Dynamic Trust in Dialogues

Gideon Ogunniye
Nir Oren and Timothy J. Norman

Department of Computing Science
University of Aberdeen

Outlines

Multi-agent Dialogues

- Within Multi-agent dialogues, participants exchange information and make decisions aimed at reaching some conclusion.

Roles of Argumentation

- Formal dialogical argumentation proposes dialogical structures to model the connectedness of utterances.
- A dialogical system consists of the following.
  1. A *set of possible moves* encoded through speech acts e.g (claim(a), retract(a), assert(a), challenge(a), etc).
  2. *Commitment stores* tracking the different propositions and arguments to which players subscribe.
  3. *Protocol rules* : regulate the set of legal moves that are permitted at each stage of a dialogue.
  4. Often, a *Logical language* is used to construct locutions.
  5. *Argumentation-based decision model* to determine justified arguments.

The Problem

Problems :

- Dialogue participants have partial information and individual preferences
- Available information pervaded with uncertainty

Approaches :

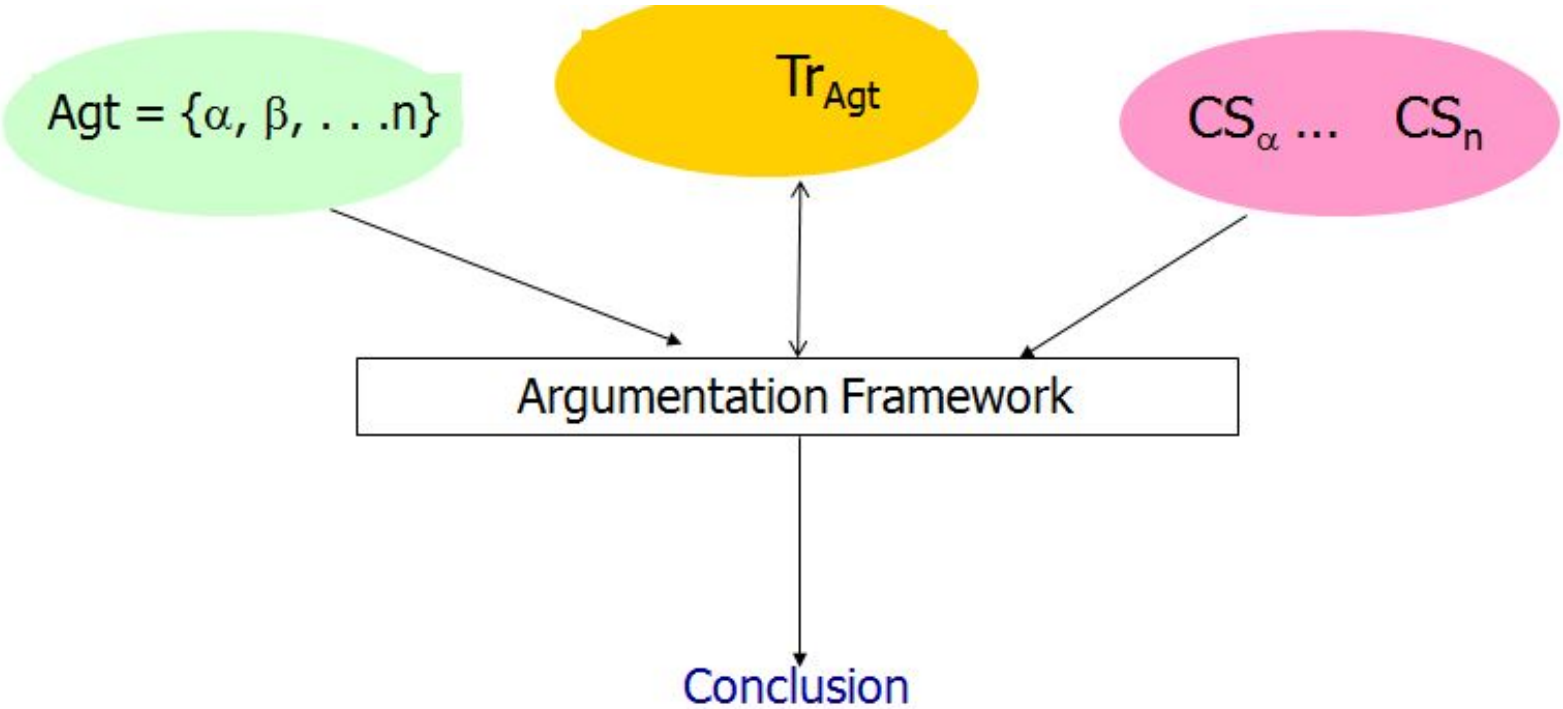
- Paglieri et al (2014) considered how trust and reputation of participants should be updated following the justified conclusions of a dialogue.
- We argue that trust in a participant can change (increase/decrease) during a dialogue. In turn, such trust should affect the conclusion of the dialogue.
- To address this, we need to formalise a dialogue system incorporating trust, and investigate its properties.

Modelling Participants

We consider a system where :

- Participants are modelled through their *commitment stores* $CS_1 \cup ... \cup CS_n \in$ A.
- There is a universal commitment store, UCS = $\cup_\alpha CS_\alpha$.
- The dialogue system consist of series of *add* and *retract* moves.(e.g., add (a, α))denotes that α adds an argument *a* to its commitment store.

The Process



The Notion of Trust

- Trust is encoded as preference ordering over dialogue participants denoted as $\succeq$.
- Arguments from more trusted sources cannot be defeated by arguments from less trusted sources.

Some Observations

- Idea : What utterances/behaviours of a dialogue participant should be penalised or rewarded ?
- Self Contradiction : A player [1] cannot contradicts or challenge its own *commitments* otherwise it looses some trust rating in a dialogue.
- Lack of Justification : A player who is unable to justify arguments in its commitment store should be less trusted.
- A player who regularly retracts arguments should be less trusted.

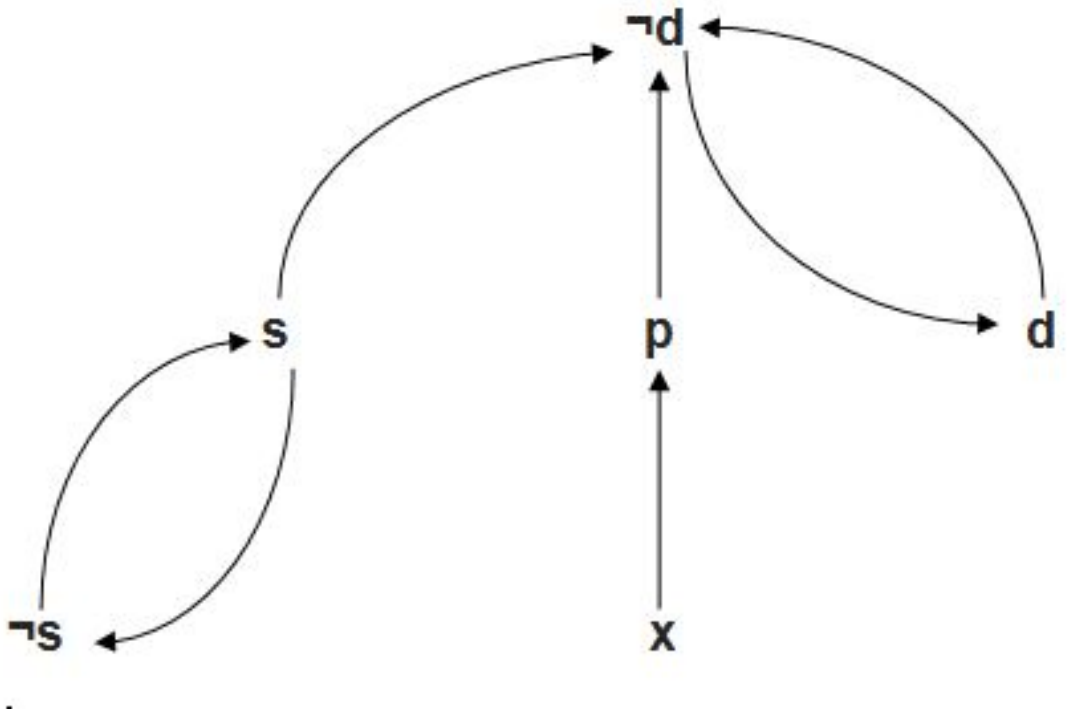
1. Note a player refers to a participant who plays a move

Computing Trust

- At any stage of the dialogue, we may compute SC_α, $LJ\alpha$ and AR_α for every agent.
- Where SC_α, $LJ\alpha$ and AR_α represent number of contradicting, unjustified and retracted arguments in CS_α respectively and,
- Trust Function $Tr : \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{R}$.

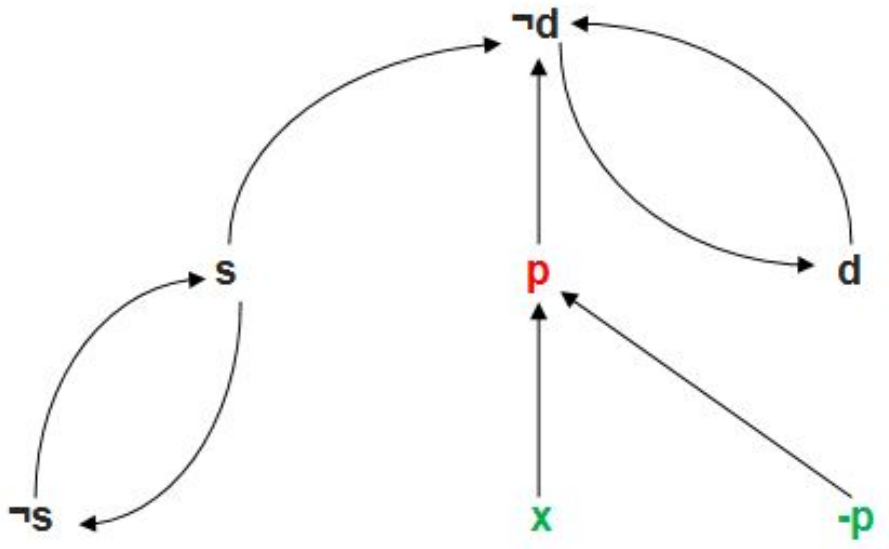
Example

How can we compute extension in this dialogue ?



$CS_\alpha = \{ \text{d, p, s} \}$
$CS_\beta = \{\neg \text{d, x, } \neg s\}$
$Tr_\alpha = 0.\ Tr_\beta = 0$

Dynamic Trust Computing



α *retracts* p

$CS_\alpha = \{$ d, s $\}$

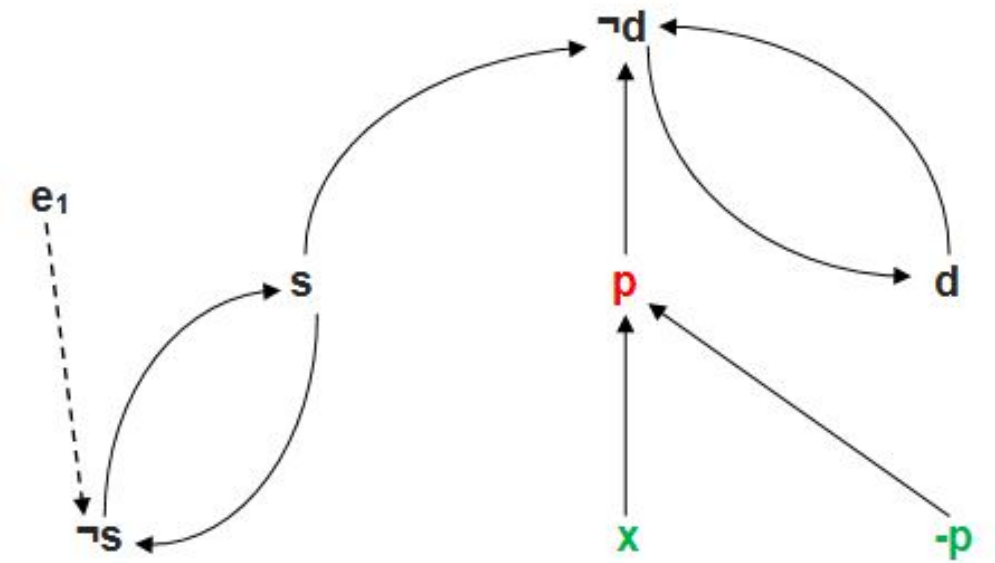
$CS_\beta = \{\neg$d, x, $\neg s\}$

Tr_α = -1. Tr_β = 0

defeat = *attack* + *preference relation* over participating agents.

Evidence

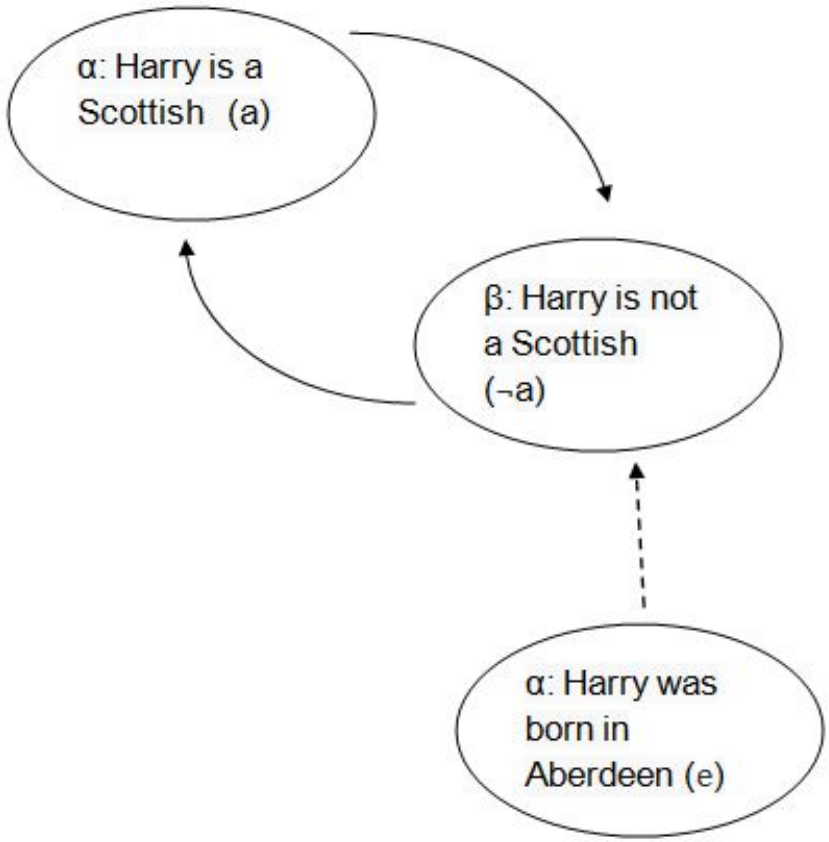
The less trusted participant must supply evidence(s) to back up its claim(s)



$CS_\alpha = \{$ d, s, e_1 $\}$. $CS_\beta = \{\neg$d, x, $\neg s\}$
Tr_α = -1. Tr_β = 0

Criteria for Good Evidence

Is evidence *e* relevant in this dialogue ?



Criteria for Strong Evidence

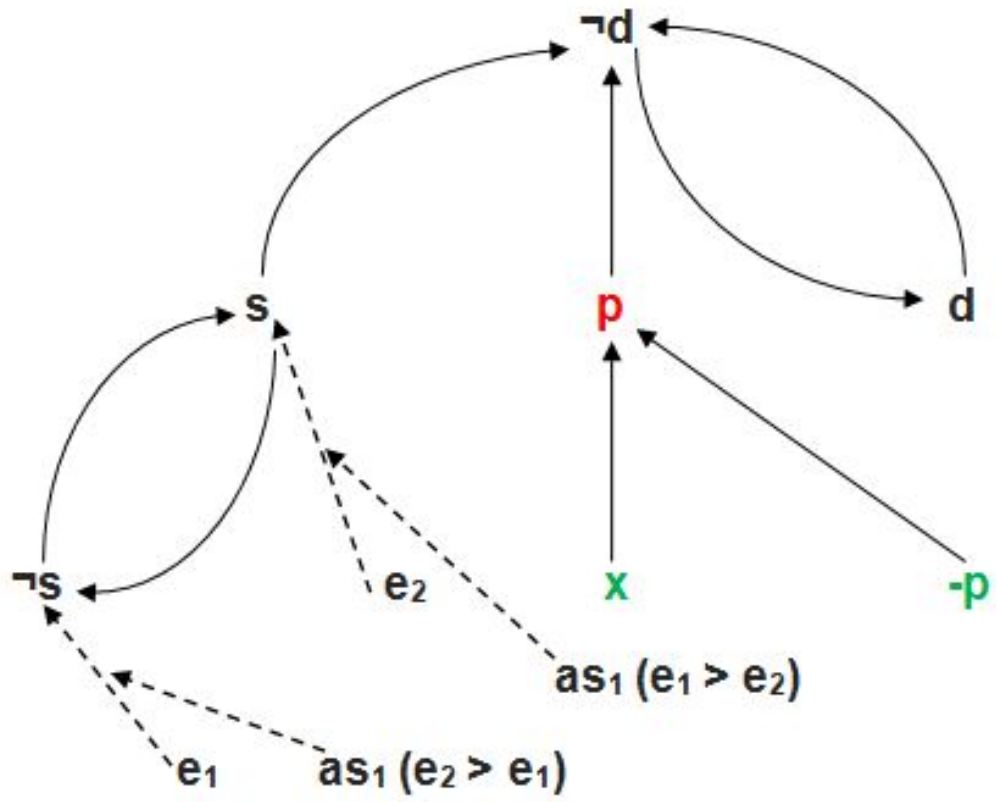
We consider two criteria for good evidence :

- Evidence must be credible (i.e it is (or very likely to be accepted) by all the parties in the dialogue to be true).
- Evidence must be relevant (i.e it makes the claim it supports probable enough).

Argument schemes as_n are used to reason about relevance of evidence.

A Possible Scenario

We are currently investigating this scenario :



$CS_\alpha = \{ \text{d}, \text{s}, e_1 \}$, $CS_\beta = \{ \neg \text{d}, \text{x}, \neg \text{s}, e_2 \}$.

Conclusions

We have described :

- A system where arguments advanced or retracted by dialogue participants affects the trust placed in them.
- How trust in turn affects participants' arguments.
- Three factors that modify trust and how extensions can be computed within a dialogue.

Future works

- Investigate under what conditions is the proposed system stable.
- Formalise argument schemes for reasoning about evidence in dialogues.
- Implement a realistic trust model for argumentative dialogues.
- Implement a complete system and evaluate its impact on argumentative dialogues.

References

[1] 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) : 321-357.

[2] Paglieri, F. ; Castlefranchi, C. ; da Costa Pereira, C. ; Falcone, R. ; Tettamanzi, A. ; and Villata, S. 2014. Trusting the messenger because of the message : feedback dynamics from information quality to source evaluation. *Computational and Mathematical Organisational Theory* 20(2) 176-194.

Thank You