COSE474-2024F: Final Project Proposal
"Self-Optimizing Character Dialogue Generation using Prompt Tuning"

2021170964 Kyoungbin Park

1. Introduction

The remarkable success of subculture games like Genshin Impact, Star Rail, Zenless Zone Zero, and Wuthering Waves demonstrates the substantial market demand for this genre. Among the numerous subculture games available, titles from HoYoverse and Kuro Games consistently achieve exceptional revenue primarily due to three key factors: distinct character personalities, compelling storylines, and immersive dialogue interactions with these characters. However, current dialogue systems in subculture games face significant limitations due to their predetermined nature. This creates two major challenges:

- **Limited Player Agency**: Players often must select from predetermined responses, sometimes forcing them into dialogue choices that don't align with their preferred interaction style.
- **Resource-Intensive Content Creation**: Companies must pre-generate all dialogue content before release, requiring substantial time and financial investment. This prevents players from engaging in new conversations with characters they've grown attached to unless the company releases updates.

These limitations lead to a gradual decline in character engagement over time, as companies lack incentive to continuously produce new content for existing characters. This project explores the potential of replacing static dialogue systems with LLM-powered conversations that maintain the richness and character-specific nuances of hand-crafted dialogue while offering dynamic, real-time interactions.

2. Problem Definition & Challenges

2.1. Challenges of the Project

In subculture games, interactions typically involve NPCs (Non-Player Characters) initiating dialogue, followed by players selecting from a predetermined set of 1-3 response options to progress the story. While different dialogue choices may not substantially alter the overall narrative direction, it is crucial that all options remain within the bounds of the intended storyline.

The core objective of this project is to implement a system where players can freely input their desired responses to NPC dialogue, with the NPCs providing real-time, tailored replies. Successfully addressing this challenge could significantly enhance player immersion and interest in gameplay, as players experience more dynamic and personalized interactions within the game world.

The primary challenges of this project include:

- **Consistency**: Maintaining character-specific speech patterns and personality traits throughout dialogues.
- **Appropriateness**: Avoiding controversial or explicit content in generated responses.
- **Coherence**: Ensuring LLM-generated responses remain relevant to the story, preventing topic drift and adhering to the narrative intentions of the game's writers.

To address these challenges, it will be necessary to develop finely-tuned LLM models for each NPC character, incorporating appropriate story elements, roles, and personality traits.

2.2. Main Purpose of the Project

Building upon the existing challenges, this project aims to develop a server-based LLM solution that can handle real-time dialogue generation while maintaining character consistency. Our key hypotheses and objectives include:

- **Cost-Effective Scaling**: Assuming a userbase of 10,000+ concurrent users, traditional API-based solutions from OpenAI or Anthropic would be prohibitively expensive. We propose fine-tuning open-source models like Mistral-7B to minimize operational costs.
- **Deployment Strategy**: We will explore two potential approaches:
  - Server-based LLM capable of handling multiple concurrent inference requests
  - Lightweight, device-local LLM that can be distributed with game installation

- **Technical Implementation**:
  - Utilize attention mechanisms to embed character personalities and contextual situations
  - Optimize dialogue generation through prompt-tuning
  - Incorporate real-time CLIP-based image processing to generate context-aware dialogue based on in-game camera views

3. Related Works

Recent advancements in character-based dialogue systems and large language models have created new opportunities for dynamic, personality-consistent conversation generation. This section examines key developments across commercial implementations and academic research that inform our approach.

3.1. Commercial Implementation Analysis

Character.ai represents a significant milestone in deployable character-based dialogue systems. Their implementation demonstrates the feasibility of maintaining consistent character personalities across multiple concurrent conversations while managing computational resources effectively. The platform's success in handling multiple simultaneous users provides valuable insights into scalable architecture design for character-based dialogue systems.

3.2. Academic Research Foundations

Recent academic work has established crucial frameworks for personality-consistent dialogue generation:

Parameter-Efficient Fine-tuning Approaches:

- The PEFT-U framework (Clarke et al., 2024) introduces a novel approach to user personalization in language models. By implementing adapter-based fine-tuning techniques, PEFT-U achieves remarkable efficiency in adapting pre-trained models to individual user characteristics while maintaining model performance. This advancement is particularly relevant for our goal of creating character-specific dialogue models with minimal computational overhead.

Character Alignment and Personality Modeling:

- "Large Language Models Meet Harry Potter" (Chen et al., 2023) presents a comprehensive framework for character alignment in dialogue systems. The study introduces innovative techniques for maintaining personality consistency through carefully constructed attribute and relation matrices. Their methodology demonstrates how to effectively capture and maintain character-specific traits across extended dialogue sequences.
- Character-LLM (Shao et al., 2023) builds upon this foundation by introducing a trainable agent specifically designed for role-playing scenarios. The system employs a novel architecture that combines transformer-based language modeling with personality embedding layers, achieving superior performance in maintaining character consistency across diverse conversation contexts.

Multi-Character Dialogue Systems:

- RoleLLM (Wang et al., 2023) provides a comprehensive benchmark for evaluating role-playing capabilities in large language models. The study introduces evaluation metrics specifically designed for assessing personality consistency and response appropriateness in character-based dialogue systems. Their findings suggest that attention-based architectures with character-specific prompt tuning achieve optimal performance in maintaining distinct personalities.
- The Neeko framework (Yu et al., 2024) introduces dynamic LoRA (Low-Rank Adaptation) techniques for efficient multi-character role-playing. Their approach demonstrates how to switch between different character personalities with minimal computational overhead, achieving a 75% reduction in parameter storage requirements while maintaining 95% of the original performance metrics.

Language-Specific Implementations:

- CharacterGLM (Zhou et al., 2023) addresses the unique challenges of implementing character-based dialogue systems in Chinese language contexts. Their work provides valuable insights into handling language-specific nuances while maintaining character consistency, achieving state-of-the-art performance in Chinese character dialogue generation.

4. Datasets

In this project, **Harry-Potter-Dialogue-Dataset** introduced by (Chen et al., 2023) is used for fine-tuning LLM Models and standard for evaluating their purposes.

Harry Potter Dialogue is a dialogue dataset that integrates with scene, attributes and relations which are dynamically changed as the storyline goes on, which is deliberately designed to be used for researches on more human-like conversational systems in practice. For example, virtual assistant,

NPC in games, etc. Moreover, HPD can both support dialogue generation and retrieval tasks.

It provides information about each character's 13 attributes such ad Gender, Age, Belongings, Hobby and Spells. Information about relations between characters is also given, which lets LLM to create more appropriate dialogues regarding to the context of the full story.

5. State-of-the-art methods and baselines

Our methodology integrates recent advances in prompt engineering, model optimization, and evaluation techniques to create a scalable, character-consistent dialogue system.

5.1. System Architecture

The proposed system follows a four-stage pipeline for dialogue generation and optimization:

A. Dataset Processing and Character Embedding

Our approach begins with comprehensive character context extraction, employing semantic role labeling techniques similar to (Fan et al., 2023). The process includes:

- Automated extraction of character-specific dialogue patterns and personality traits
- Generation of ground truth character descriptions using attribute-relation matrices
- Creation of dialogue-specific datasets with preserved contextual information

B. Model Implementation and Fine-tuning

We implement a modified Mistral-7B architecture with several key enhancements:

- PEFT-U fine-tuning (Clarke et al., 2024) for efficient parameter adaptation
- Dynamic prompt generation system using transformer-based architectures
- Character-specific embedding layers for personality consistency inspired by (Shao et al., 2023)

C. Evaluation Framework

Our evaluation system employs multiple metrics to ensure dialogue quality:

- Semantic similarity assessment using Solar Embedding Model
- Natural language quality evaluation through ROUGE-L and METEOR metrics
- Visual coherence evaluation using CLIP-based scoring
- Real-time performance monitoring for latency and resource utilization

D. Continuous Optimization

The system implements a feedback loop for ongoing improvement:

- Aggregation of evaluation metrics through weighted scoring
- Prompt optimization using advanced CoOp/CoCoOp techniques
- Dynamic adjustment of character embeddings based on interaction history

5.2. Technical Implementation Details

Our implementation addresses several key technical challenges: **Scalability and Resource Management** To handle 10,000+ concurrent users, we implement:

- Distributed inference architecture with load balancing
- Caching mechanisms for frequently accessed character embeddings
- Efficient parameter sharing across multiple character instances

Real-time Processing Pipeline The system maintains real-time performance through:

- Asynchronous processing of CLIP-based visual inputs
- Parallel computation of character embeddings and dialogue generation
- Optimized attention mechanisms for faster inference

Character Consistency Mechanisms To maintain consistent character personalities, we employ:

- Attention-based character trait preservation
- Dynamic context windows for maintaining conversation history
- Personality embedding layers with continuous updates

This comprehensive methodology enables our system to generate contextually appropriate, character-consistent dialogue while maintaining scalability and performance requirements for large-scale deployment.

5.3. Evaluation

Semantic Similarity
To assess how well the generated responses match the ground truth, we will use the Solar Embedding Model to compute semantic similarity between the generated dialogues and the reference answers.

Semantic Role Labeling
Semantic Role Labeling (SRL) will be applied to evaluate the roles of entities and their actions in the generated responses, ensuring that they align with the narrative structure.

Evaluation Metrics
We will use a variety of established evaluation metrics to measure dialogue quality:

- **METEOR** (Metric for Evaluation of Translation with Explicit ORdering)
- **ROUGE-L** (Recall-Oriented Understudy for Gisting Evaluation)
- **CIDEr** (Consensus-based Image Description Evaluation)
- **BLEU** (Bilingual Evaluation Understudy)
- **Perplexity**: a common metric for assessing the fluency of language models.

Personality Consistency Evaluation
For personality-based models, we will employ the **Big Five Inventory (BFI) Test** and **LIWC (Linguistic Inquiry and Word Count)** software to evaluate how well the generated dialogues reflect the intended personality traits.

References

Chen, N., Wang, Y., Jiang, H., Cai, D., Li, Y., Chen, Z., Wang, L., and Li, J. Large language models meet harry potter: A bilingual dataset for aligning dialogue agents with characters. Technical report, Tencent AI Lab, Hong Kong University of Science and Technology, 2023.

Clarke, C., Heng, Y., Tang, L., and Mars, J. Peft-u: Parameter-efficient fine-tuning for user personalization. Ann Arbor, MI, 2024.

Fan, J., Aumiller, D., and Gertz, M. Evaluating factual consistency of texts with semantic role labeling. Technical report, Institute of Computer Science, Heidelberg University, 2023.

Shao, Y., Li, L., Dai, J., and Qiu, X. Character-llm: A trainable agent for role-playing. Shanghai Key Laboratory of Intelligent Information Processing and Shanghai AI Laboratory, 2023.

Wang, Z. M., Peng, Z., Que, H., Liu, J., Zhou, W., Wu, Y., Guo, H., Gan, R., Ni, Z., Yang, J., Zhang, M., Zhang, Z., Ouyang, W., Xu, K., Huang, S. W., Fu, J., and Peng, J. Rolellm: Benchmarking, eliciting, and enhancing role-playing abilities of large language models. University of the Chinese Academy of Sciences and ETH Zürich and The Hong Kong Polytechnic University and Institute of Automation, Chinese Academy of Sciences and Shanghai AI Lab and Harmony.AI and Beijing University of Posts and Telecommunications and The Hong Kong University of Science and Technology, 2023.

Yu, X., Luo, T., Wei, Y., Lei, F., Huang, Y., Peng, H., and Zhu, L. Neeko: Leveraging dynamic lora for efficient multi-character role-playing agent. University of Science and Technology Beijing and Institute of Automation, CAS and University of Chinese Academy of Sciences and Beihang University, 2024.

Zhou, J., Chen, Z., Wan, D., Wen, B., Song, Y., Yu, J., Huang, Y., Peng, L., Yang, J., Xiao, X., Sabour, S., Zhang, X., Hou, W., Zhang, Y., Dong, Y., Tang, J., and Huang, M. Characterglm: Customizing chinese conversational ai characters with large language models. Lingxin AI and Dept. of Computer Sci. & Tech., Tsinghua University and Zhipu AI and Renmin University of China and Knowledge Engineering Group, DCST, Tsinghua University, 2023.