Human–Machine Social System: a Mean-Field Game Approach

Jiejun Hu-Bolz*, James Stovold

Lancaster University Leipzig, Nikolaistraße 10, 04109 Leipzig, Germany
{j.hu14, j.stovold}@lancaster.ac.uk

Abstract

The rapid development in machine intelligence, fuelled by increased data processing capabilities, advanced algorithms, and pervasive adoption, has opened new avenues for exploring the interaction between humans and machines as a social system. In this paper, we propose a mean-field game-based model to analyse large-scale interaction between humans and distributed AI-enabled machines. This paper is preliminary work in studying Human–Machine Social Systems (HMSS).

Introduction

Humans interact and obtain information from machines equipped with algorithms in daily activities, such as browsing the web, using ChatGPT, driving, trading stocks, etc. Not only has the interaction of humans and AI-enabled machines become more frequent, but also the interactions among increasingly-autonomous machines have increased, forming a Human–Machine Social System (HMSS) (Tsvetkova et al., 2024).

While there are extensive studies regarding one-to-one Human–AI systems (Walton et al., 2022), small groups of humans interacting with one AI (Leite et al., 2015), one human interacting with distributed AI systems (Kolling et al., 2016), or groups of humans and machines (Hollan et al., 2000) the reality is that we are moving towards massively-interconnected systems where many human and AI agents influence each other. It is crucial to investigate the unique interaction and collective decision-making processes of humans and machines at a large scale. For humans and machines to achieve a common goal, information flow will lead to bidirectional influence between humans and machines. For example, humans can feed new data to machines to alter their actions, which pushes the overall HMSS towards intelligence alignment; Machines can influence each other through direct or indirect interaction, similar to how persuasion and cultural evolution occur in humans. Thorough analysis and modelling are required to mitigate the negative impact of false manipulation and promote the tendency of intelligence alignment.

While models based on individual agent interactions are tractable for studying smaller-scale problems, when we consider entire societies an approach to agent-based modelling based on mean-field game theory (Lasry and Lions, 2007) becomes a better option. This approach allows us to ask questions about macroscopic behaviour of societies in a manner that is still computationally tractable, with an aim of aligning the results with smaller-scale models later.

Considering the increasing number of machines and the pervasive interaction, this research will adopt the mean-field game approach to analyze the collective decision-making processes. Ultimately, it will help us understand the conditions to achieve the best outcome of collective intelligence of humans and machines.

To capture the dynamics between humans and machines, we propose four key factors in the decision-making processes of both humans and machines: 1) information accumulation, 2) algorithm advancement, 3) hardware advancement, 4) interconnectivity. In this extended abstract, we assume the machines are autonomous artificial agents that interact with humans in the same social space. Humans use satisfaction as feedback to machines—greater satisfaction leads to more information exchange with machines. At the same time, machines are interconnected, which enables them to share information with peer machines. In the proposed HMSS, we study how humans' satisfaction as feedback to machines influences machines' cooperation dynamics and machine intelligence maturity.

Notation	Description
Satisfaction state of human	$s \in [0, 0.5]$
A set of machines	$i \in \mathcal{M}$
Intelligence maturity state	$m_i \in [0, 1]$
Mean-field term of intelligence maturity	$\theta_m \in \mathcal{N}(0.25, 0.2)$
Machine cooperation action	$c_i \in [0, 1]$
Weights of machine cost function	$\omega_1, \omega_2, \omega_3$
Weights of satisfaction dynamics	$\beta_1, \beta_2, \beta_3$

Table 1: Notations and descriptions for the equations on the following page.

Modelling

To study the impact of human satisfaction on machines, we model machine i's cost considering the cooperation with humans, the communication cost with peer machines, and reward of obtaining feedback from humans. We first define humans' satisfaction with the interaction as s. Machine i's cooperation is defined as c_i. The intelligence maturity of machine i is denoted as m_i. Please refer to Table 1 for the key notations and description. Machine i's cooperation cost is captured by the quadratic form of action c_i to ensure diminishing returns (Hu-Bolz et al., 2023). The communication cost with peer machines is defined as the product of m_i and the average maturity of i's peers, θ_m. Finally, the reward is obtained through humans' satisfaction with c_i. Hence, the cost of machine i is

$$L_i(c_i, s, \theta_m) = \omega_1 {c_i}^2 + \omega_2 \theta_m m_j - \omega_3 c_i s \tag{1}$$

where ω_1, ω_2, and ω_3 are positive weights. As the interconnectivity among humans and machines in HMSS, the interactions can affect the machine i's intelligence maturity. Hence, m_i evolves dynamically according to the average intelligence maturity θ_m, current machine maturity m_i, and humans' satisfaction s, which is defined as a Partial Differential Equation (PDE)

$$dm_i = (\beta_1 \theta_m - m_i + \beta_2 c_i + \beta_3 s)dt + \sigma dW(t) \tag{2}$$

where β_1, β_2, and β_3 are positive weights. We capture the randomness using Brownian motion, where σ is the diffusion constant and $W(t)$ a standard Wiener process (Lasry and Lions, 2007). Machine i aims to minimise the expected cost by adjusting its cooperation. We can then propose the optimisation problem of machine i

$$\min_{c_i} \quad J_i = \int_0^T \int_{\mathcal{M}} [L_i(c_i, s, \theta_m)\theta_m] dt \tag{3a}$$

s.t.

$$dm_i = (\beta_1 \theta_m - m_i + \beta_2 c_i + \beta_3 s)dt + \sigma dW(t) \tag{3b}$$

We reformulate the proposed problem in Eq. 3 using the Hamilton–Jacobi–Bellman equation and Fokker–Planck–Kolmogorov equation and then solve it by adopting the finite difference method from our previous work (Hu-Bolz et al., 2023).

Simulation

We evaluate the impact of humans' satisfaction on the mean intelligence maturity and collective cooperation of machines. Fig. 1 demonstrates that the evolution of mean intelligence maturity with respect to various satisfaction: with the increasing satisfaction, not only the mean intelligence maturity increases, but also the mean intelligence maturity increasing rate is greater.

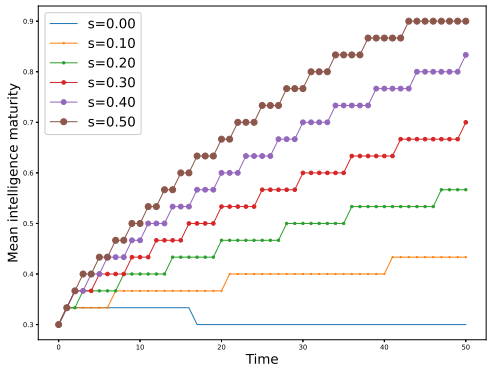

Figure 1: The evolution of mean intelligence maturity with respect to various satisfaction

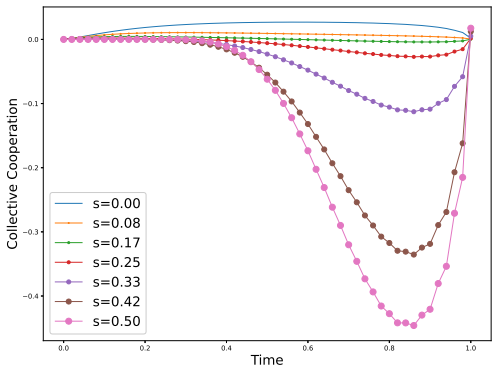

Figure 2: The evolution of collective cooperation with respect to various satisfaction

Additionally, we study the collective cooperation of machines, defined as the product of cooperation and its probability. In Fig. 2, when the satisfaction is below 0.25, the collective cooperation is positive, which guarantees positive cooperation with humans; While the satisfaction is greater than 0.25, it increases negatively, which can be viewed as disruptive (noncooperative) machine behaviour.

Conclusion

The proposed work adopts a mean-field game to model machines' cooperation with humans considering humans' satisfaction. We observe that high satisfaction levels could promote machine intelligence maturity, potentially lead to disruption. In future studies, we will analyse the cost incurred by humans interacting with intelligent machines, modeling this as a leader–follower game. Our focus will be on examining human cooperative behavior with machines, ranging from daily use to no interaction, and the associated satisfaction states within this dynamic.

Acknowledgements

Jiejun Hu-Bolz is supported within the project TRACE-V2X, which has received funding from the European Union's HORIZON-MSCA-2022-SE-01-01 under grant agreement No 101131204.

References

Hollan, J., Hutchins, E., and Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. ACM Trans. Comput.-Hum. Interact., 7(2):174–196.

Hu-Bolz, J., Farrahi, K., et al. (2023). Beyond the surface of digital contact tracing: Delving into the interconnected world of technology, individuals, and society. Authorea Preprints.

Kolling, A., Walker, P., Chakraborty, N., Sycara, K., and Lewis, M. (2016). Human interaction with robot swarms: A survey. IEEE Transactions on Human-Machine Systems, 46(1):9–26.

Lasry, J.-M. and Lions, P.-L. (2007). Mean field games. Japanese journal of mathematics, 2(1):229–260.

Leite, I., McCoy, M., Lohani, M., Ullman, D., Salomons, N., Stokes, C., Rivers, S., and Scassellati, B. (2015). Emotional storytelling in the classroom: Individual versus group interaction between children and robots. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction, HRI '15, pages 75–82. Association for Computing Machinery.

Tsvetkova, M., Yasseri, T., Pescetelli, N., and Werner, T. (2024). Human-machine social systems. arXiv preprint arXiv:2402.14410.

Walton, S. P., Rahat, A. A. M., and Stovold, J. (2022). Evaluating mixed-initiative procedural level design tools using a triple-blind mixed-method user study. IEEE Transactions on Games, 14(3):413–422.