A Cloud Based Communication System for Elders Using Dialogue Control

Kazuhiro Morita, Masao Fuketa and Jun-ichi Aoe
Institute of Technology and Science,
The University of Tokushima, Japan
E-mail: {kam,fuketa, aoe}@is.tokushima-u.ac.jp

***Abstract*—A dialogue communication between persons is efficient for elders. However elders do not have enough dialogue communications. On the other hand, many people are using mobile devices for communication. Cloud computing is a typical technologies to achieve these features efficiently. This paper presents a dialogue communication system for elders by using cloud-based approaches. The presented method proposes a dialogue controlling algorithm that can break communication. In the cloud systems, a text mining module is combined in order to analyze personal orientation from communicated utterances of individuals. From experimental results for a total of roughly 327 elders, it turns out that the presented method can make conversations smooth in comparison with the previous approaches.**

***Index Terms*—Dialogue communication, cloud computing, text mining**

I. Introduction

All over the world, the number of people aged over 65 has been increasing and the aging population has been growing rapidly. For the elders, it is very important to make their brains work actively and to stabilize themselves eventually. Conversations help to exercise elders' brains actively with listening to a conversational partner, understanding what the partner says, and thinking what to response. Thereby, it is possible to say that conversations are effective for elders [1]. But the problem is the elders are not able to have enough communications because of shortage of caregivers. Moreover, the method called reminiscence therapy is also effective for elders. The reminiscence therapy is to make elders' brains work actively by using conversation topics such as his/her old-time memories. By the reminiscence therapy, elders is to enrich QOL (Quality Of Life), and to decrease dysphonia [2][3][4][5].

On the other hand, many people are using mobile devices for communication. Cloud computing is a typical technologies to achieve these features efficiently. In cloud computing, the main part of the system is placed on a cloud. Therefore, the ubiquitous use from computer devices or system modification can be realized easily. Using this feature, it is applied to medical services [6], health monitoring systems [7], dialogue systems [8], and so on.

There are studies related to task-oriented dialogue technologies in the dialogue communication system. R. Freedman [9] developed APE (the Atlas Planning Engine) which is the reactive dialogue planner for intelligent tutoring systems. But there is no function to be conversational partners in task-oriented and dialog-based communication systems. The function to be conversational partners is to recognize conversational sentences from the human and to reply appropriate responses for less-task oriented conversations such as chats between a human and a computer.

Communication agents by Nakano [10] and Yasuda [11] are related to the dialog-based communication studies of conversational partners. The dialog-based communication agents use verbal acknowledgement based on the acoustic information such as pitch and intensity, and work with the following steps; 1) the agent ask questions to a user, 2) the agent responses positive answers for human's speeches to the questions.

This paper focuses on this QA approaches by agents because of simple architectures. From the viewpoint, one of practical estimations is the communication continuity without side effects for users. Another point is to construct a practical environment for large sets of dialogue knowledge. Therefore, this paper presents a cloud-based dialogue communication system and estimates the effectiveness from experimental results for 327 persons and for about 16,000 dialogue data.

II. Outline of Dialogue Communication Systems

The cloud-based dialogue systems to be discussed in this paper consist of the following four modules and terminal computer devices (Fig. 1).

1) DC (Dialogue Control) Module

This module generates two kinds of utterances to users by CK (Communication Knowledge) base which includes a lot of sentences. One utterance is questions to users and another is responses to answers by users to the questions. The interval of the bidirectional conversations and selection of utterances are controlled by this module. After conversations, some videos of school songs are offered if users are in the hyperexcitable state. That is to say, one cycle of dialogues is defined by questions from

Manuscript received August 30, 2013; revised September 30, 2013; accepted November 23, 2013.

doi:10.4304/jsw.9.8.2212-2216

the system, answers from users for them, and responses from the system for user's answers. Conventionally, questions, answers and responses are utilized in the following discussions if there are no confusions.

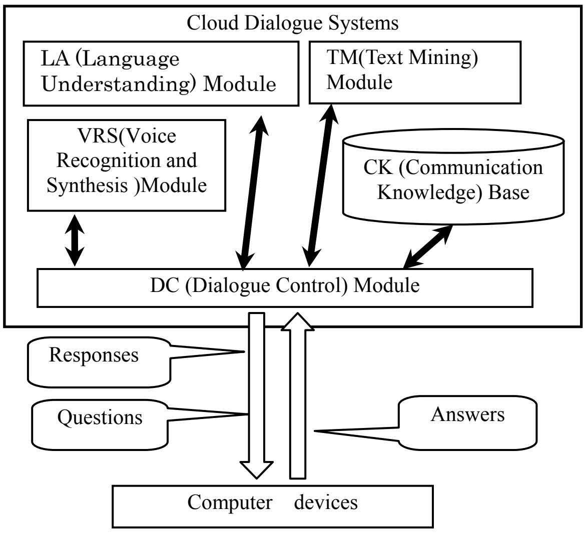

Figure 1 Relationships between a cloud dialogue system and a computer device.

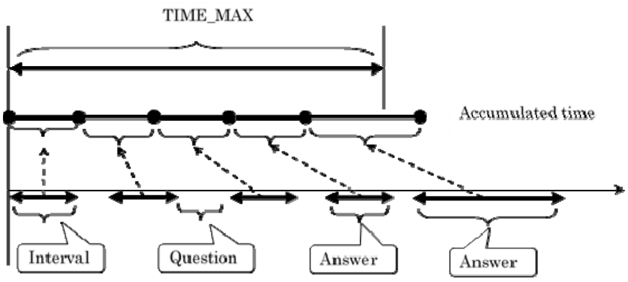

Figure 2 An illustration of the time-out scheme to be improved.

2) LA (Language Understanding) Module

This module understands answers and returns the intensions of users such as affirmation, negation, requested objects, the status of emotions, and so on. For question “Do you like an apple?”, answer “I like it” means affirmation and answer “No thanks” means negation. For question “What kinds of foods do you like?”, answer “apples” means requested objects from users. For question “How are you?”, answer “sad” means the negative status of emotion. There results can be generated by technologies [12][13].

3) TM (Text Mining) Module

All utterances from the system and from users are analyzed by the LA module and affective results are estimated in this module. The concept of this estimation is to find characteristics of users and the DC module reflects them on the conversations. Examples of characteristics include favorite foods, hobbies, interesting dialogue topics, ugly memory and so on.

4) VRS (Voice Recognition and Synthesis) Module

This module consists of different cloud-based services by voice recognition and voice synthesis systems. However, an additional signal processing module is introduced to detect whether there are users utterances or not, in order to support many voice misrecognitions. Moreover, the results of voice synthesis systems are improved by making the acoustic quality good.

The presented method proposes a dialogue controlling algorithm that can break communication if elders become hyperexcitable state. Video with school songs are introduced to reduce the side effect by hyperexcitable state. Cloud systems by using the algorithm are implemented in the DC module.

III. Dialogue Control Algorithm

A. Concept of Improvement of the DC module

Yasuda [11], Ogawa [14] and Fuketa [15] have discussed contributions by the dialogue-based communication system for elders. However, there is a problem that the system cannot terminate the communication. Therefore, it is necessary to terminate long speech by elders and to reduce the hyperexcite state. Although the problem depends on evoking their personal history, these extra impacts must be reduced. In the improvement, the maximum time TIME_MAX is introduced for the long speech. It enables us to utilize school songs for elders in hyperecitable state. Fig. 2 shows an illustration of the time-out scheme to be improved here.

In Fig 2, TIME_MAX is the total maximum time that elders are thinking and speaking in the whole conversations, but it has no question and response time from the systems. That is, accumulated time is the sum of times for interval and answering time, where interval time means waiting time from the end of questions to the beginning of answers.

B. Improvement of the DC Algorithm

The basic notations refer to the definitions by Fuketa [15]. SPEECH_AGENT(w), REPLY_ELDER(), and REPLY_AGENT(w, s) respectively mean speech action using speech content w by the agent, the function to return response contents from elders, and the function to return response contents from agents for speech contents w and elder response s.

Speech contents include triplet (x, y, z) for preliminary content x, real remembering content y and back-channel feedback content z. Examples of the contents x, y, and z are x="I’m 14 years old. I’m young, right?", y=" Please tell me how old you are." and z="I see.". A set of triplets ($x[q]$, $y[q]$, $z[q]$) for q from 1 to n controls the conversation between a user and an agent. The real remembering content y is divided into two levels by difficulties of speech contents and the levels are defined by simple and complex questions. The simple question (ex. "Please tell me how old you are.") is assuming Yes and No answers, and the complex question (ex. "How were the old days?") is assuming that elders must select one from some answer candidates after considering and remembering contents about questions.

An extended communication algorithm is below.

Improved Communication Algorithm

Input: MAX is the maximum number of questions and TIME_MAX is the maximum time to be terminated.

```
Method
(Step 1)  Initialize q by 1;
(Step 2)  Initialize t by 0 and temp_time by 0;
(Step 3)  CONFIRM(t);
(Step 4)  repeat COM(q)  until q = MAX or COM(q) = FALSE;
(Step 5)  if COM(q) = TRUE  then show school song;
(End of Algorithm)

function CONFIRM(t);
begin
   temp_time = temp_time + t;
   if temp_time >  TIME_MAX  then return TRUE;
   else return FALSE;
end;

function COM(q);
begin
   SPEECH_AGENT(x[q]);
   Wait for the interval TIME1;
   if CONFIRM(TIME1) is TRUE then return TRUE;
   if there are s = REPLY_ELDER() and t = TIME_ELDER()
   then
   begin
      if CONFIRM(t) is TRUE then return TRUE;
      REPLY_AGENT(x[q], s) ;
   end;
   if y[q] is the complex question then if SPEECH_COMPLEX() = TRUE then return TRUE;
   else
   begin
      SPEECH_AGENT(y[q]);
      Wait for the interval TIME2 ;
      if CONFIRM(TIME2) is TRUE then return TRUE;
      if there are s = REPLY_ELDER() and t = TIME_ELDER()
      then
      begin
         if CONFIRM(t) is TRUE then return TRUE;
         REPLY_AGENT(z[q], s) ;
      end;
   end ;
   return  FALSE;
end;

function SPEECH_COMPLEX()
begin
   SPEECH_AGENT(y[q]);
   Wait for the interval TIME3;
   if CONFIRM(TIME3) is true then terminate the algorithm;
   if there is s = REPLY_ELDER()
   begin
      REPLY_AGENT(z[q], s) ;
      Wait for the interval TIME4;
      if CONFIRM(TIME4) is TRUE then return TRUE;
      if there are s = REPLY_ELDER() and t = TIME_ELDER()
      then
      begin
         if CONFIRM(t) is TRUE then return TRUE;
         REPLY_AGENT(z[q], s) ;
      end;
      else break;
   end;
   else SPEECH_AGENT(y[q]);
   Wait for the interval TIME5;
   if CONFIRM(TIME5) is TRUE then return TRUE;
   if there are s = REPLY_ELDER() and t = TIME_ELDER()
   then
   begin
      if CONFIRM(t) is TRUE then return TRUE;
      REPLY_AGENT(z[q], s) ;
   end;
   else REPLY_AGENT(z[q], NULL) ;
   return FALSE;
end;
```

By a set of triplets ("I'm 14 years old. I'm young, right?", "Please tell me how old you are.", "I see.") as simple questions, steps S1) and S2) of the algorithm is explained as follows;

S1) the agent speaks preliminary content "I'm 14 years old. I'm young, right?" by SPEECH_AGENT. If the user responses something during TIME1, the agent answers "I see." by REPLY_AGENT.

S2) the agent speaks real remembering content “Please tell me how old you are.” by SPEECH_AGENT. If the user responses something during TIME2, the agent answers back-channel feedback content "I see." by REPLY_AGENT

Hereafter, by a set of triplets ("I'm 14 years old. I'm young, right?", "How were the old days?", "I see.") as complex questions, steps C1 and C2 of the algorithm is explained as follows;

C1) the agent speaks "I'm 14 years old. I'm young, right?" by SPEECH_AGENT.

C2) procedure SPEECH_COMPLEX is conducted. The agent asks complex real remembering content "How were the old days?" by SPEECH_AGENT in SPEECH_COMPLEX. If the user responses something during TIME3, the agent asks "Are there anything more?" by REPLAY_AGENT. If the user responses something during TIME4, the agent speaks "I see.” by REPLAY_AGENT. If the user responses something during TIME4, this question is finished. If the user doesn't response anything to the complex real remembering content during TIME3, the agent asks "How were the old days?" by SPEECH_AGENT again. If the user responses something during TIME5, the agent speaks "I see." by REPLAY_AGENT. If the user doesn't response anything, this means there is no response while the agent waits twice. In this case, the agent answers "Is this question difficult for you? Okay, let's go to next questions.", and this question is finished.

IV. Experimental Results

A. Dialogue Knowledge Bases

In the previous systems by Yasuda [11] and Fuketa [15], the total number of dialogues was 120, but this paper extends them to 16,163. Particularly, by observing many elders, 10320 dialogues are collected. The followings are details of each topic. "Celebrity" (Memories about old-time celebrity) is 3132, "Singers" (Old-time singers) is 2,382, "Food" in the past and in the present is 2,406 and "Place names" associating hometown is 2,400.

B. Recovering Voice Misrecognitions

As all voices from the elders are very difficult to be recognized as proper languages, default responses for answers are introduced in order to recover voice misrecognitions. The default responses are possible to keep the coherency of dialogue topics based on questions. For this reason, the default responses are suitable for affirmation, negation, and ambiguous answers; ambiguous answers means unclear and no information. As for question "Did you go to the supermarket yesterday? ", the default response includes answer "I like shopping, so I went to the supermarket yesterday." which can be used for affirmation "Yes I did.", negation "No", and unclear answer "There was a fire…". It is turned out that the rate of keeping the proper coherency is 100% from experimental results for dialogue cycles of 300 sampling questions, and the proposed method shows the outstanding improvement compared with simple response "uh huh" of the previous system.

Generated voices for the elders are very difficult to build in the voice synthesis system due to misreading and partial low acoustic quality. In order to adjust the quality to the elders, four persons have tried the system. Also, ten elders were selected to observe by the slow speed. Because confirmation of the acoustic quality requires understanding of proper meanings, five healthy persons from 30 to 61 of age participated the evaluation. The experimental results for sampling data are shown in Table I. In Table I, "A" means the number of confirmed sentences, "B" means the number of sentences with partial voice change, "C" means the rate of partial voice change by C =(100*B)/A), "D" means the number of well-understood sentences with showing written sentences, and "E" is the rate of meaning understanding by E=(100*D)/A. According to Table I, the rate of the meanings understanding is 100%, but the rate of C is very low percentages.

TABLE I.
EXPERIMENTAL RESULTS FOR GENERATED VOICE

Classification	A	B	C	D	E
preamble questions	325	99	3%	325	100%
Main questions	602	174	3%	602	100%
Affirmative responses	603	121	2%	603	100%
Negative responses	1,145	235	2%	1,145	100%
Default responses	323	44	1%	323	100%

C. Effects of Voice Dialogue Systems

The experiments by totally 4,683 questions were conducted for a total of roughly 327 elders in the hospital and care facilities. The rate of the communication continuity over ten minutes and the rate of finishing all questions are shown in Tables II and III show, respectively. The rate of the communication continuity over ten minutes or finishing all questions is shown in Table IV, where DN is the number of dialogue sets, QN is the number of questions, MORE(LESS) are the numbers of dialogue sets more and equal (less) than ten minutes, and RATE_ C(%) is the rate of dialogue stets over ten minutes to all sets by RATE_C = (100*MORE)/DN.

The results from Table II, show that the rate of the communication continuity over ten minutes is 15.9%. Meanwhile, the results from Tables III and IV show the rate of finishing all questions is 73.1% and the rate of the communication continuity over ten minutes or finishing all questions is 75.2%, respectively. From the above-mentioned results, it turns out that the presented dialogue system is very practical. Because this experiment restricted one set dialogue, the low rate of 15.9% is appeared as a result.

Next, experiments by the systems with TIME_MAX (30 minutes) carried out for fifteen elders in a nursing facility and at home. Table V shows results of the rate of the communication continuity over ten minutes and it turns out that the communication continuity over ten minutes is 73.3%. As for the maximum conversation time,

TABLE II.
RATE OF THE COMMUNICATION CONTINUITY OVER TEN MINUTES

DN	QN	MORE	LESS	RATE_C
327	4,683	52	275	15.9%

TABLE III.
RATE OF FINISHING ALL QUESTIONS

DN	QN	FINISH	NON_FINISH	RATE_F
327	4,683	239	88	73.1%

TABLE IV.
RATE OF THE COMMUNICATION CONTINUITY OVER TEN MINUTES OR FINISHING ALL QUESTIONS

DN	QN	MORE or FINISH	LESS and NON_FINISH	RATE_M
327	4,683	246	81	75.2%

TABLE V.
RATE OF THE COMMUNICATION CONTINUITY OVER TEN MINUTES

DN	QN	MORE	LESS	RATE_C
15	372	11	4	73.3%

two elders reached it, and one of them became hyperexcitable state with speaking continuously. In order to solve this situation, a school song video has been used and has helped the person to calm down. Thereby, it turned out that appropriate dialogue communication systems requires the upper bound of conversation time when users speak very actively.

V. Conclusion

This paper has presented a dialogue communication system for elders by using cloud-based approaches. The presented method has proposed a dialogue controlling algorithm that can break communication. In the cloud systems, a text mining module has combined to analyze personal orientation from communicated utterances of individuals. From experimental results for a total of roughly 327 elders, it has been verified that the presented method can make conversations smooth in comparison with the previous approaches.

In the future, some of modules must be connected by other modules on cloud environments.

References

[1] L. Dubuc and A. Blackwell, "Opportunities for augmenting conversation through technology for persons with dementia," in *Proc. International Conference on Accessible Design in the Digital World (ADDW)*, 2005.

[2] E. Bohlmeijer, M. Valenkamp, G. Westerhof, F. Smit and P. Cuijpers, "Creative reminiscence as an early intervention for depression: results of a pilot project," *Aging & Ment al Health*, vol.9, no.4, pp.302-304, 2005.

[3] E. A. Cook, "Effects of reminiscence on life satisfaction of elderly female nursing home residents," *Health Care for Women International*, vol.19, no.2, pp.109-118, 1998.

[4] B. K. Haight, "The therapeutic role of a structured life review process in home bound elderly subjects," *J. Gerontology*, vol.43, no.2, pp.40-44, 1988.

[5] Y. C. Lin, Y. T. Dai and S. L. Hwang, "The effect of reminiscence on the elderly population: a systematic review," *Public Health Nursing*, vol.20, no.4, pp.297-306, 2003.

[6] N. Z. Hu, C. Y. Lee, M. C. Hou and Y. L. Chen, "A Cloud System for Mobile Medical Services of Traditional Chinese Medicine," *J. Medical Systems*, vol.37, no.6, pp.1-13, 2013.

[7] A. Bourouis, M. Feham, A. Bouchachia, "Ubiquitous Mobile Health Monitoring System for Elderly (UMHMSE)," *International Journal of Computer Science & Information Technology*, vol3, no.3, pp.74-82, 2011.

[8] M. Fuchs, N. Tsourakis, M. Rayner, "A scalable architecture for web deployment of spoken dialogue systems," in *Proc. International Conference on Language Resources and Evaluation(LREC)*, pp.1309-1314, 2012.

[9] R. Freedman, "Using a reactive planner as the basis for a dialogue agent," in *Proc. Thirteenth International Florida Artificial Intelligence Research Society Conference (FLAIRS)*, pp.203-208, 2000.

[10] Y. Nakano, J. Hikino and K. Yasuda, "Conversational Agents for Dementia: A Field Study for a Prototype System," in *Proc. The 25th Annual Conference of the Japanese Society for Artificial Intelligence*, pp.1-4, 2011.

[11] K. Yasuda and J. Aoe, "A Computer Agent System for Conversing with Individuals with Dementia," in *Proc. The 11th International Conference on Alzheimer's & Parkinson's Diseases*, 2013.

[12] Y. Kadoya, K. Morita, M. Fuketa, M. Oono, E. S. Atlam, T. Sumitomo and J. Aoe, "A Sentence Classification Technique by Using Intention Association Expressions," *Computer Mathematics*, vol.82, no.7, pp.777-792, 2005.

[13] T. Yoshinari, E. S. Atlam, K. Morita, K. Kiyoi and J. Aoe, "Automatic acquisition for sensibility knowledge using co-occurrence relation," *Int. J. Computer Applications in Technology*, vol.33, nos.2/3, pp.218-225, 2008.

[14] T. Ogawa, H. Bando, K. Morita, M. Fuketa and J. Aoe, "A Method of Combination of Language Understanding with Touch-Based Communication Robots," *Int. J. Intelligence Science*, vol.2, no.4, pp.71-82, 2012.

[15] M. Fuketa, K. Morita and J. Aoe, "Agent-based communication systems for elders using a reminiscence therapy," *Int. J. Intelligent Systems Technologies and Applications*, vol. 12, nos. 3/4, pp.254-267, 2013.

Kazuhiro Morita is an Associate Professor at the University of Tokushima. where he was a Research Assistant from 2000 to 2006 in Information Science and Intelligent Systems. He is a member of the Information Processing Society of Japan and IEEE Computer Society. He received his BSc, MSc and PhD in Information Science and Intelligent Systems from the University of Tokushima, Japan in 1995, 1997 and 2000, respectively. His research interests are sentence retrieval from huge text databases, double-array structure and binary search trees.

Masao Fuketa received his BSc, MSc and PhD in Information Science and Intelligent Systems from University of Tokushima, Japan, in 1993, 1995 and 1998. He has been a Research Assistant from 1998 to 2000 in Information Science and Intelligent Systems from University of Tokushima, Japan. He is currently a Research Associate in the Department of Information Science and Intelligent Systems, Tokushima University, Japan. He is a member in the Information Processing Society in Japan and The Association for Natural Language Processing of Japan. His research interests in sentence retrieval from huge text databases and morphological analysis.

Jun-ichi Aoe received his BSc and MS in Electronic Engineering from the University of Tokushima, Japan, in 1974 and 1976, respectively, and PhD in Communication Engineering from the University of Osaka, Japan in 1980. Since 1976, he has been with the University of Tokushima. He is currently a Professor in the Department of Information Science and Intelligent Systems, Tokushima University, Japan. His research interest is on natural language processing, a shift-search strategy for interleaved LR parsing, robust method for understanding NL interface commands in an intelligent command interpreter, and trie compaction algorithms for large key sets. He was the Editor of the Computer Algorithm Series of the IEEE Computer Society Press. He is a member in the association for computing machinery, the association for the natural language processing of Japan.