Computational Models of Argument
H. Prakken et al. (Eds.)

doi:10.3233/FAIA200516

Increasing the Naturalness of an Argumentative Dialogue System Through Argument Chains

Niklas RACH [a] Wolfgang MINKER [a] and Stefan ULTES [b]
[a] *Institute of Communications Engineering, Ulm University, Germany*
[b] *Mercedes-Benz Research & Development, Sindelfingen, Germany*

Abstract. This work introduces chained arguments into a dialogue game for argumentation to allow a more natural and intuitive interaction with a respective system. Thus, the turn taking rules of the game are improved while still preserving the general consistency that is ensured by the framework. The improved system is used to generate artificial dialogues between two virtual agents which are assessed in a user study. The results show a significant improvement in the perceived naturalness without violating the logical consistency.

Keywords. Dialogue Games for Argumentation, Argumentative Dialogue, Argumentation Strategies

1. Introduction

Empowering virtual agents with the ability to exchange arguments and to engage in argumentative dialogue is a desirable, yet challenging task. Due to the complexity of suchlike conversations, respective systems often utilize formal modelling of the dialogue in order to achieve consistency and reasonableness in the interaction [1, 2, 3, 4].

However, despite the logical advantage of dialogue models, the resulting interactions are restricted by the formalism and can be perceived as significantly less natural than human discussion [5]. The task at hand is thus to find the balance between reasonable restrictions and a freedom of choice that allows a natural and intuitive interaction. The difficulty lies in the implications that come with this freedom, as the possibility of a more natural response may also include the possibility of responses that are neither natural nor consistent and violate the basic principles of the desired interaction. Especially modifications to an established formalism have thus to preserve the general properties of the model and extend the respective regulations rather than simply drop them.

Within this work we address this task in view of the dialogue game for argumentation introduced in [6, 7]. This choice is due to the fact that the model allows for multiple as well as postponed responses to an utterance which means that it provides a certain freedom of choices for the players by design. We built upon this freedom and increase it by modifying the underlying game protocol to allow for a chaining of multiple arguments. This is done in a way that is in line with the remaining regulations and does thus not violate the logical consistency of the resulting dialogues. Our approach is tested by generating artificial dialogues between two virtual agents that are rated in a user study with respect to their logical consistency as well as their naturalness. The setup is similar to [5], where the unmodified version of the dialogue game was applied. We show that by adapting the protocol, the perceived consistency remains whereas the naturalness is significantly increased.

The remainder of this paper is as follows: Section 2 covers the related work on argumentative dialogue systems whereas Section 3 discusses the theoretical background of dialogue games for argumentation with an emphasis on the original framework. Section 4 addresses the applied modifications and their implications whereas Section 5 includes details on the evaluation setup, the user study and the corresponding results. The work is closed in Section 6 with a conclusion and an outlook on future work.

2. Related Work

Multiple approaches to human-machine argumentation have been discussed that utilize different models to structure the interaction. In the recently introduced IBM Debater[1], the boundaries of the interaction are given by the debating rules, meaning that speaking time and turn taking are fixed by the overall setup. As a consequence, the main task of the system consists of the automatic analysis of opponent utterances with fixed length and the generation of a suitable response.

One approach to address formal issues like turn taking is to limit the system response to one argument per turn [8, 9]. Consequently, additional options like questioning the validity of an argument or chaining multiple arguments are not considered. In addition, a generative approach to argumentative chat bots was discussed in [10]. Although in this case no rules in view of the interaction are imposed, the system capacities are limited to strategies that can be derived from the training data.

Similar to our setup, dialogue games for argumentation were previously considered as an approach to model argumentative dialogue. Overviews over existing dialogue games for argumentation were presented in [11, 12] and a framework to facilitate their implementation and the development of respective applications was introduced in [13]. Even though several systems utilizing these or similar frameworks in different domains were introduced [1, 2, 3, 14, 15], the main focus of the underlying formal models is usually to preserve logical coherence. Therefore, these models enforce restrictions that can lead to interactions that are not perceived as natural when compared to human discussions [5]. In order to address this issue, the herein introduced extension focuses on this explicit property in order to increase the freedom of choices within the framework and enable a more intuitive and natural argumentation.

3. Dialogue Games for Argumentation

Within this section we recall the theoretical background on dialogue games for argumentation, following the formal description introduced in [7]. Dialogue games in general are a model of conversation, meaning that they extend the formal approach of speech acts to their effect on the listener [16]. A dialogue game for argumentation can be described as

[1] https://www.research.ibm.com/artificial-intelligence/project-debater/

Table 1. Communication language L_c of the original framework [7]. Upper-case variables denote arguments out of *Args*, lower-case variables denote elements of L_t.

Speech Act	Attacks	Surrenders
claim(a)	*why(a)*	*concede(a)*
why(a)	*argue(A)* ($conc(A) = a$)	*retract(a)*
concede(a)	-	-
retract(a)	-	-
argue(A)	*why(a)* ($a \in prem(A)$), *argue(B)* (B defeats A)	*concede(a)* ($a \in prem(A)$ or $a = conc(A)$)

tuple $(\mathscr{L}, D)$, with $\mathscr{L}$ a logic for defeasable argumentation [17] and D the so called dialogue system proper. $\mathscr{L}$ includes the set of arguments *Args* on which a (binary) defeat relation is defined. Arguments are AND-trees with nodes out of a logical language L_t. The AND-links are instantiations of inference rules out of a set R defined over L_t. The set of leaves of an argument A is called its premises (*prem(A)*) and the root is called conclusion (*conc(A)*). We call an argument B an extending argument of A if $conc(B) \in prem(A)$.

The dialogue system proper D structures the interaction and consists of a communication language L_c, a game protocol P and commitment rules C. A game is played in turns, whereas each turn includes at least one *move*. A temporally ordered sequence of moves is called a *dialogue*.

In the following, we focus on Prakken's framework for *relevant dialogues* defined in [7]. The corresponding communication language L_c is shown in Table 1, ordered by attacking and surrendering replies. Each move in the corresponding game includes one speech act out of L_c as well as a temporal identifier and replies to one specific earlier move. The game is played by two players P1 (proponent) and P2 (opponent) and is initiated by the proponent with either a *claim* or an *argue* move.

In order to ensure consistency in the responses, the protocol P determines the legality of moves in each dialogue based on a relevance criterion. In order to determine this relevance, a binary status is assigned to each played move, defining it as either *in* or *out*. A move is *out* if the dialogue includes an attack on it that is *in*. Otherwise the move is *in*. If an attack on a move m_i would change the status of the initial move, m_i is a relevant target. The player to move can only address relevant targets in his or her turn. The turn

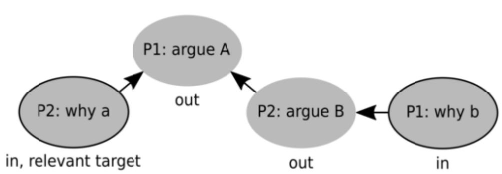

Figure 1. Illustration of the relevance criterion. Both *why* moves are not attacked and therefore *in* (indicated by black margins of the circles). Consequently, their targets are *out*. Only the *why(a)* move is a relevant target since an attack on it would change the status of the opening move *argue(A)*. Consequently, it is the turn of P1.

of each player ends once he or she manages to switch the status of the initial move in his or her favour. If a player has no legal move left and thus cannot switch the status of the

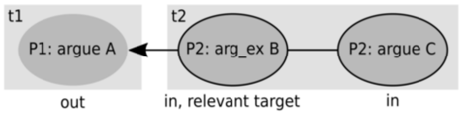

Figure 2. Illustration of a chain consisting of one *argue_extend* and one *argue* move. Grey boxes indicate the corresponding turns (t1 and t2) in the game.

initial move, he or she loses. An abstract example dialogue between two players P1 and P2 is shown in Figure 1 in order to illustrate the turn taking and the relevance criterion.

4. Extension to Chained Arguments

The main restriction of the above discussed formalism in view of naturalness lies in the inability of introducing more than one argument per turn. More precisely, a player is only allowed to extend an argument, if the corresponding move was challenged by a *why* move. This section introduces an extension which allows players to chain multiple arguments in a single turn without violating the logical consistency of the dialogue. In order to do so, we introduce an additional speech act and modifications to the protocol.

In the original formalism, a player has to move until he or she switched the status of the initial move in his or her favour, meaning that he or she plays an unspecified number of surrendering moves, followed by a single attack. After the status of the initial move is switched, the turn ends immediately. We modify this rule by allowing both players to *extend* their attack under the condition, that the attack includes an argument. An extended attack generally allows the player to introduce additional arguments to undermine his or her current move even before it was challenged. This extension does thus not reply to an actual attack but to an anticipated one. Formally, the extension is represented by an additional speech act (*argue_extend(A)*) which has the same properties as the *argue(A)* act in view of allowed attacking and surrendering replies but does not end the turn of the corresponding player. An *argue_extend* move can only be played if an extending argument is available. We call a series of *argue(_extend)* moves an (argument) *chain*.

In the following, we discuss implications and changes in the game that arise from this modification. When introduced, each move in a chain is *in*. The first move in a chain is also a relevant target since an attack on it changes the status of the initial move. The remaining moves on the other hand are not relevant targets. Moreover, challenging the relevant target in a chain only switches the status of the initial move if this challenge is not anticipated in the chain. Otherwise, the responding move in the chain becomes a relevant target and the current player is obliged to play another move. An example of a chain consisting of two moves, including status and relevance is shown in Figure 2. Attacking replies to a chain can have multiple forms as illustrated in Figure 3:

- A chain can be attacked by a series of anticipated *why* moves, followed by a *why* move that is not anticipated (Response 1).
- A chain can be attacked by a combination of anticipated *why* and *argue(_extend)* moves (Response 2).
- A chain can be attacked by an attacking reply to its first move that is not anticipated in the chain (Response 3).

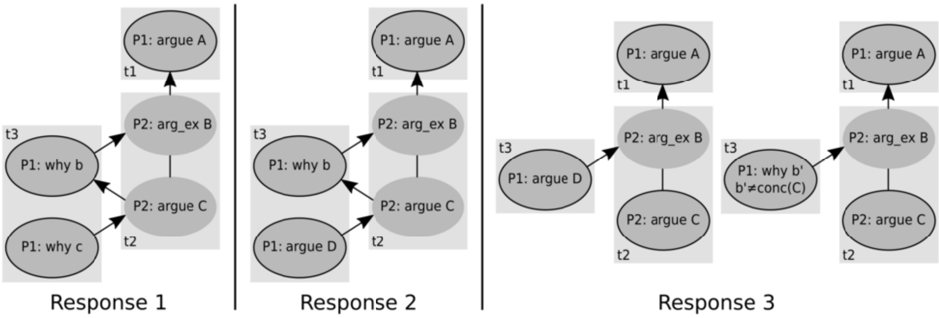

Figure 3. Illustration of the three possible response types to a chain. The arrow from the *argue C* move indicates the implicit response of the argument chain to the anticipated attack.

Generally, responses to a chain may also include a (new) chain, thus giving the players far more freedom in their choices. Nevertheless, since the legality of moves is still determined by the same principles as in the original framework, the resulting dialogues have the same formal consistency.

5. Evaluation

In order to evaluate the discussed extensions, we generate artificial dialogues between two virtual agents Alice and Bob and assess them in a user study. The evaluation setup was chosen in order to compare the results directly with the ratings for the original framework in an earlier work [5]. In order to ensure a fair comparison, the setup is as similar as possible to the original one. Thus, we employ the same multi-agent setup, including the same arguments, the same dialogue manager model and a similar natural language generation (NLG) as well as the same questionnaire for the survey.

5.1. Multi-Agent Setup

The set of arguments is derived from 72 argument components on the topic "Marriage is an outdated institution" annotated on a debate from idebate.com[2] following the argument annotation scheme introduced in [18]. The annotation scheme includes three kinds of argument components (Major Claim - MC, Claim - C, and Premise - P) and two directed relations (support and attack) between them. Each component apart from the Major Claim targets exactly one other component with a relation. Consequently, the resulting structure can be represented as a tree from which we derive the arguments of the form $A = a, so\ b$ (a supports b) and $A' = a', so\ \neg b'$ (a' attacks b').

During the interaction, the agents select their next move from the list of available options provided by the dialogue game. In order to ensure a competitive but reasonable strategy, each agent first prefers attacking moves over surrendering moves, then *argue(_extend)* moves over *why* moves and finally immediate to postponed responses. If there are multiple options with the same preference, the selection between them is random. Moreover, both agents extend their line of argumentation as long as possible. As in

[2] https://idebate.org/debatabase (last accessed 06 May 2020)

Table 2. Excerpt of an artificial discussion on the topic *Marriage is an outdated institution* including speech acts and NLG output. Italic text indicates the annotated sentences of the argument components.

Speech Acts	Utterance
argue_ex(C_1, *so* ¬MC)	It seems to me that *marriage is an important institution to religious people.*
argue(P_1, *so* C_1)	I would like to go into that a little further. You see, *there are still such huge numbers of people who practice religions to which marriage is integral.*
why(C_1)	Unfortunately I didn't find that entirely convincing. Would you mind elaborating a little further?
argue(P_2, *so* ¬P_1)	In particular, there's one aspect of your argumentation that I have some doubts about. You said that *there are still such huge numbers of people who practice religions to which marriage is integral.* It seems to me that *religion as a whole is becoming less important and, with it, marriage is becoming less important.*

the original work, we allow an $argue(a, so\ \neg MC)$ attacking reply to *claim(MC)* to cover all available arguments. This attack can also be extended in the modified framework.

The NLG is done turn wise, meaning that all moves of a turn are merged into one utterance. As in the reference work, the natural language representation of arguments is gained from the annotated sentences of the argument components. Postponed *argue(_extend)* replies include the premise and the conclusion. In the case of a direct reply, the conclusion is left implicit. For the remaining moves, a list of templates is used from which the system selects randomly. Again, the natural language representation of the argument component in the move is left implicit for direct replies and explicitly included in the case of a postponed reply. In addition, we generate a new list of connecting and opening phrases in order to concatenate multiple moves into a single utterance. This part of the NLG is an extension to the original version and may influence the user perception of the resulting dialogues. However, since this extension is only possible due to the extended framework, the advanced NLG template is a direct result of the formal extensions. An example of two utterances[3] and the speech acts of the corresponding turns is shown in Table 2.

5.2. User Study

To compare our approach with the original framework, we generated ten virtual discussions between the agents Alice (proponent) and Bob (opponent) with the new framework and evaluated them in a user survey with the same study setup as in the referenced work. In the original case, 20 dialogues were required in order to cover a majority of the available arguments, which was mainly due to the extensive use of isolated *why* moves. As those are merged into a single utterance within the modified framework, ten dialogues were sufficient to present a similar amount of arguments.

The questionnaire consists of ten questions related to the strategy, the line of argumentation and the naturalness of the dialogue. Each question was rated on a five point scale (1 completely disagree, 5 completely agree) by 61 participants from the UK with an age between 18 and 99. The survey was realized by clickworker[4] and each participant

[3]Material reproduced from www.iedebate.org with the permission of the International Debating Education Association. Copyright ©2005 International Debate Education Association. All Rights Reserved

[4]https://marketplace.clickworker.com (last accessed 06 May 2020)

Table 3. Results for the original framework (Original) and extended one (Modified). Bold lines indicate a significant difference.

Question	Original	Modified	p
The arguments presented by Bob are logically consistent responses to the utterances they refer to.	4.0	4.0	0.36
The arguments presented by Alice are logically consistent responses to the utterances they refer to.	3.5	3.0	0.81
Bob's line of argumentation is not logically consistent.	2.0	2.0	0.85
Alice's line of argumentation is not logically consistent.	2.0	2.0	0.74
It was difficult to follow the line of argumentation throughout the debate.	**3.0**	**2.0**	**0.02**
The whole debate is natural and intuitive.	**2.0**	**4.0**	**0.02**

was assigned a single randomly selected discussion in order to avoid a bias. The wording of questions that are relevant for the herein discussed topic together with the corresponding median (original and modified framework) as well as the p value achieved with a Mann-Whitney-U test are shown in Table 3. We see that the four questions related to the logical consistency of the argumentation show no significant difference to the original results, whereas the p value for both questions related to the naturalness is below the threshold of 0.05. For the sake of completeness, we report that no significant difference was found for the questions omitted in Table 3. We conclude that the herein discussed modification significantly improves the perceived naturalness of the resulting dialogues without lowering the consistency.

6. Conclusion

This work discussed the extension of an existing dialogue game for argumentation in order to enable a more natural interaction with a respective system. Our approach allows for chained arguments from both sides while preserving the regulations that ensure consistency. We evaluated the new framework by generating artificial discussions between two virtual agents that were rated in a user study. In a direct comparison to the results achieved in the reference work with the original framework, we see a significant improvement in ratings related to the naturalness and intuitiveness of the dialogue. On the other hand the perceived consistency of the dialogue remains the same.

Future work will focus mainly on exploring the generated freedom by means of optimization techniques like reinforcement learning [19]. We will also investigate if and how the game protocol can be additionally modified, in order to increase the freedom of choices for respective agents further. Finally, we want to explore the new framework in the interaction between a dialogue system and real users.

Acknowledgements: This work has been funded by the Deutsche Forschungsgemeinschaft (DFG) within the project "How to Win Arguments - Empowering Virtual Agents to Improve their Persuasiveness", Grant Number 376696351, as part of the Priority Program "Robust Argumentation Machines (RATIO)" (SPP-1999).

References

[1] Yuan T, Moore D, Grierson A. A human-computer dialogue system for educational debate: A computational dialectics approach. International Journal of Artificial Intelligence in Education. 2008;18(1):3–26.

[2] Yuan T, Moore D, Reed C, Ravenscroft A, Maudet N. Informal logic dialogue games in human–computer dialogue. The Knowledge Engineering Review. 2011;26(2):159–174.

[3] Rach N, Weber K, Pragst L, André E, Minker W, Ultes S. EVA: A Multimodal Argumentative Dialogue System. In: Proceedings of the 2018 on International Conference on Multimodal Interaction. ACM; 2018. p. 551–552.

[4] Bench-Capon TJ. Specification and implementation of Toulmin dialogue game. In: Proceedings of JURIX. vol. 98; 1998. p. 5–20.

[5] Rach N, Langhammer S, Minker W, Ultes S. Utilizing argument mining techniques for argumentative dialogue systems. In: 9th International Workshop on Spoken Dialogue System Technology. Springer; 2019. p. 131–142.

[6] Prakken H. On dialogue systems with speech acts, arguments, and counterarguments. In: JELIA. Springer; 2000. p. 224–238.

[7] Prakken H. Coherence and flexibility in dialogue games for argumentation. Journal of logic and computation. 2005;15(6):1009–1040.

[8] Rosenfeld A, Kraus S. Strategical Argumentative Agent for Human Persuasion. In: ECAI; 2016. p. 320–328.

[9] Rakshit G, Bowden KK, Reed L, Misra A, Walker M. Debbie, the debate bot of the future. In: Advanced Social Interaction with Agents. Springer; 2019. p. 45–52.

[10] Le DT, Nguyen CT, Nguyen KA. Dave the debater: a retrieval-based and generative argumentative dialogue agent. In: Proceedings of the 5th Workshop on Argument Mining. ACL; 2018. p. 121–130.

[11] Prakken H. Formal systems for persuasion dialogue. The knowledge engineering review. 2006;21(2):163–188.

[12] Prakken H. Historical overview of formal argumentation. vol. 1. College Publications; 2018.

[13] Bex F, Lawrence J, Reed C. Generalising argument dialogue with the Dialogue Game Execution Platform. In: COMMA; 2014. p. 141–152.

[14] Hunter A, Chalaguine L, Czernuszenko T, Hadoux E, Polberg S. Towards computational persuasion via natural language argumentation dialogues. In: Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz). Springer; 2019. p. 18–33.

[15] Sakai K, Higashinaka R, Yoshikawa Y, Ishiguro H, Tomita J. Hierarchical Argumentation Structure for Persuasive Argumentative Dialogue Generation. IEICE TRANSACTIONS on Information and Systems. 2020;103(2):424–434.

[16] Mann WC. Dialogue games: Conventions of human interaction. Argumentation. 1988;2(4):511–532.

[17] Prakken H, Vreeswijk G. Logics for defeasible argumentation. In: Handbook of philosophical logic. Springer; 2001. p. 219–318.

[18] Stab C, Gurevych I. Annotating Argument Components and Relations in Persuasive Essays. In: COLING; 2014. p. 1501–1510.

[19] Rach N, Minker W, Ultes S. Markov Games for Persuasive Dialogue. In: COMMA; 2018. p. 213–220.