Large Scale Multi-Agent Deep FBSDEs

Tianrong Chen [1] Ziyi Wang [2] Ioannis Exarchos [3] Evangelos A. Theodorou [4]

[1]School of ECE of GaTech [2]Center for Machine Learning of GaTech [3]Department of Computer Science of Stanford University [4]School of Aerospace Engineering of Gatech

Estimating Nash Equilibrium in Stochastic Differential Games

- HJB in Stochastic Differential Game (SDG) can be solve by a PDE solver but it does not scale beyond few dimensions.
- The deep learning model [2] suffers from limited exploration region since the dynamics is driven by Brownian motion.
- The learning algorithms might suffer from curse of many agents [3].

Scalable Fictitious Play Deep FBSDEs

The HJB-PDE of individual agent in SDG is,

$$V_t^i + h + V_x^{i\mathrm{T}}(f + G\boldsymbol{U}_{0,*}) + \frac{1}{2}\mathrm{tr}(V_{xx}^i \Sigma\Sigma^{\mathrm{T}}) = 0, \tag{1}$$

Scalable Fictitious Play Deep FBSDEs (SFPD-FBSDE) is a deep learning model which can estimate Nash Equilibrium motivated by theoretic analysis of classical FBSDEs system [?],

$$\begin{aligned} &\mathrm{d}\boldsymbol{X}_t = (f + G\boldsymbol{U}_{0,*})\mathrm{d}t + \Sigma\mathrm{d}\boldsymbol{W}_t,\ \boldsymbol{X}_{t_0} = \boldsymbol{x}_{t_0} \quad \text{(FSDE)},\\ &\mathrm{d}Y_t^i = -h_t^i\mathrm{d}t + Z_t^i\mathrm{d}W_t,\ Y_T^i = g(\boldsymbol{X}_T), \qquad \text{(BSDE)}, \end{aligned} \tag{2}$$

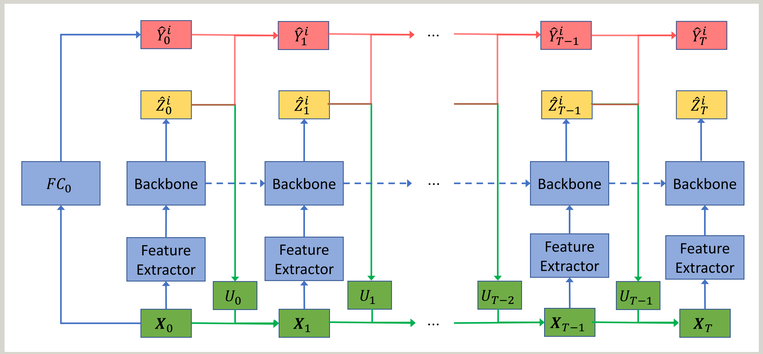

Figure 1. FBSDE Network for a single agent. The dashed arrow indicates hidden states propagation if LSTM is chosen as backbone. The dash arrow would disappear when FC is chosen. Red arrow represents for the BSDE propagation, Green arrow represents for the FSDE propagation

The solution of the BSDE is the solution of HJB-PDE almost surely and can thus be used to estimate the HJB PDE. Besides, in this paper we tackle the aforementioned difficulties:

- **Efficient Exploration** By incorporating importance Sampling technique, SFPD-FBSDE is guaranteed to have more efficient exploration compared to baselines from theoretical analysis. (modify the green arrow and the red arrow)
- **Mitigate Curse of Many Agents** SFPD-FBSDE integrates invariant layer [4] to mitigate the curse of many agents. (modify the feature extractor blue boxes)

Importance Sampling

The Importrance Sampling of FBSDEs is known as,

$$\begin{aligned} &\mathrm{d}\tilde{\boldsymbol{X}}_s = [\mu_s + \Sigma K_s]\mathrm{d}s + \Sigma\mathrm{d}\boldsymbol{W}_s,\ \tilde{\boldsymbol{X}}_t = \boldsymbol{x}_t,\\ &\mathrm{d}\tilde{Y}_s^i = [-h_s^i + \tilde{Z}_s K_s]\mathrm{d}s + \tilde{Z}_s^i\mathrm{d}W_s,\ Y_T^i = g(\boldsymbol{X}_T). \end{aligned} \tag{3}$$

The solution of modified BSDE is still aligned with the original HJB-PDE [1].

Exploration and bound for FBSDE system with drift term

Assumption 1 (informal): For a general FBSDE system,

$$\begin{aligned} \boldsymbol{X}_T^{t,\boldsymbol{x}} &= \boldsymbol{x} + \int_t^T \mu_s\mathrm{d}s + \int_t^T \Sigma_s\mathrm{d}\boldsymbol{W}_s \qquad \text{(FSDE)},\\ Y_t^{T,\boldsymbol{x}} &= g(\boldsymbol{X}_T^{t,\boldsymbol{x}}) - \int_t^T H_s\mathrm{d}s + \int_t^T Z_s\mathrm{d}\boldsymbol{W}_s \quad \text{(BSDE)}, \end{aligned} \tag{4}$$

The Lipschitz function are bounded.

Lemma 1 (informal): FBSDE system satisfying **assumption 1**, then we have

$$\begin{aligned} |\delta Y_T|^2 \le L_1|\boldsymbol{x}_1 - \boldsymbol{x}_2|^2, &|\delta Y_{t_0}|^2 \le L_2|\boldsymbol{x}_1 - \boldsymbol{x}_2|^2,\\ ||Z_t||_S^2 \le ||\Sigma||_S^2||\nabla_x Y_t||_S^2 &\le M_\Sigma L_2 \end{aligned} \tag{5}$$

Where L_1 and L_2 are the function of Lipschitz constants.

Theorem 2 (informal) Denote $\tilde{Y}_T$ is the solution of FBSDE with importance sampling, one has,

$$\begin{aligned} \max|\delta Y_T|^2 \le \max|\delta\tilde{Y}_T|^2,\\ \max|\delta Y_0|^2 \le \max|\delta\tilde{Y}_0|^2 \end{aligned} \tag{6}$$

Following table shows the similarity of training SFPD-FBSDE and traditional supervised learning task.

Items	Definitions
Training Data	random generated $\boldsymbol{x}_0$
Label	terminal cost $Y^* = \boldsymbol{g}(\boldsymbol{x}_T)$
Training Loss Function	$\|\|\hat{\boldsymbol{Y}}^T - \boldsymbol{Y}^*\|\|_2^2$
Trainable Parameters	parameters of backbones and FCs

Thus, Theorem 2 indicates that the framework with importance sampling can provide richer training data and better exploration efficiency which is critical for training a deep learning model.

Invariant Layers

Because of the symmetric problem setup, we incorporate the Invariant Layers (IL) [4] to extract permutation invariant features to improve the performance SFPD-FBSDE. It helps to mitigate the curse of many agents when the number of agent is increasing.

Experiments

We first conduct the ablation experiments on the Inter-Bank Game lending/borrowing problem to verify the effectiveness of IS and IL in fig.2

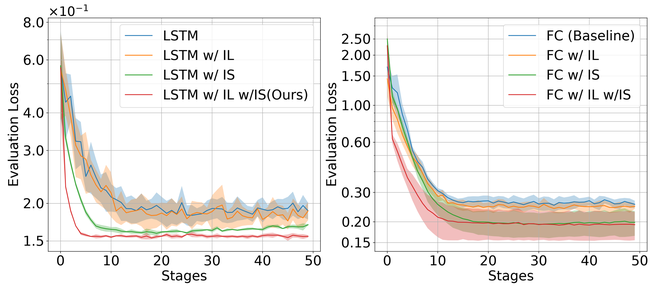

Figure 2. Ablation experiments on LSTM and FC backbone.

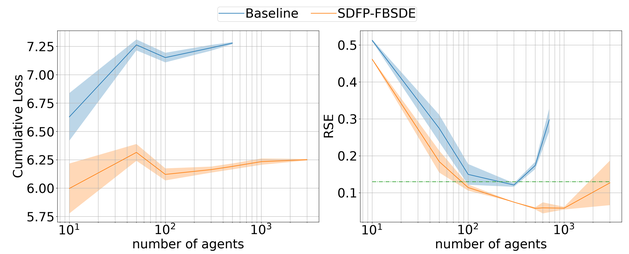

Figure 3. Comparison of SDFP-FBSDE and Baseline for inter-bank problem with different number of agents evaluated on cumulative loss and RSE.

Fig.3 shows the suprior performance of our model when the number of agents increases.

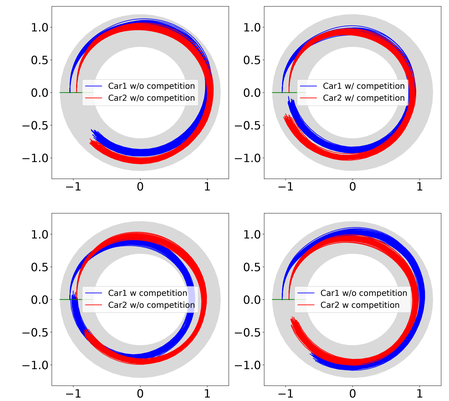

Figure 4. The plots contains 64 trials of racing. The performance varies with respect to the competition loss.

The algorithm is also executed on the partial observed autonomous racing cases in fig.4.

References

[1] Christian Bender and Thilo Moseler. Importance sampling for backward sdes. *Stochastic Analysis and Applications*, 28(2):226–253, 2010.

[2] Jiequn Han, Arnulf Jentzen, and E Weinan. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 115(34):8505–8510, 2018.

[3] Lingxiao Wang, Zhuoran Yang, and Zhaoran Wang. Breaking the curse of many agents: Provable mean embedding q-iteration for mean-field reinforcement learning. In *International Conference on Machine Learning*, pages 10092–10103. PMLR, 2020.

[4] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan Salakhutdinov, and Alexander Smola. Deep sets. *arXiv preprint arXiv:1703.06114*, 2017.

tchen429@gatech.edu