Picture information shared conversation agent: Pictgent

Miki Ueno[a], Naoki Mori[a], Keinosuke Matsumoto[a]

[a] Department of Computer Science and Intelligent Systems, Osaka Prefecture University
1-1 Gakuencho, Nakaku, Sakai, Osaka 599-8531, Japan;
e-mail: ueno@ss.cs.osakafu-u.ac.jp ; TEL: 072-254-9273; FAX: 072-257-1722

KEYWORD

Pictgent
Chatterbot
Estimating User's Interests
Picture Information
Dialogue System

ABSTRACT

Recently, the various dialogue systems have been proposed to make a natural conversation with users.

In this paper, we propose a novel dialogue system called Pictgent which utilizes "pictures" with model of situation in order to share common knowledge between users and the system. We show the basic concept and system structure of proposed Pictgent. Statistical experiments are carried out in order to confirm the effectiveness of Pictgent.

1 Introduction

Recently, the text based communication on the internet developed remarkably, various type of dialogue systems are required in lots of fields. Task-oriented dialogue systems like SHRDLU[WINOGRAD, T. 1972] or based on expert systems have been widely applied to domain specific area such as product search or travel information. However, this type of dialogue systems can not deal with free conversation. Several non-task-oriented dialogue systems have already been proposed: ELIZA[WEIZEMBAUM, J. 1966] and A.L.I.C.E [WALLACE, R. 2009] are popular one of the well-known non-task-oriented dialogue system. These systems work very well although both methods only use simple pattern match and make questions based on user's past inputs.

However, those systems have problems such as topics scattered. It is difficult to introduce personality into system and users lose interests on communicating system easily. To solve user's interest problem, we proposed a new chatterbot which can estimate user's interest during conversation [UENO, M. *et al*. 2010] [UENO, M. *et al*. 2012].

Our chatterbot and previous dialogue systems have obvious limitation for communication between user and system with only text information because nonverbal information like expression or cultural background is very important in human actual conversation[SCHANK, R. 1990]. Although lots of researchers have proposed effective methods such as using template or statistical information to make dialogue systems, there have not been any dialogue systems which can satisfy users.

In this study, we propose a novel dialogue system called *Picture Information Shared Conversation Agent* (**Pictgent**) that shares conversational background knowledge by showing prepared pictures with model to user. To make the Pictgent useable dialogue system, we also propose the representation of picture model and structure of scenario. We expect Pictgent to be able to solve problems of common dialogue systems by showing pictures.

In this paper, we first present related studies and show the position of our study in Section 2. We show the constitution of the proposed Pictgent in Section 3 and explain how to estimate user interest in Section 4. Computer experiments are described in Section 5, while Section 6 shows statistical experiments of system in scientific events. Finally, in Section 7, we present the conclusions of this study.

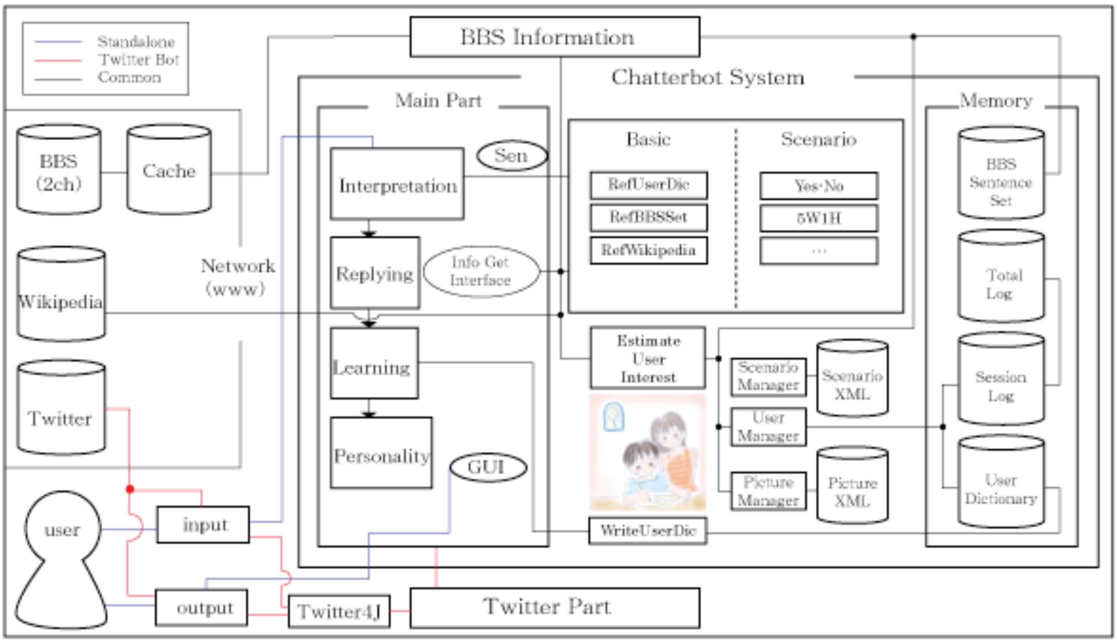


Fig. 1. Diagram of Pictgent

2 Related studies and position of our study

In this section, we describe previous studies which are related to our study.

2.1 Single/plural meaning of pictures

The method that adds text information based on semantic ontology to usual pictograms and the tool that creates new symbolic pictures have been proposed[ITO, K. *et al*. 2007]. In this method a picture made of pictograms can be translated into a sentence and a sentence can also be translated into pictures made of pictograms. In this method, the most important requirement is that created pictures must have only one meaning because target users of this method are people who can not read, but can understand significant information such as disaster prevention through the picture, easily.

Methods for determining word's semantic by showing pictures have been proposed. In these methods, ambiguity of word's semantic or variation of user's impression is not appreciated. In our study, we assume that the relation between one picture and the user's impression is one to many and this feature is a very important aspect in understanding picture. Using variation of human ideas of understanding pictures is applied in psychology for Thematic Apperception Test(TAT) [MORGAN, C. D. *et al*. 1935]. In TAT, there are neither right answers nor communication. However, our proposed Pictgent is a communication based dialogue system that accepts the ambiguity of understanding pictures.

2.2 Putting tags into pictures and structure of tags.

One of the important methods for adding text information to pictures is putting tags into pictures. The system for statistically tagging scenes by users on the Web has been proposed. ESP game[AHN, L. V. *et al*. 2004] is a method for tagging a picture with two user's opinions. User's score increase when an input tag is same as another user's input. Google Image Labeler adopted this method to put reasonable tags into pictures. Tag information is effective and easy to use in keyword search, but tag expression has no structure. Besides this, it is difficult to use tags during a conversation.

Table 1. Main tags of picture model

tag name	explanation.
<character>	person object.
<base>	unchangeable features with person on changing scene.
<scene>	information according to each scene.
<action>	behavior with target.
<state>	object’s inner state and attendant circumstances without target.
<relation>	social and static relationship with other person. (0..*)
<part>	changeable parts in the scene. (0..*)

```
    <character>
    <base id=”1” type=” human” name=”hana” sex=” fe-
male” age=”6” />
    <scene>
    <position>right</position>
    <expression> wonder</expression>
    <emotion target=”2”>thrilled</emotion>
    <action name=”see” target=”2” />
    <state>
    <physical></physical>
    <mental></mental>
    </state>
    </scene>
    <relation name=”-“ role=”-“ target=”2” />
    </character>
     (…)
    <layer>
    <part id=”1” visible=”true”>
    <part id=”2” visible=”false”>
    <part id=”3” visible=”true”>
    <part id=”4” visible=”true”>
    <part id=”5” visible=”true”>
    </layer>
```

Fig. 2. Part of picture model XML
for person information and picture parts

In our study, we also investigate which kinds of tags are required in conversation.

Table 2. DB table of picture parts

partid	sceneid	imagename	filepath
1	1	coat, jacket	coat1-1.png
1	2	coat, jacket	coat1-2.png

＊ Several image names written by comma

2.3 Positioning our study

There is no research which considers the relation between several scenes or actor and object. Putting information in passive contents such as watching movies depends on users will, it is difficult to expect users to put tags actively. However, Pictgent shows sentences of the current scene and asks users about pictures during conversation, this process makes users put tags actively.

To make a dialogue system, the most important factor is considering the human ability for understanding implicit situation. To include this property into our system, we made scenarios as sequence of several picture models in order to represent implicit intention between scenes and to help user understanding a current situation. We can introduce a story branch into scenario made of pictures, and expect that users are interested in those scenarios. Pictgent also allows any conversations which have no relation to current scenario in order to reduce the burden that users have to obey scenario.

From the point of view of understanding intention in Pictgent, we can remove ambiguity of user's intention by assuming that user's input has some relations to the current picture.

In the transition between picture models, it is enough to check if user is staying on the current picture or transiting to the next picture from user input.

Pictgent can be applied to broad fields, our first target is children. We would like to make Pictgent be testbed framework for application of conversation and picture in engineering.

3 Constitution of Pictgent

In this chapter, we describe the system outline of the Pictgent. Pictgent is made of the following 3 modules.

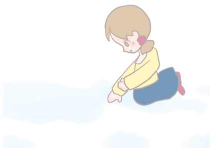


Fig. 3. Original Pic.1

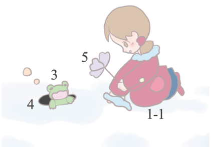


Fig. 4. Modified Pic.1

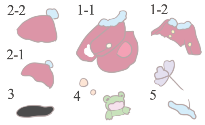


Fig. 5. Picture parts

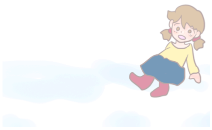


Fig. 6. Original Pic.2

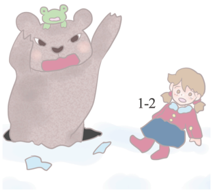


Fig. 7. Modified Pic.2

Picture Module This module manages the picture's information that is important items in this study. Each picture has its own model created by object oriented modeling technique. In this module, picture model is written in XML format.

Scenario Module This module controls transition between scenarios of pictures. In order to achieve adequate transition, this module stores user's inner state as numerical vector according to user input history. Scenarios are written in XML format containing transition map and answer example set.

Chat Module This module replies according to user's input. There are following two different reply modes in this module.

Scenario mode During user is following the scenario, the scenario mode is used. In this mode, Pictgent can talk with users utilizing picture's and scenario's XML.

Chatterbot mode When user loses interests and strays from a scenario, the mode is set to chatterbot mode. In this mode, Pictgent replies based on various topics while estimating the user interests.

Fig. 1. shows the outline of the Pictgent. The system is written in Java. We introduced multi-thread programming for input and output part in order to make Pictgent reply not only when user has finished inputs. We set an interval time for the system to manage spontaneous reply.

Table 3. Main tags in scenario XML

Tag's Name	Explanation
<unit_phase>	UnitPhase object.
<uid>	Id of UnitPhase.
<out>	Output while loading UnitPhase. (1..*)
<con>	Condition of loading Phase.
<yes>	Output if user's input is positive. (1..*)
<no>	Output if user’ s input is negative. (1..*)

Interpretation Pictgent receives an input and formats it by morphological analysis.

Replying Pictgent replies to user with an appropriate expression that is retrieved from several databases(DB) and Web information.

Learning Pictgent will ask user about user's input sentence or any unknown keywords. It can memorize new statements and revise its memory.

Personality Pictgent maintains a self-portrait and uses specific end of phrases.

3.1 Loading picture model

1. XML written as picture model is loaded by XML parser.
2. Find the upper node named character and translate all under nodes of character node to parameters by using attributes and tag names.
3. Instance of class Person is created using parameters from 2. All instances are put into an instance of ArrayList<Person>.

3.2 Picture model

In this study, we use picture information based on object oriented modeling. Not only human characters but any parts of the picture can be defined as object, we only define main characters as objects in this paper. Picture information is written in XML and is editable.

Fig. 2. shows an example of part of XML focused on a character and picture parts in Fig. 4. Table 1. shows main tags of our XML.

According to scenario and user's input, Pictgent changes picture. Figs. 3. to 7. show examples of picture and picture parts in the same scenario.

In Pictgent, there are two types of picture changes. First one is large change like completely replacing picture to represent scenario transition. On the other hand, there exists small change like decorating current picture such as Fig. 3. to Fig. 4. or Fig. 6. to Fig. 7. If a small change is static change, then the small change affects the next scene. For example, if Pictgent changes the clothes in Fig. 3. to that in Fig. 4, this change takes over the next scene in Fig. 7. In this process, shape or status of changing part depends on the scene, appropriate picture part is loaded from DB by using complex key consisted of id of picture parts and id of picture model.

Table 2. shows parts of table of picture parts DB. Each scenario also represented in XML.

Table 3. represents main tags in scenario XML. Fig. 8. represents an example of UnitPhase in scenario XML of Fig. 4.

3.3 Using picture model

The output sentence for reply sometimes contains specific tags of picture model XML. In such case, all tags are replaced to suitable string.

For example, tags like <personname: i > is replaced to name of i-th instance of Person.

3.4 Loading scenario XML

1. Scenario XML is loaded by XML parser.
2. Find the upper node named turn and translate all under nodes of turn node to parameters by using attributes and tag names.
3. Instance of class Turn is created using parameters from 2. All instances are put into an instance of ArrayList<Turn>.

3.5 Using scenario XML

1. Read current scenario XML.
2. Read picture models of current scenario XML.

```
    <unit_phase>
    <out>"Too cold... Give me back my roof!". A frog comes out from the hole on the ground covered with snow. What is the roof?</out>
    <out>What is the thing that <personname:1> has?</out>
    <con>null</con>
    <yes next="1">Yeah. The flower is frog's roof.</yes>
    <no next="2">Uh... The girl hears a strange sound under the frog.</no>
    </unit_phase>
```

Fig. 8. Example of scenario XML

3. Set the current position number of scenario to i. Load i-th element T_i of ArrayList<Turn> from 3.4 and check condition in con-node. If this condition is true, output part of T_i is pushed into out put queue. Otherwise, go to Step. 5.
4. Accept user's input. This input is labeled positive example or negative example. Both positive and negative cases have their own output string and transition-ID. Output string is push into output queue. Scenario transition based on transition-ID occurs.
5. If queue has transition-ID, this value is substituted to current position i. If queue does not have transition-ID, $i = i + 1$.
6. If $i <$ size of list of scenario XML, go to Step 3. Otherwise, end the scenario.

4 Estimating current user interests

Generally, user interests changes during conversation. If the Pictgent uses all logs equality, it is difficult to estimate current topics. To solve this problem, we propose the following method which estimates current user interests by emphasizing the current user input. Pictgent do not care about the current user interests in scenario mode because there already exists transition map. Therefore, we focus on the “chatterbot” mode of Pictgent in the following discussion.

Pictgent has a user α's interest vector $\boldsymbol{c}^{\alpha}$ which represents user's internal state. Each vector element of $\boldsymbol{c}^{\alpha}(0 \leqq \boldsymbol{c}^{\alpha}_{j} \leqq 1)$ relates to category j of stored BBS board’s information. If the degree of the user interests for the related board is maximum, the value of the vector element becomes 1. On the other hand, if the user has no interest in that board, the value becomes 0. The initial value of all elements is set to 0.

The proposed Pictgent tries to answer based on the topic of board of larger $\boldsymbol{c}^{\alpha}{}_{j}$. However, $\boldsymbol{c}^{\alpha}$ is made by all logs, so the proposed Pictgent may fail to understand the current topic. To avoid this, the proposed Pictgent decides the topic of answer as follows:

We define the latest input in time step T as S_T . Old inputs are represented by S_{T-1}, S_{T-2}, ..., S_1. We also define the changing topic vector σ_{S_T} as follows:

$$\left\{\sigma_{S_T}\right\}_j = \sum_{x=1}^{W} N_x R_x^j \tag{1}$$

where N_i is the frequency of word $_i$ in S_T and $R^j{}_i$ is the frequency word i in board j.

Next, the topic vector t^{α} which i-th element represents the current significance of board i is defined.

We set the default value of all elements of t^{α} to 0. The proposed Pictgent calculates t^{α} as follows whenever the proposed Pictgent obtains the newest input S_T.

$$t^{\alpha} = \sum_{i=1}^{T} \gamma^{i-1} \sigma_{S_{T-i+1}} \tag{2}$$

where γ, $0 \leqq \gamma \leqq 1$, is the discount rate. As γ decreases, the influence of past inputs decreases. We define normalized unit vector of t^{α} as $\hat{t}^{a}$.

Since we assume that user interests are constant in short term, we consider that the weight of past inputs and that of current inputs are the same.

Therefore $\boldsymbol{c}^{\alpha}$ is obtained by $t^{\alpha}{}_{\gamma=1}$ for $\gamma = 1$ as follows:

$$t^{\alpha}_{\gamma=1} = \sum_{i=1}^{T} \sigma_{S_{T-i+1}}, c^{a} = \hat{t}^{a}_{\gamma=1} \tag{3}$$

The proposed Pictgent decides the board to use by referring to $\boldsymbol{c}^{\alpha}$ and $\hat{t}^{a}$. We define board deciding vector $\boldsymbol{b}$ as follows:

$$b = \eta c^{a} + (1-\eta)\hat{t}^{a} \tag{4}$$

where $0 \le \eta \le 1$. The proposed Pictgent selects board i in proportion to $\boldsymbol{b}_i$.

Table 4. Experimental conditions

Focused board	Tennis, Fashion
Number of sentences sampled from board A : n	10
Number of sentences sampled from board B : m	10
Number of sentences per log : $n + m$	20
Times	100
Coefficient η	0.2, 0.5
Discount rate γ	0.2

5 Experiment 1

To evaluate the algorithm for estimating user's interest, we did a statistical experiment using text of boards from 2ch which are already categorized.

To prepare the experiment, we applied the following processing.

1. Consider *Katakana* words as noun.
2. Combine continuous nouns into one word.
3. Remove stop-words from test data.

5.1 Setting of experiment 1

1. In this experiment, we selected two boards "Tennis" and "Fashion". "Tennis" is set to board A and "Fashion" is set to board B.
2. Top 10,000 frequent words of board A and board B on Dec 1st, 2010 are defined as the group of words for estimating user interest.
3. Text data of board A and board B posted from May 26th, 2011 to June 1st, 2011 were separated into sentences and put each boards' sentences into U_A and U_B respectively. Those sentences are used as pseudo user inputs in the following experiment. We select n sentences from U_A and m sentences from U_B randomly. In experiment, n sentences from U_A are used first and then m sentences from U_B are used after finishing U_A sentences. Finally, n + m inputs are evaluated

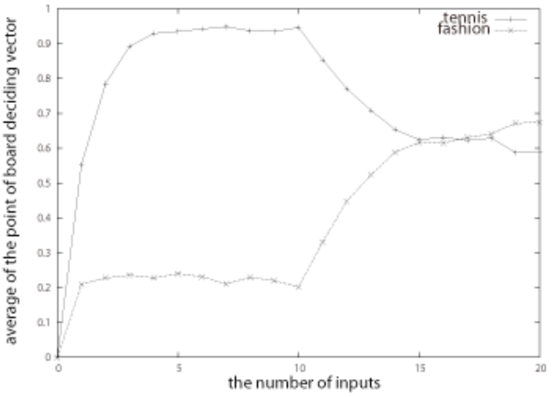


Fig. 9. Result of experiment 1-1

level of interests of the corresponding board. In this experiment, board define vector has two elements corresponding to board A and board B. The variation of each element is observed in order to check the tendency of 2ch boards and definition of boards define vector.

Table 4. displays conditions of experiment. In experiment 1-1, we set $\eta =$ 0.5, while in experiment 1-2, we set η = 0.2. η is the parameter controlling the sensitivity of latest input. 100 results of board define vector of different random seeds are obtained.

To make the effectiveness of board define vector ***b*** clear, we did a t-test(two-sided test) every step between 100 trials of the first element of ***b*** as interest level of board A and that of the second element of ***b*** as interest level of board B.

5.2 Result

In this section, the x-axis shows the step of input and the y-axis shows the average of ***b*** element in 100 trials. In step 11, the source of pseudo input is changed from $\mathrm{U_A}$ sentences to $\mathrm{U_B}$ sentences.

In Fig. 9. and Fig. 10, the variation of level of interests of board A and that of board B are shown as 2 lines.

Fig. 9. shows the results of experiment 1-1. From step 1 to step 12, there exists significant difference at the significance level 1%. On the other hand, after step 13, there was no significant difference at even the significance level of 5%.

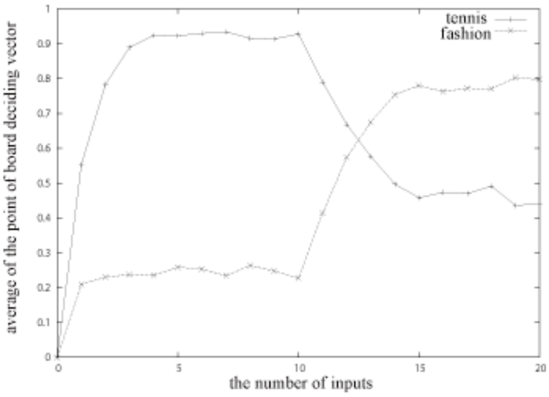


Fig. 10. Result of experiment 1-2

significant difference from step 1 to step 11 and from step 14 to step 20 at the significance level of 1%.

5.3 Consideration

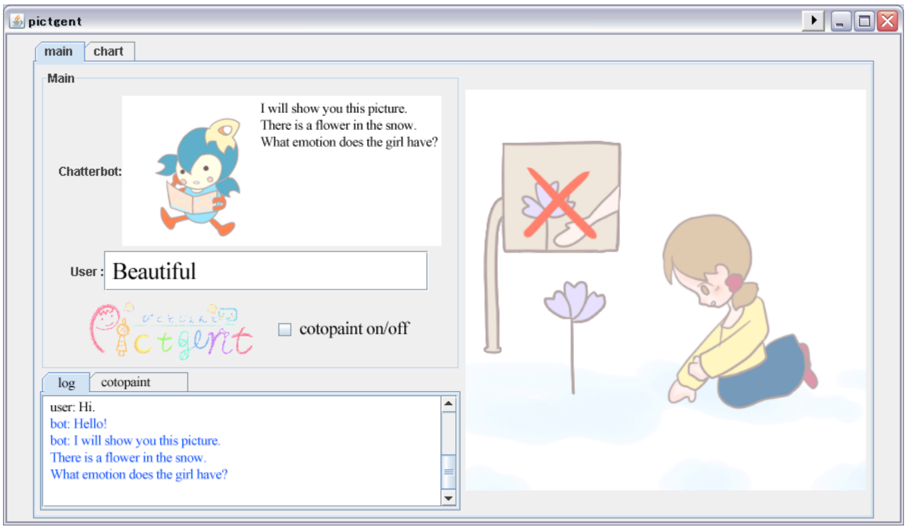
In result 1-1, we obtained that there is no significant difference at the significance level of 1% after step 13. This is because the sensitivity of the current input is low in high η. In other words, it is difficult to focus on the latest topic in high η.

In result 1-2, we obtained that there was significant difference except just after changing topics in step 11. In case of low η, latest topic has large influence and the level of interest changes rapidly. We need to define η to balance between original interest and current interest.

From the point of view of data set, 2ch data have strong relation even if there is a 6 months interval between word set for estimation and pseudo inputs. This result suggests using 2ch data for estimating user interest is reasonable.

6 Experiment 2

In Pictgent, there are 2 kinds of questions. One is aim questions which have obvious answer like numerical calculation. The other is emotional questions which involve ambiguous feelings or emotion such as character's heart in scenario. In experiment 2, we analyzed correlation between two emotional questions to confirm the effectiveness of branch in scenario mode.



I will show you this picture.
There is a flower in the snow.
What emotion does the girl have?

Beautiful

```
user: Hi.
bot: Hello!
bot: I will show you this picture.
There is a flower in the snow.
What emotion does the girl have?
```

Fig. 11. Outline of Pictgent for experiment 2

Fig. 11. shows outline of Pictgent used in scientific event for children. We customized the look of Pictgent in order that children can use it easily. 4 laptop computers installed Pictgent with one staff for helping children were arranged. The total number of users was 192 and they were between 4 and 13 years old. The event title of our booth was "Let's talk with computer! -Pictgent-".

6.1 Experiment procedure

The settings of Pictgent are as follows:

- Scenario contains 5 sub-scenario XMLs.
- The number of picture and picture model is 4.
- The number of picture parts for decorating basic pictures is 10.

It takes about 5 minutes to finish the conversation with Pictgent. We analyzed each user's log which has input string and interval between inputs. The log of input strings of users and interval times between inputs are stored for analysis. Fig. 11. shows the important picture of experiment 2.

6.2 Procedure of experiment 2

1. All users reached the scene shown in Fig. 10. after passing several pictures.
2. We focus on 2 emotional questions Q.1 and Q.2 asked in Fig. 11. We defined answers for those questions as A.1 and A.2 respectively. The details of Q.1 and Q.2 are as follows.
   - Q.1 "There is a flower on the snow. What emotions does the girl have?"
   - Q.2 "Do you think the girl will pick the flower?"
3. Chi-square test was used to determine whether there is the significant association between A.2 and following topics.
   a) User's age
   b) User's sex
   c) A.1

6.2.1 Experiment 2-1

We separated all A.2 data into 3 groups labeled yes/no/unknown manually. We investigated whether user's sex and A.2 are independent and whether user's age and A.2 are independent by chi-square test. In this experiment, user's sex is categorized into man and woman, and user's age is categorized into 4 to 7 years old group and upper 7 years old group.

e

We separated all A.1 data into 3 groups labeled positive/negative/other manually. After this procedure, we removed the "other" group from A.1 data. Because the size of the "positive" group is 92 and the size of "negative" groups is 37, so totally 92+37=129 data from A.1 were used in this experiment. We investigated whether A.1 without "other" group and A.2 were independent.

6.3 Results and consideration

The results of experiment 2-1 are that A.2 is independent from both user's age and user's sex at the significance level of 5%. The result of experiment 2-2 is that A.2 is not independent from A.1 at the significance level of 1%.

Those results indicate that user's selection in scene from Pictgent represents user's inner state better than attributes such as user's age and user's sex. We expect the input logs of Pictgent to be used to check mental condition of children.

7. Conclusion

In this paper, we proposed a novel dialogue system called Pictgent (*Picture Information Shared Conversation Agent*) which shows picture with picture model in order to share common knowledge between the user and the system. We described the constitution of Pictgent such as picture module, scenario module and chat module. It has been confirmed that the proposed Pictgent is effective to understand the inner state of users. Improving the three modules of Pictgent and creating new pictures from user's input are subjects for further study.

Acknowledgment

This research was supported in part by a Grant-in-Aid for Scientific Research(C), 22500208, 2010-2014 from the Ministry of Education, Culture, Sports, Science and Technology.

References

[AHN, L. V. *et al.* 2004] AHN, Von Luis, DABBISH, Laura. Labeling images with a computer game, CHI'2004, 319-326, 2004

[ITO, K. *et al.* 2007] ITO, Kazunari et al. SVG Pictograms with Natural Language Based and Semantic Information, SVG Open, 2007

[MORGAN, C. D. *et al.* 1935] MORGAN, C.D., MURRAY, H.A., A method of investigating fantasies. Archives of Neurology and Psychiatry, 34, 289-306, 1935

[SCHANK, R. 1990] SCHANK, Roger. Tell me a story: a new look at real and artificial memory, New York: Scribner. 1990

[UENO, M. *et al.* 2010] UENO, Miki. MORI, Naoki. MATSUMOTO, Keinosuke. Novel Chatterbot System Utilizing Web Information, Distributed Computing and Artificial Intelligence Advances in Soft Computing, Springer, Volume 79, 605-612, 2010

[UENO, M. *et al.* 2012] UENO, Miki. MORI, Naoki. MATSUMOTO, Keinosuke. Picture Information Shared Conversation Agent: Pictgent., Distributed Computing and Articial Intelligence Advances in Intelligent and Soft Computing, Springer, Volume 151, 91-94, 2012

[WALLACE, R. 2009] WALLACE, Richard. The Anatomy of ALICE. http://www.alicebot.org/anatomy.html.

[WEIZEMBAUM, J. 1966] WEIZENBAUM Joseph. ELIZA-A Computer Program For the Study of Natural Language Communication Between Man and Machine, Commun. ACM 9[1], 36-45. 1966

[WINOGRAD, T. 1972] WINOGRAD, Terry. Procedures as a Representation for Data in a Computer Program for Under-standing Natural Language. Cognitive Psychology Vol. 3 No 1, 1972