COMP7704 Dissertation

Detailed Dissertation Proposal

Name:	Ng Argens
Supervisor(s):	Dr. Dirk Schnieders
Dissertation Title:	To Play and Cooperate in Imperfect Information Games with Machine Learning
Planned Submission Semester:	2018-2019 Semester 1

Aim

The goal of my dissertation is to explore the limitation of machine learning with yet another type of game, contract bridge. This game is chosen for having a wide array of features that can be added or removed depending on the research progress, including information hiding and discovery, variable reward, cooperation, as well as other rule modification that complicates or simplifies the game. Ideally, the aim is to solve the game of bridge under the system regulation of WBF, such that the resulting agent would be eligible to compete in World Computer-Bridge Championship in the future.

Brief Literature Review

Deep reinforcement learning has been having great success in gaming in general. In 2013, Google DeepMind developed an agent which outperformed all previous approaches on six of the seven Atari 2600 games with no change in architecture and hyperparameter (except one which happens to coincide with the flash rate of laser in Space Invader) [1]. Then in 2017, Google DeepMind used similar techniques to play Go at a superhuman performance, winning 100-0 against Alpha Go, another program which had defeated a world champion in 2016 [2].

Both successes mark the versatility and power of deep reinforcement learning in computer gaming. However, one field that has yet been extensively researched is multiplayer imperfect information game, in which another technique, counterfactual regret minimization (CFR), has a head start [3, 4]. One drawback of CFR and its variations though, is it requiring the storage of (game state, action) pair values or (information set, action) pair values, subject to the level of abstraction with human intervention.

A brief look at CFR literatures suggests that the main benefit of it being the ability to greatly reduce the storage required in regret minimization. However, most abstractions are still done by human, and the power does not seem comparable to that of neural networks in the game of Go. Therefore, the hope is high that deep q network (DQN) might be able to provide a better solution in games other than two-player zero-sum complete-information games and one-player environment-interaction games.

1. V. Mnih et al., “Playing Atari with deep reinforcement learning”, arXiv:1312.5602v1 [cs], Dec. 2013.
2. D. Silver et al., “Mastering the game of Go without human knowledge,” Nature, vol. 550, no. 7676, pp. 354–359, Oct. 2017.
3. R. Gibson et al., “Generalized sampling and variance in counterfactual regret minimization”, in AAAI-12: Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, Jul. 2012, Toronto, Canada.
4. N. A. Risk & D. Szafron, “Using counterfactual regret minimization to create competitive multiplayer poker agents”, in AAMAS '10: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems, May 2010, Toronto, Canada.

Proposed Methodology

As mentioned in the aim of the dissertation, contract bridge has the benefit of having many features that can be added or removed to slightly modify the game as the research progresses.

The research will start by training an agent to play a double dummy game with no trump suit, a two-player constant-sum complete-information game. I plan to do it by training a deep q network by reinforcement learning by playing against itself. Experience replay, model freeness and on/off policy are features and choices that I would consider.

After that, trump suits would be introduced, followed by contract, which adds complexity to the rules and card-play. Lastly, one of the dummy would be closed, the agents will be assigned roles and developed accordingly as declarer and defenders. This then becomes a multiplayer imperfect-information cooperation game. To avoid overtraining, we should select random past agents as opponents during training. Also, at this stage, knowledge representation of the signals (codified meaning of cards) should be developed. Bottom-up approach would be ideal but there is more likely to be human intervention.

If time is running short, the research could be considered complete here. We could bring the agent to the test of human professionals. Two Hong Kong youth team members could provide feedbacks as to whether the agents are performing well enough. To quantitatively analyze, we can use results from external double dummy analyzers which tells the maximum number of tricks possible at each seat with different trump suits and compare the results with the game between our agent and human professionals.

If there is still time, we would further the research to the bidding aspects of the game, where 4 agents team up to communicate information to reach the optimal contract. We could start by only having 2 agents bid as partners. While the training of agents would be more difficult, the evaluation would be easier as there is a clear Nash equilibrium in the final contract biddable using (external) double dummy analyzer. Again, reinforcement learning would be used, and the knowledge representation of biddings needs to be devised. Bottom-up approach would be ideal but human intervention might be needed again.

Milestones

Week	*Tasks*	*Time Period*	*Learning Hours*	*Concurrent Activity*
1 - 2	Study game theory and counterfactual regret minimization	1/6 – 14/6	80	Nil
3 - 4	Study deep q network, reinforcement learning, experience replay and other features and choices	15/6 – 28/6	65	COMP 7904 Session 1 - 5
5 - 6	Train declarer play agent by self-playing and possibly experience replay of no-trump games	29/6 – 12/7	65	COMP 7904 Session 6 - 10
7 - 8	Add in contract and trump suit to introduce variable reward and more complexed gameplay	13/7 – 26/7	80	Nil
9	[Optional] Devise suitable knowledge representation for signaling in defense	27/7 – 2/8	40	Nil
10- 11	Close out one dummy to conduct full version and observe cooperation of agents	2/8 – 15/8	80	Nil
12	Testing of results thus far with Hong Kong youth team members	16/8 – 18/8	16	Nil
12	[Optional] Devise suitable knowledge representation for additional / basic meanings in biddings	19/8 – 24/8	40	Nil
13 – 15	Train four bidding agents teamed into two teams aiming to communicate and cooperate using biddings only	25/8 – 14/9	100	Short Trip to Mainland China
16 – 17	Testing of result with external double dummy analyzers	15/9 – 28/9	80	Nil
			Total: 646	

Deliverables

Items	
1	Website
2	Report
3	Gaming agent
4	Bidding agent
5	Table manager (Environment)
6	Database (for experience replay)
7	Visualizer (for demonstration of agents' play)

Resource Needed (Expected)

Items		*Hours Needed*	*Application Done Before*
1	Gridpoint / HPC 2015	~ 30	Week 6 (6/7)