Received 10 April 2015
Accepted 13 April 2015

Automatic annotation method on learners' opinions in case method discussion

Masaki Samejima, Daichi Hisakane and Norihisa Komoda
Graduate School of Information Science and Technology, Osaka University, Osaka, Japan

Abstract

Purpose – The purpose of this paper is to annotate an attribute of a problem, a solution or no annotation on learners' opinions automatically for supporting the learners' discussion without a facilitator. The case method aims at discussing problems and solutions in a target case. However, the learners miss discussing some of problems and solutions.

Design/methodology/approach – Because opinions about problems and solutions on the same case are similar to each other, the proposed method uses opinions that are correctly annotated in past discussions for annotating an appropriate attribute on each opinion in discussions of the same case. The annotation on each opinion is identified by Support Vector Machine learned with opinions and annotations in the past discussion.

Findings – Compared to a simple method that uses decision tree classification, this proposed method improves the recall rate and the precision rate of annotating the attribute by over 10 per cent. The proposed method is effective for automatic annotation.

Originality/value – Because the recall rate and the precision rate of annotating an attribute of a problem are over 80 per cent, it is possible to make learners aware of problems that they should discuss. On the other hand, the recall rate and the precision rate of annotating an attribute of a solution are still low. The authors discuss the research issue to improve the rates for automatic annotation.

Keywords Distance learning, E-Learning, Automatic annotation, Opinion mining, Case method discussion, Support vector machine

Paper type Research paper

1. Introduction

The case method has been widely used for problem-solving skills' training (Hammond, 1980). In a process of the case method, learners discuss an actual case with a facilitator, find every possible problem in the case and propose sufficient solutions for the problems (Brooke, 2006). When the learners miss some problems or solutions, the facilitator gives advice for leading the learners to the problems or solutions. Because facilitators are lacking for learners, the learners can not always learn through the case method with a facilitator.

An automatic facilitation system is required for effective discussions in the case method without facilitators. We have already proposed an automatic facilitation system for the case method discussion (Hisakane and Samejima, 2014). The automatic facilitation system captures learners' opinions by speech recognition with a microphone

This work was partially supported by KAKENHI: JSPS (25730205).

Interactive Technology and Smart Education
Vol. 12 No. 2, 2015
pp. 90-99

1741-5659
DOI 10.1108/ITSE-04-2015-0003

and gives facilitation based on the opinions. We are developing two functions for the facilitation. One function is to generate a diagram of opinions that are denoted by nodes and relations denoted by links. The learners can find insufficient discussions from the opinion where the node of the opinion has few linked nodes, and develop further discussions from the opinion. The other function is to emphasize opinions on problems and solutions in the diagram. Even though the learners miss proposing solutions for a problem, the function to emphasize the problems and the solutions can promote reasonable discussions to solutions.

In this paper, we address emphasizing opinions on problems and solutions for the automatic facilitation. To emphasize the nodes, the automatic facilitation system colors the nodes based on an annotation of an attribute such as a problem or a solution on each opinion. In case of learners' manual annotation, learners sometimes miss annotating opinions because it is hard for the learners to analyze opinions objectively. The purpose of our research is an automatic annotation of the attribute on the opinion during the discussion.

2. Facilitation system for case method discussion

2.1 Target facilitation system

To support case method discussions without facilitators, we have developed a facilitation system that gives instructions of facilitation to learners automatically. Here we introduce the facilitation system that has functions of diagramming discussions and annotating opinions. Figure 1 shows the target facilitation system for the case method discussion, and Figure 2 shows an example of diagrams as outputs from the system.

When the learners discuss the target case by reading a description of the case, their speech of opinions is inputted to the facilitation system through the microphone. The speeches are converted to texts by a speech recognition technology, and texts of the opinions are stored in an opinion database. A facilitator gives advice based on experience in past discussions, a description of the target case and typical examples of problems and solutions. So the facilitator inputs these data to the system in advance. The facilitation system makes a diagram of the discussion with annotations of a problem or a solution as a system output shown in Figure 2. The details of diagramming a discussion and annotating opinions are as follows:

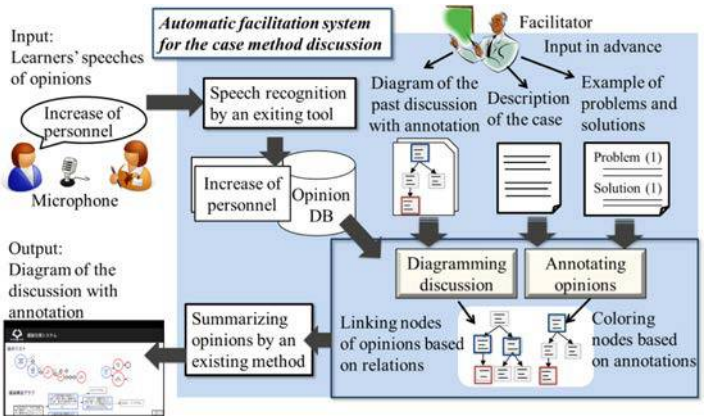

Figure 1.
Target facilitation system for the case method discussion

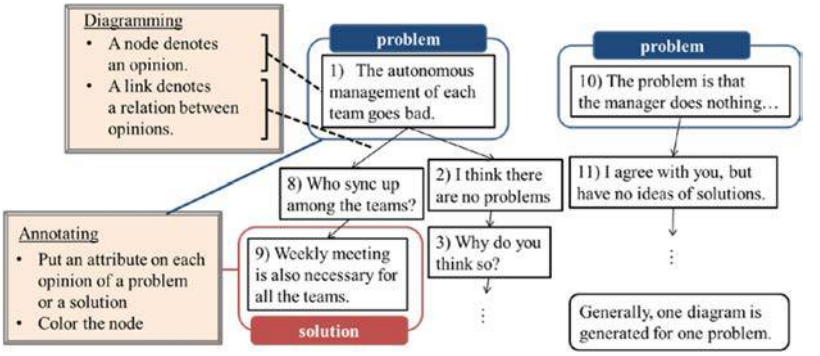

Figure 2.
An example of diagrams for a discussion

- *Diagramming a discussion*: The diagram of the discussion consists of opinions denoted by nodes and relations denoted by links. Facilitators focus on the relations, such as breakdown, complement and objection (Hisakane and Samejima, 2014; Gordon *et al.*, 2007). When a new opinion is added to the facilitation system, the system identifies such relations with existing opinions based on data that facilitators input in advance.
- *Annotating opinions*: To emphasize opinions of problems and solutions by coloring nodes of the opinions, the system annotates one of attributes: a problem, a solution and the other. After annotating an attribute on an opinion, the node of the opinion is colored with a color for the annotation.

2.2 Research purpose

In this paper, we address annotating opinions for the automatic facilitation. Currently, there are many tools for the support of manual annotation (Cunningham *et al.*, 2002). The facilitator can annotate attributes on opinions after discussions and can input annotated opinions to the system. However, it is hard for learners to annotate the attributes on the opinions during the discussion due to the following reasons:

- because learners do not have enough knowledge of the case, the learners' annotations are not always appropriate for facilitation; and
- not all learners agree upon annotations that learners subjectively put on opinions. This needs another discussion on the annotations, which interrupts the learners' discussion on the case.

To realize the annotation without depending on facilitators and learners, we aim at automatic annotation of attributes on opinions during the discussion. Figure 3 outlines the automatic annotation problem. This problem can be regarded as a three-class classification problem for a new opinion from a learner; the opinion is classified into a problem, a solution or the other. Also, it is possible to use diagrams of the past discussions with annotations, a description of the case and examples of problems and solutions:

- *The diagram of the past discussion with annotations*: As shown in Figure 2, the diagrams have opinions with annotations and links in the past discussions by

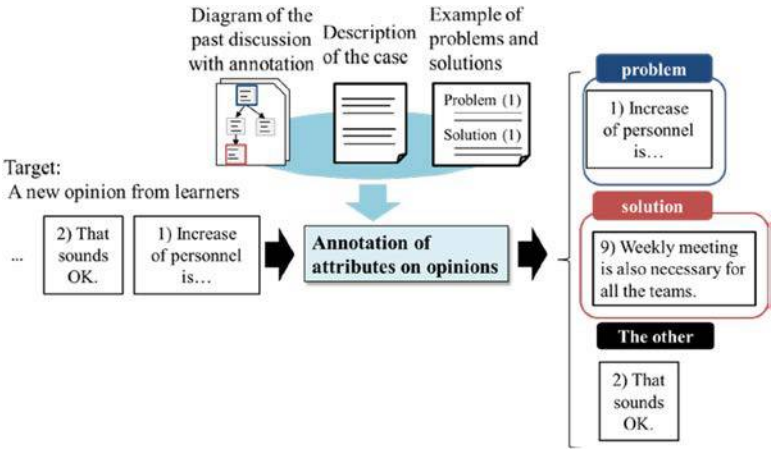

Figure 3.
An example of diagrams for a discussion

other learners. The diagrams and annotations are correctly determined by facilitators.

- *A description of the case*: A description that the learners use for the discussion consists of several pages of texts.
- *Example of problems and solutions*: Facilitators preliminarily read the description and find problems and solutions in the case. By using the problems and the solutions as examples, the facilitators can lead the discussion.

2.3 Research issue

A naïve method to solve the three-class classification problem for the automatic annotation is to use decision tree classification. Figure 4 shows the method of the automatic annotation based on the decision tree.

The decision tree classifier is constructed with opinions that have already been annotated in the past discussion. In training the classifier, the opinions are converted to word vectors that indicate whether each word is included in the opinion. The method generates feature vectors based on a bag-of-words model that is generally used for text classification. A feature vector $F_j = \{f_{i,j}\}$ $(i \in 1)$ of opinion j shows appearance of word i of all words 1 as follows:

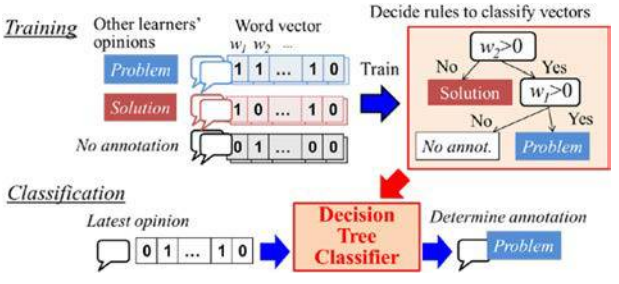

Figure 4.
Automatic annotation by the decision tree

$$f_{i,j} = \begin{cases} 1 \text{ (word } i \text{ is in opinion } j) \\ 0 \text{ (word } i \text{ is not in opinion } j) \end{cases} \quad (1)$$

Rules in the decision tree are learned to classify the opinions based on the annotations well. In case of an example in Figure 4, the rule to be applied first is to annotate "solution" if word w_1 is not in the opinion. In classifying a latest opinion with the learned decision tree classifier, the latest opinion is also converted to a word vector and the word vector is judged by the rules in the decision tree. Through applying rules, the annotation on the opinion is determined.

Because learners who join a case method discussion have different backgrounds, they give various opinions in the discussions. If there are numerous variations of words, the size of the feature vectors is large but sparse, and most components of feature values are 0. It is generally known that the classifier trained by sparse feature vectors cannot classify well. To change the sparse feature vectors to dense ones, it is necessary to choose words to be considered in the feature vectors. Therefore, it is necessary to use different word vectors for determining different annotations.

3. Automatic annotation method by Support Vector Machine

3.1 Outline of the automatic annotation method

As we described in Section 2.2, the annotation problem is a three-class classification problem. By combining binary classifiers of Support Vector Machines (SVM) based on word vectors with different features, we design the automatic annotation method to solve the classification problem. Figure 5 shows the outline of the automatic annotation method by SVMs.

Before annotating on a new opinion, the proposed method obtains supervised data from other learners' opinions with correct annotations for training SVMs. Because the proposed method aims at annotating two kinds of attributes of a problem and a solution, the other learners' opinions are classified into the following supervised data as shown at the top of Figure 5:

(1) *Supervised data for annotating "problem"*: Past opinions with annotation "problem" are used as positive instances, and past opinions with no annotations are used as negative instances.

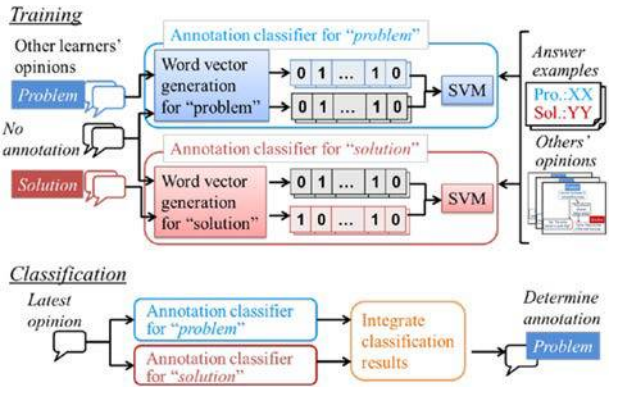

Figure 5.
Outline of the automatic annotation method

(2) *Supervised data for annotating "solution"*: Past opinions with annotation "solution" are used as positive instances, and past opinions with no annotations are used as negative instances.

Then the numbers of positive instances and negative instances are different, which causes the biased classification. To avoid the biased classification, the proposed method equalizes the numbers of positive instances and negative instance by removing excess negative instances randomly as known as Random Under Sampling (He and Garcia, 2009).

First, to process opinions by SVM, it is necessary to generate feature vectors from the opinions. Because the opinions are based on a description of the case, the opinions are expected to be similar to the examples of problems and solutions. Therefore, the feature vectors are generated from the description, the past opinions and the examples. Different feature vectors are generated for each SVM of annotating "problem" and "solution". Next, the proposed method trains SVM with the generated feature vectors before classifying a new opinion by SVM. Each SVM of annotating "problem" and "solution" outputs a classification result. Because each opinion must have one attribute in an annotation, it is necessary to choose one attribute when SVM judges that different attributes are annotated on an opinion. Therefore, the proposed method determines an appropriate attribute on an opinion by integrating classification results from SVM.

3.2 Annotation classifier for problems

Figure 6 shows the annotation classifier for opinions that are related to problems. If we consider all the words, the word vectors are often sparse and include overlapped words, which cause the wrong annotation. Before training the classifier, the proposed method focuses on verbs and nouns that represent the content of the opinion. Second, because the numbers of the verbs and the nouns are still many, the proposed method additionally chooses keywords for annotating "problem" from the verbs and the nouns. The opinions are often similar to examples of problems that are input by facilitators. Therefore, we regard words in the examples of problems as keywords, and replace the words *I* as the keywords in generating word vectors.

By generating word vectors with filtering words other than keywords, SVM classifier is trained. And, the latest opinion is also converted to a word vector by the same way, and classified by SVM.

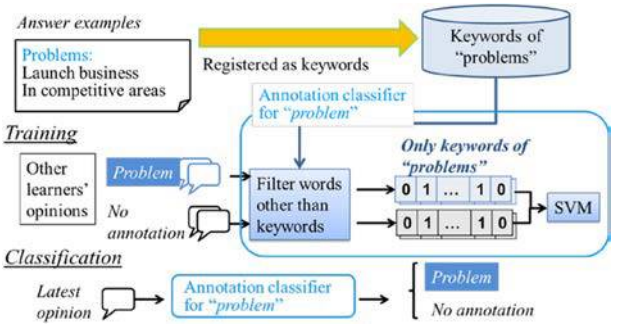

Figure 6.
Annotation classifier for opinions that are related to problems

3.3 Annotation classifier for solutions

Figure 7 shows the annotation classifier for opinions that are related to solutions. As well as the annotation classifier for opinions that are related to the problems, the annotation classifier for opinions that are related to solutions uses words in answer examples of solutions as keywords. The words I is replaced as the keywords in generating word vectors. By generating word vectors with filtering words other than keywords, SVM classifier is trained. And, the latest opinion is also converted to a word vector by the same way, and classified by SVM.

3.4 Integrating classification results

After classifying a new opinion by two SVM with the instances, the proposed method can obtain two classification results. Because each opinion has one attribute of an annotation, it is necessary to integrate the results if both results indicate that the opinion has both attributes of "problem" and "solution". Figure 8 shows integration patterns for outputs from SVM.

As discussed in Section 3.2, opinions of problems tend to include words in a description. Due to the tendency, to identify opinions of problems is easier than to identify opinions of solutions. So, the proposed method considers that the classification result by SVM of annotating "problem" is reliable. An order of reliabilities of annotations by SVM is as follows. The most reliable annotation of two classification results is put on the opinion. For example, if the proposed method obtains annotations "problem" and "solution", the annotation "problem" is put on a new opinion based on the order of the reliabilities: problem > solution > no annotation.

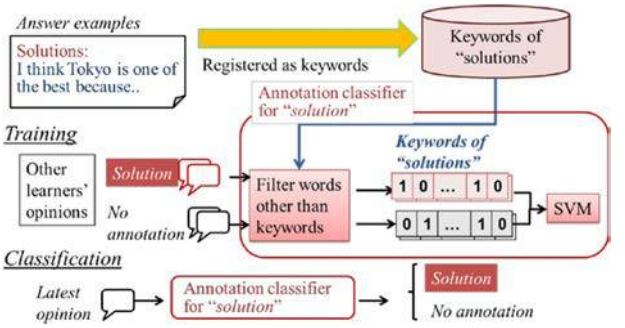

Figure 7.
Annotation classifier for opinions that are related to solutions

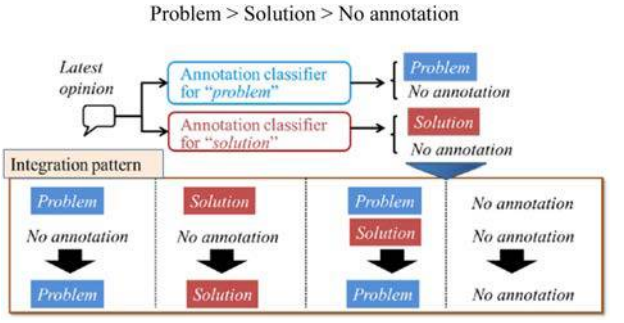

Figure 8.
Integration patterns of outputs from SVM

4. Evaluation experiment

The target cases in this experiment are a case about a project management of a software development and a case of marketing for a liquor brand. Both cases are used in an instruction course by a company and a university. We have descriptions of the cases and examples of problems and solutions that are provided by the company and the university. We collected opinions from discussions by six groups of three students that major in computer science at a university. Groups A, B and C discuss the case of the project management and Groups D, E and F discuss the case of marketing. Each group has 30 minutes for the discussion. To make supervised data, we put a correct annotation on each collected opinion.

Table I shows the numbers of all opinions, opinions with an annotation "problem" and opinions with an annotation "solution". For each case, we apply the proposed method to opinions of each group with training SVM by opinions of the other two groups. In addition, we apply the decision tree as a baseline method. Figure 9 shows the following recall rate and precision rate of annotating "problem", and Figure 10 shows the recall rate and precision rate of annotating "solution":

$$\text{Recall rate} = \frac{\text{The number of correctly annotated opinions}}{\text{The number of opinions with a true annotation}}$$

$$\text{Precision rate} = \frac{\text{The number of correctly annotated opinions}}{\text{The number of annotated opinions}}$$

Types of opinions	Case of project management			Case of marketing		
	Group A	Group B	Group C	Group D	Group E	Group F
The number of all opinions	65	50	50	54	59	59
The number of opinions with annotation "problem"	8	7	6	6	5	4
The number of opinions with annotation "solution"	6	7	9	9	10	6

Table I. The numbers of all opinions and opinions with annotations in each group

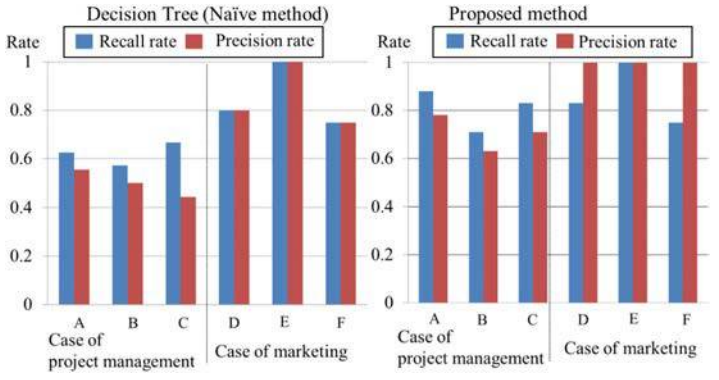

Figure 9. Recall rate and precision rate of annotating "problem"

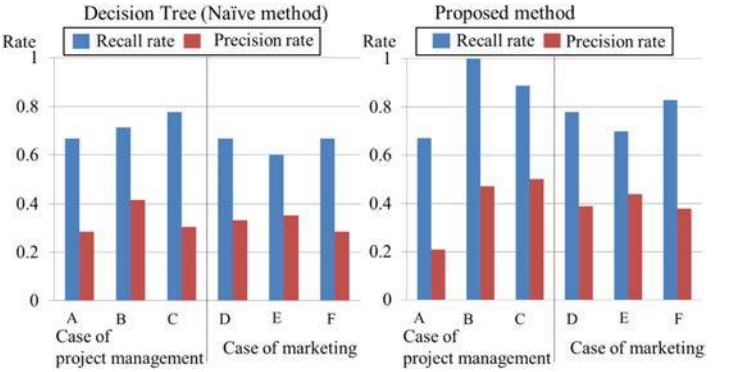

Figure 10.
Recall rate and precision rate of annotating "solution"

First, we discuss the recall rate and the precision rate of annotating "problem" in Figure 9. The proposed method annotates "problem" on the opinions at the recall rate of 83 per cent and the precision rate of 81 per cent on average. The proposed method can annotate over 10 per cent more accurately than the naïve method because the learners use words in the description as described in Section 3.2. In addition, opinions of problems tend to include more words than the other opinions. Therefore, feature vectors represent the content of the opinion well. When we listened to the learners regarding the reasons why opinions of problems often include many words, the learners answered that they prepared for starting the discussion with their finding problems. By comparing results of both cases, opinions for the case of marketing can be annotated better than opinions for the case of project management. According to the learners, the case of marketing is easy to discuss, and the learners are not familiar with various terms of marketing because they are not experts of marketing.

Next we discuss the recall rate and the precision rate of annotating "solution" in Figure 10. The proposed method annotates "solution" on the opinions at the recall rate of 81 per cent and the precision rate of 39 per cent on average. The precision rate of annotating "solution" is worse than one of annotating "problem". As the discussion progresses, the learners give opinions without referring to the description of the case to avoid redundant opinions. An opinion without referring to the description is too short to identify the content. Because opinions of solutions that the learners discuss later are also short, the proposed method can not identify whether such a short opinion should be annotated "solution", which makes the precision rate worse. Based on the result, to improve the automatic annotation, it is necessary to consider the short opinions where the words are not referred to in later opinions.

5. Conclusion

We proposed an automatic annotation on learners' opinions to realize facilitation of the case method discussions without a facilitator. The facilitation system stores learners' opinions in past discussions with correct annotations by facilitators. So the proposed method applies SVM with the past opinions for the automatic annotation. In addition, the proposed method considers that the learners' opinions tend to include words in a description of the target case and are similar to not only the past ones but also example opinions given by facilitators. Therefore, based on the description of the case, the past

opinions with annotations and the example opinions, the proposed method generates feature vectors for processing by SVM. Finally, the proposed method integrates classification results by SVM based on reliabilities of annotating. The experimental result showed that an attribute "problem" is annotated at a recall rate of 83 per cent and at a precision rate of 81 per cent but an attribute "solution" is annotated at a recall rate of 81 per cent and at a precision rate of 39 per cent. The reason why the attribute "solution" is not annotated well is that opinions of solutions do not include all representative words to avoid redundancy of the opinions. The future issue to improve annotating "solution" is to find the words that do not appear in the opinions when in applying SVM.

References

Brooke, S.L. (2006), "Using the case method to teach online classes: promoting socratic dialogue and critical thinking skills", *International Journal of Teaching and Learning in Higher Education*, Vol. 18 No. 2, pp. 142-149.

Cunningham, H., Maynard, D., Bontcheva, K. and Tablan, V. (2002), "Gate: an architecture for development of robust HLT applications", *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, Stroudsburg, pp. 168-175.

Gordona, T.F., Prakkenb, H. and Waltonc, D. (2007), "The Carneades model of argument and burden of proof", *Artificial Intelligence*, Vol. 171 Nos 10/15, pp. 875-896.

Hammond, J.S. (1980), *Learning by the Case Method*, Harvard Business School, Boston.

He, H. and Garcia, E. (2009), "Learning from Imbalanced Data", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 21 No. 9, pp. 1263-1284.

Hisakane, D. and Samejima, M. (2014), "A visualization system of discussion structure in case method learning", *Proceedings of 6th International Conference on Knowledge Management & Information Sharing*, Rome, pp. 126-132.

Corresponding author

Masaki Samejima can be contacted at: samejima@ist.osaka-u.ac.jp