part has 16 relations and the test part has 20 relations. Soon after the release of FewRel, authors presented a new version to examine the models' ability to adapt to new domains. Although FewRel was adopted by many works, the study in (Sabo et al., 2021) showed that the dataset is still far from real-world scenarios, thus authors proposed a mechanism to switch supervised datasets, such as TACRED, to apply to the few-shot training.

5.3 Relation Extraction Benchmarks

NYT (Riedel et al., 2010) is a dataset that was generated from a large New York Times articles corpus, where each item consisted of a sentence and a set of triples. Each triple is composed of subject and object entities, and a relation. **WebNLG** is a dataset that was originally generated for the Natural Language Generation (NLG) task, CopyRE (Zeng et al., 2018) customized the dataset for the triples and relations extraction tasks.

6 UNIFIED NER AND RC MODELS

In this section, we present the models that addressed both the NER and RC tasks. These models produced outputs in two formats: either as separate sets of entities and relations or as combined entities and relations represented as triples. Early RE models utilized a pipeline approach, where NER or RC was conducted at the beginning, and then the output of the process was used for running the second task. For instance, in Figure 1, the entities are extracted first, then they are fed as input for the RC task. However, this method suffers from the error propagation issue. Specifically, the errors from the first stage propagate to the second one and affect the overall performance. Thus, recent models performed a simultaneous validation for the NER and RC tasks while training the model. We explain those works in the following paragraph.

DeepStruct (Wang et al., 2022) is a supervised learning model with a zero-shot learning variant. The authors trained different language models with a triple extraction objective and then fine-tuned them for NLP downstream tasks. In LUKE (Yamada et al., 2020), authors utilized different masking and self-attention strategies to train BERT to tag entities within the text. LUKE was evaluated for different NLP tasks. The model in (Liu et al., 2022) represented text as actions to build a structure of dependencies between words for a supervised learning approach. In PL-Marker (Ye et al., 2021), markers were used in the text sequence to tag and classify entities and to extract their relations. In Set Prediction Network (SPN) (Sui et al., 2023), a non-autoregressive decoder architecture was used to jointly extract entities and their relations in the form of triples. The authors proposed a loss function to handle the prediction format of triple sets. Furthermore, the model tackled the entities overlapping problems. PURE (Zhong and Chen, 2020) is a supervised learning model that employed the pipeline approach.

7 NAMED ENTITY RECOGNITION MODELS

In this section, we provide an overview of NER models, highlighting their key characteristics. We begin with an examination of few-shot NER models, followed by a categorization of other models based on their input types. Table 2 summarizes the information in this section.

7.1 Few-Shot NER Models

The work in (Cui et al., 2021) used manually created templates of facts retrieved from different datasets to train their model. For instance, *"Bangkok is a location entity"* is a fact retrieved from the sentence "ACL will be held in Bangkok". StructShot model (Yang and Katiyar, 2020) utilized contextual representations of the labels from the support set instead of traditional approaches. However, their model was capable of detecting nested entities. ContaiNER (Das et al., 2021) employed contrastive learning for the NER task by decreasing the distance between similar entities and increasing the distance between dissimilar ones. The paper in (Hou et al., 2020) presented L-TapNet+CDT, a model that used conditional random fields (CRF) to exploit label dependencies from the source domain to the target domain in the few-shot scope. Additionally, the authors proposed L-TapNet to enlarge the gap between label embeddings, so it becomes able to detect the similarity between an input word and its label, such as *"rain"* and *"weather"*. In MUCO (Tong et al., 2021), a classifier was trained to learn to cluster entity pairs based on the non-entity class

Table 2: The main characteristics of the NER models.

Model Reference	Learning Type	Input Type	Nested Entities	Text Encoding
(Yu et al., 2020)	Supervised	Sent.	Yes	BERT
(Wang et al., 2022)	Supervised,Zero-shot	Sent.	No	GLM
(Liang et al., 2020)	Distant Supervision	Sent.	No	Roberta
(Cui et al., 2021)	Few-shot	Sent.	No	BART
(Luo et al., 2020)	Supervised	Both	No	BERT
(Yang and Katiyar, 2020)	Few-shot	Sent.	No	BERT
(Lison et al., 2020)	Weak Supervision	Doc.	No	BERT
(Wang et al., 2020a)	Supervised	Sent.	Yes	BERT
(Shen et al., 2021)	Supervised	Sent.	Yes	Variety
(Li et al., 2022)	Supervised	Sent.	Yes	BERT
(Schweter and Akbik, 2020)	Supervised	Doc.	No	Roberta,Glove
(Wang et al., 2020b)	Supervised	Sent.	No	Multiple
(Ye et al., 2021)	Supervised	Sent.	Yes	BERT variants
(Liu et al., 2022)	Supervised	Sent.	No	T5
(Yang et al., 2024)	Supervised	Sent.	Yes	Variety
(Ma et al., 2023)	Few-shot	Sent.	No	-
(Mao et al., 2024)	Supervised	Sent.	Yes	BERT,BiLSTM
(Geng et al., 2023)	Supervised	Sent.	Yes	BERT
(Chen et al., 2023b)	Supervised	Sent.	Yes	-

word that falls between any pair. Thus, the model explored common semantics between entities that belonged to the same cluster. In MAML-ProtoNet (Ma et al., 2022), a component detected text spans, and then another component labeled them with the entity type. Their approach targeted mitigating the effect of non-entity class (O-class) spans. In C2FNER (Ma et al., 2023), a model was trained on a coarse-grained class and then employed to distinguish fine-grained class using Few-shot learning.

7.2 Comprehensive NER Models

Comprehensive NER models tackle both nested and flat entities. Machine Reading Comprehension (MRC) (Liu et al., 2019) methods handled NLP problems as a question-answering task. BERT-MRC (Li et al., 2019) is an MRC model for the NER task and it was able to extract nested entities. In a different approach, the work in (Yu et al., 2020) defined the NER task as the detection of the indices of entity heads and entity tails in a sentence. The model utilized dependency parsing graph features in addition to the word representations generated by BERT (Devlin et al., 2018) with the representation of the characters. The work in (Shen et al., 2021) employed the two-stage object detector algorithm from computer vision. Pyramid (Wang et al., 2020a) is a neural network layered model that handled deep nested entities. W^2NER model (Li et al., 2022) was able to capture all types of entities: flat, nested, and discontinuous. The model leveraged the relation between entity words to identify entity boundaries. However, the model promises higher computation needs. The model in (Zheng et al., 2019) combined two components in a multi-task learning model. The first used a sequence labeling layer to detect entity boundaries without the common error propagation problem. Whereas the second employed a region classification model to classify the entity boundaries. The evaluations used biomedical datasets and German nested entities dataset. The model used character-level representation for the input. The model in (Tan et al., 2021) used embeddings combination from different language models, then a non-autoregressive decoder for the predictions. Pner (Yang et al., 2024) used a pipeline approach to tackle nested entities. In (Mao et al., 2024), a graph-based model is used for a span tagging procedure. The model tackled different types of NER tags including discontinuous entities. In (Geng et al., 2023), the model utilized self-crossing encoders to enhance the comprehensibility of overlapping information within a sentence. A comprehensive NER model is discussed in (Chen et al., 2023b). The main idea is based on the boundary regression (Chen et al., 2019; Zheng et al., 2019) model to enhance the NER recognition structured around perceptional and cognitive modules.

7.3 Flat NER Models

This section surveys models that did not address nested entities. In (Akbik et al., 2019b), pooling techniques were used to process character-level representation. In TENER (Yan et al., 2019), character level encoding was also employed and attention customization led to text context information capturing. FLERT (Schweter and Akbik, 2020) is an extension of a previous model (FLAIR) (Akbik et al., 2019a). FLERT exploited document-level features for NER. In detail, the method employed two subsets of the text that surround a sentence in the input. The model in (Wang et al., 2021) addressed two types of NER tasks: offline NER, where external resources can be used to enrich the input with related text. And the online NER, where cooperative learning minimized the distance between the input representation and the output distribution. Automated Concatenation of Embeddings (ACE) (Wang et al., 2020b) is a model that used reinforcement learning for selecting the best combination of word representations. The model addressed several NLP tasks, including NER. In TriggerNER (Lin et al., 2020), the words that surround an entity candidate were exploited for prediction. BOND (Liang et al., 2020) is a distant supervision model that utilized Wikipedia and online gazetteers to extract entities from the text at the beginning. In a second stage, two sub-models interact to enhance the model's performance recall, which is called a student-teacher framework.

7.4 Document-Level NER Models

Here we review models capable of comprehending documents for the NER task. In (Luo et al., 2020), authors proposed a sentence-level and document-level input model. The authors employed label embeddings in the sentence-level input to find a similarity score between each label and its input word. At the document-level, a key-value memory was employed for all the embeddings used during training. The input consisted of word and character representations. In (Lison et al., 2020), a weak supervision model employed external knowledge to label data with assistance from other models, such as sequence labeling and heuristic functions, then the output is aggregated for the final sequence labeling. The work in (Luo et al., 2020) proposed a model that handled sentence-level and document-level data. Authors used BERT for word-level representation and IntNet (Xin et al., 2018) for character-level representation in a hierarchical contextualized representation architecture. The work in (Lison et al., 2020) handled only document-level data in a weak supervision manner. Multiple labeling functions annotated the entities, then the output was aggregated; after that, a function was trained to label the entities in the text sequence.

8 RELATION CLASSIFICATION AND EXTRACTION MODELS

Here we describe the surveyed relation classification and relation extraction models. Table 3 summarizes the main properties of the RC models.

8.1 Few-Shot RC Models

The work in (Xie et al., 2020) used a heterogeneous graph neural network (HGNN) to predict relations as a node classification problem. The entities and sentences represent different node types in the graph. Logic-guided Semantic Representation Learning (LSRL) (Li et al., 2020b) is a supporting method that utilized two types of features from knowledge graphs. First, entity and relation embeddings to identify connections between relations. Second, relation inferring rules using rule mining methods. The features are utilized along with the word representations to connect unseen relations to seen ones. The method is model-agnostic; it was evaluated on two zero-shot models, DeViSE (Frome et al., 2013) and ConSE (Norouzi et al., 2013). TD-Proto (Yang et al., 2020) utilized relation and entity descriptions to enhance a prototypical network-based model. Prototypical networks find a prototype for classes and sentences. These networks have been adopted by several RC models and reflected good performance as they supported matching queries with prototypes (Gao et al., 2019a; Ye and Ling, 2019). ProtoNet (Ren et al., 2020) is a prototypical network-based model. The authors combined prototypical techniques from supervised learning and few-shot learning. They used a loss function targeting enlarging the distance between the relation representations in the embedding space.

8.2 Few-Shot RE Models

The work in (Peng et al., 2020) proposed a training framework that enhanced text context absorption for the RE task by applying masks to a portion of the entities. Virtual prompt pre-training (He et al., 2023) is a few-shot learning model based on a novel prompt tuning approach. The authors used GLM (Du et al., 2021) as the language model to encode text.

8.3 Supervised Learning RC Models

RECENT (Lyu and Chen, 2021) is a model-agnostic paradigm, that enhances the performance by restricting the candidate prediction relations based on the entity types. When applied to SpanBERT (Joshi et al.,

Table 3: The main characteristics of the reviewed RC/RE models.

Model Reference	Learning Type	Sent./Doc.	TextEncoding	RE/RC
(He et al., 2023)	Multi.	Sent.	GLM	RE
(He et al., 2023)	Few-shot	Sent.	Custom	RE
(Chen et al., 2023a)	Supervised	Doc.	Glove	RE
(Sui et al., 2023)	Supervised	Sent.	BERT	RE
(Ren et al., 2020)	Few-shot	Sent.	BERT	RE
(Xie et al., 2020)	Few-shot	Sent.	Glove	RC
(Chen et al., 2022)	Supervised	Sent.	Roberta	RE
(Nan et al., 2020)	Supervised	Doc.	BERT	RE
(Zhong and Chen, 2020)	Supervised	Sent.	BERT	RE
(Guo et al., 2019)	Supervised	Both	Graphencoding	RE

2020), the model achieved a new F1 score on the TACRED dataset. In TACNN (Geng et al., 2022), the authors proposed a target attention mechanism that assigned increased weights to important entities in the sentence to enhance identifying a target relation.

8.4 Supervised Learning RE Models

Unlike several works that focused on sentences and other ones for documents, DHGAT (Chen et al., 2023a) is a relation extraction model for dialog-type input. The model encoded text using Glove (Pennington et al., 2014) in addition to part-of-speech tagging and entity type features in the input. The model used a heterogeneous graph attention network to train the model, the graph contained multiple node types, such as utterance nodes, type nodes, word nodes, speaker nodes, and argument nodes. Knowprompt (Chen et al., 2022) is a supervised model that targeted enhancing the word representation by using prompt-tuning. They tackled some challenges in prompt-tuning by enriching the process with extra knowledge. For instance, the model provided entity types while fine-tuning the language model. The model encoded the input using Roberta PLM. Attention Guided Graph Convolutional Networks (AGGCN) model (Guo et al., 2019) is based on dependency parsing graphs. Latent Structure Refinement (LSR) (Nan et al., 2020) generated task-specific dependency graph structures for document-level relations. The model used iterative refinement during training to build global interaction knowledge. The model was evaluated using DocRED (Yao et al., 2019) dataset only, probably due to the lack of document-level data.

9 CONCLUSION

We present a survey of recent deep learning models that address named entity recognition and relation classification, with a focus on few-shot learning performance. In named entity recognition models, we find that entity boundary issues should be handled in the coming works since considering partial match as a correct prediction in multi-word entities is not a trusted evaluation. Furthermore, we find that models can benefit from the advances in language models' prompt-tuning to build strong architectures to achieve new state-of-the-art scores since current models either focus on proposing a complicated model design or on enhancing the word representation.

In the relation classification task, we see that researchers could direct their efforts towards cross-sentence or document-level achievements under the few-shot learning discipline since this reflects more realistic scenarios. Furthermore, there is a lack of datasets for evaluating such types of work. Additionally, efforts should consider combining linguistic features with dependency parsing information to support the reliance on language models and score new results.

REFERENCES

Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., and Vollgraf, R. (2019a). Flair: An easy-to-use framework for state-of-the-art nlp. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics (demonstrations)*, pages 54–59.

Akbik, A., Bergmann, T., and Vollgraf, R. (2019b). Pooled contextualized embeddings for named entity recognition. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 724–728.

Bragg, J., Cohan, A., Lo, K., and Beltagy, I. (2021). Flex: Unifying evaluation for few-shot nlp. *Advances in Neural Information Processing Systems*, 34:15787–15800.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Cabot, P.-L. H. and Navigli, R. (2021). Rebel: Relation extraction by end-to-end language generation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2370–2381.

Chen, H., Hong, P., Han, W., Majumder, N., and Poria, S. (2023a). Dialogue relation extraction with document-level heterogeneous graph attention networks. *Cognitive Computation*, pages 1–10.

Chen, X., Zhang, N., Xie, X., Deng, S., Yao, Y., Tan, C., Huang, F., Si, L., and Chen, H. (2022). Knowprompt: Knowledge-aware prompt-tuning with synergistic optimization for relation extraction. In *Proceedings of the ACM Web Conference 2022*, pages 2778–2788.

Chen, Y., Wu, L., Zheng, Q., Huang, R., Liu, J., Deng, L., Yu, J., Qing, Y., Dong, B., and Chen, P. (2023b). A boundary regression model for nested named entity recognition. *Cognitive Computation*, 15(2):534–551.

Chen, Y., Wu, Y., Qin, Y., Hu, Y., Wang, Z., Huang, R., Cheng, X., and Chen, P. (2019). Recognizing nested named entity based on the neural network boundary assembling model. *IEEE Intelligent Systems*, 35(1):74–81.

Cui, L., Wu, Y., Liu, J., Yang, S., and Zhang, Y. (2021). Template-based named entity recognition using bart. *arXiv preprint arXiv:2106.01760*.

Das, S. S. S., Katiyar, A., Passonneau, R. J., and Zhang, R. (2021). Container: Few-shot named entity recognition via contrastive learning. *arXiv preprint arXiv:2109.07589*.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Ding, N., Xu, G., Chen, Y., Wang, X., Han, X., Xie, P., Zheng, H.-T., and Liu, Z. (2021). Few-nerd: A few-shot named entity recognition dataset. *arXiv preprint arXiv:2105.07464*.

Du, Z., Qian, Y., Liu, X., Ding, M., Qiu, J., Yang, Z., and Tang, J. (2021). All nlp tasks are generation tasks: A general pretraining framework. *arXiv preprint arXiv:2103.10360*.

Frome, A., Corrado, G. S., Shlens, J., Bengio, S., Dean, J., Ranzato, M., and Mikolov, T. (2013). Devise: A deep visual-semantic embedding model. *Advances in neural information processing systems*, 26.

Gao, T., Han, X., Liu, Z., and Sun, M. (2019a). Hybrid attention-based prototypical networks for noisy few-shot relation classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 6407–6414.

Gao, T., Han, X., Zhu, H., Liu, Z., Li, P., Sun, M., and Zhou, J. (2019b). Fewrel 2.0: Towards more challenging few-shot relation classification. *arXiv preprint arXiv:1910.07124*.

Geng, R., Chen, Y., Huang, R., Qin, Y., and Zheng, Q. (2023). Planarized sentence representation for nested named entity recognition. *Information Processing & Management*, 60(4):103352.

Geng, Z., Li, J., Han, Y., and Zhang, Y. (2022). Novel target attention convolutional neural network for relation classification. *Information sciences*, 597:24–37.

Guo, Z., Zhang, Y., and Lu, W. (2019). Attention guided graph convolutional networks for relation extraction. *arXiv preprint arXiv:1906.07510*.

Han, X., Gao, T., Lin, Y., Peng, H., Yang, Y., Xiao, C., Liu, Z., Li, P., Sun, M., and Zhou, J. (2020). More data, more relations, more context and more openness: A review and outlook for relation extraction. *arXiv preprint arXiv:2004.03186*.

He, K., Huang, Y., Mao, R., Gong, T., Li, C., and Cambria, E. (2023). Virtual prompt pre-training for prototype-based few-shot relation extraction. *Expert Systems with Applications*, 213:118927.

Hoffmann, R., Zhang, C., Ling, X., Zettlemoyer, L., and Weld, D. S. (2011). Knowledge-based weak supervision for information extraction of overlapping relations. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, pages 541–550.

Hou, Y., Che, W., Lai, Y., Zhou, Z., Liu, Y., Liu, H., and Liu, T. (2020). Few-shot slot tagging with collapsed dependency transfer and label-enhanced task-adaptive projection network. *arXiv preprint arXiv:2006.05702*.

Huffman, S. B. (1995). Learning information extraction patterns from examples. In *International Joint Conference on Artificial Intelligence*, pages 246–260. Springer.

Joshi, M., Chen, D., Liu, Y., Weld, D. S., Zettlemoyer, L., and Levy, O. (2020). Spanbert: Improving pre-training by representing and predicting spans. *Transactions of the association for computational linguistics*, 8:64–77.

Kambhatla, N. (2004). Combining lexical, syntactic, and semantic features with maximum entropy models for information extraction. In *Proceedings of the ACL interactive poster and demonstration sessions*, pages 178–181.

Li, J., Fei, H., Liu, J., Wu, S., Zhang, M., Teng, C., Ji, D., and Li, F. (2022). Unified named entity recognition as word-word relation classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 10965–10973.

Li, J., Sun, A., Han, J., and Li, C. (2020a). A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*, 34(1):50–70.

Li, J., Wang, R., Zhang, N., Zhang, W., Yang, F., and Chen, H. (2020b). Logic-guided semantic representation learning for zero-shot relation classification. *arXiv preprint arXiv:2010.16068*.

Li, X., Feng, J., Meng, Y., Han, Q., Wu, F., and Li, J. (2019). A unified mrc framework for named entity recognition. *arXiv preprint arXiv:1910.11476*.

Li, X., Luo, X., Dong, C., Yang, D., Luan, B., and He, Z. (2021). Tdeer: An efficient translating decoding schema for joint extraction of entities and relations.

In *Proceedings of the 2021 conference on empirical methods in natural language processing*, pages 8055–8064.

Liang, C., Yu, Y., Jiang, H., Er, S., Wang, R., Zhao, T., and Zhang, C. (2020). Bond: Bert-assisted open-domain named entity recognition with distant supervision. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1054–1064.

Lin, B. Y., Lee, D.-H., Shen, M., Moreno, R., Huang, X., Shiralkar, P., and Ren, X. (2020). Triggerner: Learning with entity triggers as explanations for named entity recognition. *arXiv preprint arXiv:2004.07493*.

Lison, P., Hubin, A., Barnes, J., and Touileb, S. (2020). Named entity recognition without labelled data: A weak supervision approach. *arXiv preprint arXiv:2004.14723*.

Liu, S., Zhang, X., Zhang, S., Wang, H., and Zhang, W. (2019). Neural machine reading comprehension: Methods and trends. *Applied Sciences*, 9(18):3698.

Liu, T., Jiang, Y., Monath, N., Cotterell, R., and Sachan, M. (2022). Autoregressive structured prediction with language models. *arXiv preprint arXiv:2210.14698*.

Luo, Y., Xiao, F., and Zhao, H. (2020). Hierarchical contextualized representation for named entity recognition. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 8441–8448.

Lyu, S. and Chen, H. (2021). Relation classification with entity type restriction. *arXiv preprint arXiv:2105.08393*.

Ma, R., Lin, Z., Chen, X., Zhou, X., Wang, J., Gui, T., Zhang, Q., Gao, X., and Chen, Y. W. (2023). Coarse-to-fine few-shot learning for named entity recognition. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 4115–4129.

Ma, T., Jiang, H., Wu, Q., Zhao, T., and Lin, C.-Y. (2022). Decomposed meta-learning for few-shot named entity recognition. *arXiv preprint arXiv:2204.05751*.

Mao, T., Xu, Y., Liu, W., Peng, J., Chen, L., and Zhou, M. (2024). A simple but effective span-level tagging method for discontinuous named entity recognition. *Neural Computing and Applications*, pages 1–15.

McClosky, D., Surdeanu, M., and Manning, C. D. (2011). Event extraction as dependency parsing. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, pages 1626–1635.

Mishra, A., Verma, V. K., Reddy, M. S. K., Arulkumar, S., Rai, P., and Mittal, A. (2018). A generative approach to zero-shot and few-shot action recognition. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 372–380. IEEE.

Nan, G., Guo, Z., Sekulić, I., and Lu, W. (2020). Reasoning with latent structure refinement for document-level relation extraction. *arXiv preprint arXiv:2005.06312*.

Nasar, Z., Jaffry, S. W., and Malik, M. K. (2021). Named entity recognition and relation extraction: State-of-the-art. *ACM Computing Surveys (CSUR)*, 54(1):1–39.

Norouzi, M., Mikolov, T., Bengio, S., Singer, Y., Shlens, J., Frome, A., Corrado, G. S., and Dean, J. (2013). Zero-shot learning by convex combination of semantic embeddings. *arXiv preprint arXiv:1312.5650*.

Peng, H., Gao, T., Han, X., Lin, Y., Li, P., Liu, Z., Sun, M., and Zhou, J. (2020). Learning from context or names? an empirical study on neural relation extraction. *arXiv preprint arXiv:2010.01923*.

Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Quirk, C. and Poon, H. (2016). Distant supervision for relation extraction beyond the sentence boundary. *arXiv preprint arXiv:1609.04873*.

Ren, H., Cai, Y., Chen, X., Wang, G., and Li, Q. (2020). A two-phase prototypical network model for incremental few-shot relation classification. In *Proceedings of the 28th international conference on computational linguistics*, pages 1618–1629.

Riedel, S., Yao, L., and McCallum, A. (2010). Modeling relations and their mentions without labeled text. In *ECML PKDD*, pages 148–163. Springer.

Sabo, O., Elazar, Y., Goldberg, Y., and Dagan, I. (2021). Revisiting few-shot relation classification: Evaluation data and classification schemes. *Transactions of the Association for Computational Linguistics*, 9:691–706.

Sang, E. F. and De Meulder, F. (2003). Introduction to the conll-2003 shared task: Language-independent named entity recognition. *arXiv preprint cs/0306050*.

Schweter, S. and Akbik, A. (2020). Flert: Document-level features for named entity recognition. *arXiv preprint arXiv:2011.06993*.

Shen, Y., Ma, X., Tan, Z., Zhang, S., Wang, W., and Lu, W. (2021). Locate and label: A two-stage identifier for nested named entity recognition. *arXiv preprint arXiv:2105.06804*.

Stoica, G., Platanios, E. A., and Póczos, B. (2021). Re-tacred: Addressing shortcomings of the tacred dataset. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 13843–13850.

Sui, D., Chen, Y., Liu, K., Zhao, J., Zeng, X., and Liu, S. (2020). Joint entity and relation extraction with set prediction networks. arxiv 2020. *arXiv preprint arXiv:2011.01675*.

Sui, D., Zeng, X., Chen, Y., Liu, K., and Zhao, J. (2023). Joint entity and relation extraction with set prediction networks. *IEEE Transactions on Neural Networks and Learning Systems*.

Sun, Q., Liu, Y., Chua, T.-S., and Schiele, B. (2019). Meta-transfer learning for few-shot learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 403–412.

Tan, Z., Shen, Y., Zhang, S., Lu, W., and Zhuang, Y. (2021). A sequence-to-set network for nested named entity recognition. *arXiv preprint arXiv:2105.08901*.

Tang, W., Xu, B., Zhao, Y., Mao, Z., Liu, Y., Liao, Y., and Xie, H. (2022). Unirel: Unified representation and interaction for joint relational triple extraction. *arXiv preprint arXiv:2211.09039*.

Tong, M., Wang, S., Xu, B., Cao, Y., Liu, M., Hou, L., and Li, J. (2021). Learning from miscellaneous other-class words for few-shot named entity recognition. *arXiv preprint arXiv:2106.15167*.

Wang, C., Liu, X., Chen, Z., Hong, H., Tang, J., and Song, D. (2022). Deepstruct: Pretraining of language models for structure prediction. *arXiv preprint arXiv:2205.10475*.

Wang, J., Shou, L., Chen, K., and Chen, G. (2020a). Pyramid: A layered model for nested named entity recognition. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5918–5928.

Wang, X., Jiang, Y., Bach, N., Wang, T., Huang, Z., Huang, F., and Tu, K. (2020b). Automated concatenation of embeddings for structured prediction. *arXiv preprint arXiv:2010.05006*.

Wang, X., Jiang, Y., Bach, N., Wang, T., Huang, Z., Huang, F., and Tu, K. (2021). Improving named entity recognition by external context retrieving and cooperative learning. *arXiv preprint arXiv:2105.03654*.

Weischedel, R., Palmer, M., Marcus, M., Hovy, E., Pradhan, S., Ramshaw, L., Xue, N., Taylor, A., Kaufman, J., Franchini, M., et al. (2013). Ontonotes release 5.0 ldc2013t19. *Linguistic Data Consortium, Philadelphia, PA*, 23.

Xie, Y., Xu, H., Li, J., Yang, C., and Gao, K. (2020). Heterogeneous graph neural networks for noisy few-shot relation classification. *Knowledge-Based Systems*, 194:105548.

Xin, Y., Hart, E., Mahajan, V., and Ruvini, J.-D. (2018). Learning better internal structure of words for sequence labeling. *arXiv preprint arXiv:1810.12443*.

Yadav, V. and Bethard, S. (2019). A survey on recent advances in named entity recognition from deep learning models. *arXiv preprint arXiv:1910.11470*.

Yamada, I., Asai, A., Shindo, H., Takeda, H., and Matsumoto, Y. (2020). Luke: Deep contextualized entity representations with entity-aware self-attention. *arXiv preprint arXiv:2010.01057*.

Yan, H., Deng, B., Li, X., and Qiu, X. (2019). Tener: adapting transformer encoder for named entity recognition. *arXiv preprint arXiv:1911.04474*.

Yang, H., Zhang, Q., and Kwon, H.-C. (2024). Pner: Applying the pipeline method to resolve nested issues in named entity recognition. *Applied Sciences*, 14(5):1717.

Yang, K., Zheng, N., Dai, X., He, L., Huang, S., and Chen, J. (2020). Enhance prototypical network with text descriptions for few-shot relation classification. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 2273–2276.

Yang, Y. and Katiyar, A. (2020). Simple and effective few-shot named entity recognition with structured nearest neighbor learning. *arXiv preprint arXiv:2010.02405*.

Yao, Y., Ye, D., Li, P., Han, X., Lin, Y., Liu, Z., Liu, Z., Huang, L., Zhou, J., and Sun, M. (2019). Docred: A large-scale document-level relation extraction dataset. *arXiv preprint arXiv:1906.06127*.

Ye, D., Lin, Y., Li, P., and Sun, M. (2021). Packed levitated marker for entity and relation extraction. *arXiv preprint arXiv:2109.06067*.

Ye, Z.-X. and Ling, Z.-H. (2019). Multi-level matching and aggregation network for few-shot relation classification. *arXiv preprint arXiv:1906.06678*.

Yu, J., Bohnet, B., and Poesio, M. (2020). Named entity recognition as dependency parsing. *arXiv preprint arXiv:2005.07150*.

Zeng, X., Zeng, D., He, S., Liu, K., and Zhao, J. (2018). Extracting relational facts by an end-to-end neural model with copy mechanism. In *ACL*, pages 506–514.

Zhang, Y., Zhong, V., Chen, D., Angeli, G., and Manning, C. D. (2017). Position-aware attention and supervised data improve slot filling. In *Conference on Empirical Methods in Natural Language Processing*.

Zheng, C., Cai, Y., Xu, J., Leung, H., and Xu, G. (2019). A boundary-aware neural model for nested named entity recognition. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics.

Zhong, Z. and Chen, D. (2020). A frustratingly easy approach for entity and relation extraction. *arXiv preprint arXiv:2010.12812*.

Zou, X. (2020). A survey on application of knowledge graph. In *Journal of Physics: Conference Series*, volume 1487, page 012016. IOP Publishing.