Games and Agents - Project Proposal

Katarzyna Michałowska (5951925), Bror Emil Baklund Krogsrud (5978602), Maayan Shvo (5920833), Breixo Soliño Fernández (6063314), Thijs van Zwieten (3388751)

Goals

The goal of this project is to build our own negotiation algorithm, on top of the existing BANDANA framework. BANDANA is an extension of the DipGame framework; however, it provides a new negotiation server and uses a simplified negotiation language. The existing ontology in DipGame (and by extension in BANDANA) covers relationships between players and negotiations. This includes transfer of knowledge (sharing of information or feelings), negotiating a deal, explaining and arguing.

We would like to leverage the ontologies exposed by DipGame and the negotiation framework offered by BANDANA in order to write diplomacy-playing agents that are able to express several aspects of human behavior. Our agents should incorporate notions such as trust (and violation thereof...), long-term planning, decision-making, negotiation and deceit. These agents are to play against each other, while we observe their actions. If possible, we would like to implement a theory-of-mind aspect, wherein an agent reasons about its opponent’s motivations.

The specific modules that we would like to implement are trust and adversity between agents and a reasoning module, which takes those interdependencies into account. Moreover, each agent would have its own personality which would impact the way they negotiate, and features like whether they are trustworthy, if they trust other agents etc.

Eventually, we hope to create an agent that is versatile (able to express different play-styles), capable of negotiation, of lying (speaking false truths) and that displays interesting personality traits. We also hope to evaluate the agents’ performance and interaction with one another (depending on their personalities), both qualitatively (assessing the different agents’ personalities and observing interesting behavior) and quantitatively (comparing the performance of the agents when playing against one another, based on clearly defined metrics of performance).

AI and agent Aspects Needed

The negotiation algorithm will be implemented on top of the existing strategy framework BANDANA - a Java framework for the development of automated agents that play the game of Diplomacy. The python-based Parlance framework will be used to support networked play and to observe games.

Our goal requires a two or three-legged rule-based system. This would serve three functions:

- Goal formulation (long-term and short-term)
  - Could possibly be formulated by specifying game phases.
  - Each game phase would be associated with a predefined goal set.
  - End goal is win condition.
- Action planning
- Reasoning framework
  - Plan recognition / Goal recognition
  - Why is player X doing this?
The reasoning of agents influences parameters in goal-formulation system, resulting in changing goals. This system needs to be populated with rules (and their associated preconditions and effects) that determine the agents’ actions. To that end we need to implement at least three different systems/modules:

- Trust framework (Can I trust this opponent?)
- Adversity framework (Do I like this opponent?)
- Emotions/Personality (Resp. short vs. long-term modifiers to parameters that influence goal selection and thereby actions...)

### Planning

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>May 9</td>
<td>Literature research and project proposal</td>
</tr>
<tr>
<td>May 30</td>
<td>Working prototype</td>
</tr>
<tr>
<td>June 6</td>
<td>Literature presentation</td>
</tr>
<tr>
<td>June 15</td>
<td>Testing and evaluation</td>
</tr>
<tr>
<td>June 22</td>
<td>Project presentation</td>
</tr>
<tr>
<td>June 22</td>
<td>Project final submission</td>
</tr>
</tbody>
</table>

### Working prototype

At this time, we intend to have a working prototype. Agents should be able to perform basic negotiations and model trust and adversity values as a consequence of events over the course of the game. At this time, trust and adversity do not yet have an impact on the actions agents take.

### Testing and evaluation

The success of our implementation can be determined by evaluating agent play styles over the course of several games. Our intent is to model several play-styles - distinguished by (for instance):

- Different degrees of trust and adversity.
- Non-identical negotiation styles
- Quality of proposals
  - Truthful vs. deceitful
  - Generosity: How optimal is the proposal for the receiver?

The frequency with which agents select a truthful over a deceitful message is an example of a quantitative measure we could use to assess style differences between agents.

### Preliminary literature review
