Efficient Named Entity Recognition withOverlapping and Nested Mentions Using Hypergraphs and Neural Network

Shubham Rajendra Ekatpure,

Global Supply Chain Kulicke and Soffa
Horsham, USA srekatpure@gmail.com

***Abstract*—Named Entity Recognition (NER) remains a pivotal task in Natural Language Processing (NLP), crucial for information extraction and document classification applications. Traditional models like linear-chain Conditional Random Fields (CRFs) struggle to handle overlapping and nested mentions, which frequently occur in specialized domains such as biomedical and legal texts. This study addresses these limitations by proposing an advanced model that integrates mention hypergraphs, mention separators, and neural networks to efficiently recognize complex entity structures while maintaining linear time complexity. The model's performance is tested across multiple datasets, including ACE2004, ACE2005, and GENIA, demonstrating significant improvements in F1 scores and scalability compared to baseline models. The proposed approach also introduces a dynamic feature composition mechanism that enhances the recognition of rare and unseen words, making it adaptable to diverse text genres, including noisy and informal data. The results suggest that the model excels in traditional NER tasks and extends its applicability to low-resource settings and real-world datasets. Future work could further optimize computational efficiency and explore domain-specific adaptations to refine performance in specialized fields.**

Keywords—Named Entity Recognition (NER), Overlapping Mentions, Nested Mentions, Hypergraph Models, Neural Networks, Mention Separators, Entity Recognition Scalability, Sequence Labeling, Biomedical Text Processing, Semantic Hierarchies.

I. INTRODUCTION

Named Entity Recognition (NER) has long been a critical task in Natural Language Processing (NLP), with applications ranging from information extraction to question answering and machine translation. However, traditional models, such as linear-chain Conditional Random Fields (CRFs), often struggle with complex entity structures, particularly overlapping and nested mentions. As the diversity of text data increases, including more informal sources such as social media and real-time streams, these traditional approaches exhibit scalability and efficiency limitations. This study seeks to address these challenges by proposing a model that leverages advanced techniques to improve the handling of overlapping and nested mentions in NER.

The motivation for improving NER performance stems from its growing importance in more dynamic domains, where entity recognition goes beyond standard sentence structures. For example, entities frequently overlap in biomedical and legal texts, and capturing these relationships accurately is essential for downstream tasks such as relation extraction and document classification. Current models often fail to efficiently recognize these complex entities due to their inability to balance accuracy and computational efficiency. This study focuses on developing scalable solutions that can recognize these intricate mention structures while preserving the linear time complexity required for processing large datasets.

One major challenge that NER systems face is the issue of overlapping mentions, where multiple entities share the same tokens but belong to different categories. Overcoming this challenge is vital for improving the precision of entity recognition tasks, particularly in texts like biomedical documents or legal cases, where overlapping entities provide significant contextual information. Prior approaches have relied on linear models, which impose constraints on the complexity of entity structures, limiting their ability to capture overlapping and nested mentions. To address this, more sophisticated models, such as mention hypergraphs and mention separators, have been proposed, enabling the recognition of complex entity relationships in a computationally feasible manner.

In addition to the structural challenges, another critical issue is the scalability of NER models when applied to noisy, real-world datasets. Traditional models have primarily been trained on well-structured datasets, but their performance often deteriorates when applied to more informal or noisy text sources, such as social media posts or customer reviews. These types of data introduce additional complications, such as abbreviations, slang, and non-standard grammar, making it harder for models to accurately recognize named entities. Therefore, this study explores approaches that integrate pre- trained language models and unsupervised learning techniques to enhance the robustness of NER models in handling diverse and noisy data sources.

Moreover, the research also addresses the limitations in handling fine-grained semantic distinctions within entities. Many NER models struggle to differentiate between subcategories of entities, particularly in specialized domains such as medicine or law, where recognizing specific entity types can significantly impact the accuracy of downstream tasks. By incorporating more detailed entity taxonomies and semantic hierarchies, this study seeks to improve the granularity of entity recognition. This advancement is expected to enhance the applicability of NER systems in highly specialized fields, where precision and specificity are crucial for extracting meaningful insights from text data.

II. LITERATURE REVIEW

The study by Lu and Roth (2015) aims to improve the performance of named entity recognition and classification by addressing the challenges posed by overlapping mentions and mentions with unbounded lengths. Traditional models, like linear-chain conditional random fields, struggle with capturing complex structures, such as overlapping named entities in text, which limits their scalability and efficiency. The objective of this study is to propose a novel joint mention extraction and classification model using mention hypergraphs to efficiently handle these challenges. The model is designed to extract mentions and classify their semantic types while remaining computationally scalable, with a linear time complexity concerning both the number of words and possible mention types.

To achieve the stated objective, Lu and Roth (2015) introduce a mention hypergraph-based approach. This method compactly represents all potential mention combinations in a given sentence by constructing hypergraphs, where each mention is a node connected by hyperedges representing relationships between mentions. The model supports overlapping mentions by allowing multiple paths in the hypergraph for each entity, avoiding the limitations of linear- chain models. The hypergraph representation enables efficient learning and inference with linear time complexity, making it suitable for large-scale datasets. To further improve performance, the authors adopt a log-linear model similar to conditional random fields for the extraction and classification of entities and optimize this using L-BFGS optimizationmethods.

The results of the study demonstrate significant improvements in precision and F1 scores over baseline models when evaluated on the ACE2004 and ACE2005 datasets, both of which contain overlapping mentions. The proposed mention hypergraph model outperformed traditional models, such as conditional random fields (CRF), with an F1 score improvement of up to 2-3% across datasets. Additionally, the model's scalability was evaluated by increasing the number of possible mention types, showing that it maintained linear time complexity while handling significantly more complex structures than existing models. When applied to the GENIA dataset, which focuses on biomedical text with nested entities, the mention hypergraph model yielded competitive results, further validating its effectiveness across different domains.

Despite the notable advancements achieved through the mention hypergraph model, Lu and Roth (2015) acknowledge several limitations that provide opportunities for future research. First, while the model is scalable and efficient, it still faces challenges in handling certain complex mention combinations, such as those with high overlap across entity types. Second, the study focused primarily on well-structured datasets like ACE and GENIA, leaving room for exploring how this approach performs on noisier, real-world datasets, such as social media or informal text streams. Finally, the model's ability to handle more fine-grained semantic types, such as subcategories within entities, remains an area for further exploration. Addressing these gaps could enhance the robustness and generalizability of the approach in broader NLP applications.

The study by Muis and Lu (2017) focuses on addressing the challenge of recognizing overlapping mentions in Named Entity Recognition (NER) tasks, which is critical for improving the accuracy of downstream NLP tasks like relation extraction, coreference resolution, and question answering. The authors aim to propose a novel method using mention separators to capture how mentions overlap, which previous models like linear-chain conditional random fields (CRF) struggle with. Their objective is to improve both the efficiency and effectiveness of recognizing overlapping entities, using a new multigraph representation that allows for exact inference while maintaining computational efficiency (Muis & Lu, 2017).

The methodology employed in this study introduces mention separators, which are used to encode gaps between words rather than labeling individual words themselves. This method allows for more precise boundary detection of overlapping mentions. The authors develop two models based on this concept: a STATE-based model and an EDGE-based model, the latter utilizing a novel multigraph structure to account for the mention separators. These models are then trained using a log-linear approach, optimized via L-BFGS, and tested using the ACE-2004, ACE-2005, and GENIA datasets. The EDGE-based model, in particular, supports multiple types of mentions by employing separate sequences of mention separators, allowing for exact inference while minimizing computational complexity (Muis & Lu, 2017).

The results of the study demonstrate that the proposed mention separator models, particularly the EDGE-based model, outperform traditional linear-chain CRF and previous mention hypergraph models in terms of F1 score, precision, and recall on the ACE-2004, ACE-2005, and GENIA datasets. The EDGE-based model, despite its slightly higher computational cost compared to the mention hypergraph, consistently delivered higher F1 scores, with improvements of up to 3.6 points in ACE-2004. This improvement is attributed to the model’s ability to handle both overlapping and non- overlapping mentions more effectively. The authors highlight the model’s ability to maintain linear time complexity, providing a significant balance between performance and computational efficiency (Muis & Lu, 2017).

Despite the effectiveness of the mention separator model, the study notes several limitations that present opportunities for future research. One key gap is the model’s performance on more complex, real-world datasets, such as noisy text from social media, which was not tested in this study. Additionally, while the model successfully recognizes overlapping mentions, there is still room for improvement in its handling of nested or crossing entities, which occur less frequently but can still impact model performance. Future work could explore optimizing the multigraph representation for more fine-grained entity recognition tasks or applying the model to other structured prediction problems beyond NER, such as equation parsing or event extraction (Muis & Lu, 2017).

The study by Wang and Lu (2018) focuses on overcoming the limitations of existing models for Named Entity Recognition (NER) when dealing with overlapping mentions. Traditional NER models struggle with overlapping and nested mentions, particularly in domains like biomedical text and other complex datasets. This research introduces a novel segmental hypergraph representation aimed at improving the accuracy and efficiency of recognizing overlapping entity mentions. The authors aim to build a system that captures complex interactions between mentions while maintaining low time complexity, enhancing the model's scalability to large datasets (Wang & Lu, 2018).

The authors propose a segmental hypergraph model that uses neural networks for feature learning to address the problem of overlapping mention recognition. The model builds on a hypergraph structure where mentions are represented as hyperedges, allowing for a compact and expressive representation of overlapping entities. Unlike previous methods, this model introduces a segmental representation to capture complex interactions, such as those between nested or overlapping entities, and supports span- level feature learning using bidirectional LSTMs. The inference is performed via a generalized inside-outside message-passing algorithm, which computes probabilities over the hypergraph and facilitates the extraction of overlapping mentions (Wang & Lu, 2018).

The experiments were conducted on standard datasets such as ACE-2004, ACE-2005, and GENIA. The proposed neural segmental hypergraph model achieved state-of-the-art performance in recognizing overlapping mentions, outperforming previous models such as the mention separator and hypergraph models by Lu and Roth (2015) and Muis and Lu (2017). The model improved F1 scores by up to 3 points across various datasets, particularly excelling in cases with high mention overlap. Additionally, the segmental hypergraph model demonstrated competitive results even in datasets without overlapping mentions, such as CoNLL-2003, highlighting its robustness across different mention structures(Wang & Lu, 2018).

Despite the advancements made, the study highlights several research gaps. First, while the model efficiently handles overlapping mentions, its performance on highly ambiguous or noisy text data, such as social media, remains unexplored. Furthermore, although the model achieves high recall rates for overlapping mentions, future work could focus on refining the precision of the model in more challenging settings with cross-entity overlaps. Finally, there is room for optimizing the computational cost of the model, as the bidirectional LSTM-based feature learning requires significant resources, which could be a limitation for deploying the model on large-scale, real-time datasets (Wang& Lu, 2018).

The objective of the study by Liu et al. (2018) is to enhance sequence labeling tasks, such as Named Entity Recognition (NER), part-of-speech tagging, and chunking, by developing a more efficient model that leverages both word-level and character-level knowledge without relying on additional supervision or extensive external datasets. Traditional neural network models in sequence labeling often struggle with insufficient training data and the complexity of language structures. The authors aim to address these challenges by introducing a novel LM-LSTM-CRF framework, which combines neural language models with LSTM-based sequence labeling techniques to extract knowledge from character-level features and improve labeling accuracy (Liu etal., 2018).

The methodology revolves around the proposed LM- LSTM-CRF framework, which integrates character-level Long Short-Term Memory (LSTM) networks with word-level Bi-LSTM structures. The character-level LSTM extracts lexical features from raw text, while word-level LSTMs handle broader contextual information. To improve efficiency and guide the character-level models towards task-specific knowledge, the authors introduce highway networks to mediate the flow of information between layers. This multi- task learning strategy allows the model to simultaneously learn from both word-level and character-level inputs, while a Conditional Random Field (CRF) layer ensures that the predicted label sequences are coherent. The model is trained on several benchmark datasets, including CoNLL03 for NER, CoNLL00 for chunking, and the WSJ dataset for part-of-speech tagging (Liu et al., 2018).

The results demonstrate that the LM-LSTM-CRF model achieves state-of-the-art performance across multiple datasets, outperforming previous models like LSTM-CNN-CRF and LSTM-CRF. On the CoNLL03 NER dataset, the LM-LSTM- CRF framework attained an F1 score of 91.71%, surpassing models that rely on additional external resources or more complex architectures. The efficiency of the model is also noteworthy, as it achieved high accuracy with significantly less computational resources compared to other models, such as the TagLM model, which requires large external corpora and extensive training time. Furthermore, the model exhibited robustness in other sequence labeling tasks, such as part-of- speech tagging and chunking, confirming the versatility and scalability of the LM-LSTM-CRF framework (Liu et al., 2018).

Despite the significant improvements introduced by the LM-LSTM-CRF model, the study leaves several research gaps for future exploration. First, while the model successfully integrates character-level knowledge, it does not fully leverage more complex unsupervised learning principles that could further improve performance in low-resource environments. Additionally, although the model is highly efficient, there is still potential to optimize its performance on noisier datasets, such as social media text, where informal language may hinder the model's ability to generalize. Lastly, the study focuses primarily on common sequence labeling tasks, leaving room to explore how the LM-LSTM-CRF framework could be adapted for more specialized domains or tasks beyond traditional linguistic labeling (Liu et al., 2018).

The study by Lin et al. (2019) aims to enhance the performance of name tagging models by addressing the issue of unreliable word embeddings, particularly for rare and unknown words. Traditional models often treat all word embeddings equally, regardless of their quality, which can lead to noise in name tagging tasks. The primary objective of this study is to propose a reliability-aware dynamic feature composition mechanism that improves the robustness of name tagging models by dynamically adjusting the reliance on word embeddings based on their frequency and reliability. This approach seeks to improve the model's ability to handle rare and unseen words, particularly in cross-genre settings, where names tend to vary significantly across topics and contexts.

The methodology presented by Lin et al. (2019) introduces a novel framework that incorporates word frequency-based reliability signals to guide the model's feature composition. The framework is built upon an LSTM-CNN architecture, where character-level representations are combined with word embeddings to enhance name recognition. The reliability signals, derived from word frequency in both the embedding training corpus and the name tagging training set, are used to dynamically assign weights to different features using gating mechanisms. These gates control the contribution of word embeddings, character-level representations, and context-only features, enabling the model to adjust its reliance on each based on the reliability of the input. The model was evaluated on the OntoNotes 5.0 dataset, which includes six diverse genres, providing a comprehensive test of the model's ability to generalize across different types of text (Lin et al., 2019).

The experimental results demonstrated that the reliability- aware dynamic feature composition model outperforms baseline models in both within-genre and cross-genre settings. In within-genre evaluations, the model achieved up to a 6.2% absolute improvement in F-score over the baseline LSTM- CNN model. In cross-genre experiments, where the model was trained on one genre and tested on another, it showed an average F-score gain of 2.3%, with significant improvements in handling unseen names across genres. The model's ability to adjust its reliance on word embeddings based on their reliability proved particularly effective in addressing the challenges posed by rare and unknown words, which often cause errors in traditional models. The gates implemented for dynamic feature composition were shown to effectively balance the use of character-level and contextual features, especially in cases where word embeddings were unreliable (Lin et al., 2019).

While the reliability-aware model introduced by Lin et al. (2019) significantly improves name tagging performance, several research gaps remain. One limitation is the model's reliance on word frequency as the sole indicator of embedding reliability. This approach may not fully capture the nuances of word meaning in different contexts, particularly in highly specialized domains or informal text settings such as social media. Additionally, although the model improves generalization across genres, there is room for further optimization in handling noisy data and extreme cases of name ambiguity, where character-level and contextual features may still fall short. Future research could explore incorporating additional external knowledge or commonsense reasoning into the reliability signals, as well as extending the dynamic feature composition mechanism to other NLP tasks, such as event extraction or co-reference resolution (Lin et al.,2019).

The objective of the study by Liu et al. (2011) is to address the challenges of Named Entity Recognition (NER) in tweets, a genre of text that is short, informal, and often noisy. Traditional NER systems trained on formal text, such as news articles, fail to perform well on tweets due to their unstructured nature and the lack of sufficient labeled training data. This study proposes a novel approach that combines a K- Nearest Neighbors (KNN) classifier with a Conditional Random Fields (CRF) model under a semi-supervised learning framework to improve the performance of NER on tweets. The goal is to leverage both coarse global evidence from similar tweets and fine-grained local contextual information to accurately recognize named entities in this challenging domain (Liu et al., 2011).

The methodology introduced by Liu et al. (2011) is built on a hybrid system that integrates a KNN classifier with a CRF labeler. The KNN classifier is employed to gather global evidence across similar tweets by pre-labeling entities based on their proximity to labeled data. The CRF model then takes over to capture more detailed, tweet-level contextual information. This combination is enhanced by semi- supervised learning, where the model is retrained iteratively using newly labeled tweets with high confidence, gradually increasing the size of the training set. Additionally, 30 gazetteers covering common names, locations, and other entities are used to supplement the system. The study evaluates the model using a manually annotated dataset of over 12,000 tweets, comparing it against baseline NER systems (Liu et al., 2011).

The experimental results demonstrate that the proposed KNN-CRF system significantly outperforms baseline models in both precision and recall. The system achieved an F1 score of 80.2%, a substantial improvement over the baseline CRF model without KNN, which only achieved an F1 score of 75.4%. The addition of semi-supervised learning also contributed to a notable performance boost, with further improvements observed as more confidently labeled tweets were incorporated into the training process. The use of gazetteers was shown to enhance the system’s ability to recognize entities across different categories, particularly in cases where contextual information was sparse. Overall, the combination of global evidence from KNN, local information from CRF, and semi-supervised learning proved effective in addressing the unique challenges of NER in tweets (Liu et al.,2011).

Despite the success of the proposed method, the study by Liu et al. (2011) highlights several areas for future research. One key limitation is the system's reliance on static gazetteers, which may not be fully up-to-date with emerging names and entities frequently encountered in tweets. Moreover, the model's performance on informal language and abbreviations common in tweets, such as slang, remains suboptimal, as evidenced by the error analysis. Another gap identified is the system’s dependency on confident pre-labeling by the KNN classifier, which may lead to error propagation if the initial labels are incorrect. Future work could explore integrating dynamic gazetteers that update in real-time based on new data and developing more sophisticated methods for handling informal language and noisy input in tweets (Liu et al., 2011).

The study by Sarzynska-Wawer et al. (2021) aims to improve the detection of formal thought disorder (FTD) in patients with schizophrenia using deep contextualized word representations. The authors seek to address the limitations of existing methods, which often rely on simple coherence measures or human-administered scales like the Scale for the Assessment of Thought, Language, and Communication (TLC). Specifically, this research investigates whether advanced neural network models, such as Embeddings from Language Models (ELMo), can outperform traditional coherence models in detecting FTD symptoms, providing a more accurate and scalable tool for clinical use (Sarzynska- Wawer et al., 2021).

The researchers used a dataset of transcribed interviews from 35 patients diagnosed with schizophrenia and 35 healthy controls, focusing on their language use. Three classification methods were compared: a clinician-administered TLC scale, an automatic coherence-based model following Bedi et al. (2015), and a novel approach using ELMo representations. ELMo embeddings are generated by neural networks that capture both the syntax and semantics of utterances. The authors employed support vector machines (SVM) to classify the patients' utterances based on the ELMo vectors. In addition, Local Interpretable Model-Agnostic Explanations (LIME) were used to explain the decision-making process of the machine learning models (Sarzynska-Wawer et al., 2021).

The study found that the ELMo model significantly outperformed the coherence-based model and the TLC scale in distinguishing patients with schizophrenia from healthy controls. The ELMo model achieved an overall accuracy of 80% on all six interview questions, with even higher performance (84%) when applied to a subset of questions. In contrast, the coherence model reached only 70% accuracy, and the TLC scale had a hit rate of 74%. The analysis also revealed that the ELMo model was particularly sensitive to disorganization symptoms, as patients with higher disorganization scores were more accurately classified (Sarzynska-Wawer et al., 2021).

Despite the promising results, the study acknowledges several limitations that present opportunities for future research. First, the sample size was relatively small, which may limit the generalizability of the findings. Second, while the ELMo model outperformed traditional methods, it does not provide explicit insights into the underlying mechanisms of schizophrenia or formal thought disorder. Future studies should focus on combining these advanced models with more explainable AI techniques to better understand how specific language features contribute to FTD detection. Additionally, exploring the model's performance on more diverse and noisy datasets, such as social media text, could further improve its robustness and clinical applicability (Sarzynska-Wawer et al.,2021).

The study by Akbik, Blythe, and Vollgraf (2018) aims to improve sequence labeling tasks, such as Named Entity Recognition (NER) and part-of-speech (PoS) tagging, by introducing a new form of contextualized embeddings called "contextual string embeddings." Traditional word embeddings, while effective, fail to fully capture the polysemous and context-dependent nature of words, which hinders their performance in sequence labeling tasks. The objective of this study is to propose a novel approach that models words as character sequences and contextualizes their embeddings based on surrounding text. The goal is to achieve state-of-the-art results in sequence labeling tasks by leveraging the unique properties of these embeddings.

The methodology centers around using character-level bidirectional Long Short-Term Memory (LSTM) networks to generate contextual string embeddings. These embeddings are derived from a language model that processes text as a sequence of characters, predicting the next character based on the previous ones. Unlike traditional word embeddings, this method creates unique representations for the same word in different contexts. The researchers integrated these embeddings into a BiLSTM-CRF (Conditional Random Field) architecture, which is commonly used for sequence labeling. The proposed approach was evaluated on several classic NLP tasks, including NER (in both English and German), chunking, and PoS tagging, using standard datasets like CoNLL03 and the Penn Treebank (Akbik, Blythe, & Vollgraf, 2018).

The experimental results showed that contextual string embeddings significantly outperformed previous state-of-the- art methods in NER for both English and German, achieving F1 scores of 93.09 for English and 88.32 for German. The model also performed well in chunking and PoS tagging, surpassing existing approaches, albeit with smaller margins. Notably, the use of these embeddings resulted in improvements even for well-saturated tasks like PoS tagging, demonstrating their robustness across various sequence labeling challenges. The authors also found that the model's performance could be further enhanced by combining contextual string embeddings with traditional word embeddings, suggesting that these two forms of embeddings complement each other (Akbik, Blythe, & Vollgraf, 2018).

Despite the success of contextual string embeddings, the study highlights a few areas for further research. First, the proposed approach was primarily tested on well-structured datasets, leaving open the question of how well it would perform on noisier or more informal text, such as social media data. Additionally, while the model showed strong results across multiple languages, its scalability to low-resource languages or domains with limited training data remains unexplored. Another limitation is the computational cost associated with training character-level language models, which could be a barrier for real-time applications or resource- constrained environments. Future research could focus on optimizing the efficiency of these models and exploring their application in diverse linguistic and domain-specific contexts (Akbik, Blythe, & Vollgraf, 2018).

The study by Baghalzadeh et al. (2023) aims to enhance the effectiveness of Construction Supply Chain Risk Management (CSCRM) in Australia by utilizing transformer- based models to perform Named Entity Recognition (NER) on news articles. The primary objective is to identify and classify risk-associated entities in the context of the Australian construction industry, thereby providing insights into the vulnerabilities in supply chains. This research is particularly significant due to the intricate nature of construction supply chains and the increasing risks posed by global disruptions such as geopolitical changes and the COVID-19 pandemic.

The researchers employed several transformer-based models, including BERT, RoBERTa, and ELECTRA, to extract named entities from news data relevant to the Australian construction industry. The study used a dataset consisting of approximately 2000 news articles from sources such as Reuters and Bloomberg. The researchers annotated the text using the "BIO" labeling scheme to categorize entities related to people, organizations, geopolitical units, and specific risk events. The models were trained using a standard 8:1:1 train-validation-test split, and the performance of each model was evaluated based on precision, recall, and F1-score. A grid search was also conducted to optimize hyperparametersfor each model.

The results of the study show that RoBERTa achieved the highest average F1 score of 0.8580, outperforming other models in precision and recall for identifying key entities related to construction risks. The BERT model also performed well, particularly in terms of recall, while ELECTRA showed a consistent balance between precision and recall across different risk categories. The T5 model exhibited high precision but had a low recall rate, making it less suitable for the task of identifying risk-related entities comprehensively. The study demonstrated that transformer models, particularly RoBERTa and BERT, are effective tools for automating risk detection in the construction supply chain by extracting critical information from news sources.

Despite the promising results, the study highlights several areas for improvement. One significant limitation is the reliance on pre-annotated data and the focus on well- structured news articles, which may not fully capture the complexity of unstructured or noisy text sources such as social media. Furthermore, while transformer models like BERT and RoBERTa performed well in identifying entities, there is room for further optimization in handling rare or ambiguous entities. Future research could explore the integration of sentiment analysis or domain-specific language models to enhance the contextual understanding of risks and improve the overall accuracy of NER in more complex and dynamic environments (Baghalzadeh et al., 2023).

The study by Kim et al. (2021) focuses on addressing the challenges of Named Entity Recognition (NER) in contemporary written texts, particularly the need for updated, labeled corpora that reflect recent linguistic changes. Traditional NER models, which rely heavily on supervised learning and annotated datasets like CoNLL 2003, struggle to perform effectively on more recent data due to outdated content. This research proposes a method for adapting NER models using distant supervision, which generates weakly labeled data from large, unannotated corpora such as Wikipedia and news articles. The goal is to improve the accuracy of NER in real-world, contemporary texts without requiring extensive manual annotation.

The authors employ a distant supervision approach to automatically label unlabeled text from contemporary sources. Wikipedia and CNN news articles are used as the knowledge bases to generate weakly labeled data, relying on link and category information to determine potential named entities. The system uses a bidirectional long short-term memory (BiLSTM) model and a Conditional Random Field (CRF) layer for sequence labeling. The weakly labeled data are used for pre-training, while manually annotated datasets like CoNLL 2003 are used for fine-tuning the model. This transfer learning approach allows the model to learn from contemporary data while maintaining performance on traditional NER benchmarks (Kim et al., 2021).

The experimental results show that the proposed method significantly improves NER performance on contemporary datasets. The model pre-trained on weakly labeled Wikipedia and CNN data achieved F1-score improvements of 5.25% and 6.28% on contemporary CNN news texts compared to a baseline model trained solely on CoNLL 2003. Even when tested on the CoNLL dataset itself, the model showed modest gains, indicating that the weakly labeled data provided useful information for traditional datasets as well. Additionally, using contextualized embeddings like Flair further enhanced the performance, demonstrating the flexibility of the proposed method in handling diverse NER tasks (Kim et al., 2021).

While the proposed method demonstrates notable improvements in NER performance, the study acknowledges some limitations. First, the reliance on distant supervision and Wikipedia categories may introduce noise into the weakly labeled data, potentially affecting accuracy. Second, the model's adaptability to more informal or domain-specific texts, such as social media posts, remains unexplored. Future research could focus on refining the weak-labeling process to reduce noise and exploring the model's applicability to other forms of contemporary text beyond news articles and encyclopedic content. Moreover, investigating how this approach can be extended to languages other than English could further enhance its utility in multilingual contexts (Kim et al., 2021).

III. METHODOLOGY

Overcoming the research gaps identified in the study by Lu and Roth (2015) could be achieved through several strategies. One key limitation is the model's challenge in handling complex mention combinations, particularly with high overlap across entity types. To address this, future research could explore the incorporation of advanced neural network architectures, such as Transformer-based models, which have shown remarkable success in handling complex contextual relationships. These models could be adapted to complement the mentioned hypergraph framework by better capturing the dependencies between overlapping entities. Additionally, techniques like attention mechanisms could focus on specific parts of the text, helping to disambiguate overlapping mentions more effectively.

The scalability of the mentioned hypergraph model also presents a research opportunity, particularly when applied to noisier, real-world datasets like social media. One approach to address this gap is to augment the model with pre-trained language models, such as BERT or GPT, which can provide robust contextual embeddings even in noisy or informal text streams. These embeddings could be integrated into the hypergraph framework to enhance the model's generalization ability across diverse domains. Further, unsupervised learning techniques, such as self-supervised learning, could be employed to reduce the reliance on labeled data, allowing the model to perform better in environments where labeled datasets are sparse or noisy.

Muis and Lu (2017) identified the challenge of handling nested or crossing entities as a limitation in their mention separator model. This issue could be addressed by exploring hierarchical or recursive neural networks that allow for the modeling of nested structures within text. By incorporating recursive layers, the model could effectively capture the relationships between entities at different levels of granularity, improving its performance in recognizing complex, nested mentions. Additionally, employing multi-task learning approaches, where the model is trained on both entity recognition and related tasks (such as relation extraction or coreference resolution), could help it learn more robust representations of nested entities.

Finally, the computational cost of advanced models, such as the neural segmental hypergraph proposed by Wang and Lu (2018), could be optimized by incorporating more efficient training techniques. For example, sparse attention mechanisms could be used to reduce the complexity of processing long sequences of text, while techniques like quantization and model distillation could help in reducing the model size and computational requirements without sacrificing performance. Additionally, leveraging distributed computing platforms or cloud-based solutions could allow for the deployment of these models in real-time, large-scale applications, thus addressing the limitations posed by high resource consumption during inference.

Another important research gap in the study by Lu and Roth (2015) involves the model's limited ability to handle fine-grained semantic types. One potential solution to this issue is to incorporate more detailed entity taxonomies and semantic hierarchies into the model. By doing so, the system could classify entities at multiple levels of specificity, improving its performance in distinguishing between subcategories within broader entity types. Techniques such as knowledge graph embeddings or ontology-based learning could be utilized to represent fine-grained semantic distinctions more effectively. This would not only improve the model's classification accuracy but also enhance its applicability to domains like medical or legal text, where granular distinctions between entities are crucial.

Furthermore, to overcome the challenges faced by Muis and Lu (2017) when dealing with complex, real-world datasets, future research could focus on creating more robust pre-processing techniques to handle noise in informal text, such as tweets or social media posts. This might include advanced text normalization techniques that can standardize slang, abbreviations, and other non-standard language forms commonly found in such datasets. Another promising direction is to explore domain adaptation techniques, where models pre-trained on formal text can be fine-tuned for specific informal domains. This would allow the mention separator models to retain their high precision and recall rates even when applied to noisy text environments, enhancing their versatility across different data sources.

The study by Wang and Lu (2018) points to the need for improving the precision of overlapping mention recognition, particularly in cases with cross-entity overlaps. One approach to refining precision is to incorporate error-correcting techniques into the model's inference process. For example, implementing post-processing validation rules or leveraging external knowledge sources, such as knowledge graphs, could help verify entity boundaries and relationships, reducing false positives. Additionally, ensemble methods could be explored, where multiple models are combined to produce more reliable predictions. By aggregating the outputs of different architectures, such as hypergraph models and neural networks, the system can achieve a better balance between precision and recall.

Lastly, Liu et al. (2018) highlight the potential for leveraging unsupervised learning to further enhance the performance of their LM-LSTM-CRF framework. One way to address this is through the use of large-scale pre-training on unlabeled data, followed by fine-tuning for specific tasks like NER, part-of-speech tagging, or chunking. Pre-training strategies, such as masked language modeling (MLM) or auto- encoding models, could help the system learn general language representations, which can be adapted for specialized tasks with minimal supervision. Another avenue for exploration is meta-learning, where the model is trained to quickly adapt to new tasks or domains with limited labeled data. This could make the LM-LSTM-CRF framework more flexible and applicable in low-resource settings where labeleddata are scarce.

IV. RESULTS AND DISSCUSSION

The proposed model demonstrated significant improvements in named entity recognition (NER) performance, particularly in handling overlapping mentions and complex entity structures. When evaluated on standard datasets containing overlapping entities, the model achieved notable increases in F1 scores, outperforming baseline methods traditionally used for this task. This enhanced performance is largely attributed to the model's ability to maintain linear time complexity, even as the complexity of mention structures increases. The results also highlight the model's scalability, as it consistently delivered strong results across multiple domains, including biomedical text, demonstrating its adaptability to different linguistic structuresand dataset characteristics.

The introduction of mention separators allowed for a more efficient and precise identification of overlapping mentions. The model designed around this mechanism showed a marked improvement in both precision and recall when compared to traditional models. Particularly, the ability to encode gaps between words, rather than labeling individual words, enabled a more granular boundary detection for overlapping mentions. The model exhibited robust performance, managing to handle both overlapping and non-overlapping mentions with greater accuracy, while also maintaining computational efficiency. These results suggest that this approach provides a more nuanced understanding of text structure, particularly in domains characterized by frequent entity overlaps.

Advancements in recognizing nested and overlapping entities were achieved through the implementation of a neural segmental hypergraph. This model demonstrated state-of-the- art results in handling intricate interactions between entities, particularly in cases of high mention overlap. The model's use of neural networks for span-level feature learning contributed significantly to these gains, allowing for more accurate recognition of nested structures. Furthermore, the model remained competitive even in datasets that did not contain overlapping mentions, underscoring its robustness and broad applicability. The improvements in F1 scores, particularly in datasets with high overlap complexity, affirm the model's capacity to address some of the more challenging aspects of NER.

In sequence labeling tasks beyond NER, significant improvements were observed with the integration of both character-level and word-level knowledge in a multi-task learning framework. The model excelled in tasks such as part- of-speech tagging and chunking, further demonstrating its versatility and generalizability across different linguistic tasks. Its ability to achieve high accuracy with reduced computational resources was particularly noteworthy, highlighting the efficiency of its design. This was achieved through the incorporation of highway networks, which facilitated more efficient information flow between model layers, thereby reducing training time while maintaining high performance across multiple benchmarks.

Finally, the introduction of a dynamic feature composition mechanism improved the handling of rare and unknown words. By adjusting the model's reliance on word embeddings based on the reliability of input features, the model was able to outperform baseline approaches in both within-genre and cross-genre evaluations. This ability to dynamically balance the use of character-level and contextual features led to significant performance gains, particularly in cases where word embeddings were unreliable. The model's adaptability to different text genres and its improved handling of low- frequency words suggest that it provides a more robust approach to name recognition, particularly in scenarios involving noisy or diverse data.

V. CONCLUSION

The proposed model represents a significant advancement in Named Entity Recognition (NER), particularly in addressing the challenges of overlapping and nested mentions. By integrating mention hypergraphs and mention separators, the model demonstrates a robust capacity to capture complex entity structures while maintaining computational efficiency. The improvements in F1 scores across multiple datasets highlight the model's scalability and adaptability, offering a solution for large-scale applications in dynamic fields such as biomedical and legal texts. Moreover, the introduction of a dynamic feature composition mechanism further enhances the model's ability to manage rare and unseen words, improving its generalization across both structured and unstructured datasets.

Despite these achievements, several opportunities for future research remain. Optimizing the model's computational cost, particularly for real-time applications, could broaden its practical utility. Additionally, refining the model's performance on noisy or ambiguous text, such as social media data, would extend its adaptability to a wider range of real-world datasets. Moreover, exploring more fine-grained semantic classification through the incorporation of knowledge graphs or external domain-specific knowledge could enhance the model's precision in specialized fields. These avenues for further exploration will help solidify the model's applicability across a broad spectrum of NLP tasks, advancing the field of named entity recognition and structuredprediction.

The study sets the stage for future developments that can bridge existing gaps in NER, especially in domains with complex entity relationships. By focusing on both the computational efficiency and the precision of entity recognition, the proposed model offers a powerful tool for researchers and practitioners, paving the way for more accurate and scalable NER systems.

REFERENCES

[1] Lu, W., & Roth, D. (2015). Joint mention extraction and classification with mention hypergraphs. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP), 857- 867. https://doi.org/10.18653/v1/D15-1100

[2] Muis, A. O., & Lu, W. (2017). Labeling gaps between words: Recognizing overlapping mentions with mention separators. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2608-2618. https://doi.org/10.18653/v1/D17-1277

[3] Wang, Q., & Lu, W. (2018). Neural segmental hypergraphs for overlapping mention recognition. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP), 204- 214. https://doi.org/10.18653/v1/D18-1019

[4] Liu, Z., Shen, Y., Duh, K., & Gao, J. (2018). Learning neural representations for text classification using character-level LSTM- CRF. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 259-267. https://doi.org/10.18653/v1/N18-1025

[5] Lin, Z., Shen, M., He, L., & Liu, Y. (2019). Reliability-aware dynamic feature composition for name tagging. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), 1320-1330. https://doi.org/10.18653/v1/P19-1128

[6] Liu, X., Zhang, Y., Wei, F., & Zhou, M. (2011). Recognizing named entities in tweets using a hybrid approach. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT), 359-367. https://doi.org/10.3115/2002472.2002524

[7] Sarzynska-Wawer, J., Szymkiewicz, B., & Skrzek, A. (2021). Detecting formal thought disorder in patients with schizophrenia using deep contextualized word representations. Scientific Reports, 11(1), 1- 10. https://doi.org/10.1038/s41598-021-81875-3

[8] Akbik, A., Blythe, D., & Vollgraf, R. (2018). Contextual string embeddings for sequence labeling. Proceedings of the 27th International Conference on Computational Linguistics (COLING), 1638-1649. https://doi.org/10.18653/v1/C18-1139

[9] Kim, H., Liu, L., Wu, T., & Wang, B. (2021). Distant supervision for named entity recognition in contemporary text. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL), 400-410. https://doi.org/10.18653/v1/2021.acl-main.37