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

Citation for published version:

Link:
Link to publication record in Heriot-Watt Research Portal

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics

General rights
Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Evaluating Persuasion Strategies and Deep Reinforcement Learning methods for Negotiation Dialogue agents

Simon Keizer\(^1\), Markus Guhe\(^2\), Heriberto Cuayahuitl\(^3\), Ioannis Efstathiou\(^1\), Klaus-Peter Engelbrecht\(^1\), Mihai Dobre\(^2\), Alex Lascarides\(^3\) 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

\(\{s.keizer,ie24,o.lemon\}@hw.ac.uk\)
\(\{m.guhe,m.s.dobre,alex\}@inf.ed.ac.uk\)
\(hcuayahuitl@lincoln.ac.uk\)

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 (Afan tenos 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 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) (Cuayáhuítl et al., 2015a; Cuayáhuítl 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
Table 1: Example trade negotiation chat.

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

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.

<table>
<thead>
<tr>
<th>Game strategy</th>
<th>Negot. strategy</th>
<th>Games</th>
<th>Human WinRate</th>
<th>AvgVPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Orig Persuasion</td>
<td>10</td>
<td>70.0%</td>
<td>7.8</td>
<td></td>
</tr>
<tr>
<td>2. Impr Original</td>
<td>17</td>
<td>29.4%</td>
<td>8.4</td>
<td></td>
</tr>
<tr>
<td>3. Impr Persuasion</td>
<td>15</td>
<td>26.7%</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>4. Impr DeepRL</td>
<td>11</td>
<td>18.2%</td>
<td>6.5</td>
<td></td>
</tr>
<tr>
<td>5. Impr RandForest</td>
<td>9</td>
<td>44.4%</td>
<td>8.7</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>62</td>
<td>37.7%</td>
<td>7.8</td>
<td></td>
</tr>
</tbody>
</table>

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.
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


