No Press Diplomacy: Modeling Multi-Agent Gameplay

Philip Paquette, Yuchen Lu, Steven Bocco, Max O. Smith, Satya Ortiz-Gagne, Jonathan K. Kummerfeld, Satinder Singh, Joelle Pineau, Aaron Courville

Université de Montréal, University of Michigan, Université McGill, Mila, webDiplomacy

Overview

We present **the first human-competitive system** for the seven-player, non-stochastic game Diplomacy. The game, shown below, requires agents to both **collaborate and compete** in order to win. We:

- Built a dataset of **150,000** games (available, NDA required).
- Developed a policy model for No Press Diplomacy.
- Explored training with **supervised learning** and **self-play**.
- Beat **state-of-the-art** rule-based agents.
- Ran a tournament against **almost 100 humans**.

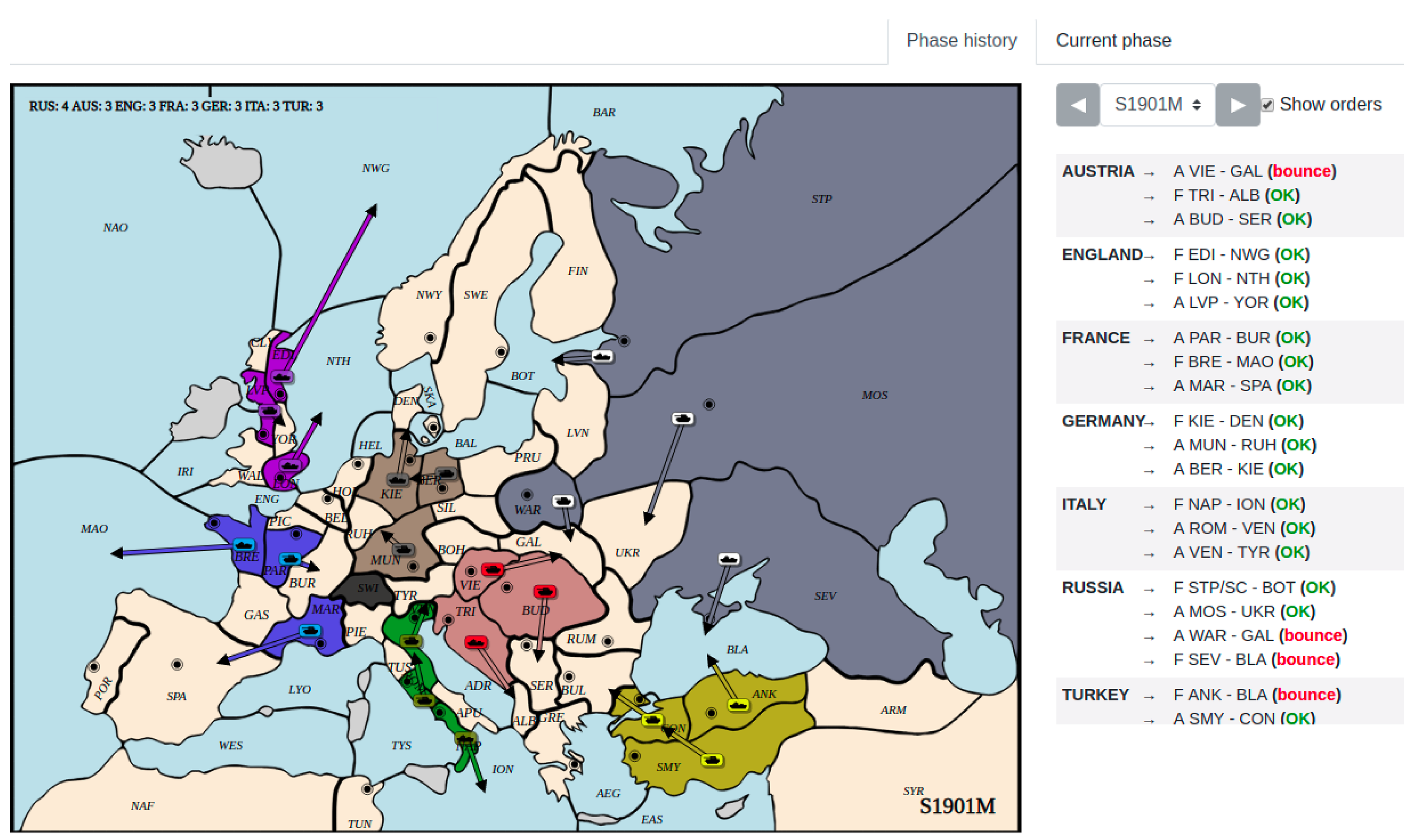

Figure 1: Game Overview

Dataset

We trained using a mix of games with (106k) and without (33k) human communication on the standard map and 16,600 games on other maps.

	Win%	Draw%	Defeated%	Survival rate for opponents AUS	ENG	FRA	GER	ITA	RUS	TUR
Austria	4.3%	33.4%	48.1%	100%	79%	62%	55%	40%	29%	15%
England	4.6%	43.7%	29.1%	47%	100%	30%	16%	49%	33%	80%
France	6.1%	43.8%	25.7%	40%	26%	100%	22%	45%	59%	77%
Germany	5.3%	35.9%	40.4%	44%	26%	39%	100%	61%	27%	80%
Italy	3.6%	36.5%	40.2%	15%	65%	56%	61%	100%	56%	25%
Russia	6.6%	35.2%	39.8%	25%	52%	77%	38%	63%	100%	42%
Turkey	7.2%	43.1%	26.0%	9%	78%	71%	56%	23%	31%	100%
Total	39.9%	60.1%		37%	59%	65%	49%	51%	50%	64%

Table 1: Dataset statistics

Data is available on request by contacting webdipmod@gmail.com.

DipNet: A Generative Model of Unit Orders

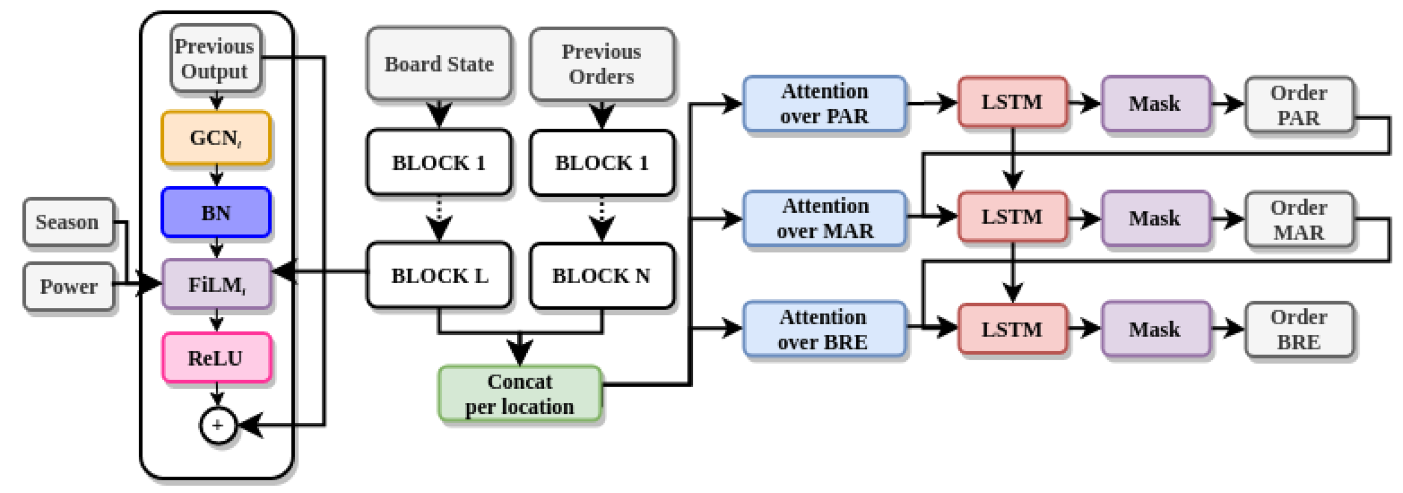

Figure 2: Architecture

We treat each location as a node and use a graph convolution network to process the map. We use conditioned batch normalization for information such as the current power, the season, and locations. Finally, we decode the unit orders per location and mask out impossible orders according to the game rules.

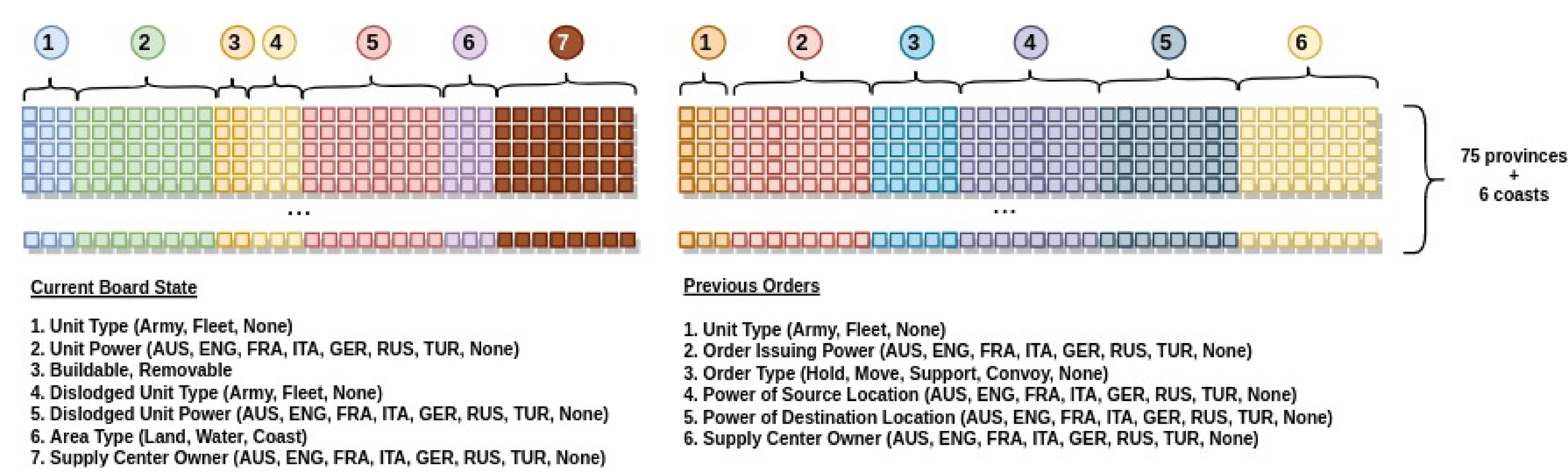

Figure 3: Board Representation

Results: Ablation Study

The best model is able to predict **61.3%** of human orders correctly. Errors are more common in the late game and when there are a larger number of units to provide orders for.

Model	Accuracy per unit-order		Accuracy for all orders	
	Teacher forcing	Greedy	Teacher forcing	Greedy
DipNet	**61.3%**	**47.5%**	**23.5%**	**23.5%**
Untrained	6.6%	6.4%	4.2%	4.2%
Without FiLM	60.7%	47.0%	22.9%	22.9%
Masked Decoder (No Board)	47.8%	26.5%	14.7%	14.7%
Board State Only	60.3%	45.6%	22.9%	23.0%
Average Embedding	59.9%	46.2%	23.2%	23.2%

Table 2: Evaluation of supervised models: Predicting human orders.

	Support Accuracy	
	1st location	16th location
DipNet	**40.3%**	**32.2%**
Board State Only	38.5%	25.9%
Without FiLM	40.0%	30.3%
Average Embedding	39.1%	27.9%

Table 3: Comparison of the models' ability to predict support orders

Results: SelfPlay and TrueSkill

We train DipNet with self-play using A2C, with the same model for all powers and shared updates. The supervised model performs better than rule-based agents, but there is no significant difference between the SL and RL models.

Agent A (1x)	Agent B (6x)	TrueSkill A-B	% Win	% Most SC	% Survived	% Defeated
SL DipNet	Random	28.1 - 19.7	100.0%	0.0%	0.0%	0.0%
SL DipNet	GreedyBot	28.1 - 20.9	97.8%	1.2%	1.0%	0.0%
SL DipNet	Dumbbot	28.1 - 19.2	74.8%	9.2%	15.4%	0.6%
SL DipNet	Albert 6.0	28.1 - 24.5	28.9%	5.3%	42.8%	23.1%
SL DipNet	RL DipNet	28.1 - 27.4	6.2%	0.3%	41.4%	52.1%
Random	SL DipNet	19.7 - 28.1	0.0%	0.0%	4.4%	95.6%
GreedyBot	SL DipNet	20.9 - 28.1	0.0%	0.0%	8.5%	91.5%
Dumbbot	SL DipNet	19.2 - 28.1	0.0%	0.1%	5.0%	95.0%
Albert 6.0	SL DipNet	24.5 - 28.1	5.8%	0.4%	12.6%	81.3%
RL DipNet	SL DipNet	27.4 - 28.1	14.0%	3.5%	42.9%	39.6%

Table 4: Results when playing different models against eachother.

To probe variations in behavior, we consider agent collaboration:

- *X-support-ratio*: the fraction of support orders from a model that are attempting to support other agents.
- Eff-X-support-ratio: the fraction of X-support orders that succeed.

We find that the model is able to collaborate using support orders, but not as effectively as humans.

		X-support-ratio	*Eff-X-support-ratio*
Human Games	No Communication	14.7%	7.7%
	Public Comm.	11.8%	12.1%
	Public & Private Comm.	14.4%	23.6%
Agents Games	RL DipNet	9.1%	5.3%
	SL DipNet	7.4%	10.2%
	Board State Only	7.3%	7.5%
	Without FiLM	6.7%	7.9%
	Masked Decoder (No Board)	12.1%	0.62%

Table 5: Coalition Analysis

Human-Competitive

We hosted a tournament in collaboration with the website webDiplomacy. There were close to 100 participants, and 300 tournament games. In addition, over 4,200 games have been played with humans on the site.

- Bots are reasonably strong, but **not yet human level**. Humans win 3̃3% of games against six bots, much higher than the 14% we would expect if all players won equally.
- The bots do well in the opening and mid-game, but struggle in the endgame.
- The bots can get stuck in some positions that require careful coordination to disrupt the human position.
- The bots are very **loyal**, rarely backstabbing.

Come and play with the system on webdiplomacy.net!