Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD28271

Application of Reinforcement Learning in Non-player Character Intelligent Dialogue System

Xuezhi Ding

Southwest Jiaotong University, Chengdu, China
dxz1012@my.swjtu.edu.cn

Abstract. It is of great significance to develop non-player character (NPC) dialogue systems in games through reinforcement learning, which can enhance the immersion and interactivity of the game. NPC dialogue has characteristics different from those of other dialogue systems, the most obvious of which is its task orientation. This paper outlines the mainstream implementation methods of NPC dialogue systems based on reinforcement learning in recent years. It analyses the application of reinforcement learning in dialogue strategy optimisation, hierarchical and multi-agent frameworks, knowledge graphs, and other methods. It introduces the dialogue datasets and evaluation metrics commonly used in the game field and summarises and compares the performance of current mainstream methods. Finally, this paper discusses the limitations of existing methods and evaluation criteria and looks forward to possible future improvements. This paper allows readers to quickly understand how to implement an NPC intelligent dialogue system and how to make the dialogue system more personalised.

Keywords: Reinforcement Learning, Non-player Character, Dialogue System, Game Development, Artificial Intelligence.

1. Introduction

Non-player characters (NPCs) are an important part of all kinds of games. The dialogue between players and NPCs plays many important roles in the game, such as advancing the story and guiding players. Unlike Artificial Intelligence (AI) customer service or chatbots, the content output by NPCs must conform to the game's unique plot settings and reflect the character's personality and worldview [1]. The content of the conversation between the player and the NPC should match the game's plot, character background, etc., to avoid breaking the player's immersion.

In most games in the past, the content said by NPCs was preset by the developers. Although this method can provide players with some diverse choices through dialogue branches, the content of what the NPC says is still fixed. Game developers have long been trying to increase the freedom of NPC dialogue. Due to the limitations of technological development, they still face problems such as limited achievable degrees of freedom and insufficient computing power [2]. In recent years, the development of reinforcement learning (RL) and large language models (LLMs) has made it possible for NPCs to have more flexible conversations with players. Reinforcement learning can directly optimise the long-term rewards of dialogue policies; therefore, it is an important method to

DOI: 10.54254/2755-2721/2025.LD28271

realise the intelligent dialogue of NPCs. This paper will start from the perspective of the dialogue system, focus on how to train NPC dialogue strategies through reinforcement learning, and realise intelligent dialogue between players and NPCs that conforms to the game content. Through this paper, readers can understand the current mainstream NPC intelligent dialogue system implementation methods, including Reinforcement Learning for Dialogue Policy Learning, Hierarchical and Multi-Agent Reinforcement Learning Frameworks, Knowledge Graph and Memory-Augmented Methods, and Large Pre-trained Language Models with Reinforcement Learning from Human Feedback (RLHF). At the same time, readers can also learn about some commonly used data sets in this field.

Galingana et al.'s experiments show that dynamic NPC interactions can significantly enhance the player's immersion [3]. The intelligent NPC dialogue system allows players to get the real feeling of communicating with real people, which is incomparable to the current scripted NPC dialogue. Players can get personalised experience and diversified game plots through intelligent dialogue, which can greatly improve player participation and satisfaction [1]. Intelligent dialogue of game NPCs is a very meaningful research direction. Its development will feed back to other dialogue systems in the future and realise innovation in the personalisation of dialogue systems.

2. Mainstream methods

In recent years, research on game NPC dialogue has focused on combining reinforcement learning with deep generative models, multi-agent methods, etc., to learn dialogue strategies.

2.1. Scenario-adaptive design (reinforcement learning for dialogue policy learning)

NPC dialogues in game scenarios are very different from other dialogue systems. For example, NPC dialogues are usually highly task-oriented. The scenario-adaptability design of NPC dialogue is to make the NPC's behaviour match the game situation and realise the most basic functions of the NPC. In traditional task-based dialogue systems, RL is often used to optimise dialogue policies, enabling the system to achieve its goals in multiple rounds of interactions. For game NPCs, dialogue-action can be regarded as a Markov decision process, and multiple rounds of dialogue can be regarded as sequential decision problems. The model can be trained by designing a reward function. $M = (S, A, P, R, \gamma)$; State space s , usually represented as a sequence of text in the current conversation history; Action space a , that is, the next response or token generated; $P(s'|s, a)$ means that after executing action a in state s , the state transitions to the new state s' ; Reward function $R(s, a)$ defines the quality of NPC responses, such as whether the mission objectives are achieved, whether the player's response is positive; Discount factor γ , used to balance immediate rewards and future rewards; In this framework, the learning goal of the dialogue policy $(a|s)$ is to maximise the expected cumulative return:

$$J(\pi) = E_{\pi}[\sum_{t=0}^{T} \gamma^t R(s_t, a_t)] \tag{1}$$

This modelling approach can view NPC multi-round interactions as a sequential decision-making problem, and the RL algorithm can learn to optimise the long-term interaction quality while generating behaviours. Jaques et al. used Batch RL to train a model on static conversation data to optimise the emotion and coherence in the conversation. This study shows that RL can directly optimise non-differentiable indicators such as topic coherence and player feedback [4]. However, the action space of dialogue is huge, and it is difficult to train an end-to-end generative model based on

DOI: 10.54254/2755-2721/2025.LD28271

pure RL. Joint Optimisation for Information Extraction and Policy Learning (JOIE) is a joint optimisation framework that integrates natural language understanding, information extraction, and dialogue policy learning into an end-to-end system. On the MultiWOZ dataset, JOIE achieved a task success rate of 91%, and its average dialogue rounds and rewards were significantly better than baseline models such as DQN [5].

2.2. Personalised design (hierarchical and multi-agent reinforcement learning frameworks)

The NPC's dialogue content must be consistent with the game's world view, storyline, and character background. At the same time, the NPC's words must be consistent with the game settings in theme and style [1]. Different NPCs need to have different personalities and characteristics. In order to achieve this characteristic while reducing the action space, researchers adopted strategies such as hierarchical RL and multi-agent RL. For example, the multi-agent RL framework proposed by Wang & Wong decomposes the dialogue action into different parts, which are generated by different agents, thereby reducing the action space of a single agent [5]. Saleh et al. used hierarchical RL, where the Manager (high-level policy) is responsible for speech-level decision-making and the Worker (low-level policy) is responsible for word generation, making training more stable and improving conversation quality [4]. This method models the dialogue policy as a two-layer structure, high-level policy $\pi_H(g_t|s_t)$ selects a sub-goal g_t based on the current state s_t at time t ; low-level policy $\pi_L(a_t|s_t, g_t)$ generates specific language actions a_t based on the current state and high-level goal; the overall goal is to maximize the expected return:

$$J(\pi_H, \pi_L) = E_{s_0,g_0,a_0,\ldots}[\sum_{t=0}^{T} \gamma^t R(s_t, a_t)] \tag{2}$$

High-level policies can be used to model the personality tendencies of NPCs, allowing them to exhibit stronger personalities. In gaming scenarios, multi-agent RL can also be used to model the interactive behaviours between players and NPCs, allowing NPCs to dynamically respond to player actions during conversations, which may greatly increase the playability of some text-based games.

2.3. Information adaptability design (knowledge graph and memory-augmented methods)

The dialogue between players and NPCs is often related to gameplay and plot advancement, and its content needs to have functions such as information prompts and task guidance [6]. The difficulty in implementing this technology lies in the fact that the game NPC dialogue system not only needs to generate natural language, but also needs to understand and utilise multimodal information such as game status and world view background. Zhou et al.'s dialogue shaping method uses information obtained from the dialogue to construct a knowledge graph and use it to guide RL policy learning, thereby accelerating the convergence of RL agent policies [6]. In the dialogue knowledge acquisition framework proposed by Santamaria et al., the agent actively asks questions in the dialogue and uses RL to select graph patterns to expand the knowledge base [7]. This type of method makes the generated dialogue more relevant to the game context by explicitly maintaining states or the knowledge graph, while assisting RL optimisation.

DOI: 10.54254/2755-2721/2025.LD28271

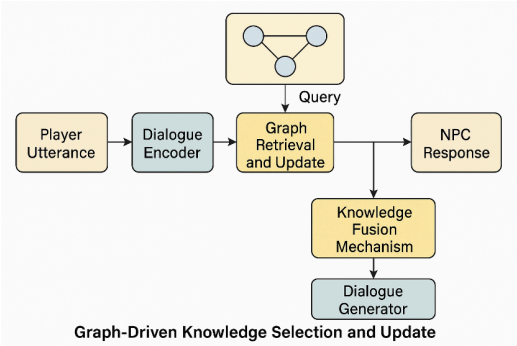

Figure 1. Graph-driven knowledge selection and update (picture credit: original)

As shown in Figure 1, the knowledge graph is embedded into the dialogue system and combined with RL policies for knowledge selection and dynamic updating. This allows NPCs to respond to player dialogue more intelligently, with contextual memory and reasoning capabilities. The knowledge graph stores information known to NPCs in the form of triples, such as <scene, character, event>. The input information is used as a query to retrieve or update the knowledge graph. RL policies (such as PPO) can be used to optimise graph usage behaviour, such as quickly skipping redundant paths. This method can give NPCs the ability to perceive the game state, allowing them to respond dynamically as the game progresses.

2.4. Interactive adaptability design (large pre-trained language models with reinforcement learning from human feedback)

Another approach is to use a LLM to generate conversation scripts. Gao & Emami tried using GPT-3 to generate NPC dialogue scripts, and the results showed that they were both diverse and realistic [1]. On this basis, RLHF can be used to fine-tune the model so that its output is more in line with player preferences. The above ideas have already appeared in commercial solutions such as NVIDIA ACE [8]. The core of these methods is to combine the optimisation objectives of RL with the powerful language generation capabilities of large language models, and they are characterised by strong interactivity. It is worth mentioning that the large language model is powerful and can also be used to schedule multi-agent systems, improve the efficiency of collaboration between agents, and thus cope with some complex tasks [9].

In summary, current methods are mostly based on hybrid approaches. Developers first use supervised learning or LLM to generate dialogues, and then use RL for subsequent optimisation. Pure RL end-to-end training of NPC dialogue is difficult, mainly due to data scarcity.

3. Analysis of mainstream methods

Compared with traditional dialogue systems, such as intelligent customer service, datasets specifically for game NPC dialogues are relatively scarce. Commonly used benchmark training sets include open-domain PersonaChat, DailyDialog, and task-based MultiWOZ [10]. Urbanek et al. constructed the LIGHT dataset based on a fantasy text adventure game, which includes character dialogues and environment descriptions [11]. In the Light dataset, the model needs to generate discourse based on the scene, objects, and character status, realising the joint learning of dialogue and perception environment. In addition, Alavi et al. constructed the MCPDial (Minecraft Persona-driven Dialogue) dataset, which generates long player-NPC dialogues containing a large number of character descriptions through LLM [12]. MCPDial emphasises the combination of dialogue,

DOI: 10.54254/2755-2721/2025.LD28271

character background, and calling game functions, providing a useful reference for future dataset work.

The indicators of the three datasets are shown in Table 1.

Table 1. Indicators of the datasets

	LIGHT	MCPDial	MultiWOZ
Construction Environment	Text Fantasy RPG Game World	Minecraft Simulator	Multi-domain task-based dialogue system
Data scale	About 110K dialogue rounds	269 full dialogues	About 115K dialogue rounds
Number of scenes	663 RPG scenes	-	7 real-life scenario areas
Data format (JSON)	<persona, setting, objects, actions, utterance>	<persona, player_act, npc_utterance, func_call>	<belief_state, domain, slots, user/system utterances>
Sample structure	Character setting + environment background + behaviour + NPC response	NPC character design + player behaviour + NPC dialogue + optional function call	User request + system intent + belief state + slot-value pair
Data characteristics	Multimodal (language + action + emotion), suitable for story-driven NPC dialogue	Personality drive, action perception, support function call logic	Clarify the slot-filling structure to support belief tracking and policy optimisation

Table 2. Analysis of four mainstream methods

Method	Dataset/Environment	Success rate / pass rate	Other key indicators	Human evaluation/subjective quality
Reinforcement Learning for Dialogue Policy Learning	MultiWOZ	DQN: 11%; JOIE: 91%	Average number of rounds: 8.45, Average Reward: 50.59	-
	Reddit conversation (Saleh et al.)	-	Positive emotionality ↑ Question rate ↑ Repetition rate ↓	-
Hierarchical and Multi-Agent Reinforcement Learning Frameworks	MultiWOZ	3-agent: 90%; 2-agent: 88%	Average number of rounds: 8.1, Average Reward: 50.6	Manual scoring: 4.6 / 5
	Custom game environment	-	The number of training rounds is greatly reduced (RL optimisation is faster)	Performance close to human level
Knowledge Graph and Memory-Augmented Methods	LIGHT	Action prediction: 41%; Emotion prediction: 26.5%	Dialogue Recall@1: 73.3%	-
	Graph construction tasks	-	Graph structure indicators converge well	-
Large Pre-trained Language Models with RLHF	Turing Quest	GPT pass rate: 64.6% (some paragraphs reach 75%)	The probability of human-written NPC scripts being misidentified as AI: 75%	High quality but no policy-level goals

DOI: 10.54254/2755-2721/2025.LD28271

Table 2 shows the analysis of four mainstream methods. In general, Knowledge Graph and Memory-Augmented Methods are both feasible and perform well. The NPC under this method has better perception ability and can dynamically adjust its language according to the progress of the game, thereby better achieving functions such as promoting the plot and guiding tasks. Knowledge Graph and Memory-Augmented Methods enable NPCs to efficiently synchronise information through shared knowledge graphs, which is important. If NPCs do not have the ability to communicate with each other, they may not be able to respond appropriately to the external environment, thereby reducing the playability of the game [13].

In game NPC dialogues, the commonly used evaluation criteria for evaluating dialogue systems have many special features compared to traditional indicators. Since game NPC dialogues have unique functionality, the evaluation criteria will be different in different game types. In other words, the most appropriate evaluation system should include both subjective and objective aspects. Especially for Knowledge Graph and Memory-Augmented Methods, the results of a comprehensive evaluation from both subjective and objective aspects are more valuable for reference [14]. Lubis et al. proposed a method to evaluate dialogue performance using offline RL, which is to train an RL critic model as an evaluator to improve the consistency and generalisation of the evaluation [15]. The evaluation results obtained by this method may be more in line with the actual needs of users. However, the current evaluation of game NPC dialogue is still mainly based on manual testing and limited objective indicators. In the future, developers still need to design comprehensive evaluation methods that are closer to the player experience.

4. Conclusions

Intelligent dialogue between players and game NPCs is an important direction of dialogue system research and has great research value. Reinforcement learning can be used to optimise dialogue policies for long-term goals and enhance players' interactive experience. This paper summarises the characteristics of game NPC dialogues and their differences from other scenarios from the perspective of dialogue systems. It also outlines the main reinforcement learning methods in recent years and analyses their current levels and challenges. Future research can develop in the direction of multi-agent collaboration, knowledge graphs, and memory enhancement to improve the consistency and flexibility of the dialogue system, while designing an evaluation system that better meets the needs of current players. This paper analyses and compares some existing datasets in related fields and proposes that the key factor currently restricting the development of reinforcement learning in NPC intelligent dialogue systems is the lack of data. The development of dialogue systems has become very mature in some fields, and it is a general trend to bring them into the field of games. This technology still has a lot of room for development in the gaming industry, and the complexity of the gaming field will drive breakthroughs in artificial intelligence dialogue systems.

References

[1] Gao, Q. C., & Emami, A. (2023, July). The turing quest: Can transformers make good npcs?. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop) (pp. 93-103).

[2] Li Yang. (2017). Design and research of intelligent agent decision-making system based on ID3 algorithm (Master's thesis, Hefei University of Technology).

[3] Galingana, C. D., & Cagadas, R. B. (2023). Enticing Player-NPC Dialogue System in a Visual Novel Game Using Intent Recognition with BERT. East Asian Journal of Multidisciplinary Research, 2(7), 2999-3014.

[4] Saleh, A., Jaques, N., Ghandeharioun, A., Shen, J., & Picard, R. (2020, April). Hierarchical reinforcement learning for open-domain dialog. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 05, pp.

DOI: 10.54254/2755-2721/2025.LD28271

8741-8748).
[5] Wang, H., & Wong, K. F. (2021, November). A collaborative multi-agent reinforcement learning framework for dialog action decomposition. In Proceedings of the 2021 Conference on empirical methods in natural language processing (pp. 7882-7889).
[6] Zhou, W., Peng, X., & Riedl, M. (2023). Dialogue shaping: Empowering agents through npc interaction. arXiv preprint arXiv: 2307.15833.
[7] Santamaria, S. B., Wang, S., & Vossen, P. (2024). Knowledge acquisition for dialogue agents using reinforcement learning on graph representations. arXiv preprint arXiv: 2406.19500.
[8] Aslan, P. (2024). Understanding LLMs for Game NPCs: An exploration of opportunities.
[9] Gong, R., Huang, Q., Ma, X., Vo, H., Durante, Z., Noda, Y., ... & Gao, J. (2023). Mindagent: Emergent gaming interaction. arXiv preprint arXiv: 2309.09971.
[10] Budzianowski, P., Wen, T. H., Tseng, B. H., Casanueva, I., Ultes, S., Ramadan, O., & Gašić, M. (2018). Multiwoz--a large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling. arXiv preprint arXiv: 1810.00278.
[11] Urbanek, J., Fan, A., Karamcheti, S., Jain, S., Humeau, S., Dinan, E., ... & Weston, J. (2019). Learning to speak and act in a fantasy text adventure game. arXiv preprint arXiv: 1903.03094.
[12] Alavi, S. H., Rao, S., Adhikari, A., DesGarennes, G. A., Malhotra, A., Brockett, C., ... & Dolan, B. (2024). Mcpdial: A minecraft persona-driven dialogue dataset. arXiv preprint arXiv: 2410.21627.
[13] Wei Linghua, Zhang Dongbing, & Fan Qi. (2014). Research on NPC behavior modeling based on swarm intelligence. Journal of Huaibei Normal University: Natural Science Edition, 35(3), 4.
[14] Wang Jisheng, & Qiao Junfu. (2024). Analysis of the application of generative artificial intelligence in non-player character dialogue. Computer Knowledge and Technology, 20(9), 22-26.
[15] Lubis, N., Geishauser, C., Lin, H. C., van Niekerk, C., Heck, M., Feng, S., & Gašić, M. (2022). Dialogue evaluation with offline reinforcement learning. arXiv preprint arXiv: 2209.00876.