Argumentation in the Agreement Negotiation Process: A Model that Involves Natural Reasoning

Mare Koit and **Haldur Õim** [1]

Abstract. This paper describes a computational model that we are implementing in an experimental dialogue system. Agreement negotiation process is modelled where one participant is trying to influence his/her partner to agree to do an action. Argumentation is used to direct reasoning of the partner. Our goal is to model natural dialogue where computer as a dialogue participant follows norms and rules of human-human communication.

1 INTRODUCTION

When one person initiates communication with another (s)he mainly proceeds from the fact that the partner is a human being who, first, feels, reasons and has wishes and plans like every human being and, secondly, as this particular individual person. In order to be able to foresee what processes will be triggered in the partner after a dialogue act, the agent must know the inner workings of the partner's psychological mechanisms. When aiming at a certain goal in communication, the agent must know how to direct the functioning of these mechanisms (actually, how to argue) in order to bring about the intended result in the partner.

In our work we have dealt with interactions where the goal of one of the partners, A, is to get another partner, B, to carry out a certain action D. Such communication process can be treated as exchange of arguments (and counter-arguments). This type of dialogue constitutes one kind of so-called agreement negotiation dialogues [10]. Such dialogue can be considered, on a more general level, as rational behaviour of conversation agents which is based on beliefs, desires and intentions of agents, at the same time being restricted by their resources [4], [11].

Because of this, we have modelled the reasoning processes that people supposedly go through when working out a decision whether to do an action or not. In a model of conversation agent it is necessary to represent its cognitive states as well as cognitive processes. One of the most well-known models of this type is the BDI model [1], [2].

Our model is implemented as an experimental dialogue system and can be used, among other applications, as a "communication trainer".

2 MODELLING THE PROCESS OF COMMUNICATION

Let us consider conversation between two agents – A and B – in a natural language. In the goal base of one participant (let it be A) a certain goal G^A related to B's activities gets activated and triggers in A a reasoning process. In constructing his/her first turn A must plan the dialogue acts (DA) and determine their verbal form as a turn r_1. This turn triggers a reasoning process in B where two types of procedures should be distinguished: the interpretation of A's turn and the generation of his/her response r_2. B's response triggers in A the same kind of reasoning cycle in the course of which (s)he has to evaluate how the realization of his/her goal G^A has proceeded, and depending on this (s)he may activate a new sub-goal of G^A, and the cycle is repeated: A builds a new turn r_3. Dialogue comes to an end, when A has reached his/her goal or abandoned it.

[1] University of Tartu, Estonia email: mare.koit@ut.ee

CMNA IV, 4^{th} *Workshop on Computational Models of Natural Argument*
Edited by Floriana Grasso, Chris Reed and Giuseppe Carenini
ECAI 2004

2.1 Model of Conversation Agent

In our model a conversation agent is a program that consists of 6 (interacting) modules, cf. [6]: $(PL, PS, DM, INT, GEN, LP)$, where PL – planner, PS – problem solver, DM – dialogue manager, INT – interpreter, GEN – generator, LP – linguistic processor. PL directs the work of both DM and PS, where DM controls communication process and PS solves domain-related tasks. The task of INT is to make semantic analysis of partner's utterances and that of GEN is to generate semantic representations of agent's own contributions. LP carries out linguistic analysis and generation. Conversation agent uses in its work goal base GB and knowledge base KB which consists of 4 components: $KB = (KB_W, KB_L, KB_D, KB_S)$, where KB_W contains world knowledge, KB_L – linguistic knowledge, KB_D – knowledge about dialogue and KB_S – knowledge about interacting subjects. KB_D contains definitions of dialogue acts (declarative knowledge) and algorithms that are applied to reach communicative goals - communicative strategies and tactics (procedural knowledge). KB_S contains knowledge about evaluative dispositions of participants towards the action(s) (e.g. what do they consider as pleasant or unpleasant, useful or harmful), and, on the other hand, algorithms that are used to generate plans for acting on the world. A necessary precondition of interaction is existence of shared (mutual) knowledge of agents.

2.2 Reasoning Model

After A has expressed his/her wish to B that B does D, B can respond with agreement or rejection, depending on the result of his/her reasoning. Rejection can be supported with an argument. These arguments can be used as giving information about the reasoning process that brought B to the given decision.

In general lines our reasoning model follows the ideas realised in the BDI model. But it has a certain particular features we would like to stress [7]. We want to model a "naive" theory of reasoning, a "theory" that people themselves use when they are interacting with other people and trying to predict and influence their decisions. That is, we depart from what psychologists call "theory theory of mind" [3].

The reasoning model consists of two functionally linked parts: 1) a model of human motivational sphere; 2) reasoning schemes. In the motivational sphere three basic factors that regulate reasoning of a subject concerning D are differentiated. First, subject may wish to do D, if pleasant aspects of D for him/her overweight unpleasant ones; second, subject may find reasonable to do D, if D is needed to reach some higher goal, and useful aspects of D overweight harmful ones; and third, subject can be in a situation where (s)he 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-factors, respectively.

Resources of the subject concerning D constitute any kinds of internal and external circumstances which create the possibility to perform D and which are under the control of the reasoning subject.

The values of the dimension obligatory/prohibited are in a sense absolute: something is obligatory or not, prohibited or not. On the other hand, the dimensions pleasant/unpleasant, useful/harmful have a scalar character: something is pleasant or useful, unpleasant or harmful to a certain degree. For simplicity's sake, it is supposed that these aspects have numerical values and that in the process of reasoning (weighing the pro- and counter-factors) these values can be summed up.

In reality people do not operate with numbers but, rather, with some fuzzy sets. On the other hand, existence of certain scales also in human everyday reasoning is apparent. For instance, for the characterisation of pleasant and unpleasant aspects of some action there are specific words: *enticing, delightful, enjoyable, attractive, acceptable, unattractive, displeasing, repulsive* etc. Each of these adjectives can be expressed quantitatively.

We have represented the model of motivational sphere of a subject by the following vector of weights: w = ($w(resources)$, $w(pleas)$, $w(unpleas)$, $w(use)$, $w(harm)$, $w(obligatory)$, $w(prohibited)$, $w(punish)$, $w(punish\ -\ not)$). In the description, $w(pleas)$, $w(unpleas)$, $w(use)$, $w(harm)$ mean weight of pleasant, unpleasant, useful, harmful aspects of D, $w(punish)$ – weight of punishment for doing D if it is prohibited and $w(punish\ -\ not)$ – weight of punishment for not doing D if it is obligatory. Here $w(resources) = 1$, if subject has resources necessary to do D (otherwise 0); $w(obligatory) = 1$, if D is obligatory for the reasoning subject (otherwise 0); $w(prohibited) = 1$, if D is prohibited (otherwise 0). The values of other weights are non-negative natural numbers. If we consider many actions $D_1, ..., D_n$ instead of one action D, then similar components must be added into the vector of weights for all these actions.

The second part of the reasoning model consists of reasoning schemes, that supposedly regulate human action-oriented reasoning. A reasoning scheme represents steps that the agent goes through in his/her reasoning process; these consist in computing and comparing the weights of different aspects of D; and the result is the decision to do or not to do D.

How does the reasoning itself proceed? It depends on the determinant which triggers it (WISH, NEEDED or MUST). In addition, a reasoning model, as a naive theory of mind, includes some principles which represent the interactions between determinants and the causal connection between determinants and the decision taken. For instance, the principles fix such concrete preferences as:

- People want pleasant states and do not want the unpleasant ones.
- People prefer more pleasant states to less pleasant ones.

We do not go into details concerning these principles here. Instead, we refer to [7].

As an example, let us present a reasoning procedure which is triggered by NEEDED-determinant, that is, if the subject believes that it would be useful (needed) to do D in order to reach a goal.

```
Input considerations: w(use) > w(harm).

Are there enough resources for doing D?
  If not then do not do D.
Is w(pleas) > w(unpleas)?
  If not then go to 1.
Is D prohibited? If not then do D.
Is w(pleas)+w(use) > w(unpleas)+ w(harm)+
  w(punish)?
  If yes then do D. Otherwise do not do D.
1: Is D obligatory? If not then do not do D.
Is w(pleas) + w(use) + w(punish-not-D) >
  w(unpleas) + w(harm)?
  If yes then do D. Otherwise do not do D.
```

In the case of other input determinants (WISH, MUST) the general structure of the algorithm is analogous, but there are differences in concrete steps.

The reasoning model is connected with the general model of conversation agent in the following way. First, the planner PL makes use of reasoning schemes and second, the KB_S contains the vector w^A (A's subjective evaluations of all possible actions) as well as vectors w^{AB} (A's beliefs concerning B's evaluations, where B denotes agent(s) A may communicate with). The vectors w^{AB} are used as partner models.

When comparing our model with BDI model, then beliefs are represented by knowledge of the conversation agent with reliability less than 1; desires are generated by the vector of weights w^A; and intentions correspond to goals in GB. In addition to desires, from the weights vector we also can derive some parameters of the motivational sphere that are not explicitly conveyed by the basic BDI model: needs, obligations and prohibitions.

Some wishes or needs can be stronger than others: if $w(pleasD_i)$ - $w(unpleasD_i) > w(pleasD_j)$ - $w(unpleasD_j)$, then subject's wish to do D_i is stronger than the wish to do D_j. In the same way, some obligations (prohibitions) can be stronger than others, depending on the weight of the corresponding punishment.

3 DIALOGUE KNOWLEDGE

3.1 Dialogue Acts

In the descriptions of dialogue acts two types of knowledge are represented: first, the structure of the corresponding DA (static part), and second, the procedures that make up the reasoning processes that underlie the generation and interpretation of the corresponding DA (dynamic part). The frame formalism is used [8], [9]. For example, let us consider the argument frame: author A grounds (argues) an assertion q by an assertion p (which could be called as argument for q).

```
ARGUMENT (author A, recipient B)

I. Static part
SETTINGS:
(1) A believes that p
(2) A believes that q
(3) A believes that if p then q
```

```
(4) A believes that B believes that
    if p then q
GOAL: B believes that q
PLOT: A informs B that p
CONSEQUENCES:
(1) B believes that p
(2) B believes that q

II. Dynamic part
Generation procedures (implemented by A):
(1) Inform B that p or
(2) inform B that p, and if p then q
Interpretation-generation procedures
(implemented by B):
(1) Agree or
(2) reject (+ counter-argument)
```

In our case, argumentation is used only for increasing/decreasing weights of various aspects of actions. Thus asserting e.g. *The nature is very beatiful in Venice*, one tries to increase the weight of pleasantness of the action to travel to Venice.

3.2 Communicative Strategies and Tactics

In a general case, a communicative strategy is an algorithm used by a participant for achieving his/her goal in interaction. An agent can realise a communicative strategy by means of several communicative tactics (this concept more closely corresponds to the concept of communicative strategy as used in some other approaches, cf. [5].

There is one relevant aspect of human-human communication which is relatively well studied in pragmatics of human communication and which we have included in our model as the concept of communicative space. Communicative space is defined by a number of coordinates that characterise the relationships of participants in a communicative encounter. Communication can be collaborative or confrontational, personal or impersonal; it can be characterised by the social distance between participants; by the modality (friendly, ironic, hostile, etc.) and by intensity (peaceful, vehement, etc.). Just as in case of motivations of human behaviour, people have an intuitive, "naive theory" of these coordinates.

In our model the choice of communicative tactics depends on the "point" of the communicative space in which the participants place themselves. The values of the coordinates (social distance, intensity etc.) are again given in the form of numerical values.

In our case a communicative strategy can be presented as the following algorithm.

1. Choose a communicative tactic.
2. Choose an initial point in the communicative space.
3. Implement the tactic to generate an utterance: inform the partner of the communicative goal (agreeing to do an action D).
4. Did the partner agree to do D? If yes then finish (the communicative goal has been achieved).
5. Give up? If yes then finish (the communicative goal has not been achieved).
6. Change the point in the communicative space? If yes then choose a new point.
7. Change the communicative tactic? If yes then choose a new tactic.
8. Implement the tactic to generate an argument.
9. Go to the step 4.

The participant A can realize his/her communicative strategy in different ways (using different arguments for): stress pleasant aspects of D (i.e. entice B), stress usefulness of D for B (i.e. persuade B), stress punishment for not doing D if it is obligatory (threaten B). We call communicative tactics these concrete ways of realization of a communicative strategy. Actually, communicative tactics are ways of argumentation. The participant A, trying to direct B's reasoning to the positive decision (to do D), proposes various arguments for doing D while B, when opposing, proposes counter-arguments.

There exist 3 tactics for A in our model which are connected with 3 reasoning procedures (WISH, NEEDED, MUST). By tactics of enticing the reasoning procedure WISH, by tactics of persuading the procedure NEEDED and by tactics of threatening the procedure MUST will be tried to trigger in the partner.

For illustration, let us present a schematic description of the tactics of persuasion, based on the reasoning procedure NEEDED (cf. above).

The general idea underlying this tactic is that A proposes arguments for usefulness of D trying to keep the weight of usefulness for B high enough and the possible negative values of other aspects brought out by B low enough so that the sum of positive and negative aspects of D would bring B to the decision to do D.

```
WHILE B is rejecting AND A is not giving up DO

CASE B's answer of
no resources: present a counter-argument
  in order to point at the possibility
  to gain the resources,
  at the same time showing that the cost
  of gaining these resources is lower than
  the weight of the usefulness of D

much harm: present a counter-argument
  to decrease the value of harmfulness
  in comparison with the weight of usefulness

much unpleasant: present a counter-argument
  in order to downgrade the unpleasant aspects
  of D as compared to the useful aspects of D

D is prohibited and the punishment is great:
  present a counter-argument in order
  to downgrade the weight of punishment
  as compared to the usefulness of D
END CASE

Present an argument to stress the usefulness
of D.
```

4 IMPLEMENTATION

An experimental dialogue system is implemented which in interaction with a user can play the role of both A or B. At the moment the computer operates with semantic representations of linguistic input/output only, the surface linguistic part of interaction is provided in the form of a list of ready-made utterances (sentences in Estonian) which are used both by the computer and user. These sentences are only classified semantically according to their possible functions and contributions in a dialogue. For example, sentences informing about the communicative goal (*The firm offers you to trip to Venice*, here

D = *to trip to Venice*), affirming sentences (*Good, I shall go*), and sentences that can be used as arguments for stressing/downgrading the pleasant/unpleasant/useful etc. aspects of an action (*The nature is very beatiful there, You must pay the travel costs yourself*, etc.). The work on linguistic processor is in progress.

Playing A's role, the computer chooses tactics (of enticing, persuading or threatening) and generates (randomly) a model of the partner, according to which the corresponding reasoning procedure (WISH, NEEDED or MUST) yields a positive decision, i.e. the computer presupposes that the user can be influenced this way. A dialogue begins by an expression of the communicative goal (this is the first utterance r_1). If the user refuses (after his/her reasoning by implementing a normal human reasoning which we are trying to model here) the computer recognizes on the basis of the utterance r_2 the step where the reasoning forked into the "negative branch". Then it determines the aspect of D the weight of which does not match the reality, and changes this weight in the user model so that a new model will give a negative result as before but it is an extreme case: if we increased this weight (in case of positive aspects of D) or decreased it (in case of negative ones) we should get a positive decision. In current implementation, each argument will change the corresponding weight exactly by one unit. On the basis of a valid reasoning procedure (tactics) the computer chooses a (counter-)argument r_3 from the set of sentences for increasing/decreasing this weight in the partner model by 1. A reasoning procedure based on the new model will yield a positive decision. Now the user must choose his/her utterance (argument), and the process can continue in a similar way. Every argument can be used only once (cf. the following example; A – computer, B – user).

A: *Do you like to travel to Venice and conclude a contract there?*

B: *I don't have enough experience.*

The user said that there are not enough resources for D: the value of $w(resources)$ was incorrect in the user model.

A: *You can press the right button at the right moment.*

The computer supposes that after its reply the value of w(resources) will be correct.

B: *I can't see any use of this trip.*

The user indicated little usefulness of the action. Thus, the weight $w(use)$ must be corrected in the user model. The computer chooses the tactic of threatening.

A: *The boss knows a lot about you, and he can use it against you.*

B: *This trip is too intense.*

The user pointed out the unpleasantness of D. The computer corrects the value of $w(unpleas)$ in the user model.

A: *There are many people interested in getting your job.*

B: *Well, I'll go.*

5 CONCLUSION

As we have so far mostly dealt with agreement negotiation dialogues, we have planned as one of the practical applications of the system as a participant in communication training sessions. Here the system can establish certain restrictions on argument types, on the order in the use of arguments and counter-arguments, etc.

At present our implementation represents just a prototype realisation of our theoretical ideas and we are working on refining it. In addition, in the present model the partners represent simply certain "abstract" people. There would be very interesting possibilities to take into account the (predetermined) roles of the partners (e.g. chief - subordinate) or even to bring in certain personality traits of the communicating partners.

ACKNOWLEDGEMENTS

We would like to thank the Estonian Science Foundation for supporting this work (grants 5685 and 5534).

REFERENCES

[1] J. Allen, *Natural Language Understanding*, The Benjamin/Cummings Publ. Comp. Inc. , 2nd edn. , 1994.

[2] G. Boella. and L. van der Torre, 'BDI and BOID Argumentation', *Proc. of CMNA-03* http://www.computing.dundee.ac.uk/staff/-creed/research/previous/cmna/finals/boella-final.pdf (used 29/04/2004)

[3] M. Davies and T. Stone, *Folk psychology: the theory of mind debate*, Blackwell, Oxford, Cambridge, Massachusetts, 1995.

[4] K. Jokinen, 'Rational Agency: Concepts, Theories, Models, and Applications', *Proc. of the AAAI Fall Symposium*. Ed. M. Fehling, MIT, Boston, 89–93, (1996).

[5] K. Jokinen, 'Cooperative Response Planning in CDM: Reasoning about Communicative Strategies', *TWLT11. Dialogue Management in Natural Language Systems*. Ed. S. LuperFoy, A. Nijholt G. Veldhuijzen van Zanten, Universiteit Twente, Enschede, 159–168, (1996).

[6] M. Koit and H. Õim, 'Dialogue management in the agreement negotiation process: a model that involves natural reasoning', *The 1st SIGdial Workshop on Discourse and Dialogue*. Ed. L. Dybkjaer, K. Hasida, D. Traum. HongKong, 102–111, (2000).

[7] M. Koit and H. Õim, 'Reasoning in interaction: a model of dialogue', *TALN 2000. 7th Conference on Automatic Natural Language Processing*. Ed. E. Wehrli. Lausanne, Switzerland, 217–224, (2000).

[8] M. Minsky, 'A Framework for Representing Knowledge', *The Psychology of Computer Vision*. Ed. P.H. Winston, New York, 211–277, (1975)

[9] H. Õim and M. Saluveer. 'Frames in linguistic descriptions'. *Quaderni de Semantica* , 6(2), 295–305, (1985)

[10] T. Yuan and D. Moore and Alec Grierson, 'Human-Computer Debate: a Computational Dialectics Approach', *Proc. of CMNA-02*. http://www.csc.liv.ac.uk/ floriana/CMNA/YuanMooreGrierson.pdf (used 29/04/2004)

[11] B. Webber, 'Computational Perspectives on Discourse and Dialogue', *The Handbook of Discourse Analysis*. Ed. D. Schiffrin, D. Tannen, H. Hamilton, Blackwell Publishers Ltd., 798–816, (2001).