Toward Gamified Knowledge Contents Refinement
Case Study of a Conversation Partner Agent

Takayuki Iwamae[1], Kazuhiro Kuwabara[2] and Hung-Hsuan Huang[2]
[1] *Graduate School of Information Science and Engineering, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan*
[2] *College of Information Science and Engineering, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan*

Keywords: Crowdsourcing, Gamification, Knowledge Refinement, Conversation Partner Agent.

Abstract: Rich knowledge contents are necessary to develop an intelligent agent that interacts with people and supports their communication or their activities. In this paper, we choose a conversation partner agent for people with aphasia as an example and propose a method that interactively acquires and refines knowledge contents for the agent. The proposed method is invoked when a problem is found in the knowledge contents and utilizes the concept of gamified crowdsourcing. Gamified tasks verify the data input by a user. By utilizing a crowdsourcing approach, we strive for more accurate knowledge contents. The paper presents its game design and an example scenario.

1 INTRODUCTION

A large amount of knowledge contents is required to develop an intelligent agent to support human communication or activities. Knowledge contents may be constructed from data available on the Internet using machine learning technologies, for example. On the other hand, the concept of crowdsourcing can also be applied to construct various knowledge contents by harnessing the power of hundreds or thousands of people. In such crowdsourcing platforms as Amazon Mechanical Turk (MTurk)[1], many microtasks, often called Human Intelligence Tasks (HITs), are shared with many workers over networks. An example of such a task is labeling an image.

In crowdsourcing, giving proper incentives to workers is important for eliciting both qualitative and quantitative better performances. A typical reward is monetary, which can be considered an extrinsic motivation. The importance of intrinsic motivation has also been pointed out (Ryan and Deci, 2000). Gamification, which is defined as the use of game design elements in non-gaming contexts (Deterding et al., 2011), is a popular method to provide intrinsic motivation. In crowdsourcing, the idea of gamification is widely utilized (Morschheuser et al., 2016).

In this paper, we propose a method that utilizes a gamified crowdsourcing approach for knowledge contents refinement and focuses on a case of fixing problems in knowledge contents. The proposed method is intended to be applied when a problem is found while using the target system. Users provide the initial data for revising the knowledge contents, and through a gamified crowdsourcing process, many users can contribute to its refinement.

As an example application, we use a conversation partner agent for people with aphasia (Kuwabara et al., 2016). One of the conversation partner agent's functions is to assist the retrieval of a word that a person with aphasia is having difficulty recalling. This assisting process uses knowledge contents to present a series of questions for the person with aphasia. From his answers, the conversation partner agent suggests the word he wants to express.

The rest of this paper is organized as follows. The next section describes related work, followed by a description of a conversation partner agent, which is the target domain of this work. Section 4 explains our proposed method that refines knowledge contents, and Section 5 describes an application of the gamification concept. We conclude this paper in Section 6.

2 RELATED WORK

Many applications of the gamification concept exist in crowdsourcing tasks. One early application was *Games with a purpose* (GWAPs), where a game ele-

[1] http://www.mturk.com/

Iwamae T., Kuwabara K. and Huang H.
Toward Gamified Knowledge Contents Refinement - Case Study of a Conversation Partner Agent.
DOI: 10.5220/0006243103020307
In *Proceedings of the 9th International Conference on Agents and Artificial Intelligence (ICAART 2017)*, pages 302-307
ISBN: 978-989-758-219-6

ment was introduced in a task (von Ahn and Dabbish, 2008). Introducing a proper game rule makes it possible to produce meaningful results from game-playing. GWAPs are applied in various fields including semantic knowledge acquisition (Siorpaes and Hepp, 2008).

The gamification idea is also used with paid crowdsourced microtasks. An incentive model for such a case was proposed (Feyisetan et al., 2015). *Quizz* is another gamified crowdsourcing system for knowledge curation (Ipeirotis and Gabrilovich, 2014). To maintain the quality of the knowledge, it uses a *calibration quiz* whose answers are known beforehand to estimate a worker's competence.

In addition, a web-based game called Common Consensus (Lieberman et al., 2007) collects and validates commonsense knowledge for understanding goals in everyday human life. Similarly, in Robot Trainer (Rodosthenous and Michael, 2016), not only factual knowledge is collected but also knowledge in rule form.

In this paper, we target the knowledge contents that are used in the domain of word retrieval assistance for people with aphasia and focus on the refinement of the pieces of collected knowledge contents. In the refinement process, we handle not only conventional answers but also obscure ones and let workers input their own level of confidence to maintain the quality of the final results.

3 CONVERSATION PARTNER AGENT

Conversation partner agents are intended to be used as a kind of interpreter for people with aphasia. One of their main functions is to support the word retrieval process for such sufferers. That is, a person with aphasia often has a problem recalling the proper word even though he knows what he wants to say. In such a situation, a human supporter (*conversation partner*) acts as an interpreter to elicit the word. For example, when a person with aphasia wants a particular type of food, the human conversation partner might ask, *Is it a fruit?* or *Is it a snack?*. Depending on the answer, the human conversation partner narrows down to identify the word the person with aphasia is thinking about.

The conversation partner agent is designed to present an appropriate question for the person with aphasia, and based on the answer, candidates are narrowed down (Fig. 1). When a word with a certain possibility is identified, the agent presents it. The knowledge contents are utilized that contain possible words to be guessed and questions to ask. Note that the conversation partner agent is not meant to completely replace human conversation partners; it can also give assistance by providing proper questions to ask and reduce his burden.

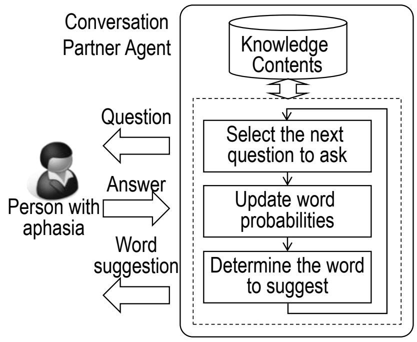

Figure 1: Conversation partner agent overview.

This word retrieval assistance process resembles a popular game called Akinator[2], which guesses the identity of a character being thought of through a series of questions and answers. It also allows a player to input the correct answer when the system fails to find an answer. This game, however, focuses on characters that are either fictional or real, whereas conversation partner agents must handle a variety of topics. In addition, in the Akinator game, some questions such as *Does the name begin with 'A'?* assume that the player already knows the word (name) of the answer (character). This kind of question is not appropriate for word retrieval assistance.

3.1 Knowledge Contents

The knowledge contents for the word retrieval assistant consists of words, questions, and the answers to the given question for the word the person with aphasia is thinking of. Let w ($\in D$) denote a target word of the word retrieval assistant, where D denotes a set of possible words. Let q_i ($1 \leq i \leq n$) denote a question to ask to identify the word. We assume n multiple choice type questions. Let N_i denote the number of answer choices for question q_i, and let $c_{ik}(1 \leq k \leq N_j)$ denote an answer choice for question q_i.

We also assume that multiple answers are possible for word w. As we will explain below, we exploit the concept of information gain to select the next question. The knowledge contents contain the probability that answer choice c_{ik} is selected when question q_i is asked and w is the correct answer. This probability is denoted by $p(c_{ik}|w)$. If there is no possibility that c_{ik} will be chosen, $p(c_{ik}|w) = 0$.

[2]http://en.akinator.com

3.2 Interaction with Users

The word retrieval assistant's goal is to guess the word the user is thinking about. In this regard, we calculate the probability that word w is the correct answer, which is denoted by $p(w)$. Based on the answers to the questions, the values of $p(w)$ are updated. When the probability of the word exceeds a certain threshold, it is selected as the one the user has in mind.

The next question is determined based on information gain (Arima et al., 2015; Kuwabara et al., 2016). To calculate the expected information gain of asking question q_i, first we set the probabilities of words, $p(w)$. The average information entropy, H, can be defined as follows:

$$H = -\sum_{w \in D} p(w) \log p(w) \quad .$$

The probability of w when answer choice c_{ik} is selected for question q_i is represented using the Bayesian theorem:

$$p(w|c_{ik}) = \frac{p(c_{ik}|w)}{p(c_{ik})} p(w) \quad .$$

The probability that c_{ik} is selected for question q_i is calculated as follows:

$$p(c_{ik}) = \sum_{w \in D} p(c_{ik}|w) p(w) \quad .$$

The values of $p(c_{ik}|w)$ are taken from the knowledge contents. The expected entropy after c_{ik} is selected when question q_i is asked, H_{ik}, is given as:

$$H_{ik} = -\sum_{w \in D} p(w|c_{ik}) \log p(w|c_{ik}) \quad ,$$

and the expected entropy after question q_i is asked is calculated as:

$$H_i = \sum_{k=1}^{N_i} p(c_{ik}) H_{ik} \quad .$$

Thus, the information gain of question q_i, IG_i, is given as:

$$IG_i = H - H_i \quad .$$

To determine the next question, we calculate the priority of question q_i, PR_i. The priority value takes into consideration the number of answer choices for question q_i in addition to its information gain. Generally the information gain becomes higher when there are many answer choices, but it is inadvisable to ask a question with too many answer choices from the start, especially for people with aphasia who might have difficulty understanding too many answer choices. PR_i is defined as follows:

$$PR_i = \frac{1}{N_i - \beta} IG_i \quad ,$$

where β denotes the parameter that represents the effect of the number of answer choices of the question and is set to 1.5 in the prototype.

The question with the highest priority value is selected and asked. When c_{ik} is selected as its answer, $p(w)$ is updated with $p(w|c_{ik})$. This process continues until the probability of a certain word exceeds a predefined threshold. If all the probabilities of words become 0, the system reports that no answer is found. If the questions are exhausted before it narrows down the options to a single word, a list of possible answers is returned.

Table 1: Example knowledge contents.

	q_1: *Is it a fruit?*	q_2: *What color is it?*	q_3: *What does it taste like?*
strawberry	*Yes*	*red*	*sweet*
banana	*Yes*	*yellow*	*sweet*
orange	*Yes*	*orange*	*sweet,sour*
lettuce	*No*	*green*	—
carrot	*No*	*orange*	*sweet,bitter*
tomato	*No*	*red*	—

3.3 Example Scenario

As an example, consider the simple data shown in Table 1. In this example, there are three questions and six words. Each cell in the table defines an answer for the word when a question is asked. When there is only one answer in the cell, 1 is the probability that that answer will be selected. For example, an answer to question q_2 (*What color is it?*) for *strawberry* is *red*, and $p(\textit{red}|\textit{strawberry}) = 1$ for q_2. If the cell contains two possibilities, such as the answer to q_3 (*What does it taste like?*) for *orange*, they are treated as if the possibility of each answer is 0.5. If there is no definite answer or an answer is unknown, "—" is inserted in the table.

Assume that the user (person with aphasia) is thinking about *strawberry*. The system first calculates the priorities of the three questions and selects q_1: *Is it a fruit?*. In reply, answer *Yes* is given, and then q_2 is asked next: *What color is it?*. If answer *red* is given, *strawberry* will be chosen as the word the user is thinking of since only *strawberry* remains as a possibility in the example knowledge contents.

4 KNOWLEDGE CONTENTS REFINEMENT

An overview of our proposed method of knowledge contents refinement is shown in Fig. 2. When a word

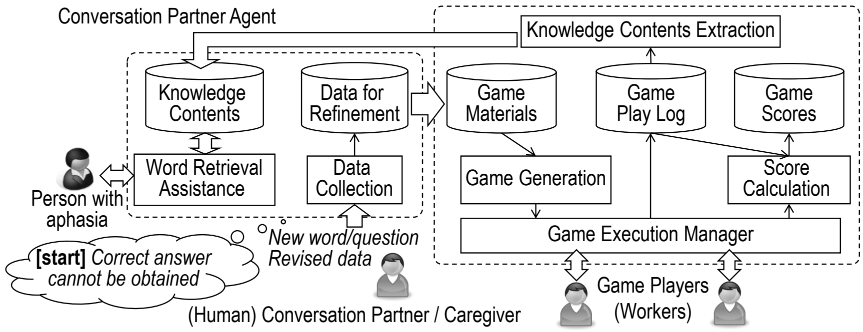

Figure 2: Knowledge contents refinement overview.

retrieval assistant system fails to produce a correct answer, it asks for data with which to revise the knowledge contents. We assume that a human caregiver such as a conversation partner who uses the system with a person with aphasia will input the correct answer. Using the input data, gamified tasks are generated and executed by other users. Based on their results, the knowledge contents are updated.

4.1 Invocation of a Refinement Process

When the word retrieval assistant system fails to find a word with high probability or to narrow a list of words to one, it is treated as a failure. The followings are the possible causes of failures: 1) The target word is not in the knowledge contents; 2) The target word is in the knowledge contents, but its data are not correct; 3) There are not enough questions to narrow the possible words to one.

In our proposed method, the data for correction are input by the user. When a word cannot be found or a word comes out that is different from what the user had in mind, the correct answer (word) is asked to be entered. When the word cannot be narrowed to one, another question must be entered.

The data input for revision can be represented as a triple $\langle w, q_i, c_{ik} \rangle$. The system contains the history of the user's responses, which can be used as part of the data when a new word is input. However, with this information, we cannot determine the value of the probability of $p(c_{ik}|w)$ required by the knowledge contents. The gamified tasks are generated from these triples to determine $p(c_{ik}|w)$ values.

4.2 Example Scenario (cont'd)

Continuing the example scenario explained above, assume that *apple* is the word being thought of by the user. As before, based on a question's priority value, q_1 is asked first: *Is it a fruit?*. The user answers with *Yes*. Then the next question (*What color is it?*) is answered with *red*, and *strawberry* is suggested by the system, which is not correct in this case. This invokes the revising mode. The correct answer *apple* is entered as shown in Fig. 3. By utilizing the history of a user's responses, additional data for *apple* are also obtained, as shown in Table 2.

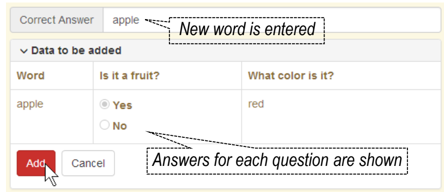

Figure 3: Entering a new word.

Table 2: Data added for a new word.

Word	Question	Answer
apple	q_1: *Is it a fruit?*	*Yes*
apple	q_2: *What color is it?*	*red*

Assume that these additional data are added to the knowledge contents. In this case, after the responses are obtained for q_1 and q_2, the possible words that remain are *strawberry* and *apple*. Question q_3 is asked next: *What does it taste like?*. However, *apple* remains one possibility since it has no answer data for question q_3, and consequently, it cannot be eliminated from the possible word list.

In such a case, a new question is needed to distinguish between *apple* and *strawberry*. Suppose that a new question is input with the answers for these two words: *How big is it?*. Then the data shown in Table 3 are obtained.

Further, assume that the word the user has in mind is a specific kind of apple, for example, a green apple. In this case, for the first question q_1 (*Is it a fruit?*), *Yes* is the answer, and the next question q_2 (*What color is it?*) is presented with the following possible answer choices: *red*, *orange*, and *yellow*. Since the answer choices do not contain the correct response, the user must choose none of the above. In this case, the correct answer, *apple*, is entered in the revising mode,

Table 3: Data of new question.

Word	Question	Answer
strawberry	q_4: *How big is it?*	*small*
apple	q_4: *How big is it?*	*medium*

and the answer to question q_2 (*What color is it?*) is entered as *green*.

At this state, question q_2 (*What color is it?*) will have two possible answers for *apple*: *red* and *green*. The possibilities of these answers are determined with the gamified tasks explained in the next section.

5 GAMIFIED REFINEMENT PROCESS

5.1 Task Description

We designed a task to determine the possibilities of an answer for a given question. The task presents a statement that can be answered by either *Yes* or *No*. A worker inputs his degree of agreement with the presented statement and also her *confidence value*. The statement of the task, denoted by $s_{w,ik}$, is generated from the data to revise the knowledge contents: $\langle w, q_i, c_{ik} \rangle$. Worker u enters the degree of agreement to statement $s_{w,ik}$, denoted by $A_u(s_{w,ik})$ and the confidence value, denoted by $B_u(s_{w,ik})$. Both values range between 0 and 1.

Let M denote a set of workers of the task. We first calculate the weighted average of the degree of agreement regarding statement $s_{w,ik}$, where the weight is determined based on the confidence value. Let $v_{w,ik}$ denote this weighted average. We determine the value of $p(c_{ik}|w)$ by considering every possible c_{ik} for question q_i. By denoting a set of possible values of c_{ik} by C_i, we have the following:

$$\begin{aligned} v_{w,ik} &= \frac{\sum_{u \in M} A_u(s_{w,ik}) B_u(s_{w,ik})}{\sum_{u \in M} B_u(s_{w,ik})} \\ p(c_{ik}|w) &= \frac{v_{w,ik}}{\sum_{c_{ik} \in C_i} v_{w,ik}} \ . \end{aligned}$$

Continuing the above example, for the triple of word *apple*, question *What color is it?*, and answer *red*, the generated statement is *The color of the apple is red*. A worker enters his degree of agreement to this statement and his confidence value (Fig. 4).

Assume that after the tasks are executed by different workers, the results shown in the three most left columns of Table 4 are obtained. Note that the number of entries is kept small for the example's clarity. From these inputs we get $v_{w,ik} = 0.816$ for *red* and $v_{w,ik} = 0.177$ for *green*. From these values, we get $p(\textit{red}|\textit{apple}) = 0.821$ and $p(\textit{green}|\textit{apple}) = 0.178$.

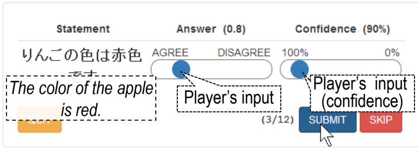

Figure 4: Screenshot of playing the game.

5.2 Scoring Rules

Next we introduce the concept of gamification into the above tasks. Gamification's main objective is to motive workers (or game players) to execute tasks.

We present a set of tasks to a worker under a time limit. For example, one game round might consist of 12 tasks that must be executed in 60 seconds. By setting a time limit, a worker is expected to make more intuitive answers. Points are basically given based on the number of executed tasks. However, if the points given are only determined by the number of executed tasks, workers are not motivated to input appropriate answers.

To address this problem, we also give bonus points to workers as an incentive to deliver more plausible answers. Bonus points, which are calculated after the value of $p(c_{ik}|w)$ is determined, increase when degree of agreement $A_u(s_{w,ik})$ is closer to $p(c_{ik}|w)$. When $A_u(s_{w,ik})$ and $p(c_{ik}|w)$ are the same, the bonus points are the highest; a worker who submits the most different values from $p(c_{ik}|w)$ receives the fewest bonus points. In addition, we consider confidence value $B_u(s_{w,ik})$ input by a worker.

More specifically, we calculate the number of bonus points given to a worker, denoted by R_u, as follows. Let err_u denote the absolute difference between $A_u(s_{w,ik})$ and $p(c_{ik}|w)$. If err_u is smaller than half of $\max_{u \in M} err_u$, the bonus points are increased. Otherwise, they are decreased. The amount of increase or decrease is determined by confidence value $B_u(s_{w,ik})$:

$$\begin{aligned} err_u &= |A_u(s_{w,ik}) - p(c_{ik}|w)| \\ R_u &= BP(1 - K(err_u - \frac{\max_{u \in M} err_u}{2}))B_u(s_{w,ik}) \ , \end{aligned}$$

where K denotes a parameter that represents the effect of the confidence value to the bonus points and BP denotes the ratio of the internal value to them.

Continuing the earlier example, a worker's bonus points are calculated as shown in the right most column of Table 4, where K is set to 10 and BP is set to 100. A worker who submitted a degree of agreement value closer to $p(c_{ik}|w)$ is given more bonus points, and a worker who submitted the value farthest from $p(c_{ik}|w)$ receives fewer bonus points.

The worker with higher confidence values receives more bonus points even if he enters the same

Table 4: Example of worker inputs and bonus points.

Worker input			Bonus points	
Worker	Degree of agreement	Confidence value	Difference with $p(c_{ik}\|w)$	Bonus points R_u
The color *of the* apple *is* red.				
u_1	0.9	0.9	0.084	77
u_2	0.8	0.3	-0.016	113
u_3	0.8	0.8	-0.016	134
u_4	0.7	0.5	-0.116	71
The color *of the* apple *is* green.				
u_1	0.2	0.4	0.022	116
u_2	0.3	0.3	0.122	82
u_3	0.1	0.6	-0.078	90

degree of agreement values (see u_2 and u_3 in Table 4). Since the confidence values are reflected in determining both $p(c_{ik}|w)$ and bonus points R_u, if we assume that a worker wants to gain more points, he will be more motivated to input more plausible answers.

6 CONCLUSIONS AND FUTURE WORK

This paper presented a method of refining knowledge contents to be used with a conversation partner agent for people with aphasia. Our proposed method deals with such problems of knowledge contents as missing words or questions. The information that revise the knowledge contents is requested to be entered by a user, and then the input data are refined by applying the concept of gamified crowdsourcing.

Currently we are implementing a prototype as a web application with conventional gamification elements such as a leader board or a badge system. We plan to conduct evaluation experiments to show the effectiveness of our proposed approach, especially where a gamified approach can effectively provide better incentives to workers.

In this paper, the conversation partner agent is used as the target of our case study. The word retrieval assistance process can be viewed as guessing an item a user is consciously or unconsciously thinking of. It can be viewed as recommending an item a user wants through a series of questions and answers. The proposed knowledge contents refining method can be used for such kinds of applications. We also plan to apply the proposed method to other application domains.

ACKNOWLEDGEMENTS

This work was partially supported by JSPS KAKENHI Grant Number 15K00324.

REFERENCES

Arima, S., Kuroiwa, S., Horiuchi, Y., and Furukawa, D. (2015). Question-asking strategy for people with aphasia to remember food names. *Journal on Technology & Persons with Disabilities*, 3:10–19.

Deterding, S., Dixon, D., Khaled, R., and Nacke, L. (2011). From game design elements to gamefulness: Defining ”gamification”. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, pages 9–15. ACM.

Feyisetan, O., Simperl, E., Van Kleek, M., and Shadbolt, N. (2015). Improving paid microtasks through gamification and adaptive furtherance incentives. In *Proceedings of the 24th International Conference on World Wide Web*, WWW ’15, pages 333–343. International World Wide Web Conferences Steering Committee.

Ipeirotis, P. G. and Gabrilovich, E. (2014). Quizz: Targeted crowdsourcing with a billion (potential) users. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW ’14, pages 143–154. ACM.

Kuwabara, K., Iwamae, T., Wada, Y., Huang, H.-H., and Takenaka, K. (2016). Toward a conversation partner agent for people with aphasia: Assisting word retrieval. In *Intelligent Decision Technologies 2016: Proceedings of the 8th KES International Conference on Intelligent Decision Technologies (KES-IDT 2016) – Part I*, pages 203–213. Springer International Publishing.

Lieberman, H., Smith, D., and Teeters, A. (2007). Common consensus: a web-based game for collecting commonsense goals. In *ACM Workshop on Common Sense for Intelligent Interfaces*.

Morschheuser, B., Hamari, J., and Koivisto, J. (2016). Gamification in crowdsourcing: A review. In *49th Hawaii International Conference on System Sciences (HICSS)*, pages 4375–4384.

Rodosthenous, C. and Michael, L. (2016). A hybrid approach to commonsense knowledge acquisition. In *proceedings of the 8th European Starting AI Researcher Symposium (STAIRS 2016)*, pages 111–122. IOS Press.

Ryan, R. M. and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, 55(1):68–78.

Siorpaes, K. and Hepp, M. (2008). OntoGame: Weaving the semantic web by online games. In *The Semantic Web: Research and Applications*, volume 5021 of *Lecture Notes in Computer Science*, pages 751–766. Springer Berlin Heidelberg.

von Ahn, L. and Dabbish, L. (2008). Designing games with a purpose. *Commun. ACM*, 51(8):58–67.