将知识融入 NLP 模型的方法总结

December 4, 2022

Abstract

在大规模预训练模型出现之后，NLP 开始越来越频繁地谈论“知识”。知识是模型获取常识、逻辑和其他外部信息的关键。文本对近几年 SOTA 工作将知识应用于 NLU、NLG 和常识推理的方法进行整理和总结。

1 知识的定义

近年来，通过结合更大的模型、更好的训练策略和更多的数据，NLP 的发展得到了极大的推动，如 BERT、RoBERTa 和 GPT。这些预训练模型可以有效地从文本中学到语言模式并生成高质量的、上下文感知的表征。然而，这些模型的输入只有原文本，因而很难获取到和概念、关系、常识相关的外部世界知识。本文用“知识”表示**对模型预测目标输出很重要但是在当前模型输入中缺失的外部信息**。知识应该被融入到模型的训练和推理中，因为知识对语言表征非常重要，它是实现高阶智能不可或缺的组成部分；其次，仅仅在输入文本上进行统计学习是无法获取到知识的。

2 知识增强的 NLU

NLU 任务的目的是在输入文本的基础上对词、短语、句子、篇章的属性做预测，比如情绪分析、命名实体识别和文本推理。从知识资源的维度可将知识增强 NLU 大致分为两类，一类是结构化知识（如知识图谱），另一类是非结构化知识（如文本语料）。

将结构化知识融入到 NLU 的工作又可分为两种，一种是基于概念或者实体嵌入的显式方法 [1, 2, 3, 4, 5]，一种是通过实体遮掩预测的隐式方法 [6, 7, 8, 9]。比如 ERNIE[1] 使用 TransE 在知识图谱上显式地预训练实

体嵌入，而 EAE[10] 将其作为模型参数来学习。KEPLER[9] 基于描述文本使用预训练模型来隐式地计算实体嵌入。最近，一些工作提出联合训练知识图谱模块和语言模型。比如 JAKET[11] 提出使用知识模块来生成文本中实体的嵌入，而用语言模型来生成知识图谱中实体和关系的上下文感知的初始嵌入。Yu 等人 [12] 和 Xu 等人 [13] 提出使用词典描述作为额外的知识源来做 NLU 和常识推理任务。

将非结构化知识融入 NLU 模型，一般需要一个文本检索模块来从知识语料中获取相关文本。使用非结构化知识有很多方法，尤其是对于开放领域的问答任务。比如 Lee 第一个通过 ICT(inverse cloze task) 来训练 retriever，然后联合训练 retriever 和 reader 用于开放领域的问答；DPR 通过监督学习训练 retriever 在开放领域问答上取得了更好的成绩；REALM 预测遮掩的包含重要实体的 span 来联合预训练 reader 和 retriever；KG-FiD 提出在检索阶段通过检索文章之间的结构化关系来过滤噪声文章。

3 知识增强的 NLG

NLG 的目标是从各种形式的语言或非语言数据（如文本数据、图像数据和结构化知识图谱）中生成人类语言的可理解文本。和 NLU 方法不同的是，NLG 方法一般都采用 encoder-decoder 框架，在生成过程中，如果在解码下一个 token 时引入知识会面临很多挑战。

当前将知识融入 NLG 模型的方法大概分为三种：

- 通过模型结构融入知识：知识关联的注意力机制，知识关联的拷贝/指针机制；
- 通过训练框架融入知识：posterior regularization, constraint-driven learning, semantic loss, knowledge-informed weak supervision；
- 通过推理方法融入知识，在推理时增加不同的知识约束来指导解码，比如 lexical constraints, task-specific objectives, global inter-dependency；

如果从知识源的维度划分，结构化的知识融入方法可以划分以下四种：

- 将预先计算的知识嵌入注入语言生成；
- 通过三元组信息将知识迁移到语言模型；
- 通过路径寻找策略来实现知识图谱推理；

- 用图神经网络来提高图嵌入。

非结构化的知识融入方法可以划分为以下两种：

- 用检索信息来指导生成；
- 将背景知识建模到文本生成中。

4 NLP 常识和推理

略。

5 文章列表

- 知识增强 NLU：[1],[2],[3],[14],[15],[11]
- 知识增强 NLG：[16],[17],[18],[19],[20]
- 常识和推理：[21],[22],[23],[20],[24],[25]
- 相关综述：[4],[26],[27],[28]

6 代表人物

1. Chenguang Zhu
2. Yichong Xu
3. Xiang Ren
4. Bill Yuchen Lin
5. Meng Jiang
6. Wenhao Yu

References

[1] Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, and Qun Liu. Ernie: Enhanced language representation with informative entities. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1441–1451, 2019.

[2] Matthew E Peters, Mark Neumann, Robert Logan, Roy Schwartz, Vidur Joshi, Sameer Singh, and Noah A Smith. Knowledge enhanced contextual word representations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 43–54, 2019.

[3] Weijie Liu, Peng Zhou, Zhe Zhao, Zhiruo Wang, Qi Ju, Haotang Deng, and Ping Wang. K-bert: Enabling language representation with knowledge graph. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 2901–2908, 2020.

[4] Wenhao Yu, Mengxia Yu, Tong Zhao, and Meng Jiang. Identifying referential intention with heterogeneous contexts. In *Proceedings of The Web Conference 2020*, pages 962–972, 2020.

[5] Qingkai Zeng, Wenhao Yu, Mengxia Yu, Tianwen Jiang, Tim Weninger, and Meng Jiang. Tri-train: Automatic pre-fine tuning between pre-training and fine-tuning for sciner. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4778–4787, 2020.

[6] Yu Sun, Shuohuan Wang, Yukun Li, Shikun Feng, Xuyi Chen, Han Zhang, Xin Tian, Danxiang Zhu, Hao Tian, and Hua Wu. Ernie: Enhanced representation through knowledge integration. *arXiv preprint arXiv:1904.09223*, 2019.

[7] Tao Shen, Yi Mao, Pengcheng He, Guodong Long, Adam Trischler, and Weizhu Chen. Exploiting structured knowledge in text via graph-guided representation learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8980–8994, 2020.

[8] Wenhan Xiong, Jingfei Du, William Yang Wang, and Veselin Stoyanov. Pretrained encyclopedia: Weakly supervised knowledge-pretrained language model. 2020.

[9] Xiaozhi Wang, Tianyu Gao, Zhaocheng Zhu, Zhengyan Zhang, Zhiyuan Liu, Juanzi Li, and Jian Tang. Kepler: A unified model for knowledge embedding and pre-trained language representation. *Transactions of the Association for Computational Linguistics*, 9:176–194, 2021.

[10] Thibault Févry, Livio Baldini Soares, Nicholas Fitzgerald, Eunsol Choi, and Tom Kwiatkowski. Entities as experts: Sparse memory access with entity supervision. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4937–4951, 2020.

[11] Donghan Yu, Chenguang Zhu, Yiming Yang, and Michael Zeng. Jaket: Joint pre-training of knowledge graph and language understanding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 11630–11638, 2022.

[12] Wenhao Yu, Chenguang Zhu, Yuwei Fang, Donghan Yu, Shuohang Wang, Yichong Xu, Michael Zeng, and Meng Jiang. Dict-bert: Enhancing language model pre-training with dictionary. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1907–1918, 2022.

[13] Yichong Xu, Chenguang Zhu, Ruochen Xu, Yang Liu, Michael Zeng, and Xuedong Huang. Fusing context into knowledge graph for commonsense question answering. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1201–1207, 2021.

[14] Ming Ding, Chang Zhou, Qibin Chen, Hongxia Yang, and Jie Tang. Cognitive graph for multi-hop reading comprehension at scale. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2694–2703, 2019.

[15] Shangwen Lv, Daya Guo, Jingjing Xu, Duyu Tang, Nan Duan, Ming Gong, Linjun Shou, Daxin Jiang, Guihong Cao, and Songlin Hu. Graph-based reasoning over heterogeneous external knowledge for commonsense question answering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8449–8456, 2020.

[16] Hao Zhou, Tom Young, Minlie Huang, Haizhou Zhao, Jingfang Xu, and Xiaoyan Zhu. Commonsense knowledge aware conversation generation with graph attention. In *IJCAI*, pages 4623–4629, 2018.

[17] Houyu Zhang, Zhenghao Liu, Chenyan Xiong, and Zhiyuan Liu. Grounded conversation generation as guided traverses in commonsense knowledge graphs. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2031–2043, 2020.

[18] Haozhe Ji, Pei Ke, Shaohan Huang, Furu Wei, Xiaoyan Zhu, and Minlie Huang. Language generation with multi-hop reasoning on commonsense knowledge graph. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 725–736, 2020.

[19] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474, 2020.

[20] Han Wang, Yang Liu, Chenguang Zhu, Linjun Shou, Ming Gong, Yichong Xu, and Michael Zeng. Retrieval enhanced model for commonsense generation. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3056–3062, 2021.

[21] Kaixin Ma, Jonathan Francis, Quanyang Lu, Eric Nyberg, and Alessandro Oltramari. Towards generalizable neuro-symbolic systems for commonsense question answering. In *Proceedings of the First Workshop on Commonsense Inference in Natural Language Processing*, pages 22–32, 2019.

[22] Zhihao Fan, Yeyun Gong, Zhongyu Wei, Siyuan Wang, Yameng Huang, Jian Jiao, Xuan-Jing Huang, Nan Duan, and Ruofei Zhang. An enhanced knowledge injection model for commonsense generation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2014–2025, 2020.

[23] Ye Liu, Yao Wan, Lifang He, Hao Peng, and S Yu Philip. Kg-bart: Knowledge graph-augmented bart for generative commonsense reason-

ing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 6418–6425, 2021.

[24] Jian Guan, Yansen Wang, and Minlie Huang. Story ending generation with incremental encoding and commonsense knowledge. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6473–6480, 2019.

[25] Jian Guan, Fei Huang, Zhihao Zhao, Xiaoyan Zhu, and Minlie Huang. A knowledge-enhanced pretraining model for commonsense story generation. *Transactions of the Association for Computational Linguistics*, 8:93–108, 2020.

[26] Jian Yang, Gang Xiao, Yulong Shen, Wei Jiang, Xinyu Hu, Ying Zhang, and Jinghui Peng. A survey of knowledge enhanced pre-trained models. *arXiv preprint arXiv:2110.00269*, 2021.

[27] Zhihan Zhang, Wenhao Yu, Mengxia Yu, Zhichun Guo, and Meng Jiang. A survey of multi-task learning in natural language processing: Regarding task relatedness and training methods. *arXiv preprint arXiv:2204.03508*, 2022.

[28] Xiaokai Wei, Shen Wang, Dejiao Zhang, Parminder Bhatia, and Andrew Arnold. Knowledge enhanced pretrained language models: A compreshensive survey. *arXiv preprint arXiv:2110.08455*, 2021.