Opponent Modelling for Optimising Strategic Dialogue

Verena Rieser, Oliver Lemon, and Simon Keizer
School of Mathematical and Computer Sciences (MACS)
Heriot-Watt University
Edinburgh EH14 4AS, UK.
{v.t.rieser, o.lemon, s.keizer}@hw.ac.uk

Abstract

Within the wider context of the STAC project, we are developing new models of non-cooperative strategic conversation. We concentrate on learning optimised negotiation strategies (such as deception and information hiding) from real data collected in the domain of "Settlers of Catan", a multi-player board game. This paper illustrates how multi-agent reinforcement learning techniques can be used to model strategic dialogue behaviour. In particular, we discuss novel probabilistic models, called "interactive POMDPs", which combine game theoretic opponent modelling with Partially Observable Markov Decision Processes.

1 Introduction

Within the wider context of the STAC project (2012-2017) we are developing models of non-cooperative strategic conversation[1]. While other partners explore the linguistic and game-theoretic underpinnings of non-Gricean behaviour (Asher and Lascarides, 2008), we focus on learning negotiative dialogue strategies from real data.

The STAC project is collecting data on human trading strategies while playing a modified online version of the board game "Settlers of Catan" (Thomas and Hammond, 2002) where players negotiate trades via a chat interface (Guhe and Lascarides, 2012).

In the following we illustrate how multi-agent reinforcement learning (RL) can be used to optimise strategic trading actions such as deception and information hiding. Previous work has explored single-agent RL for negotiation strategies (Georgila and Traum, 2011; Heeman, 2009), using very limited amounts of data and limited strategic reasoning.

2 Opponent Modelling for Strategic Trading

Single-agent RL approaches were successfully applied to handle uncertainty in Spoken Dialogue Systems, see e.g. (Rieser and Lemon, 2011). However, when considering non-cooperative bargaining domains such as resource negotiation in Settlers, a new type of uncertainty has to be modelled: agents can lie, deceive, bluff, and hide information (Osborne and Rubinstein, 1990). This type of partial observability falls outside the scope of current Partially Observable Markov Decision Processes (POMDPs) approaches to dialogue (Williams and Young, 2007), which focus on uncertainty derived from speech recognition errors.

Examples from an initial data collection (Guhe and Lascarides, 2012) show that human Settlers players employ elaborate strategic conversational moves: On the one hand, players deflect by providing misleading implicatures (Example 1b), hold back information by not answering a question (1c), or tell explicit lies. On the other hand, seemingly cooperative strategies, such as volunteering information, can be observed (Guhe and Lascarides, 2012). Furthermore, offers as in Example (1a) are also often under-specified or "partial", i.e. instead of explicitly specifying how many resources are offered and how many are needed, this information is only revealed strategically in the course of the dialogue.

(1) a. A: Do you have rock?
b. B: I've got lots of wheat [in fact, B has a rock]
c. C: [*silence*]

[1] http://www.irit.fr/STAC/

In order to account for this type of strategic dialogue behaviour, we are exploring novel probabilistic models which combine game-theoretic and POMDP control strategies. In game theory, the process of inferring strategies of other players is also known as "k-level thinking" or opponent modelling (Leyton-Brown and Shoham, 2008). The RL community has adapted these ideas for multi-agent adversarial learning using minimax Q-learning (Littman, 1994) or interactive Partially Observable Markov Decision Processes (iPOMDPs) (Gmytrasiewicz and Doshi, 2005). We extend iPOMDPS for extensive-form games with sequential actions, see (2).

$$I\text{−}POMDP =< IS_i, A_i, T_i, \Omega_i, O_i, R_i > \quad (2)$$

$IS_i = S \times M_j$ is a set of interactive states, where S is the set of states of the physical environment, and $\{M_{j...m}\}$ is the set of possible models of agents $j...m$. $\{A_i\}$ describes agent i's set of actions. $T_i : IS \times A_i \times IS \rightarrow [0, 1]$ is a transition function which describes results of an action. Ω_i is the set of observations the agent i can make. $O_i : IS \times A_i \times \Omega_i \rightarrow [0, 1]$ is the agent's observation function which specifies probabilities of observations given agents' actions and resulting states. Finally, $R_i : S \times A_i \rightarrow R$ is the reward function representing agent i's preferences.

By formulating an "interactive state" which includes explicit possible behavioural models of other agents, iPOMDPs recognise that agents are not playing against distributions like in single-agent RL, "but other players who understood the rules and were prepared to leverage them against slower players" (Wunder et al., 2011).

3 Action Set for Learning

We will employ iPOMDPs to model and reason about hidden states and strategic conversational behaviour of other players. In particular, we aim to learn optimal behaviour for the following decisions:

1. How to make a strategic offer and how much information do I expose?
2. How do I reply to an offer and how sincere is my reply?

For learning the first decision, we modify the JSettlers system, where an artificial trading agent can only pose a fully specified offer ("*[A], I'll give you 1 wheat for 1 sheep.*") via a graphical interface. We will handle partial offers (see Section 2) as well as disjunctive offers and requests (e.g. "wheat or sheep").

For optimising the latter decision, STAC has developed an annotation scheme which distinguishes between observable replies and their sincerity based on the logged game state. We plan to evaluate the success of these strategic dialogue capabilities against the original graphical version.

Acknowledgments

The research leading to this work has received funding from the European Community's FP7 programme (FP7/2011-16) under grant agreement no. 269427 (STAC). http://www.irit.fr/STAC/

References

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

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

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

Guhe, Markus and Alex Lascarides. 2012. Trading in a multiplayer board game: Towards an analysis of non-cooperative dialogue. In *Proc. of CogSci*.

Heeman, Peter. 2009. Representing the reinforcement learning state in a negotiation dialogue. In *Proc. of ASRU*.

Leyton-Brown, Kevin and Yoav Shoham. 2008. *Essentials of Game Theory: A Concise, Multidisciplinary Introduction*. Morgan & Claypool Publishers.

Littman, Michael L. 1994. Markov games as a framework for multi-agent reinforcement learning. In *Proc. ICML*, pages 157–163.

Osborne, Martin J. and Ariel Rubinstein. 1990. *Bargaining and markets*. Academic Press.

Rieser, Verena and Oliver Lemon. 2011. *Reinforcement Learning for Adaptive Dialogue Systems*. Theory and Applications of Natural Language Processing. Springer.

Thomas, R. and K. Hammond. 2002. Java settlers: a research environment for studying multi-agent negotiation. In *Proc. of IUI '02*, pages 240–240.

Williams, J. and S. Young. 2007. Partially Observable Markov Decision Processes for Spoken Dialog Systems. *Computer Speech and Language*, 21(2):231–422.

Wunder, Michael, Michael Kaisers, Michael Littman, and John Robert Yaros. 2011. Using iterated reasoning to predict opponent strategies. In *The 10th International Conference on Autonomous Agents and Multiagent Systems*.