A Dual-Approach Framework for Relation Extraction: A Unified Hybrid Model and a BERT-based Method with Multi-Level Distillation and Mention-Span Enhancement

Michael Seiranian, Divy Solanki, Arden Mendonca, Moin Chalkoo

1 Introduction

Relation extraction (RE) is a fundamental task in natural language processing that aims to identify semantic relationships between entities in text. RE is crucial for applications such as knowledge graph construction, information retrieval, and question answering (Zhang, Qi and Manning, 2017). For this work, we have chosen the TACRED dataset—a large-scale, well-annotated corpus that has become a standard benchmark in RE research due to its rich set of relation labels and challenging examples (Zhang, Qi and Manning, 2017). In this paper, we describe and evaluate two distinct approaches to RE: a unified hybrid model that integrates a deep learning branch (based on a BiLSTM with attention), a symbolic rule-based component, and traditional SVM-style features; and a novel BERT-based method enhanced with Multi-Level Distillation (MLD) and Mention-Span Enhancement (MSE).

2 Dataset Description

The TACRED dataset (Zhang, Qi and Manning, 2017) is a widely used benchmark for relation extraction. It consists of over 106,000 examples annotated with a variety of semantic relations. Each example includes a sentence, two entities (with their respective types and positions), and a relation label. TACRED's diverse and challenging examples make it ideal for evaluating models that need to capture both fine-grained and coarse semantic distinctions.

3 Related Work

Early RE systems primarily relied on hand-crafted features and rule-based techniques (Hong, 2005; Rink and Harabagiu, 2010). The advent of deep learning brought significant improvements, with models such as BiLSTMs (Zhang et al., 2015; Valette, 2019) learning contextualized representations from raw text. More recently, transformer-based models like BERT have set new performance records (Devlin et al., 2019). Approaches such as knowledge distillation (Hinton et al., 2015) have been applied to improve model generalization, while auxiliary tasks for entity boundary detection have also been explored (Lample et al., 2016). The two methods presented in this paper build on these developments by fusing multiple paradigms and introducing novel training strategies to further enhance performance.

4 Methodology

4.1 Unified Hybrid Model (Method 1)

The unified hybrid model leverages three complementary components:

BiLSTM with Attention:
This deep learning branch processes token embeddings through a bidirectional LSTM combined with an attention mechanism. It captures contextual dependencies and outputs a fixed-length representation (Zhang et al., 2015).

Symbolic Rule-Based Extraction:
A set of regular expression patterns is used to extract lexical cues (e.g., "founded", "located in") that indicate specific relations. This component produces one-hot encoded vectors corresponding to the predicted relation based on these cues.

Traditional SVM-style Features:
TF–IDF representations are computed over concatenated segments of the input (subject, context between entities, and object) and then projected into a common space. This branch captures statistical properties typically exploited by support vector machines (Hong, 2005).

These representations are concatenated and fed into a feed-forward neural network for final relation classification. The design is motivated by the hypothesis that integrating diverse sources of information can overcome the limitations of individual approaches.

4.2 Novel BERT-based Method with MLD and MSE (Method 2)

This approach extends a standard BERT-based classifier with two novel components:

Multi-Level Distillation (MLD): Instead of relying solely on the final hidden layer, hidden states from multiple BERT layers are combined. A self-distillation loss encourages lower layers to approximate higher-level representations, thereby regularizing the model and enhancing semantic abstraction (Hinton et al., 2015).

Mention-Span Enhancement (MSE): An auxiliary prediction head is added to the model to predict the start and end offsets of subject and object entities. This task forces the model to learn more precise token-level signals, improving its ability to delineate entity boundaries (Lample et al., 2016).

These enhancements are jointly optimized with the primary relation classification objective, resulting in a model that is both more robust and capable of finer-grained entity localization.

5 Experimental Setup

Preprocessing:

Unified Hybrid Model:
Tokenization is performed using spaCy with a custom vocabulary. GloVe embeddings initialize the BiLSTM.

BERT-based Model:
Input sentences are augmented with special entity markers (e.g., [E1], [/E1], [E2], [/E2]) using an extended BERT tokenizer.

Training:

The unified model uses CrossEntropy loss with label smoothing and gradient clipping.

The BERT-based model uses mixed precision training, a linear learning rate scheduler with warmup, and incorporates both MLD and MSE losses.

Computational Environment:
Both models are trained and evaluated on the TACRED dataset using RTX 3060 GPU to accelerate deep learning computations.

6 Method Comparison

In this section, we compare the two approaches, the Unified Hybrid Model (Method 1) and the BERT-based Method with Multi-Level Distillation and Mention-Span Enhancement (Method 2), by discussing their advantages and limitations.

The Unified Hybrid Model offers several advantages, notably its interpretability and modular design. The inclusion of a symbolic rule-based component provides human-understandable cues, which facilitates debugging and error analysis. Additionally, the integration of deep learning through a BiLSTM with attention alongside traditional TF–IDF features allows for a flexible combination of statistical and contextual information, while also being more resource-efficient compared to larger transformer-based architectures.
However, this method exhibits significant drawbacks. Its low macro F1 performance (0.1010) suggests that it struggles to capture less frequent or more nuanced relation types.
Furthermore, the reliance on manually defined rules can limit generalization across diverse datasets, and the use of traditional features may restrict the model's ability to fully capture complex semantic nuances.

In contrast, the BERT-based method with Multi-Level Distillation (MLD) and Mention-Span Enhancement (MSE) demonstrates superior overall performance, as evidenced by higher accuracy (0.8655) and macro F1 scores (0.4451). This approach benefits from enhanced representation learning, as the MLD component allows the model to leverage hidden states from multiple BERT layers, resulting in richer semantic representations. The auxiliary MSE head further improves performance by explicitly predicting entity boundaries, which sharpens token-level signals and aids in precise relation classification. Despite these advantages, the BERT-based method comes with increased computational complexity and model size, which may hinder its deployment in resource-constrained environments. Additionally, the increased complexity from the additional MLD and MSE components can lead to greater sensitivity to hyperparameter settings and a higher risk of overfitting if not carefully managed.

7 Evaluation Methodology

Both approaches are evaluated on the TACRED dataset using standard metrics. The evaluation protocol includes:

Accuracy: The overall percentage of correctly classified instances.

Macro F1-Score: The harmonic mean of precision and recall computed on a per-class basis and averaged equally.

For Method 1, the training involves CrossEntropy loss with label smoothing, along with gradient accumulation and clipping to ensure stable convergence. In contrast, Method 2 employs mixed precision training, a linear learning rate scheduler with warmup, and incorporates the MLD and MSE losses into the overall training objective.

8 Error Analysis

An in-depth error analysis reveals that the unified hybrid model (Method 1) struggles with rare relations and cases where rule-based patterns do not trigger. In contrast, the BERT-based method (Method 2) shows improved performance on these challenging cases due to its enhanced internal representations and auxiliary supervision. However, the BERT-based method sometimes misclassifies relations when entity boundaries are ambiguous, indicating room for further refinement in the mention-span component.

9 Results and Discussion

The experimental results for the two approaches on TACRED are as follows:

Method 1 (Unified Hybrid Model):

Test Accuracy: 0.7592

Test Macro F1-Score: 0.1010

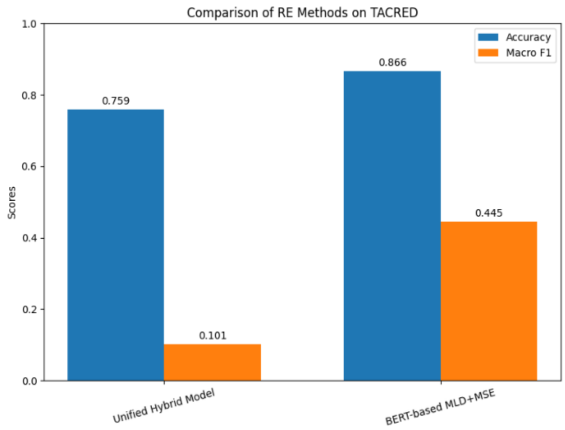
Method 2 (BERT-based with MLD and MSE):

Test Accuracy: 0.8655

Test Macro F1-Score: 0.4451

The results indicate that while the unified hybrid model (Method 1) demonstrates reasonable accuracy, its macro F1-score is significantly lower, suggesting difficulties in capturing less frequent or more nuanced relation types. In contrast, the BERT-based method (Method 2) achieves higher accuracy and a markedly improved macro F1-score. The multi-level distillation likely contributes to better regularization, and the mention-span enhancement aids in refining entity boundaries—both factors that are critical for improving relation classification performance.

However, the BERT-based approach also incurs higher computational costs and model complexity.



10 Future Work

Hybrid Integration: Combining the strengths of both approaches to leverage the interpretability of symbolic features and the high performance of transformer-based models.

Fine-Tuning Auxiliary Losses: Experimenting with different weighting schemes for the MLD and MSE losses to further optimize performance.

Dataset Expansion: Evaluating the models on additional datasets to assess generalizability across various RE tasks.

Error Correction Mechanisms: Incorporating post-processing steps or ensemble techniques to address misclassification in ambiguous cases.

11 Conclusion

This paper has presented and evaluated two approaches to relation extraction on the TACRED dataset. The unified hybrid model combines BiLSTM, symbolic rule-based extraction, and TF–IDF features to leverage diverse information sources. Meanwhile, the novel BERT-based approach with Multi-Level Distillation and Mention-Span Enhancement introduces regularization and auxiliary tasks to improve performance. Although Method 2 outperforms Method 1 in both accuracy and macro F1-score, each method offers unique advantages. The unified model provides modularity and interpretability, while the BERT-based approach benefits from large-scale pre-training and sophisticated internal representations. Future work could explore integrating the strengths of both methods to further advance the state-of-the-art in relation extraction.

12 References

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL-HLT*.

Hong, G., 2005. Relation Extraction Using Support Vector Machine. In *Proceedings of the Second International Joint Conference on Natural Language Processing: Full Papers*.

Hinton, G., Vinyals, O. and Dean, J., 2015. Distilling the Knowledge in a Neural Network. *arXiv preprint arXiv:1503.02531*.

Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K. and Dyer, C., 2016. Neural Architectures for Named Entity Recognition. In *Proceedings of NAACL-HLT*.

Rink, B. and Harabagiu, S., 2010. UTD: Classifying Semantic Relations by Combining Lexical and Semantic Resources. In *ACL 2010* (pp. 256-259).

Valette, M., 2019. Simple Relation Extraction with a Bi-LSTM Model. Available at: https://medium.com/southpigalle/simple-relation-extraction-with-a-bi-lstm-model-part-1-682b670d5e11 [Access Date: 22/02/25].

Zhang, S., Zheng, D., Hu, X. and Yang, M., 2015. Bidirectional Long Short-Term Memory Networks for Relation Classification. In *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation* (pp. 73-78).

Zhang, Y., Qi, P. and Manning, C.D., 2017. TACRED: A Large-Scale Relation Extraction Dataset. In *Proceedings of NAACL-HLT*.