Automatic model construction for the behavior of human crowds

Jinghui Zhong [a,b,*], Wentong Cai [b], Michael Lees [c,d], Linbo Luo [e]

[a] *South China University of Technology, School of Computer Science and Engineering, Guangzhou, China*
[b] *Nanyang Technological University, School of Computer Science and Engineering, Singapore*
[c] *University of Amsterdam, Computational Lab, Amsterdam, Netherlands*
[d] *National Research University ITMO, St. Petersburg, Russia*
[e] *Xidian University, School of Cyber Engineering, Xi'an, China*

ABSTRACT

Designing suitable behavioral rules of agents so as to generate realistic behaviors is a fundamental and challenging task in many forms of computational modeling. This paper proposes a novel methodology to automatically generate a descriptive model, in the form of behavioral rules, from video data of human crowds. In the proposed methodology, the problem of modeling crowd behaviors is formulated as a symbolic regression problem and the self-learning gene expression programming is utilized to solve the problem and automatically obtain behavioral rules that match data. To evaluate its effectiveness, we apply the proposed method to generate a model from a video dataset in Switzerland and then test the generality of the model by validating against video data from the United States. The results demonstrate that, based on the observed movement of people in one scenario, the proposed methodology can automatically construct a general model capable of describing the crowd dynamics of another scenario in a different context (e.g., Switzerland vs. U.S.) as long as that the crowd behavior patterns are similar.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Agent-based modeling is a popular and powerful crowd modeling paradigm that has a range of applications in digital game design, military training, crowd evacuation planning, etc. [1–6]. In the agent-based paradigm, individuals are modeled as autonomous agents that can perceive information and make decisions based on individual behavioral rules. A fundamental issue in all agent-based modeling is how to design suitable behavioral rules so that the simulation can match the desired crowd behaviors (e.g., those observed in videos). Once the correct behavioral rules are obtained, the crowd simulation model can be used for prediction and performing "what-if" analysis. However, designing correct behavioral rules to generate a desired crowd behavior remains a challenging task because an agent's behavior is affected by its environment as well as the surrounding agents. Typically, this is done in a fairly ad-hoc manner, and relies on expert intuition. However, intuition can often be misleading as crowds are complex systems and as such the relationships between the individual behavioral rules and the collective crowd dynamics are often unintuitive.

There are a number of existing models that have taken a scientifically Popperian approach to understanding crowd dynamics. A theory is developed which can then be falsified against real data. Helbing et al. [7,8] suggested using virtual force to guide pedestrians' motions. They designed mathematical equations to formulate the attractive and repulsive forces from goals, other pedestrians and obstacles. Yamaguchi et al. [9] designed behavioral rules using an energy function that considers multiple personal and social factors such as damping and grouping. Similarly, Pellegrini et al. [10], Scovanner and Tappen [11] also designed energy functions by considering different personal and social factors to determine movements of pedestrians. These manual methods usually require a model calibration process to tune coefficients of the initialized formulas, so that the model dynamics can match the desired crowd dynamics [9,12,10,11,13,14]. However, if the initialized formulas are not well-defined, the calibration will not be able to generate the desired simulation results. In addition, the model calibration is also a time consuming and tedious process.

Recently, more Baconian, or data-driven approaches, have been proposed to automate the generation of crowd behaviors. The key idea is to learn examples (mostly in terms of state-action pairs) from video data. These examples are then used as behavioral rules in the simulation to generate agents' movement [15–20]. In the computer vision community, several approaches have been proposed that learn global motion patterns (e.g., certain

movement curves) of crowds from video data [21–25]. The global motion patterns can be integrated with a crowd simulation model to generate realistic crowd behaviors. However, the knowledge learned by these methods (e.g., examples and global motion patterns) are typically highly scenario specific where training data is available. Once the application scenario changes, the learned knowledge may become incapable of generating the desired crowd dynamics.

In this paper and in previous work [26] we aim to develop a way of merging the Baconian and Popperian philosophies, by developing ways to automate model (or rule) generation. Effectively we allow the computer to find sets of plausible theories (models) that we can read and understand and perhaps manipulate later. In summary the core question of this paper is: *"Can we design an automatic method to learn generic behavior rules from video data?"* Here the term "generic" means that the behavioral rules are applicable not only to the scenario where training data are obtained from but also to other new scenarios. Also the notion of rule is critical, whereas a pure data-mining approach may be able to derive a model (e.g., trained neural network), the notion of a rule implies that humans can understand and perhaps modify the model.

Specifically, this paper focuses on simulating crowd behaviors in an area with one or multiple source regions (SRs) and destination regions (DRs). Each pedestrian in the given area keeps moving toward its own DR and leaves the area via the DR. Such a scenario is quite common in public places, e.g., train stations, airports, etc. We assume that the environment contains obstacles and that pedestrians move toward their DRs casually, following certain curved paths (e.g., "S" and "C" shape curves). The movement curves are determined by specific environment factors such as the positions of obstacles and the destinations. Our goal is to automatically learn a generic behavioral rule (labeled as Γ) from video data which describes the way pedestrians move toward their DRs while avoiding obstacles.

In the proposed methodology, an abstract dual-layer architecture is used to model individual behaviors. The bottom layer adopts the commonly used social force model (SFM) [8] to model the collision avoidance behaviors of pedestrians. Meanwhile, in the top layer, a behavioral rule that defines velocity fields is used to determine the strategic or goal-directed movement of pedestrians. A velocity field is a set of position-velocity pairs that describe the velocities of pedestrians in the simulation regions [27,28,20]. Based on the dual-layer modeling architecture, the problem of finding a suitable behavioral rule to generate the desired crowd behaviors is formulated as a symbolic regression problem. By using a genetic programming (GP) [29] based method to solve the symbolic regression problem, a behavioral rule that leads to the desired simulation results can be obtained automatically. To validate its effectiveness, the proposed methodology is applied to learn Γ from video data of a scenario in Switzerland. The learned Γ is then used to generate crowd dynamics in a new scenario in United Stated. The results have demonstrated that the proposed methodology is capable of distilling suitable behavioral rule that can generate realistic crowd dynamics in the new scenario.

This paper extends two of our previous papers on data-driven crowd modeling [26,20]. In the first paper [26], we have conducted a preliminary research on learning goal selection rules of agents from simulation data. This paper extends [26] by addressing a complementary problem – to distill generic behavioral rules to define the velocity fields that can be used to drive agents toward their goals. In the second paper [20], we have proposed a data-driven method to learn behavioral patterns (e.g., the goal selection patterns and the velocity fields) from videos so as to generate realistic crowd dynamics. However, the behavioral patterns learned in [20] are only applicable to the specific scenario where training data is obtained from. In this paper, we aims to distill generic behavior

rules that can be applied directly to different scenarios that have similar crowd behavior patterns.

In general, the major contributions of the paper are as follows. First, we have formulated the problem of distilling suitable crowd behavior rule from crowd data as a symbolic regression problem. Second, we have developed a genetic programming based framework to solve the formulated problem. The proposed framework is generic such that the solution obtained from one scenario can be applicable to other scenarios that have similar crowd behavior patterns. In addition, two real world datasets have been used to validate the effectiveness and generality of the proposed framework.

The rest of the paper is organized as follows: Section 2 describes the related work. Section 3 presents the outline of the proposed framework. Section 4 describes the procedure of applying Self-Learning Gene Expression Programming (SL-GEP) to distill behavioral rules from video data. Section 5 presents the simulation studies. Finally, Section 6 presents the conclusions and discussions.

2. Related work

Over the past decades, various methods have been proposed to help automate the process of modeling human crowds. These works can generally be classified into two groups. The first group focuses on automatic parameter calibration. The objective of this class of methods is to use optimization algorithms such as genetic algorithms (GA) to tune parameters of a model so that the simulations match the desired crowd behaviors (e.g., those observed in the video). For example, Johansson et al. [12] proposed an evolutionary algorithm (EA) to calibrate the parameters of the social-force model (SFM) [7]. They utilized the microscopic motions of pedestrians such as the moving speed and direction to evaluate the quality of simulation results. Wolinski et al. [13] suggested using EAs (e.g., GA with a local search) to calibrate crowd models such as SFM and RVO2 based on video data. They defined several metrics (e.g., speed-density relationships) for simulation evaluation. Pellegrini et al. [10] proposed a dynamic model named Linear Trajectory Avoidance (LTA) model for multi-target tracking. The LTA model consists of an energy function with six parameters to estimate the next velocities of pedestrians. A GA is used to optimize the parameters of the LTA model based on video data so that the prediction errors of the LTA model are minimized. Yamaguchi et al. [9] proposed an agent-based behavioral model to improve tracking performance. They designed an energy function that encodes various personal, social and environmental factors to determine the next velocity of pedestrians. They used a simplex algorithm to learn optimal parameters of the energy function based on training data extracted from videos. Similarly, Scovanner and Tappen [11] proposed a new energy function with 17 parameters to predict movements of pedestrians. They utilized a modified Newton's method to optimize model parameters so that the simulated movement match tracks in the videos. In [30] a density-based evaluation method was proposed to evaluate the quality of simulation output. Differential evolution (DE) was adopted to calibrate parameters of two modified SFMs so that the simulated crowd density matched real crowd density calculated based on video data. In [14], an enhanced DE algorithm that included sensitivity analysis and a local search procedure was proposed to improve the efficiency of parameter calibration. The main difference between the work described in this paper and the existing ones described above, is that these existing works focus on using data to calibrate parameters of models that are designed by humans; whereas in this paper we aim to automatically distill behavioral rules from data.

The second group of methods focuses on automating the generation of crowd behaviors from data. The core idea is to derive typical