Group Formation by Group Joining and Opinion Updates via Multi-Agent Online Gradient Ascent

Chuang-Chieh Lin, *Tamkang University, Taiwan*,
Chih-Chieh Hung, *National Chung Hsing University, Taiwan**
Chi-Jen Lu, *Academia Sinica, Taiwan*
Po-An Chen, *National Yang Ming Chiao Tung University, Taiwan*

***Abstract*—This article aims to exemplify best-response dynamics and multi-agent online learning through group formation. We introduce the idea of best-response and how online learning algorithms can help stabilize the entire system composed of strategic agents. This extended abstract provides a summary of the full paper in IEEE Computational Intelligence Magazine on the special issue *AI-eXplained* (AI-X). The full paper includes interactive components to facilitate interested readers to grasp the idea of pure-strategy Nash equilibrium and how the system converges to a stable state by means of decentralized online gradient ascent.**

I. Introduction

GAME theory has been applied in a variety of aspects due to its predictability of outcomes in real world. It can also be used in solving problem, such as the saddle-point optimization and training generative adversarial network models [1]. A game is a system which consists of strategic agents, each of which acts rationally to maximize its own reward (or utility) or minimize its cost. A Nash equilibrium is a stable state composed of strategies of all agents such that none of the agents wants to change its own strategy unilaterally. Hence, such a stable state is possibly achievable or even predictable. Yet, how to achieve a Nash equilibrium in a game may not be quite straightforward, especially when agents behave in a "decentralized" way. Indeed, while an agent's reward functions has dependency on strategies of the other agents, a maximizer of one agent's reward function is not necessarily a maximizer for any other agent.

In this article, we study the group formation of a system of strategic agents as an illustrating example. A strategic agent can either join a group or change its opinion to maximize its reward. The eventual equilibrium of the game hopefully suggests predictable outcomes of the whole society. For the case that agents apply group joining strategies we consider *pure-strategy Nash equilibria* (PNE) as the solution concept, where a pure strategy means a strategy played with probability 1. For the case that agents change their opinions, we assume that each agent plays an *online gradient ascent algorithm*, which guarantees the time-average convergence to a hindsight optimum for a single agent (see [2] for the cost-minimization case), in a decentralized way and then we look into the convergent state of the system.

*Corresponding author.
This work is supported by the Taiwan Ministry of Science and Technology under grant no. NSTC 110-2222-E-032-002-MY2, NSTC 111-2221-E-005-047- and NSTC 111-2410-H-A49-022-MY2.

II. Group and Opinion Formation

In this game setting, we are given a set V of n agents, each of which v_i is represented as a *public preference vector* z_i and a *private preference vector* s_i, such that the former (we call it an *opinion*) corresponds to the preference revealed to all the agents while the latter corresponds to its *belief* which is unchangeable. We consider $s_i, z_i \in \mathcal{K}$ such that $\mathcal{K} := \{x \in [-1,1]^k : \|x\|_2 \le 1\} \subset \mathbb{R}^k$ is the feasible set. One can realize that each dimension of the domain stands for a certain social issue such that -1 maps to the far-left politics while 1 maps to the far-right politics. The bounded 2-norm constraint is in line with the bounded rationality of a person, or bounded budget for a group. We use $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ and $\mathbf{s} = (s_1, s_2, \ldots, s_n)$ to denote two profiles which include each agent's opinion and belief, respectively. Each agent is initially regarded as a group. The *opinion of a group* is the average of the opinions of its members. Like the *monotone* setting in [3, 4], a group wins with higher odds if its opinion brings more utility to all of the agents. The *reward* (i.e., payoff) of an agent is the expected utility it can get from all the groups. Specifically, assume that we have currently $m \le n$ groups $G_1, G_2, \ldots, G_m$, and denote by $|G_i| = n_i$ the number of members in group G_i. Let $\mathcal{G} = (G_1, G_2, \ldots, G_m)$ denote the profile of groups. To ease the notation, we denote by $\tau = (\mathbf{z}, \mathbf{s}, \mathcal{G})$ the *state* of the game. The reward function of agent i is $r_i(\tau) = \sum_{j=1}^m p_j(\tau)\langle s_i, \bar{g}_j\rangle$, where $\bar{g}_j = \sum_{v\in G_j} z_i/|n_j|$ represents the average opinion of group G_j and the winning probability $p_j(\tau)$ of group G_j is

$$p_j(\tau) = \frac{e^{n_j\langle \bar{g}_j, \sum_{v\in V} s_v\rangle}}{\sum_{i\in[m]; n_i>0} e^{n_i\langle \bar{g}_i, \sum_{v\in V} s_v\rangle}},$$

where $[m]$ denotes $\{1, 2, \ldots, m\}$. We consider the following strategic behaviors of an agent in such a game:

- Group Joining:
  - Seeking for a specific group which hopefully maximizes the agent's reward and join the group.
- Opinion Updating without Regularization:
  - Each agent in a certain group tries to maximize its reward through changing its own opinion.
- Opinion Updating with Regularization:
  - Each agent in a certain group tries to maximize its reward through changing its own opinion, while the

reward includes $-\|s_i - z_i\|_2^2$ as a regularizer which hopefully limits how strategic an agent can be by preventing it from moving too far from its own belief.

III. Group Joining and Pure Nash Equilibria

A. 1D Representation

When the opinions and beliefs are assumed to be in $\mathbb{R}$, we can illustrate these vectors as well as the dynamics of changes on a real line. For example, in Fig. 2 we have five agents v_1, v_2, v_3, v_4, v_5. By assuming v_1, v_2, v_3, and v_4 to have their public preference vectors z_1, z_2, z_3, z_4 fixed, we can observe the changes of winning probability of v_5 by moving z_5 from -1 to 1, and hence the rewards as well.

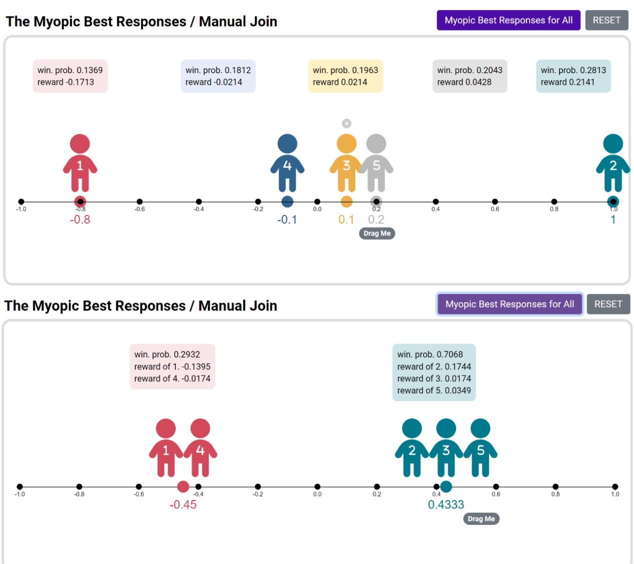

Fig. 1. 1D Representation: myopic best-responses and a PNE.

B. Group Joining: Myopic Best-Responses

We assume that agent i decides to join G_j for which $j = \arg\max_\ell p_\ell(\tau) \cdot \langle g_\ell, s_i \rangle$. We call this strategy a *myopic best-response*. An agent joins a group by considering not only its winning probability but also the utility that the agent can get from the group before its joining. In Fig. 2 we can check that the state below is a PNE.

IV. Opinion Updates by Online Learning

A. 2D Representation

We illustrate the opinions and beliefs as well as the dynamics of opinion changes in $\mathcal{K} := [-1, 1]^2$. The 2-norm constraint that $\|z_i\|_2, \|s_i\|_2 \leq 1$ correlates both the dimensions and projection of the opinion is required when it is not feasible.

B. Online Gradient Ascent

We consider the setting that each agent tries to maximize its own reward by "changing its opinion" without deviating from the group that it belongs to. Each agent runs the online gradient ascent algorithm to iteratively update their opinions so as to maximize its reward. The update is done by adding a certain quantity (tuned by the learning rate η) toward the direction of the gradient. A "projection" which projects x onto the feasible set $\mathcal{K}$ by dividing its 2-norm is performed if necessary.

Algorithm Multi-Agent Online Gradient Ascent

Input: feasible set $\mathcal{K}$, T, learning rate η.

```
for t ← 1 to T do
    for each agent i do
        observe reward r_i(τ), where state τ = (z, s, 𝒢)
        z_{i,t+1} ← Π_𝒦(z_{i,t} + η∇_{z_i} r(τ))
    end for
end for
```

C. Online Gradient Ascent with Regularization

The reward function for agent i including the regularizer is defined as $r_i(\tau) = \sum_{j=1}^{m} p_j(\tau)\langle s_i, \bar{g}_j \rangle - \|z_i - s_i\|_2^2$. Since $-\|z_i - s_i\|_2^2$ is always non-positive, an agent will be constrained to consider "not being too far from its private preference". Our experimental illustrations show that such a regularization helps the game converge to a state where agents will not have their opinions too far from their beliefs.

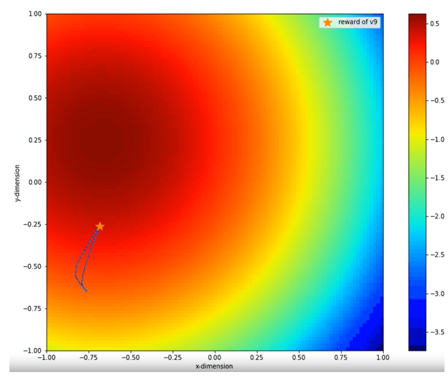

Fig. 2. Opinion updates via online gradient ascent with regularization.

V. Conclusion

This article presents a preliminary study on the dynamics of group formation. From the illustrations, readers can realize what a pure-strategy Nash equilibrium in a system of multi-agents is and also learn how online gradient ascent algorithms can help reach a stable state.

Acknowledgments

We thank Victorien Yen for helping us implement the Javascript codes in the immersive article.

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, pp. 139–144, 2020.

[2] F. Orabona, "A modern introduction to online learning," 2022.

[3] C.-C. Lin, C.-J. Lu, and P.-A. Chen, "How good is a two-party election game?" *Theoretical Computer Science*, vol. 871, pp. 79–93, 2021.

[4] ——, "On the efficiency of an election game of two or more parties: How bad can it be?" 2023.