

目录

1	引言	2
2	实现方法	2
2.1	指令微调	2
2.2	数据构造	3
2.3	函数调用	3
2.4	规则匹配	4

智能采购助手

摘要

采购员在采购过程中有时没有明确的采购物料型号，但有明确的采购物料要求，例如物料的工作场景，关键技术属性等。此时采购员难以在工业品平台中快速搜索到他所需要的物料。我们希望制作一个了解工业品知识的采购助手，能够直接解析用户的采购需求，直接根据需求推荐满足他们需求的 SKU。我们尝试利用大语言模型强大的对话能力和上下文信息提取能力，实现大模型与采购员的关于物料的专业对话，并从对话提取采购员要求的物料信息，最后根据提取的物料信息进行具体商品推荐。完成项目之后，最终效果表明以自然语言大模型为搜索引擎的方案能够准确有效的对采购员的需求进行提取，这种搜索方式必将成为传统搜索方式的一个强有力的替代方案。

引言

随着互联网和数字化技术的迅速发展，大数据变得越来越普遍。这为使用大模型进行信息检索和知识管理提供了基础。用户层面对搜索引擎的期望不断提高，不仅仅满足于基本的关键词搜索，更期望搜索引擎能够理解其搜索意图，提供更加个性化，更加智能的搜索服务。而自然语言大模型凭借其强大的文本理解以及文本生成能力，使得能更好地理解复杂的查询意图和语境，提供更加准确和深入的用户意图分析，使得搜索过程朝向智能化发展。

实现方法

2.1 指令微调

我们使用指令微调 [1] 的作为核心方法，如图 1所示，我们期望通过该方法使大模型可以扮演客服与用户进行引导性对话，也就是产生符合上下文内容的输出信息。同时期望大模型拥有一些常见的工业品的基本常识，并在对话中有所体现。在对话结束之后，

模型可以将对话中的用户采购需求进行结构化的总结并输出，并要求模型具有一定鲁棒性，从而能够较好的适应实际的应用场景。

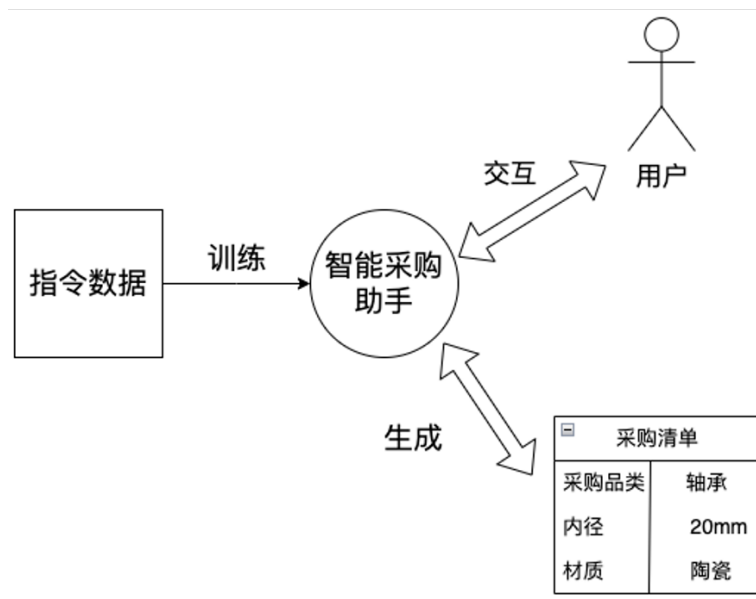


图 1: 指令微调

2.2 数据构造

由于缺乏工业品销售的知识背景，且没有可靠的训练数据来源，如图 2 我们采用对先进大模型进行知识提取，来产生尽可能贴近现实的工业品采购对话数据。我们提取欧

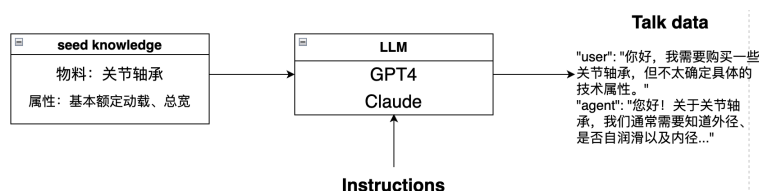


图 2: 知识提取

贝杯提供商品数据作为 seed knowledge 来保证商品类型的多样性。通过人工编写工业品对话数据，使用 few shot 方法来让模型生成的数据更有连续性和逻辑性，同时我们选用尽可能强大的大语言模型作为蒸馏对象，来保证整个数据的质量。

2.3 函数调用

由于现实对话情况复杂等原因，模型不能稳定的按照客户的意图来触发采购清单总结功能，如图 3 我们借鉴大模型 tool 调用 [2] 的方法，将总结过程抽象为函数参数提取

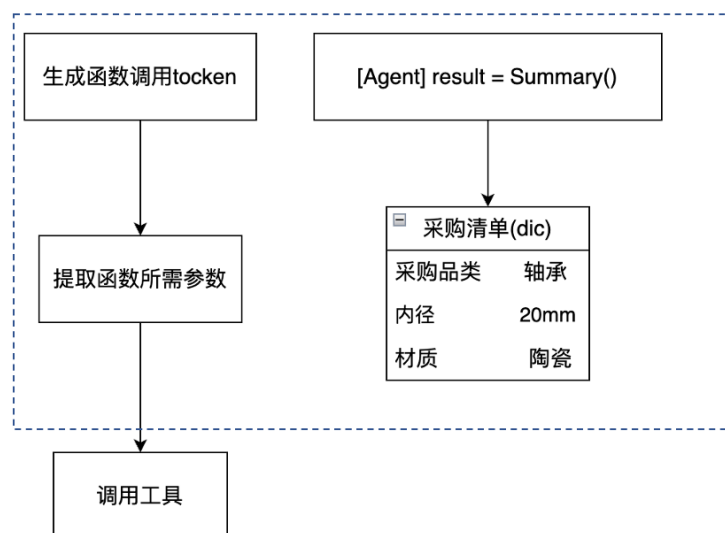


图 3: 工具调用

过程，从而实现总结过程的稳定触发。实验表明，由于对应的训练数据较少，对于较小的模型而言，已经拥有 tool 调用能力的模型能展现出良好的泛化能力，对于中模型而言，无论是否拥有函数调用能力，都能拥有非常好的清单提取能力。

2.4 规则匹配

虽然大模型已经能够实现强大的采购清单总结能力，理想情况下简单的基于规则的属性匹配算法就可以实现就好的推荐功能，但考虑到现实对话场景的复杂情况，如图 4所示，错别字、同义词等情况始终存在，我们采用 text2vector 的思想，将不同的词组转换为向量，基于语义相似性来增强规则匹配的鲁棒性。

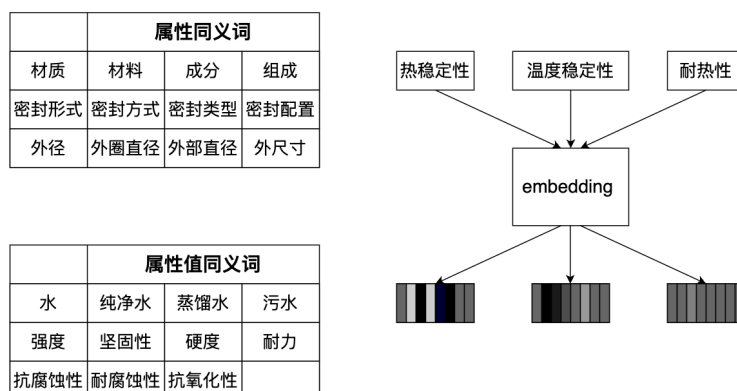


图 4: 同义词现象 (左) 与 text2vector(右)

参考文献

- [1] Swaroop Mishra, Daniel Khashabi, Chitta Baral, and Hannaneh Hajishirzi. Cross-task generalization via natural language crowdsourcing instructions. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3470–3487, Dublin, Ireland, May 2022. Association for Computational Linguistics.
- [2] Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language models can teach themselves to use tools, 2023.