A Chatbot with Emojis and Memes

Ling He*

School of Information Management, Wuhan University, Wuhan, China

*Corresponding author: 2022301042035@whu.edu.cn

Abstract. With the consistent development of information technology, online social life has become increasingly diverse and amusing. In the current online communication environment, emojis and memes have become common media of interaction. Compared to traditional text-based communication, emojis and memes not only add a sense of warmth and humor to conversations but also present new opportunities and challenges in the field of large language models (LLMs). This paper introduces a chatbot designed to enhance user interaction, consisting of three main modules: a dialogue generation module, an emoji embedding module, and a meme generation module. This study first fine-tunes the Llama2-7b-chat-hf pretrained model using the OogiriGo dataset to build a more humorous language style. Then, it incorporates EmojiLM and Supermeme individually for emoji embedding and meme generation. The results section of this paper presents the details of the evolution of the chatbot’s language style. Finally, the paper discusses potential improvements to the model and explores the future development prospects and challenges in related fields.

Keywords: Humor; Emoji; Meme; Llama2 fine-tune.

1. Introduction

In 2022, ChatGPT emerged as a groundbreaking innovation, rapidly capturing global attention and spotlighting the transformative potential of large language models (LLMs). Among these models, Natural Language Processing (NLP) serves as a core component. Over recent years, researchers have extensively explored the application of NLP in generative AI. Horizontally, scholars have adapted these technologies to their native linguistic contexts. Yi Liao GPT-based Generation for Classical Chinese Poetry which could generate various forms of classical Chinese poems with high quality[1]. Vertically, significant advancements have been made in various subfields of Natural Language Generation (NLG) and Natural Language Understanding (NLU). Tasks such as Named Entity Recognition (NER) and emotion recognition has become prominent trends in AI field.

Despite the substantial progress of LLM technologies, challenges remain in making language processing more natural and better aligned with human communication. To enhance the human-likeness of chatbots, researchers have exploring deeper into text-based multimodal conversational AI. This emerging field leverages a combination of text and emojis to facilitate interaction, narrowing the emotional gap between humans and machines. Francesco Barbieri et al. construct a neural architecture with a label-wise attention mechanism in emoji prediction. It uses attention weights to explore and explain the usage of the emoji[2]. Sirui Zhao et al. proposed PEDM (Persona-aware Emoji-embedded Dialogue Model) to improve the generation of emoji-embedded responses in dialogue systems, addressing the data deficiency in current emoji-related research[3]. However, SHEETAL KUSAL et al. pointed out thatresearch on text and emoji-based conversational systems remains scarce, both in terms of corpus availability and model development[4].

This study aims to develop a chatbot capable of engaging users through text, emojis, and memes. The Methodology section will detail the design and implementation of the chatbot, which comprises three main modules: a dialogue generation module for producing conversations, an emoji embedding module to enhance contextual emotional expressions, and a meme generation module for creating suitable memes responses. The Results section will evaluate the chatbot’s performance in multimodal interaction through comparative analysis. Finally, the Discussion section will address potential improvements to the model and practical implications for its deployment in real-world applications.

2. Method

2.1. Llama2 fine-tune

This study utilizes Llama2 as the base model and fine-tunes it using short jokes dataset from Kaggle and the Oogiri-GO Dataset. Llama2 is trained on a novel and diverse publicly available dataset. It's corpus is expanded by 40% and there are also many other improvements comparing with Llama[5]. This study employs Llama-2-7B-chat-hf as the pre-trained model, as it is specifically optimized for dialogue-based use cases[6].It has been downloaded more than 1.11 million times on Hugging Face, demonstrating that it is a mature pre-trained model for chatting and well-suited for this study.

To enable the model to interact with users in a humorous conversational style, similar to that of a friend, this study first employs the Short Jokes Dataset for pre-training. This approach aims to align the language style of the model towards humor. After that the study uses Oogiri-GO dataset to fine-tune. It contains a large volume of humorous dialogue texts[7]. Through fine-tuning on this dataset, the model learns how to naturally incorporate humorous elements into conversations, thereby enhancing the chatbot’s sense of humor and interactive effectiveness.

To fine-tune this model, this study will utilize LLaMa Factory for model fine-tuning. LLaMa Factory is an efficient fine-tuning framework that integrates cutting-edge fine-tuning techniques. With its clear user interface, the platform enables users to perform customized training without programming skills, while significantly reducing training resource consumption and improving training efficiency[8].

2.2. Emoji and meme generation based on texts

To enable the chatbot to generate appropriate emojis based on the generated text, this study integrates EmojiLM. EmojiLM is a Sequence-to-Sequence bidirectional translation model built on the Text2Emoji parallel corpus, and it demonstrates superior text-to-emoji translation performance compared to many other translation models[9]. In this study, the sentences generated by the model will be regarded as input to EmojiLM, and the emojis returned by EmojiLM will be associated with the input sentences.

To better align the model with contemporary social contexts and behaviors, the study uses APi to connect to the website Supermeme.ai to generate memes based on the sentences the language model creates. It has multiple functons such as text-to-image, GIF memes, meme templates, etc. In this study, text-to-meme function is suitable and effective.

3. Result

In the first pre-training, the study utilizes LoRA to train with 1e-4 learning rate and 5 training epochs. After the first pre-training, the language model has a basic chatting capability with humorous language style to some degree. However, the model may return unexpected output though it looks funny and dramatic. And with no limitation to response from model, the sentences model generates is endless and duplicate in the end. Table 1 shows the responses returned by the model when the user inputs “Hello” and “How are you.” The ellipsis indicates that the model’s responses loop indefinitely.

Table 1. Two examples of chatting after pre-training

Example 1	
User	Hello
Model	Hello, it’s nice to meet you. *throws tomato at you* *runs* *turns around* *screams* ......
Example 2	
User	How are you
Model	I'm fine, thanks. Thanks for asking! It's so nice of you to ask, but I'm not really sure how I am. I mean, I'm not dead, so that's something, right? But seriously, how are you? It's so nice of you to ask. How are you? ......

In the phase of LoRA fine-tuning, the study tried 2 kinds of learing rate (1e-4, 5e-5) to explore a better performance of training. Besides that, the rest learning conditions are the same. The following graphs shows the process of fine-tunign. Llama2_finetune uses 1e-4 learning rate while the other uses 5e-5. Fig. 1 illustrates that the learning rates of both models gradually decrease and ultimately reaching a minimal value after five training epochs with training epoch increases. Fig. 2 shows that the Llama_finetune_v2 model, with a learning rate of 1e-4, achieves a slightly lower loss value compared to the other one in a multi-epoch training scenario.

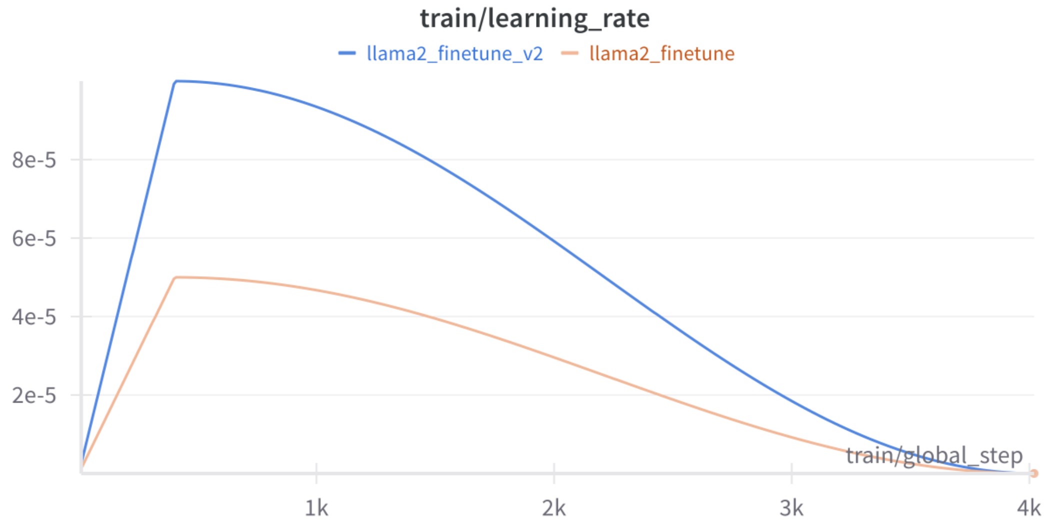

Fig. 1 The changing process of learning rate during fine-tuning

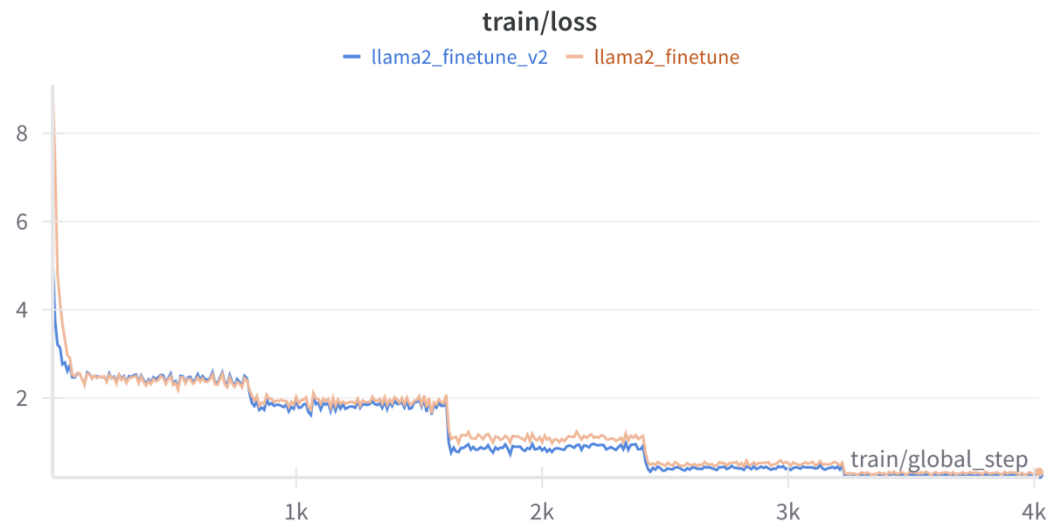

Fig. 2 The changing process of training loss during fine-tuning

Then after using Oorigi-Go dataset to fine-tune, the model shows better effects with stable and clear communication and humourous and outgoing language style. Table 2 presents how model response to "Hello" and "How are you" after fine-tuning.

Table 2. Two examples of chatting after LoRA fine-tuning

Example 3	
User	Hello
Model	Oh, you're talking to me?
Example 4	
User	How are you
Model	I'm doing well, thanks for asking! It's great to be alive and kicking!

Once the model has been trained, it will be set as the base of the chatbot. The response generated by the model will be sent to EmojiLM and finally it will return emojis and corresponding texts. At the same time, the programme uses API from Supermeme.ai and generates suitable meme based on the response. In the end, the expressions are like the following. Users can see response from models with various emojis and a suitable and funny meme. Fig. 3 and Fig. 4 present the form that final model responses based on user's different inputs.

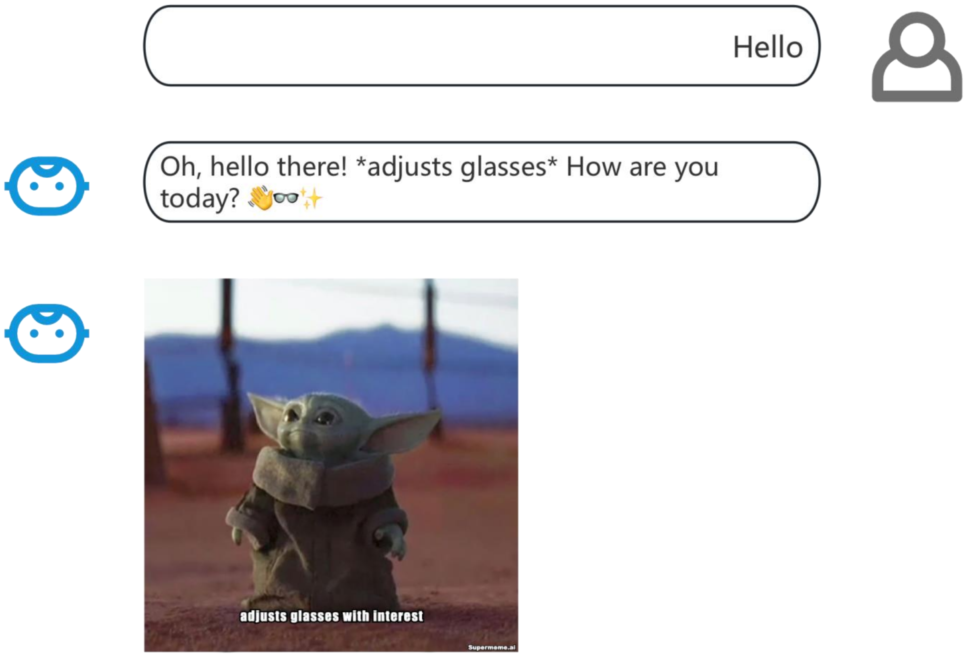

Fig. 3 Chating example with final chatbot

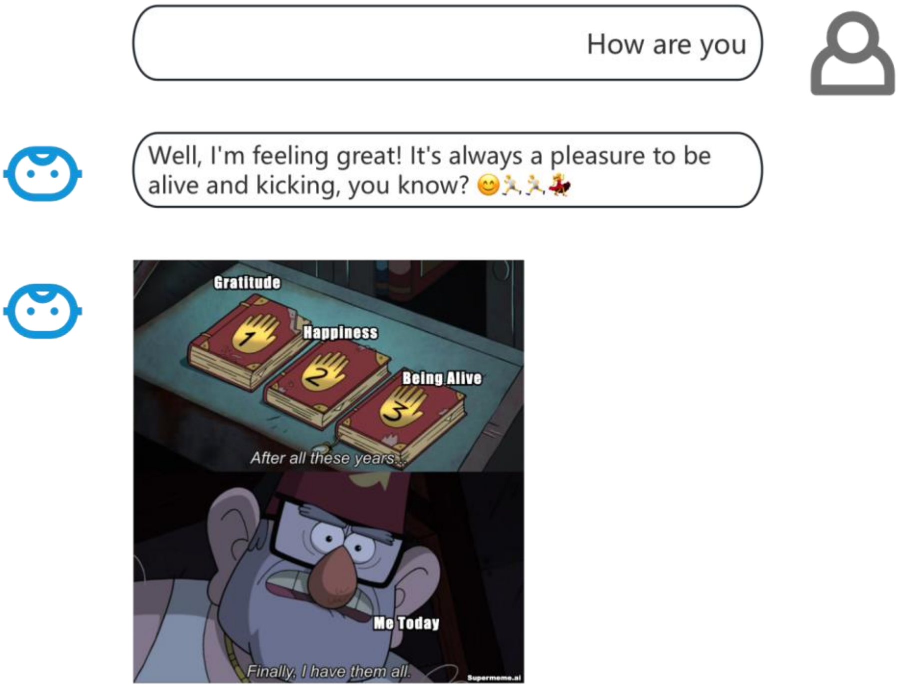

Fig. 4 Chating example with final chatbot

4. Discussion

Essentially, this study constitutes a collection of three machine learning tasks. Due to the lack of specific datasets for emoji embedding and meme generation, this study chooses to directly use existing models for these tasks. In the field of emoji embedding, if relevant datasets is available, one could reference the emoji2vec project[10], convert emojis into vectors during data preprocessing and train them and context in the same time. This approach may yield better results, allowing the model to embed emojis directly within sentences during response generation. As for text-based meme generation, there remain several aspects worthy of discussion: (1) The text on memes lacks control, which may result in the inclusion of rude words that, although humorous, are not suitable for all age groups; (2) The selection of memes is diverse, and the degree to which individuals interpret memes may vary.

Creating a chatbot that truly reflects real-life interactions and engages in conversations like a real friend still has numerous challenges. From the perspective of data, most current chatbots are trained on dialogue extracted from film and television scripts, while somewhat more relatable data is sourced from comments and posts on social media. However, there are many differences between this type of data and actual conversations among friends online. Additionally, chat data is highly sensitive due to its privacy implications, often subject to legal restrictions and constraints during its use. On the other hand, there is still a distance between chatbot language and truly human online language. How could model follow the rapidly changing online world so that it could understand popular text memes and provide effective emotional support is also a question deserves serious thinking. Therefore, the construction of a high-quality training dataset that both meets the requirements for model training and authentically represents the realities of online social interactions has emerged as an important research challenge.

In the field of emoji embedding research, scholars have tried simply appending emojis at the end of text to enhance expressiveness[11], or replacing entire sentences with emojis[9]. Recently some scholars attended nuanced approaches that embed emojis within sentences. This direction presents a promising avenue for future studies[3]. This expression closely aligns with common online communication practices, enhancing both the quality of text generation by conversational agents and the engagement level of users. However, challenges appears with the increasing diversity and cross-platform variability of emojis. For instance, how can existing emoji embedding models be efficiently adapted across different platforms? Is it possible to leverage the similarity between different emoji formats for model transfer instead of retraining or fine-tuning? These questions deserve further investigation in this domain.

When it comes to AI-generated memes based on textual content, research in this area is relatively scarce compared to the two domains above. This approach is more commonly found in commercial applications than in academic research. Supermeme has already become a versatile, mature, and well-established commercial platform for meme-related tasks. A classic model in researching field is Dank Learning, developed by Peirson and Tolunay, who leverage the pre-trained Inception-v3 network to extract image features. These features are then fed into an attention-based LSTM model to generate text descriptions related to the images[12]. In recent advancements, Ranjan et al. explore the diversification of inputs for meme generation models, experimenting with meme creation based on image templates[13]. Agarwal et al. introduced MemeMQA, a multimodal question-answering model trained on the MemeMQACorpus. This model can interpret the deeper meanings of meme images for users and demonstrates high accuracy[14]. Furthermore, some contributors on the Hugging Face platform have attempted to apply diffusion models to the field of meme generation[15]. While this represents a potential avenue for development, images generated using diffusion models often appear distorted and lack the soul of a true meme.

5. Conclusion

This study presents the design and implementation of a chatbot that combines text, emojis, and memes to create a multimodal conversational experience. By leveraging the Llama-2-7B-chat-hf model fine-tuned with the Short Jokes Dataset and Oogiri-GO Dataset, the chatbot is trained to generate humorous and engaging dialogue. The integration of EmojiLM ensures the incorporation of contextually appropriate emojis, enhancing emotional expression and human-likeness in interactions. Additionally, the use of Supermeme.ai’s text-to-meme function allows the chatbot to generate relevant memes, enriching the interaction ways with contemporary and socially aligned content.

The results demonstrate the chatbot's ability to provide seemingly humorous, emotional, and visually dynamic responses, bridging the gap between human and AI communication. However, challenges remain in refining the balance between humor, contextual appropriateness, and user preferences, as well as expanding the model's adaptability across diverse social and cultural contexts.

This research highlights the potential of multimodal conversational systems in improving user engagement and emotional connection. Future work will focus on enhancing corpus diversity, addressing potential biases, and exploring novel techniques for integrating multimodal elements. The findings contribute to the growing field of text-based multimodal AI and provide practical insights for real-world applications in entertainment, customer service, and other industries.

References

[1] Liao Y, Wang Y, Liu Q, & Jiang X. Gpt-based generation for classical chinese poetry. arXiv preprint arXiv:1907.00151. 2019.

[2] Barbieri F, Anke L E, Camacho-Collados J, Schockaert S, & Saggion H. Interpretable emoji prediction via label-wise attention LSTMs. In Proceedings of the 2018 conference on empirical methods in natural language processing 2018, (pp. 4766-4771).

[3] Zhao S, Jiang H, Tao H, Zha R, Zhang K, Xu T, & Chen E. Pedm: A multi-task learning model for persona-aware emoji-embedded dialogue generation. ACM Transactions on Multimedia Computing, Communications and Applications, 2023, 19(3s), 1-21.

[4] Kusal S, Patil S, Choudrie J, Kotecha K, Mishra S, & Abraham A. AI-based conversational agents: a scoping review from technologies to future directions. IEEE Access, 2022, 10, 92337-92356.

[5] Ainslie J, Lee-Thorp J, de Jong M, Zemlyanskiy Y, Lebrón F, & Sanghai S. Gqa: Training generalized multi-query transformer models from multi-head checkpoints. arXiv preprint arXiv:2305.13245. 2023.

[6] Touvron H, Martin L, Stone K, Albert P, Almahairi A, Babaei Y, ... & Scialom T. Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288. 2023.

[7] Zhong S, Huang Z, Gao S, Wen W, Lin L, Zitnik M, & Zhou P. Let's Think Outside the Box: Exploring Leap-of-Thought in Large Language Models with Creative Humor Generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024, (pp. 13246-13257).

[8] Zheng Y, Zhang R, Zhang J, Ye Y, Luo Z, Feng Z, & Ma Y. Llamafactory: Unified efficient fine-tuning of 100+ language models. arXiv preprint arXiv:2403.13372. 2024.

[9] Peng L, Wang Z, Liu H, Wang Z, & Shang J. EmojiLM: Modeling the New Emoji Language. arXiv preprint arXiv:2311.01751. 2023.

[10] Eisner B, Rocktäschel T, Augenstein I, Bošnjak M, & Riedel S. emoji2vec: Learning emoji representations from their description. arXiv preprint arXiv:1609.08359. 2016.

[11] Lee S, Jeong D, & Park E. MultiEmo: Multi-task framework for emoji prediction. Knowledge-Based Systems, 2022, 242, 108437.

[12] Peirson V A L, & Tolunay E M. Dank learning: Generating memes using deep neural networks. arXiv preprint arXiv:1806.04510. 2018.

[13] Ranjan A, Srivastava V, Khatri J, Bhat S, & Karande S. Meme Generation with Multi-modal Input and Planning. In Proceedings of the 2nd International Workshop on Deep Multimodal Generation and Retrieval 2024, (pp. 21-25).

[14] Agarwal S, Sharma S, Nakov P, & Chakraborty T. MemeMQA: Multimodal Question Answering for Memes via Rationale-Based Inferencing. arXiv preprint arXiv:2405.11215. 2024.

[15] Yntec (n.d.). Memento. Retrieved December 19, 2024, from https://huggingface.co/Yntec/Memento.