# SG-Net: Syntax-Guided Machine Reading Comprehension

Zhuosheng Zhang, Yuwei Wu, Junru Zhou, Sufeng Duan, Hai Zhao*, Rui Wang*
Shanghai Jiao Tong University & National Institute of Information and Communications Technology (NICT)
zhangzs@sjtu.edu.cn, will8821@sjtu.edu.cn, zhaohai@cs.sjtu.edu.cn

## Introduction

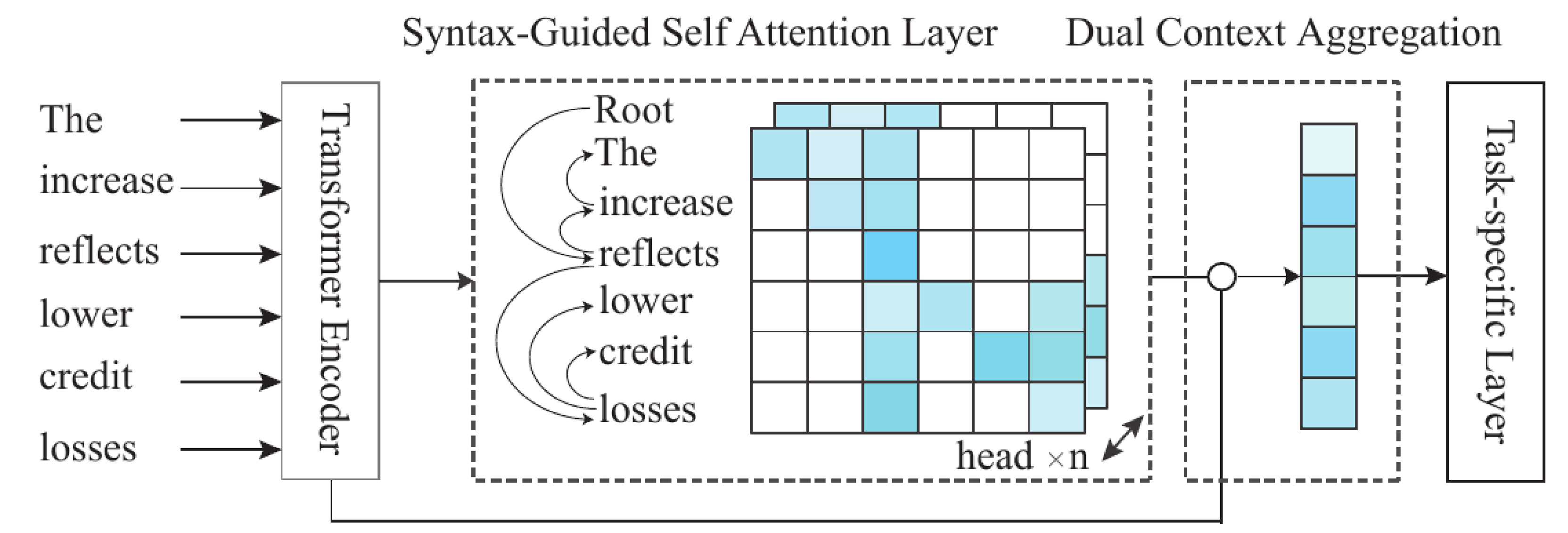
**SG-Net:**
- uses syntax to guide the text modeling by incorporating explicit syntactic constraints into attention mechanism
- introduces syntactic dependency of interest (SDOI) design into the self-attention network (SAN) to form an SDOI-SAN with syntax-guided self-attention
- helps achieve substantial performance improvement over strong baselines on popular benchmarks including SQuAD2.0 and RACE

**Motivation:**
- A person reads most words superficially and pays more attention to the key ones during reading and understanding sentences.
- The linguistic knowledge from the detail-riddled and lengthy passages and getting ride of the noises is essential for better reading comprehension.
- Traditional attentive models attend to all words without explicit constraint, which results in inaccurate concentration on some dispensable words.

**Paper Link**: https://arxiv.org/abs/1908.05147
**Code Link**: https://github.com/cooelf/SG-Net

## Syntax-Guided Network

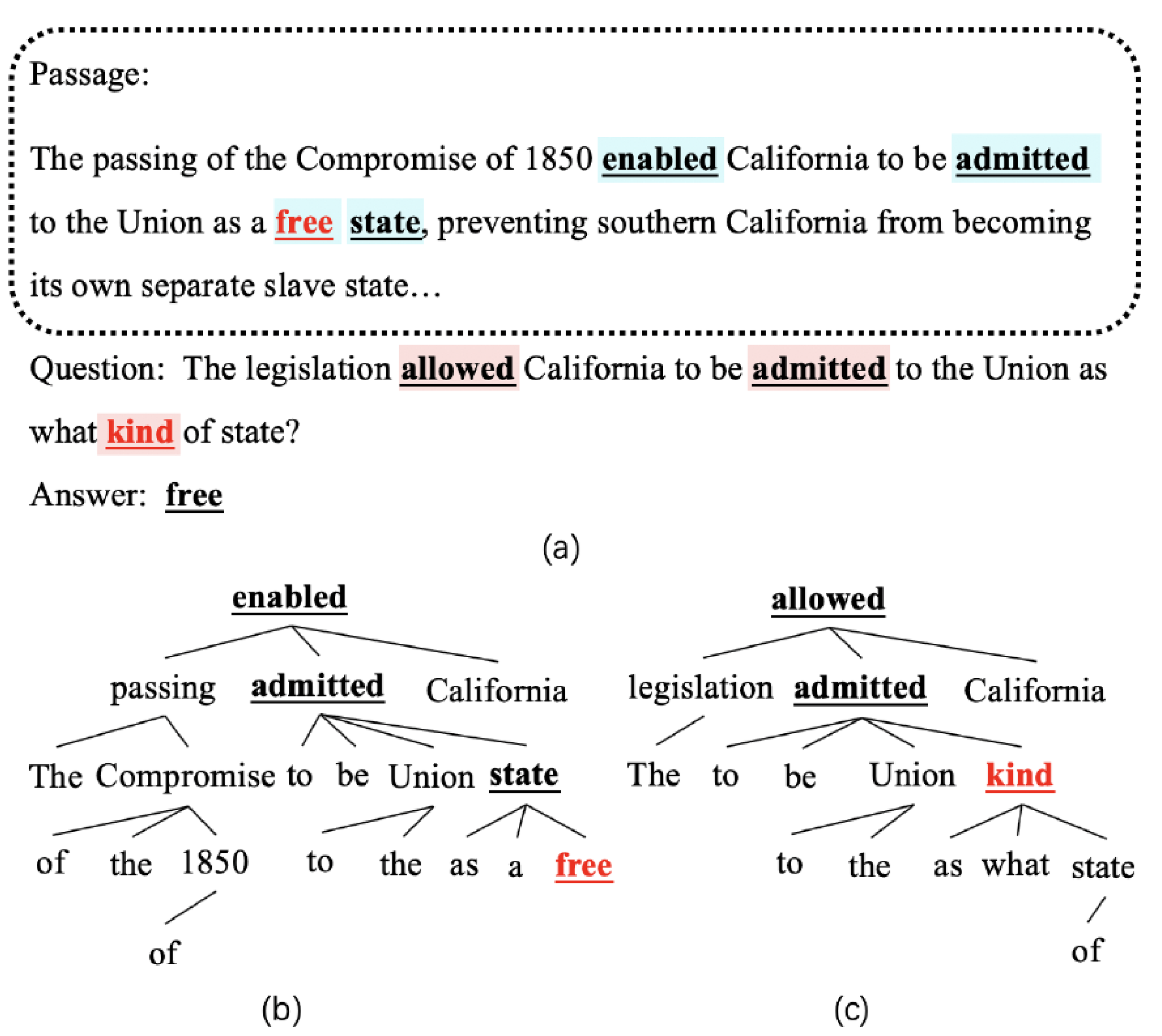
**Syntax-Guided Network**
- We pass the encoded representation from the Transformer encoder to a syntax-guided self-attention layer.
- The corresponding output is aggregated with the original encoder output to form a syntax-enhanced representation.




**Syntax-Guided self-attention Layer**
- pre-train a syntactic dependency parser to annotate the dependency structures for every sentence.
- restrain the scope of attention only between word and all of its ancestor head words using the SDOI mask.

## Method

**Syntactic dependency of interest**
- adopt pre-trained dependency syntactic parse tree structure
- produce the related nodes for each word in a sentence, namely syntactic dependency of interest (SDOI)
- regard each word as a child node and the SDOI consists all its ancestor nodes and itself in the dependency parsing tree.




**SDOI mask**
- organized as n*n matrix
- elements in each row denote the dependency mask of all words to the row-index word.

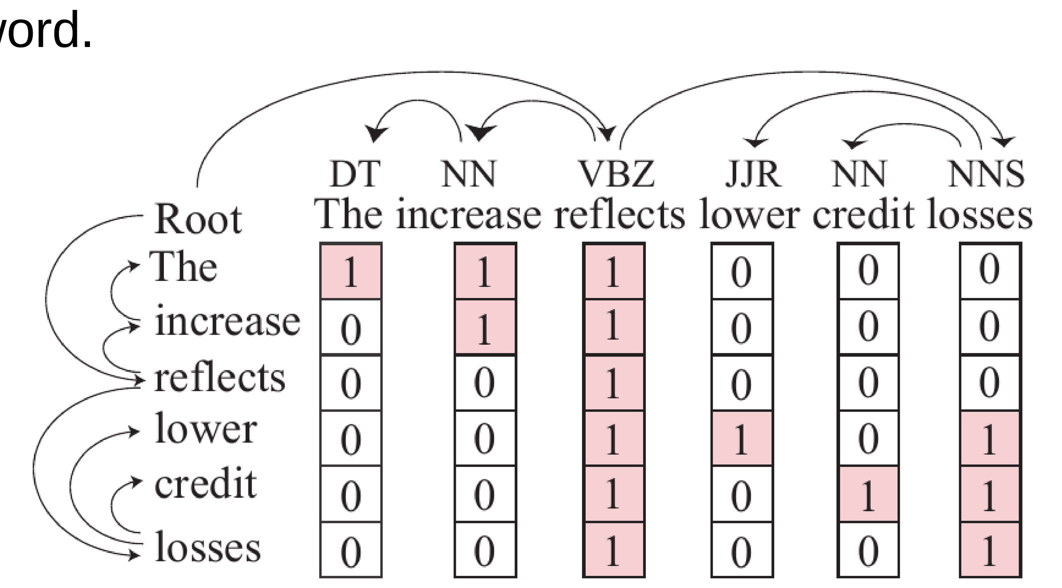



Figure 3: An example of the syntactic dependency of interest (SDOI) mask.

## Experiments

**Datasets**: SQuAD2.0 and RACE

| Model | Dev EM | Dev F1 | Test EM | Test F1 |
|---|---|---|---|---|
| *Regular Track* | | | | |
| Joint SAN | 69.3 | 72.2 | 68.7 | 71.4 |
| U-Net | 70.3 | 74.0 | 69.2 | 72.6 |
| RMR + ELMo + Verifier | 72.3 | 74.8 | 71.7 | 74.2 |
| *BERT Track* | | | | |
| Human | - | - | 86.8 | 89.5 |
| BERT + DAE + AoA† | - | - | 85.9 | 88.6 |
| BERT + NGM + SST† | - | - | 85.2 | 87.7 |
| BERT + CLSTM + MTL + V† | - | - | 84.9 | 88.2 |
| SemBERT† | - | - | 84.8 | 87.9 |
| Insight-baseline-BERT† | - | - | 84.8 | 87.6 |
| BERT + MMFT + ADA† | - | - | 83.0 | 85.9 |
| $\text{BERT}_{\text{LARGE}}$ | - | - | 82.1 | 84.8 |
| Baseline | 84.1 | 86.8 | - | - |
| **SG-Net** | **85.1** | **87.9** | - | - |
| **+Verifier** | **85.6** | **88.3** | **85.2** | **87.9** |

*Results for SQuAD2.0*

| Model | EM | F1 |
|---|---|---|
| baseline | 84.1 | 86.8 |
| + Vanilla attention only | 84.2 | 86.9 |
| + Syntax-guided attention only | 84.4 | 87.2 |
| + Dual contextual attention | **85.1** | **87.9** |
| Concatenation | 84.5 | 87.6 |
| Bi-attention | 84.9 | 87.8 |

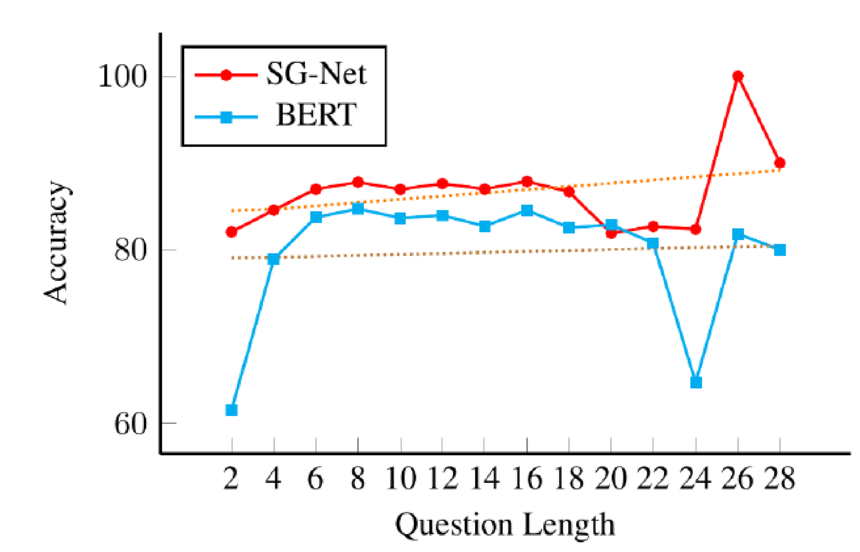
*Ablation on aggregation methods*

| Model | RACE-M | RACE-H | RACE |
|---|---|---|---|
| *Human Performance* | | | |
| Turkers | 85.1 | 69.4 | 73.3 |
| Ceiling | 95.4 | 94.2 | 94.5 |
| *Leaderboard* | | | |
| DCMN | 77.6 | 70.1 | 72.3 |
| $\text{BERT}_{\text{LARGE}}$ | 76.6 | 70.1 | 72.0 |
| OCN | 76.7 | 69.6 | 71.7 |
| Baseline | 78.4 | 70.4 | 72.6 |
| **SG-Net** | **78.8** | **72.2** | **74.2** |

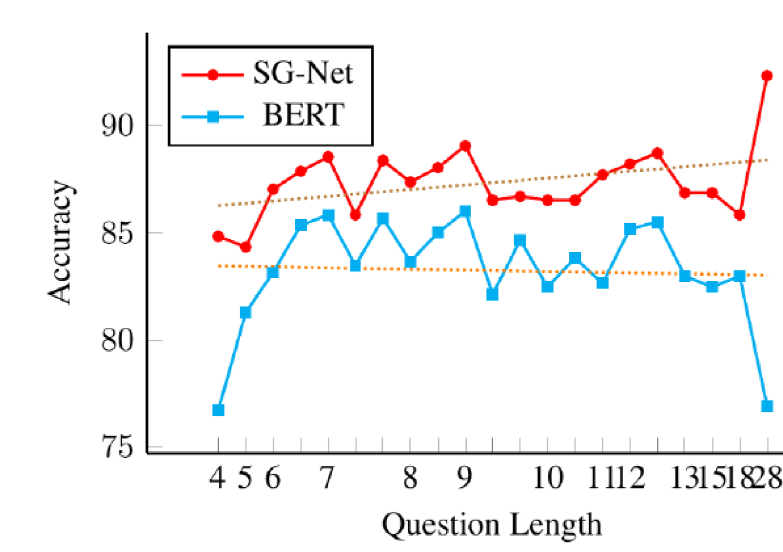
*Results for RACE*

**Results:**
- outperforms all the published works and achieves the 2nd place on the leaderboard when submitting SG-Net.
- adding an extra answer verifier module could yield better result.

**Effect of Answering Long Questions**
- the performance of the baseline drops heavily when encountered with long questions
- our proposed SG-Net works robustly, even showing positive correlation between accuracy and length.




(a) Split by equal range of question length




(b) Split by equal amount of questions