Toward a Unified Theory of Human-Agent Modeling: A Position Paper

Elizabeth Sklar[1,2]

[1] The Graduate Center, City University of New York, New York, NY 10016, USA
[2] Brooklyn College, The City University of New York, Brooklyn, NY 11210, USA
sklar@sci.brooklyn.cuny.edu

Abstract. Approaches to *Human-Agent Modeling* can broadly be categorized as the use of agents to emulate (1) the *outcome* and (2) the *process* of human behavior. The first category can be further sub-divided into emulating the outcomes produced by *individual* humans and by *groups* of humans. This position paper proposes this categorization as the core of a unified theory of human-modeling.

1 Introduction

Over the last ten and more years, we have conducted research that has touched on the idea of human-agent modeling, as the focus of multiple separate projects. This position paper attempts to bring this work together and formulate a unified theory of *Human-Agent Modeling*. We believe that approaches fall broadly into two categories: (1) the use of agents to emulate the *outcome* of human behaviors; and (2) the use of agents to emulate the *process* of human behaviors. The first category can be further sub-divided into emulating the outcomes produced by (1a) individual humans; and (1b) groups of humans.

We consider environments that can be modeled as Markovian, either fully or partially observable. A human or agent starts in an initial state, S_0, and performs some sequence of actions, $a_0, a_1, \ldots, a_{n-1} \in \mathcal{A}$, to arrive at a goal state, S_g. Our aim is to build agent models of human behaviors that share the same S_0 and S_g, as well as the same intermediate states, $S_0 \rightarrow S_1 \rightarrow \ldots \rightarrow S_g$. In addition, we would like to emulate the sequence of actions taken by human(s), if we have available information about their actions. We define four different types of models, at varying levels of abstraction:

I. *start-end-state model* (the most abstract): the agent mimics the human's start state, S_0, and end state, S_g
II. *all-states model*: the agent mimics the human's complete sequence of states: $S_0 \rightarrow S_1 \rightarrow \ldots \rightarrow S_{g-1} \rightarrow S_g$
III. *states-actions model*: the agent mimics the human's complete sequence of state-action pairs: $S_0 \stackrel{a_0}{\rightarrow} S_1 \stackrel{a_1}{\rightarrow} \ldots \stackrel{a_(n-2)}{\rightarrow} S_{g-1} \stackrel{a_(n-1)}{\rightarrow} S_g$
IV. *action-selection model* (the least abstract): the agent mimics not only the complete set of state-action pairs (as above), but also the human's *decision-making* process for selecting which actions to take in each state

Our work has primarily focused on building models of type III and IV, which are the emphasis in this paper, though we have done some work constructing type II models [1]. We have also found that type II models are useful in evaluating the results of our modeling efforts [2]. (See Section 3.)

The approaches that emulate the *outcomes of individual* humans' behaviors (category 1a) produce models of type I, II or III. A variety of representations and machine learning techniques can be used to acquire and store such models. For example, a Bayesian network [3] could represent the probability that a particular person executes an action in a given state, and a model could learn the probabilities associated with each state-action pair. (See Section 2.) The approaches that emulate the *outcomes of groups* of humans' behaviors (1b) employ a strategy similar to modeling individuals and also produce type I, II or III models. However, the data used to train the model is based on the collective behavior of multiple individuals. The collective behaviors may be clustered in various ways, e.g., according to behavior characteristics or demographics. (See Section 3.) The approaches that emulate the *process* of human behaviors (2) produce type IV models. These tend to be centered on individuals (though these approaches could model groups where all members exhibit the same, or similar, processes), so we do not differentiate between "one" and "many" trainers in this category. The idea is to model the human's *decision-making* process for action selection—not just what the outcome was, but how she made the choice of action(s) that led to the outcome. (See Section 4.)

Generally speaking, there are two different representation strategies employed for storing the models produced, each following one of the two historically opposing camps from within the Artificial Intelligence community [4]: *symbolic* methods that require manual construction of rules which describe behaviors, versus *connectionist* methods that require machine learning procedures to evolve weights or probabilities that guide selection of actions based on input states. The first strategy involves *engineering* a policy for an agent to follow when selecting an action given a state, whereas the second involves *learning* the policy[3]. The more successful applications of the first strategy stem from work done by behavioral specialists, such as developmental psychologists, in which they observe human subjects engaging in activities and then create theories about the subjects' behaviors by analyzing the observations made. Following this procedure, computational models can then be built that reproduce the subjects' behaviors. While the second strategy could also be based on theoretical models from the literature, the connectionist methods more typically skip that step and learn directly from quantitative or categorical data that is collected during a period of human subject activity. For example, a human expert could explain to a programmer how she plays the game of Backgammon and the programmer could engineer a rule set that encodes the expert's behavior; or a connectionist program (e.g., neural network) could learn how to play Backgammon by playing many games against itself [5], another computer player or a human opponent, and adjusting its weights as it plays.

[3] Note that the second also involves engineering the connectionist representation.

representation strategy	(1a) modeling outcomes of individuals	(1b) modeling outcomes of groups	(2) modeling processes
symbolic	SimEd, type IV [6]	SimEd, type IV [7]	MBTI, type IV [8]
connectionist	Tron, type III [9] robotics, type III [11]	Tron, type III [10] SimEd, type II & III [2]	

Table 1. Categorization of approaches to human-agent modeling for projects discussed in this paper: SimEd (Sections 2 and 3); MBTI (§4); Tron (§2 and §3); robotics (§2).

Our work has primarily focused on using connectionist strategies, though we have also explored symbolic strategies. Table 1 illustrates how some of our different projects, undertaken since 1997, fit within the categorization described above. The remainder of this paper highlights these examples; and we close with a brief discussion of ongoing and future work.

2 Modeling the outcomes of individuals' behaviors

In 1997, we introduced a simple human-vs-agent video game inspired by the movie *Tron* and demonstrated one of the first Internet-based examples of collective intelligence [12]. In our Tron game, both players move continuously within a toroidal arena, each leaving an impenetrable trail marking its path (neither player can cross these trails). The trails do not decay, so eventually the arena fills up with "walls" and it becomes hard to find safe moves. The agent player has sensors that measure the distance to the nearest obstacles. Both players have a choice of 3 actions: turn left, turn right, or go straight. The goal is to maximize the length of an episode, by avoiding crashing into walls and staying "alive" longer than the other player. We trained agents, controlled by genetic programs, to play Tron by engaging in a co-evolutionary learning process that matched agents against humans. While this experiment did not model human behaviors directly, the game logs were mined in subsequent experiments, described next.

Follow-on work [9] introduced the idea of using logs of humans' actions, collected during the game play mentioned above, to train agent-based models of the human players. These "trainees" were represented as neural networks with inputs corresponding to agents' sensors and outputs corresponding to actions. The training process reconstructed previously-played games, following the moves of both players recorded in game logs. The human player is designated as the "trainer", and the "trainee" network tries to replicate the human's actions. Before each action taken by the human trainer, the current state of the game arena is input to the trainee agent's network, which predicts which action to take. Then the prediction is compared to the trainer's action recorded in the log. If the actions match, then the training sequence proceeds to the next move in the game. If the actions do not match, then the trainee's network weights are adjusted (using back-propagation [13]), in proportion to the prediction error. This process continues until the trainee can predict its trainer's actions with accuracy.

We produced 58 different human-trained networks in this manner. For comparison, we produced a control group of networks trained by deterministic agents (i.e., players that always perform the same action in a given state). To evaluate the trainees, we ran test games pitting the human-trained networks, their trainers, the control group and their trainers against a common group of deterministic-behavior opponents (different from the control group). In addition, we also ran test games pitting the common group against a set of random players (i.e., players that select their actions randomly). We looked for correlation between the actions selected by trainee-trainer pairs, as well as pairs between the trainees and the random players. The correlation between trainees and their trainers was significantly higher than the correlation between trainees and random players, for both types of trainees. The correlation for trainees that trained on deterministic players was higher than for those that trained on human (non-deterministic) players. This is not unexpected, since humans naturally produce noisy behaviors. We did some comparative analysis of the variance in the trainees' behavior patterns, next to that of the human trainers, and we found that the trainees were able to filter out spurious actions of the humans and behave less erratically while still exhibiting a comparable range of normal behaviors.

After the initial Tron work, we subsequently applied this method, training agents from game logs, to other environments, including educational games [14]. Current work is applying this technique to modeling humans playing more complex games (Tetris [15] and Poker [16]) and to modeling behaviors of humans performing a task in which they assign keywords to natural language data sets that discuss medical conditions. In addition, we are applying this approach using a *learning from demonstration* algorithm [17] in a robotics domain, facilitating a human teacher to train a robot to avoid obstacles [11].

The *SimEd classroom* project [6] involves emulating the outcomes of humans in a simulated learning environment in which different types of behaviors are exhibited by agent-based models of students and teachers. The knowledge being acquired by these simulated students is an abstract set of concepts, represented symbolically: $\{c_0, c_1, \ldots c_{n-1}\} \in \mathcal{C}$. A state in this environment is defined by the subset of concepts ($\subseteq \mathcal{C}$) that the simulated teacher has presented and the simulated student has acquired. (The teacher agents are assumed to "know" all the concepts in $\mathcal{C}$.) The teacher performs two tasks: *act*, by selecting a concept $c_i \in \mathcal{C}$ to present to the student, with the ultimate goal of helping the student acquire all elements in $\mathcal{C}$; and *react*, by evaluating the student's response to c_i. There are $n = |\mathcal{C}|$ choices of action and 2 choices of reaction ("right" or "wrong"). In between, the student acts by selecting a response to c_i which results in either acquiring c_i or not, or acquiring some portion of c_i. The student's action choices are abstracted into degrees of correct knowledge about the current c_i, depending on the value of a *learning rate* parameter.

We applied a symbolic approach to modeling the behavior of individual humans in the *SimEd classroom* project [6]. Agent models of students and teachers were derived from theories of teaching and learning described in pedagogical literature (e.g., [18–20]). Three different teacher behavior models were compared.

In the *lecture model*, the teacher agent selects new c_i's regardless of students' progress. In the *lecture-feedback model*, the teacher selects new c_i's when the majority of students had acquired previously-presented c_j's ($i \neq j$). In the *tutorial model*, the teacher selects customized c_i's for students, based on individual progress with previously-presented c_j's. The student model was derived from the *trilogy of mind* theory of knowledge acquisition [18], where the probability of action selection is computed according to the student's current set of beliefs (*cognition*) and attitude (*emotion* and *motivation*). The simulation results were not surprising: the tutorial model produced the best learning outcomes, in terms of acquisition of more concepts by more students; and students with higher cognitive ability and stronger motivation to learn acquired more concepts.

3 Modeling the outcomes of group behaviors

One of our primary motivations for investigating human behavior modeling is not to mimic individuals, but rather to emulate *groups* of humans. We analyzed data collected in the first 15 months of the Tron experiment, examined the "win rates" of the human players (i.e., the number of games won divided by the number of games played) and categorized them into clusters based on win rate (e.g., humans who win 10% of their games, 50%, 90%, etc.). We found that the distribution of clusters remained relatively constant over time, despite the steady increase in the number of different people playing games on the site; i.e., although the number of players increased, the proportions of "terrible" (10% win rate), "fair" (50% win rate) and "expert" (90+% win rate) players remained the same. We wanted to see if we could train agents to imitate each group—could we create a terrible, a fair and an expert agent player? If we could, then we would have a suite of players who could be matched against humans who were learning to play and provide a scaffolded series of opponents to aid human learners.

We repeated the process used for training on individuals' data, but instead of using data from one human, we clustered human players based on their win rates and used the combined game logs from all the humans in the same cluster as input [10]. These training data sets were naturally larger than those used for training on individuals' data, and the results were better. Despite the fact that the training data was a merged set of actions from multiple people's games, the result was a suite of agents that could reliably play at the levels matching their trainers. We repeated this process within the educational games environment mentioned above [10, 14]. The results surpassed those from Tron because the educational games were in static environments and the state space was smaller.

Subsequently, also part of the *SimEd* project, we applied the technique to modeling children's behaviors in an interactive educational assessment environment [2] and experimented with different ways of grouping the human trainers. The domain, an educational assessment instrument, consisted of a sequence of questions and was administered to elementary school children in order to screen for learning disabilities. The assessment was adaptive: after an initial fixed sequence, questions were chosen dynamically, based on students' answers. In the data we studied, students responded to a subset of 7-8 questions, out of 94. A student's pattern of right and wrong answers governed her "trajectory" through the assessment landscape (i.e., the subset and sequence of questions answered).

We built agent-based models of students from logs that recorded 117 students' answers to questions in the assessment, and experimented with two different measures for clustering the training sets. One method was to create a 94-element feature vector in which each question was encoded as "right", "wrong" or "not seen" and classify students based on the Euclidean distance between feature vectors. The second method was to measure the distance between trajectories in the landscape using the Hausdorff geometric distance measure [21]. We applied a hierarchical clustering technique to both measures, generating two different partitions of student transaction logs. Then we trained agents to emulate the outcomes of the students in each group. The agents were represented using probabilistic influence networks [3], and were evaluated by deploying in a simulated version of the educational assessment and logging answers. We compared the correlation between trainee-trainer pairs resulting from each clustering measure, indicating how closely the trainees replicated the question-answering behavior of their trainers. Our results showed that the trajectory-based clustering produced superior correlation to the feature-based clustering. We believe this is because the trajectory method takes into account the sequence and dependencies between questions, whereas the feature method views questions independently.

In follow-on work to the SimEd classroom project, we simulated learning in groups of human students [7]. Based on multiple pedagogical theories of group learning, we engineered a model in which simulated students are presented with concepts (as in [6]) and interact in groups in order to acquire the concepts. We experimented with different compositions of "low" and "high" ability learners in groups and different reward structures ("individual", "competitive" and "cooperative"), to determine the combination that produced the best results in terms of acquisition of more concepts by more students. The competitive and cooperative reward structures assigned the same reward to all members of the group based on how the group performed in relation to other groups. The results demonstrated that cooperative reward structures help low ability learners progress as part of a heterogeneous group more rapidly than other reward structures or homogeneous group assignments. This mirrors reports in pedagogical literature that describe observations of human classrooms.

4 Modeling the process of individuals' behaviors

Although we have primarily focused on constructing models that emulate the *outcomes* of human behaviors, some of our more recent work has involved emulating the *process* of humans selecting actions (i.e., model type IV). In our *MBTI* project [8, 22], we apply the *Myers-Briggs Type Indicator* theory of human personality [23] to agents acting in a simple artificial life environment [24]. This theory defines human personalities along 4 axes: extroversion versus introversion; sensing versus intuition; thinking versus feeling; and judging versus perceiving.

Each of these dichotomies influence how humans interpret input, interact with others and make decisions. By recognizing a tendency for one extreme within each dichotomy, the MBTI theory classifies people into sixteen different "personality types". In our simulated environment, we created agents of each type, who make different action-selection decisions according to their personality type. Our results demonstrate different outcomes that reflect those personality types. For example, agents with introverted personality types, who lose energy when interacting with others, are less efficient in completing tasks in densely populated environments as compared to sparsely populated environments.

5 Summary

This position paper has proposed a unified theory of human-agent modeling that defines two major categories of approaches: emulating the *outcome* of human behaviors and emulating the *process* of human behaviors. The outcome-based approaches can be further sub-divided into models based on individuals or groups of humans. Four types of models have been described, ranging in levels of abstraction from the most coarse, where only start and end states are modeled, to most specific, where the process of action-selection is mimicked. Two techniques for model representation were discussed: symbolic and connectionist. Examples of each were highlighted from previous and ongoing projects.

A number of key lessons have been learned. First, training agents on human transaction logs results in controllers that filter out anomalous actions and produce more consistent patterns of behavior, i.e., with less variance in action selection than human trainers. Second, constructing computational models from theories suggested in literature requires careful validation efforts, which are not necessarily easy to design and may be controversial. Finally, creating graphical representations of state and action landscapes opens the door to a wide range of methodologies that can be used to compare and assess behavior models.

Current efforts involve deeper exploration of this unified theory by reviewing a wide range of behavior modeling projects and assessing the broader applicability of our categorization. As well, application of the theory to more complex domains—some mentioned here—is underway. This includes investigation of a mixed-model-type approach where models of types III and IV are combined by overlaying personality type on agents in the SimEd and gaming environments.

Acknowledgments

I am grateful for the invaluable input of the many graduate students who have worked with me on the projects described in this paper (Kenneth Auyeung, Mathew Davies, Ilknur Icke, J. Pablo Munoz, Jeffery Raphael, Jordan Salvit, and Maartje Spoelstra), as well as my collaborators on the original Tron project (Pablo Funes, Hugues Juillé, Jordan Pollack and Alan Blair). This work has been supported by a number of grants from the National Science Foundation (ITR-0113317, ITR-0219347, ITR-0552294, IIP-0637713, IIS-1117000).

References

1. Tang, Y., Parsons, S., Sklar, E.: An agent-based model that relates investment in education to economic prosperity. In: Autonomous Agents and Multiagent Systems (AAMAS). (2007)
2. Sklar, E., Icke, I.: Using simulation to evaluate data-driven agents. Multi-agent Based Simulation (MABS) **LNAI 5269** (2009)
3. Pearl, J.: Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann (1988)
4. Russell, S., Norvig, P.: Artificial Intelligence. Prentice Hall (2003)
5. Pollack, J.B., Blair, A.D.: Co-evolution in the successful learning of backgammon strategy. Machine Learning **32** (1998)
6. Sklar, E., Davies, M.: Multiagent simulation of learning environments. In: Autonomous Agents and Multi Agent Systems (AAMAS). (2005)
7. Spoelstra, M., Sklar, E.: Using simulation to model and understand group learning. Agent Based Systems for Human Learning, International Transactions on Systems Science and Applications **4**(1) (2008)
8. Salvit, J., Sklar, E.: Toward a myers-briggs type indicator model of agent behavior. Multi-Agent-Based Simulation (MABS) **LNAI 6532** (2010)
9. Sklar, E., Blair, A.D., Funes, P., Pollack, J.: Training Intelligent Agents Using Human Internet Data. In: Intelligent Agent Technology (IAT). (1999)
10. Sklar, E., Blair, A.D., Pollack, J.B.: Ch 8: Training Intelligent Agents Using Human Data Collected on the Internet. In: Agent Engineering. World Scientific (2001)
11. Munoz, J.P., Ozgelen, A.T., Sklar, E.: Learning from Demonstration in Spatial Exploration. In: AAAI Robotics Workshop and Exhibition. (2011)
12. Funes, P., Sklar, E., Juillé, H., Pollack, J.B.: Animal-Animat Coevolution: Using the Animal Population as Fitness Function. In: Simulation of Adaptive Behavior (SAB), MIT Press (1998)
13. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Nature **323** (1986)
14. Sklar, E.: CEL: A Framework for Enabling an Internet Learning Community. PhD thesis, CS Dept, Brandeis Univ, Advisor: J. Pollack (2000)
15. Raphael, J.: Masters thesis proposal: Social learning as a method of knowledge transfer for a tetris agent. CIS Dept, Brooklyn College, CUNY, Advisor: E. Sklar (Jan 2012)
16. Auyeung, K.Y.: Masters thesis proposal: A neural network approach to developing robust poker strategies. CIS Dept, Brooklyn College, CUNY, Advisor: E. Sklar (Jan 2012)
17. Chernova, S., Veloso, M.: Interactive policy learning through confidence-based autonomy. JAIR **34**(1) (2009)
18. Lazarus, R.S.: Cognition and Motivation in Emotion. American Psychologist **46**(4) (1991)
19. Gardner, H.: Frames of Mind. BasicBooks (1983)
20. Papert, S., Harel, I.: Constructionism. Ablex (1991)
21. Rote, G.: Computing the minimum hausdorff distance between two point sets on a line under translation. Information Processing Letters **38** (1991)
22. Salvit, J.: A Myers-Briggs Type Indicator Model of Agent Behavior. Master's thesis, CIS Dept, Brooklyn College, CUNY, Advisor: E. Sklar (2012)
23. Myers, I.B., Myers, P.B.: Gifts Differing. Consulting Psychologists Press (1980)
24. Resnick, M.: Turtles, Termites, and Traffic Jams: Explorations in Massively Parallel Microworlds. MIT Press (1997)