Abstract

In recent years, "Internet+" has become the next phase of the Internet development trend, "Internet+" simply means" Internet + traditional "industry." With the development of information and communication technology (ICT), the Internet can be integrated with traditional industries to create new development opportunities. With the promotion of "Internet+", all industries have given birth to new models and new vitality. Thus, Internet+ education occupies a significant position in the education industry. Compared with the traditional classroom, the online classroom takes up an increasing proportion in the education field because of its advantages, such as abundant resources, fewer time and place limitations, and the possibility of personalized teaching. With the development of robotics, there is a growing expectation that robots can help people learn. This includes companion robots that accompany students in their studies, as well as functional robots that help with learning, such as learning English and programming, so these education-supporting robots are getting more and more attention. So far, a variety of learning-supporting robots have been proposed. For example, Taki et al. proposed Interactive Emotional Communication (IEC), in which humans and robots communicate by exchanging emotional actions. Yu et al. developed a robot called an "intelligent learning partner" that can interact with learners using personalized technologies. Such as conversational agents, question-answering systems, and emotion recognition. In these researches, the robots followed predetermined rules and interacted with the learner based on their emotions and behavior. However, the effect of the robot's interactions is different on the different learners.

This research proposes an interaction network model for learning partner robots to follow such individual differences and realize personalized interactions. The network consists of three layers representing action, words, and speech rate. It is a 3*4 fully connected network where the initial weights are set manually in advance. The weights are translated into a probabilistic form, which determines the content of the robot's interaction. We also define the "good or bad" of the previous interaction based on the learner's feedback, decide to increase or decrease the weight of the interaction network, and update the interaction network weights so that the robot can change the interaction content through the learner's feedback. Thus, it enables the robot to give the most personalized interaction to different learners.

As the partner robot in this research, we used the communication robot Sota, which has a camera, microphone, speaker, and network functions to interact with the learner with

words and actions. Sota can change the rate of speech, pitch, and volume level and customize the movements' design. For the learner's engagement prediction task, we used V. Huynh's method, a sub-challenge of the 7th Emotion Recognition in the Wild Challenge (EmotiW 2019). His method involves three basic steps: feature extraction, regression, and model combination. First, the input video is divided into multiple overlapping segments, and features are extracted for each instance. A combination of long and short-term memory (LSTM) and fully connected layers was used to capture temporal information, and regression was performed on the features from the previous step to obtain the engagement intensity. In the last step, Explain elements for this fusion to achieve better performance. In this model, the engagement intensity is divided into four levels, namely highly-engaged, engaged, barely-engaged, and disengaged. Then, the interaction network generates Sota's interactions based on the engagement intensity predicted model. After the interactions, the model re-predicts the learner's engagement intensity, determines the interaction contents are good or bad, and updates the interaction network weights to adapt to the learner's preference. Finally, we gathered 20 students and let them watch two 30-minutes instructional videos, one of which was with Sota and the other was Self-study. By comparing the conditions with and without Sota, the 20 subjects were divided into four groups to reduce the influence of the differences in the video content and learning order: Group 1: Self-study + video 1 → Sota + video 2; Group 2: Self-study + video 2 → Sota + video 1; Group 3: Sota + video 1 → Self-study + video 2; Group 4: Sota + video 2 → Self-study + video 1.

In this experiment, there are three primary objectives. 1. To evaluate the accuracy of the engagement intensity detection model. So after watching each video, the subjects were asked to refer to their facial video and score their engagement intensity at 0, 5, 10, 15, 20, 25, and 30 minutes in the learning process. The scoring criteria were 0-disengaged, 1-barely-engaged, 2-engaged, 3-highly-engaged. The data judged by the model were then compared with the data answered by the subject. The results show that out of 280 judgments, the number of correct judgments was 150, with a correct rate of 53.6%. 2. To evaluate how well the Sota performs the interaction. The goodness of Sota's interaction was evaluated from the following three perspectives. a) Which condition kept the engagement higher in Sota vs. Self-study. The data indicate that their engagement intensity was higher in the Sota condition in both cases that learners scored themselves and the neural network estimated. b) Comparison of the number of times engagement was restored during the course in Sota vs Self-study. We counted the number of changes from low to high engagement intensity in the two conditions, respectively, as a way to assess whether Sota had a positive effect on the learner's

recovery of engagement intensity. The results show that in both the subject judgment and the neural network judgment, the engagement intensity was recovered more often in the condition of Sota. We can conclude that the presence of Sota helps the learner to improve the engagement intensity. c) We use the questionnaire’s answer to judge whether the timing and content of Sota interactions are appropriate and whether Sota is useful for helping learners maintain a good learning state. More than half of Sota's interaction timing was positively accepted for the interaction timing. Around half of Sota's interaction contents were positively accepted for the interaction content. Furthermore, around 40% of Sota's interaction affected the maintenance of learning engagement. 3. To evaluate the effectiveness of the proposed update algorithm for the interaction model from the following two perspectives. a)According to the responses of the questionnaire, the number of Sota interactions was divided into the first third, the middle third, and the last third. To evaluate the effectiveness of the proposed update algorithm was verified by comparing the average of the subject's satisfaction with the content of these three parts. From the results, we can conclude that the subject satisfaction increases significantly in the number of interactions at the last third period as the experiment progresses. b) We summarized absolute weight changes between each engagement intensity for the 20 subjects. We found considerable variation in the weights of disengagement-criticism, engagement-criticism, and engagement-praise.

In this research, the accuracy rate of engagement intensity was 53.6%. This makes some mismatches for Sota interaction. This model also has a certain delay and is not so sensitive. Therefore, our next step is to develop a more accurate and sensitive model for future research. In future work, we expect to change the robot's interaction content based on the learner's interest in the learning content and the learning effect. In this research, each session was only 30 minutes, and we are interested in how the robot works well when learners watch longer videos.

Keyword:Deep Learning, Interaction,Interaction network,E-learning,Robot,Engagement