Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

IDENTIFYING CORRELATES OF EMERGENT BEHAVIORS IN AGENT-BASED SIMULATION MODELS USING INVERSE REINFORCEMENT LEARNING

Faraz Dadgostari
Samarth Swarup

Biocomplexity Institute and Initiative
University of Virginia
Town Center Four 994 Research Park Boulevard
Charlottesville, VA 22911, USA

Stephen Adams
Peter Beling

Virginia Tech National Security Institute
Virginia Polytechnic Institute and State University
900 North Glebe Rd (MC 0379)
Arlington, VA 22203, USA

Henning S. Mortveit

Department of Engineering Systems and Environment
University of Virginia
Thornton Hall, 351 McCormick Road
Charlottesville, VA 22904, USA

ABSTRACT

In large agent-based models, it is difficult to identify the correlate system-level dynamics with individual-level attributes. In this paper, we use inverse reinforcement learning to estimate compact representations of behaviors in large-scale pandemic simulations in the form of reward functions. We illustrate the capacity and performance of these representations identifying agent-level attributes that correlate with the emerging dynamics of large-scale multi-agent systems. Our experiments use BESSIE, an ABM for COVID-like epidemic processes, where agents make sequential decisions (e.g., use PPE/refrain from activities) based on observations (e.g., number of mask wearing people) collected when visiting locations to conduct their activities. The IRL-based reformulations of simulation outputs perform significantly better in classification of agent-level attributes than direct classification of decision trajectories and are thus more capable of determining agent-level attributes with definitive role in the collective behavior of the system. We anticipate that this IRL-based approach is broadly applicable to general ABMs.

1 INTRODUCTION

Most of the modern world is composed of highly dynamic and complex systems formed through interacting intelligent agents and their adaptive decisions. A key challenge in this regard is to understand how the heterogeneous characteristics of their decentralized, local dynamics translate to collective system behavior and dynamics (Fenichel et al. 2011; Funk et al. 2015).

Heterogeneity of individual level characteristics of agents and their local embedding within the system is reflected in the heterogeneity of the adaptive interaction among agents and thus drives the aggregate global dynamics of the system. Due to heterogeneity of the agents and systemic complexities, it is difficult to empirically or analytically identify if and how the emerging global dynamics of the system are driven by which structural attributes of agents' (demographics, activity patterns, and system embedding), as well as their behavioral heterogeneity (which depend on descriptive and injunctive norms (Bicchieri et al. 2021), perceptions of risk and efficacy of different behaviors in risk reduction).

Some existing methods use machine learning (ML) algorithms to train reduced-form mapping structures between the agent-based model (ABM) parameters and the emergence dynamics (Lamperti et al. 2018; Thorve et al. 2020; Vahdati et al. 2019). A key goal of these approaches is to build faithful surrogates of ABMs to reduce the computational cost of ABMs, and these techniques are by design not well suited to provide mechanistic insight relating local properties and behaviors to global system dynamics.

In this paper, we aim to take a step toward overcoming this shortcoming by proposing a systematic approach to investigate what are the key individual-level characteristics of agents that govern their emerging behaviors in a heterogeneous multiagent setting. Our proposed method is based on inverse modeling of agents' observed sequential decision-making process and allows us to study whether an individual-level characteristic, such as an agent's behavioral model or a demographic attributes, or their (local) embedding in a heterogeneous complex system is a key driver of their realized behavior in a multi-agent setting. To evaluate the efficiency of our method in identifying the key drivers that *directly* or *indirectly* contribute to agents' emerging behaviors, we experiment with ABMs that incorporate 1) individual-level attributes that are defined explicitly/directly in the definition of an agent (e.g. agents' behavioral models), and 2) individual-level attributes that are implicitly/indirectly baked in the structure of the social contact network of agents (e.g. agents' age, gender, household size and income, population density, etc.).

Inverse reinforcement learning (IRL) (Adams et al. 2022) is a ML technique that estimates the reward function of a Markov decision process (MDP) from observed agent behavior. IRL has shown the capability of replicating agent behaviors in multi-agent settings (Lin et al. 2017; Lin et al. 2019). In this paper, we formulate the sequential decision-making of agents as MDPs where each agent's decision-making behavior is driven by a reward function aimed at maximizing her expected future reward. We use IRL to relate the underlying decision-making drivers of agents to their observed behaviors in multi-agent settings, which is a step toward a systematic methodology to connect the local and global dynamics of complex systems.

The IRL modeling frameworks used by Lee et al. (2017) and Rucker et al. (2021) are similar to our work in that the agents' behavior rules are modeled as MDPs, and IRL is used to extract the parameters of the agents' reward functions. However, we use model-free IRL to study large-scale ABMs, whereas Lee et al. (2017) and Rucker et al. (2021) use IRL techniques that require a model to study relatively small and simplified simulation environments. Furthermore, Lee et al. (2017)'s proposed IRL-based method aims to construct a simplified copy of ABMs using the simulation output of the original ABMs. Thus, the main point of the study is to evaluate the capacity of IRL to construct simplified copies of agents that can faithfully replicate the behavior of the original agents, whereas in this paper, we use IRL as a representation learning tool that is used to investigate if and how the local and agent-level attributes relate to the emerging dynamics of the ABMs. On the other hand, Rucker et al. (2021) study the capacity of IRL to identify the pre-defined strategies of agents in an adversarial environment, whereas in our work, the aim is to examine whether an agent-level attribute plays a driving role in the agents' emerging behaviors. This would not be possible to study in the experimental setting of Rucker et al. (2021), since in their simulated models, all individual-level attributes are identical, other than the agents' pre-defined strategies. thus, the only possible driver of the behavioral differences of agents would be due to the differences of the pre-defined strategies.

We have developed the proposed method in the context of the BESSIE epidemic simulator (Mortveit et al. 2022) which combines a detailed epidemic process with agent decision-making. Indeed, large-scale simulations, and specifically ABMs, are widely used to study systems of agents with behavioral models. However, there is no systematic framework to gain insight about the relevance and importance of local and agent-level characteristics to the emerging behavior of agents in response to for example an epidemic outbreak, public health policies, and interventions.

To evaluate how well the reward functions capture the underlying dynamics that relate local and agent-level characteristics to the agents' behavior in a multi-agent setting, we use agents' reward functions as a basis to identify different agent classes, such as their behavioral, age, income level and employment classes. Our computational experiments involve three different agent-based models and two widely used classification methods, Decision Trees and Multi-Layer Neural Networks, to compare the performance of

the reward functions with a baseline that uses the agent's raw decision-making trajectories as a basis to identify different agent classes.

Based on our results, conventional classification models such as Decision Trees and Multi-Layer Neural Networks have significantly better performance if IRL based reward functions are used for their training compared to raw decision trajectories or demographic attributes of agents. Our results are consistent across different parametric settings of BESSIE simulation; with and without explicit behavioral models formulated in the definition of agents, suggesting that IRL-based reward functions are better formulation of ABMs simulation output for investigation of the agent-level and local characteristics that have definitive role in the collective behavior of the systems.

In Section 2 we briefly review some preliminaries, including Markov Decision Processes (MDP) and Inverse Reinforcement Learning (IRL). In Section 3, we describe our case study of an ABM of a Covid-like outbreak. Then, we formally present the sequential decision-making modeling framework we used to formulate the agents' observed decision trajectories in Section 4. In Section 5 we discuss our experimental setting and the numerical analysis of the results. Finally, we conclude in Section 6 and discuss potential opportunities for future research.

2 INVERSE REINFORCEMENT LEARNING

The goal in IRL is to estimate a reward function for a Markov decision process from observed agent behavior. An MDP is a model for sequential decision making and is represented by the tuple $(\mathscr{S}, \mathscr{A}, T, R, d_0, \gamma)$, where $\mathscr{S}$ represents a set of states, and $\mathscr{A}$ represents a set of actions. T is the transition function, where $T(s,a,s') = P(s_{t+1} = s' \mid s_t = s, a_t = a)$ for $s, s' \in \mathscr{S}, a \in A$. R is the reward function, where $R(s,a)$ represents the reward for taking action a in state s. The initial state distribution is represented by d_0, and γ is the discount factor. In many IRL formulations, it is assumed that $R(s,a) = \sum_{i=1}^{k} \theta_i f_i(s,a)\ \forall (s,a)$, where f_i represents a state feature, θ_i represents a corresponding feature weight, and $i = 1...k$. A policy π is a deterministic or stochastic function that maps states to actions.

The expected return of π is the sum of expected rewards that will be received acting under π, that is, $J(\pi) = \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \mid d_0, \pi, T]$. An optimal policy maximizes $J(\pi)$. The expectation of f_i under π can be written as $f_i^{\pi} = \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t f_i(s_t, a_t) \mid d_0, \pi, T]$. Following this, $J(\pi)$ can be written as a linear combination of the feature expectations $J(\pi) = \sum_{i=1}^{k} \theta_i f_i^{\pi}(s,a)$. Let τ represent a trajectory of state-action pairs, and let $\mathscr{T}$ be the set of trajectories. The expected return over the set of all trajectories following π is $J(\pi) = \sum_{\tau \in \mathscr{T}} P(\tau \mid \pi, T) \sum_{i=1}^{k} \theta_i f_i^{\tau}$, where $f_i^{\tau} = \sum_t \gamma_t f_i(s_t, a_t)$.

The aim of IRL is to find a reward function that produces a policy π where the expected return is at least as good as the expert's policy π^E. However, $J(\pi)$ and $J(\pi^E)$ cannot be directly compared without the reward function. Ng and Russell (2000) proposed the first IRL algorithm to learn an agent's reward function and then use it to recover the expert's policy. Most of the existing IRL algorithms focus on solving this problem assuming that the transition matrix of the agent is known, however, as discussed earlier, in ABMs the transition dynamics are complicated and may not be easily formulated as a transition matrix. One approach to overcome this limitation is to use algorithms such as the model-based maximum entropy proposed by Ziebart et al. (2008, 2010), where a transition matrix is learned from sampled trajectories. However, these methods require a large number of sampled trajectories, otherwise they may lead to learning reward functions that are completely different from the true ones. Since it is computationally expensive to sample a large number of trajectories for each agent in a large-scale ABMs, these algorithms are not efficient to learn agent reward functions. Therefore, for learning agents' reward functions, we use the relative entropy IRL algorithm proposed by Boularias et al. (2011) which builds on Ziebart et al. (2010)'s maximum entropy algorithm but is capable of learning high quality reward functions from small sample of trajectory instances.

The relative entropy IRL algorithm (Boularias et al. 2011) minimizes the KL divergence between the empirical distribution of $\mathscr{T}$ under a baseline policy, represented by q, and the distribution of the trajectories under a policy that matches the reward feature counts of the agent's decision trajectories, represented by p,

$$\min_p \mathrm{KL}(p||q) = \min_p \sum_{\tau \in \mathscr{T}} p(\tau) \ln \frac{p(\tau)}{q(\tau)}.$$

Constraints are added to the optimization problem to minimize the difference between the observed feature counts and the expected feature counts under the estimated reward, and to ensure that the probability of a trajectory is non-zero and that the probabilities of all trajectories sum to 1. Boularias et al. (2011) demonstrate that solution to this constrained optimization problem can be found using the Lagrangian and KKT conditions. However, this gradient can only be calculated if the transition function is known. Boularias et al. (2011) propose using importance sampling to estimate the gradient when the transition function is unknown.

3 CASE STUDY: AN AGENT-BASED MODEL OF AN EPIDEMIC OUTBREAK

We use agent-based simulation as a proxy of the real world, where the adaptive interactions of heterogeneous agents drive both individual-level and aggregate dynamics of the system. While different goals may require different kinds of models, recovering behavioral models from simulation data requires two components: adequate detail based on real data (Swarup 2019) and true agency (adaptive coupling with the environment, normativity, etc. (Barandiaran et al. 2009)). To capture adequate detailed information of the real world data we used a synthetic population that statistically represent real world properties of the modeled population. We also utilized the behavioral modeling capacity of the BESSIE tool (Mortveit et al. 2022) to simulate adaptive and behavioral interactions of individuals in response to a Covid-like epidemic process. We also integrated local and global observables into the agents' decision making processes to capture how local embedding of agents and their asymmetric access to information influences their adaptive interactions.

Synthetic Population. We use a *synthetic population* (SP), a statistically accurate representation of a population of a given region (Chen et al. 2020; Adiga et al. 2015). For a given region R, an SP has the following components:

- P is the set of *individuals* of R. Each individual is grouped into a *household*, and each person p has demographic attributes such as age, gender, household income, and a simplified version of the NAICS classification.
- Each person has an *activity sequence* that specifies the time and type of activity they perform. The activity sequence covers a week. The types of activities are `Home`, `Work`, `Shopping`, `Other`, `School`, `College` and, `Religion`.
- The set of *locations* is divided into *residence locations* and *activity locations*. Households are assigned to residence locations and an activity a for each person is assigned an activity location.

Using the activity sequences of of the SP, one can determine which individuals visited the same location at the same time during the week.

Epidemic Process. An *epidemic process* takes place over the population, where transmission can occur when a susceptible person comes into contact with an infectious person at one or more locations. The BESSIE tool (Mortveit et al. 2022) implements an agent-based model for the spread of an epidemic that integrates the synthetic population with the epidemic process.

Local and Global Observables. For each person and for each time step, the following data are available:

- Local observables: The person's attributes and her own current *health state*. For each activity type, the following counts are recorded as the person enters a location, and the most recent from prior iterations: (i) count of people in the disease states, (ii) the total count of visitors at the location, (iii) the count of people wearing masks at the location, and (iv) the count of people engaging in social distancing at the location.
- Global observables: counts of the number of people in each disease state at the beginning of each day (iteration) after initialization.

Behavioral Models. We use the customizable *behavioral modeling* feature of BESSIE to allow each person to adopt a set of measures. At the beginning of each time step, the behavioral model for each person selects from the following actions: (1) wear a mask, (2) social-distance, or (3) modify her visit schedule to refrain from selected activities. If an activity is skipped, it is replaced by the `Home` activity. The behavioral model for each agent may use the agent's global observables (e.g., total number of symptomatic people), or local observables (e.g., demographic information about the person, information about things they have seen at previous visits such as the number of symptomatic cases observed when shopping).

4 MARKOV DECISION PROCESS FORMULATION OF AGENT DECISION TRAJECTORY

This section outlines the formulation of states and actions to model the sequential decisions of an individual in response to an epidemic outbreak. Formally, we define each person's decision-making time steps by $t \in T$ to represent the steps of her scheduled activities during the observation. The state of an agent is composed of the features described in Table 1, and therefore, the states defined for each person can be represented as $S_t = \left(\bar{\alpha}_t, \alpha_{t+1}, N_t, I_t, O_t, M_t\right)$. Here, the index of the agent, p, is dropped to simplify the notation

Table 1: Description of MDP features for each agent decision making process.

State feature	
α_t	Planned activity for time t according to the activity schedule
$\bar{\alpha}_t$	Realized activity at t (may differ from α_t by the location being mapped to the person's home)
N_t	Observed number of people for most recent activity of type (e.g., work) matching α_{t+1}
I_t	Observed number of infected people for most recent activity type matching α_{t+1}
O_t	Observed number of symptomatic people for most recent activity type matching α_{t+1}
M_t	Observed number of people wearing masks for most recent activity type matching α_{t+1}
Action feature	
$a_{1,t}$	Binary variable; 1, if the agent conducts the activity defined by α_{t+1} in person; 0 if at home
$a_{2,t}$	Binary variable; 1, if the agent wears a mask when conducting the activity α_{t+1}; 0, otherwise
$a_{3,t}$	Binary variable; 1, if the agent social distances when conducting the activity α_{t+1}; 0, otherwise

Note that the second state feature, α_{t+1}, is the planned activity defined by the activity schedule for the next time step, $t+1$, which may be realized or not, depending on the agent action at t.

Next, let us define the MDP action for each agent at time t using the action features described in Table 1. All action features are binary decision variables where $a_{\cdot} = 1$ and $a_{\cdot} = 0$ represent taking and not taking the defined action, respectively. With that, $a_t = \left(a_{1,t}, a_{2,t}, a_{3,t}\right)$ formulates the MDP action of each person. Note that there are some restrictions under this action definition. For example, if $a_{1,t} = 1$, then the only possible action is $[1,0,0]$ because we assume that mask wearing and social distancing will not be implemented at home and the home activity is done in person.

The transition function does not need to be characterized, as we will explore model-free methods. The discount rate is a parameter that can be adjusted. Different reward functions will be estimated using IRL, so this section simply assumes that the reward function is a function of states and actions $R(s,a)$ or a function of states $R(s)$. Therefore, the state and action construction must contain all the information required to characterize the reward function.

5 NUMERICAL ANALYSIS AND DISCUSSION

For the numerical analysis of our approach, we used the synthetic population data of 4000 people in Charlottesville, Virginia, for a period of 75 days. We defined three different behavioral models for an agent, 1) risk-averse, 2) risk-neutral, and 3) risk-seeking. Behavioral models formulate the probability

of an agent deciding to take an action at $t+1$ based on their state at t and form a set of available actions. Following the formulation of random utility models of population choice behavior by (Lerman and Manski 1981; McFadden and Train 2000), the probability of an agent participating in their next scheduled activity is defined as a logistic function of their observables, such as the relative number of people wearing mask, social distancing and being symptomatic at the location of her next scheduled activity, α_{t+1} as defined $p(a_{1,t}) = Logit(\beta_0 + \beta_1 * \frac{O_t}{N_t} + \beta_2 * \frac{M_t}{N_t} + \beta_3 * \frac{I_t}{N_t})$. Similarly, an agent's probability of wearing a mask $p(a_{2,t})$ and probability of social distancing $p(a_{3,t})$ if participating in α_{t+1} is determined. The parametric setting of the risk-averse, risk-neutral, and risk-seeking behavioral models in our experiments are set as $\beta = (-5, 0.2, 0.2, 0.2)$, $\beta = (-1, 0.1, 0.1, 0.1)$, and $\beta = (-10, 0.02, 0.02, 0.02)$, respectively. We used the behavioral modeling capability of BESSIE (see Section 3) to incorporate these models into agent-based modeling of the pandemic outbreak in three different ABMs. In Model I, we assign the behavioral models based on the age and household size attributes of each agent. More specifically, we assigned a risk-averse model to agents who are 35 years old or older, and a risk-neutral model if the agent's age is equal to or less than 20 years old. For agents that are between 20 and 35 years old, if the household size is less than or equal to 2, we assign risk-seeking model, otherwise risk-neutral model. In Model II, the behavioral models are assigned arbitrarily. In Model III, risk-neutral behavioral model is assigned to all agents so that each agent's behavior model is the same. For Model III, the observed behavioral difference should be solely attributed to demographics or location in the simulation. This allows us to investigate if we can identify any individual level attribute of agents that are not explicitly formulated in the behavioral model, but indirectly drives or correlate with the emerging behavior of agents.

We use the simulation outputs of these three agent-based models, referring to them as "raw decision trajectories" of agents during a period of 75 days as input data for the Relative Entropy IRL algorithm, which in turn learns the reward functions of agents.

Finally, we use the trained reward functions for qualitative and quantitative evaluation of the IRL's capacity to associate agent-level attributes to emergent behavior of agents within the multi-agent setting.

5.1 Qualitative Assessment

Here we use t-SNE (t-Distributed Stochastic Neighbor Embedding), a nonlinear dimension reduction and data visualization method, to visually explore the potential capacity of the reward functions to identify underlying behavioral models of agents and if it helps to associate agent-level attributes to emergent behavior of the agents. Reward function of each agent is a vector of 440 weights. The size of the reward function vector is associated with the number of all observed states in all agent decision trajectories. Using t-SNE we reduced the dimensionality of the approximated weight vectors of the reward functions from 440 to 2, which in turn was used to visualize the trained reward functions of the population.

Figure 1a illustrates the 2-dimensional embedding of the reward functions of the population that are simulated using the three behavioral models, assigned to individual agents based on their age and household size attributes. We color-labeled the reward functions of the individuals based on their assigned behavioral model. It is easy to see that different clustering structures appear even in the two-dimensional illustration of the reward functions, where the reward functions of the individuals with the yellow-labeled behavioral model are distinct from the other two sets of individuals. This pattern consistently appears independent of the parametric setting of the t-SNE, when the assigned behavioral models are formulated based on the agents' age and household population attributes. This suggests that the agents' reward functions have some capacity to capture the effect of agent-level attributes in their emerged behavior within the multi-agent setting, even after dimensionality reduction from 440 to 2.

We further investigate this idea by simulating two other agent-based models, Model II and Model III, that have less structured behavioral models formulated in the agent's definition. In Model II, we assign the behavioral models to agents arbitrarily. In Model III, we used only one behavioral model, risk-neutral, for all agents. Therefore, in Model III, agents are identical in their definition and differ only in terms of their location within the agent-based model. In other words, although the third agent-based model does not

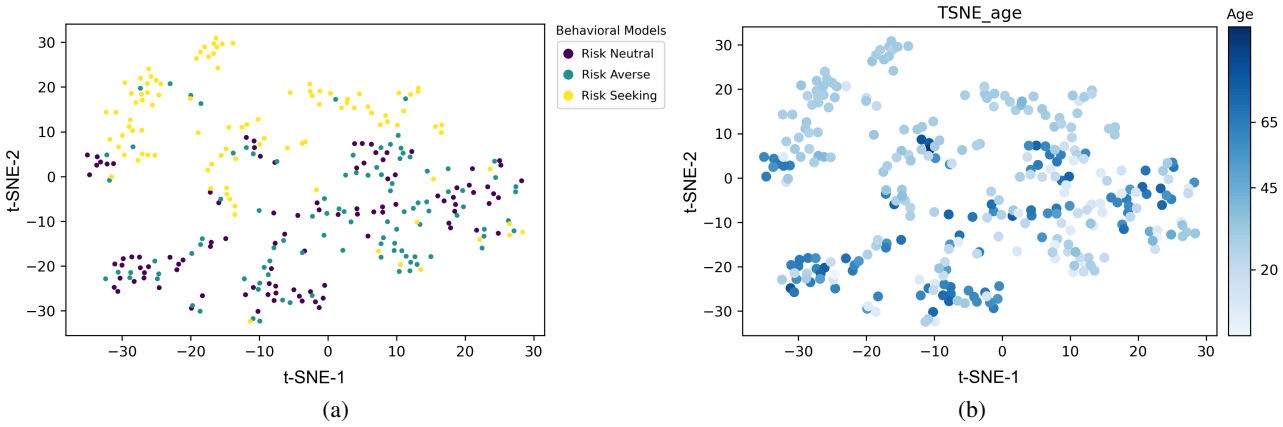
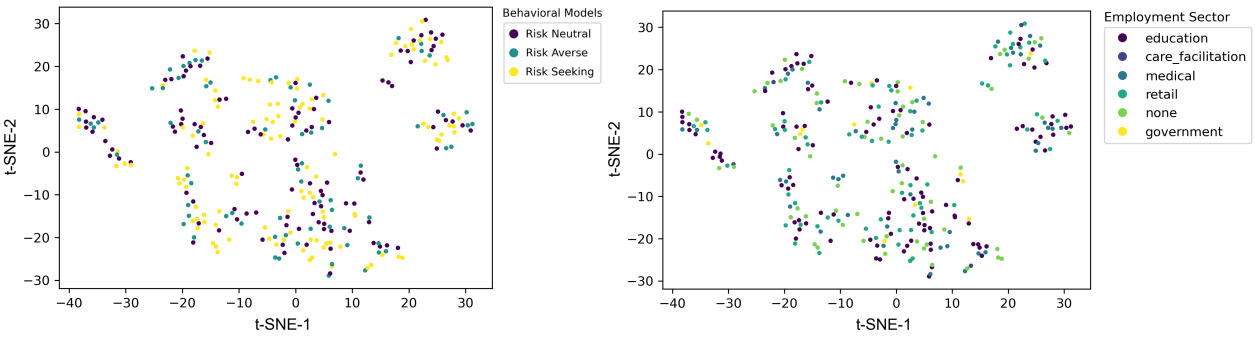


Figure 1: Three behavioral models formulated based on the age and household size of the agent (Model I). Weight vectors of the reward functions after dimentionality reduction from 440 to two by the t-SNE algorithm (a) Illustrate the behavioral model assigned to the agents by color labels, The appeared clustering patterns suggest that even though agents are assigned with three equality different behavioral models, the emerging behavior of risk-neutral and risk-averse agents are more similar compered to emerging behaviors of risk-seeking agents; (b) This shows how correlation of age with risk seeking model partially contributes to the differentiation of emerging behaviors.

directly use individual-level attributes in the definition of the agents, and since agents are embedded in the model based on the contact network of the synthetic population, their location within the agent-based model would be the source of heterogeneity of their behaviors. Then, if there are any differences in the emerging decision-making behaviors, one would associate them indirectly to agent attributes, as they contribute to agent locations within the synthetic population and therefore the agent-based model.

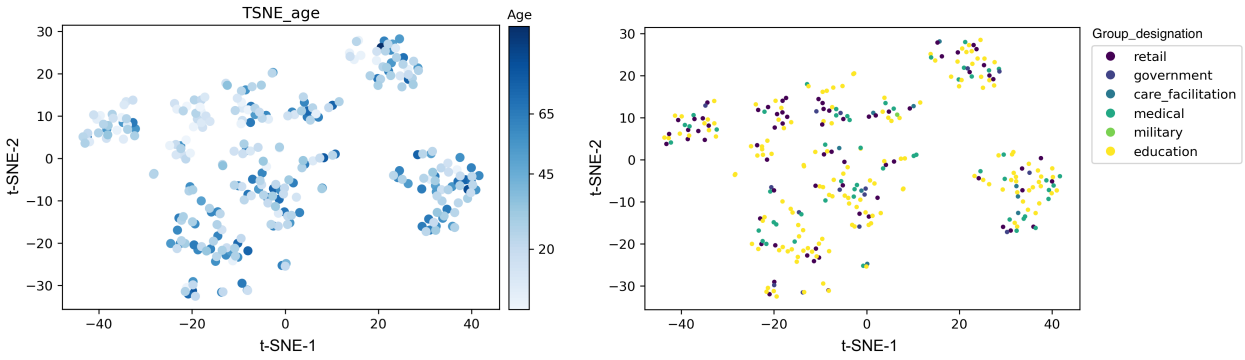
As shown in Figure 2 and Figure 3, some clustering patterns appear consistently in the t-SNE visualizations of agent reward functions trained using the simulation outputs of Model II and Model III. However, it is difficult to relate these clustering patterns to any agent-level attributes, including the assigned behavioral models or demographic characteristics, whereas in the t-SNE visualizations of the simulation outputs of Model I, we can recognize the correlation between age and household size attributes and the behavioral model of agents. We omit including the t-SNE visualizations of the agents' reward functions labeled by other demographic attributes, including race, employment status, household size, income level, etc; here, since similar to Figure 2 and Figure 3, it is difficult to relate the clustering patterns to the agent attributes.

The qualitative assessment of the reward functions trained using the simulation outputs of the first agent-based model implies that the combination of an agent's age and household size correlated with her behavioral model is associated with the agent's emerging behavior. But for the second and third agent-based models (Model II and Model III), the qualitative assessment is not conclusive. This could be due to two reasons: 1) the t-SNE visualization of IRL trained reward functions are not capable of capturing such associations or 2) the agent-level attributes like the agent's behavioral model, age, household size, etc. do not have any definitive association or definitive role in shaping the emergent decision-making behavior of agents. Note that the first possible reason, itself, could be due to loss of information by dimentionality reduction of t-SNE visualization or incapability of the IRL trained reward functions to capture such correlations in the first place. In the next section, we show that more complicated models are capable of using IRL trained reward functions to appropriately separate interactions among dimensions, provide reasonable classification accuracy, and reveal correlates of emergent behaviors in our experiments.



(a) (b)

Figure 2: Weight vectors of the reward functions after dimentionality reduction by t-SNE algorithm with arbitrary assignment of behavioral classes (Model II) (a) Color labels illustrate the behavioral classes assigned to the agents; (b) Color labels illustrate the employment sector of the agents.



(a) (b)

Figure 3: Weight vectors of the reward functions after dimentionality reduction by t-SNE algorithm with risk-neutral behavioral classes assigned to agents (Model III) (a) Darker blue illustrates higher ages of the agents.; (b) Color labels illustrate the employment sector of the agents.

5.2 Quantitative Assessment

Given that the IRL-trained reward functions of agents are (sequential decision) model-based reformulations of the raw decision trajectories of agents (raw output of simulations of agent-based models), we expect them to be a more informative representation of agents, especially in terms of the underlying drivers of their emergent behavior. To investigate this hypothesis, in this section, we quantitatively examine whether agent reward functions are more informative compared to the raw output of the simulations, especially in terms of the associations of agent-level attributes with the emerging behaviors of agents.

We use the performance of two conventional classification models, Decision Trees and Multi-Layer Neural Networks, as a basis to compare the informative capacity of raw decision trajectories versus IRL-trained reformulation of raw decision trajectories (referred to as reward functions) in predicting agent-level characteristics such as behavioral models, age, household size, employment group, and income level, based on each agent's observed trajectory during the ABM simulations. Note that for classification training, we convert numerical attributes such as age, household size, and income level to categorical attributes In other words, we train the same classification model (say Decision Trees) once using the agents' raw decision trajectories and once again using their reward functions. With this we can examine whether the

reformulation of the simulation output using IRL is a better representation of the agent, as it can increase the performance of the classification algorithms in recognizing the associations of agent-level attributes with the emerging behaviors of agents. Note that here the only source of difference between the classification models' performance is due to the model-based prepossessing of the input data using IRL, and thus any improvement in the prediction performance of the classification models could be associated with the higher capacity of the IRL-based representation of agents' observed behavior.

In the first part of this section, we apply two widely used classification methods, Decision Trees and Multi-Layer Neural Networks, to predict agent behavioral classes of Models I and II.

Table 2: Behavioral class prediction accuracy of Model I and Model II.

		Model I (% weighted averages)			Model II (% weighted averages)		
Data	Classifier	Precision	Recall	f1-score	Precision	Recall	f1-score
Decision Trajectories	Decision Trees	69	70	69	49	52	50
	Multi-Layer NN	41	40	40	47	45	45
Reward Functions	Decision Trees	90	89	89	68	64	65
	Multi-Layer NN	91	90	90	86	74	67

This allows us to evaluate the informative capacity of the reward functions to recover the behavioral models of agents from their emerging behaviors in a multi-agent setting, in comparison to the informative capacity of raw output of agent-based model simulations.

In the second part, we apply the same classification methods to investigate whether the data generated by agent-based simulations could be used to predict different demographic attributes of agents. Performance of a predictive model of an agent-level attribute based on simulation outputs could be interpreted as a lower bound for how much is the definitive role of that attribute in shaping the emergent decision-making behavior of agents.

Based on the results of the qualitative assessment section, we expect that reward functions would be capable of predicting behavioral classes as well as ages and household sizes of the agents, for the agent-based model I. Table 3 verifies the implications of the qualitative assessments in 5.1 and also shows that other demographic attributes could be recovered with significantly higher accuracy if reward functions are used for classification compared to the raw simulation outputs.

Table 3: Demographic attribute prediction of Model I.

		Accuracy (%)			
Data	Classifier	Age	Household Size	Employment Type	Income Level
Decision Trajectories	Decision Trees	52	69	57	39
	Multi-Layer NN	37	60	39	53
Reward Functions	Decision Trees	67	83	78	76
	Multi-Layer NN	75	78	78	78

Table 4 and 5 also show that either of the classification methods, decision trees and Multi-Layer Neural Networks, could recover demographic attributes of the agents with significantly higher accuracy if reward functions are used to train the classification models compared to the raw simulations outputs. This gives us a better picture of how well the trained reward functions capture the underlying dynamics of the simulation, compared to the raw simulations outputs such as decision trajectories of agents.

The results of Table 5 may seem counterintuitive, since IRL is supposed to pick up the behavioral models of individuals and since there is no explicit behavioral model formulated in Model III, the resulting

Table 4: Demographic attribute prediction of Model II.

		Accuracy (%)			
Data	Classifier	Age	Household Size	Employment Type	Income Level
Decision Trajectories	Decision Trees	62	56	58	54
	Multi-Layer NN	49	49	43	43
Reward Functions	Decision Trees	83	60	74	77
	Multi-Layer NN	88	62	75	75

Table 5: Demographic attribute prediction of Model III.

		Accuracy (%)			
Data	Classifier	Age	Household Size	Employment Type	Income Level
Decision Trajectories	Decision Trees	41	64	54	52
	Multi-Layer NN	38	57	42	42
Reward Functions	Decision Trees	79	81	73	72
	Multi-Layer NN	68	84	70	70

reward functions should be similar with no specific structure. However, these results support the idea that the emergent behaviors of individuals are also formed (indirectly) by their demographic attributes, since demographic attributes relate to the agents' location within the heterogeneous structure of the underlying contact networks of the synthetic population. Thus, despite the fact that the demographics of individuals do not play any role in the formulation of agents in models II and III, the underlying heterogeneity and dynamics of the agent-based models result in emergence of differences in agents behaviors.

6 CONCLUSIONS AND FUTURE WORK

In this paper we showed that reformulating the observed behavior of agents in a multi agent setting based on a behavioral inverse sequential decision modeling framework as a basis to identify the key individual level attributes of agents that correlate with their emerging behaviors. We used the Relative Entropy IRL algorithm to train representations of agents' reward functions in a large-scale ABM of a COVID-like epidemic process, where agents must make sequential decisions in response to an epidemic outbreak.

Our results illustrate the capacity and performance of the IRL-based representations (reward functions) to identify static and dynamic attributes that correlate with the emerging dynamics of large-scale multi-agent systems in different experimental settings.

This is a step forward toward a systematic framework for analyzing how agent-level attributes as well as local structures relate to the emerging behavior of the agents in large-scale agent-based models. But given the limitations of IRL methodologies, in terms of interpretability of the trained reward functions, it is hard to understand why IRL based reformulation of decision trajectories improves the performance of classification methods and thus how reward function capture the relationship of the agents emerged behaviors with the agent attributes. Therefore, exploration of interpretable IRL algorithms that are capable of learning meaningful reward functions in a multi-agent setting and from small samples of decision trajectories is an avenue of future work. This would require experimentations with ABMs that explicitly formulate the goal-seeking nature of agents' decision-making to validate the credibility of proposed interpretable IRL algorithms. Another avenue for future research is to develop an experimental design framework to rigorously analyze the significance and causal dependencies of hypothetical key drives of the complex system's global dynamics. It is also important to examine the robustness of the proposed inverse sequential

decision modeling as well as the IRL techniques used in this paper against noise, partially observable information, and a more diverse set of behavioral models.

ACKNOWLEDGMENTS

This work was partially supported by the Global Infectious Diseases Institute grant "Machine Learning Efficient Behavioral Interventions for Novel Epidemics" and NSF Expeditions in Computing CCF-1918656.

REFERENCES

Adams, S., T. Cody, and P. A. Beling. 2022. "A Survey of Inverse Reinforcement Learning". *Artificial Intelligence Review*:1–40.

Adiga, A., A. Agashe, S. Arifuzzaman, C. L. Barrett, R. J. Beckman, K. R. Bisset, J. Chen, Y. Chungbaek, S. G. Eubank, S. Gupta, M. Khan, C. J. Kuhlman, E. Lofgren, B. L. Lewis, A. Marathe, M. V. Marathe, H. S. Mortveit, E. Nordberg, C. Rivers, P. Stretz, S. Swarup, A. Wilson, and D. Xie. 2015. "Generating a Synthetic Population of the United States". Technical Report NDSSL 15-009, Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute, Virginia Tech.

Barandiaran, X. E., E. Di Paolo, and M. Rohde. 2009. "Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-Temporality in Action". *Adaptive Behavior* 17(5):367–386.

Bicchieri, C., E. Fatas, A. Aldama, A. Casas, I. Deshpande, M. Lauro, C. Parilli, M. Spohn, P. Pereira, and R. Wen. 2021. "In Science We (Should) Trust: Expectations and Compliance Across Nine Countries During the COVID-19 Pandemic". *Plos One* 16(6):e0252892.

Boularias, A., J. Kober, and J. Peters. 2011. "Relative Entropy Inverse Reinforcement Learning". In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, 182–189. JMLR Workshop and Conference Proceedings.

Chen, J., A. Vullikanti, S. Hoops, H. Mortveit, B. Lewis, S. Venkatramanan, W. You, S. Eubank, M. Marathe, C. Barrett, and A. Marathe. 2020. "Medical Costs of Keeping the US Economy Open During COVID–19". *Nature Scientific Reports* 10:18422.

Fenichel, E. P., C. Castillo-Chavez, M. G. Ceddia, G. Chowell, P. A. Gonzalez Parra, G. J. Hickling, G. Holloway, R. Horan, B. Morin, C. Perrings, M. Springborn, L. Velazquez, and C. Villalobos. 2011. "Adaptive Human Behavior in Epidemiological Models". *Pnas* 108(15):6306–6311.

Funk, S., S. Bansal, C. T. Bauch, K. T. D. Eames, W. J. Edmunds, A. P. Galvani, and P. Klepac. 2015. "Nine Challenges in Incorporating the Dynamics of Behaviour in Infectious Disease Models". *Epidemics* 10:21–25.

Lamperti, F., A. Roventini, and A. Sani. 2018. "Agent-Based Model Calibration Using Machine Learning Surrogates". *Journal of Economic Dynamics and Control* 90:366–389.

Lee, K., M. Rucker, W. T. Scherer, P. A. Beling, M. S. Gerber, and H. Kang. 2017. "Agent-Based Model Construction Using Inverse Reinforcement Learning". In *Proceedings of the 2017 Winter Simulation Conference (WSC)*, edited by V. W. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. H. Page, 1264–1275. Institute of Electrical and Electronics Engineers, Inc.

Lerman, S., and C. Manski. 1981. "On the Use of Simulated Frequencies to Approximate Choice Probabilities". *Structural Analysis of Discrete Data With Econometric Applications* 10:305–319.

Lin, X., S. C. Adams, and P. A. Beling. 2019. "Multi-Agent Inverse Reinforcement Learning for Certain General-Sum Stochastic Games". *Journal of Artificial Intelligence Research* 66:473–502.

Lin, X., P. A. Beling, and R. Cogill. 2017. "Multiagent Inverse Reinforcement Learning for Two-Person Zero-Sum Games". *IEEE Transactions on Games* 10(1):56–68.

McFadden, D., and K. Train. 2000. "Mixed MNL Models for Discrete Response". *Journal of Applied Econometrics* 15(5):447–470.

Mortveit, H. S., S. C. Adams, F. Dadgostari, S. Swarup, and P. A. Beling. 2022. "BESSIE: A Behavior and Epidemic Simulator for Use With Synthetic Populations". *CoRR* abs/2203.11414.

Ng, A. Y., and S. J. Russell. 2000. "Algorithms for Inverse Reinforcement Learning.". In *Proc. ICML*, Volume 1, 663–670.

Rucker, M., S. Adams, R. Hayes, and P. A. Beling. 2021. "Inverse Reinforcement Learning for Strategy Identification". In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 3067–3074.

Swarup, S. 2019. "Adequacy: What Makes a Simulation Good Enough?". In *Proceedings of the Spring Simulation Conference (SpringSim)*. Tucson, AZ.

Thorve, S., Z. Hu, K. Lakkaraju, J. Letchford, A. Vullikanti, A. Marathe, and S. Swarup. 2020. "An Active Learning Method for the Comparison of Agent-Based Models". In *Proceedings of the 19th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*.

Vahdati, A. R., J. D. Weissmann, A. Timmermann, M. S. P. de León, and C. P. Zollikofer. 2019. "Drivers of Late Pleistocene Human Survival and Dispersal: An Agent-Based Modeling and Machine Learning Approach". *Quaternary Science Reviews* 221:105867.

Ziebart, B. D., J. A. Bagnell, and A. K. Dey. 2010. "Modeling Interaction via the Principle of Maximum Causal Entropy". In *Proceedings of the International Conference on Machine Learning*.

Ziebart, B. D., A. L. Maas, J. A. Bagnell, A. K. Dey et al. 2008. "Maximum Entropy Inverse Reinforcement Learning.". In *Proceedings of the AAAI International Conference on Artificial Intelligence*, Volume 8, 1433–1438. Chicago, IL, USA.

AUTHOR BIOGRAPHIES

FARAZ DADGOSTARI is a postdoctoral research associate in the Biocomplexity Institute and Initiative at the University of Virginia. His research interests are in the area of adaptive decision-making in large-scale socio-economic, cyber-physical, and human-AI systems, with emphasis on theoretically-grounded, model-centric, and data-informed solutions for complex engineering, business, and policy problems. His email address is fd4cd@virginia.edu.

SAMARTH SWARUP is a research associate professor in the Biocomplexity Institute and Initiative at the University of Virginia. His research interests are in large-scale agent-based simulations and machine learning applied to problems in public health and social science. His email address is swarup@virginia.edu.

STEPHEN ADAMS is a research associate professor in the Intelligent Systems Division of the Virginia Tech National Security Institute. His research focuses on applications of machine learning and artificial intelligence in real-world systems. His email address is scadams21@vt.edu.

PETER BELING is a professor in the Grado Department of Industrial and Systems Engineering at Virginia Tech and associate director of the Intelligent Systems Division in the Virginia Tech National Security Institute. His research interests lie at the intersections of systems engineering and AI and include reinforcement learning, transfer learning, and digital engineering. His email address is beling@vt.edu.

HENNING S. MORTVEIT is an associate professor in the Biocomplexity Institute and Initiative and the Department of Engineering Systems and Environment at the University of Virginia. His research interests include the mathematical structures, theory, software design and computational architectures involved in the modeling and analysis of coupled, co-evolving, massively interacting, networked systems. His email address is Henning.Mortveit@virginia.edu.