**论文题目：Not All Countries Celebrate Thanksgiving: On the Cultural Dominance in Large Language Models**

**原敬闰\_2024.10.19\_论文笔记**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **基本信息** | 发表刊物 | The Association for Computational Linguistics | | 发表年份 | 2024 | 第一完成单位（国内） | 香港中文大学 |
| 作者 | Wenxuan Wang, Wenxiang Jiao, Jingyuan Huang, Ruyi Dai, Jen-tse Huang , Zhaopeng Tu, Michael R. Lyu | | | | | |
| 关键词（中文） | | 大语言模型，文化主导，训练，提示，文化多样性 | | | | |
| 关键词（英文） | | LLMs，the Cultural Dominance，Cultural diversity，train | | | | |
| **论文内容** | 解决的问题（如有实际应用场景请说明） | | 由于当前比较先进的大语言模型基本都是使用英文的训练数据，因此即使使用非英文来询问大模型，大模型的回答也都会偏向于英语世界的文化色彩，例如中文询问有哪些节日给出的回答会包含“圣诞节，感恩节，平安夜”等。 | | | | |
| 解决问题的方法（采用什么模型框架等） | | 解决方法分两种：   1. 使用更多样化的数据来训练大模型（本文使用ERNIE Bot） 2. 也可以提示大模型来基于哪一种文化来回答 | | | | |
| 仍旧存在的问题（注明论文中说明的问题或自己认为存在的问题） | | 本文认为的：1. 采用的语言种类并不特别多，不够全面。2. （对于较抽象的问题是否有文化偏见的时候）比较依赖现有的社会调查  我认为仍然存在的：本文提出的解决方案中第一种由于成本问题并不方便推广，第二种只是治标不治本且对抽象问题效果不好。 | | | | |
| **实验内容** | 实验采用的数据集 | | 本文研究数据集来源是世界价值观调查和政治坐标测试的各个国家测试结果以及多种大模型 | | | | |
| 数据集内容是否和待解决问题模型对应 | | 是的 | | | | |
| 实验是否涉及实际应用场景 | | 涉及，现实中使用大模型就会出现偏向英语文化的情况。 | | | | |
| 实验采用的对比方法 | | 分为具体和抽象两种，具体的让大模型列出一些实体（电影，等），然后判断是否符合提出的语言，抽象的则通过一些测试来确认其在价值分布图的坐标 | | | | |
| 实验任务 | | 尝试降低大模型中存在的文化主导力现象。 | | | | |
| 实验衡量指标 | | 具体的：通过看列出的结果是否和使用的语言相符合  抽象的：通过询问问题来得到其在价值分布空间中的坐标，将其与对应语言的人类参考结果和英语的人类参考结果的欧氏距离进行对比判断。 | | | | |
| 实验说明所提出方法的优点 | | 提示的方法比较省成本  采用多样化的数据来训练能够根本上解决问题 | | | | |
| **思考内容（阅读论文后自己思考填充）** | 论文的主要优点是什么 | | 敏锐的注意到了大模型的训练数据主导语言对非英语文化产生的文化入侵等社会学问题 | | | | |
| 论文仍然可以改进的地方是什么 | | 1. 首先可以采用更多的语言来测试 2. 其次可以研究数据之外对大模型的文化主导现象的因素 3. 提出的解决问题并不很实用，可能会由于成本问题不被使用 | | | | |
| 选择读这篇论文的原因是什么 | | 分配给我了… | | | | |
| 以此论文为出发点，如果需要你做一篇和其相关的顶会论文，你需要的资源是什么？数据，硬件，技术支持等 | | 需要高性能计算机和高速运算芯片，需要各种社会调查的结果。 | | | | |
| 所选这篇论文和目前自己在做的内容能够想到的相关点 | | 没有在做的项目… | | | | |
| 其他想要补充说明的内容 | | 无 | | | | |