Simultaneous Job Interview System Using Multiple Semi-autonomous Agents

Haruki Kawai, Yusuke Muraki, Kenta Yamamoto,
Divesh Lala, Koji Inoue, and **Tatsuya Kawahara**
Graduate School of Informatics, Kyoto University, Japan
[kawai,muraki,yamamoto,lala,inoue,kawahara]
@sap.ist.i.kyoto-u.ac.jp

Abstract

In recent years, spoken dialogue systems have been used in job interviews where an applicant talks to a system that asks pre-defined questions, called on-demand and self-paced job interviews. We propose a simultaneous job interview system, where one interviewer can conduct one-on-one interviews with multiple applicants simultaneously by cooperating with multiple autonomous interview dialogue systems. However, it is challenging for interviewers to monitor and understand all parallel interviews done by the autonomous system simultaneously. To address this issue, we implement two automatic dialogue understanding functions: (1) response evaluation of each applicant's responses and (2) keyword extraction for a summary of the responses. In this system, interviewers can intervene in a dialogue session when needed and smoothly ask a proper question that elaborates the interview. We have conducted a pilot experiment where an interviewer conducted simultaneous job interviews with three candidates.

1 Introduction

Owing to the widespread use of online job interviews during the COVID-19 situation, spoken dialogue systems supporting job interviews to make them more efficient are being investigated. In conventional face-to-face job interviews, interviewers conducted interviews with many applicants one by one, which was time-consuming. Therefore, on-demand interviews have been widely adopted as an alternative to face-to-face interviews, such as *Hirevue* [1] and *Modern Hire* [2]. In this style, job applicants answer predefined typical questions and then submit video recordings of interviews. However, there is a lack of the much needed interaction between interviewers and applicants since applicants only respond to predefined questions. Therefore, to elicit sufficient information from applicants for their selection becomes difficult.

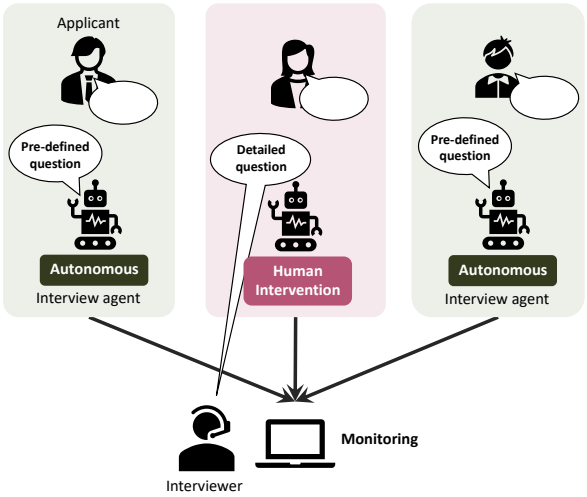

Figure 1: Concept of simultaneous job interview system

In this study, we propose a new framework for a spoken dialogue system that makes job interviews more interactive and efficient than that of on-demand interviews. The proposed framework is a cooperation between system and humans, namely semi-autonomous agents. With this framework, job interviewers can conduct multiple job interviews simultaneously. Specifically, a human job interviewer (operator) cooperates with multiple autonomous job interview agents to conduct one-on-one interviews with multiple applicants simultaneously (Figure 1). For most of the session, an autonomous agent conducts a job interview with each job applicant, and the human interviewer (operator) monitors them. The interviewer can intervene in any of the dialogues when necessary and then asks specific follow-up questions that cannot be generated by the autonomous agent. These follow-up questions are necessary to make job interviews more interactive and substantial. In this paper, we describe the framework of the proposed system and report a pilot experiment.

[1] https://www.hirevue.com/
[2] https://modernhire.com/

Proceedings of the SIGdial 2022 Conference, pages 107–110
Heriot-Watt University, Edinburgh, UK. 07-09, September, 2022 ©2022 Association for Computational Linguistics

2 Simultaneous job interview system

First, we introduce the one-on-one autonomous job interview dialogue system which is a basic component of the proposed framework. This system only asks predefined questions one by one such as motivation, strengths, and weaknesses. Similar to the existing on-demand job interview systems, no follow-up questions are asked after the responses. Although several works exist on follow-up question generation in the job interview domain (Su et al., 2019; Inoue et al., 2020), the questions automatically generated by the system are not necessarily appropriate or what the interviewer actually wants to know.

Next, we describe the proposed simultaneous job interview system. In this system, each applicant is interviewed by the above autonomous agent, and the human interviewer observes these multiple interview sessions. If the human interviewer wants to directly interact with any applicant, the interviewer can switch from the autonomous agent and then interact with the applicant. For example, the interviewer can ask specific follow-up questions that cannot be generated by the autonomous agent. Then, after the interviewer ends the intervention, the autonomous agent continues the session.

In this system, the interviewer is required to comprehend each applicant's answer and then ask proper follow-up questions and also decide on the timing of intervention. However, due to the cognitive ability of humans, it is not possible to understand the contents of multiple dialogues simultaneously. Even if each log of automatic speech recognition is generated and shown to the interviewer, it is difficult to follow all of them. It is necessary to summarize the information of each session. Therefore, we introduce response evaluation and keyword extraction that enable the interviewer to follow the dialogues done by multiple agents, as follows.

2.1 Response evaluation

We implemented a model that automatically evaluates the quality of each applicant's response. First, we conducted an annotation of response quality using a job interview dialogue corpus containing 86 mock job interview sessions (Inoue et al., 2020). The following three metrics were evaluated on the 3-point scale, from 0 (low) to 2 (high), for each response from the corpus.

Table 1: Number of annotated samples for response evaluation (0: insufficient, 1: middle, 2: sufficient)

Evaluation item	0	1	2
Appropriateness	18	26	464
Concreteness	164	190	154
Conciseness	112	311	85

- Appropriateness (Does the response fulfill what was asked?)
- Concreteness (Is the response concrete? Does the response contain any evidence and specific episodes?)
- Conciseness (Is the response brief?)

The numbers of annotated samples for each score and item are summarized in Table 1.

For each evaluation item, we made a binary classifier with BERT where the input is the concatenation of the system's question and applicant's response. The pre-trained BERT model [3] was fine-tuned with the three class labels of each item. The five-fold cross-validation was conducted and the macro F1-scores were 64.2%, 71.6%, and 76.0% for appropriateness, concreteness, and conciseness, respectively. A sample input is shown below, and the response evaluation models correctly assign each score of 2.

> (What is your strength?)
> "I have a degree in education, so I know a lot about how to help children learn while having fun. I also studied specialized content in my master's program, which I believe will be useful in creating teaching materials. I am also well versed in special needs education, and I think my strength lies in my ability to work with a wide variety of children."

The sum of the three scores is presented to the operator and used for evaluation of applicants. The operator can choose to intervene the applicant who is given high scores. On the other hand, when the system is used for interview practice, the operator might intervene against applicants with low scores.

2.2 Keyword extraction

Keyword extraction was implemented using the same response data. We annotated keywords using

[3] https://github.com/cl-tohoku/bert-japanese

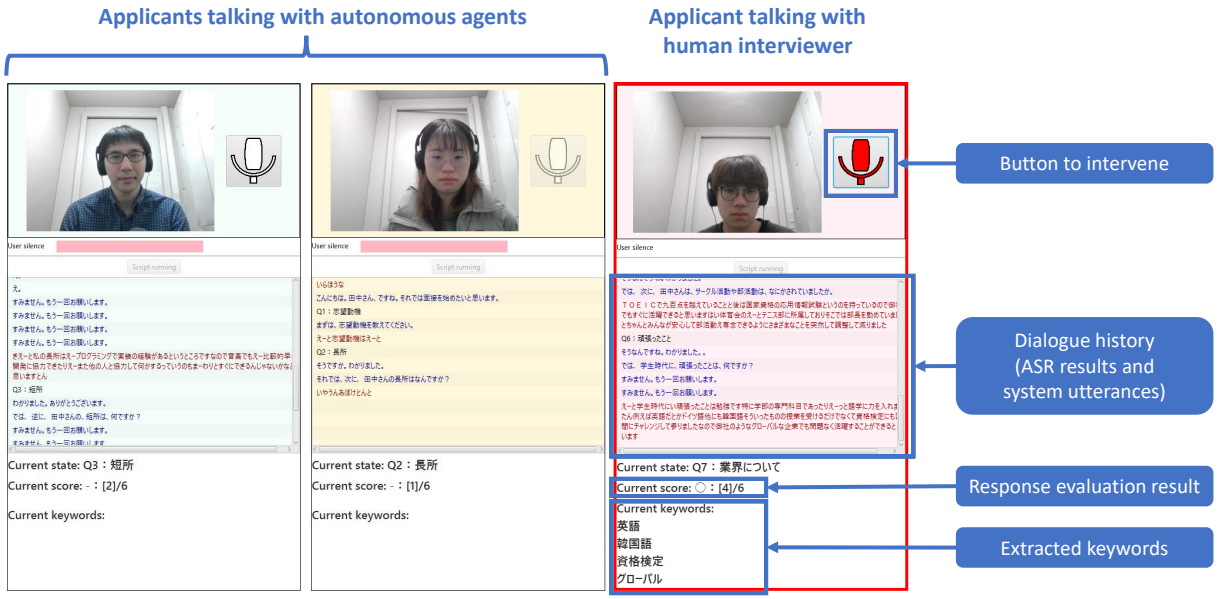

Figure 2: Interface for interviewer

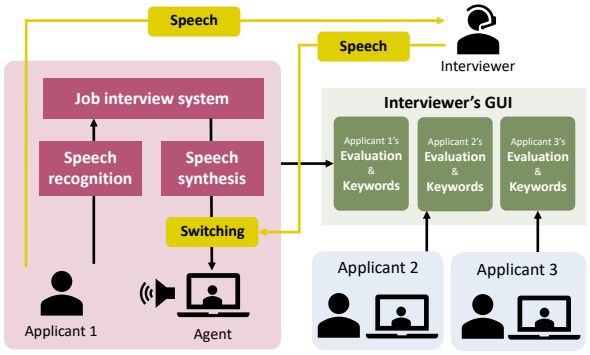

Figure 3: System configuration

the criterion of "words (or compound nouns) that represent the applicant's ability and experience". A character-based BiLSTM-CRF (Akbik et al., 2018) was used as a keyword extraction model. The benchmark result showed that the F1-score was 61.9%. For example, keywords extracted from the same input response as in Section 2.1, were "degree in education", "how to help", "specialized content", and "a wide variety of children". These keywords are presented to the interviewer as summary of the responses as a help for follow-up questions.

3 System implementation

Figure 3 depicts the configuration of the proposed system. The autonomous job interview system runs for each applicant. The input speech is segmented by a pause and fed into an automatic speech recognition with the sub-word-based attention mechanism. The recognition results are concatenated within the same turn and then used for the response evaluation and keyword extraction. The interface of the job interviewer agent is realized by MMDAgent (Lee et al., 2013). The system utterances are played with a text-to-speech engine.

Figure 2 shows the GUI for an interviewer where they can monitor multiple dialogues. This interface consists of mainly three items: (1) the dialogue history of each applicant, (2) the results of response evaluation, and (3) the results of keyword extraction. The human interviewer can select any applicant they want to intervene by clicking a button in the GUI. Once the interviewer selects an applicant, they can interact with each other directly, meanwhile, autonomous agents talk with the other applicants simultaneously.

4 Pilot experiment

We conducted a pilot experiment to confirm if the proposed system can handle multiple job interviews. In this experiment, a within-subject comparison was made between the fully autonomous system without human intervention (baseline) and the proposed system with three applicants. The subjects were 30 undergraduate and graduate students as applicants in the setting of "a student who participates in a first-round interview of some company." Note that the company was selected by each participant freely and independently. They were divided into groups of three persons in the condition

Table 2: Evaluation result in pilot experiment (5-point scale from 1:low to 5:high)

Evaluation items		Baseline		Proposed		p-value
		Mean	STD	Mean	STD	
(Q1)	The dialogue was smooth	4.14	1.53	4.32	0.82	.153
(Q2)	The system's responses were natural	4.07	1.18	4.14	1.02	.301
(Q3)	You participated in the interview seriously	4.36	0.91	4.43	0.77	.245
(Q4)	You were nervous during the interview	3.29	1.34	3.61	1.14	.030*
(Q5)	You talked well about yourself	3.64	1.13	3.93	1.03	.066+
(Q6)	You felt the interviewer listened your answers	3.29	1.40	4.14	0.35	<.001**
(Q7)	The interviewer understood you	3.11	1.43	3.64	0.83	.005**

(+ $p < .10$, * $p < .05$, ** $p < .01$)

of the proposed system. The evaluation items are listed in Table 2 where each was rated on a 5-point scale from 1 to 5. This experiment was conducted in Japanese.

Table 2 summarizes the evaluation results. The one-tailed paired *t*-test was conducted for each evaluation item, and the proposed system received significantly higher scores on the three items "You were nervous during the interview", "You felt the interviewer listened to your answers", and "The interviewer understood you". A significant trend was also observed for the item "You talked well about yourself". Although no significant trend was observed, the proposed method was rated higher than the baseline method for the other three items. Therefore, the proposed system improved the quality of interaction through the intervention of the interviewer, and can conduct efficient multiple job interviews with three applicants simultaneously.

We present some comments given by the subjects after the experiment. Following were the comments regarding the proposed system.

> "I thought it was efficient to let the machine ask the typical questions that have to be asked during the interview and let a human engage in interaction more advanced."
>
> "I get very nervous when the questions are asked back. It is good to have a realistic sense."

The baseline fully autonomous system received the following comments.

> "I did not feel like I was being listened to."
>
> "I did not really feel like I was being interviewed because I was always told "I see" after each answer. I did not feel like I was being interviewed very much."

5 Conclusions

We propose a simultaneous job interview system that allows human interviewers to interact with multiple applicants in real-time based on response evaluation and keyword extraction. For the interface of interviewers, the response evaluation and keyword extraction were implemented for making efficient intervention. In the pilot experiment, we showed the effectiveness of the proposed system and confirmed the proposed architecture would potentially be accepted as a new framework for future job interviews.

Acknowledgement

This work was supported by JST, Moonshot R&D Grant Number JPMJPS2011.

References

Alan Akbik, Duncan Blythe, and Roland Vollgraf. 2018. Contextual string embeddings for sequence labeling. In *COLING*, pages 1638–1649.

Koji Inoue, Kohei Hara, Divesh Lala, Kenta Yamamoto, Shizuka Nakamura, Katsuya Takanashi, and Tatsuya Kawahara. 2020. Job interviewer android with elaborate follow-up question generation. In *ICMI*, pages 324–332.

Akinobu Lee, Keiichiro Oura, and Keiichi Tokuda. 2013. MMDAgent–A fully open-source toolkit for voice interaction systems. In *ICASSP*, pages 8382–8385.

Ming-Hsiang Su, Chung-Hsien Wu, and Yi Chang. 2019. Follow-up question generation using neural tensor network-based domain ontology population in an interview coaching system. In *Interspeech*, pages 4185–4189.