Pedestrian Behavior Mining from Data*

Juan Galán-Páez and Joaquín Borrego-Díaz
Department of Computer Science and Artificial Intelligence – University of Seville, Seville, Spain
juangalan@us.es, jborrego@us.es

Abstract

A general -observation-based- qualitative framework to extract agent-based pedestrian behavior is presented. To extract qualitative rules from data we use tools, from Formal Concept Analysis, for implicational reasoning.

Introduction

Analysis of agent movement is a key research line in the study of (biological) complex systems. For example, the understanding the behavior of pedestrians is very important in urban design Guo and Huang (2008). This kind of problems has become an attractive field of study (see e.g. Helbing and Johansson (2014)). For example, in zones where pedestrian flow is dense, any small change in urban planning can have extreme consequences on pedestrian mobility. As it is stated in Canca et al. (2013), one of the most important goals in studies on pedestrian mobility and behavior is to evaluate the effect of new policies on pedestrian facilities before its implementation. A robust model to simulate pedestrian behavior is a good tool to prevent potential difficulties.

From the point of view of Artificial Intelligence, the pedestrian, as an agent, selects the next action from its own knowledge. Our thesis is that behavior has qualitative nature and is based on intuitive (geometrical, social, goal-driven) attributes. Thus it is interesting to explore how the reasoning with this kind of features can provide knowledge bases for modeling pedestrian behavior. The proposed methodology is based in both, Agent Based Modeling (ABM) and Formal Concept Analysis (FCA). In this work pedestrian behavior is considered individually by means of discrete ABM, where pedestrian flow emerges from interactions between pedestrians (agents) and the urban environment. The modeling is carried out from the pedestrian point of view in qualitative terms, allowing the use of reasoning and concept-mining methods in order to analyze pedestrian flow dynamics.

The aim of this paper is to show how to exploit knowledge extracted from observations (of real or artificial systems) to study and explain -in a qualitative formalism- pedestrian behavior. The result of this process is, itself, a knowledge-based system. By using this system as a deliberative module for agents, we have implemented a general simulation framework for natural and artificial models of mobility.

Basic Agent-Based Model for pedestrians

In order to show how method works, consider as simulation environment an orthogonal grid where agents can make discrete movements. In each time step agents can move to any neighboring cell (Moore Neighborhood), where the chosen movements depends on local information agents obtain from their neighborhood and possibly additional information on urban planning. The environment consists of: **free cells** (any cell on which agents can move. In the basic model two agents cannot take up the same cell. Therefore, in the basic model, an occupied cell will be considered as an obstacle), **obstacles** (cells representing buildings, street furniture and other elements), and **exits** (representing possible pedestrians destination; these can be buildings/streets that are out of the area under study which are called *exits* as agents *leave* the simulation area through them).

The agent selects the best action (movement) according to their knowledge and the information he has about their neighborhood. To exemplify the general simulation framework, three basic agent behaviors will be considered: *best movement* (agent moves always to the adjacent cell closest to destination. This behavior can lead to blocked agents in certain scenarios.), *best movement with uncertainty* (agent makes the *best movement* with probability p or a random movement with probability $1 - p$), and em any *good* movement: (agent randomly moves to any of the adjacent cells toward destination. That is to say, any cell reducing agent's Manhattan distance to destination).

It is important to note that, due to spatial distribution within the environment, it is possible that the best movement in the short term (locally) will not be the best movement in the long term. The basic model can be improved in a number of ways: 1) Larger agents' range of vision (larger neighborhood). 2) Agents have memory. They can keep information of a number of past movements. 3) Agents have the ability

*Partially supported by TIC-6064 Excellence project (*Junta Andalucía*) and TIN2013-41086-P t (Spanish Ministry of Economy and Competitiveness), cofinanced with FEDER funds

to communicate with and/or follow other agents. All these extensions can be added to the basic model by considering attributes specifying agents' new knowledge and abilities. However, the state description of an agent would be more complex and this could complicate the analysis of simulation results and the extraction of useful conclusions.

In order both to extract qualitative knowledge and to reason with this, **Formal Concept Analysis** (FCA) is used. This approach allows to extract and reasoning with knowledge from data. See Ganter and Wille (1997) and Aranda-Corral et al. (2013b) where details about the use of FCA to represent knowledge on cellular automata is presented.

Representing agent's knowledge. In the basic model the agent receives only information about the distance to its exit and neighboring obstacles. This first one is the *potential* of each cell with respect of agent destination, which evaluates the goodness of each possible movement of the agent, with respect to cells' distance to destination. In this regard, potential can be positive (the cell is closer to destination), negative (the cell is farther) or neutral. If any of the cells is an obstacle, it will not be taken into account in agent's decision.

In order to validate the proposal, experiments using two of the basic ABM for pedestrians (Best movement and Any *good* movement) have been carried out. Due to space limitations it has not been possible to consider more elaborate models on this article. Likewise two different attribute sets, based on cells' potentials, will be used as agents' (local) knowledge representation[1]:

Detailed potentials (58 attributes): These estimate how good the potential of each neighboring cell is. Despite the potential being an abstract concept, in this case, it is based only on agents' distance till destination, for example to quantify potentials in terms of cell's *Manhattan distance* till destination.

Attributes *Will-Move-To-XX(Target)* contains information on agent's next movement. These will be target attributes for the reasoning system for prediction, in which a model is built from past information on agents' behavior.

Simplified potentials (42 attributes): This set provides information only on whether potentials are positive, negative or neutral without quantifying how positive or negative they are. The rest of attributes are the same in both attribute sets.

In both attribute sets, each (neighboring) cell is identified by its relative position with respect to the agent (that is to say, {*TL, TC, TR, CL, CR, BL, BC, BR*}). For instance *TL* refers to top-left cell, *CR* refers to center-right cell and *BC* refers to bottom-center cell.

Contextual selection for pedestrians

The reasoning system from Aranda-Corral et al. (2013b) allows the use of FCA-based tools for carrying out pedestrian behavior simulations. From a contextual selection (formal context) it is possible to extract a knowledge base to be used by the reasoning system. The **contextual selection** contains information items similar (based on its context, that is, time, space and other properties) to the ones under study. Thus reasoning with this selection will provide more reliable entailment. For instance, let a be a pedestrian whose current position is known. Contextual selection to predict the next movement of a consists of:

Spatial dimension: Depending on the nature of the scenario under study it is possible to consider the whole pedestrian set or a pedestrian subset containing only those pedestrians closer to a (for big or heterogeneous scenarios), for instance, those pedestrians located in the same street as a.

Temporal dimension: A contextual selection for pedestrian dynamics usually contains information of agents' movements for more than one past time step. In order to estimate the next movement of a pedestrian a in a time step T the contextual selection for a in T will consist on other pedestrian movements in a recent time period of length W. In this way, the time window considered for the contextual selection would be $[T - W, T)$.

Attribute selection: Since different attribute sets can be used, for each setting the attribute set most suitable to be used as knowledge representation for a and its environment would be selected. In this concrete case (the pedestrian basic model), one of the two proposed attribute sets will be selected. In more complex scenarios the attribute set can include other attributes specific for the environment where a is located (spatial, temporal or from other nature).

FCA-based simulation of pedestrian flow

The process for computing the next movement, in a certain time step M, of a group of pedestrians for which past movements' information until a certain time step N, is known (where $N < M$), is as follows:

Stage 1: A formal context (the contextual selection) is built containing information on the W most recent time steps (movements) with respect to the target M. This formal context will contain, for each time step w_i and for each pedestrian, an object describing pedestrian's neighborhood at time step w_i and its next movement at time step (w_{i+1}).

Stage 2: From this formal context, the knowledge base is extracted. According to the nature of the experiment, Stem basis or Luxenburger basis can be used.

Stage 3: In order to predict agent's next movement, a reasoning process is carried out. The initial facts consist of an attribute set describing agent's neighborhood at time M. The next movement at time step $M + 1$ is extracted from the entailment obtained by automated reasoning.

Soundness of the model. The standard way to specify (deterministic, reactive) agents includes $\langle S, T, Act, P, Do, acc\rangle$ with $Perceive : S \rightarrow T$, $Do : Act \times S \rightarrow S$ and $acc : T \mapsto Act$. Here S is the set

[1]Standard qualitative attributes have been selected to show the method. The attribute selection can be expanded by adding any (computable) attribute the observer finds relevant

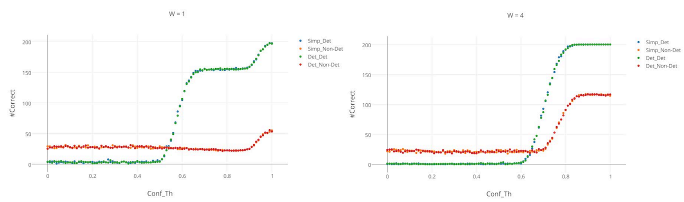

Figure 1: Results with different temporal windows

of states, T is a partition of S (due to perception features of the agent), Act is the actions set, and $Perceive, Do, acc$ are functions: P determines the agent state, Do determines the effect of an action on a state (reaching other state) and acc selects the action to be executed in a certain state. Agent execution from $\sigma(0)$ is the sequence $\{\sigma(t)\}_{t\in\mathbb{N}}$ where $\sigma(t+1) = Do(acc(Perceive(\sigma(t))), \sigma(t))$.

Given an attribute selection A, it is said that A is *descriptive* for the agent specification if each state of $t \in T$ can be interpreted as a set of attributes t^A of A, and for each $\alpha \in Act$ an attribute $\alpha^A \in A$.

Let $\mathcal{S}_n^K$ denote the subset of implications with positive support of the Stem basis for the context M_n^K which contains every observation of the history of the system from the initial state till the $\sigma(n)$ in $I_K = [-K, K] \times [-K, K] \subset \mathbb{Z}^2$ with a descriptive attribute set A. A distribution is a map $\Delta : \mathbb{Z}^2 \to \{\, agent,\ obstacle,\ free,\ exit\}$.

Lastly, denote by $s \in_\Delta X$, where $s \in T$ and $X \subset \mathbb{Z}^2$, the fact that there exists a cell c in X such that if an agent is located in $c \in \mathbb{Z}^2$ with the distribution $\Delta(X)$ then the agent perceives s. A distribution Δ is T-complete if $\{s : s \in_\Delta \Delta(\mathbb{Z}^2)\} = T$. A preliminary result on the soundness of the model can be stated as follows:

Theorem. 1. *Let Δ be a distribution and A be a finite descriptive attribute set for T (T finite). Suppose that Δ is T-complete and agents share the specification $\langle S, T, Act, P, Do, acc\rangle$. Then there exists $K > 0$ such that for all $n \in \mathbb{N}$*

$$acc(Perceive(\sigma(n)))^A \in S_n[Perceive(\sigma(n))^A]$$

Therefore the implication basis is sound to simulate agent behavior by means of a deliberative process. Note that the result does not give any estimation for $K = K(\Delta, acc)$.

Experiments and conclusions

In order to experiment with the methodology above described, a simulation platform has been developed. This consists of two modules, the first one comes with a NetLogo-based simulation viewer and is used for preliminary tests, and the second one focuses on computing massive simulations and is used for full experiments.

Due to the lack of space, only few of experiments(Fig. 1) are mentioned on a squared grid with 625 cells (25 per side) populated by 200 agents. In order to show the importance of the amount of information considered, results of experiments for different window sizes W are provided (see Fig. 1, where $W = 1$ (left) and $W = 4$ (right)). Results of four different simulations are shown, one for each of the two possible knowledge representations (detailed or simplified) and one for each of the two possible pedestrian models (*best movement* or *any good movement*).

In each experiment a knowledge base is built from observable information in time steps $[T - W, T)$, and used (after selecting the implications with confidence greater than a certain threshold C_{th}) to predict agents' movement in time step $T + 1$. Results show the mean number of properly predicted movements. Each experiment is repeated for different values of the confidence threshold ($C_{th} = [0, 1]$) and $N = 100$ times for each value, in order to obtain a reliable estimate.

From experiments we can conclude that there is not substantial difference between the two representations used. It is worthy to note that a small uncertainty in agents' behavior (*any good movement*) leads to a great increase in the error. The reason is that a step-by-step performance evaluation is too strict for non-deterministic behaviors.

The work is based on a general hybrid approach to phenomenological reconstruction of Complex Systems (CS), using FCA as main tool for conceptual data mining (see Aranda-Corral et al. (2013a)). In Aranda-Corral et al. (2013b), the idea was applied to a classic CA (Conway's game of Life). The approach presented in this work specifies and implement a general method for movable agents. The key advantage of this method is that the observer can select (computable, qualitative) attributes in order to understand (and model) pedestrian behavior. The selection can comprise any feature on both, pedestrians and streets. From this selection, our method provides a qualitative model.

References

Aranda-Corral, G. A., Borrego-Díaz, J., and Galán-Páez, J. (2013a). Complex concept lattices for simulating human prediction in sport. *J. Syst. Sci. and Complexity*, 26(1):117–136.

Aranda-Corral, G. A., Borrego-Díaz, J., and Galán-Páez, J. (2013b). Qualitative reasoning on complex systems from observations. In *Hybrid AI Systems*, volume 8073 of *Lecture Notes in Computer Science*, pages 202–211. Springer.

Canca, D., Zarzo, A., Algaba, E., and Barrena, E. (2013). Macroscopic attraction-based simulation of pedestrian mobility: A dynamic individual route-choice approach. *European J. Op. Res.*, 231(2):428 – 442.

Ganter, B. and Wille, R. (1997). *Formal Concept Analysis: Mathematical Foundations*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1st edition.

Guo, R.-Y. and Huang, H.-J. (2008). A modified floor field cellular automata model for pedestrian evacuation simulation. *Journal of Physics A*, 41(38):385104.

Helbing, D. and Johansson, A. (2014). Pedestrian, crowd, and evacuation dynamics. In Meyers, R. A., editor, *Encyclopedia of Complexity and Systems Science*, pages 1–28. Springer.