9^{th} International Conference on Computational Social Science IC^{2}S^2
July 17-20, 2023, Copenhagen, Denmark

Quantitative Evaluation of Multi-agent Simulation using Generative Adversarial Network

Keywords: Multi-agent simulation, Generative adversarial network, Evaluation, Financial markets, Social Simulation

Introduction Multi-agent simulation is a promising instrument for social sciences. However, validating these social simulations is often tricky because of their quantitative evaluation difficulties. Fundamentally, social phenomena are usually difficult to quantify because of their qualitativeness. Although it is difficult, quantitative evaluation should be addressed because qualitative evaluation has its limitation in terms of simulation parameter tunings, simulation comparison, and so on. Hence, in this study, we address this issue to enable quantitative evaluation of multi-agent simulations, especially in artificial market simulations, i.e., multi-agent simulations for financial markets. While these qualitative evaluations seem challenging, we propose that a generative adversarial network (GAN) can convert human qualitative evaluation into a quantitative evaluation through distributional evaluation.

Method: GAN Evaluator Figure 1 shows the outline of the core concept. We aimed to propose an alternative method to human qualitative evaluation of simulations with the quantitative evaluation of GAN to enable quantitative evaluation. The assumption is that GAN can be an alternative because GAN can learn the vague distributional characteristics of real data through its learning architecture. In our proposed method, only the critic of the trained GAN was used for the simulation evaluations. Simulations generate the data, and the critic classifies it as fake or real.

Data and Model Implementation In our experiments, we targeted financial markets. As a GAN model, we employed PGSGAN [2], a state-of-the-art GAN for financial market order generation that can consider the discreteness of orders. We used the same data as in [2] from the Tokyo Stock Exchange (TSE) between January and September 2019 for the training of the GAN. As a simulation model, we used the stylized financial market simulation proposed in [4], which is widely used as an artificial financial market simulation.

Experiments In the first step, we tuned the parameters of the artificial market simulations using the GAN evaluator. The tuning algorithm we employed was a tree-structured Parzen estimator (TPE) [3]. Then we tested the following stylized facts presented in [1] in simulations: (1) Absence of autocorrelations, (2) Heavy tails property, (3) Aggregational Gaussianity, (4) Volatility clustering, and (5) kurtosis and skewness of log returns of each time lag. Moreover, to validate our GAN evaluator, we changed some parameters from the optimal values and checked the replications of those stylized facts (we call this "ablation study"). Because the replication of some of those stylized facts is difficult to determine, we employed a three-tier evaluation:

- +: fairly observed
- -: not observed. But the judgment could include our bias or be subjective.
- -- : fairly not observed. The judgment is confident.

Table 1: Ablation study results. Rows are sorted by GAN evaluator value. The top row is the best tuned case and the following rows are cases in which one simulation parameter is changed. The fourth column represents the overall evaluation based on the stylized facts following the column. ACF, Agg. Gaus., Vol. Clus., and Kurt./Skew. mean autocorrelation function (the same as the absence of autocorrelation), aggregational Gaussianity, volatility clustering, and kurtosis/skewness, respectively.

				The details of stylized facts				
Param.	*Value Change*	*GAN*	*All*	*ACF*	*Heavy Tail*	*Agg. Gaus.*	*Vol. Clus.*	*Kurt./Skew.*
–	–	0.96524	+	+	+	+	+	+
$\tau_{\min}$	$15.1 \to 1$	0.96523	+	+	+	+	+	+
$\tau^*_{\max}$	$64700 \to 669$	0.95020	+	+	+	+	+	+
$k_{\min}$	$0.0314 \to 0.0552$	0.91626	+	+	+	+	+	+
$\tau^*_{\min}$	$669 \to 0$	0.91145	+	+	+	+	+	+
w_C	$2.29 \to 50$	0.86420	+	+	+	+	+	+
n_{agent}	$929 \to 20$	0.84225	-	+	+	+	-	+
$k_{\max}$	$0.0552 \to 1$	0.68891	-	+	+	+	-	+
n_{agent}	$929 \to 2$	0.64020	--	No price movement occurs (--)				
$k_{\min}$	$0.0314 \to 0$	0.51658	-	+	+	-	-	-
$\tau_{\max}$	$10470 \to 15.1$	0.31392	--	--	+	-	+	-
w_F	$36.6 \to 0$	0.02262	--	+	+	-	-	--
σ	$5.52 \times 10^{-6} \to 0.1$	-1	--	Price goes infinity (--)				

Findings and the Impacts Table 1 shows the results of the ablation study. This table shows which parameter is changed and how much it is changed in each row. The top row is the best tuned case and the following rows are cases in which one simulation parameter is changed. According to this result, there seems to be a reasonable rank correlation between the output of GAN evaluator and traditional qualitative evaluation based on stylized facts.

Those results support the usefulness of the GAN evaluator as an alternative to the traditional evaluation using stylized facts. The realization of quantitative evaluation using GAN as an alternative to the traditional qualitative evaluation may expand the usage of multi-agent simulation. For future works, more varied simulations should be tested, and more precise evaluations should be addressed.

References

[1] Rama Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

[2] Masanori Hirano, Hiroki Sakaji, and Kiyoshi Izumi. Policy gradient stock gan for realistic discrete order data generation in financial markets. 2022. https://doi.org/10.48550/arXiv.2204.13338.

[3] Yoshihiko Ozaki, Yuki Tanigaki, Shuhei Watanabe, and Masaki Onishi. Multiobjective tree-structured parzen estimator for computationally expensive optimization problems. pages 533–541. Association for Computing Machinery, 2020.

[4] Takuma Torii, Kiyoshi Izumi, and Kenta Yamada. Shock transfer by arbitrage trading: analysis using multi-asset artificial market. *Evolutionary and Institutional Economics Review*, 12(2):395–412, 2015.

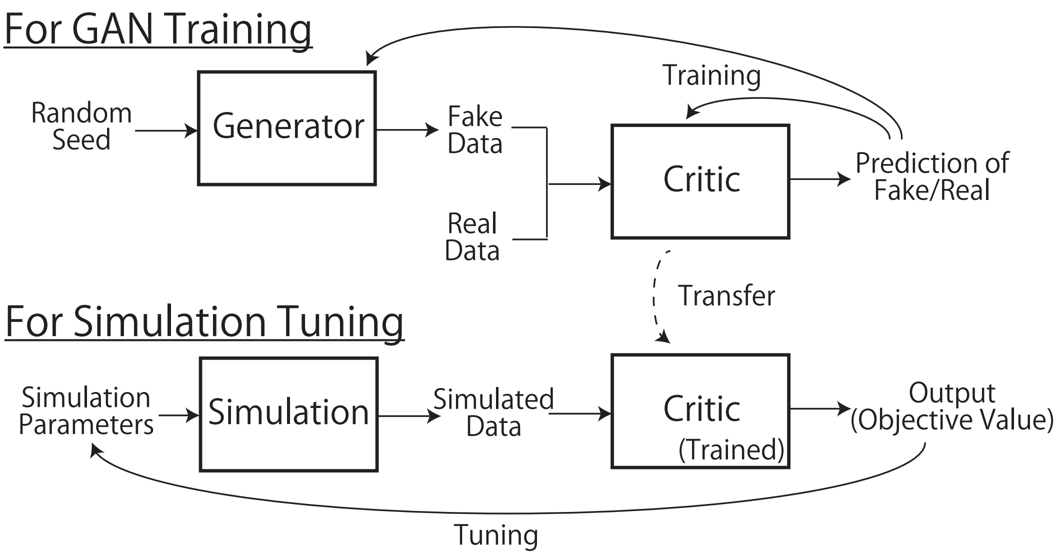

Figure 1: Outline of the GAN Evaluator. The first step is GAN training using actual data. In the training phase, GAN is trained adversarially. The GAN generator accepts random seeds for generation and generates fake data. Then, the critic of GAN classifies the fake data and the actual data. This phase is the same as the usual GAN training procedure, and any GAN models and the actual data can be applicable. In the second step, the only critic of GAN is used for simulation evaluation. Simulations generate the data, and the critic classifies it as fake or real.