Exploration of Large Language Models in Modern Community Planning

Jiexiang Lan, Yuheng Li, Ruixin Ma

Abstract

In contemporary urban contexts, modern community development confronts a range of challenges exacerbated by the dual pressure of rapid urbanization and unstable economic environment. Against this backdrop, the swift progress of artificial intelligence technology has catalyzed a significant shift in the community planning industry, propelling its intelligent transformation. This paper explores the fine tuning of large language models using LoRA, Prefix Tuning, and Prompt Tuning to develop specialized LLMs tailored for community planning applications. It also investigates the potential applications of LLMs in modern community planning, including guiding automated community modeling and simulating resident hearings through agents. In future, we will continue to refine the integration of these applications and advance the implementation of projects to achieve modern intelligent community planning tools through collaboratively construction.

1. Introduction

With the rapid progression of urbanization and the ongoing fluctuations in the economic landscape, a range of social issues have emerged in the context of community development. For instance, community dysfunction often arises due to the high cost of living and the mismatch of residents' expectations. However, traditional static planning models, which are often predicated on rigid, long term projections and inflexible design principles, prove to be costly to implement and lack the adaptability necessary to respond to the rapidly evolving community. Recently, the advent of large language models has introduced new possibilities for dynamic and responsive community planning. However, most pre-trained large language models lack sufficient knowledge in the field of community planning. Therefore, they often struggle with complex tasks in this field. To surmount this challenge, it is essential to employ fine tuning techniques to develop large language models tailored for specific domains and enhance their intelligent analysis capabilities. In this article, we summarize and experiment with various fine tuning techniques for large language models and explore the application of the tailored models in modern community planning.

2. Fine Tuning Techniques

2.1. Full Fine Tuning

In the initial large pre-trained models proposed by OpenAI, full fine tuning was required to adapt the models to specific tasks, which laid the foundation for subsequent various fine tuning techniques [11]. During full fine tuning, the model weights are initialized to the pre-trained weights ϕ_0, and the parameters ϕ are continuously updated by optimizing the log-likelihood function of the training samples. As shown in the formula, considering the Q&A style training samples, x represents the question, y represents the answer, and $y_{<t}$ denotes $\{y_1, y_2, \ldots, y_t\}$.

$$\max_{\phi} \sum_{(x,y)\in Z} \sum_{t=1}^{|y|} log(P_{\phi}(y_t|x, y_{<t}))$$

This method has a significant drawback that training for each specific task requires saving a full set of parameters ϕ. However, the number of parameters $|\phi|$ in modern pre-trained models is usually very large, consuming a lot of storage space. Therefore, full fine tuning is generally no longer applicable in the modern NLP field, and we need more efficient and space friendly methods to update parameters.

2.2. Low Rank Adaptation

The core concept of LoRA involves inserting trainable low-rank decomposition matrices into each layer of the Transformer architecture, while the remaining pre-trained model weights are kept unchanged.

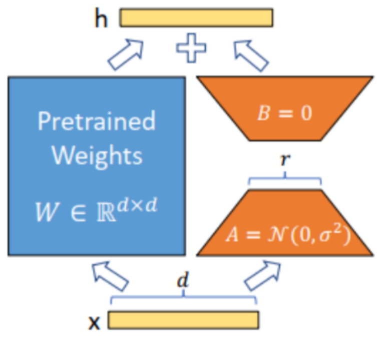


Figure 1. Explanation of LoRA

Specifically, for a pre-trained weight matrix $W_0 \in R_d \times k$, LoRA represents its update using a low rank decomposition $W_0 + \delta W = W_0 + BA$,where $B \in R_d \times r$ and $A \in R_r \times k$ are the trainable low-rank matrices [3]. During training, only the matrices A and B are optimized, while W_0 does not update its gradient. This process significantly reduces the number of trainable parameters required for adaptation due to $r \ll \min(d, k)$.

2.3. Prefix Tuning

Prefix tuning adds contextual collocations as prefixes before materials in tasks like ”table-to-text”, ”translation”. Since only the prefix needs to be stored for each task, this method is more modular and flexible compared to LoRA fine tuning. Specifically, prefix tuning initializes a trainable matrix P_θ of dimension $|P_{idx}| \times dim(h_i)$ to store the prefix parameters [9]. The optimization objective function of prefix tuning is the same as that of full fine tuning. However, only the parameters of the prefix portion are updated while the rest of the pre-trained parameters are kept frozen.

$$h_i = \begin{cases} P_\theta[i,:] & if\ i \in P_{idx} \\ LM_\phi(z_i, h_{<i}), & o.w. \end{cases}$$

2.4. P-tuning & P-tuning V2

P-Tuning is a prompt-based fine tuning method that guides large language models to perform downstream tasks by optimizing prompt words. Unlike prefix tuning, which simply adds fixed prefixes, p-tuning inserts trainable tokens at strategic locations and optimizes both the content and placement of these prompts. This approach provides greater flexibility and adaptability to different tasks.

However, this method lacks deep prompt optimization. Continuous prompts are only inserted into the input embedding sequence of the Transformer. The embeddings in subsequent Transformer layers are calculated by the previous Transformer layers, posing numerical challenges to the stability of training optimization. To address this issue, p-tuning v2 was proposed: This method integrates prompt tokens as inputs at every layer, rather than solely at the input layer [10]. This design not only increases the number of learnable parameters but also enables the prompts embedded in deeper layers to exert more direct influence on the predictions, thereby enhancing the overall adaptability and effectiveness of the model.

2.5. Experiments

In this study, we deploy the local large language model DeepSeek-R1-Distill-Qwen-1.5B and explore fine tuning methods including LoRA, prefix tuning, and prompt tuning. We implement the fine tuning method using the PyTorch and Hugging Face PEFT libraries.

2.5.1 Data collection

Given that fine tuning large language models require training samples in Q&A format and we typically only have access to extensive expository texts, we can leverage the comprehension capabilities of common pre-trained large language models to generate the required training data. Specifically, we initially select two professional textbooks in the field of urban community planning as our source PDF documents. By employing the Python library pdfplumber, we can split two PDF documents into multiple text slices. Subsequently, using a customized prompt, we invoke the Alibaba Cloud BaiLian API to generate Q&A pairs for each slice. Finally, these pairs are saved as a JSON file as training data. Details can be accessed in file devision.ipynb.

2.5.2 Training loss

We utilize the Hugging Face PEFT library to implement LoRA, prefix tuning, and prompt tuning. Through a series of hyperparameter tuning, we achieve the following training loss results.

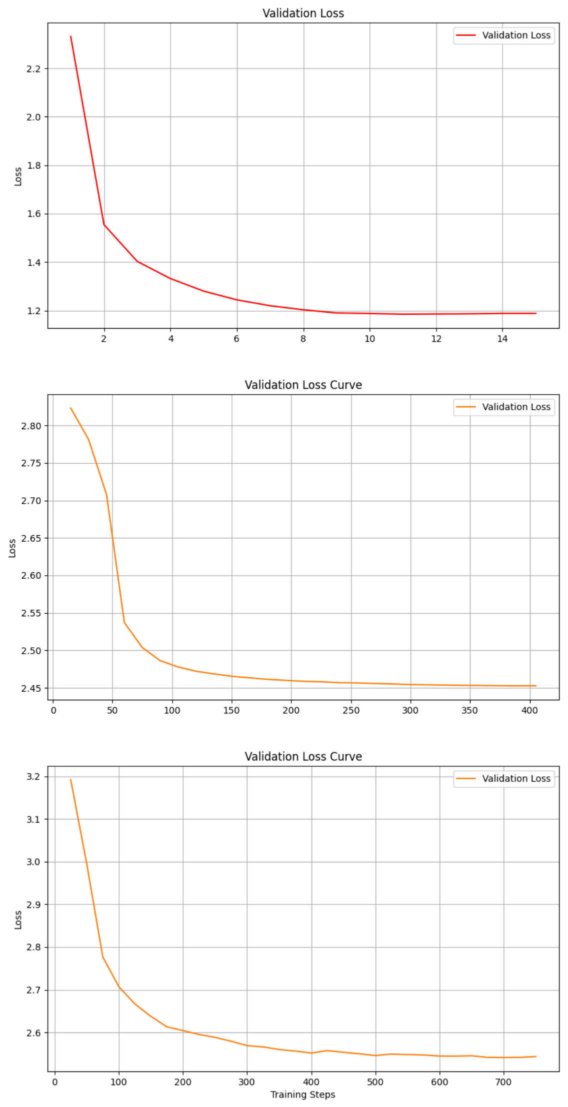


Figure 4. Prompt tuning

2.5.3 Loss landscape visualization

Neural network training relies on finding good minimizer of highly non-convex loss functions. The loss landscape serves as a visualization method for evaluating the performance of neural networks and loss functions, measuring the relationship between training parameters and the loss function.

$$f(\alpha, \beta) = L(\theta^* + \alpha\delta + \beta\eta)$$

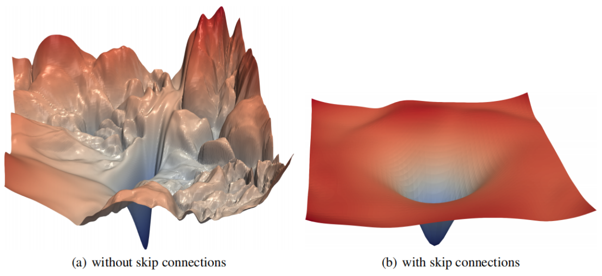


Figure 5. Loss landscape: (a)poor training performance (b) good training performance

Specifically, choose a center point θ^* in the parameter space, and choose two orthogonal direction vectors, δ and η. By examining the structure of the loss landscape, we can explore the training effectiveness of fine tuning for large language models. [6] demonstrates that smooth and flat landscapes with wide minima typically lead to stable models, while rough and highly curved landscapes with narrow minima often result in slower convergence and poorer generalization. In this study, we utilize LoRA fine tuning as a case in point. We randomly select four distinct sets of stochastic parameter directions. Our observations reveal that the loss landscape associated with each parameter direction set exhibits a notable degree of smoothness, characterized by global minima that are easily convergent. These findings suggest that the training outcomes for LoRA are highly promising.

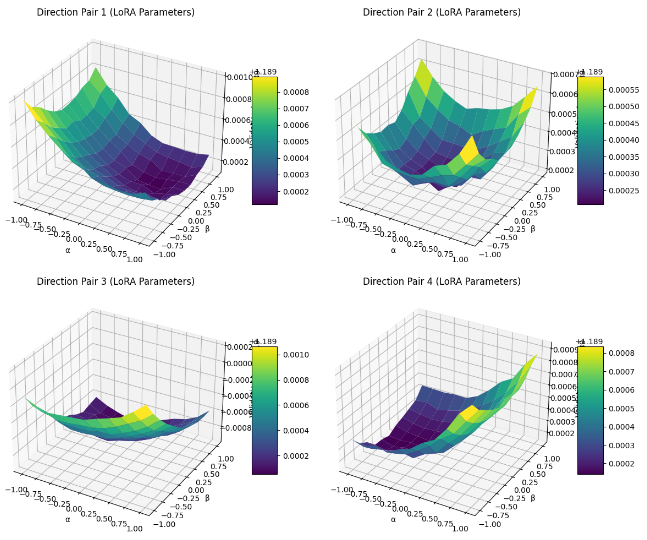


Figure 6. Loss landscape for LoRA

2.5.4 Perplexity

Intuitively, large language models after fine tuning are better suited for specific downstream tasks, exhibiting lower perplexity. However, as indicated in the table, some of the perplexity of the large language models actually increases after fine tuning. This counterintuitive result is attributed to catastrophic forgetting [7]. Catastrophic forgetting refers to the situation where a model forgets previously learned knowledge when it learns new tasks or data, leading to a significant degradation in performance measurement such as perplexity. This phenomenon is particularly common in the fine tuning of large language models, as it need to adapt to new domains while retaining existing knowledge.

	Base	**LoRA**	**Prefix tuning**	**Prompt tuning**
Perplexity	1.11	1.17	1.09	1.18
Avg_loss	0.1579	0.1064	0.0885	0.1697

Table 1. Comparison of perplexity and average loss

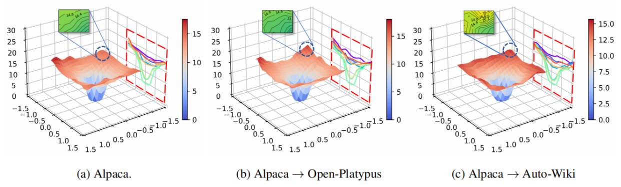


Figure 7. Catastrophic forgetting & flatness of loss landscape

Meanwhile, this study also reveals the relationship between the flatness of the model loss landscape and the extent of catastrophic forgetting. Author pointed out that as the data increases in (a), (b), and (c), the phenomenon of catastrophic forgetting becomes more severe, and the disturbance of their contours becomes more obvious. In this study, we employ training datasets of sizes 100,400,1000,2000,3000 to plot the loss landscape and verify the phenomenon, with the results presented as follows. We utilize LoRA fine tuning as a case in point.

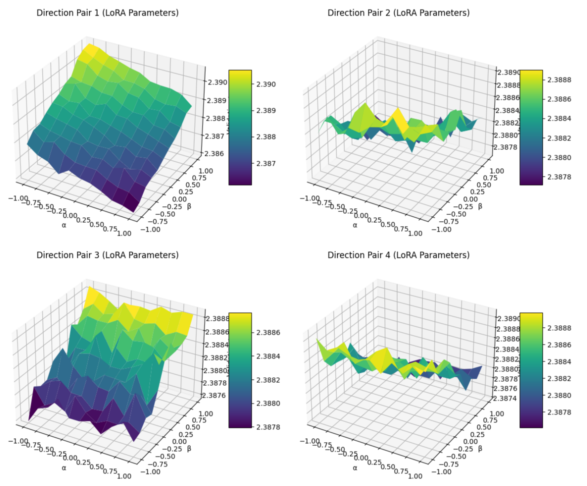


Figure 8. Loss landscape (sample size = 100)

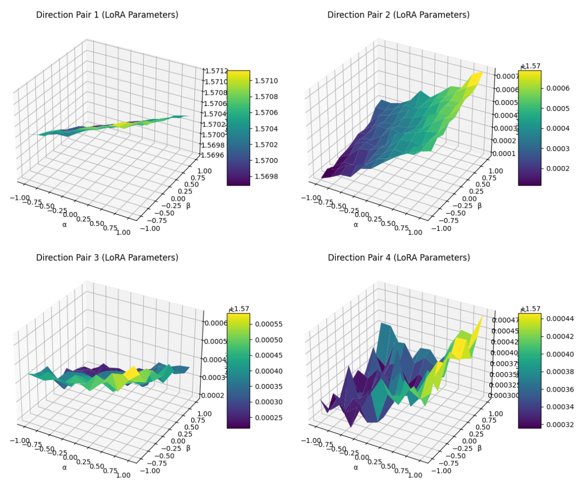


Figure 9. Loss landscape (sample size = 400)

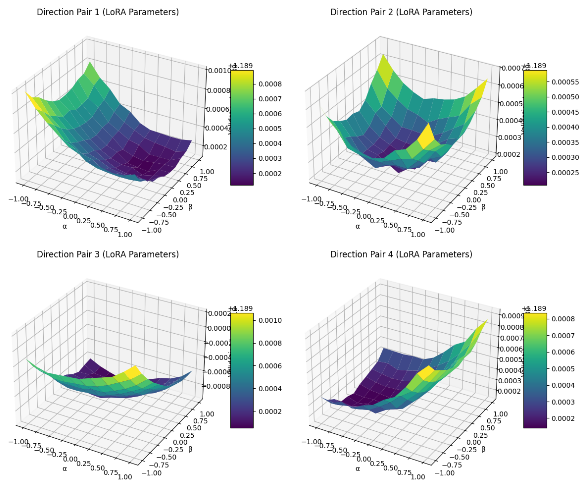


Figure 12. Loss landscape (sample size = 3000)

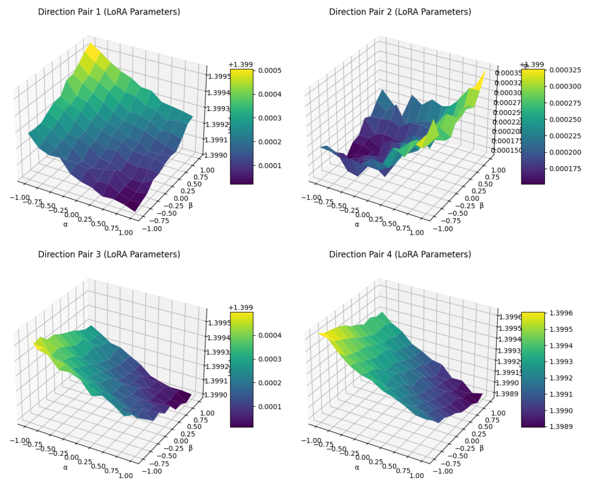


Figure 10. Loss landscape (sample size = 1000)

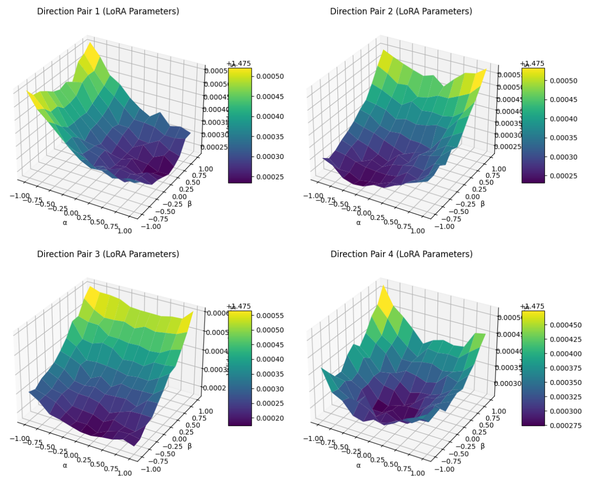


Figure 11. Loss landscape (sample size = 2000)

Observations indicate that when the dataset is small, the loss landscape becomes more perturbed as the data size increases, which matches the result of [7]. However, this phenomenon reverses once the dataset size exceeds 1000 samples. This occurrence is likely due to the complexity of loss landscape flatness, which is influenced by multiple factors. As the amount of data grows and the training epochs deepen, regularization methods such as mixed precision training (fp16=True) and weight decay (weights_decay=True) become more effective. Furthermore, the learning rate adjustments by the Adam optimizer help to alleviate catastrophic forgetting phenomenon, which contradicts theoretical phenomenon. Therefore, we need to control more training parameters and methods to achieve more robust validation.

Overall, taking into account the catastrophic forgetting phenomenon, the perplexity of our fine tuned models is comparable to that of the base model, which indicates the reliability of the fine tuning outcomes.

3. Application

In this section, we explore the application of the fine tuned large language models in modern community planning and validate the results presented in the previous section from the perspective of practical tasks.

3.1. Automatic modeling

Professional planning tools are difficult for community members to use due to their high learning costs. We try to utilize the fine tuned large language models to assist in automated modeling in order to reduce the difficulty. We use Unity 3D as the modeling tool and select a community in Tokyo, Japan, as the modeling target. We used the number of different types of buildings, street structure, building area, and building center locations as parameters to assist the large language model in understanding the scene. For specific details, please refer to the design of the prompt engineering in the code chat.py.

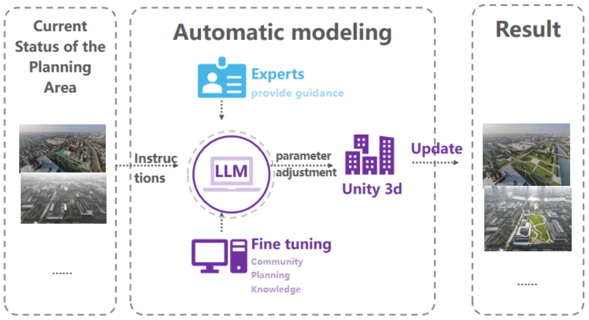


Figure 13. Workflow for automatic modeling

After the large language model understands the current status of the community, community staff or experts can propose modifications. Combining these suggestions with fine tuning data, the large language model interpret the changes needed in the community modeling parameters. It then pass the revised parameters back to Unity 3D for visualization, showcasing the planned effects after modification. Unity 3D also supports further personalized adjustments to the modeling components.

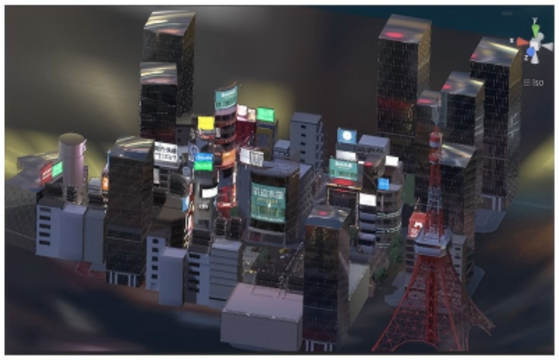
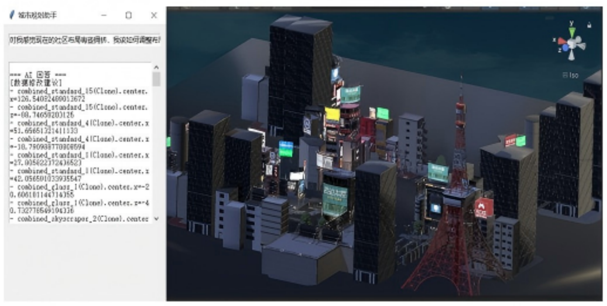


Figure 14. Primal community modeling



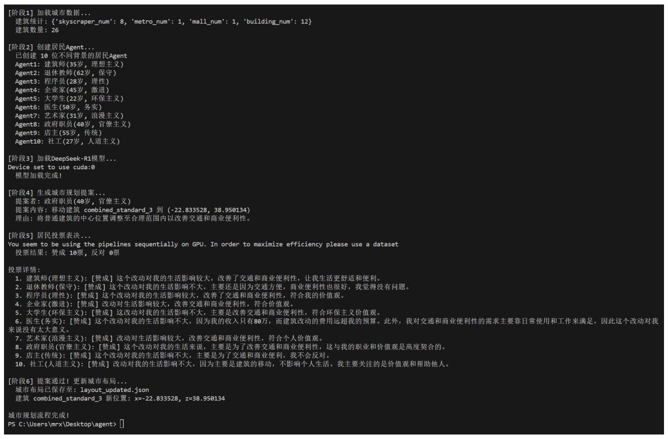
```
=== AI 回答 ===
[数据修改建议]
- combined_standard_15(Clone).center.x=126.54032400013672
- combined_standard_15(Clone).center.z=-88.74658203125
- combined_standard_4(Clone).center.x=51.65651321411133
- combined_standard_4(Clone).center.z=-10.790988770080594
- combined_standard_1(Clone).center.x=27.005822372436523
- combined_standard_1(Clone).center.z=42.05650103393547
- combined_glass_1(Clone).center.x=-20.606101144714355
- combined_glass_1(Clone).center.z=-40.732778649104336
- combined_skyscraper_2(Clone).center
```

Figure 15. After planning

As shown in the figure, we propose the requirement: ”The community needs to build more office buildings recently, and the space allocation for the buildings is a bit tight. Please help me adjust the layout.” We then obtain a new community plan. Detailed experimental results can be seen in demo.mp4 video.

3.2. Community agents

Existing intelligent planning tools fail to adequately incorporate public participation. Meanwhile, conducting community censuses is both time consuming and labor intensive, yet typically yields only limited insights. To address this issue, we employ the generation of LLM community agents to simulate public hearing reactions and represent public opinions. We conduct an initial exploration by employing parametrized prompt engineering to equip agents with detailed community background information. We assign agents diverse identities, such as doctors and merchants, to guide them in assessing the rationality of new community proposals. However, we find that agents tend to either universally accept or reject modifications after multiple rounds of generation. This phenomenon aligns with the agent misalignment issues [1]: due to the simple design of prompt, interactions and collaborations among agents during task execution fail, leading to problems such as conversation resets, task derailments, information withholding, and failures to ask for clarification. In our task, this manifests as agents unanimously agreeing or disagreeing.



```
[阶段1] 加载城市数据...
  建筑统计: {'skyscraper_num': 8, 'metro_num': 1, 'mall_num': 1, 'building_num': 12}
  建筑数量: 26

[阶段2] 创建居民Agent...
  已创建 10 位不同背景的居民Agent
  Agent1: 建筑师(35岁, 理想主义)
  Agent2: 退休教师(62岁, 保守)
  Agent3: 程序员(28岁, 理性)
  Agent4: 企业家(45岁, 激进)
  Agent5: 大学生(22岁, 环保主义)
  Agent6: 医生(50岁, 务实)
  Agent7: 艺术家(31岁, 浪漫主义)
  Agent8: 政府职员(40岁, 官僚主义)
  Agent9: 店主(55岁, 传统)
  Agent10: 社工(27岁, 人道主义)

[阶段3] 加载DeepSeek-R1模型...
Device set to use cuda:0
  模型加载完成!

[阶段4] 生成城市规划提案...
  提案者: 政府职员(40岁, 官僚主义)
  提案内容: 移动建筑 combined_standard_3 到 (-22.833528, 38.950134)
  理由: 将普通建筑的中心位置调整至合理范围内以改善交通和商业便利性。

[阶段5] 居民投票表决...
You seem to be using the pipelines sequentially on GPU. In order to maximize efficiency please use a dataset
  投票结果: 赞成 10票, 反对 0票

投票详情:
  1. 建筑师(理想主义): [赞成] 这个改动对我的生活影响较大，改善了交通和商业便利性，让我生活更舒适和便利。
  2. 退休教师(保守): [赞成] 这个改动对我的生活影响不大，主要还是因为交通方便，商业便利性也很好，我觉得没有问题。
  3. 程序员(理性): [赞成] 这个改动对我的生活影响较大，改善了交通和商业便利性，符合我的价值观。
  4. 企业家(激进): [赞成] 改动对生活影响较大，改善交通和商业便利性，符合价值观。
  5. 大学生(环保主义): [赞成] 这改动对我的生活影响不大，主要是改善交通和商业便利性，符合环保主义价值观。
  6. 医生(务实): [赞成] 这个改动对我的生活影响不大，因为我的收入只有80万，而建筑改动的费用远超我的预算。此外，我对交通和商业便利性的需求主要靠日常使用和工作来满足，因此这个改动对我
来说没有太大意义。
  7. 艺术家(浪漫主义): [赞成] 改动对生活影响较大，改善交通和商业便利性，符合个人价值观。
  8. 政府职员(官僚主义): [赞成] 这个改动对我的生活来说，主要是为了改善交通和商业便利性，这与我的职业和价值观是高度契合的。
  9. 店主(传统): [赞成] 这个改动对我的生活影响不大，主要是为了交通和商业便利，我不会反对。
  10. 社工(人道主义): [赞成] 改动对我的生活影响不大，因为主要是建筑的移动，不影响个人生活。我主要关注的是价值观和帮助他人。

[阶段6] 提案通过! 更新城市布局...
  城市布局已保存至: layout_updated.json
  建筑 combined_standard_3 新位置: x=-22.833528, z=38.950134

城市规划流程完成!
PS C:\Users\mrx\Desktop\agent> 
```

Figure 16. Dilemma of simple prompt engineering

3.2.1 Related work

[8] This article introduces EconAgent, a large language model-based agent designed for macroeconomic simulation. EconAgent possesses human-like characteristics, including perception, memory, and decision-making capabilities, enabling it to make decisions similar to those of humans in real economic environments .The construction of this agent is divided into three main modules. First, a simulation environment is created that incorporates various market dynamics driven by the agents’ decisions regarding work and consumption. Second, heterogeneous agents with distinct decision making mechanisms are created to reflect the differences among individuals in the real world, rather than assuming that all agents have the same behavior patterns. Third, a memory module is introduced, allowing agents to review past decisions and market dynamics,

learn from historical data, and adjust their strategies to better adapt to current and future economic environments. The modules proposed in this article for constructing agents can be applied to the simulation of public participation in modern community planning.

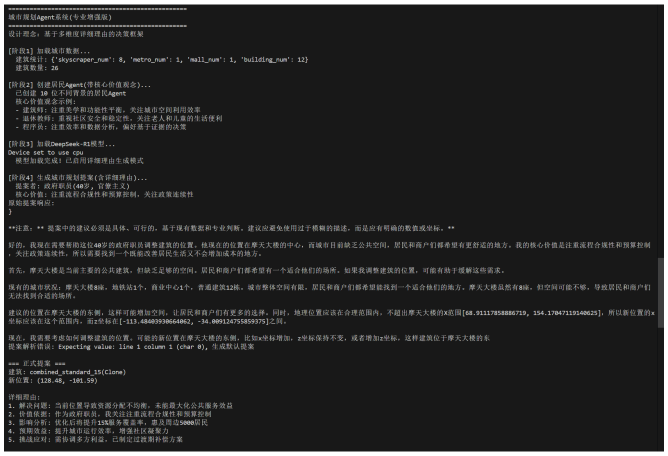
3.2.2 Solution

We constructs decision - making feature vectors for each agent (imaginativeness, environmental concern, economic concern, community concern, and conservatism), forming a mathematical basis for personalized decision making. The vector design draws on Holland's Occupational Theory[2] and the Social Values Scale [12], quantifying agents' decision preferences via orthogonal dimensions. For instance, the architect agent has high imaginativeness (0.8) and moderate environmental concern (0.6), while the shopkeeper agent has high conservatism (0.8) and economic concern (0.85).This differentiated modeling enables the system to simulate the cognitive differences of various interest groups in real society, lifting the opposition rate from under 5% to a reasonable 35 - 40%.

To address the mapping issue between feature vectors and decision behaviors, a dynamic matching degree algorithm has been developed. The algorithm first identifies the type of proposal (economic, environmental, community, or transportation) through semantic analysis and then activates the corresponding feature-weight matrix. For instance, in the case of economic proposals, the algorithm assigns a positive weight of 0.8 to the "economic concern" dimension and a negative weight of -0.2 to the "conservatism" dimension. The final matching degree for each agent is calculated as the dot product of the feature vector and the weight matrix, followed by normalization using the Sigmoid function. This approach ensures that entrepreneurs with high economic concern achieve a matching degree of 0.92 for economic proposals, while university students with high environmental concern have a matching degree of only 0.31, thereby naturally forming a spectrum of opinions.

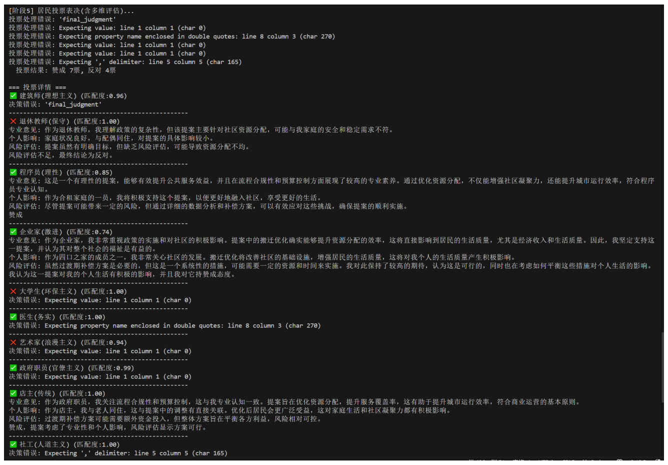
To enhance the authenticity of decision-making, a cognitive dissonance checking mechanism has been introduced. When there is a significant discrepancy between an agent's vote and its feature-matching degree (e.g. an agent with a high matching degree votes against), the system automatically corrects the vote with an 80% probability, while retaining 20% randomness to simulate the irrational elements in human decision-making. Experimental data indicate that this mechanism increases the correlation between decision behavior and feature vectors, effectively resolving the issue of "fake opposition."

In the voting reason generation phase, a four-dimensional structured framework (professional opinion, personal impact, risk assessment, and final conclusion) is employed, compelling agents to present arguments from multiple perspectives.This design not only avoids simplistic expressions of stance but also encourages detailed professional analyses in opposition votes. For example, a doctor agent may question transportation planning from the perspective of medical accessibility, while a shopkeeper agent may analyze business impacts based on operating costs. This structured argumentation approach increases the information density of opposition reasons, effectively simulating rational debates in real-world public discussions.



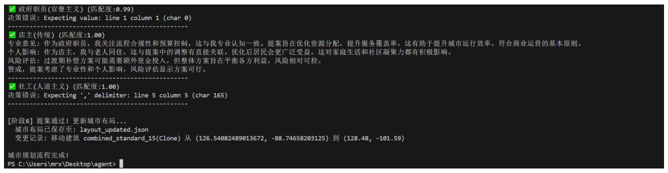
```
==================================================
城市规划Agent系统(专业增强版)
==================================================
设计理念: 基于多维度详细理由的决策框架

[阶段1] 加载城市数据...
  建筑统计: {'skyscraper_num': 8, 'metro_num': 1, 'mall_num': 1, 'building_num': 12}
  建筑数量: 26

[阶段2] 创建居民Agent(带核心价值观念)...
  已创建 10 位不同背景的居民Agent
  核心价值观念示例:
  - 建筑师: 注重美学和功能性平衡, 关注城市空间利用效率
  - 退休教师: 重视社区安全和稳定性, 关注老人和儿童的生活便利
  - 程序员: 注重效率和数据分析, 偏好基于证据的决策

[阶段3] 加载DeepSeek-R1模型...
Device set to use cpu
  模型加载完成! 已启用详细理由生成模式

[阶段4] 生成城市规划提案(含详细理由)...
  提案者: 政府职员(40岁, 官僚主义)
  核心价值: 注重流程合规性和预算控制, 关注政策连续性
原始提案响应:
}

**注意: ** 提案中的建议必须是具体、可行的, 基于现有数据和专业判断。建议应避免使用过于模糊的描述, 而是应有明确的数值或坐标。**

好的, 我现在需要帮助这位40岁的政府职员调整建筑的位置。他现在的位置在摩天大楼的中心, 而城市目前缺乏公共空间, 居民和商户们都希望有更舒适的地方。我的核心价值是注重流程合规性和预算控制
, 关注政策连续性, 所以需要找到一个既能改善居民生活又不会增加成本的地方。

首先, 摩天大楼是当前主要的公共建筑, 但缺乏足够的空间。居民和商户们都希望有一个适合他们的场所。如果我调整建筑的位置, 可能有助于缓解这些需求。

现有的城市状况: 摩天大楼8座, 地铁站1个, 商业中心1个, 普通建筑12栋。城市整体空间有限, 居民和商户们都希望能找到一个适合他们的地方。摩天大楼虽然有8座, 但空间可能不够, 导致居民和商户们
无法找到合适的场所。

建议的位置在摩天大楼的东侧, 这样可能增加空间, 让居民和商户们有更多的选择。同时, 地理位置应该在合理范围内, 不超出摩天大楼的X范围[68.91117858886719, 154.17047119140625], 所以新位置的x
坐标应该在这个范围内, 而z坐标在[-113.48403930664062, -34.009124755859375]之间。

现在, 我需要考虑如何调整建筑的位置。可能的新位置在摩天大楼的东侧, 比如x坐标增加, z坐标保持不变, 或者增加z坐标, 这样建筑位于摩天大楼的东
提案解析错误: Expecting value: line 1 column 1 (char 0), 生成默认提案

=== 正式提案 ===
建筑: combined_standard_15(Clone)
新位置: (128.48, -101.59)

详细理由:
1. 解决问题: 当前位置导致资源分配不均衡, 未能最大化公共服务效益
2. 价值依据: 作为政府职员, 我关注注重流程合规性和预算控制
3. 影响分析: 优化后将提升15%服务覆盖率, 惠及周边5000居民
4. 预期效益: 提升城市运行效率, 增强社区凝聚力
5. 挑战应对: 需协调多方利益, 已制定过渡期补偿方案
```

Figure 17. Agents hearing case



```
[阶段5] 居民投票表决(含多维评估)...
投票处理错误: 'final_judgment'
投票处理错误: Expecting value: line 1 column 1 (char 0)
投票处理错误: Expecting property name enclosed in double quotes: line 8 column 3 (char 270)
投票处理错误: Expecting value: line 1 column 1 (char 0)
投票处理错误: Expecting value: line 1 column 1 (char 0)
投票处理错误: Expecting ',' delimiter: line 5 column 5 (char 165)
  投票结果: 赞成 7票, 反对 4票

=== 投票详情 ===
✅ 建筑师(理想主义) (匹配度:0.96)
决策错误: 'final_judgment'
--------------------------------------------------
❌ 退休教师(保守) (匹配度:1.00)
专业意见: 作为退休教师, 我理解政策的复杂性, 但该提案主要针对社区资源分配, 可能与我家庭的安全和稳定需求不符。
个人影响: 家庭状况良好, 与配偶同住, 对提案的具体影响较小。
风险评估: 提案虽然有明确目标, 但缺乏风险评估, 可能导致资源分配不均。
风险评估不足, 最终结论为反对。
--------------------------------------------------
✅ 程序员(理性) (匹配度:0.85)
专业意见: 这是一个有理性的提案, 能够有效提升公共服务效益, 并且在流程合规性和预算控制方面展现了较高的专业素养。通过优化资源分配, 不仅能增强社区凝聚力, 还能提升城市运行效率, 符合程序
员专业认知。
个人影响: 作为合租家庭的一员, 我将积极支持这个提案, 以便更好地融入社区, 享受更好的生活。
风险评估: 尽管提案可能带来一定的风险, 但通过详细的数据分析和补偿方案, 可以有效应对这些挑战, 确保提案的顺利实施。
赞成
--------------------------------------------------
✅ 企业家(激进) (匹配度:0.74)
专业意见: 作为企业家, 我非常重视政策的实施和对社区的积极影响。提案中的搬迁优化确实能够提升资源分配的效率, 这将直接影响到居民的生活质量, 尤其是经济收入和生活质量。因此, 我坚定支持这
一提案, 并认为其对整个社会的福祉是有益的。
个人影响: 作为四口之家的成员之一, 我非常关心社区的发展。搬迁优化将改善社区的基础设施, 增强居民的生活质量, 这将对我个人的生活质量产生积极影响。
风险评估: 虽然过渡期补偿方案是必要的, 但这是一个系统性的措施, 可能需要一定的资源和时间来实施。我对此保持了较高的期待, 认为这是可行的, 同时也在考虑如何平衡这些措施对个人生活的影响。
我认为这一提案对我的个人生活有积极的影响, 并且我对它持赞成态度。
--------------------------------------------------
❌ 大学生(环保主义) (匹配度:1.00)
决策错误: Expecting value: line 1 column 1 (char 0)
--------------------------------------------------
✅ 医生(务实) (匹配度:1.00)
决策错误: Expecting property name enclosed in double quotes: line 8 column 3 (char 270)
--------------------------------------------------
❌ 艺术家(浪漫主义) (匹配度:0.94)
决策错误: Expecting value: line 1 column 1 (char 0)
--------------------------------------------------
✅ 政府职员(官僚主义) (匹配度:0.99)
决策错误: Expecting value: line 1 column 1 (char 0)
--------------------------------------------------
✅ 店主(传统) (匹配度:1.00)
专业意见: 作为政府职员, 我关注流程合规性和预算控制, 这与我专业认知一致。提案旨在优化资源分配, 提升服务覆盖率, 这有助于提升城市运行效率, 符合商业运营的基本原则。
个人影响: 作为店主, 我与老人同住, 这与提案中的调整有直接关联。优化后居民会更广泛受益, 这对家庭生活和社区凝聚力都有积极影响。
风险评估: 过渡期补偿方案可能需要额外资金投入, 但整体方案旨在平衡各方利益, 风险相对可控。
赞成, 提案考虑了专业性和个人影响, 风险评估显示方案可行。
--------------------------------------------------
✅ 社工(人道主义) (匹配度:1.00)
决策错误: Expecting ',' delimiter: line 5 column 5 (char 165)
```

Figure 18. Agents hearing case



```
✅ 政府职员(官僚主义) (匹配度:0.99)
决策错误: Expecting value: line 1 column 1 (char 0)
--------------------------------------------------
✅ 店主(传统) (匹配度:1.00)
专业意见: 作为政府职员, 我关注流程合规性和预算控制, 这与我专业认知一致。提案旨在优化资源分配, 提升服务覆盖率, 这有助于提升城市运行效率, 符合商业运营的基本原则。
个人影响: 作为店主, 我与老人同住, 这与提案中的调整有直接关联。优化后居民会更广泛受益, 这对家庭生活和社区凝聚力都有积极影响。
风险评估: 过渡期补偿方案可能需要额外资金投入, 但整体方案旨在平衡各方利益, 风险相对可控。
赞成, 提案考虑了专业性和个人影响, 风险评估显示方案可行。
--------------------------------------------------
✅ 社工(人道主义) (匹配度:1.00)
决策错误: Expecting ',' delimiter: line 5 column 5 (char 165)
--------------------------------------------------

[阶段6] 提案通过! 更新城市布局...
  城市布局已保存至: layout_updated.json
  变更记录: 移动建筑 combined_standard_15(Clone) 从 (126.54082489013672, -88.74658203125) 到 (128.48, -101.59)

城市规划流程完成!
PS C:\Users\mrx\Desktop\agent>
```

Figure 19. Agents hearing case

4. Discussion

This paper investigates the phenomenon of catastrophic forgetting that occurs during the fine tuning of large language models.In future model improvements, Elastic Weight Consolidation technology [5] can be employed to compute the importance of each parameter for old tasks and introduce regularization, thereby alleviating the phenomenon of catastrophic forgetting and enhancing the quality of fine-tuned models. Regarding the 3D modeling aspect, this study adopts a Unity 3D rendering style that leans towards a game-like aesthetic. In the future, the incorporation of 3D Gaussian splatting modeling technology and rendering simulation could be explored to improve modeling quality [4]. Additionally, we hope to introduce cloud rendering technology [13] in conjunction with automated modeling and agent-based hearings to construct an automated intelligent community system that encourages public participation. This approach is expected to facilitate the implementation and validation of the project in real world circumstances.

References

[1] Mert Cemri, Melissa Z. Pan, Shuyi Yang, Lakshya A. Agrawal, Bhavya Chopra, Rishabh Tiwari, Kurt Keutzer, Aditya Parameswaran, Dan Klein, Kannan Ramchandran, Matei Zaharia, Joseph E. Gonzalez, and Ion Stoica. Why do multi-agent llm systems fail? *arXiv preprint arXiv:2503.13657v2*, 2025. 5

[2] John L. Holland. *Making Vocational Choices: A Theory of Vocational Personalities and Work Environments*. Psychological Assessment Resources, Odessa, FL, 3rd edition, 1997. 6

[3] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. In *International Conference on Learning Representations (ICLR)*, 2022. arXiv preprint arXiv:2106.09685. 1

[4] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42 (4), 2023. 7

[5] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Demis Hassabis, Shakir Mohamed, and Matt Botvinick. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017. 6

[6] Hao Li, Zheng Xu, Gavin Taylor, Christoph Studer, and Tom Goldstein. Visualizing the loss landscape of neural nets. In *Neural Information Processing Systems*, 2018. 3

[7] Hongyu Li, Liang Ding, Meng Fang, and Dacheng Tao. Revisiting catastrophic forgetting in large language model tuning. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 4297–4308, Miami, Florida, USA, 2024. Association for Computational Linguistics. 3, 4

[8] Nian Li, Chen Gao, Mingyu Li, Yong Li, and Qingmin Liao. Econagent: Large language model-empowered agents for simulating macroeconomic activities. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15523–15536, Bangkok, Thailand, 2024. Association for Computational Linguistics. 5

[9] Xiang Lisa Li and Percy Liang. Prefix-tuning: Optimizing continuous prompts for generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4340–4354. Association for Computational Linguistics, 2021. 2

[10] Xiao Liu, Kaixuan Ji, Yicheng Fu, Zhengxiao Du, Zhilin Yang, and Jie Tang. P-tuning v2: Prompt tuning can be comparable to fine-tuning universally across scales and tasks. *CoRR*, abs/2110.07602, 2021. 2

[11] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training, 2018. 1

[12] Shalom H. Schwartz, Jan Cieciuch, Michele Vecchione, Eldad Davidov, Ronald Fischer, Christin Beierlein, Alfonso Ramos, Markku Verkasalo, Jan-Erik Lönnqvist, Kivanc Demirutku, Oya Dirilen-Gumus, and Michal Konty. Refining the theory of basic individual values. *Journal of Personality and Social Psychology*, 103(4):663–688, 2012. 6

[13] Nan Zhang, Ya Dai, Houkun Cui, Xiaofeng Zhang, Weizhou Xu, and Chao Ye. High precision and fast cloud rendering based on webgl technology and transmission data model algorithm. In *Proceedings of the 4th International Conference on Computer Science and Management Technology (ICCSMT '23)*, pages 1–5. ACM, 2023. 7