Evaluating Persuasion Strategies and Deep Reinforcement Learning methods for Negotiation Dialogue agents

Simon Keizer[1], Markus Guhe[2], Heriberto Cuayáhuitl[3],
Ioannis Efstathiou[1], Klaus-Peter Engelbrecht[1], Mihai Dobre[2], Alex Lascarides[2] and Oliver Lemon[1]

[1]Department of Computer Science, Heriot-Watt University
[2]School of Informatics, University of Edinburgh
[3]School of Computer Science, University of Lincoln
[1]{s.keizer,ie24,o.lemon}@hw.ac.uk
[2]{m.guhe,m.s.dobre,alex}@inf.ed.ac.uk
[3]hcuayahuitl@lincoln.ac.uk

Abstract

In this paper we present a comparative evaluation of various negotiation strategies within an online version of the game "Settlers of Catan". The comparison is based on human subjects playing games against artificial game-playing agents ('bots') which implement different negotiation dialogue strategies, using a chat dialogue interface to negotiate trades. Our results suggest that a negotiation strategy that uses persuasion, as well as a strategy that is trained from data using Deep Reinforcement Learning, both lead to an improved win rate against humans, compared to previous rule-based and supervised learning baseline dialogue negotiators.

1 Introduction

In dialogues where the participants have conflicting preferences over the outcome, Gricean maxims of conversation break down (Asher and Lascarides, 2013). In this paper we focus on a non-cooperative scenario – a win-lose board game – in which one of the components of the game involves participants negotiating trades over restricted resources. They have an incentive to agree trades, because alternative means for getting resources are more costly. But since each player wants to win (and so wants the others to lose), they not only make offers and respond to them, but also bluff, persuade, and deceive to get the best deal for themselves at perhaps a significant cost to others (Afantenos et al., 2012).

In recent work, computational models for non-cooperative dialogue have been developed (Traum, 2008; Asher and Lascarides, 2013; Guhe and Lascarides, 2014a). Moreover, machine learning techniques have been used to train negotiation strategies from data, in particular reinforcement learning (RL) (Georgila and Traum, 2011; Efstathiou and Lemon, 2015; Keizer et al., 2015). In particular, it has been shown that RL dialogue agents can be trained to strategically select offers in trading dialogues (Keizer et al., 2015; Cuayahuitl et al., 2015c), but also to bluff and lie (Efstathiou and Lemon, 2015; Efstathiou and Lemon, 2014).

This paper presents an evaluation of 5 variants of a conversational agent engaging in trade negotiation dialogues with humans. The experiment is carried out using an online version of the game "Settlers of Catan", where human subjects play games against artificial players, using a Natural Language chat interface to negotiate trades. Our results suggest that a negotiation strategy using persuasion (Guhe and Lascarides, 2014b) when making offers, as well as a strategy for selecting offers that is trained from data using Deep Reinforcement Learning (Cuayahuitl et al., 2015c), both lead to improved win rates against humans, compared to previous rule-based approaches and a model trained from a corpus of humans playing the game using supervised learning.

2 Task domain

"Settlers of Catan" is a complex multi-player board game[1]; the board is a map consisting of hexes of different types: hills, mountains, meadows, fields and forests. The objective of the game is for the players to build roads, settlements and cities on the map, paid for by combinations of re-

[1] See www.catan.com for the full set of game rules.

sources of five different types: clay, ore, sheep, wheat and wood, which are obtained according to the numbers on the hexes adjacent to which a player has a settlement or city after the roll of a pair of dice at each player's turn. In addition, players can negotiate trades with each other in order to obtain the resources they desire. Players can also buy Development Cards, randomly drawn from a stack of different kinds of cards. Players earn Victory Points (VPs) for their settlements (1 VP each) and cities (2 VPs each), and for having the Longest Road (at least 5 consecutive roads; 2 VPs) or the Largest Army (by playing at least 3 Knight development cards; 2 VPs). The first player to have 10 VPs wins the game.

2.1 The JSettlers implementation

For testing and evaluating our models for trade negotiation, we use the JSettlers[2] open source implementation of the game (Thomas, 2003). The environment is a client-server system supporting humans and agents playing against each other in any combination. The agents use complex heuristics for the board play—e.g., deciding when, what and where to build on the board—as well as what trades to aim for and how to negotiate for them.

2.2 Human negotiation corpus

With the aim of studying strategic conversations, a corpus of online trading chats between humans playing "Settlers of Catan" was collected (Afantenos et al., 2012). The JSettlers implementation of the game was modified to let players use a chat interface to engage in conversations with each other, involving the negotiation of trades in particular. Table 1 shows an annotated chat between players W, T, and G; in this dialogue, a trade is agreed between W and G, where W gives G a clay in exchange for an ore. For training the data-driven negotiation strategies, 32 annotated games were used, consisting of 2512 trade negotiation dialogue turns.

3 Overview of the artificial players

For all the artificial players ('bots'), we distinguish between their *game playing* strategy (**Game Strategy**) and their *trade negotiation* strategy (**Negot. Strategy**), see Table 2. The game playing strategy involves all non-linguistic moves in the game: e.g., when and where to build a settlement, where to move the robber when a 7 is rolled and who to steal from, and so on. The negotiation strategy, which is triggered when the game playing strategy chooses to attempt to trade with other players (i.e. the trade dialogue phase), involves deciding which offers to make to opponents, and whether to accept or reject offers made by them. This strategy takes as input the resources available to the player, the game board configuration, and a 'build plan' received from the game playing strategy, indicating which piece the bot aims to build (but does not yet have the resources for).

One of the bots included in the experiment uses the **original** game playing strategy from JSettlers (Thomas, 2003), whereas the other 4 bots use an **improved** strategy developed by Guhe and Lascarides (2014a). We distinguish between the following negotiation strategies:

1. the **original** strategy from JSettlers uses hand-crafted rules to filter and rank the list of legal trades;

2. an enhanced version of the original strategy, which includes the additional options of using **persuasion** arguments to accompany a proposed trade offer (rather than simply offering it)—for example "If you accept this trade offer, then you get wheat that you need to immediately build a settlement"—and **hand-crafted rules** for choosing among this expanded set of options (Guhe and Lascarides, 2014a);

3. a strategy which uses a legal trade re-ranking mechanism trained on the human negotiation corpus described in (Afantenos et al., 2012) using supervised learning (**Random Forest**) (Cuayáhuitl et al., 2015a; Cuayáhuitl et al., 2015b; Keizer et al., 2015); and

4. an offer selection strategy that is trained using **Deep Reinforcement Learning**, in which the feature representation and offer selection policy are optimised simultaneously using a fully-connected multilayer neural network. The state space of this agent includes 160 non-binary features that describe the game board and the available resources. The action space includes 70 actions for offering trading negotiations (including up to two giveable resources and only one receivable resource) and 3 actions (accept, reject and counteroffer) for replying to offers from opponents. The reward function is

[2] jsettlers2.sourceforge.net

Speaker	Utterance	Game act	Surface act	Addressee	Resource
W	*can i get an ore?*	Offer	Request	all	Receivable(ore,1)
T	*nope*	Refusal	Assertion	W	
G	*what for.. :D*	Counteroffer	Question	W	
W	*a wheat?*	Offer	Question	G	Givable(wheat,1)
G	*i have a bounty crop*	Refusal	Assertion	W	
W	*how about a wood then?*	Counteroffer	Question	G	Givable(wood,1)
G	*clay or sheep are my primary desires*	Counteroffer	Request	W	Receivable((clay,?) OR (sheep,?))
W	*alright a clay*	Accept	Assertion	G	Givable(clay,1)
G	*ok!*	Accept	Assertion	W	

Table 1: Example trade negotiation chat.

based on victory points—see (Cuayahuitl et al., 2015c) for further details.

4 Experiment

The evaluation was performed as an online experiment. Using the JSettlers environment, an experimental setup was created, consisting of a game client that the participants could download and use to play online games, and a server for running the bot players and logging all the games.

We decided to compare the five bot types described in Section 3 in a between-subjects design, as we expected that playing a game against each of the 5 bot types would take more time than most participants would be willing to spend (about 4 hours) and furthermore would introduce learning effects on the human players that would be difficult to control. Each participant played one game against three bots of the same type. The bot was chosen randomly.

In order to participate, the subjects registered and downloaded the game client. Next, they were asked to first play a short training game to familiarise themselves with the interface (see Fig. 1), followed by a full game to be included in the evaluation. The training game finishes when the subject reaches 3 VPs, i.e., when they have built at least one road and one settlement in addition to the two roads and two settlements (making 2 VPs) each player starts with. Although subjects were allowed to play more games after they completed their full game, we only used their first full game in the evaluation to avoid bias in the data through learning effects.

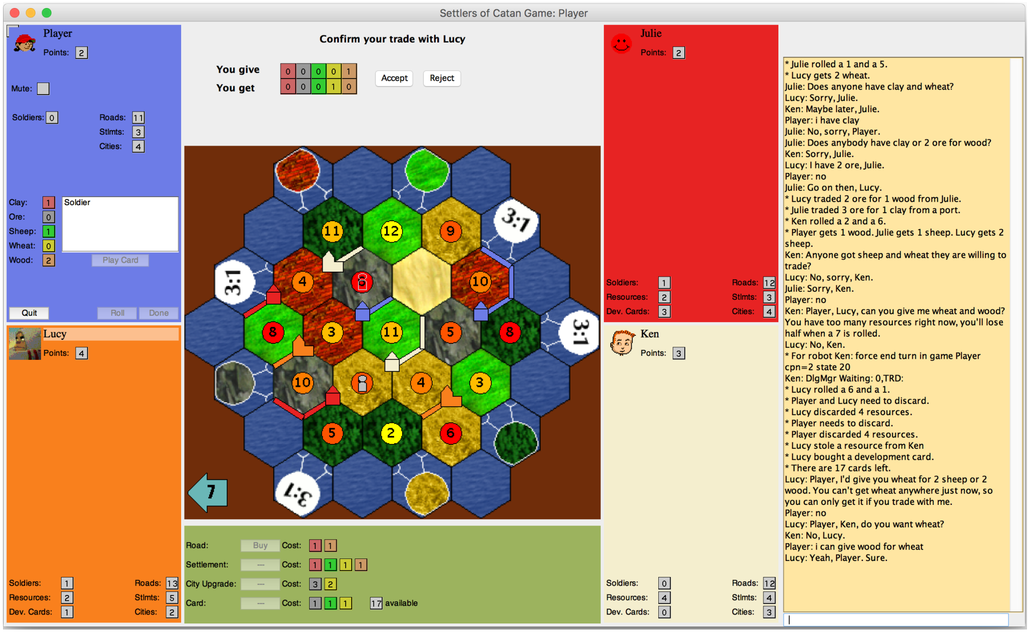
We advertised the experiment online through university mailing lists, twitter, and "Settlers of Catan" forums. We also hung out posters at the university and in a local board gaming pub. We particularly asked for experienced Settlers players, who had played the game at least three times before, since the game is quite complex, and we expected that data from novice players would be too noisy to reveal any differences between the different bot types. Each subject received a £10 Amazon UK voucher after completing both training and full game, and we included two prize draws of £50 vouchers to further encourage participation.

5 Results

After running the experiments for 16 weeks, we collected 212 full games in total (including the training ones), but after only including the first full game from each subject (73 games/subjects), and removing games in which the subject did not engage in any trade negotiations, we ended up with 62 games.

The evaluation results are presented in Table 2 and Fig. 2, which show how the human subjects fared playing against our different bots: the numbers of Table 2 refer to the performance of the humans, but of course measure the performance of the bots. Indicated in the table are the percentage of games won by the humans (WinRate, so the lower the WinRate the stronger the bot's performance on the task) and the average number of victory points the humans gained (AvgVPs). Since JSettlers is a four-player game, each human plays against 3 bots, so a win rate of 25% would indicate that the humans and bots are equally good players.

Although the size of the corpus is too small to make any strong claims about the relative strength of the different bots, we are encouraged by the results so far. The results confirm our expectation, based on game simulations in which one agent with the 'improved' game strategy beat 3 original opponents by significantly more than 25% (Guhe and Lascarides, 2014b), that the improved game strategy is superior to the original strategy against



```
* Julie rolled a 1 and a 5.
* Lucy gets 2 wheat.
Julie: Does anyone have clay and wheat?
Lucy: Sorry, Julie.
Ken: Maybe later, Julie.
Player: i have clay
Julie: No, sorry, Player.
Julie: Does anybody have clay or 2 ore for wood?
Ken: Sorry, Julie.
Lucy: I have 2 ore, Julie.
Player: no
Julie: Go on then, Lucy.
* Lucy traded 2 ore for 1 wood from Julie.
* Julie traded 3 ore for 1 clay from a port.
* Ken rolled a 2 and a 6.
* Player gets 1 wood. Julie gets 1 sheep. Lucy gets 2
sheep.
Ken: Anyone got sheep and wheat they are willing to
trade?
Lucy: No, sorry, Ken.
Julie: Sorry, Ken.
Player: no
Ken: Player, Lucy, can you give me wheat and wood?
You have too many resources right now, you'll lose
half when a 7 is rolled.
Lucy: No, Ken.
* For robot Ken: force end turn in game Player
cpn=2 state 20
Ken: DlgMgr Waiting: 0,TRD:
* Lucy rolled a 6 and a 1.
* Player and Lucy need to discard.
* Lucy discarded 4 resources.
* Player needs to discard.
* Player discarded 4 resources.
* Lucy stole a resource from Ken
* Lucy bought a development card.
* There are 17 cards left.
Lucy: Player, I'd give you wheat for 2 sheep or 2
wood. You can't get wheat anywhere just now, so
you can only get it if you trade with me.
Player: no
Lucy: Player, Ken, do you want wheat?
Ken: No, Lucy.
Player: i can give wood for wheat
Lucy: Yeah, Player. Sure.
```

Figure 1: Graphical interface of the adapted online Settlers game-playing client, showing the state of the board itself, and in each corner information about one of the four players, seen from the perspective of the human player sitting at the top left (playing with blue; the other 3 players are bots). The human player is prompted to accept the trade displayed in the top middle part, as agreed in the negotiation chat shown in the panel on the right hand side.

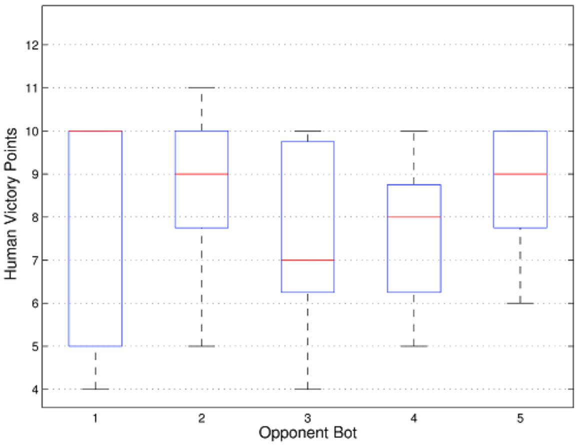


Figure 2: Box plots representing the victory points (VPs) scored by humans against each bot (as shown on Table 2). Humans scored lower against the bots 3 and 4 (i.e. on Table 2 the bots of the 3rd and 4th row respectively). Red line: median VPs.

Game strategy	Negot. strategy	Games	Human WinRate	AvgVPs
1. Orig	Persuasion	10	70.0%	7.8
2. Impr	Original	17	29.4%	8.4
3. Impr	Persuasion	15	26.7%	7.5
4. Impr	DeepRL	11	18.2%	6.5
5. Impr	RandForest	9	44.4%	8.7
Overall		**62**	**37.7%**	**7.8**

Table 2: Results of human subjects playing a game against 3 instances of one of 5 different bot types. **Human WinRate** is the percentage of games won by human players, and **AvgVPs** is the (mean) average number of VPs gained by the human players. If the humans were equally strong as the bots, they would achieve approximately a 25% win rate.

human opponents (70.0% vs. 26.7%). Improving the game strategy is important because negotiation is only a small part of what one must do to win this particular game.

The lowest win rates for humans are achieved when playing against the Deep Reinforcement Learning (DRL) negotiation strategy (18.2%). This confirmed its superiority over the supervised learning bot (RandForest) against which it was trained (18.2% vs. 44.4%, using the same game playing strategy). This confirms previous results in which the DRL achieved a win rate of 41.58% against the supervised learning bot (Cuayahuitl et al., 2015c). Since the win rate is also well below the 25% win rate one expects if the 4 players are of equal strength, the deep learning bot beats the human players on average. As described in Section 3, the DRL bot uses a large set of input features and uses its neural network to automatically learn the patterns that help finding the optimal negotiation strategy. In contrast, human players, even experienced ones, have limited cognitive capacity to adequately oversee game states and make the best trades.

Against the bots using a negotiation strategy with persuasion, the human players achieved lower win rates than against the bot with the original, rule-based negotiation strategy (26.7% vs. 29.4%), and much lower win rates than the bot with the supervised learning strategy (26.7% vs. 44.4%). In terms of average victory points, both persuasion and deep learning bots outperform the rule-based and supervised learning baselines.

6 Conclusion

We evaluated different trading-dialogue strategies (original rule-based/persuasion/random forest/deep RL) and game-playing strategies (original/improved) in online games with experienced human players of "Settlers of Catan". The random forest and deep RL dialogue strategies were trained using human-human game-playing data collected in the STAC project (Afantenos et al., 2012). The results indicate that the improved game strategy of (Guhe and Lascarides, 2014a) is beneficial, and that dialogue strategies using persuasion (Guhe and Lascarides, 2014b) and deep RL (Cuayahuitl et al., 2015c) outperform both the original rule-based strategy (Thomas, 2003) and a strategy created using supervised learning methods (random forest). The deep RL dialogue strategy also outperforms human players, similarly to recent results for other (non-dialogue) games such as "Go" and Atari games (Silver et al., 2016; Mnih et al., 2013). More data is being collected.

Acknowledgements

This research was funded by the European Research Council, grant number 269427, STAC project https://www.irit.fr/STAC/

References

Stergos Afantenos, Nicholas Asher, Farah Benamara, Anaïs Cadilhac, Cédric Dégremont, Pascal Denis, Markus Guhe, Simon Keizer, Alex Lascarides, Oliver Lemon, Philippe Muller, Saumya Paul, Vladimir Popescu, Verena Rieser, and Laure Vieu. 2012. Modelling strategic conversation: model, annotation design and corpus. In *Proc. Workshop on the Semantics and Pragmatics of Dialogue (SemDIAL)*.

Nicholas Asher and Alex Lascarides. 2013. Strategic conversation. *Semantics and Pragmatics*, 6(2):1–62.

Heriberto Cuayáhuitl, Simon Keizer, and Oliver Lemon. 2015a. Learning to trade in strategic board games. In *Proc. IJCAI Workshop on Computer Games (IJCAI-CGW)*.

Heriberto Cuayáhuitl, Simon Keizer, and Oliver Lemon. 2015b. Learning trading negotiations using manually and automatically labelled data. In *Proc. 27th IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*.

Heriberto Cuayahuitl, Simon Keizer, and Oliver Lemon. 2015c. Strategic dialogue management via deep reinforcement learning. In *Proc. NIPS workshop on Deep Reinforcement Learning*.

Ioannis Efstathiou and Oliver Lemon. 2014. Learning non-cooperative dialogue behaviours. In *Proc. Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.

Ioannis Efstathiou and Oliver Lemon. 2015. Learning non-cooperative dialogue policies to beat opponent models: "the good, the bad and the ugly". In *Proc. Workshop on the Semantics and Pragmatics of Dialogue (SemDIAL)*.

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

Markus Guhe and A Lascarides. 2014a. Game strategies for The Settlers of Catan. In *Proc. IEEE Conference on Computational Intelligence and Games (CIG)*.

Markus Guhe and Alex Lascarides. 2014b. Persuasion in complex games. In *Proc. Workshop on the Semantics and Pragmatics of Dialogue (SemDIAL)*.

Simon Keizer, Heriberto Cuayahuitl, and Oliver Lemon. 2015. Learning trade negotiation policies in strategic conversation. In *Proc. Workshop on the Semantics and Pragmatics of Dialogue (SemDIAL)*.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. 2013. Playing atari with deep reinforcement learning. In *Proc. NIPS Deep Learning Workshop*.

David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529.

Robert Shaun Thomas. 2003. *Real-time decision making for adversarial environments using a plan-based heuristic*. Ph.D. thesis, Northwestern University.

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