

A Model Structure in Details

Figure 1 provides details of our model structure with the output dimensions of each layer. $|T|$ is the number of tokens in the input sentence. The dimensions of the final output are $\mathbb{R}^{|T| \times 3}$, where the three parameters represents the scores for blank, uninformative, and informative, respectively. We did not tune the model structure nor the dimensions of the output of each layer.

B Hyperparameters tuning

We tuned a number of hyperparameters as listed in Table 1. We used manual tuning for each hyperparameters with three trials for each setting, and chose the best settings. Note that all of these parameters are not critical. In our experiments, the F1 scores differ by no more than 0.01.

C Runtime

In our experiments, each epoch in the training process took approximately three minutes with the number of training data set to 7,000, and trialled up to 50 epochs, so the total runtime for each trial was 2.5 hours. Our experiments were conducted on a Linux server using Xeon E5-1620 v4 (3.50 GHz) CPU with 256 GB of memory, and we used a single GeForce GTX 1080 Ti GPU with 11GB of memory for each trial.

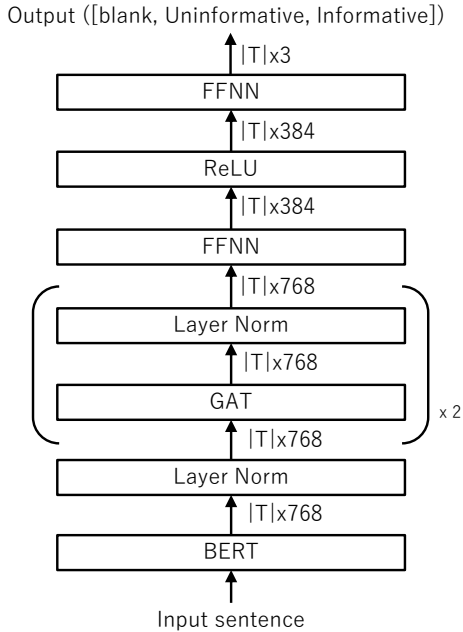


Figure 1: Model structure of our method with the dimensions of output of each layer.

Table 1: Hyperparameters tuning. The numbers with **bold face** represent the setting that we finally used in the paper.

Hyperparameter	Trials
Learning rate	{ 0.001 , 0.0005}
L2 regularization coefficient	{ 0.1 , 0.001, 0.00001}
Dropout ratio	{ 0.1 , 0.2, 0.3, 0.5}
Word dropout ratio	{0.1, 0.2 , 0.3}
Label smoothing ratio in CTCTC	{0.1, 0.2 , 0.3}
Token smoothing ratio in CTCTC	{0.1, 0.2 , 0.3}
Leaking ratio in CTCTC	{0.1, 0.2 , 0.3}