A Distantly-Supervised Relation Extraction Method Based on Selective Gate and Noise Correction

Zhuowei Chen [1] Yujia Tian [1] Lianxi Wang✉ [1,2] Shengyi Jiang [1,2]

[1]Guangdong University of Foreign Studies, Guangzhou, China, 510006
[2]Guangzhou Key Laboratory of Multilingual Intelligent Processing, Guangzhou, China, 510006

Introduction

Distantly supervised relation extraction (DSRE) aims to address the scarcity of annotated data in relation extraction tasks. However, the distant supervision method introduces noise, negatively impacting the relation extraction model's performance, especially when it becomes overfitted to noisy data.

To tackle the challenge of noisy data, this research proposes a novel approach that applies a selective gate mechanism and an end-to-end noise correction framework. The selective gate generates a bag-level representation by selectively combining sentence features. The noise correction framework rectifies misclassified labels during the training process. Furthermore, in this era of dominant large language models, we also introduce a pre-trained language model to our method, enhancing sentence comprehension and performance.

Experiments on different datasets demonstrate that the proposed method outperforms baseline models, significantly improving the performance of distant supervised Entity Relation Extraction. The innovative combination of selective gate, noise correction, and PLMs promises to bring new insights to the field of Entity Relation Extraction.

Main contributions

- We propose a DSRE method, named PLMG-Pencil, which combines PLM and selective gate and introduces an end-to-end noise correction training framework called pencil. Selective gate prevents the propagation of noisy representations and pencil corrects noise labels during the training process, reducing the impact of noise on the dataset and improving the performance of the DSRE model.
- We present a novel algorithm for DSRE that combines selective gate mechanism and pencil framework within a three-stage training process. This process involves training the backbone model, gradually correcting noisy labels, and subsequently fine-tuning our model using the corrected data. Empirical experiments demonstrate the robustness and effectiveness of our proposed method.
- Our experiments on three different Chinese and English datasets demonstrate that effective sentence-level feature construction methods and training methods, combined with noise correction, are crucial for improving the performance of models on DSRE tasks.

Experiment Results

We evaluate our proposed model on three different datasets and use AUC and P@N values as evaluation metrics. AUC measures the area under the ROC curve, P@N indicates the average accuracy of top N instances, and P@M represents the average of these three P@N results.

These results suggest that the selective gate has a positive impact on constructing sentence bag features and improving model performance. Furthermore, the pencil framework effectively corrects for noisy samples during training, leading to improved performance.

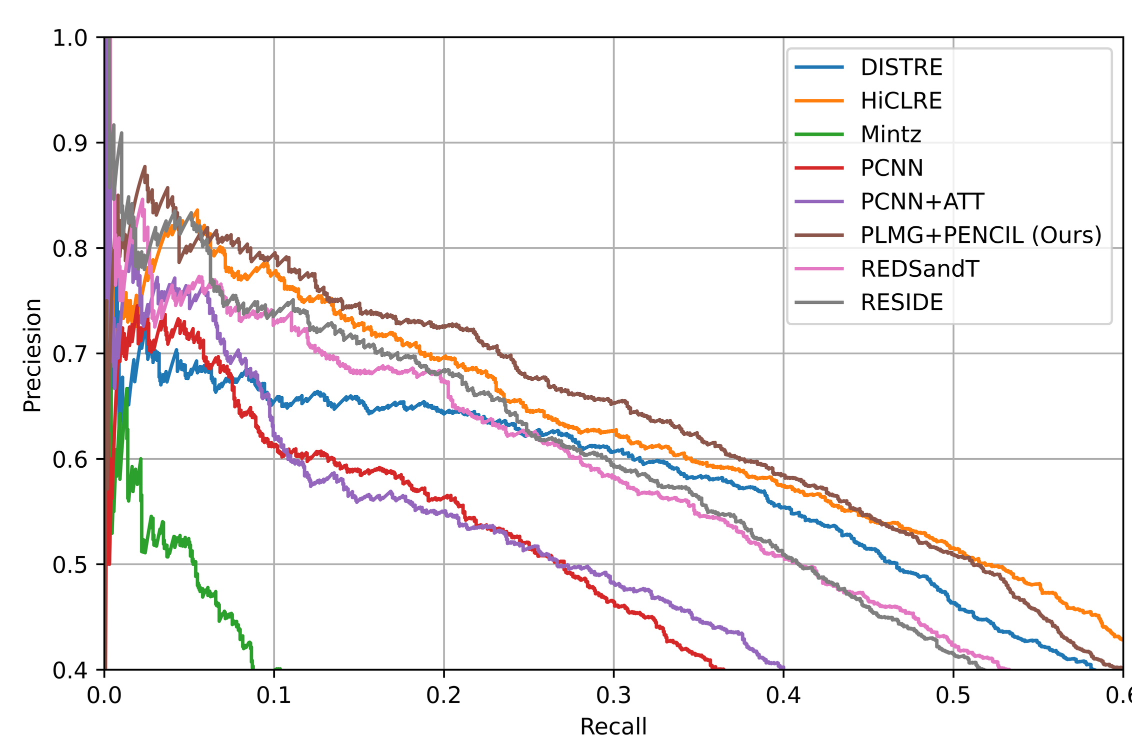

Figure 1. PR Curve on NYT-10. Our model clearly outperforms the baselines, particularly compared to the DISTRE model, which also uses pre-train language model and multi-instances learning.

Dataset	Models	AUC	P@500	P@1K	P@2K	P@M
GDS	PCNN-ATT† (2016)	79.9	90.6	87.6	75.2	84.5
	MTB-MIL† (2019)	88.5	94.8	92.2	87.0	91.3
	RESIDE† (2019)	89.1	94.8	91.1	82.7	89.5
	REDSandT† (2021)	86.1	95.6	92.6	84.6	91.0
	DISTRE† (2019)	89.9	97.0	93.8	87.6	92.8
	CIL† (2021)	90.8	97.1	94.0	87.8	93.0
	HiCLRE (2022)	90.8	96.6	93.8	88.8	93.1
	PLMG-Pencil (OURS)	91.0	95.4	94.1	88.8	92.8
	without pencil (PLMG)	90.8	95.4	94.0	89.0	92.8

Table 1. Model Performances on GDS. (†) marks the results are reported in the previous research. On the GDS dataset, the AUC values of our model reach comparable levels with CIL and HiCLRE.

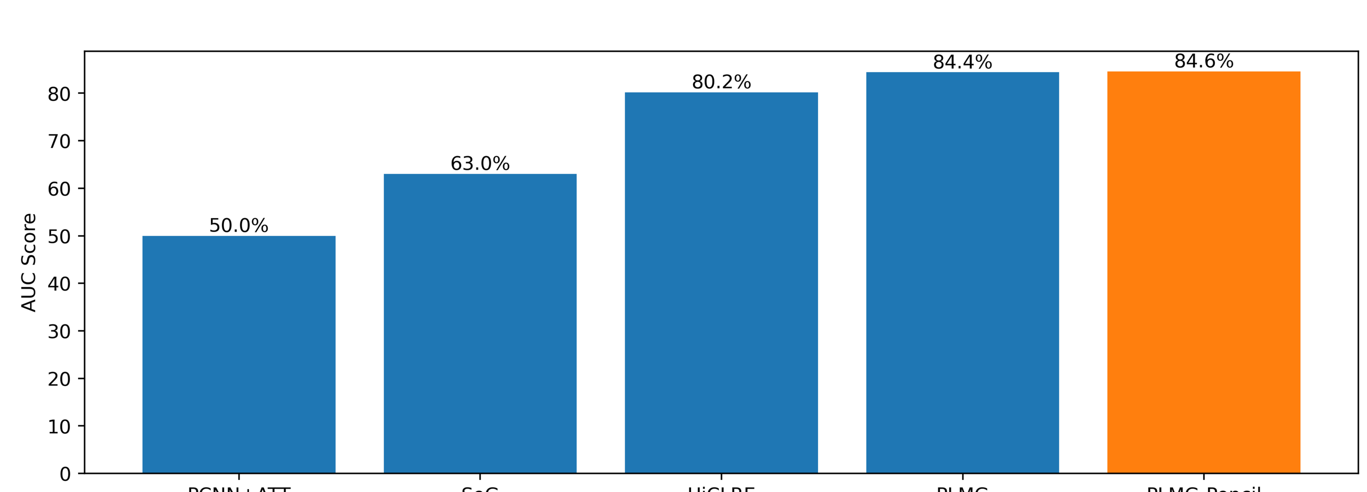

Figure 2. AUC on SanWen.

Our model exhibits superior performance compared to HiCLRE, which utilizes the contrast learning framework, with a notable increase of 4.4% in AUC values. Furthermore, when compared to the SeG model that employs the selective gate mechanism, our PLMG-Pencil model, which incorporates the pencil approach, demonstrates a significant enhancement in AUC values. The ablation experiment further validates the effectiveness and robustness of our method.

Backbone Model

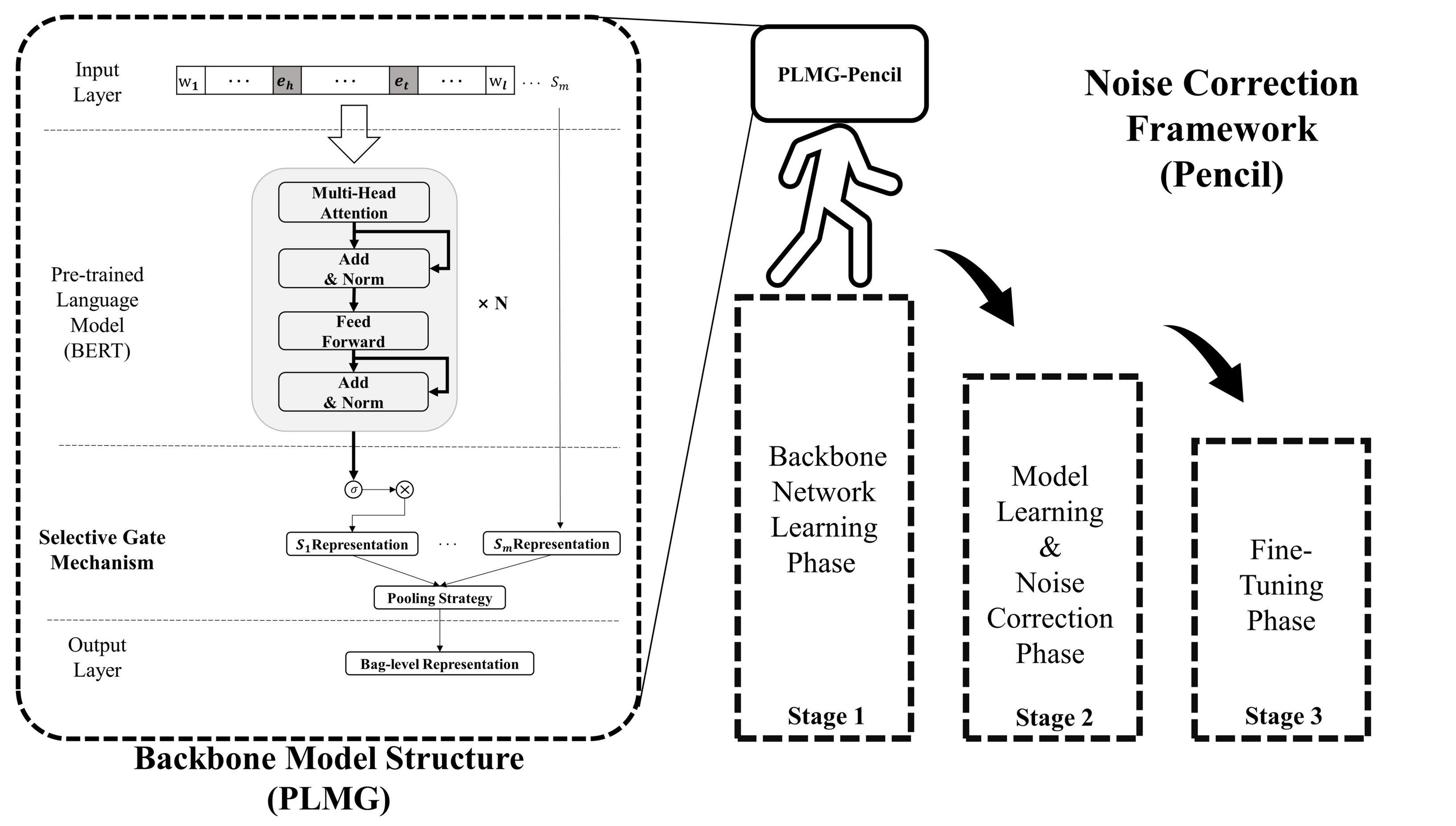

Figure 3. Overview of PLMG-Pencil

This paper proposes the PLM-based selective gate as the backbone model, inspired by the Entity-aware Self-attention Enhanced selective gate (SeG) framework proposed by Li (2020b). The primary architecture of our model is presented in Figure 3, and it comprises two main components:

- PLM: Structured to encode sentence, entity, and location features for semantic enhancement.
- Selective Gate: Enhances the representation of bag-level features by assigning weights to different sentences in the bag. The selective gate mechanism reduces the impact of noise on the model by weighing the contribution of each sentence in the bag.

Noise Correction Framework

In our research, we introduce the Pencil framework, a noise correction approach based on the end-to-end noise-labeled learning correction framework. The framework is illustrated in Figure 4, with solid arrows representing forward computation and dashed arrows indicating backward propagation.

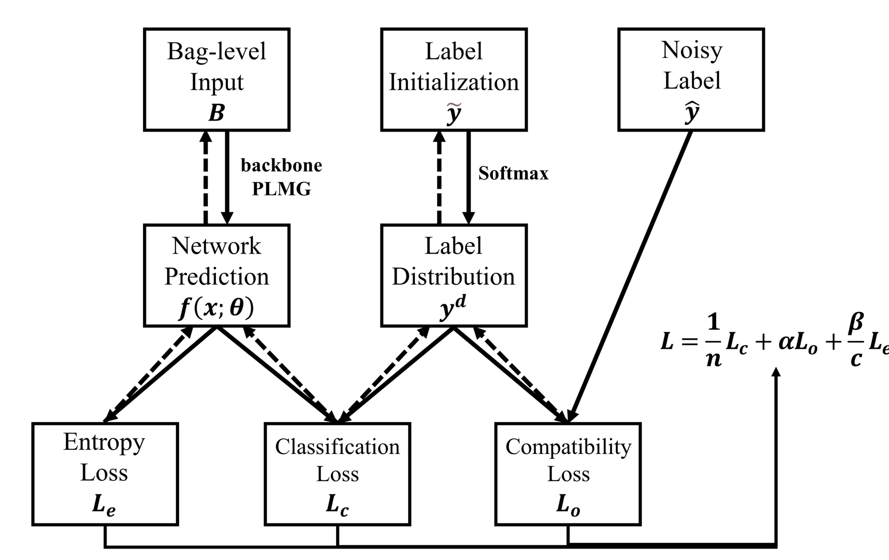

Figure 4. Pencil Framework

The Pencil framework is designed to update both the network parameters and the data labels simultaneously using gradient descent and backpropagation.

To achieve this, the model generates a vector $\tilde{y}$ by Softmax function to construct soft labels . Futhermore, an intricately devised loss function Eq.1 is employed to correct the noise labels during the model training procedure, with L_e and L_o as penalty terms and L_c as the classification loss. This loss function incorporates two hyperparameters, denoted as α and β, which can be flexibly adjusted to accommodate diverse datasets with varying proportions of noisy data.

$$L = \frac{1}{c}L_c + \alpha L_o + \frac{\beta}{c}L_e \tag{1}$$

PLMG-Pencil Relation Extraction Method

- Stage 1 - Backbone Network Learning Phase: Initially, the PLMG-Pencil network is trained from scratch with a larger fixed learning rate. The noise in the data is not processed in this stage.
- Stage 2 - Model Learning and Noise Correction Phase: In this stage, the network parameters and label distributions are updated together using the model, thus, noisy labels can be corrected.
- Stage 3 - Final Fine-Tuning Phase: The label distribution learned by the model in the previous stages are utilized to fine-tune the network in this stage. Sample labels in the training set are not updated, and the network parameters are updated using the classification loss as the loss function of the model.

Algorithm 1: PLMG-Pencil Distantly Supervised Relation Extraction Algorithm

Input: Dataset $D = x_i, \tilde{y}_i (1 < i < n)$, epoch of stages T_1, T_2.

Stage 1:
Initialization: $t \leftarrow 1$.
while $t \leq T_1$ **do**
  Train and update the model parameters θ, while calculating the loss in equation (14) with $\alpha = 0$ and $\beta = 0$. Hold off on using $\tilde{y}_i$;
  $t \leftarrow t + 1$;

Stage 2:
Initialization: $\tilde{y}_i = K\hat{y}_i$.
while $T_1 \leq t \leq T_2$ **do**
  Train and update the model parameters θ and y_i^d;
  $t \leftarrow t + 1$;

Stage 3:
while $T_2 \leq t$ **do**
  Train and update the model parameters θ and y_i^d;
  Train and update the model parameters θ, while calculating the loss in equation (14) with $\alpha = 0$ and $\beta = 0$. Do not update sample labels.
  $t \leftarrow t + 1$;

Output: θ, noise-corrected labels.