SMP2019中文人机对话

肖达，李童俊，方文达，钱胜杰，崔春来，李卓恒，李佳芮，高原

s1 黄昏，肃静的时光。我喜欢这种环境。

s2 有时间您也来逛逛吧

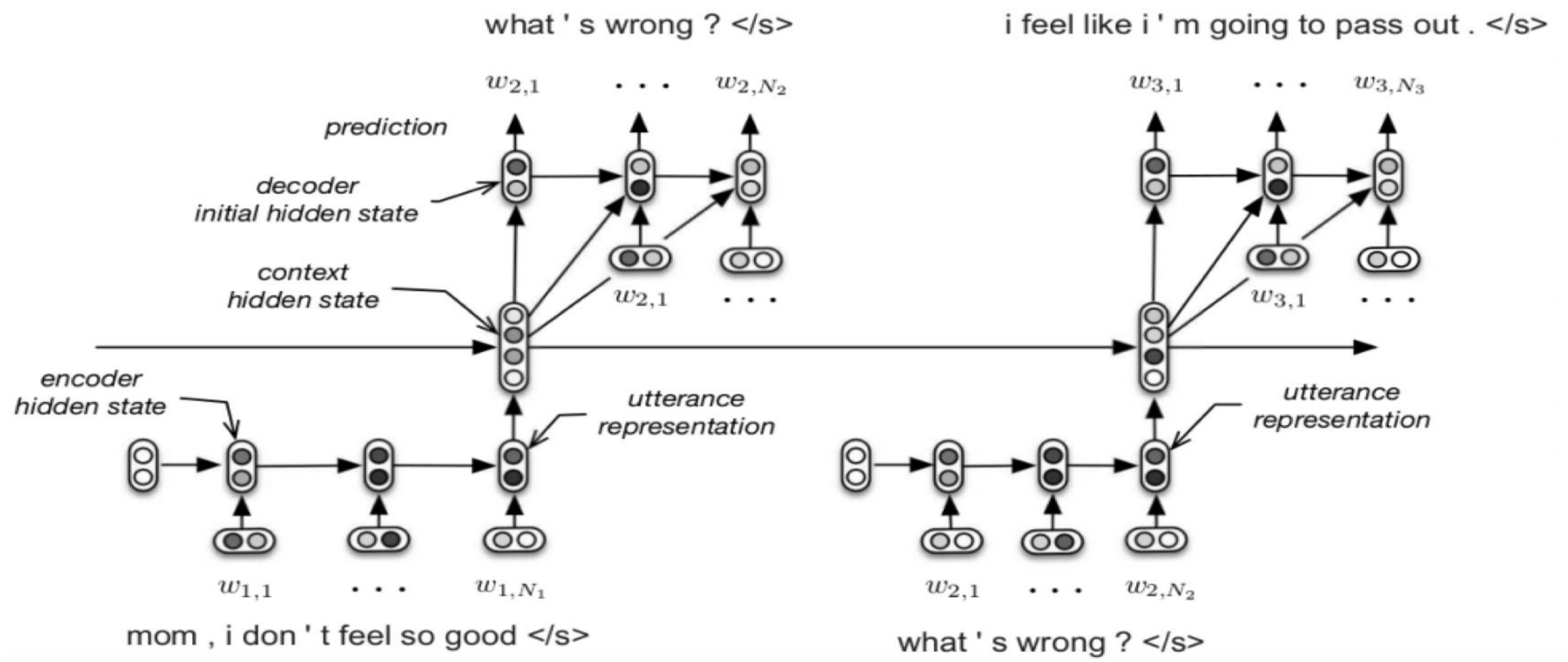
s1 好的，我都希望回中国。还希望下次在中国常驻工作时一个星期起马可以休息一天最好连休！这样就可以多看中国的美丽风景。

s2 那一定不是广本

p1:(性别：男，年龄：80后，地域：'海外 日本'，爱好标签：'重口味')

p2:(性别：女，年龄：90后，地域：'广东 广州'，爱好标签'音乐;自由;旅游;吃货')

S1和S2是对话信息，p1和p2是双方的个性化信息。模型在生成一个回复的时候，不仅需要考虑对话的上文信息，还需要考虑当前说话者的个人信息，例如图1中，在生成第三句话的时候就需要考虑当前说话者的地域信息。



encode RNN 把每个 dialog 变成 real-valued 的向量 u，context RNN 把每次对话里的 u 作为输入变成向量 c，最后把 c 交给 deocde RNN 生成下一个 dialog

预训练GPT（小说数据）

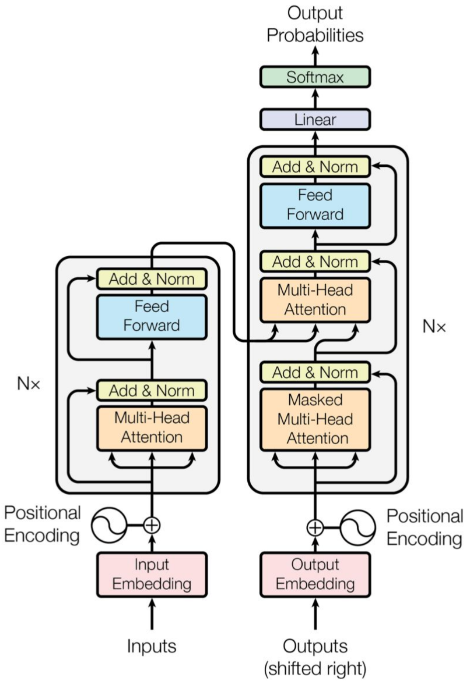
对话历史 word	男	上海	音乐	[SEP]	你	喜欢	什么	[EOS]	我	最爱	音乐	[EOS]
token type embedding	0	0	0	1	1	1	1	2	2	2	2	2
fuse position embedding	0	1	2	0	1	2	3	4	5	6	7	8

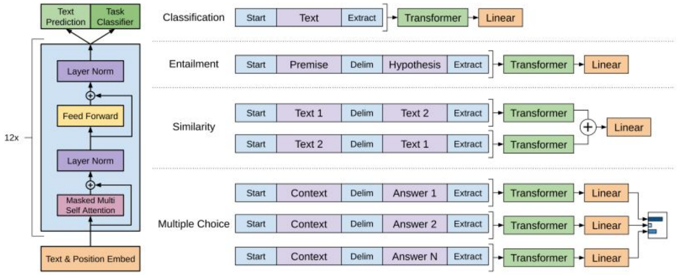
Top-P	Temperature	Bleu	Distinct
0.9	0.7	0.00529	0.243
0.9	1	0.002	0.378
TOP-K			
3	1	0.004417	0.1918
3	0.7	0.007	0.1886
3	0.9	0.0074	0.1885
5	1	0.00632	0.2166
Beam-size			
5	1	0.01	0.16
5	0.7	0.00959	0.1718
5	0.9	0.00996	0.1712
10	0.7	0.0101	0.1753

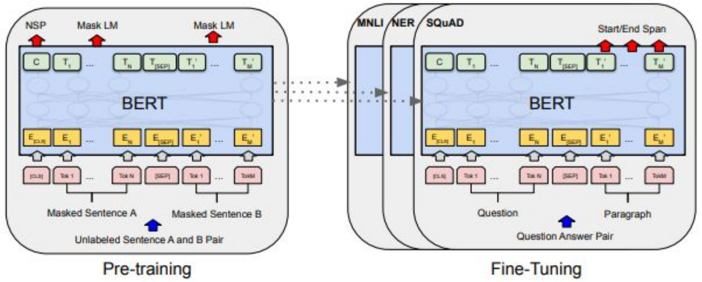
表 2

	Bleu	diversity	Ppl
V1	0.01063	0.1647	135.5
V2	0.0105	0.163	120.3
V3	0.0106	0.113	123.1
V1&V2&V3 Ensemble	0.0116	0.1429	107.5

Transformer-based Generation Model

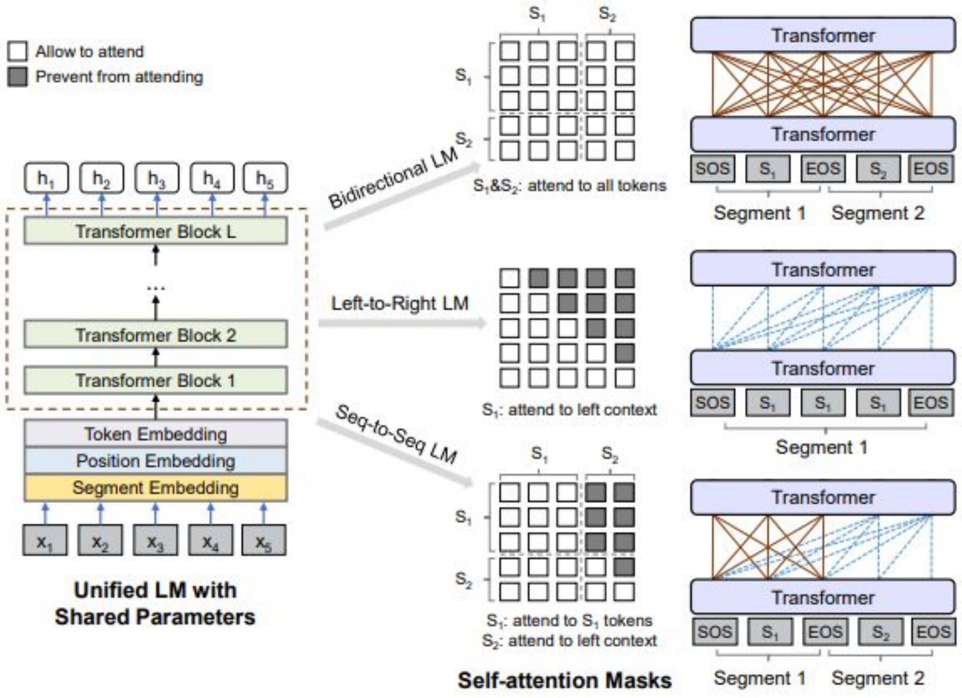






Transformer VS GPT VS BERT

UniLM



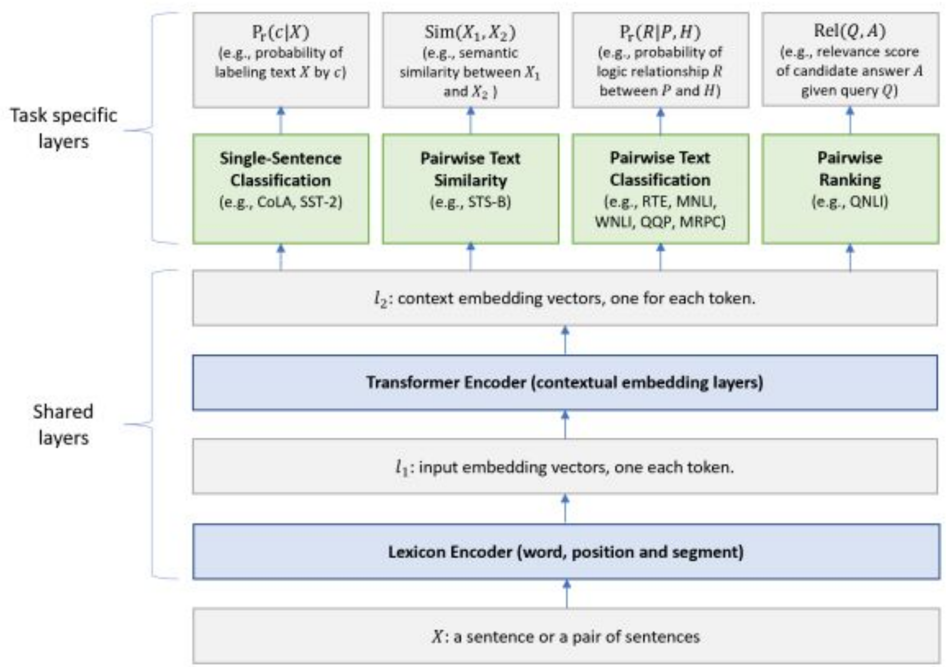
Unified Language Model Pre-training for Natural Language Understanding and Generation-https://arxiv.org/abs/1905.03197

Roberta

- RoBERT在BERT的基础上进行了更细致的探究，包括损失函数、超参数、数据量、训练方法等：
  - 使用动态mask
  - FULL-SENTENCES/DOC-SENTENCES without NSP
    - FULL-SENTENCES。直接从文档中连续sample句子直到塞满512的长度，且丢掉NSP。当到达一个文档结尾时，先增加一个文档分隔符，再从后面的文档接着sample。
    - DOC-SENTENCES。和上面一样，只是不得(may not)跨文档。同样增大batch size 使得总字符数和上面差不多。
  - 更大的batch size 和 training data
  - 更大的byte-level BPE
  - Optimizer参数调整

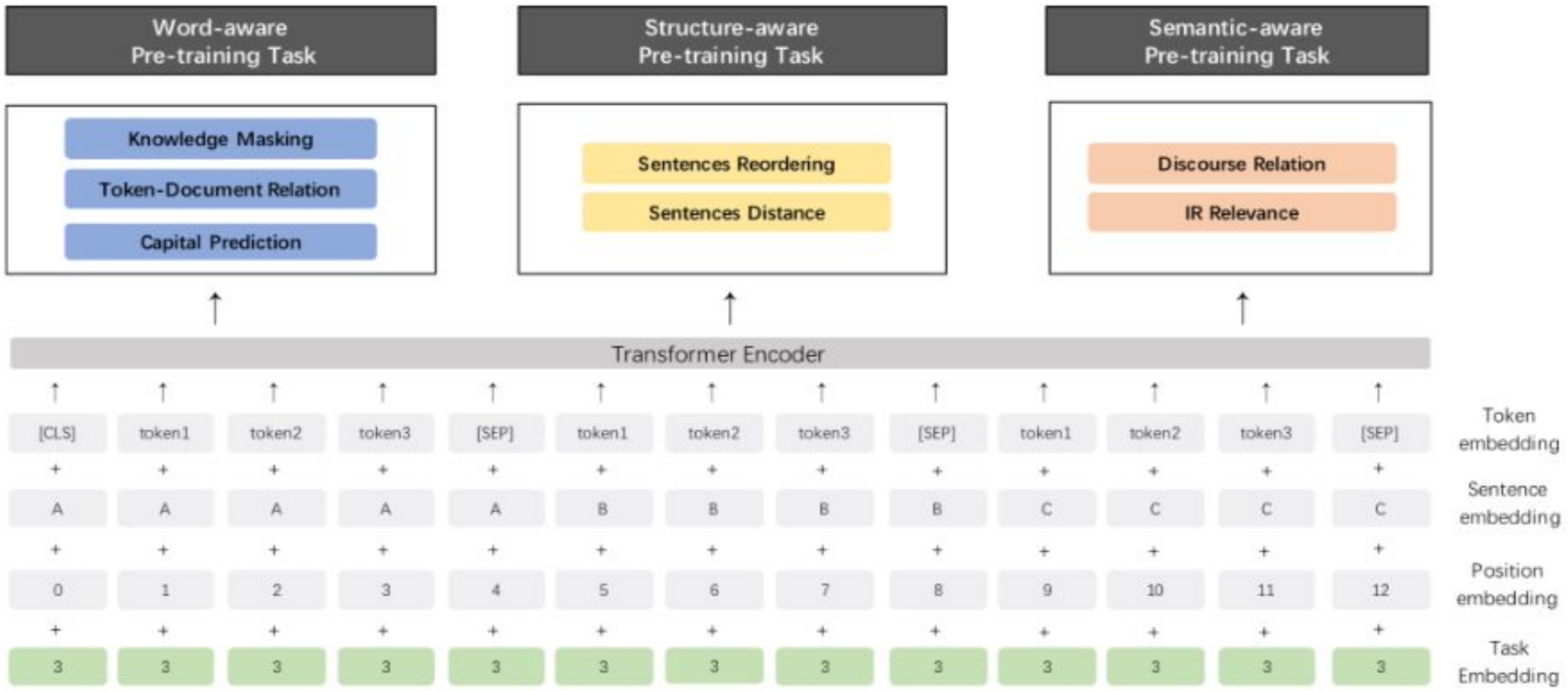
RoBERTa: A Robustly Optimized BERT Pretraining Approach- https://arxiv.org/abs/1907.11692

MT-DNN



- Multi-Task Deep Neural Networks for Natural Language Understanding-https://arxiv.org/abs/1901.11504

ERNIE 2



Word-aware Tasks

- Knowledge Masking Task　实体tokens
- Capitalization Prediction Task 预测每个词是否是大小写
- Token-Document Relation Prediction Task 判断一个词是否在其他的段落中出现

Structure-aware Tasks

- Sentence Reordering Task　预测被打乱的句子的原始顺序
- Sentence Distance Task　句子位置关系 (邻近句子、文档内非邻近句子、非同文档内句子)

Semantic-aware Tasks

- Discourse Relation Task　句对 (sentence pairs) 间的修辞关系
- IR Relevance Task　一个文档是否被用户点击

Cross-lingual Language Model

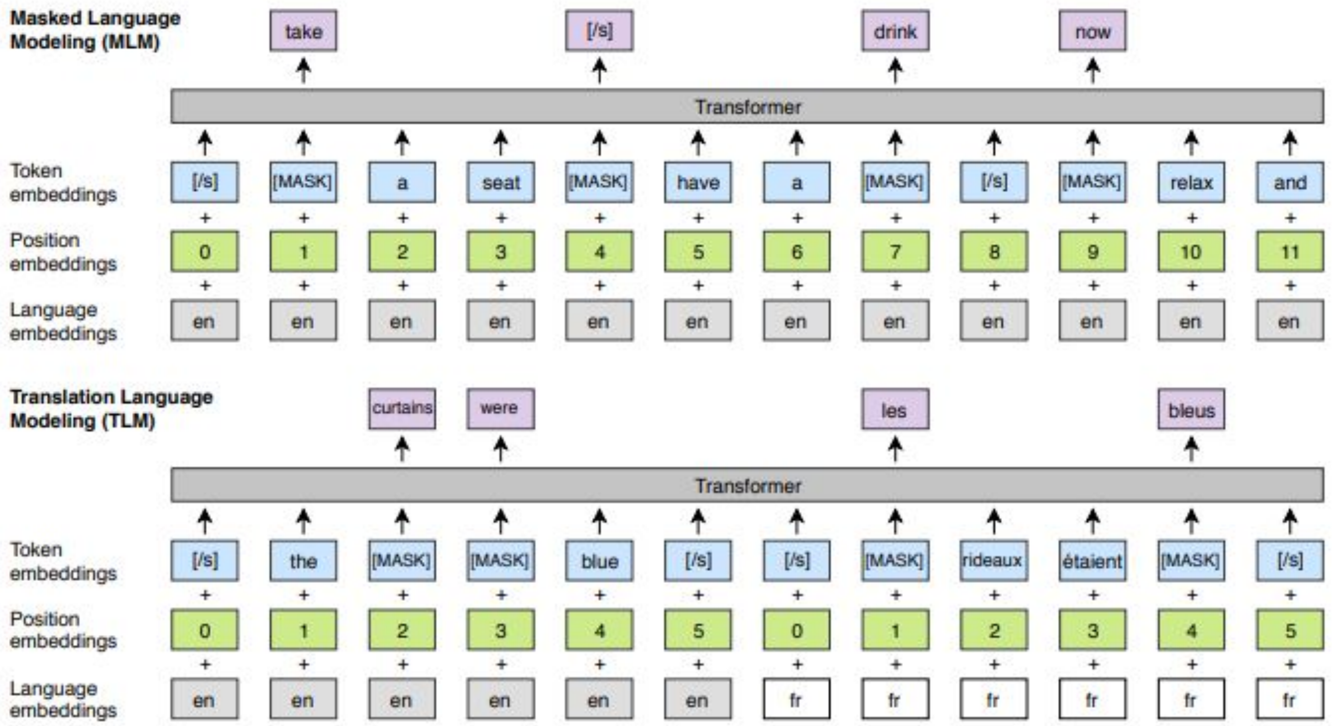


Figure 1: **Cross-lingual language model pretraining.** The MLM objective is similar to the one of Devlin et al. (2018), but with continuous streams of text as opposed to sentence pairs. The TLM objective extends MLM to pairs of parallel sentences. To predict a masked English word, the model can attend to both the English sentence and its French translation, and is encouraged to align English and French representations. Position embeddings of the target sentence are reset to facilitate the alignment.

MASS: Masked Sequence to Sequence Pre-training for Language Generation

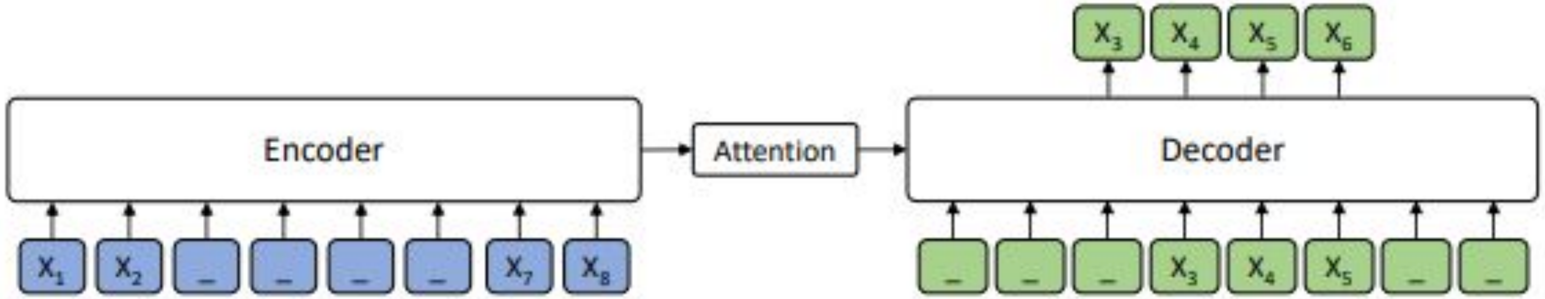


Figure 1. The encoder-decoder framework for our proposed MASS. The token "_" represents the mask symbol [M].

Thanks