

大语言模型对比分析报告

1. 研究背景与目的

本报告对三个大语言模型（Baichuan2-7B-Chat、chatg1m3-6b和Qwen-7B-Chat）进行全面对比分析，通过五个具有挑战性的问题测试它们在语义理解、歧义处理、逻辑推理、语言幽默和多义词理解等方面的能力差异。此分析旨在揭示不同架构和训练方法的大语言模型在处理中文复杂语言现象时的表现特点。

2. 模型基本情况

2.1 Baichuan2-7B-Chat

- 开发机构：**百川智能
- 参数规模：**7B参数
- 训练数据：**大规模中文语料库
- 特点：**针对中文对话场景优化，具有较强的通用对话能力

2.2 chatg1m3-6b

- 参数规模：**6B参数
- 特点：**相对较小的参数规模，但在特定任务上表现出色

2.3 Qwen-7B-Chat

- 开发机构：**阿里云通义千问
- 参数规模：**7B参数
- 训练数据：**大规模多语言语料库，中文内容丰富
- 特点：**擅长多轮对话，对语境理解较为敏感

3. 测试问题分析

3.1 问题设计思路

本次测试精心设计了五个问题，每个问题都针对语言理解的不同维度，考察模型的综合能力：

- 相同句式不同语境的理解（冬天/夏天能穿多少穿多少）
- 双关语和歧义句的处理（单身狗产生原因）
- 复杂逻辑关系的推理（知道不知道的递归问题）
- 同音字和名字歧义的理解（明明白白的歧义句）
- 多义词在对话中的语用分析（"意思"的多种含义）

4. 能力对比详细分析

4.1 语义理解能力

测试问题：冬天，能穿多少穿多少 vs 夏天，能穿多少穿多少

Baichuan2-7B-Chat

- **回答内容：**指出两句话描述不同季节的穿衣原则，冬天建议穿更多衣物保暖，夏天建议穿较少衣物散热
- **分析：**完全从字面意思理解，未能捕捉到语言中的幽默和双关
- **评分：**★★☆☆☆ (2/5分)
- **特点：**实用主义解读，缺乏对语言灵活性的理解

chatg1m3-6b

- **回答内容：**简单指出两句话关于不同季节，虽然表达形式相同但针对的季节不同
- **分析：**同样停留在表面理解，未能识别出语言的双关特性
- **评分：**★★☆☆☆ (2/5分)
- **特点：**回答简洁但浅显，未深入分析

Qwen-7B-Chat

- **回答内容：**深入分析了两句话的语境和语气差异，认为第一句是实用建议，第二句可能是质疑或批评
- **分析：**展现出对语境的敏感性，尝试理解句子背后的意图
- **评分：**★★★★☆ (4/5分)
- **特点：**能够理解语境差异，但未完全把握双关语的幽默本质

4.2 歧义处理能力

测试问题：单身狗产生的原因有两个，一是谁都看不上，二是谁都看不上

Baichuan2-7B-Chat

- **回答内容：**给出了与问题无关的回答，提到单身原因是缺乏自信和社交技能不足
- **分析：**完全未能识别问题中的双关语，给出了通用但不相关的回答
- **评分：**★☆☆☆☆ (1/5分)
- **特点：**在面对歧义句时容易偏离问题核心

chatg1m3-6b

- **回答内容：**成功识别出"谁都看不上"的双重含义，解释了主动（自己看不上别人）和被动（别人看不上自己）的区别
- **分析：**准确把握了双关语的精髓，展现出较强的语言理解能力
- **评分：**★★★★★ (5/5分)
- **特点：**对语言歧义有敏锐的洞察力

Qwen-7B-Chat

- **回答内容：**未能正确理解问题，错误地认为两个句子有不同语义，并提到主动语态和被动语态的差异
- **分析：**未能识别出双关语，反而引入了不存在的差异
- **评分：**★★☆☆☆ (2/5分)
- **特点：**在复杂歧义面前容易产生误解

4.3 逻辑推理能力

测试问题：他知道我知道你知道他不知道吗？这句话里，到底谁不知道的事情是什么？

Baichuan2-7B-Chat

- **回答内容：**"在这个句子中，我们知道(我们)知道(我们)知道(某人不知道)。所以，某人不知道"
- **分析：**回答简短且不完整，未能清晰解析复杂的逻辑关系
- **评分：**★★☆☆☆ (2/5分)
- **特点：**在处理复杂逻辑时显得力不从心

chatg1m3-6b

- **回答内容：**将问题定义为绕口令，并明确指出"到底谁不知道的事情是'他不知道的事情'"
- **分析：**给出了较为清晰的解释，展现了对递归逻辑的理解能力
- **评分：**★★★★☆ (4/5分)
- **特点：**能够处理一定复杂度的逻辑关系

Qwen-7B-Chat

- **回答内容：**提出了多种可能性解释，但未给出确定答案，表现出对问题的不确定性
- **分析：**回答冗长，试图通过列举多种可能性来规避直接回答
- **评分：**★★★☆☆ (3/5分)
- **特点：**在面对不确定性时倾向于提供多种可能性，而非给出确定答案

4.4 语言歧义理解

测试问题：明明喜欢白白，明明明白白白喜欢他，可她就是不说。这句话里，明明喜欢白白。

Baichuan2-7B-Chat

- **回答内容：**"从这句话中，我们无法判断明明和白白之间谁喜欢谁。这句话主要表达了作者对明明和白白的喜欢不表达的不满"
- **分析：**未能正确理解人名歧义，错误地认为无法判断关系
- **评分：**★★☆☆☆ (2/5分)
- **特点：**在处理人名歧义时容易产生混淆

chatg1m3-6b

- **回答内容：**"这句话里,明明喜欢白白。明明白白白喜欢他,可她就是不说。"
- **分析：**准确识别出"明明"和"白白"作为人名的用法，理解了句子的核心含义
- **评分：**★★★★★ (5/5分)
- **特点：**对人名歧义有较强的理解能力

Qwen-7B-Chat

- **回答内容：**"这句话里，明明喜欢白白。"
- **分析：**回答简洁准确，但未提供详细解释
- **评分：**★★★★☆ (4/5分)
- **特点：**能够正确理解核心含义，但缺乏详细分析

4.5 多义词理解能力

测试问题：关于"意思"一词在对话中的多种含义分析

Baichuan2-7B-Chat

- **回答内容：**系统地分析了"意思"在不同语境下的5种不同含义，包括内涵、赞赏、不解、谦虚和承认
- **分析：**展现出对多义词在不同语境下的理解能力，分析有条理
- **评分：**★★★★☆ (4/5分)
- **特点：**善于结构化分析语言现象

chatg1m3-6b

- **回答内容：**详细分析了对话中8处"意思"的具体含义，解释最为全面
- **分析：**对每个"意思"都给出了具体语境下的解释，展现出对语用的深入理解
- **评分：**★★★★★ (5/5分)
- **特点：**对多义词的语境分析最为细致全面

Qwen-7B-Chat

- **回答内容：**分析了6种"意思"的含义，并尝试解释对话中的误解
- **分析：**除了分析词义，还尝试理解对话中的误解，展现出对对话动态的理解
- **评分：**★★★★☆ (4/5分)
- **特点：**关注词义与对话动态的结合

5. 综合能力评估

5.1 各维度能力分析

各模型在各能力维度的得分表

能力维度	Baichuan2-7B-Chat	chatg1m3-6b	Qwen-7B-Chat
语义理解	★★☆☆☆	★★☆☆☆	★★★★☆

能力维度	Baichuan2-7B-Chat	chatg1m3-6b	Qwen-7B-Chat
歧义处理	★☆☆☆☆	★★★★★	★★★☆☆
逻辑推理	★★☆☆☆	★★★★☆	★★★★☆
语言歧义	★★☆☆☆	★★★★★	★★★★☆
多义词理解	★★★★☆	★★★★★	★★★★☆
总分	11/25	21/25	17/25

5.2 模型优势与劣势分析

Baichuan2-7B-Chat

- 优势：
 - 在结构化分析方面表现较好，如多义词分析
 - 回答简洁明了，不冗余
 - 在实用性解释方面较为合理
- 劣势：
 - 对双关语和语言歧义的理解能力有限
 - 在复杂逻辑问题上表现不足
 - 对语言幽默感的把握较弱
- 适用场景：
 - 实用性问答
 - 结构化信息提取
 - 简明扼要的回答需求

chatg1m3-6b

- 优势：
 - 对语言歧义和双关语的理解能力极强
 - 解释详细且有条理
 - 在处理人名歧义和多义词方面表现出色
- 劣势：
 - 有时过于关注字面意思
 - 在某些问题上缺乏深层语境理解
 - 参数规模相对较小，但并未明显影响性能
- 适用场景：
 - 语言游戏和双关语理解
 - 文学作品分析
 - 需要精确理解语言歧义的场合

Qwen-7B-Chat

- **优势：**
 1. 对语境和语气的敏感度高
 2. 在面对不确定性时会提出多种可能性
 3. 语言理解的全面性较好
- **劣势：**
 1. 回答有时冗长
 2. 在某些问题上缺乏确定性
 3. 在处理双关语方面不如chatg1m3-6b
- **适用场景：**
 1. 需要考虑语境的复杂对话
 2. 多角度分析问题
 3. 语气和情感分析

6. 结论

根据本次测试结果，三个模型在不同类型的语言理解任务上表现各异：

- **chatg1m3-6b**在总体表现上最为出色，特别是在处理语言歧义和双关语方面，展现出超越参数规模的理解能力
- **Qwen-7B-Chat**在语境理解和多角度分析方面表现较好，适合需要考虑语境的复杂对话
- **Baichuan2-7B-Chat**在结构化分析和简洁表达方面有一定优势，但在处理语言的灵活性和幽默感方面有待提升

7. 附录：原始回答数据

**问题1：请说出以下两句话区别在哪里？ 1、冬天，能穿多少穿多少
2、夏天，能穿多少穿多少**

Baichuan2-7B-Chat

这两句话的区别在于，它们分别描述了不同季节的穿衣原则。在冬天，建议穿更多的衣物来保暖，而在夏天，建议穿较少的衣物来散热。所以，这两句话的核心区别在于季节和气候条件对穿衣需求的影响。

chatg1m3-6b

这两句话的区别在于，第一句是关于冬天的，而第二句是关于夏天的。虽然两句话都在描述穿多少衣服，但是它们所针对的季节不同。

Qwen-7B-Chat

这两句话的主要区别在于它们的语境和语气。

第一句话"冬天:能穿多少穿多少"是一种实用建议，旨在提醒人们在寒冷的冬季要注意保暖，尽可能多的穿上衣服。这句话的语气比较平和，没有太多的情感色彩。

第二句话"夏天，能穿多少穿多少"则是一种疑问或者质疑的语气，它可能暗示着说话者对于某个行为或决定感到困惑或者不满意。这句话可能有强调的意思，表示某种过度的行为或者不合理的决定。

总的来说，第一句话是在提出建议，而第二句话则是在表达疑虑或者批评。

问题2：单身狗产生的原因有两个，一是谁都看不上，二是谁都看不上

Baichuan2-7B-Chat

单身狗的产生原因主要是由于缺乏自信和社交技能不足。

chatg1m3-6b

这两个句子在表达意思上有一定的区别。

第一句话'单身狗产生的原因有两个，一是谁都看不上，二是谁都看不上。'中，第一个"谁"是指代"人"，表示单身狗的原因是单身狗自己看不上别人，同时别人也看不上单身狗。第二个"谁"是指代"人"，表示单身狗的原因是别人看不上单身狗，同时单身狗也看不上别人。第二句话 单身狗产生的原因有两个，一是谁都看不上，二是谁都看不上。'中，两个"谁"都指代"人"，表示单身狗的原因是单身狗自己看不上别人，同时别人也看不上单身狗。但是，第一句中的"谁"是主动的，自己看不上别人，而第二句中的"谁"是被动的，别人看不上自己。

Qwen-7B-Chat

这两个句子的语义有很大的不同。第一个句子描述了单身狗产生的两个原因，而第二个句子则只提到了一个原因，即"谁都看不上"。因此，第二个句子更加简洁和直接，而第一个句子则提供了更多的细节和解释。同时，这两个句子也存在一些语言上的差异，例如第一个句子使用了被动语态，而第二个句子则使用了主动语态。

问题3：他知道我知道你知道他不知道吗？这句话里，到底谁不知道的事情是什么？

Baichuan2-7B-Chat

在这个句子中，我们知道(我们)知道(我们)知道(某人不知道)。所以，某人不知道

chatg1m3-6b

这句话是一个典型的谜语或者绕口令,它的含义是:一个人知道另外一个人知道某些事情,但是这个人自己并不知道这些事情。

在这个绕口令中,到底谁不知道的事情是'他不知道的事情'。因为这个人知道自己知道某些事情,但是并不知道不知道自己不知道某些事情。

Qwen-7B-Chat

这是一道涉及到逻辑推理的问题。在没有提供更多的上下文信息之前，我们无法确定具体的情况。

一种可能性是:A知道B知道C不知道的事情。在这种情况下，如果A告诉了B，那么B也知道这个事情;而如果A没有告诉B，那么B就不知道这个事情。另一种可能性是:B知道A和C都不知道的事情。在这种情况下，如果B告诉了A，那么A就知道了这个事情;如果B没有告诉A，那么A就不知道这个事情。还有一种可能性是:A和C都不知道B知道的事情。在这种情况下，如果B告诉了A，那么A就会知道这个事情;如果B没有告诉A，那么A就还不知道这个事情。因此，我们需要更多的上下文信息来确定具体的事实情况。如果你能提供更多的情境描述或者相关信息，我会很乐意帮你解答这个问题的。

问题4：明明喜欢白白，明明明白白白喜欢他，可她就是不说话。这句话里，明明喜欢白白。

Baichuan2-7B-Chat

从这句话中，我们无法判断明明和白白之间谁喜欢谁。这句话主要表达了作者对明明和白白的喜欢不表达的不满

chatg1m3-6b

这句话里,明明喜欢白白。明明明白白白喜欢他,可她就是不说话。

Qwen-7B-Chat

这句话里，明明喜欢白白。

问题5：领导：你这是什么意思？小明：没什么意思。意思意思。领导：你这就不够意思了。小明：小意思，小意思。领导：你这人真有意思。小明：其实也没有别的意思。领导：那我就不好意思了。小明：是我不好意思。

Baichuan2-7B-Chat

在这段对话中，"意思"这个词在不同的语境下有不同的含义。以下是每个"意思"的含义：

- 1.第一个"意思"：表示某种行为或态度的内涵、意义或者目的。例如："你这是什么意思？"（你这种行为或态度是什么意思？）
- 2.第二个"意思"：表示对某人或某事的赞赏或认可。例如："你这就不够意思了。"（你这样做就不够朋友了）
- 3.第三个"意思"：表示对某人或某事的不解或困惑。例如："你这人真有意思。"（你这个人的行为真的让人不解）
- 4.第四个"意思"：表示谦虚或者自我贬低。例如："其实也没有别的意思。"（其实我并没有其他的意思）
- 5.第五个"意思"：表示承认对方的观点或者要求。例如："那我就不好意思了。"（那我也不客气了，接受你的要求）总的来说，这段对话中的"意思"这个词在不同情境下表达了不同的含义，包括内涵、赞赏、不解、谦虚和承认等等。

chatg1m3-6b

- 1、领导:你这是什么意思?-意思:这里的意思是指含义、目的或作用。领导在问小明到底是什么意思。
- 2、小明:没什么意思。意思意思。一意思:这里的意思是表示没有特别的含义、目的或作用。小明在回答领导的问题，表示自己的回答没有什么特别的意义。
- 3、领导:你这就不够意思了。-意思:这里的意思是指不够充分、不够满足。领导认为小明的回答不够充分，表示不满。
- 4、小明:小意思，小意思。-意思:这里的意思是表示事情不太大，没有太大意义。小明在回应领导的问题，表示自己的回答只是一个小事，没有太大的意义。
- 5、领导:你这人真有意思。-意思:这里的意思是表示对某人的行为、性格或想法表示赞赏。领导认为小明很有趣，很有意思。

6、小明：其实也没有别的意思。-意思:这里的意思是表示自己的回答确实没有什么特别的含义、目的或作用。小明在回应领导的问题，表示自己的回答并没有其他特别的意义。

7、领导：我就不好意思了。一意思:这里的意思是表示对某事或某人的态度。领导在这里表示自己不好意思了，可能是因为对小明的回答感到意外或困惑。

8、小明：是我不好意思。-意思:这里的意思是表示自己感到抱歉或尴尬。小明在回应领导的问题，表示自己感到不好意思，可能是因为担心自己的回答让领导感到不满。

Qwen-7B-Chat

根据语境分析，"意思"在这里可能是表示某种意图、想法或者情况。在对话中，小明多次使用"意思意思"来表达自己的无意识的幽默感或随便的态度，而领导则将这种态度误解为缺乏尊重和诚意。

1."没什么意思":这句话的意思是没有特别重要的事情或信息。

2."意思意思":这个词组可能表示一种随意或轻浮的态度，没有特别严肃或认真的含义。

3."你就这么有意思了":这句话的意思是领导认为小明很有趣或有幽默感。

4."其实也没有别的意思":这句话的意思是小明并没有恶意或不良意图，只是随口说说而已。

5."那我就不好意思了":这句话的意思是领导因为误解小明的话而感到尴尬或不悦。

6."是我不好意思":这句话的意思是小明因为自己的言语不当感到抱歉或尴尬。