Learning to manage risk in non-cooperative dialogues

Ioannis Efstathiou
Interaction Lab
Heriot-Watt University
ie24@hw.ac.uk

Oliver Lemon
Interaction Lab
Heriot-Watt University
o.lemon@hw.ac.uk

Abstract

We investigate statistical dialogue agents which learn to perform non-cooperative dialogue moves in order to complete their own objectives in a stochastic trading game. We show that, when given the ability to perform both cooperative and non-cooperative dialogue moves, such an agent can learn to bluff and to lie so as to win games more often – against a variety of adversaries, and under various conditions such as risking penalties for being caught in deception. Here we present new results showing how learned non-cooperative dialogue strategies change depending on a) how severe the penalty is for being caught being non-cooperative, and b) how risky the non-cooperative behaviour is (i.e. the probability of being caught). For example, we show that a non-cooperative dialogue agent can learn to win an additional 4.5% of games against a strong rule-based adversary, even when there is an additional 10% chance of being caught (exposed) every time it attempts a non-cooperative (manipulative) move, when the penalty for being caught is that the adversary will no longer trade.

1 Introduction

Non-cooperative dialogues, where an agent may act to satisfy its own goals rather than those of other participants, are of practical and theoretical interest (Georgila and Traum, 2011), and the game-theoretic underpinnings of non-Gricean behaviour are actively being investigated (Asher and Lascarides, 2008). For example, it may be advantageous for an automated agent not to be fully cooperative when trying to gather information from a human, and when trying to persuade, argue, or debate, when trying to sell them something, when trying to detect illegal activity (for example on internet chat sites), or in the area of believable characters in video games and educational simulations (Georgila and Traum, 2011; Shim and Arkin, 2013). Another arena in which non-cooperative dialogue behaviour is desirable is in negotiation (Traum, 2008; Nouri and Traum, 2014), where hiding information (and even outright lying) can be advantageous. Indeed, Dennett argues that deception capability is required for higher-order intentionality in AI (Dennett, 1997).

A complementary research direction in recent years has been the use of machine learning methods to automatically optimise *cooperative* dialogue management - i.e. the decision of what dialogue move to make next in a conversation, in order to maximise an agent's overall long-term expected utility, which is usually defined in terms of meeting a user's goals (Young et al., 2010; Rieser and Lemon, 2011). This research has shown how robust and efficient dialogue management strategies can be learned from data, but has only addressed the case of cooperative dialogue. These approaches use Reinforcement Learning with a reward function that gives positive feedback to the agent only when it meets the user's goals.

An example of the type of non-cooperative dialogue behaviour which we are generating in this work is given by agent B in the following dialogue:
A: "I will give you a sheep if you give me a wheat"
B: "No"
B: "I really need rock" [B actually needs wheat]
A: "OK... I'll give you a wheat if you give me rock"
B: "OK"

Here, A is deceived into providing the wheat that B actually needs, because A believes that B needs rock rather than wheat. Similar behaviour can be observed in trading games such as Settlers

Exp.	Learning Agent policy	Adversary policy	LA win	Adversary win
	Random	Baseline	32%	66%
a	SARSA	Baseline	49.5%	45.555%
b	SARSA + Manipulation	Baseline+Gullible	59.17%*	39.755%
1.1	SARSA+Manipulation	Basel.+ Gull.+Expos(10%).(no trade)	50.86%*	46.33%
1.2	SARSA+Manipulation	Basel.+ Gull.+Expos(5%).(no trade)	51.785%*	45.595%
2	SARSA+Manipulation	Basel.+ Gull.+Expos(10%).(win game)	49.7%	46.225%

Table 1: *Performance (% wins) in testing games (*= significant improvement over baseline, $p < 0.05$)*

of Catan (Afantenos et al., 2012).

1.1 Non-cooperative dialogue and implicature

Our trading dialogues are linguistically cooperative (based on the Cooperative Principle (Grice, 1975)) since their linguistic meaning is clear from both sides and successful information exchange occurs. Non-linguistically though they are non-cooperative, since they they aim for personal goals. Hence they violate Attardo's Perlocutionary Cooperative Principle (PCP) (Attardo, 1997).

In our non-cooperative environment, the manipulative utterances such as "I really need sheep" can imply that "I don't really need any of the other two resources", as both of the players are fully aware that three different resources exist in total and more than one is needed to win the game, so therefore they serve as scalar implicatures (Vogel et al., 2013). We have previously shown that the LA learns how to include scalar implicatures in its dialogue to successfully deceive its adversary by being cooperative on the locutionary level and non-cooperative on the perlocutionary level (Efstathiou and Lemon, 2014).

2 The Trading Game

To investigate non-cooperative dialogues in a controlled setting we created a 2-player, sequential, non-zero-sum game with imperfect information called "Taikun", between a Learning Agent (LA) and an adversary. See (Efstathiou and Lemon, 2014) for details.

Trade occurs through trading proposals that may lead to acceptance from the other player. In an agent's turn only one '1-for-1' trading proposal may occur for each resource, or nothing . Agents respond by either saying "No" or "OK" in order to reject or accept the other agent's proposal. Three manipulative actions are added to the learning agent's set of actions, of the form "I really need X" where X is a resource type. The adversary might believe such statements, resulting in modifying their probabilities of making certain trades.

2.1 Risk of exposure: Experiment 1

In this case when the Learning Agent (LA) is exposed by the adversary then the latter *does not trade* for the rest of the game. We have explored two different cases, one with a 10% chance of exposure (1.1) which gradually increases to 100% at the 10th attempt and another one (1.2) with a chance of 5%, increasing to 100% at the 20th attempt. See table 1. The results show that the LA managed to locate a successful strategy that balances the use of the manipulative actions and the normal trading actions with the risk of exposure.

2.2 Risk of exposure: Experiment 2

In this case if the LA becomes exposed by the adversary then it *loses the game*. Here we also have a 10% chance of exposure which gradually increases to 100% at the 10th attempt. See table 1. The LA learned a strategy that is similar to that of our baseline case, and it never uses manipulative actions since they are now so dangerous.

3 Conclusion & Future Work

In our previous work (Efstathiou and Lemon, 2014) we showed that a statistical dialogue agent can learn to perform non-cooperative dialogue moves in order to enhance its performance in trading negotiations. In this paper, we show that the agent can further learn how to successfully perform such moves in environments where the risk of the deception's exposure is high and the cost means either rejection of all future trades or even an instant win. Alternative methods will also be considered such as adversarial belief modelling with the application of interactive POMDPs (Partially Observable Markov Decision Processes) (Gmytrasiewicz and Doshi, 2005).

References

Stergos Afantenos, Nicholas Asher, Farah Benamara, Anais Cadilhac, Cedric Degremont, Pascal Denis, Markus Guhe, Simon Keizer, Alex Lascarides, Oliver Lemon, Philippe Muller, Soumya Paul, Verena Rieser, and Laure Vieu. 2012. Developing a corpus of strategic conversation in The Settlers of Catan. In *Proceedings of SemDial 2012*.

N. Asher and A. Lascarides. 2008. Commitments, beliefs and intentions in dialogue. In *Proc. of SemDial*, pages 35–42.

S. Attardo. 1997. Locutionary and perlocutionary cooperation: The perlocutionary cooperative principle. *Journal of Pragmatics*, 27(6):753–779.

Daniel Dennett. 1997. When Hal Kills, Who's to Blame? Computer Ethics. In *Hal's Legacy:2001's Computer as Dream and Reality*.

Ioannis Efstathiou and Oliver Lemon. 2014. Learning non-cooperative dialogue strategies. In *Proceedings of SIGDIAL 2014*.

Kallirroi Georgila and David Traum. 2011. Reinforcement learning of argumentation dialogue policies in negotiation. In *INTERSPEECH*.

Piotr J. Gmytrasiewicz and Prashant Doshi. 2005. A framework for sequential planning in multi-agent settings. *Journal of Artificial Intelligence Research*, 24:49–79.

Paul Grice. 1975. Logic and conversation. *Syntax and Semantics*, 3.

Elnaz Nouri and David Traum. 2014. Initiative taking in negotiation. In *Proceedings of SIGDIAL 2014*.

Verena Rieser and Oliver Lemon. 2011. *Reinforcement Learning for Adaptive Dialogue Systems: A Data-driven Methodology for Dialogue Management and Natural Language Generation*. Theory and Applications of Natural Language Processing. Springer.

J. Shim and R.C. Arkin. 2013. A Taxonomy of Robot Deception and its Benefits in HRI. In *Proc. IEEE Systems, Man, and Cybernetics*.

David Traum. 2008. Computational models of non-cooperative dialogue. In *Proc. of SIGdial Workshop on Discourse and Dialogue*.

Adam Vogel, Max Bodoia, Christopher Potts, and Dan Jurafsky. 2013. Emergence of gricean maxims from multi-agent decision theory. In *Proceedings of NAACL 2013*.

Steve Young, M. Gasic, S. Keizer, F. Mairesse, J. Schatzmann, B. Thomson, and K. Yu. 2010. The Hidden Information State Model: a practical framework for POMDP-based spoken dialogue management. *Computer Speech and Language*, 24(2):150–174.