

Developing a Formal Model of Argumentation-based Dialogue

Mare Koit

Institute of Computer Science, University of Tartu, J. Liivi 2, Tartu, Estonia

Keywords: Dialogue, Argument, Dialogue Model, Dialogue Corpus, Knowledge Representation.

Abstract: We are considering dialogues in natural language where the participants (A and B) are arguing for and against of doing an action D by B . The participants can have similar or opposite communicative goals. If both A and B have the same goal („ B will do D “ or, respectively, „ B will not do D “) then they are cooperatively looking for arguments that will eliminate possible obstacles before achieving the goal. If the goals are opposite then the participants exchange arguments and counterarguments and one of them has finally to abandon his or her initial communicative goal. A model of dialogue has been developed which includes a model of argument. An analysis of human-human dialogue corpus is carried out in order to give a preliminary evaluation of the introduced model. A limited version of the model is implemented on the computer. Full implementation is planned as a future work.

1 INTRODUCTION

Many researchers have been modelling argumentation on the computer.

Rahwan et al (2004) discuss three approaches to automated negotiation: game-theoretic, heuristic-based and argumentation-based. Argumentation-based approaches to negotiation allow agents to ‘argue’ about their beliefs and other mental attitudes during the negotiation process.

Besnard and Hunter (2008) formalize argumentation by using classical logic and define an argument as a pair $\langle \Phi, \alpha \rangle$ where Φ is a set of formulas (a subset of the knowledge base) and α is a formula such that (1) Φ is consistent; (2) Φ entails α ; (3) Φ is a minimal subset of the knowledge base which satisfies 2. If $\langle \Phi, \alpha \rangle$ is an argument, it is said that it is an argument for α and it is also said that Φ is a support for α . Here α is called the claim of the argument.

Logical models of argument support decision making by participants, guide negotiation and allow to reach agreements (Amgoud and Cayrol, 2002).

Rahwan and Larson (2011) explore the relationships between mechanism design and formal logic, particularly in the design of logical inference procedures when knowledge is shared among multiple participants.

Hadjinikolis et al. (2012) provide an argumentation-based framework for persuasion dialogues, using a logical conception of arguments, that an agent may undertake in a dialogue game, based on its model of its opponents.

Overviews of the state of art in modelling argumentation can be found e.g. in (Chesñevar et al., 2000) and (Besnard and Hunter, 2008).

We are studying the interactions in natural language between two participants (A and B) where A is convincing B to do or, respectively, not to do an action D . We have worked out a dialogue model which includes a reasoning model as its part and implemented it in a simple dialogue system (Koit and Õim, 2000; 2014; Koit, 2015).

In the current paper, we will further develop the model. The participants of dialogue exchange arguments for and against of doing D . They can also ask and answer questions in order to make choices among the arguments for averting the partner’s counterarguments.

The rest of the paper is structured as follows. Section 2 introduces our current model of argumentation-based dialogue. Section 3 gives the results of analysis of human-human dialogues, in order to justify the model. Section 4 discusses some questions related to the concepts of argumentation, negotiation and debate in human-human interaction and in our computational model. Conclusions will be made in Section 5.

2 DIALOGUE MODEL

2.1 The Structure of Dialogue

Let us consider a dialogue in natural language between two participants (humans or artificial agents) – A and B (Koit and Öim, 2014). Let the initiator of dialogue be A , and let his communicative goal be “ B will do an action D ” or, respectively, “ B will not do D ”. B ’s communicative goal can conform to A ’s one or it can be opposite. In interaction, A is influencing B to make a decision about doing D which coincides with his communicative goal. The following cases can occur:

- (1) A ’s goal is “ B will do D ”, and B ’s goal is “ B will do D ”;
- (2) A ’s goal is “ B will do D ”, and B ’s goal is “ B will not do D ”;
- (3) A ’s goal is “ B will not do D ”, and B ’s goal is “ B will do D ”;
- (4) A ’s goal is “ B will not do D ”, and B ’s goal is “ B will not do D ”.

In the cases (1) and (4) A and B have the same goal and in interaction, they are cooperatively looking for reasons (arguments) why to do (respectively, not to do) D and how to overcome possible obstacles before doing D or, respectively, to prevent possible undesirable results of not doing D .

In the cases (2) and (3), A and B have opposite goals and in interaction, the initiator A is proposing arguments which should influence B to accept A ’s goal and to abandon her own initial goal. At the same time, B can propose counterarguments which should force A to accept B ’s goal and to abandon his own initial goal.

A as the initiator has a partner model in his disposal – an image about B which gives him an opportunity to suppose that B will agree to accept his communicative goal (to do or, respectively, not to do the action D). In constructing his first turn, A must plan the dialogue acts (e.g. proposal, request, question, proposal together with an argument, etc. depending on his image of B) and determine their verbal form (i.e. utterances). The partner B interprets A ’s turn and before generating her response, triggers a reasoning procedure in her mind in order to make a decision – to do D or not. In the reasoning process, B weighs her resources for doing D , positive and negative aspects of doing D and its consequences and finally, makes a decision. Then she in her turn will plan the dialogue acts (e.g. agreement, refusal, refusal with argument, etc.) and their verbal form in order to inform A about her decision. If B agrees to accept A ’s

goal then the dialogue finishes (A has reached his communicative goal). If B ’s response is refusal then A must change his partner model (it did not correspond to the reality because A supposed that B will agree to accept A ’s goal) and find out new arguments in order to convince B to make a positive decision.

Our **reasoning model** has been introduced in (Koit and Öim, 2000; 2014). It consists of two parts: (1) a model of human motivational sphere; (2) reasoning procedures.

In the motivational sphere three basic factors are differentiated that regulate reasoning of a subject concerning an action D . First, a subject may *wish* to do D if the pleasant aspects of D for him/her overweight the unpleasant ones; secondly, a subject may find it reasonable to do D if D is *needed* to reach some higher goal, and the useful aspects of D overweight the harmful ones; and thirdly, a subject *must* (is obliged) to do D – if not doing D will lead to some kind of punishment. We call these factors WISH, NEEDED and MUST determinants, respectively.

If the subject is reasoning about **not doing** D then the basic factors which trigger the reasoning are analogous: first, the subject *does not wish* to do D if unpleasant aspects of D overweight the pleasant ones; secondly, doing D is *not needed* for him/her if harmful aspects of D overweight the useful aspects; and thirdly, doing D is *prohibited* (not allowed) for him/her and will cause some punishment. We call these factors NO-WISH, NOT-NEEDED and NOT-ALLOWED determinants, respectively.

Let us represent the model of motivational sphere of a subject concerning an action D by the following vector of ‘weights’ (with numerical values of its components): $w_D = (w(\text{resources}_D), w(\text{pleasant}_D), w(\text{unpleasant}_D), w(\text{useful}_D), w(\text{harmful}_D), w(\text{obligatory}_D), w(\text{prohibited}_D), w(\text{punishment-do}_D), w(\text{punishment-not}_D))$.

In the description, $w(\text{pleasant}_D)$, etc. mean the weight of pleasant, etc. aspects of D ; $w(\text{punishment-do}_D)$ – weight of punishment for **doing** D if it is prohibited, and $w(\text{punishment-not}_D)$ – weight of punishment for **not doing** D if it is obligatory. Further, $w(\text{resources}_D) = 1$, if subject has all the resources necessary to do D (otherwise 0); $w(\text{obligatory}_D) = 1$, if D is obligatory for the reasoning subject (otherwise 0); $w(\text{prohibited}_D) = 1$, if D is prohibited (otherwise 0). The values of other weights can be non-negative natural numbers. In the following, we suppose that the action D is fixed and do not indicate it explicitly in the vector.

The second part of the reasoning model consists of reasoning procedures that supposedly regulate human action-oriented reasoning. A reasoning procedure depends on the determinant which triggers it (in our model, WISH, NEEDED, MUST, or respectively, NO-WISH, NOT-NEEDED, NOT-ALLOWED). As an example, let us present a procedure triggered by the NOT-ALLOWED determinant.

Presumption: D is prohibited.

1) Are there enough resources for doing D ? If not then goto 8.

2) Is $w(\text{pleasant}) > w(\text{unpleasant})$? If not then 8.

3) Is $w(\text{pleasant}) > w(\text{unpleasant}) + w(\text{punishment-do})$? If not then goto 8.

4) Is $w(\text{pleasant}) > w(\text{unpleasant}) + w(\text{punishment-do}) + w(\text{harm})$? If not then goto 8.

5) Is $w(\text{pleasant}) + w(\text{useful}) > w(\text{unpleasant}) + w(\text{punishment-do}) + w(\text{harm})$? If not then goto 8.

7) Decide: do D . End.

8) Decide: do not do D .

The vector \mathbf{w}^{AB} (A 's beliefs concerning B 's evaluations in relation to the action D) is used as a partner model while the vector \mathbf{w}^B – the model of B herself – represents B 's actual evaluations of D 's aspects (which exact values A does not know).

A **communicative strategy** is an algorithm used by a participant for achieving his/her goal in the interaction. The initiator (participant A) can realize his communicative strategy in different ways: stress pleasant or respectively, unpleasant aspects of D (i.e. *entice* the partner B), stress usefulness or, respectively, harmfulness of D for B (i.e. *persuade* B), stress punishment for not doing D if it is obligatory or respectively, punishment for doing D if it is prohibited (*threaten* B), etc. These concrete ways of realization of a communicative strategy we call **communicative tactics**. A , trying to direct B 's reasoning to the desirable decision, proposes arguments for doing D (respectively, not doing D) while B , when opposing, proposes counterarguments.

When influencing B in interaction, A can bring out different aspects of D . Implementing certain communicative tactics in a systematic way A will choose one aspect of D (the 'title' aspect of the fixed tactics) and proposes arguments for stressing it.

In order to achieve B 's decision **to do D** , A can stress the following 'title' aspects:

- pleasantness of D (i.e. to trigger B 's reasoning procedure by the WISH determinant)

- usefulness of D (to trigger the reasoning procedure by the NEEDED determinant)
- punishment for not doing D if D is obligatory for B (to trigger the reasoning procedure by the MUST determinant).

Similarly, in order to achieve B 's decision **not to do D** , A can stress unpleasantness, harmfulness or punishment for doing D .

The **knowledge base** for the agent A includes (1) reasoning algorithms, (2) communicative strategies and tactics, (3) the partner model \mathbf{w}^{AB} , (4) a list of dialogue acts which A can use (proposal, question, assertion, etc.), (5) a list of utterances which he can use for verbalizing the dialogue acts.

The knowledge base for B includes similar knowledge, the only difference is that is \mathbf{w}^B (the model of B herself) is used instead of the partner model \mathbf{w}^{AB} .

When interacting about an action, A and B exchange arguments. The general structure of A 's **argument** is as follows, cf. (Amgoud and Cayrol, 2002; Besnard and Hunter, 2008; Koit, 2015):

$\langle \{R, T, \mathbf{w}^{AB}_i, \text{proposition}^A\}, \text{claim}^A \rangle$,

where

- R is the reasoning procedure which A is trying to trigger in B
- T is the communicative tactics used
- $\mathbf{w}^{AB}_i = (w^{AB}_i(\text{resources}), w^{AB}_i(\text{pleasant}), w^{AB}_i(\text{unpleasant}), w^{AB}_i(\text{useful}), w^{AB}_i(\text{harmful}), w^{AB}_i(\text{obligatory}), w^{AB}_i(\text{prohibited}), w^{AB}_i(\text{punishment-do}), w^{AB}_i(\text{punishment-not}))$ is the current partner model (at turn i of the dialogue)
- proposition^A denotes the utterance chosen by A in order to influence one of the weights in the partner model, after what R will supposedly give B 's positive decision on the changed model (which coincides with A 's communicative goal) its weight is $w(\text{proposition}^A)$
- $\text{claim}^A = \text{"}B \text{ will do } D\text{"}$ or, respectively, $\text{"}B \text{ will not do } D\text{"}$.

The proposition^A chosen by A in interaction yields a new partner model \mathbf{w}^{AB}_{i+1} (at turn $i+1$):

- if $\text{proposition}^A \in P_{\text{increase_resources}}$, then $w^{AB}_{i+1}(\text{resources}) := 1$
- if $\text{proposition}^A \in P_{\text{increase_pleasantness}}$, then $w^{AB}_{i+1}(\text{pleasant}) := w^{AB}_i(\text{pleasant}) + w(\text{proposition}^A)$,

etc.

Here $P_{\text{increase_resources}}$ denotes the set of propositions (utterances) that can be used for indicating that there exist resources for doing D ; $P_{\text{increase_pleasantness}}$ denotes

the set of utterances for increasing the pleasantness of D , etc.

The structure of B 's argument is analogous:

$\langle \{R^B, T^B, \mathbf{w}^B, \text{proposition}^B\}, \text{claim}^B \rangle$,

where

- the reasoning algorithm R^B gives the decision “do not do D ” or respectively, “do D ” (claim^B) on the model \mathbf{w}^B
- proposition^B indicates the aspect of D which (too small or too big) value causes this decision
- T^B is the current communicative tactics of B .

Here B 's proposition^B gives to A information for choosing his next proposition (as argument) in interaction. For example, if A is arguing for doing D and $\text{proposition}^B \in P_{\text{missing_resources}}$, then the actual value of $w_{i(\text{resources})}^{AB}$ is 0 and the next utterance will be chosen by A from the set $P_{\text{increasing_resources}}$ (after that, $w_{i+1(\text{resources})}^{AB} = 1$ will hold) and another proposition will be chosen from the set of propositions which correspond to the title aspect of the reasoning algorithm R which A is trying to trigger in B using the communicative tactics T .

In order to choose the next proposition (counterargument), B triggers her current reasoning procedure R^B in her model \mathbf{w}^B , and finally, B is able to determine the aspect of D which brought her to the negative decision. For example, she can choose an utterance indicating to missing resources, e.g. by saying *I don't have so much money as needed to do D* but she can also refuse by saying *I do not do D* . In the last case, A cannot avert any counterargument but he has to make a choice among the utterances for stressing the title aspect of the implemented communicative tactics T .

2.2 Argumentation-based Dialogue

If A and B have contradictory goals when starting interaction then they are involved into debate (e.g. A 's communicative goal is “ B will do D ”, B 's goal is “ B will not do D ”). One participant will achieve his or her communicative goal (‘wins’ debate) and another has to abandon her or his initial goal (‘loses’ debate).

If A and B have common communicative goals then they are cooperatively looking for arguments that support achieving this collective goal. Still, for example, B can indicate to obstacles which do not allow achieve the goal. Then A has to find arguments for showing how the obstacles can be eliminated. The final result of discussion is whether achieving the collective goal or its withdrawal if some of the obstacles cannot be eliminated.

Let us suppose that both A and B have a common set of reasoning procedures. We also suppose that both A and B can use fixed sets of dialogue acts (e.g. proposal, question, agreement, refusal, statements for increasing or decreasing the values of different components of the vector of motivational sphere which will be used as arguments for doing or not doing D) and corresponding utterances which are classified semantically, e.g. $P_{\text{increasing_resources}}$ for indicating that there exist resources for doing D , $P_{\text{increasing_pleasantness}}$ for stressing pleasantness of D , $P_{\text{missing_resources}}$ for indicating that some resources for doing D are missing, $P_{\text{decreasing_pleasantness}}$ for decreasing pleasantness of D , etc.

Starting interaction, A fixes a partner model \mathbf{w}^{AB} using his pre-knowledge about B , and determines the communicative tactics T which he will use, i.e. he accordingly fixes a reasoning algorithm R which he will try to trigger in B 's mind. B has her own model \mathbf{w}^B . She determines a reasoning procedure R^B which she will use in order to make a decision about doing D .

The structure of argumentation-based dialogue looks like follows (the dialogue acts in parentheses can miss):

```
A: proposal (+ argument)
REPEAT
(
  B: question
  A: answer/giving information
)
B: agreement OR refusal (+ argument)
(
  A: question
  B: answer/giving information
)
A: argument
UNTIL a finishing condition is fulfilled.
```

Whether A or B can indicate that a finishing condition is fulfilled. The finishing conditions are: (1) the communicative goal is already achieved, (2) the participant gives up (2.1) regardless of having utterances for expressing new arguments, or (2.2) there are no utterances to continue the fixed communicative tactics but no new tactics will be chosen regardless of having some tactics not implemented so far, or (2.3) all the tactics are already implemented and all the utterances are used but the communicative goal is not achieved.

Questions can be asked by participants in order to make choices between different utterances which can be used in argumentation.

3 ANALYSIS OF HUMAN-HUMAN DIALOGUES

Does the structure of actual human-human dialogues coincide with the structure presented in Section 2.2? We carried out an analysis of dialogues taken from the Estonian dialogue corpus (Hennoste et al., 2008): (1) 22 everyday calls, and 4 face-to-face conversations between acquaintances, and (2) 24 calls of a customer who is planning a trip with a travel agent.

Let us consider two examples. The first example is an everyday phone call between mother and daughter. The second example is a face-to-face conversation in a travel agency. Transcription of Conversation Analysis (Sidnell and Stivers, 2012) is used in the examples.

Example 1. Here, the mother *A* presents several arguments in order to increase her daughter's wish to bake gingersnaps (the action *D*). *A*'s last argument (*I will not be at home*) turns out to be sufficient for making a positive decision by *B*.

/---/

A: .hhhhh kas sulle pakuks 'pinget 'piparkookide 'küpsetamine.

would you like to bake gingersnaps

proposal

B: .hhhhhhh ma=i='tea vist 'mitte.

I don't know, perhaps not

refusal

A: ja=sis gla'suurimine=ja='nii.

and then glazing and so on

proposition^{A1}

(0.6)

B: 'ei, 'ei, 'ei ei='ei.

no, no, no, no, no

refusal

(0.9)

A: me saaksime nad 'vanaema=jurde 'kaasa võtta.

we could take them with us when going to visit grandmother

proposition^{A2}

(0.4)

B: 'präägu ei='taha.

I don't want just now

refusal

/---/

A: ma mõtlen: kui mind kodus ei='ole.

I suppose when I will not be at home

proposition^{A3}

B: aa.

ah

(0.5) .hhh et 'lähen ostan 'tainast=vä.

then I'll go to buy paste, yes

agreement

Example 2. The travel agent *A* presents several arguments attempting to indicate that the proposed trip (which is here the action *D*) is interesting/ useful for the customer. *B* asks questions in order to make a decision.

/---/

A: m:eil on 'sellel aastal (.) uus 'reis välja pakkuda, see on Sit'siilia.

we offer a new trip this year to Sicily

proposal

(.) see peaks teid kindlasti 'huvitama, see on nimelt niisuguse omapärane mt=.hh 'kant I'taalias.

you should like it, this is an original place in Italy

proposition^{A1}

(0.6)

B: ee (0.6) mis:=mis:=ee (0.4) mis=a- aja- 'aegadel teil on

which time do you offer

question

/---/

A: @ te näete antiik ja ba'rokkunsti ja saate suurepärase

'võimaluse {-} 'puhata Dürreeni mere 'ran[nikul.] @

you will see ancient and baroque art and you will have an excellent chance to take a rest at the coast of Tyrrhenian sea

proposition^{A2}

/---/

B: et=ee (.) kas see nagu 'väljasõidud ja=kõik=e (.) kas ned=on=nagu: 'hinna 'sees kohe või net: tuleb 'eraldi arvestada.

are the set-offs included in the price or have they to be paid separately

question

/---/

The results of the corpus analysis show that the introduced model can be in general lines suitable for analysis of Estonian human-human dialogues and it can be taken as a basis of a dialogue system.

4 DISCUSSION

We are considering dialogues where two participants argue about doing an action *D* by one of them. Here we would like to explain our understanding of the relationships between such concepts as argumentation, negotiation, and debate as used in the paper.

Argumentation (as a discussion in which reasons are advanced for and against some proposition or proposal) constitutes a necessary part of negotiations and debates. Both in negotiation and in debate there are clearly fixed 'sides' with different goals when considering the outcome of the communicative event. However, negotiation covers much more divergent possible variants than debate. "Negotiation is a form of interaction in which a group of agents with

conflicting interests and a desire to cooperate try to come to a mutually acceptable agreement on the division of scarce resources“ (Rahwan et al., 2004). The main uniting feature of all variants of negotiation is that the participants start the communicative event with the ultimate aim to reach an agreement which is seen as a compromise, that is, all sides are ready to accept some losses. Debate is an adversarial event from the start: the participants have conflicting goals and the aim of each participant is to promote his or her goal only.

The model presented in Section 2 covers a certain limited kind of negotiations about doing an action. If A and B are pursuing the same communicative goal then they start discussion in order to explain that there are no obstacles before doing the action D or, respectively, no undesirable consequences follow after D will not be done. They do not necessarily achieve their joint communicative goal. The model does not consider the situations where the initial goal will be modified. If the goals are opposite then A and B are involved into debate where one participant wins and another loses.

The structure of argument used in the model is adapted to the limited kind of negotiations considered here. When arguing, a participant presents only one part of argument – proposition(s); the remaining parts are implicit (cf. the examples in Section 3).

5 CONCLUSION AND FUTURE WORK

We introduced a model of argumentation-based dialogue which includes exchange of arguments. A model of argument is presented which consists of a partner model for A (or, respectively, a model of herself for B), a reasoning procedure which A tries to trigger in B (or what B is implementing herself), communicative tactics and (a set of) proposition(s) (utterances) which all together would bring A and/or B to a desirable conclusion. The conclusion (a decision about doing D by B) is interpreted as a claim in the structure of argument.

We evaluated our model on actual human-human dialogues taken from a dialogue corpus. The corpus study gives an opportunity to believe that the introduced model can be used for the analysis of human-human dialogues and modelling them in a dialogue system.

We have implemented on the computer a simple argumentation-based dialogue (debate) where A 's communicative goal is “ B will do D ” and B 's goal is,

on the contrary, “do not D ” (Koit, 2015). Our future work includes implementation of the whole model.

ACKNOWLEDGEMENTS

This work was supported by the Estonian Research Council (project IUT-2056).

REFERENCES

- L. Amgoud, and C. Cayrol. 2002. A Reasoning Model Based on the Production of Acceptable Arguments. In *Ann. Math. Artif. Intell.* 34(1-3): 197–215.
- P. Besnard, and A. Hunter. 2008. *Elements of Argumentation*, MIT Press, Cambridge, MA.
- C. Chesñevar, A. Maguitman, and R. Loui. 2000. Logical Models of Argument. In *ACM Computing Surveys*, 32(4), 337–383.
- C. Hadjinikolis, S. Modgil, E. Black, P. McBurney, and M. Luck. 2012. Investigating Strategic Considerations in Persuasion Dialogue Games. In *STAIRS*, 137–148.
- T. Hennoste, O. Gerassimenko, R. Kasterpalu, M. Koit, A. Rääbis, and K. Strandson. 2008. From Human Communication to Intelligent User Interfaces: Corpora of Spoken Estonian. In *Proc. of the 6th International Language Resources and Evaluation (LREC'08)*. Marrakech, Morocco: European Language Resources Association (ELRA), 2025–2032.
- M. Koit. 2015. Communicative Strategy in a Formal Model of Dispute. In *Proc. of the International Conference on Agents and Artificial Intelligence: 7th International Conference on Agents and Artificial Intelligence (ICAART)*, Lisbon, Portugal, SCITEPRESS, 489–496.
- M. Koit, and H. Õim. 2014. A Computational Model of Argumentation in Agreement Negotiation Processes. In *Argument & Computation*, 5 (2-3), 209–236, Taylor & Francis Online. DOI: 10.1080/19462166.2014.915233
- M. Koit, and H. Õim. 2000. Developing a Model of Natural Dialogue. In *From spoken dialogue to full natural interactive dialogue-theory, Empirical analysis and evaluation. LREC2000 Workshop proceedings*, 18–21.
- I. Rahwan, and K. Larson. 2011. Logical Mechanism Design. In *The Knowledge Engineering Review*, 26(1), 61–69.
- I. Rahwan, S. D. Ramchurn, N. R. Jennings, P. Mcburney, S. Parsons, and L. Sonenberg. 2004. Argumentation-Based Negotiation. In *The Knowledge Engineering Review*, Vol. 18:4, 343–375. Cambridge University Press. DOI: 10.1017/S0269888904000098
- J. Sidnell, and T. Stivers (eds.). 2012. *Handbook of Conversation Analysis*, Boston: Wiley-Blackwell.