

# Integrating Local Context Aware Self-Attention with Part of Speech Attention for Question Classification

Yurong Xie (✉ [sapphire\\_xyr@163.com](mailto:sapphire_xyr@163.com))

Shanghai Maritime University <https://orcid.org/0000-0001-8124-0126>

Jin Liu

Shanghai Maritime University <https://orcid.org/0000-0001-7249-698X>

Yujie Zhou

Shanghai Maritime University

---

## Research Article

**Keywords:** question classification, self-attention, neighbor relation, local context representation

**Posted Date:** May 24th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-547138/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Question Answering (QA) is an important task in Natural Language Processing (NLP). An effective and efficient question classification model is essential to question answering, which can not only reduce answers search space but also guide the QA system to select the correct knowledge base and search strategy. In the literature, most existing approaches adopt self-attention mechanisms to classify the questions. Thus, they only capture the global dependencies, ignoring the influence of adjacent local representations. To address this issue, this paper proposes a novel self-attention distribution network to enhance the local context awareness by exploiting neighbor relations (SNR) among representations in the text. We introduce a learnable Gaussian bias to obtain an adaptive self-attention distribution via a dynamic window. This learnable Gaussian bias is then integrated into the original attention distribution to capture both the long-term dependency and local representation of semantic relations among words in different sentences. Furthermore, a novel position-aware part of speech (POS) attention embedding layer is proposed, which can contribute to finding the appropriate word corresponding to the original word as the central word. Extensive experiments are conducted on Experimental Data for Question Classification (EDQC) dataset and Yahoo! Answers Comprehensive Questions and Answers 1.0, the results demonstrate that our model significantly improves the performance, achieving 95.59% in coarse-grained level accuracy and 92.91% in fine-grained level accuracy respectively.

# Full Text

This preprint is available for [download as a PDF](#).

# Figures

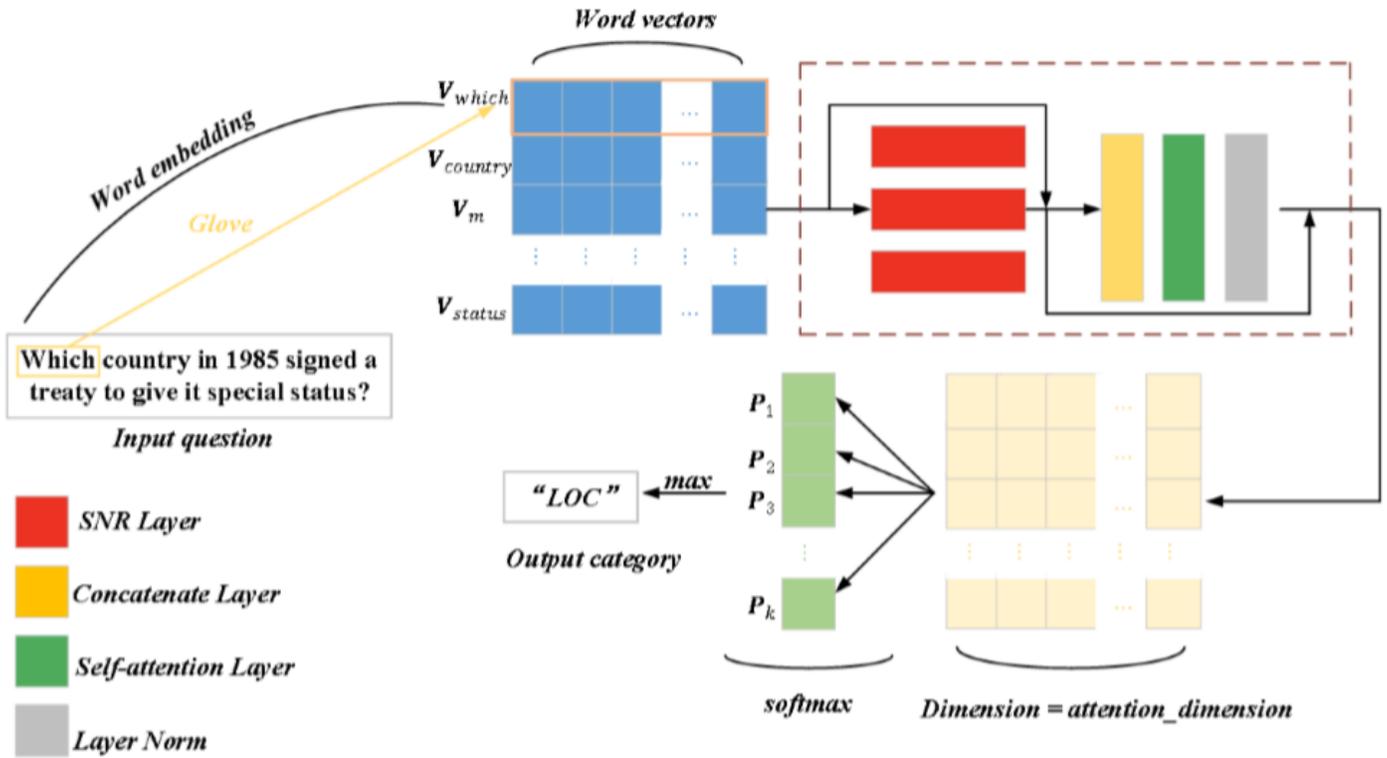


Figure 1

The whole architecture of the proposed SNR model.

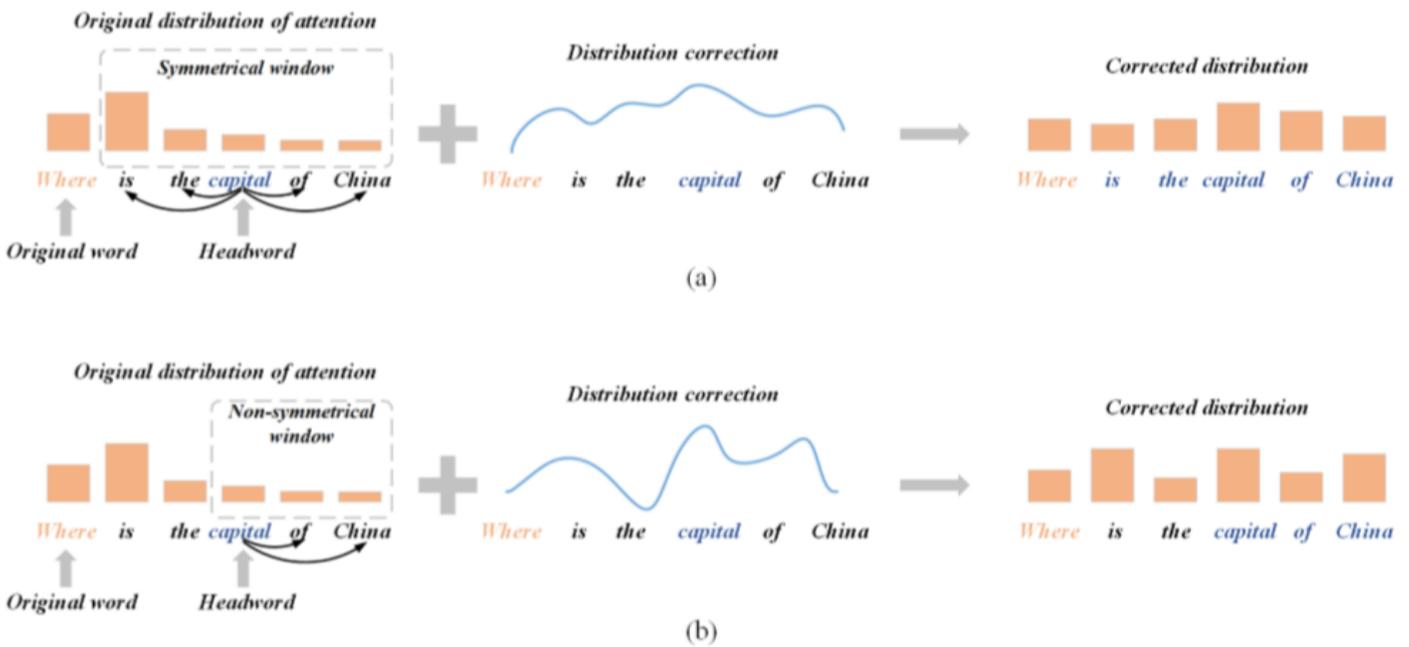


Figure 2

Distribution correction of symmetrical window and non-symmetrical window.

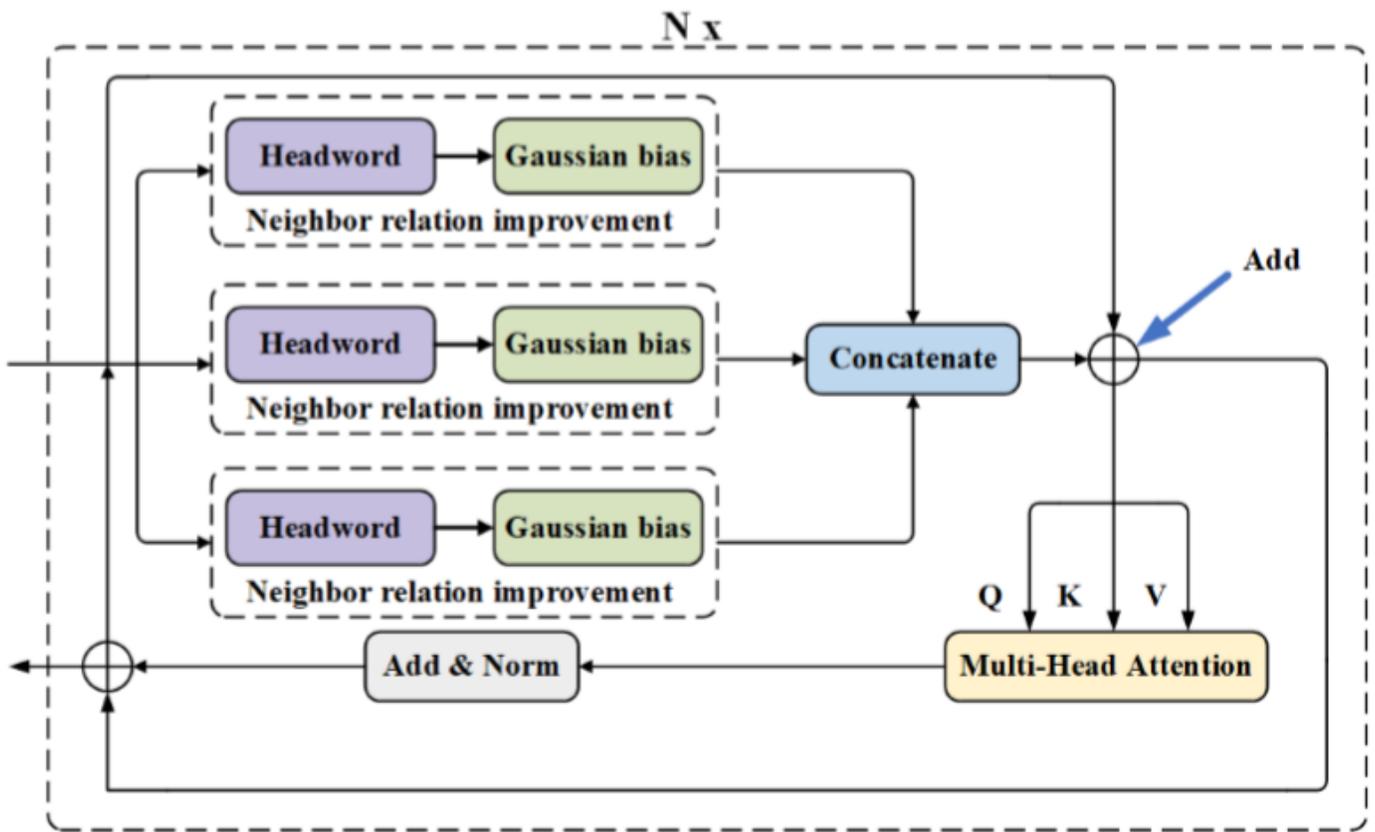


Figure 3

The structure of SNR module.

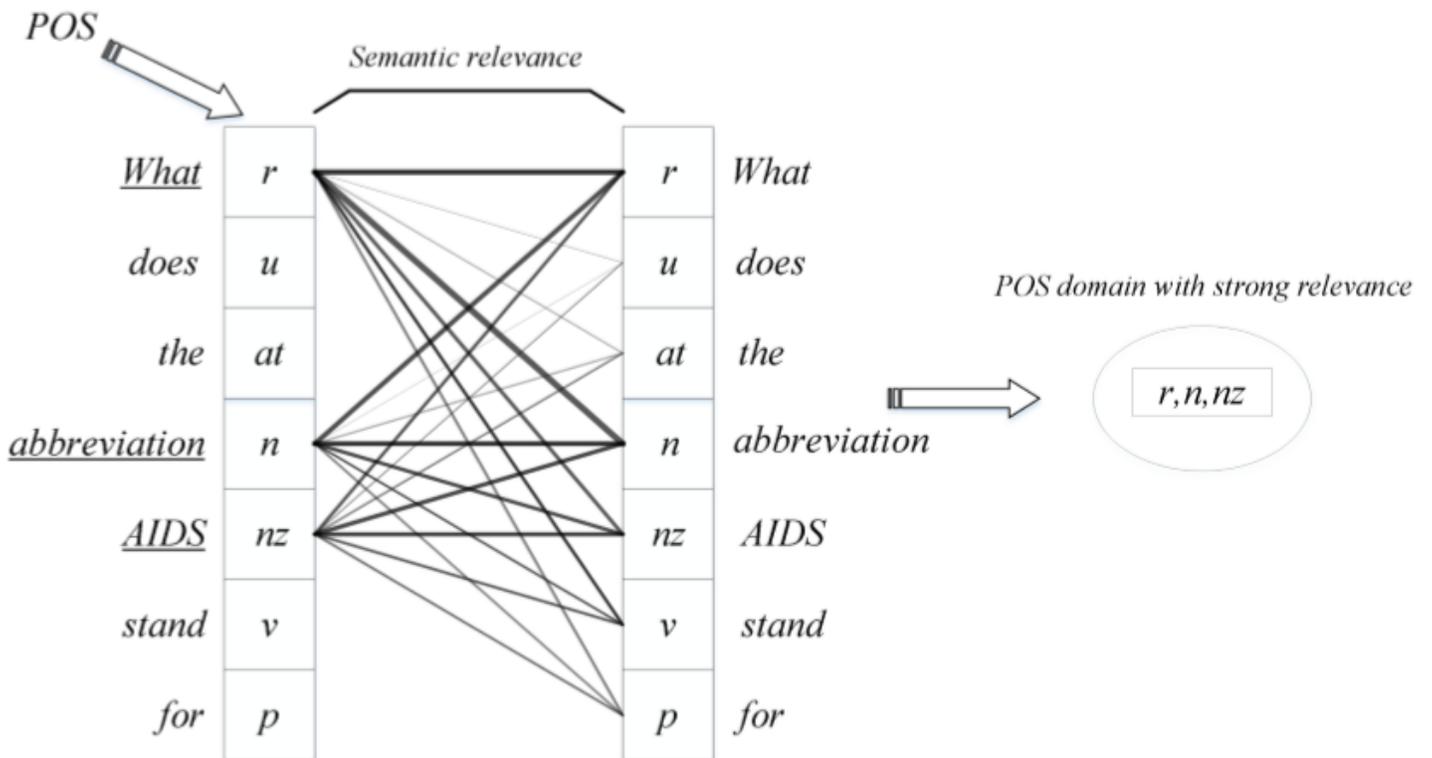


Figure 4

The relation between semantic relevance and POS.

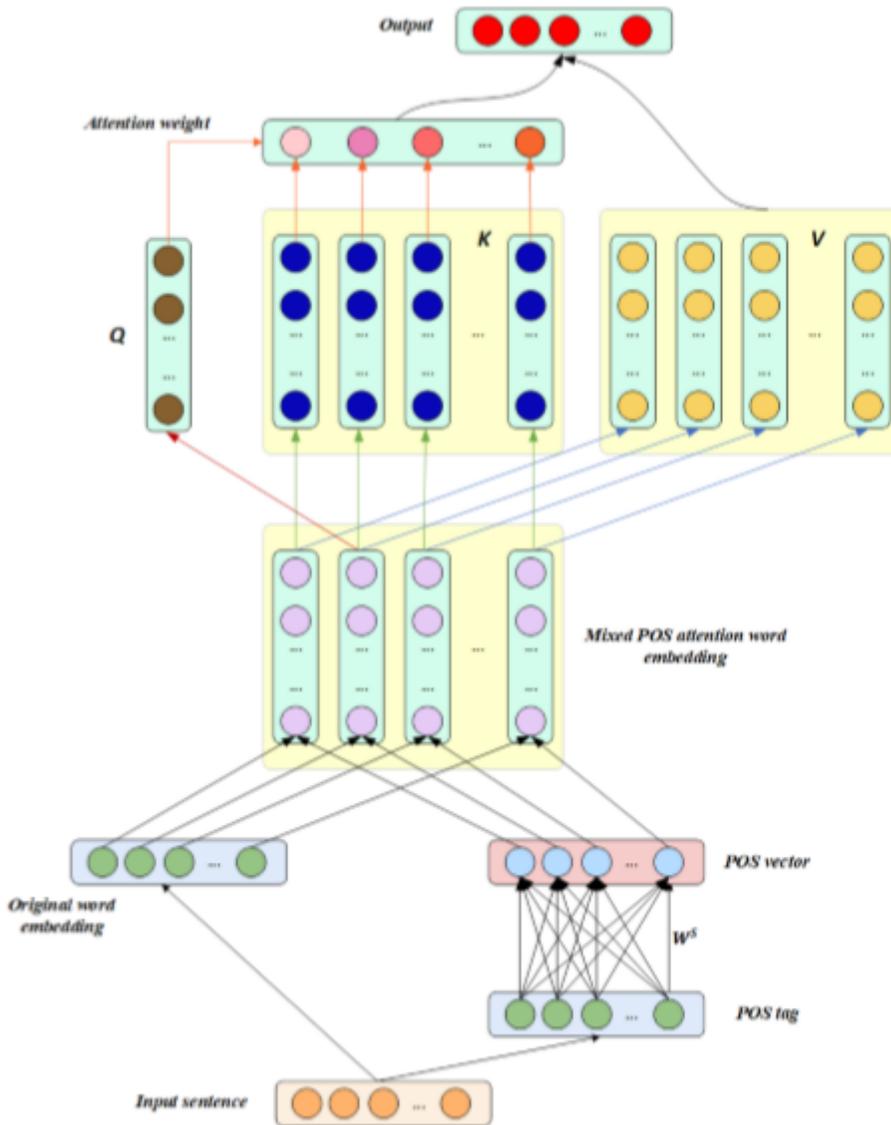
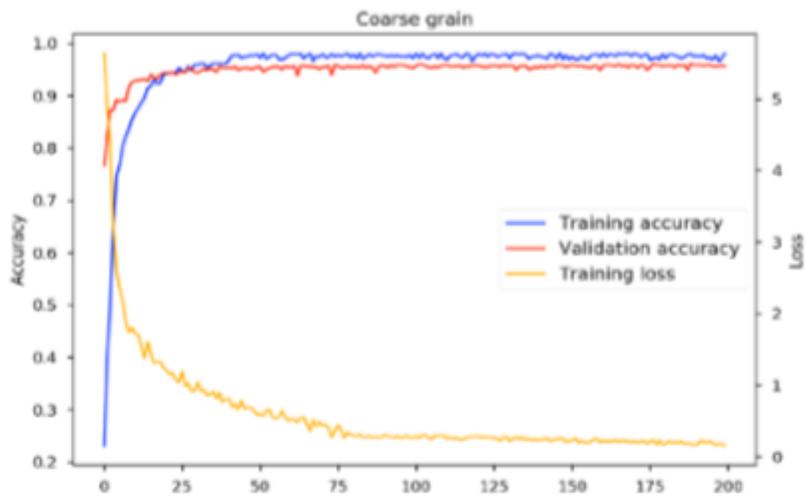
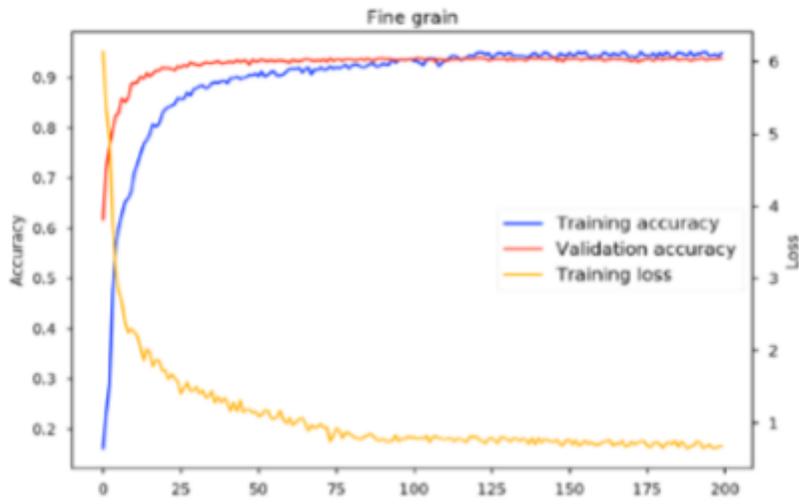


Figure 5

The detailed process of improved self-attention.



(a)



(b)

Figure 6

Visualization of the training accuracy.

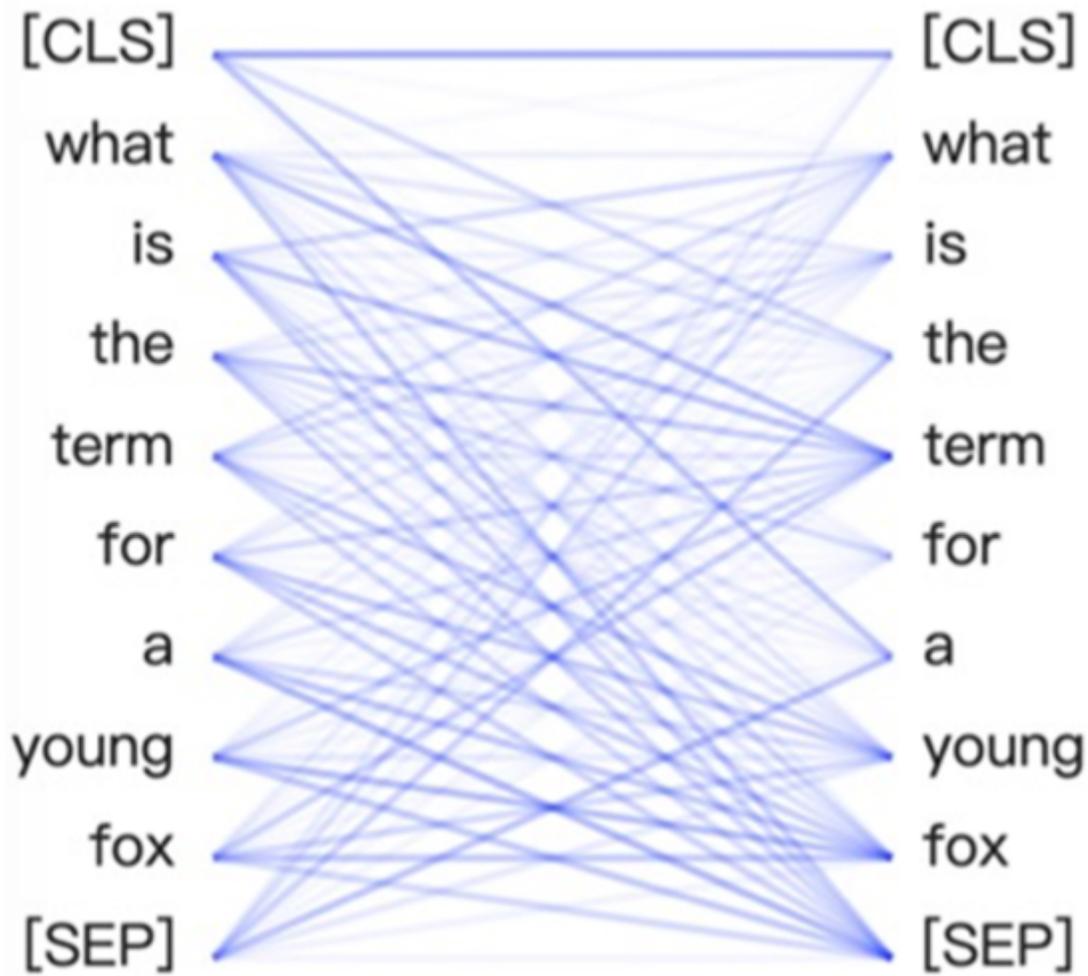


Figure 7

The association of each word in the sentence.

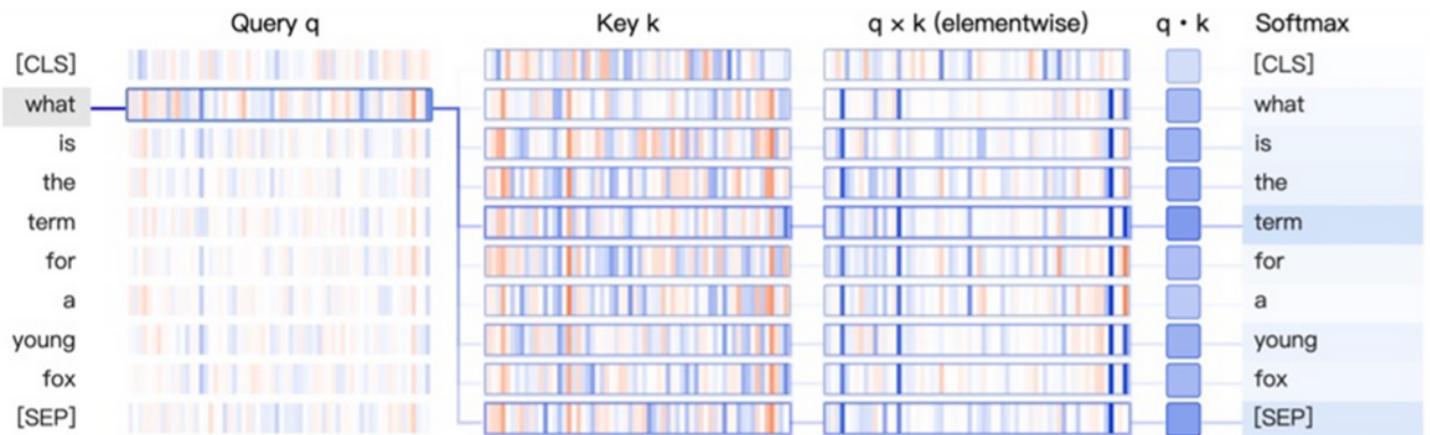


Figure 8

The visualized result of the improved self-attention. [CLS] and [SEP] are two delimiters mentioned in (Radford et al., 2018) and have no semantic information.

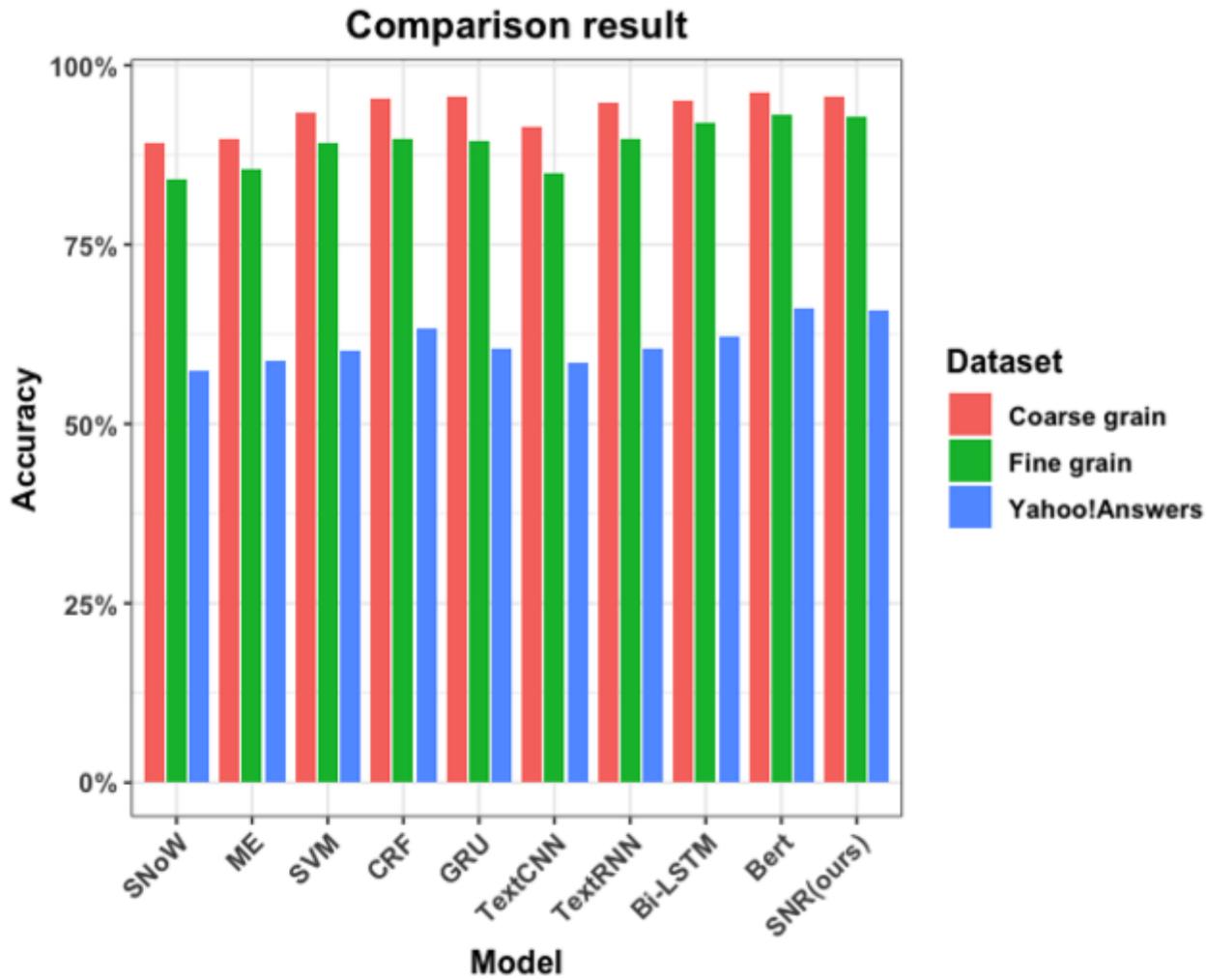


Figure 9

The comparison results of various models.

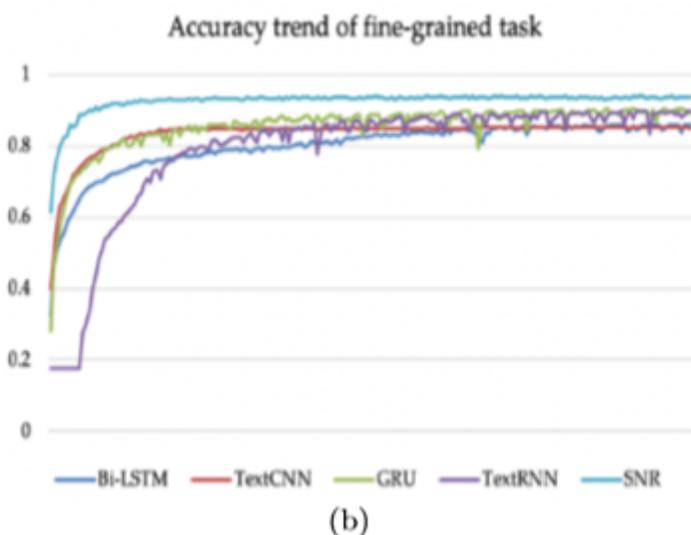
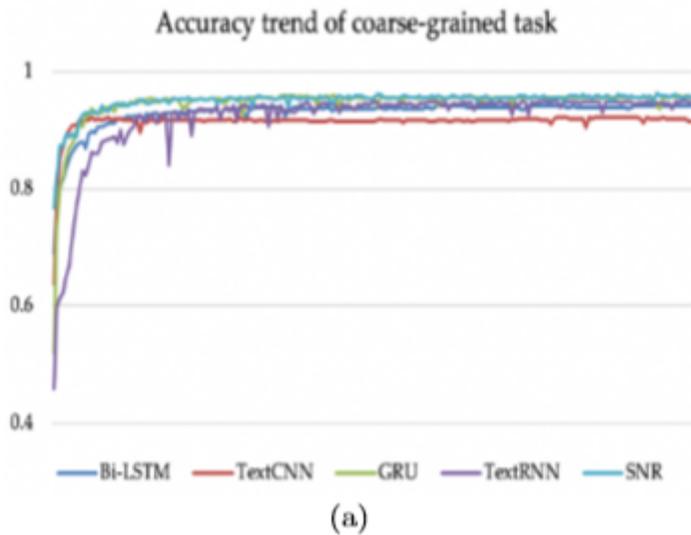


Figure 10

The coarse-grained and ne-grained accuracy trend.

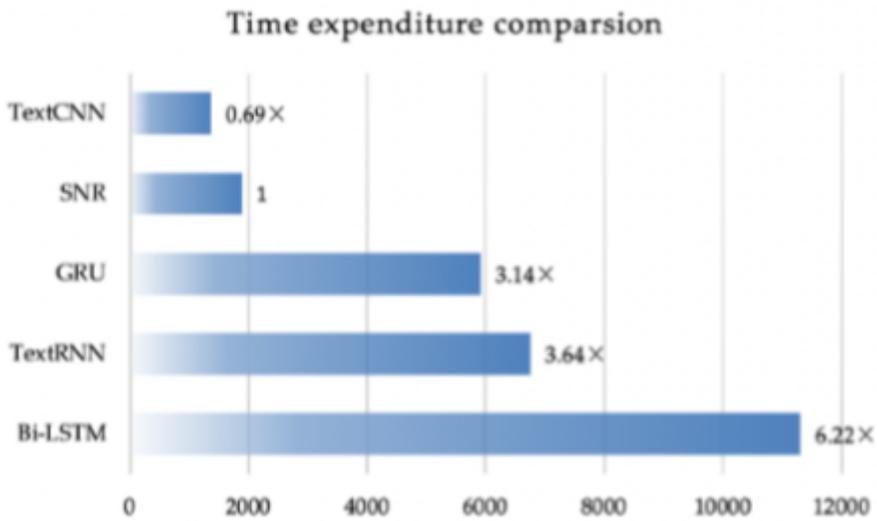


Figure 11

The difference of various models time expenditure.

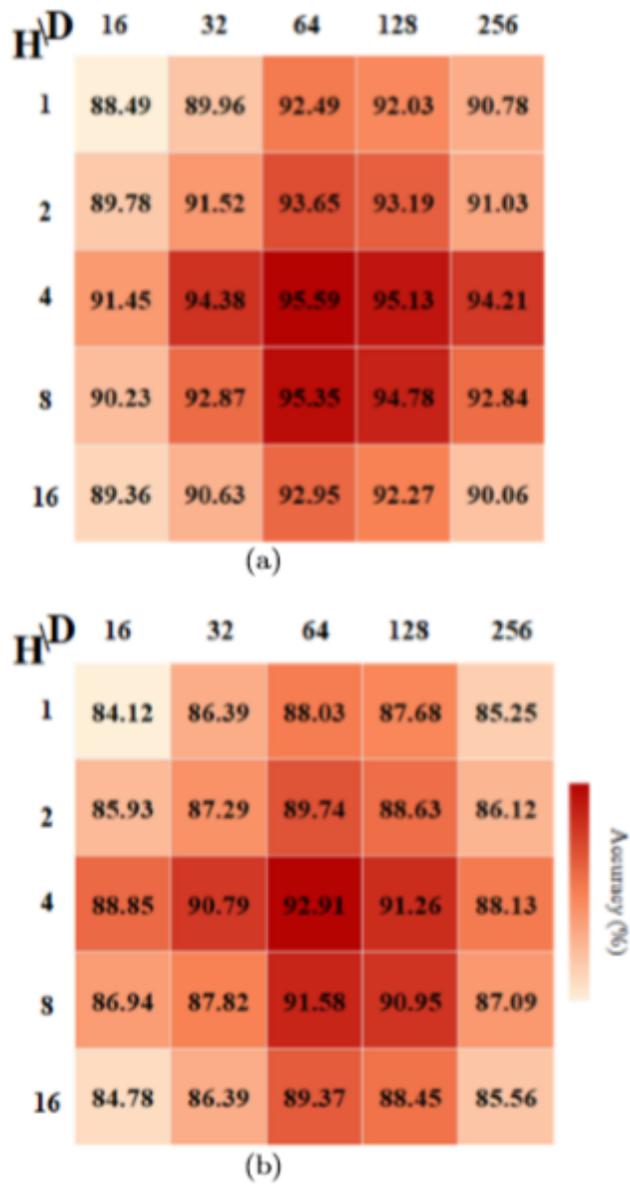
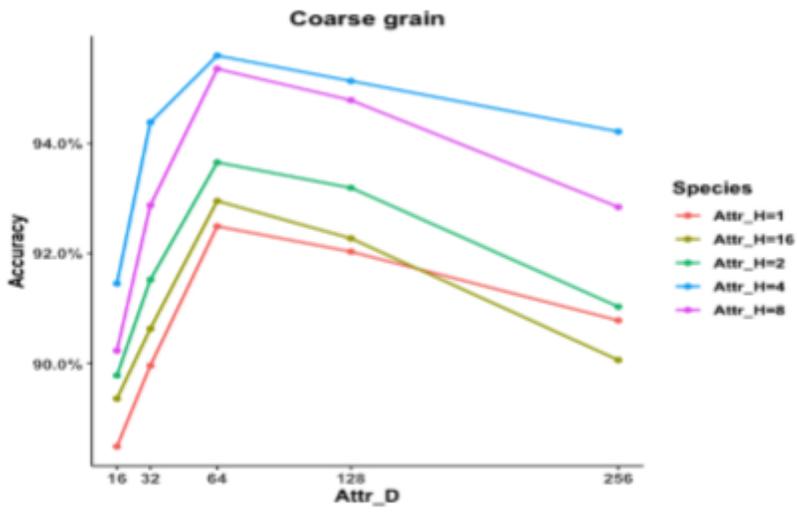
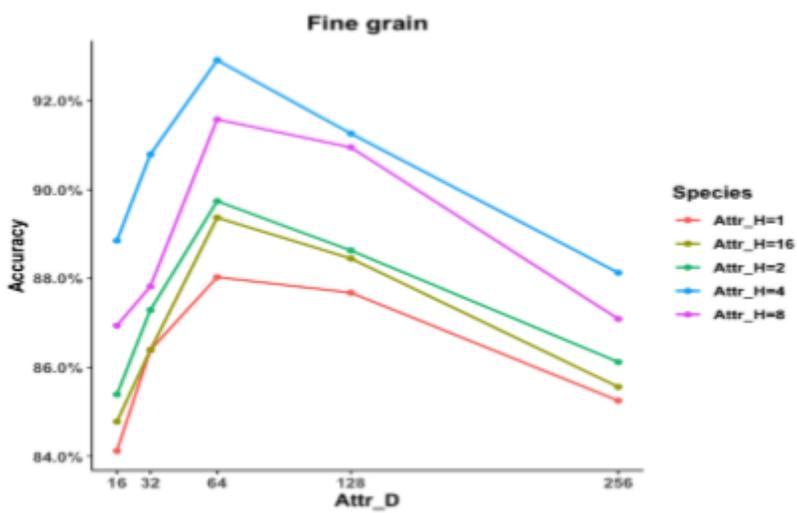


Figure 12

Heatmap of accuracy (%) on EDQC of different configurations.



(a)



(b)

Figure 13

Results on EDQC with different Attr-H and Attr-D.