Evaluation of Jurisprudence Arguments Based on Annotations of Logical Structures and Speech Acts

Shumpei Kubosawa, Shogo Okada and Katsumi Nitta

Interdisciplinary Graduate School of Science and Engineering
Tokyo Institute of Technology
4259 Nagatsuta-cho, Midori-ku, Yokohama, Japan 226–8502
Email: kubosawa@ntt.dis.titech.ac.jp, {okada,nitta}@dis.titech.ac.jp

***Abstract*—In jurisprudence education, discussion trainings by mock arbitrations, mock mediations and so on are conducted. On those trainings, students discuss an issue and experienced supervisors evaluate the discussion by observing them and make comments with respect to argumentation and other contextual abilities. However, evaluation consume much time of experts, and objective criteria for evaluation are not established yet. Besides, existing tools for supporting them are insufficient to analyze a total discussion log and compare them. Furthermore, there are hardly any objective method to analyze contextual abilities. In order to support evaluation of actual discussion records, we propose a method and supporting tools for discussion evaluation. The presented method is based on annotations of logical structure of total arguments and speech acts attached to each utterance on discussion logs. We compared evaluation result of our method to expert's manual evaluation and showed that evaluation results by our method was similar to experts' result.**

Keywords–argument analysis; jurisprudence education; argumentation framework; Toulmin model.

I. INTRODUCTION

In education for students majoring in law, argumentation training through mock trials or arbitration is introduced with the goal of enhancing their operational capability of law and arbitration skills. In this type of training, the participants proceed with a conflict-case discussion as the parties concerned. The teacher evaluate the discussion and gives them counsel with respect to argumentation and discussion skills after the discussion. For the purpose of giving more opportunities to make comments to as many students as possible, recently, online discussion training programs have been introduced by utilizing electronic bulletin boards.

In addition to this kind of training, argumentation competitions including Intercollegiate Negotiation Competition (INC) [1] are held in order to bring about further enhancement of argumentation skills. They develop discussions in two forms, a mock arbitration and a mock negotiation, conducted in about 20 different locations. The evaluation items for each discussion are described on a score sheet. At each location, legal experts score and compare these evaluation items in order to evaluate and rank each team.

This type of arbitration training requires that a limited number of teachers need to observe and evaluate a number of discussions. Not only observe each individual discussion in detail, but they also need to compare multiple discussions in order to evaluate the relative merits. Therefore, teachers are placed under significant burden and pressure, while support based on computers becomes very necessary.

However, the focus of previous studies on discussion record analysis (described in Section 2) remains to have only partial support for analysis, such as the visualization of argument structures. These methods are insufficient for determining the discussion quality and characteristics. Furthermore, objective criteria for evaluation of argumentation and contextual abilities are not established yet. Contextual abilities cover attitudes of mediators, smoothness of discussions, deadlock response, and other skills. For instance, one of the important contextual skills for mediators is reframing, while workshops targeting settlement committees have been conducted. Reframing is a type of utterance which can potentially resolve deadlocks. In spite of the importance, there are hardly any objective analysis with respect to reframing in conventional studies to our knowledge. Therefore, those works have not met the needs for discussion training support.

We introduce a novel method and a support tool for discussion analysis. In order to calculate proposing evaluation indices and detect deadlocks for supporting reframing analysis, our method utilize sequence of tags attached to each utterance in discussion record. Logical structure tags and speech act tags mainly correspond to argument structure analysis and discussion skill analysis respectively. Section 2 describes related studies regarding discussion analysis. Section 3 handles the outline of the discussion analysis support tool, while Section 4 shows the discussion analysis results obtained by using this tool and shows that the result is similar to experts' evaluations.

II. RELATED WORK

A. Plotting of discussion structures and tagging

Many researchers have studied how to analyze logical structures of arbitration by visualizing them based on diagrams. For example, Toulmin proposed the model (Toulmin model) which visualizes arguments as diagrams by dividing them into each element of claim, data, warrant, backing, qualifier, and rebuttal [2] (Figure 1). The Toulmin model is commonly known among legal experts, while it is widely accepted that sound arguments can be divided into these elements.

In addition, there have been developed many systems based on Toulmin-model-based diagrams, such as an online system that supports discussions while presenting discussion

 ISBN: 978-1-61208-461-9

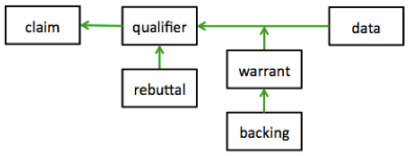

Figure 1. Toulmin model

structures on a real-time basis, and a system that analyzes logical structures of a discussion [3][4][5][6].

Furthermore, in order to extract logical structures from discussion records, tagging the role of each individual word, section, or paragraph are useful. The concept of Global Document Annotation (GDA) proposed by Hashida et al. provides a universal tagging method [7]. The tagging method proposed by Tanaka et al. categorizes the roles of each utterance into claims, agreements, denials, complements, reasons, withdrawals, questions, answers, proposals, requests for reasons, and transfer of the right. Tagged utterances are then used for automatic extraction of the factors of argument and the search for similar discussion scenes [8].

B. Argumentation frameworks

An argumentation framework is a framework that is provided in order to select valid arguments based on attack relationships between arguments. Dung defined the possibility of accepting each argument within a discussion as the principle that "The one who has the last word laughs best." Namely, he defines this possibility as a dialogical relationship based on the concept that "although argument A is attacked by argument B, if argument B attacks (countercharges) argument A, argument A is not defeated by argument B." Based on this definition, he proposed the following argumentation frameworks (AF) [9].

$$AF = (Arg, attacks) \tag{1}$$

The AF (1) is a pair consisting of a set of arguments Arg and $attacks \subseteq Arg \times Arg$, which indicates the attack relationships between arguments. In the AF, the possibility of accepting each argument is evaluated based on the attack relationships between arguments and multiple semantics are defined.

Prakken proposed the framework referred to as ASPIC+ framework which incorporated logical expressions into argumentation frameworks [10]. This framework provides a method to interpret discussion structures based on AF by transforming argument trees consisting of strict rules and defeasible rules into AF.

C. Discussion analysis based on points of dispute

In the field of support for argumentation education, Ashley and Aleven proposed discussion models based on exchanges of legal precedents which were expressed as a list of points (dimensions or factors) of dispute [11][12]. The method proposed by Ashley et al. analyzes those sets of precedents that are preliminarily available in order to extract dimensions or factors that characterize individual precedents in advance.

Expanding this method, Tanaka et al. developed an automatic method to recognize points of dispute by recognizing the factors included in individual comments made during a discussion based on pre-prepared lists of factors [8].

D. Reframing

Reframing is one of the discussion skills required of the committee of settlement in an arbitration case [13]. Reframing is a type of utterances that can change the frame of thinking of other parties. This is done in order to change the thought of them from negative things to positive things, or in order to persuade them [14]. Reframing is an effective technique for arbitrations when the discussion stagnates and seems to go nowhere due to emotional reactions expressed by both parties or other reasons. However, it is difficult to decide what to say and when to say for reframing since they depend on the context of discussion. Therefore, automated detection of reframing portins in a discussion record is also difficult, while reframing utterances would tend to exist after deadlocks by its nature.

III. Discussion Evaluation Methods and an Evaluation Support Tool, Corte

A. Outline of Corte

The purpose of our discussion analysis support is to support discussion education in the law schools of universities as described in Section 1. Those jurisprudence discussions have three characteristics: (1) facts on the disputed case are preliminarily given, (2) factors correspond to each claim are predicted in advance, and (3) there are multiple discussion records regarding the same case. In order to support education for this kind of discussions, we developed a discussion analysis tool called Corte with the following functions described below.

- Analysis of arguments in discussion records in terms of mathematical argumentation theories in order to determine which party succeeded in refuting the discussion, which argument becomes the key to conclude the discussion, and which argument needs to be discussed further.
- Calculate indices related to discussion skills using tags attached to each utterance in discussion record in order to evaluate:
  - "the composition of every argument" (whether the facts that should be claimed and legal theories were properly and clearly claimed or not), which is one of evaluation items in INC,
  - arbitration directions of mediators (whether they are facilitative or evaluative), which are considered to be the arbitration skills of mediators, and
  - the smoothness of arbitration (whether the discussion is rehashed or not), as another evaluation index for arbitration skills.
- Deadlock detection for supporting reframing analysis.

Corte inputs the discussion records, factor lists, and deadlock patterns.

Discussion records are text data where the utterance of participants are aligned in order of time. Factor lists are sets of factors that indicate the facts, laws, and claims that are included in the case to be analyzed. These lists are provided by the case maker or by discussion analyzers during the tagging

 ISBN: 978-1-61208-461-9

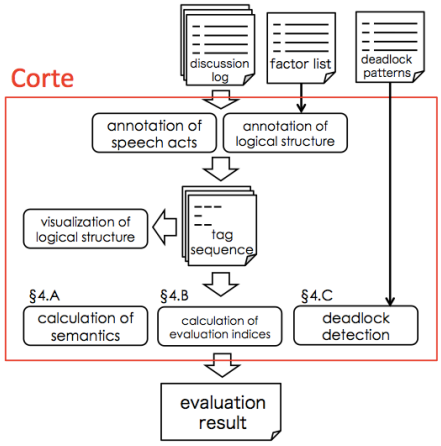

Figure 2. Workflow of discussion evaluation supported by Corte

process as described later. Deadlock patterns are chains of utterance types that indicate deadlocks to be extracted.

As shown by the flow in Figure 2, users of Corte annotate discussion records with speech act tags and argument structure tags. Based on this two-tag system, Corte evaluates discussions from various angles.

As speech act tags, in order to calculate evaluation indexes that will be described later and to detect candidates of deadlock portions, we used the following 11 categories: Claim, argument, agreement, denial, complement, close-ended-question (CEQ), open-ended-question (OEQ), reply, request, proposal, and "other".

Argument structure tags are argument elements where arguments included in discussions are dissolved based on the Toulmin model. However, the attack relationships between arguments are separately defined; the following three elements are sufficient when individual arguments are dissolved: Claim, warrant, and data. Each element corresponds to a factor on the factor list. Furthermore there are two types of relationships between elements, namely, support links and attack links. An argument consists of one or more elements and support links (a claim is supported by warrants and data). An element would be connected with other element via attack link since each element itself can be an argument. In the rest of this paper, we call this argument structure model as "argument model."

B. Annotation

Figure 3 shows the annotation screen of Corte. On the list of utterances shown in Figure 3, utterances are displayed using colors corresponding to each speaker. Factor lists show a list of argument elements based on the Toulmin model, while Section 3-A describes argument elements. Selecting each utterance shows its details on the top of the screen. In Figure 2 of "annotated factor", each user reads each utterance and adds elements on the diagram screen (in the center of the screen of Figure 3) while dividing utterances into argument structures. Doing this adds elements on the factor list shown on the right side of the screen. Annotating these added elements to the relevant portions can give the utterance log an argument structure tag. Although where there exist support or attack relationships between elements, the relationships between elements are defined by adding or deleting the arcs on the graph. With this tool, users can input argument structures just by editing the graph intuitively. At the same time, it can visualize argument structures (visualization of argument structures within Figure 2). On the screen shown by Figure 3, the elements included in the selected utterance (top of the screen) are indicated as nodes in thick color on the graph screen. In addition, between argument logs regarding a common case, sharing the argument lists and argument structure graphs can make it possible to compare discussions.

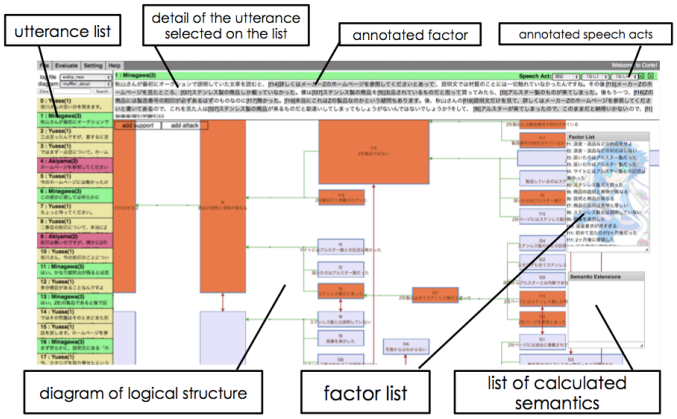

Figure 3. Screenshot of Corte

IV. Discussion Analysis by Using Corte

After annotation, this tool calculates argumentation semantics (section IV.A), calculates the evaluation indices for discussion skills (section IV.B), and detects deadlock portions (section IV.C). The section below explains the analysis methods based on the use of discussion logs of the actual mock arbitration and mock mediation, along with the results of evaluating the analysis methods. We used three mock arbitration logs and eight mock mediation logs for our experiments. Each log was evaluated by using the evaluation scales and evaluated by experts manually. We then compared the evaluation results. It should be noted that these educational-purpose discussions are rarely released to the public, while detailed evaluation conducted by experts are even less frequently publicized. Therefore, the number of logs to be evaluated is limited. For this reason, the section below describes the tendencies of evaluation based on the evaluation scale and evaluation conducted by experts. Note that we used discussion logs written in Japanese for that experiment, however, Corte supports Unicode files therefore other languages are also supported.

A. Semantic calculation

Semantic calculation is used to determine which argument can be established and not be established in a discussion where arguments and counter-evidence are entangled complicatedly. Here, while each argument structure for argument models is stored, semantic calculation is conducted by applying the argument model to the ASPIC+ framework [Prakken11] in order to generate the set of arguments and a set of attack relationships.

The argumentation system in the ASPIC+ framework can be defined by the following tuple

 ISBN: 978-1-61208-461-9

$$AS = (\mathcal{L}, \bar{}, \mathcal{R}, \leq) \quad (2)$$

where $\mathcal{L}$ indicates logical language, which is a set of inference elements, $\bar{} \subseteq \mathcal{L} \times \mathcal{L}$ indicates counter-evidence relationships between inference elements, $\mathcal{R} \subseteq \mathcal{L} \times \mathcal{P}(\mathcal{L}) \setminus \varnothing$ indicates inference regulations consisting of the strict inference regulations and the defeasible inference regulations, and $\leq$ indicates the partial order relationships in the inference regulations. In order to convert argument models in this paper into the argument system, we introduced the following conversion regulations.

1) Elements in the argument model are set to the logical language in the corresponding argument system.
2) Considering the warrant element to be one of the data elements, the claim elements to be the conclusion part and the data element to be the premise part, these elements are added to the defeasible rules in the corresponding argument system.
3) Attack relationships are added to the counter-evidence relationships of the argument system.

In the argument model, all elements are regarded as arguments, and there are possibilities that attack relationships could be generated between these elements. Therefore, all support relationships are interpreted to be defeasible rules. The order relationships in the inference regulations are considered to be always empty, because the order relationships of support relationships in the argument model are not considered. This simplifies the model. Figure 4 shows an example of converting the argument model into the argumentation framework by means of the procedure described above.

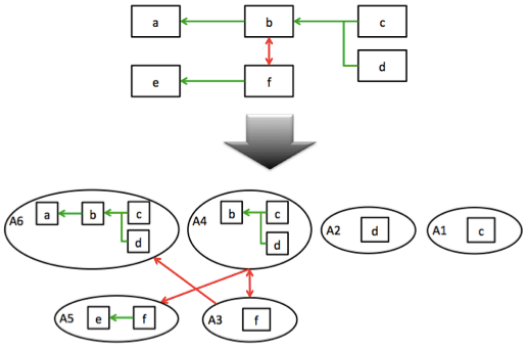

Figure 4. An example of extraction of argumentation framework from argument model

Figure 4 shows that the argument model which consists of six discussion points shown above changed into the argumentation framework which consists of six arguments shown below. After the argumentation framework has been obtained, this tool calculates each extension in the argumentation framework. In the case of Figure 4, preferred extensions are calculated to be A1, A2, A3, A5 and A1, A2, A4, A6. Each of these arguments originates in each discussion point of the argument model. Therefore, the list of discussion points corresponding to each semantic is presented on the list of extensions in the lower right of Figure 3. Figure 5 shows an example where arguments included in one of the preferred extensions as nodes emphasized in yellow. This figure overlaps argument graphs obtained from all discussion logs regarding the common case, while selecting and showing one of the preferred extensions in selected discussion logs. There are multiple preferred extensions. Therefore, this function is used to determine what argument is valid and what argument should be attacked so that the discussion can become advantageous.

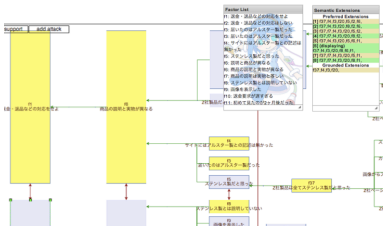

Figure 5. Nodes in the selected preferred extension is highlighted in yellow

B. Evaluation of discussion skills

Comparison and analysis of multiple discussions (match ups) based on common dispute cases can make it possible to clarify the issues for the participants in each discussion and evaluate discussions in a relative way. With that, introducing three indexes for evaluating discussion skills, we considered the effectiveness of these indexes.

1) Evaluation of the strictness (corresponds to the composition of every argument): The strictness is an index that indicates the composition of every discussion point (whether the facts that should be claimed and legal theories are properly and clearly claimed or not). Usually, the composition of issues is complicated in a negotiation competition, while multiple discussion points are included in one issue. According to each discussion point, the parties concerned extract discussion elements from the shared facts and previous knowledge, and present the arguments that attack the claims of the opponent party. By doing so, they need to increase the validity of their own claims. When discussions of both parties are engaged, arguments and counter-evidence arguments are repeatedly exchanged regarding the discussion point concerned, while the diagrams become quite dense. On the other hand, if discussions are not engaged, the claims of both parties remain superficial, while diagrams become non-dense. We obtained three discussion records (A, B, and C) on the first day (arbitration) of the INC in 2011. Figures 6 and 7 show the argument graphs that indicate the discussion points that were referred to during discussions A and C.

Figure 6. Factors referred on discussion A

These figures show diagrams of two discussions by overlapping them. The arguments appeared in discussion A are indicated in red nodes in Figure 6, while those appeared

Figure 7. Factors referred on discussion C

in discussion C are indicated in red nodes in Figure 7. Comparison of both figures clearly shows that Figure 6 has a greater number of discussion points than that of Figure 7.

In order to evaluate the strictness of this kind of discussions, we introduced the following scales.

- No. of support relationships: To what extent do the arguments justified by the warranty argument exist?
- No. of attack relationships: To what extent do the counter-evidence arguments against the opponent party's claim exist?
- Factor coverage (equals to No. of factors in the target discussion divided by No. of factors of all discussion logs): When other discussions are compared, to what extent is the argument elements presented?
- Warrant coverage (equals to No. of warranty elements that appear in the target discussion divided by No. of warranty elements that appear in all discussion logs): When other discussions are compared, to what extent are the legal regulations including contracts in operation?

TABLE I. STRICTNESS CORRESPOND TO EACH DISCUSSION

index	*A*	*B*	*C*
support relationships	55	39	31
attack relationships	28	18	18
Factor coverage	93	59	52
Warrant coverage	79	36	36

Table I shows the calculation results of the evaluation scales in the third discussion of the INC of 2011. This shows that as these values became greater, as the discussion proceeded strictly and logically. Table 1 shows the order of the strictness according to the discussion as A$\ll$B$\leq$C. This shows that discussion A was stricter than discussions B or C.According to the evaluation results of INC, team X participated in discussion A received the top prize. The other team that participated in discussion B was ranked lower than team X, but was awarded the top prize. Additionally, team Z participated in discussion C but never received the top prize. The evaluation done by legal experts shows that discussion A was highly evaluated in the item "the composition of every discussion point," while there was no significant difference between discussion B and discussion C. This result shows that the evaluation values of the strictness of discussion performed by the experts showed similar tendencies in the evaluation of "the composition of every discussion point." This suggests that strictness can be used for evaluating discussion skills.

2) Evaluation for mediator's attitude: The attitude shown by mediators during an arbitration case can be classified into facilitative or evaluative. Being facilitative is an attitude that facilitates the parties concerned to make utterances while showing respect for the utterances made by the participants concerned. Being evaluative is an attitude that evaluates the utterances made by the parties concerned while the discussions proceeds using leadership. Generally, mediators are required to be facilitative. In other words, even though mediators have more expertise than the client, they are prohibited to give advantageous information to either side. They have to protect their neutral position.

It is necessary for mediators to show facilitative attitude in an arbitration case. In a mock arbitration, however, many students play the roles of mediators. Therefore, they need to have proper instructions. Determination of mediators' attitude from arbitration records can be used for reference to teach arbitration skills to mediators.

Differences in attitudes are determined focusing on the speech act tag, proposal. Facilitative mediators behave so that the parties concerned can make positive utterances. However, evaluative mediators present their own evaluation, or sometimes they even take the lead in forming consensus. On the other hand, in the process of forming consensus, an utterance with the intention to form a proposal appears. For this reason, the attitude of mediators is evaluated by using the percentage of the number of proposal utterances made by the mediators. Namely, this percentage is expressed by No. of proposal utterances made by the mediators divided by No. of all utterances in each discussion log. Although the arbitration issue is the same, the number of utterances may vary significantly. Therefore, the proposal utterances are divided by all utterances in order to reduce this influence by normalization.

TABLE II. MEDIATOR'S ATTITUDES ON EACH DISCUSSION

index	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Proposals by mediator (1)	0	3	1	0	3	2	0	5
Total utterances (2)	76	68	91	25	70	55	67	62
(1) / (2) [%]	0	4.4	1.1	0	4.3	3.6	0	8.0
expert's evaluation	F	E	F	F	E	F	E	E

By using eight discussion records of mock arbitration cases regarding products between sellers and bidders for buying and selling automobile mufflers in an online auction, we verified the evaluation scale that indicates whether the mediators were evaluative or facilitative. Table II shows the calculation results of evaluation scale and the evaluation results of each discussion done by legal experts. According to Table II, this evaluation index value shows the same tendency of the evaluation done by experts. It also suggests the possibility of automatic determination by the mediators' attitude.

3) Evaluation for smooth operation of arbitration: One more index for measuring the skills of mediators is whether the arbitration discussion proceeds smoothly or not. In arbitration, after the parties concerned have confirmed the facts and their desires, each party proposes solutions based on the facts recognized by them, in order to reach consensus.

A certain order where consensus is formed after the facts have been confirmed is very important for smooth consensus formation. When facts are reconfirmed during the discussion

 ISBN: 978-1-61208-461-9

of consensus formation, consensus items formed during the past consensus formation processes could be invalid. As to the scale for measuring whether the discussion goes back to fact confirmation after the consensus formation discussion has started, the value normalized by the number of proposal utterances after the first proposal utterance has been made with the number of argument elements presented after the first proposal utterance has been made. Namely, as the smooth operation level of each discussion log, No. of factors referred to after the first proposal utterance has been made divided by No. of utterances made after the first proposal utterance has been made is introduced. Here, the discussion of consensus formation is interpreted as to have started by the first proposal utterance.

TABLE III. SMOOTHNESS INDICES ON EACH DISCUSSION

index	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Factors after the first prop. (1)	0	15	0	0	0	3	4	0
Utterances after the first prop. (2)	32	53	68	8	24	32	28	28
(1) / (2) [%]	0	28	0	0	0	9	14	0
expert's evaluation	A	C	B	B	A	C	C	B

Table III shows the calculation results of evaluation scales regarding the smooth operation level of mediators by using the same logs to evaluative attitude and the verification results by using the evaluation results of each discussion done by experts. The experts' evaluation show that evaluation A (excellent) was smoother than evaluation B (good) and C (poor).

According to Table III, the experts determined that discussions with evaluation indexes more than 0 are not smooth. However, distinguishing between good and excellent was not determined. However, discussions which were not extremely smooth corresponded to the evaluation done by experts.

C. Deadlock detection

Discussions could go into deadlock when both parties do not yield. In order to eliminate deadlocks, reframing utterances to change viewpoints are effective. Deadlocks have a wide variety of patterns; it is difficult to enumerate these patterns.

Corte users register deadlock patterns, and Corte detect portions that match with the deadlock patterns. By doing so, this tool offers deadlock recognition support. The following pattern is a sample pattern. A typical deadlock pattern is that a proposal made by one party is rejected by the other party without having any particular reasons. This phenomenon is repeated again and again. In other words, where the parties concerned (speakers) are A and B, the factors concerned are X and Y, and the following tuples (speaker, speech act tags, and logical structure tags) are used, the utterance systems below are repeated.

1) (A, proposal, X) OR (A, proposal+argument, X)
2) (B, denial, $\neg$X) OR (B, denial+argument, Y$\Rightarrow \neg$X)

TABLE IV. PERFORMANCE OF DEADLOCK DETECTION

Index	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*Total*
Detected by Corte	3	3	1	1	2	2	3	0	15
Detected manually	1	0	2	0	1	1	3	2	10
Precision	1/3	0/3	0/1	0/1	1/2	1/2	1/3	0/0	4/15
Recall	1/1	0/0	0/2	0/0	1/1	1/1	1/3	0/2	4/10

Table IV shows the results of deadlock detection from discussion logs of mock arbitrations by applying deadlock patterns. While this tool successfully detect a part of deadlocks, its accuracy is not enough high. Therefore, it is clear that additional patterns or features are necessary for detection.

V. CONCLUSION AND FUTURE WORK

In this paper, we introduced objective evaluation methods to evaluate the entire discussion by giving logical structure tags and speech act tags to discussion utterance logs, methods to analyze and detect deadlock portions, and the discussion evaluation support tool Corte. We also evaluated this proposed tool by using the actual mock arbitration and mock mediation logs. Our evaluation results showed that the tool output showed a similar tendency to the evaluation results obtained by experts. On the other hand, when it comes to detecting deadlock portions, detection by using tags is partially effective, while deadlock patterns need to be analyzed more.

Future issues include the enhancement of analysis and evaluation discussion logs. Annotation was done manually in the tool proposed this time. We are actually working on semi-automatic annotation [15]. Automatic extraction of tag patterns of deadlock portions needs to achieve further energy-saving measures for users.

REFERENCES

[1] Intercollegiate Negotiation Competition. http://www.osipp.osaka-u.ac.jp/inc/index.html.

[2] Toulmin, S. 1958. The Uses of Argument, Cambridge University Press.

[3] Reed, C. and Rowe, G. 2001. Araucaria: Software for Puzzles in Argument Diagramming and XML. Technical Report. Dept. of Applied Computing. University of Dundee.

[4] Nitta, K. et. al. 2002. A Legal Negotiatiton Support System Based on A Diagram (in Japanese), Transactions of the Japanese Society for Artificial Intelligence, Vol.17, No.1 D, pp.32-43.

[5] Shibata, Y. and Yamaguchi, K. 2011. SPURI: Argument Analysis Framework (in Japanese), Journal of Information Processing, 52(3), pp. 1395-1411.

[6] Lynch, C., Ashley, K. D., and Falakmassir, M. 2012. Comparing Argument Diagrams. In JURIX 2012: The Twenty-Fifth Annual Conference on Legal Knowledge and Information Systems. pp. 81-90.

[7] Hashida, K. 1998. GDA Versatile and Intelligent Contentware with Semantic Annotation (in Japanese). Transactions of the Japanese Society for Artificial Intelligence, Vol.13, No.4, pp.528-535.

[8] Tanaka, T. 2007. Case Based Online Mediation Support System (in Japanese). Doctoral Thesis. Tokyo Institute of Technology.

[9] Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence, 77, (2): pp. 321-357.

[10] Prakken, H. 2011. An overview of formal models of ar- gumentation and their application in philosophy. Studies in logic. Vol 4, no 1: pp. 65-86.

[11] Ashley, K. 1991. Modeling legal arguments: Reasoning with cases and hypotheticals. MIT Press.

[12] Aleven, V. 1997. Teaching case based argumentation through an example and models. Doctoral Thesis. The University of Pittsburgh.

[13] Levin, K., H. 2007. Chouteisya Handobukku - Choutei No Rinen To Gihou (in Japanese). Shinzansya.

[14] Hall, L., M. and Bodenhamer, B., G. 2011. MIND-LINES: Lines For Changing Minds, 5th ed (Yuile, Y. trans.). Empowerment Technologies (Original work published in 1997).

[15] Kubosawa, S., Lu, Y., Okada, S. and Nitta, K. 2012. Argument Analysis with Factor Annotation Tool. In JURIX 2012: The Twenty-Fifth Annual Conference on Legal Knowledge and In- formation Systems. 61-70.

 ISBN: 978-1-61208-461-9