ERNIE: Enhanced Representation through Knowledge Integration

百度 ERNIE, 与清华 ERNIE ERNIE-Enhanced-Language-Representation-with-Informative-Entities 同名

Motivation

现有的表征学习方法在学习 token embedding 时没有考虑语句中的先验知识，作者认为如果模型能够学习到先验知识，则可以获得更可靠的语言表征。

Methodology

ERNIE 使用 knowledge masking 策略隐式地学习先验知识，除了标准的随机 masking 方法之外，ERNIE 引入了两种特殊的 maksing 策略：phrase-level 和 entity-level，即分别将一个短语或一个实体作为一个基本单元进行 masking。
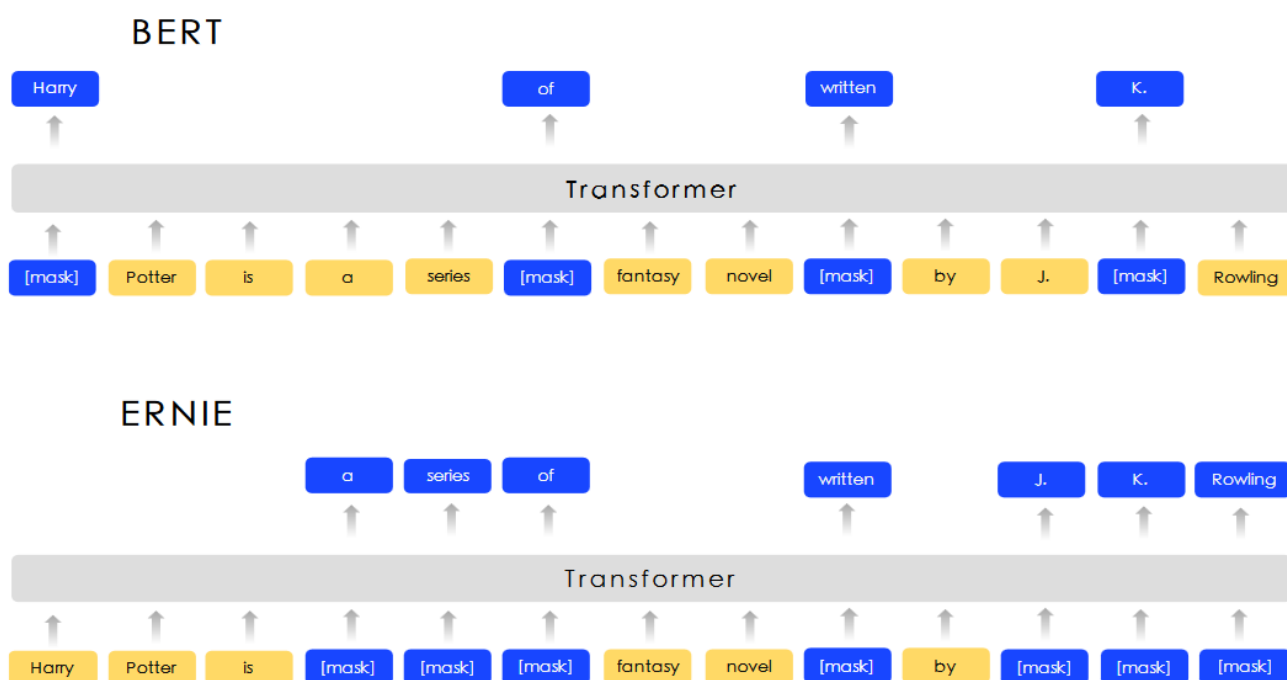

Figure 1: The different masking strategy between BERT and ERNIE

BERT as Backbone

语言表征的基础编码器采用和 BERT 等模型相同的多层 Transformer 结构。针对中文语料，作者将每个字用空格分开，并采用 WordPiece 进行 tokenize。每个 token 的输入 embedding 由 token embedding, segment embedding 和 position embedding 相加构成（与 BERT 相同），每个输入序列的第一个 token 是特殊的分类 token ($[CLS]$)（与 BERT 相同）

multi-stage knowledge masking strategy

Base-Level Masking 第一阶段将输入语句视为基本的语言单元的序列(对于英文，基本单元是词；对于中文，基本单元是字)。这一阶段直接使用 BERT 的 Masking 方法。作者认为此阶段难以对高阶语义知识建模。

Phrase-Level Masking 第二阶段将语句中的**短语**视为 Mask 的基本单元。随机选择句子中的一部分短语进行 Masking，然后预测被 Masking 的短语包含的所有基本语言单元。

Entity-Level Masking 第三个阶段从语句中分析出命名实体的存在(使用额外工具)，然后对实体进行 Masking 并预测实体所包含的所有语言单元。

实验

对于对话语言模型任务(DLM)，ERINE 使用 dialogue embedding 来标示对话中的角色。另外还构建了假样本用于训练模型分辨一组多轮对话数据是真的对话还是假的对话。

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)

No	Text	Predict by ERNIE	Predict by BERT	Answer
1	2006年9月，＿＿＿＿＿与张柏芝结婚，两人婚后育有两儿子——大儿子Lucas谢振轩，小儿子Quintus谢振南；	谢霆锋	谢振轩	谢霆锋
	In September 2006, _____ married Cecilia Cheung. They had two sons, the older one is Zhenxuan Xie and the younger one is Zhennan Xie.	Tingfeng Xie	Zhenxuan Xie	Tingfeng Xie
2	戊戌变法，又称百日维新，是＿＿＿＿＿、梁启超等维新派人士通过光绪帝进行的一场资产阶级改良。	康有为	孙世昌	康有为
	The Reform Movement of 1898, also known as the Hundred-Day Reform, was a bourgeois reform carried out by the reformists such as ____ and Qichao Liang through Emperor Guangxu.	Youwei Kang	Shichang Sun	Youwei Kang
3	高血糖则是由于＿＿＿＿＿分泌缺陷或其生物作用受损，或两者兼有引起。糖尿病时长期存在的高血糖，导致各种组织，特别是眼、肾、心脏、血管、神经的慢性损害、功能障碍。	胰岛素	糖糖内	胰岛素
	Hyperglycemia is caused by defective _____ secretion or impaired biological function, or both. Long-term hyperglycemia in diabetes leads to chronic damage and dysfunction of various tissues, especially eyes, kidneys, heart, blood vessels and nerves.	Insulin	(Not a word in Chinese)	Insulin
4	澳大利亚是一个高度发达的资本主义国家，首都为＿＿＿＿＿。作为南半球经济最发达的国家和全球第12大经济体、全球第四大农产品出口国，其也是多种矿产出口量全球第一的国家。	墨尔本	墨悉本	堪培拉
	Australia is a highly developed capitalist country with _____ as its capital. As the most developed country in the Southern Hemisphere, the 12th largest economy in the world and the fourth largest exporter of agricultural products in the world, it is also the world's largest exporter of various minerals.	Melbourne	(Not a city name)	Canberra (the capital of Australia)
5	＿＿＿＿＿是中国神魔小说的经典之作，达到了古代长篇浪漫主义小说的巅峰，与《三国演义》《水浒传》《红楼梦》并称为中国古典四大名著。	西游记	《小》	西游记
	_____is a classic novel of Chinese gods and demons, which reaching the peak of ancient Romantic novels. It is also known as the four classical works of China with Romance of the Three Kingdoms, Water Margin and Dream of Red Mansions.	The Journey to the West	(Not a word in Chinese)	The Journey to the West
6	相对论是关于时空和引力的理论，主要由＿＿＿＿＿创立。	爱因斯坦	卡尔斯所	爱因斯坦
	Relativity is a theory about space-time and gravity, which was founded by _____.	Einstein	(Not a word in Chinese)	Einstein