

# 论文阅读笔记

## Step5

MF1833063, 史鹏, spwannasing@gmail.com

2019 年 7 月 3 日

# 1 ERNIE: Enhanced Language Representation with Informative Entities

此文章是对bert的一次扩展，提出了知识图谱中的多信息实体（informative entity）可以作为外部知识改善语言表征。

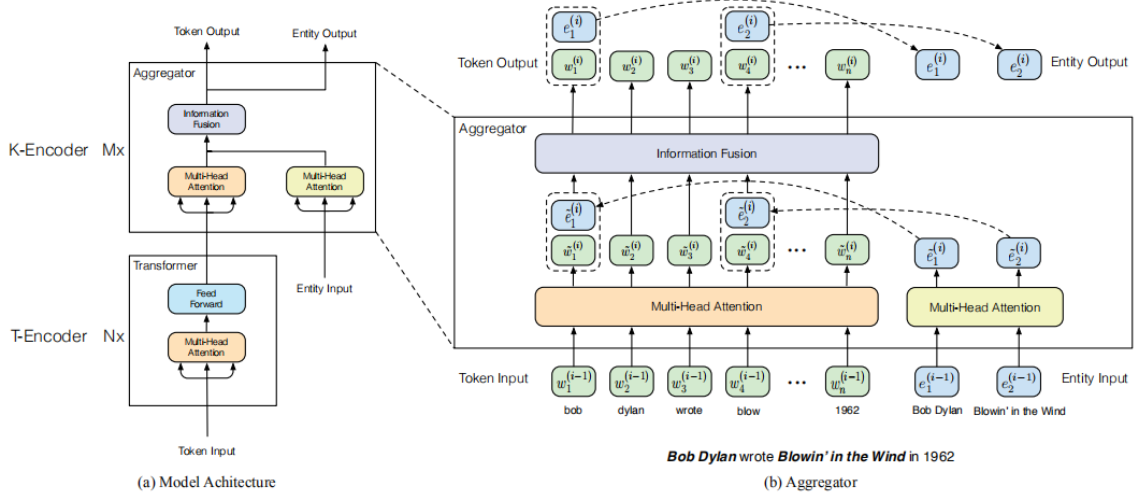


图 1: overview

**Mark Twain** wrote **The Million Pound Bank Note** in 1893.

Input for Common NLP tasks:

[CLS] [ ] mark twain [ ] wrote [ ] the million pound bank note [ ] in 1893 . [SEP]

Input for Entity Typing:

[CLS] [ENT] mark twain [ENT] wrote [ ] the million pound bank note [ ] in 1893 . [SEP]

Input for Relation Classification:

[CLS] [HD] mark twain [HD] wrote [TL] the million pound bank note [TL] in 1893 . [SEP]

图 2: finetune

Knowledgeable Encoder:

$$\{\tilde{w}_1^{(i)}, \dots, \tilde{w}_n^{(i)}\} = \text{MH-ATT}(\{w_1^{(i-1)}, \dots, w_n^{(i-1)}\}) \quad (1.1)$$

$$\{\tilde{e}_1^{(i)}, \dots, \tilde{e}_m^{(i)}\} = \text{MH-ATT}(\{e_1^{(i-1)}, \dots, e_m^{(i-1)}\}) \quad (1.2)$$

对于和entity对齐的token:

$$\begin{aligned} h_j &= \sigma(\tilde{\mathbf{W}}_t^{(i)} \tilde{w}_j^{(i)} + \tilde{\mathbf{W}}_e^{(i)} \tilde{e}_k^{(i)} + \tilde{\mathbf{b}}^{(i)}) \\ w_j^{(i)} &= \sigma(\mathbf{W}_t^{(i)} h_j + \mathbf{b}_t^{(i)}) \\ e_k^{(i)} &= \sigma(\mathbf{W}_e^{(i)} h_j + \mathbf{b}_e^{(i)}) \end{aligned} \quad (1.3)$$

else:

$$\begin{aligned} \mathbf{h}_j &= \sigma \left( \tilde{\mathbf{W}}_t^{(i)} \tilde{\mathbf{w}}_j^{(i)} + \tilde{\mathbf{b}}^{(i)} \right) \\ \mathbf{w}_j^{(i)} &= \sigma \left( \mathbf{W}_t^{(i)} \mathbf{h}_j + \mathbf{b}_t^{(i)} \right) \end{aligned} \quad (1.4)$$

对于引入的信息的pre-training目标:

$$p(e_j | w_i) = \frac{\exp(\text{linear}(\mathbf{w}_i^o) \cdot \mathbf{e}_j)}{\sum_{k=1}^m \exp(\text{linear}(\mathbf{w}_i^o) \cdot \mathbf{e}_k)} \quad (1.5)$$

## 2 ERNIE: Enhanced Representation through Knowledge Integration

ERNIE 通过建模海量数据中的词、实体及实体关系，学习真实世界的语义知识。相较于 BERT 学习原始语言信号，ERNIE 直接对先验语义知识单元进行建模，增强了模型语义表示能力。这里我们举个例子：

Learnt by BERT：哈 [mask] 滨是 [mask] 龙江的省会，[mask] 际冰 [mask] 文化名城。

Learnt by ERNIE：[mask] [mask] [mask] 是黑龙江的省会，国际 [mask] [mask] 文化名城。

在 BERT 模型中，我们通过『哈』与『滨』的局部共现，即可判断出『尔』字，模型没有学习与『哈尔滨』相关的任何知识。而 ERNIE 通过学习词与实体的表达，使模型能够建模出『哈尔滨』与『黑龙江』的关系，学到『哈尔滨』是『黑龙江』的省会以及『哈尔滨』是个冰雪城市。

训练数据方面，除百科类、资讯类中文语料外，ERNIE 还引入了论坛对话类数据，利用 DLM（Dialogue Language Model）建模 Query-Response 对话结构，将对话 Pair 对作为输入，引入 Dialogue Embedding 标识对话的角色，利用 Dialogue Response Loss 学习对话的隐式关系，进一步提升模型的语义表示能力。

我们在自然语言推断，语义相似度，命名实体识别，情感分析，问答匹配 5 个公开的中文数据集上进行了效果验证，ERNIE 模型相较 BERT 取得了更好的效果。

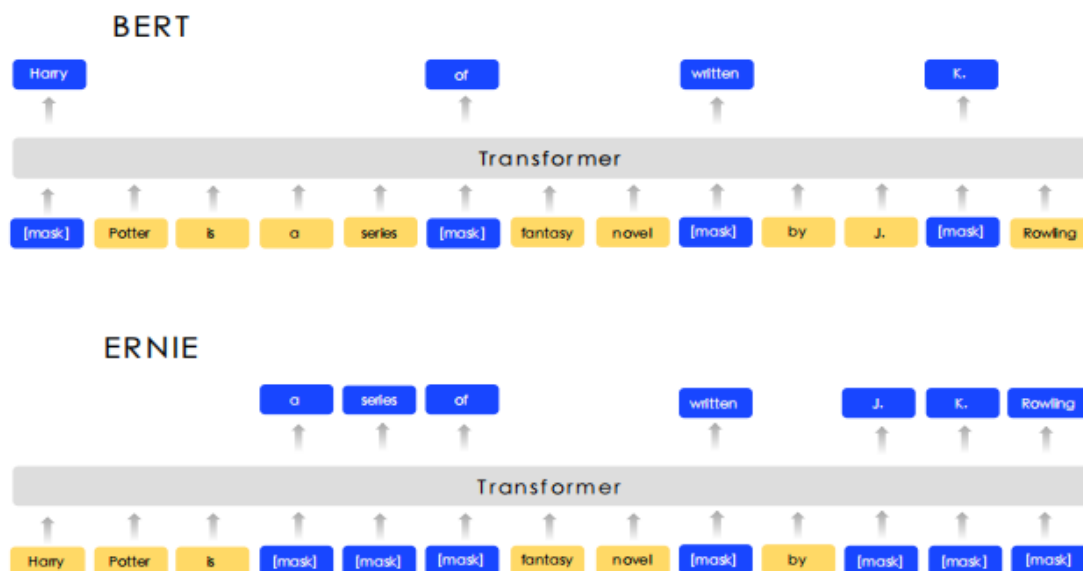


Figure 1: The different masking strategy between BERT and ERNIE

Sentence	Harry	Potter	is	a	series	of	fantasy	novels	written	by	British	author	J.	K.	Rowling
Basic-level Masking	[mask]	Potter	is	a	series	[mask]	fantasy	novels	[mask]	by	British	author	J.	[mask]	Rowling
Entity-level Masking	Harry	Potter	is	a	series	[mask]	fantasy	novels	[mask]	by	British	author	[mask]	[mask]	[mask]
Phrase-level Masking	Harry	Potter	is	[mask]	[mask]	[mask]	fantasy	novels	[mask]	by	British	author	[mask]	[mask]	[mask]

Figure 2: Different masking level of a sentence



Figure 3: Dialogue Language Model. Source sentence: [cls] How [mask] are you [sep] 8 . [sep] Where is your [mask] ? [sep]. Target sentence (words the predict): old, 8, hometown)

### 3 Learning to Ask Unanswerable Questions for Machine Reading Comprehension

提出一种数据增强技术，根据与包含答案的相应段落配对的可回答问题自动生成相关的无法回答的问题。所提出的结构为“pair-to-sequence”。

<b>Title:</b> Victoria (Australia)
<b>Paragraph:</b> ...Public schools, also known as state or government schools, are funded and run directly by the <b>Victoria Department of Education</b> . Students do not pay tuition fees, but some extra costs are levied. Private fee-paying schools include parish schools ...
<b>Ans. Question:</b> What organization runs <i>the public schools</i> in Victoria?
<b>UnAns. Question:</b> What organization runs <i>the waste management</i> in Victoria?
<b>(Plausible) Answer:</b> <b>Victoria Department of Education</b>

Figure 1: An example taken from the SQuAD 2.0 dataset. The annotated (plausible) answer span in the paragraph is used as a pivot to align the pair of answerable and unanswerable questions.

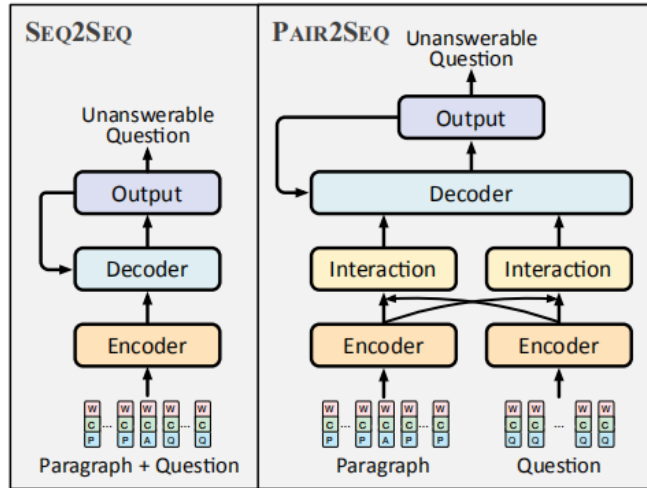


Figure 2: Diagram of the proposed pair-to-sequence model and sequence-to-sequence model. The input embeddings is the sum of the word embeddings, the character embeddings and the token type embeddings. The input questions are all answerable.

## 4 Course Concept Expansion in MOOCs with External Knowledge and Interactive Game

随着大规模在线开放课程(MOOC)的日益普及,为MOOC用户自动提供课外知识成为可能。语义漂移和知识缺乏在复杂的MOOC环境下,现有的方法不能有效地扩展课程概念。本文首先在通过外部知识库搜索新概念的过程中建立一个新的边界,然后利用异构特征来验证高质量的结果。具体算法没有特别仔细地看。

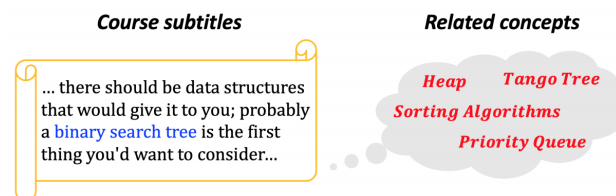


Figure 1: An example of “out-of-teaching” concepts in the course “*Data Structure and Algorithm*”.

## 5 XLNet: Generalized Autoregressive Pretraining for Language Understanding