Human-to-Human Interaction Using Virtual Agent Posing as Another Person

Kaoru Sumi
Future University Hakodate
Hakodate, Hokkaido 041-8655, Japan
Email: kaoru.sumi@acm.org
www.fun.ac.jp

Abstract

Opportunities for online distance education have greatly expanded in recent years, and agent-based interactions in virtual spaces have attracted attention in this context. In this talk, I will discuss the various educational possibilities of using virtual spaces and agents, by presenting examples from several studies. I will also introduce our own development of systems using game-based learning and a game-based story generation system that automatically generates scripts in real time on the basis of players' emotions and actions. Finally, I will discuss persuasive technology that systems can use to influence human behavior, along with impressions and applications of facial expressions and gestures, which are expressions of agents, and I will discuss what can happen when a virtual agent interacts with other users in a virtual space.

Keywords: virtual agents, virtual spaces, emotions, affective computing, persuasive technology, online education

1. Introduction

In recent years, remote teaching has rapidly become common in university education. However, it mainly entails watching a video of the speaker's face and viewing teaching materials as videos and slides.

In the field of artificial intelligence in education, a hot topic is learning systems that interactively handle learners by estimating their emotions. Our group's laboratory conducted research on a system that estimates a learner's emotions from his or her facial expressions, gestures, and operation history to assist in learning programming [1].

Another hot topic is metaverse systems, in which we become virtual agents that differ from ourselves and communicate in a virtual space. Regarding communication between virtual agents, in the past couple of years, there have been attempts for users to participate in conferences and open campuses as virtual agents from remote locations. These users choose prepared virtual agents to participate in events as different characters from themselves and communicate with other virtual agents.

This scenario poses several questions. For learning in a virtual space, what kind of learning effect is involved when an instructor and learner become virtual agents that differ from themselves? Does the kind of agent that will be more effective for learning depend on the learner's personality? Would estimation of the learner's current, real-world state be useful in a virtual world?

In this talk, we will examine various educational possibilities using virtual spaces and agents, while introducing examples from our group's research.

I will also introduce our own development of systems using game-based learning [2], and a game-based story generation system [3] [4] that automatically generates scripts in real time on the basis of players' emotions and actions.

In recent years, we have come to be able to use virtual spaces for distance learning. In the future, learning by experience in a virtual space is also expected to become feasible. When using virtual agents, the impact on communication should be considered. I will discuss persuasive technology [5] that systems can use to influence human behavior, along with impressions and applications of facial expressions and gestures, which are expressions of virtual agents [6]; and I will discuss what can happen when a virtual agent interacts with other users in a virtual space.

We need to understand the novel experience of using a virtual space and a virtual agent than differs from oneself. To provide preliminary knowledge and an example for the lecture, this paper describes a recently presented

experiential learning system in a virtual space, which was developed in our laboratory. Please refer to the related papers [3][4] for a detailed survey.

2. Learning in a virtual space

The proposed system incorporates a player's real-time emotions and actions to facilitate game-based story generation. By using a webcam to acquire facial images of the player, the system performs real-time emotion recognition. Specifically, the player uses an Oculus Touch for motion recognition. The Oculus Touch controller uses two sensors to track its position in a three-dimensional virtual space. While the sensors respond to the controller's movements, the virtual space shows a virtual hand that moves just like the player's real hand. The system then classifies the player's actions by tracking his or her hand movements.

The system's virtual space comprises an office and a break room, like a closed space found in everyday life. As the office and break room aren't separated by a wall, the player can move freely between them.

This system was inspired by a short film called "The Black Hole." In the film, a company employee in an office prints a mysterious sheet of paper containing a "black hole." The hole turns out to be a "magical tool" that enables a human hand to pass through objects and take objects on the other side. For example, after placing the black hole on a vending machine, the employee could pass his hand through the machine and take items inside. The cunning employee used the hole to take snacks from the vending machine without paying. He then used it to pass his hand through a door and steal something.

In our proposed system, the player uses the Oculus Touch controller to perform various actions: "take out," "throw," "put," "get," and "eat/drink." The specific action is determined from the controller's acceleration and what the player touches. For example, if the player touches the desk in the office, he will take a document, book, or snack, depending on his emotion, which could be "fear," "anger," or "joy." As a result of the player operating the tool while experiencing emotions, the system generates a story.

The system uses Affdex with the Facial Action Coding System (FACS) [7] to perform real-time emotion recognition. FACS encodes emotional facial expressions by combining action units (AUs), which are the smallest units of facial expressions that can be distinguished visually. In this way, the system can acquire seven emotions: joy, fear, disgust, sadness, anger, surprise, and contempt. Because a player wearing a special sensor device to acquire emotions might be tense and unable to exhibit accurate emotions, the player's emotions are obtained via a camera.

The system generates story scripts by combining four components: (1) the action performed with the Oculus Touch, (2) the player's recognized emotion, (3) the object acquired in the virtual space, and (4) information about the player's emotion based on knowledge of the object. The text is generated by incorporating the above components into a sentence structure. In Table 1, rows 8–13 give examples of generated terms. Note that the system can also generate text automatically.

For example, suppose that the system generates and displays this sentence: "You nervously take the cola out of the vending machine." Here, it generates "You nervously" because it detects "fear" as the player's emotion. Similarly, "vending machine" is the target object of the player's virtual motion, "cola" is obtained from knowledge of the vending machine's contents, and "take out" is obtained from the player's action with the Oculus Touch, with modification based on the player's emotion.

The system visualizes and displays the results obtained by applying the knowledge-based method with the acquired actions and emotions. Here, I explain the sentence generation process in terms of the sentences listed in Table 1. Rows 1-7 contain introductory sentences that are displayed at the bottom of the screen to instruct the player how to use the system. In rows 8, 10 and 12, the player "takes out" an object from the desk, and the object corresponds to the emotion determined from knowledge of the target object (i.e., the desk). In rows 9, 11 and 13, the player "puts" the object somewhere after obtaining it. The story ends when the player performs a specific action a certain number of times within 5 min; for example, it may end when he or she "eats" three times. Likewise, the story may end if the player obtains an object a certain number of times. Finally, rows 14-16 contain ending sentences. Once the system has chosen the story's ending, it darkens the screen and outputs the ending.

This system outputs the user's game experience as a text history, and we believe that this reflects the learning experience itself with real-time emotions such as joy and surprise. With such a learning experience, we believe that reviewing the history of the experience later will also support the player's learning.

Table 1. *An Example of a Generated Story*

1	You had been sleeping for some time because you were tired from work.
2	You woke up at your own workplace with no one around.
3	You noticed that the copier is moving, and you go to the front of copier.
4	A black circle was printed on a piece of paper that emerges from the copier.
5	You felt a sense of strangeness and touched the black circle.
6	Suddenly, your hand was sucked into the circle.
7	You were surprised and remove your hand.
8	You took the book out of the disk with surprise.
9	You put the book with surprise.
10	You took out documents from the desk with a lack of interest.
11	You put the documents with a lack of interest.
12	You took the book out of the desk with surprise.
13	You put the book with surprise.
14	You were holding a paper in your hand, but you did not feel like doing anything.
15	You decided to spend the day at office today.
16	In the morning, the paper was back to its original state, and I told my colleagues what I had experienced.

3. Conclusion

In the future, systems will emerge that allow users to learn through experience in a virtual space, as in the system described here. Such systems still involve many unknown methodologies, and this is the crux of our research problem. There are also many unknown factors in learning through interaction using virtual agents, which means that there are research issues to address. Through this talk, I have sought to discuss various possibilities for using virtual spaces and agents in learning, by introducing several research cases.

References

1. Thomas James Z. Tiam-Lee and Kaoru Sumi: Analysis and prediction of student emotions while doing programming exercises. International Conference on Intelligent Tutoring Systems, Lecture Notes in Computer Science, 2019. (Best Full Paper Award).
2. Kaoru Sumi and Kodai Kasai: A Serious Game for Learning Social Networking Literacy by Flaming Experiences, Lecture Notes in Computer Science, Volume 178, Social Informatics and Telecommunications Engineering series, Springer, 2016.
3. Kaoru Sumi & Shusuke Sato: Experiences of Game-Based Learning and Reviewing History of the Experience Using Player's Emotions, Frontiers in Artificial Intelligence, 2022.
4. Shusuke Sato & Kaoru Sumi: Story Generation System Using Player's Emotions for Review in Gamed-Based Learning. In 28th International Conference on Computers in Education, 2020. (Best Technical Design Paper award).
5. B. J. Fogg: Persuasive Technology –Using Computers to Change What We Think and Do-, Elsevier, 2003.
6. K. Sumi and M. Nagata, "Evaluating a virtual agent as persuasive technology, psychology of persuasion, janos csap´o and andor magyar eds," 2010.
7. Ekman P, Friesen W: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto: Consulting Psychologists Press, 1978.

Authors Introduction

Prof. Kaoru Sumi

Kaoru Sumi is a professor in Future University Hakodate, Japan. She received her Ph.D. in engineering from the University of Tokyo. She is currently working on Human Agent Interaction, Persuasive technology, Artificial Intelligence in Education, Affective Computing, Digital storytelling. She previously worked at ATR MI&C Research Laboratories, Communications Research Laboratory, and Osaka University, where she researched human-computer interaction, knowledge engineering, and the application of artificial intelligence. After Prof. Sumi worked on media informatics and human-agent interaction at the National Institute of Information and Communications Technology (NICT), and Hitotsubashi University. She was a visiting professor in British Columbia, Canada.