Verbal and Nonverbal Communications of Chat Robot

Xingping Yu
Peking University
2100017812@stu.pku.edu.cn

Abstract

Embodied chat robot is a hot topic recently. One advantage of embodied chat robor is that it can use some nonverbal communicative expressions instead of just verbal expression. So what kinds of nonverbal expressions should it have? This essay proposes some basic nonverbal expressions like facial expressions, gazing and pointing. In this essay, I want to talk about some features of human communication and what we can learn from human communication. Inspired by this, I will explain why these nonverbal expressions should be given to chat robot. Finally, this essay talks about when and how will the chat robot combine verbal and nonverbal expressions.

1   Introduction

Recently many chat robots based on large language models appear, among which ChatGPT performs very well in text generation tasks. And it is also believed to have some recognitive and reasoning abilities to some degree. However, most chat models are virtual, which means thay can only chat with people through verbal expressions. It produces some limitations of chatting. If we design an embodied chat robot, we should give it some nonverbal expressions to make it more complex. These nonverbal expressions can help us understand better. Before considering the embodied chat robot, we should consider the features of human communication first.

2   Features of human communication

When we humans communicate with others, there is a recognitive base behind the communication. We call it common ground. The common ground includes many common knowledges and recognition. One important feature is that we can understand a certain sentence based on the context. The ability to combining context with current sentences is regarded as the difference between humans and apes[3]. Another important feature is that both sides of the communication should have joint attention. This may be related to the sociality of humans. But we don't discuss the social behavior studies in this essay. Instead, we only focus on the model of human communication.

According to the book $Origins\ of\ human\ communication$ [3], the model of human communication is based on joint attention and prosocial motives. First, human communicators and recipients create the joint intention of successful communication, adjusting for one another as needed. Human communicatives acts are grounded in joint attention and shared understanding of the situation at hand. Human acts are performed for fundamentally prosocial motives such as informing others of things helpfully and sharing emotions and attitudes with them freely. Then, human communicators operate in all of this with shared assumptions of cooperative between participants. In addition, human linguistic conventions, as the crowning pinnacle of human discourse, are fundamentally shared in the sense that we both know together that we are both using a convention in the same manner.

For humans, the model is quite simple because the common ground is easily built based on our sociality and culture. But for chat robot, it needs some training process to acquire the common ground. To build an embodied chat robot, we should not only imporve the verbal expression ability, but also give it some nonverbal expressions to allow it express itself in a more human-like way.

3 Nonverbal expressions of chat robot

To simulate the human communication model above, in my opinion, an embodied chat robot should have facial expressions, gazing and pointing gestures.

First, facial expressions are important because they can express many complex emotions. For a chat robot, we may not understand what it is actually "thinking" through linguistic ways. But facial expressions can help us understand the robot rapidly and directly. Most times, facial expressions are the complement of verbal language. And the cost is quite small. Because the facial expressions have less ambiguity and are easy to map to the emotion, we don't need to design complex neural networks to train the robot, only limited training is enough.

Moreover, the chat robot needs to gaze at someone to grasp attention before it starts communicating. It is part of the joint attention in human communication. When we talk to someone, we usually have eye-contact with each other to get a good communication experience. In some culture, gazing at someone while communicating is regarded as a polite action. So the embodied chat robot should have the eye-contact ability. It is a good assistance of verbal language.

When it comes to pointing gestures, this nonverbal expression is so strong that it can be independent from verbal language. Research shows that infants are able to point by the age of about twelve months. They can not only notice where others are pointing, but also attract other's attention by pointing [2]. Another research show that the development of pointing and language have horizontal and vertical relationships. Pointing is the key joint-attention hehavior to language acquiring [1].

In general, pointing gestures can be divided into two types: imperative and declarative. In terms of specific intentions, they can also be divided into three categories. First, requesting. For example, a person may point to a cup of water to ask another person to bring the cup. Second, informing. For example, a person guides the way to others on the street, pointing in one direction to indicate that it's the right direction. Third, sharing. Usually sharing your opinion about something. For example, point to a flower to show that the flower is beautiful.

Imperative	Declarative	
Requesting	Informing	Sharing
Give me the cup of water	Guide the way	The flower is beautiful

Figure 1: categories of pointing gesture

As one of the origins of human communication, pointing gestures play an important role in human communication. Pointing gestures can complement, and sometimes even be more intuitive than, verbal language. For example, when someone asks for directions on the street, pointing in the right direction is often simpler and more effective than giving a verbal explanation. Pointing gestures are a form of social intelligence, which means it needs to be generated and understood in an interaction. Although each person has a different understanding and interaction policy, the sender and receiver of pointing gestures need to have common ground to understand and use gestures. So it is important to help the chat robot acquire pointing gestures.

4 Combination of verbal and nonverbal expressions

Although some nonverbal expressions like pointing gestures can be independent from verbal languages, most times we use them in the same time. We combine the verbal and nonverbal expressions to help communicators understand each other.

In my opinion, there are two scenes that the chat robot will combine the verbal and nonverbal expressions together. First, when verbal languages have ambiguity. In this scene, nonverbal expressions

are necessary. For example, when we are not sure whom the robot is talking to, we can infer from the gazing action of the chat robot. The second scene is when verbal languages are to complex to understand. For example, if the chat robot is guiding the ways to a passer-by, pointing to the right direction is often easier that verbal explaining.

5 Conclusion

In this essay, I proposed some nonverbal expressions an embodied chat robot should have: facial expressions, gazing actions and pointing gestures. They are key joint-attention behaviors to build the common ground. And when verbal languages are hard to understand, the robot should combine verbal and nonverbal langueges together.

References

[1] Cristina Colonnesi, Geert Jan JM Stams, Irene Koster, and Marc J Noom. The relation between pointing and language development: A meta-analysis. *Developmental Review*, 30(4):352–366, 2010. 2

[2] Eleanor H Leung and Harriet L Rheingold. Development of pointing as a social gesture. *Developmental psychology*, 17(2):215, 1981. 2

[3] Michael Tomasello. *Origins of human communication*. MIT press, 2010. 1