Synthetizing Qualitative (Logical) Patterns for Pedestrian Simulation from Data

Gonzalo A. Aranda-Corral[1], Joaquín Borrego-Díaz[1], and Juan Galán-Páez[2,3](✉)

[1] Department of Information Technology, Universidad de Huelva, Palos de La Frontera, Spain
{garanda,jborrego}@us.es
[2] Department of Computer Science and AI, Universidad de Sevilla, Sevilla, Spain
juangalan@us.es
[3] Datrik Intelligence, Sevilla, Spain
http://www.datrik.com

Abstract. This work introduces a (qualitative) data-driven framework to extract patterns of pedestrian behaviour and synthesize Agent-Based Models. The idea consists in obtaining a rule-based model of pedestrian behaviour by means of automated methods from data mining. In order to extract qualitative rules from data, a mathematical theory called Formal Concept Analysis (FCA) is used. FCA also provides tools for implicational reasoning, which facilitates the design of qualitative simulations from both, observations and other models of pedestrian mobility. The robustness of the method on a general agent-based setting of movable agents within a grid is shown.

Keywords: Agent-based modelling · Knowledge acquisition · Qualitative spatial reasoning · Formal concept analysis

1 Introduction

Understanding the behaviour (movement) of pedestrians in various situations is an important task in order to design and improve public places such as waiting rooms in railway stations, streets, etc. [15]. The study of problems related to pedestrian flow has become an attractive field of study (see e.g. [16]), due to the fast development (in terms of size and population density) of big cities in recent years. 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 [9], one of the most important goals in studies on pedestrian mobility and behaviour is to evaluate the effect of new policies on pedestrian facilities before its implementation. A robust model able to simulate pedestrian behaviour is a good tool to prevent future difficulties.

According to [10], Cellular Automata (CA) and ordinary differential equation (ODE) models present two major differences: movement in CA is restricted to a

-

grid and navigation is achieved by moving directly in the desired direction. Force-based ODE models operate in a continuous space and navigation is computed indirectly through the acceleration vector. A great number of existing models use simulations methods inspired by physical notions applied to agents [17], while others are based in Multi-agent simulations (even by exploiting complex systems emergence phenomena), see e.g. [20].

A plethora of models and systems have been designed, focused on a number of features and considerations (cf. [21]). The qualitative nature of pedestrians' behaviour invites to combine agent based models with rational deliberative empowerment. From the point of view of (symbolic) Artificial Intelligence, the pedestrian, as an agent, selects the next action from its own knowledge. The idea is that pedestrian behaviour has qualitative nature and is based on intuitive (geometrical, social, goal-driven) attributes. Therefore it is interesting to explore how the reasoning with this kind of features can provide knowledge bases for modelling pedestrian behaviour in a deliberative manner. The proposed methodology is based in both, Agent Based Modelling (ABM) and Formal Concept Analysis (FCA).

In this work, pedestrian behaviour is considered individually by means of discrete ABM, where pedestrian flow emerges from interactions between agents and the urban environment. The modelling is carried out from the pedestrian point of view in qualitative terms, allowing the use of reasoning and concept-mining methods in order to analyse pedestrian flow dynamics.

1.1 Aim of the Paper

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 behaviour. The result of this process is, itself, a knowledge-based system that is also useful to simulate the source 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.

This work is based on the intensive use of Formal Concept Analysis [12] which provides mathematical tools for detecting qualitative concepts, useful in the *phenomenological reconstruction* of CS [5] (in this case, related to pedestrian mobility).

1.2 Structure of the Paper

The structure of the paper is as follows. The next section is devoted to recall the basic pedestrian model based on cells. In Sect. 3 basic concepts on Formal Concept Analysis are summarized, as well as the notion of *attribute selection*. Section 4 presents FCA as a tool for agents' knowledge modelling. The notion of contextual selection (induced attribute selection and spatial-temporal features) is described in Sect. 5. The FCA-based simulation model (and some considerations on experimentation) is presented in Sect. 6. Section 7 shows the robustness of the model. Finally, related and future work enumerated (Sects. 8 and 9).

2 Basic Agent-Based Model for Pedestrians

The simulation environment consists of an orthogonal grid where agents can make discrete movements. In each time step agents can move to any neighbouring cell (Moore Neighbourhood), where the chosen movements depends on local information that agents obtain from their neighbourhood and possibly additional information on urban planning (see Fig. 3). The simulation environment consist of the following elements:

- **Free cells:** Any of the cells to which agents can move. In the basic model two agents cannot take the same cell up. Therefore, in the basic model, an occupied cell will be considered as an obstacle.
- **Obstacles:** Cells representing buildings, street furniture and other elements to which agents cannot move.
- **Exits (destination):** Cells representing possible pedestrians destination. These can be buildings or streets that are out of the area under study. In this work, these destinations are called *exits* as agents *leave* the simulation area through these cells.
- **Agent (pedestrian):** they select the best action (movement) according to their knowledge and the information they have about their neighbourhood. In order to validate the general simulation framework, three basic agent behaviours will be considered:
  - **Best movement:** Agent moves always to the adjacent cell closest to destination (the best choice in the short term). This behaviour 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 $1 - P$.
  - **Any *good* movement:** Agent randomly moves to any of the adjacent cells towards destination. That is to say, any cell reducing agent's Manhattan distance to destination.

It is important to note that, due to some elements' spatial distribution within the environment, it is possible that the best movement in the short term (locally) results in a bad movement the long term. The basic model can be improved in a number of ways: (1) Larger agents' range of vision (larger neighbourhood). (2) Agents have memory. They can keep information of a number of past movements. (3) Agents have the ability 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. Finally, it is worthy to mention that the deduction algorithm is based on Horn-like propositional clauses.

3 Background: Formal Concept Analysis

In this section, we summarize the basic notions of FCA. For detailed information, we refer to [12]. The input to FCA is an object-attribute data table describing which objects have which attributes. Such table can be identified with $\mathbb{K} = (G, M, I)$ where G is a finite non-empty set of objects, M is a finite non-empty set of attributes, and $I \subseteq G \times M$ is an (object-attribute) relation. The objects and attributes correspond to table rows and columns, respectively, and $(g, m) \in I$ indicates that object g has attribute m. In terms of FCA, $\mathbb{K}$ is called a formal context. Figure 1 shows a formal context describing fish (objects) living on different aquatic ecosystems (attributes).

Every data table (G, M, I) induces a pair of so-called concept-forming operators, defined for each $A \subseteq G$ and $B \subseteq M$ by

$$A' = \{a \in M \mid (o, a) \in I \text{ for all } o \in X\}$$

and

$$B' = \{o \in G \mid (o, a) \in I \text{ for all } a \in Y\}$$

A (formal) concept is a pair (X, Y) such that $X' = Y$ and $Y' = X$. The set of concepts of a context given M can be endowed with the lattice structure by means of the "subconcept" relationship [12]. For example, the concept lattice from the formal context on fish of Fig. 1 (attributes are understood as "live in") is shown in Fig. 2. In this representation each node is a concept, and its intension

	River	Coast	Sea
Carp	X		
Escatofagus	X	X	
Bream		X	X
Sparus		X	X
eel	X	X	X

Fig. 1. Formal context on fish

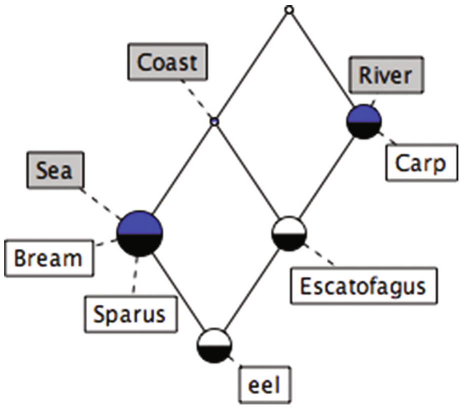

Fig. 2. Concept lattice associated with the formal context of Fig. 1

(resp. extension) is formed by the set of attributes (resp. objects) included along the path to the top (resp. bottom) concept. For example, the bottom concept

$$(\{eel\}, \{Coast, Sea, River\})$$

is the concept *euryhaline fish*.

3.1 Implication Basis

Knowledge Bases in FCA are formed by *implications between attributes*. An attribute implication (over a set M of attributes) is an expression $L = A \to B$ where $A, B \in 2^M$ (2^X stands for the power set of X). The set of implications over a set M is denoted by $Imp(M)$. It is denoted by $att(L) = A \cup B$, and being $\mathcal{L}$ a set of attributes, $att(\mathcal{L}) = \bigcup_{L \in \mathcal{L}} att(L)$.

It is said that $A \to B$ is valid in a set $T \subseteq M$ of attributes (or T is a model of the implication, also it is called T respects the implication), written $T \models A \to B$, if the following condition is satisfied: If $A \subseteq T$ then $B \subseteq T$. $\mathbb{K} \models A \to B$ if $\{g\}' \models A \to B$ for any $g \in G$. In this case it is said that implication $A \to B$ is valid in the context $\mathbb{K}$.

By simplicity we suppose throughout the paper that implications $A \to B$ satisfy $A \cap B = \emptyset$. $\top = \emptyset \to \emptyset$ denotes an implication that is always true. In fact, if $\mathcal{L} \models \top$ then $\mathcal{L}$ only contains implications like $A \to \emptyset$.

Definition 3.1: Let $\mathbb{K} = (G, M, I)$ be a formal context, $\mathcal{L} \subseteq Imp(M)$ and L be an implication.

(1) L follows from $\mathcal{L}$ (or L is consequence of $\mathcal{L}$, denoted by $\mathcal{L} \models L$) if each subset of M modelling $\mathcal{L}$ also models L.
(2) $\mathcal{L}$ is complete for $\mathbb{K}$ if the following condition is satisfied for every implication L:

$$\text{If } \mathbb{K} \models L \text{ then } \mathcal{L} \models L$$

(3) $\mathcal{L}$ is non-redundant if for each $L \in \mathcal{L}$, $\mathcal{L} \setminus \{L\} \not\models L$.
(4) $\mathcal{L}$ is a (implication) basis for $\mathbb{K}$ if $\mathcal{L}$ is complete for $\mathbb{K}$ and non-redundant.

Given $L \in Imp(M)$, the set of models of L is denoted by $Mod(L)$ (resp. $Mod(\mathcal{L})$ for a set of implications). Therefore L is **consequence** of a set of formulas $\mathcal{L}$ if and only if $Mod(L) \subseteq Mod(\mathcal{L})$. Given $\mathcal{L}'$ other implication set, it is denoted by $\mathcal{L}' \trianglelefteq \mathcal{L}$ when $\mathcal{L} \models \mathcal{L}'$, and $\mathcal{L}' \triangleleft \mathcal{L}$ if $\mathcal{L} \models \mathcal{L}'$ but $\mathcal{L}' \not\models \mathcal{L}$. It is said that $\mathcal{L}$ and $\mathcal{L}'$ are equivalent, $\mathcal{L}' \equiv \mathcal{L}$, if $\mathcal{L} \models \mathcal{L}'$ and $\mathcal{L}' \models \mathcal{L}$.

A particular basis is the *Duquenne-Guigues* or so called *Stem* Basis (SB) [14]. In order to work with formal contexts, stem basis and association rules, the Conexp[1] software has been selected. In forecasting and other data analysis tasks within Complex System dynamics, given in [2], the reasoning is performed by means of a production system that works with sets of implications. The entailment relationship based on this method is denoted by $\vdash_p$. Formally:

[1] http://sourceforge.net/projects/conexp/.

Definition 3.2: Let $\mathcal{L} = \{A_i \to C_i,\ i \in I\} \subseteq Imp(M)$ and $H \subseteq M$. The **implicational closure** of H with respect to $\mathcal{L}$, $\mathcal{L}[H]$, is the smaller set $B \subseteq M$ such that:

- $H \subseteq B$
- If there exists $i \in I$ such that $A_i \subseteq B$, then $C_i \subseteq B$

The relation $\vdash_p$ is the inference relationship induced by the closure above defined: given $A \subseteq M$,

$$\mathcal{L} \cup H \vdash_p A \quad \overset{def.}{\Longleftrightarrow} \quad A \subseteq \mathcal{L}[H]$$

The computing of $\vdash_p$-closure is carried out, given $\mathcal{L}$ y H, by computing the sets:

- $\mathcal{L}_0 = H$
- $\mathcal{L}_{k+1} = \mathcal{L}_k \cup \{a \in A\ :\ \exists i \leq n \text{ such that } A_i \subseteq \mathcal{L}_k \text{ and } a \in C_i\}$

Then $\mathcal{L}[H] = \bigcup \mathcal{L}_i = \{a \in M\ :\ \mathcal{S} \cup H \vdash_p a\}$.

The soundness and completeness with respect to the entailment is based on the following result (see [1] for details):

Theorem 3.3: Let $\mathcal{S}$ be a basis for $\mathbb{K} = (G, M, I)$ and $\{a_1, \ldots, a_n\} \cup Y \subseteq M$. The following statements are equivalent:

(1) $\mathcal{S} \cup \{a_1, \ldots a_n\} \vdash_p Y$.
(2) $\mathcal{S} \models \{a_1, \ldots a_n\} \to Y$
(3) $\mathbb{K} \models \{a_1, \ldots a_n\} \to Y$.

3.2 Armstrong Rules

The so-called *Armstrong rules* which were introduced in the design of relational databases to determine functional dependencies [7], facilitates implicational reasoning. These rules are:

$$R1 : \frac{}{X \to X},\ R2 : \frac{X \to Y}{X \cup Z \to Y}, R3 : \frac{X \to Y,\ Y \cup Z \to W}{X \cup Z \to W}$$

A set of implications is closed (contains all its consequences) if and only if the set is closed by Armstrong rules [7]. As a consequence of the former, if $\vdash_A$ denotes the proof notion associated to Armstrong rules, implicational bases are $\vdash_A$-complete, that is to say:

Theorem 3.4: Let $\mathcal{L}$ be an implicational basis for $\mathbb{K}$, and L an implication. Then $\mathbb{K} \models L$ if and only if $\mathcal{L} \vdash_A L$

3.3 Association Rules and Luxenburger Basis

The development of logical reasoning methods for association rules is a relatively recent promising line of research [8]. In FCA, association rules are also implications between sets of attributes. Confidence and support are defined as usual in data mining. The analogous to Stem Basis for association rules is the Luxenburger basis [19]. The reasoning system for SB can be adapted for reasoning with Luxenburger basis [1]. The section is devoted to introduce such bases.

A set Y is **closed** if $Y'' = Y$, and given Y_1, Y_2 closed, it is denoted $Y_1 \prec Y_2$ when there is not Y closed such that $Y_1 \subset Y \subset Y_2$.

In FCA, association rules are also implications between attributes. Confidence and support are defined as usual in data mining:

Definition 3.5: Let be $\mathbb{K} = (G, M, I)$ a formal context and $Y, Y_1, Y_2 \subset M$.

- The support of an attribute set $Y \subseteq M$ is

$$supp(Y) = \frac{|Y'|}{|G|}$$

- The support of an implication $L = Y_1 \to Y_2$ is

$$supp(L) = \frac{|(Y_1 \cup Y_2)'|}{|G|}$$

  (Which can also be interpreted as an estimate of the probability $P(Y_2|Y_1)$, that is to say, the probability for an object to satisfy every attribute of Y_2 under the condition that it also satisfies every attribute of Y_1)
- The confidence of L is $conf(L) = \dfrac{supp(Y_1 \cup Y_2)}{supp(Y_1)}$

Definition 3.6: Given γ and δ, the Luxemburger basis of a context $\mathbb{K}$ with confidence γ and support δ, is

$$\mathcal{L}(\mathbb{K}, \gamma, \delta) := \{L : Y_1 \to Y_2 \mid Y_1, Y_2 \text{ closed}, Y_1 \prec Y_2, \\ conf(L) \geq \gamma,\ sup(L) \geq \delta\}$$

In order to simplify the notation, let us suppose that SB (a subset of implications having confidence equal to one, they are always true within the context) is contained in that basis. Implications of Luxenburger basis work as association rules from classic data mining setting. For the example from Fig. 1, two Luxenburger bases are depicted in Fig. 1.

3.4 Simulation Process by Means of Attribute Selection

The overall simulation process includes four steps (see Fig. 4 right): First step, data collection. Second step, the observer selects attributes he/she finds relevant to explain agent behaviour. In the third step, attribute values are computed for each agent (possibly by selecting thresholds for continuous values). A formal context is built with these observations and attributes. Lastly, the predicted action of the agent is obtained by reasoning with the basis extracted from the context.

Table 1. Two Luxemburger bases for example from Fig. 1

$$\mathcal{L}(\mathbb{K},0.5,2/5)=\left\{\begin{array}{l}\mathcal{L}(\mathbb{K},0.8,1)=\left\{\begin{array}{|l|l|l|}\hline Implication & Confidence & Support\\ \hline Sea\rightarrow Coast & 1 & 1\\ \{\,\}\rightarrow Coast & 4/5 & 1\\ \hline\end{array}\right.\\ \qquad\qquad\qquad\qquad\;\;\begin{array}{|l|l|l|}\hline Coast\rightarrow Sea & 3/4 & 4/5\\ River\rightarrow Coast & 2/3 & 1/3\\ \{\,\}\rightarrow River & 3/5 & 3/5\\ River, Coast\rightarrow Sea & 1/2 & 2/5\\ \hline\end{array}\end{array}\right.$$

4 Representing Agent's Knowledge by Means of FCA

Figure 3 shows the information each agent has about its environment. In the basic model the agent only receives information about its distance to destination and the neighbouring obstacles. This information is the *potential* of each cell with respect to each agent target (destination).

As it is shown in Fig. 3, the potential assesses 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 considered in agent's decision.

In order to validate the proposal, experiments using two of the basic Agent-Based Models for pedestrians (Best movement and Any *good* movement) have been carried out. Due to space limitations it has not been possible to

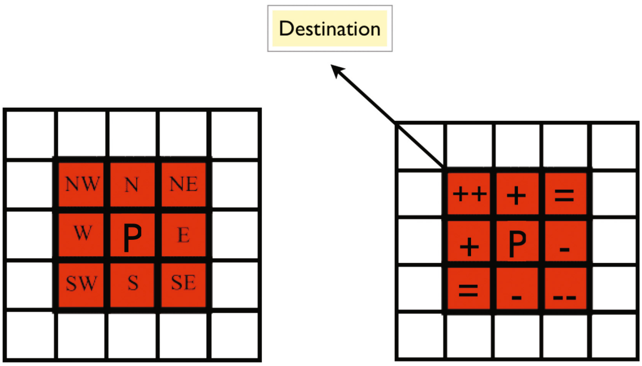

Fig. 3. Agent's (visible) neighbourhood (top, left), *potentials* of agent neighbouring cells

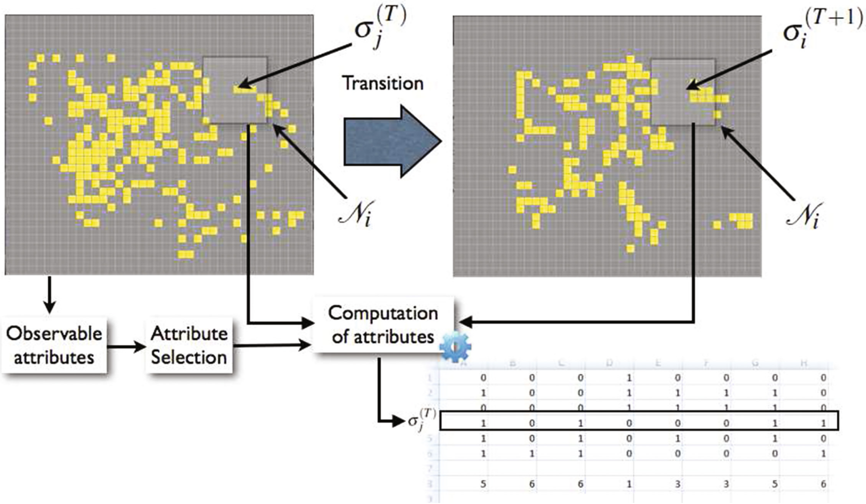

Fig. 4. FCA-based modelling of pedestrians behaviour

consider more elaborate models on this article. Likewise, two different attribute sets, based on cells' potentials, will be used as agents' (local) knowledge representation[2]:

4.1 Detailed Potentials (8 Attributes)

This attribute set considers how good or bad the potential of each neighbouring cell is. Despite the potential being an abstract concept, in this case, it is based only on agents' distance to destination. Therefore it is possible to quantify potentials in terms of cell's *Manhattan distance* to destination.

For instance, movement top-left for the agent of Fig. 3 have a potential of (+2), since this movement would decrease in 2 cells the agent's Manhattan distance to destination. Similarly, movement bottom right has a potential of (−2) (distance increases 2 cells). Finally, neutral movements are those not increasing nor decreasing the distance. Table 2 summarizes this attribute set.

The nine attributes *Will-Move-To-XX(Target)* contains information on agent's next movement. These will be the target attributes during the reasoning process in prediction experiments, in which a model is built from past information on agents' behaviour.

[2] For simplicity, standard qualitative attributes have been selected to show the method. The attribute selection can be expanded by adding any (computable) attribute the observer finds important for pedestrian mobility in its workspace.

Table 2. Detailed potentials attribute set

16 attributes for positive potentials:	
8 attributes for cells with potential (+2)	{*TL-Potential-POS-2*, ..., *BR-Potential-POS-2*}
8 attributes for potential (+1)	{*TL-Potential-POS-1*, ..., *BR-Potential-POS-1*}
8 attributes for cells with neutral potentials	{*TL-Potential-0*, ..., *BR-Potential-0*}
16 attributes for negative potentials:	
8 attributes for cells with potential (−2)	{*TL-Potential-NEG-2*, ..., *BR-Potential-NEG-2*}
8 attributes for potential (+1)	{*TL-Potential-NEG-1*, ..., *BR-Potential-NEG-1*}
8 attributes for obstacles	{*TL-Obstacle*, ..., *BR-Obstacle*}
One attribute for *agent in exit cell*	*Is-On-Exit*
9 attributes for agent's next movement	
8 attributes for each possible movement	{*Will-Move-To-TL(Target)*, ..., *Will-Move-To-BL(Target)*}
Agent will not move	*Wont-Move(Target)*

4.2 Simplified Potentials (42 Attributes)

This attribute set provides information only on whether potentials are positive, negative or neutral (see Table 3) without quantifying how positive or negative they are. The rest of attributes are the same in both attribute sets.

In both attribute sets, each (neighbouring) 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 centre-right cell and *BC* refers to bottom-centre cell.

Table 3. Simplified potentials attribute set

8 attributes for positive potentials	{*TL-Potential-0*, ..., *BR-Potential-0*}
8 attributes for cells with neutral potentials	{*TL-Potential-POS*, ..., *BR-Potential-POS*}
8 attributes for negative potentials	{*TL-Potential-NEG*, ..., *BR-Potential-NEG*}
8 attributes for obstacles	{*TL-Obstacle*, ..., *BR-Obstacle*}

5 Contextual Selection for Pedestrians

The reasoning system presented in [3] allows using FCA-based tools to carry out pedestrian behaviour simulations on both, artificial and real environments. It is possible to extract a knowledge base, to be used by the reasoning system, from a contextual selection (formal context). A contextual selection $\mathbb{K} = (G, M, I)$ for the study of pedestrian mobility dynamics consist of pedestrian movements (objects O) and properties (attributes A) describing both, pedestrians' current state (and its neighbourhood) in a certain time step and its next movement.

The contextual selection contains information items (based on its context, that is, time, space and other properties) similar to the ones under study. Therefore, reasoning with this selection will provide more reliable entailments. For instance, let Z be a pedestrian whose current position is known. A contextual selection to predict the next movement of Z would be defined by:

- **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 Z (for big or heterogeneous scenarios), for instance, those pedestrians located in the same street as Z.
- **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 Z in a time step T the contextual selection for Z in T will consist of 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, the attribute set most suitable to be used as knowledge representation for Z and its environment, would be selected for each setting. In this specific case (the pedestrian basic model), one of the two proposed attribute sets can be selected. In more complex scenarios the attribute set can include other attributes specific for the environment where Z is located (spatial, temporal or from other nature).

6 FCA-based Simulation of Pedestrian Flow

When working with a specific bounded scenario, it is not necessary to compute a new contextual selection for every pedestrian since all pedestrians located in the same area share facets of their environments (context). Therefore the same contextual selection (and its associated knowledge base) can be used for pedestrians located in the same area in a given time step. The process for simulating/predicting the next movement, in a certain time step M, of a group of pedestrians for which past movements', until a certain time step N, are known (where $N < M$), is as follows:

(1) A formal context (the contextual selection) is built containing information on the W most recent time steps (movements) with respect to the target. This formal context will contain, for each time step w_i and for each pedestrian, an

object describing pedestrian's neighbourhood at time step w_i and its next movement at time step (w_{i+1}).

(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.

(3) Finally, in order to predict pedestrian's next movement, a reasoning process is carried out. Initial facts for this process consist in an attribute set describing pedestrian neighbourhood at time step M. A next movement estimate at time step $M + 1$ can be extracted from the reasoning process entailment.

6.1 Analysing Simulation Results

In order to experiment with the methodology above described, a simulation platform has been developed. This platform consists of two modules, the first one comes with a NetLogo-based simulation viewer (see Fig. 5) and is used for preliminary tests. The second one focuses on computing massive simulations and is used for complete experiments.

Although many experiments have been carried out, due to the lack of space, only few of them (see Fig. 6) are mentioned in this work. The setting for these experiments consists in a squared grid with 625 cells (25 per side) populated

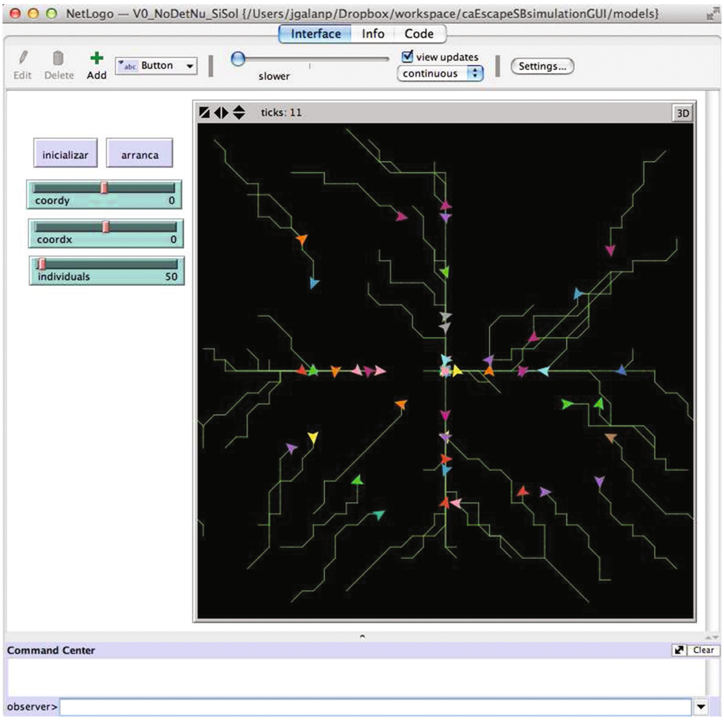

Fig. 5. NetLogo-based simulation viewer for preliminary tests

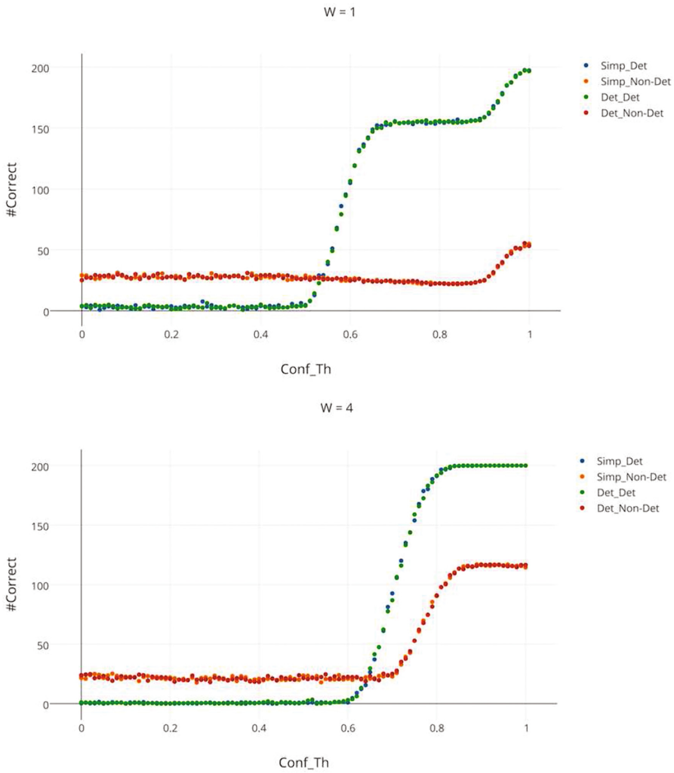

Fig. 6. Experimental results with different temporal windows

by 200 agents. In order to show the importance of the amount of information considered (see Sect. 7), results of experiments for different window sizes W are provided (see Fig. 6, where $W = 1$ (top) and $W = 4$ (bottom)). Results of four different simulations are shown in each plot, 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 the observable information collected within the time interval $[T - W, T)$, and used (after selecting the implications with confidence greater than a certain threshold C_{th}) to predict agents' next 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 ($C_{th} \in [0, 1]$) and $N = 100$ times for each value, in order to obtain a reliable estimate. Figure 7 shows the basis $\mathcal{L}(\mathbb{K}, 0.92, 0.26)$ for the formal context generated by a temporal window of size 5.

After the experiments, we can conclude that there is not a substantial difference between the two knowledge representations used. It is worthy to note that a small uncertainty in agents' behaviour (*any good movement*) leads to a great

```
1 < 10748 > CL_Potential_NEG_1 TC_Potential_NEG_1 =[100%]=> < 10748 > TL_Potential_NEG_2;
2 < 10794 > CL_Potential_NEG_1 BC_Potential_NEG_1 =[100%]=> < 10794 > BL_Potential_NEG_2;
3 < 10675 > TR_Potential_NEG_2 =[100%]=> < 10675 > TC_Potential_NEG_1 CR_Potential_NEG_1;
4 < 10600 > BR_Potential_NEG_2 =[100%]=> < 10600 > BC_Potential_NEG_1 CR_Potential_NEG_1;
5 < 10675 > TC_Potential_NEG_1 CR_Potential_NEG_1 =[100%]=> < 10675 > TR_Potential_NEG_2;
6 < 10748 > TL_Potential_NEG_2 =[100%]=> < 10748 > CL_Potential_NEG_1 TC_Potential_NEG_1;
7 < 10600 > BC_Potential_NEG_1 CR_Potential_NEG_1 =[100%]=> < 10600 > BR_Potential_NEG_2;
8 < 10794 > BL_Potential_NEG_2 =[100%]=> < 10794 > CL_Potential_NEG_1 BC_Potential_NEG_1;
9 < 19184 > CL_Potential_POS_1 =[100%]=> < 19157 > CR_Potential_NEG_1;
10 < 19304 > BC_Potential_POS_1 =[100%]=> < 19265 > TC_Potential_NEG_1;
11 < 19227 > TC_Potential_POS_1 =[100%]=> < 19187 > BC_Potential_NEG_1;
12 < 19371 > CR_Potential_POS_1 =[100%]=> < 19330 > CL_Potential_NEG_1;
13 < 19880 > BL_Potential_0 =[93%]=> < 18515 > TR_Potential_0;
14 < 20775 > CL_Potential_NEG_1 =[93%]=> < 19330 > CR_Potential_POS_1;
15 < 20602 > CR_Potential_NEG_1 =[93%]=> < 19157 > CL_Potential_POS_1;
16 < 20047 > TL_Potential_0 =[93%]=> < 18636 > BR_Potential_0;
17 < 20734 > TC_Potential_NEG_1 =[93%]=> < 19265 > BC_Potential_POS_1;
18 < 20656 > BC_Potential_NEG_1 =[93%]=> < 19187 > TC_Potential_POS_1;
19 < 20095 > BR_Potential_0 =[93%]=> < 18636 > TL_Potential_0;
20 < 20020 > TR_Potential_0 =[92%]=> < 18515 > BL_Potential_0;
```

Fig. 7. The basis $\mathcal{L}(\mathbb{K}, 0.92, 0.26)$ for an experiment with temporal window $[1, 5]$. The format for an implication $Y_1 \to Y_2$ is $< |Y_1'| > \to < |(Y_1 \cup Y_2)'| >$. The context has 40,000 observations

increase in the error. The reason is that a step-by-step performance evaluation is too strict for non-deterministic behaviours.

7 On the Robustness of the Model

In order to state a result on the soundness of the model, it is necessary (in this case) to have a prefixed agency model. The classic model to specify (deterministic, reactive) agents includes (see [13]):

$$\langle S, T, Act, P, Do, acc \rangle$$

where $Perceive$ is the function that determines the agent situation

$$Perceive : S \to T$$

where S is the set of states, T is a partition of S (the situations, due to perception features of the agent), acc selects the action to be executed in a certain state

$$acc : T \mapsto Act$$

and Do determines the effect of an action on a state (reaching other state)

$$Do : Act \times S \to S$$

The execution of the agent from an initial state $\sigma(0)$ 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 $\mathcal{S}_n^K$ be a subset of implications with positive support, that can be obtained from the Stem basis of a context M_n^K containing 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, let us 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 robustness of the model could be stated in a particular case as follows: If the context is large enough and the distribution of obstacles is not biased (relative to environments based on the Moore Neighbourhood. For example, a uniform distribution), then the knowledge base ($\mathcal{S}_n^K$) considers every possible situation and provides an action that agrees with the behaviour selected by Acc.

Theorem 7.1: Let Δ be a distribution and A be a finite descriptive attribute set for T (being T finite). It can be supposed that:

- Δ is T-complete
- Agents share the specification $\langle S, T, Act, P, Do, acc\rangle$

Then there exists $K > 0$ such that for all $n \in \mathbb{N}$

$$acc(Perceive(\sigma(n)))^A \in \mathcal{S}_n^K[Perceive(\sigma(n))^A]$$

Proof: On the one hand, since A is descriptive, it holds that $Perceive(\sigma(n))$ characterizes the state $\sigma(n)$ for any n. On the other hand, since Δ is T-complete, $s \in_\Delta (I_{K_0})$ for some K_0 large enough. Then the object corresponding to the transition from K_0 to $K_0 + 1$ respects the implication

$$\sigma((n))^A \to \sigma(n+1))^A$$

Since any transition between states $s \mapsto s'$ only depends on s^A, this implication is respected by the former objects. Therefore, it is true within the context, so by completeness of Stem Basis

$$\mathcal{S}_n^K \models \sigma(n))^A \to \sigma(n+1))^A$$

for $K = K_0 + 1$, and

$$acc(Perceive(\sigma(n)))^A \in \sigma(n+1)^A$$

Therefore the implication basis is sound to simulate agents' behaviour by means of a deliberative process. The result shown above does not give any estimation of K (actually $K = K(\Delta, acc)$). Likewise, it would be interesting to state a similar result for non-deterministic agents and Luxenburger basis.

8 Related Work

Recently, the interest on automated synthesis of agent behaviour from raw observational data (e.g. [20]) has increased greatly. Behaviour mining can produce sound simulations in this field. Thus there exists a number of ongoing works on which our proposal could be applied. This would provide alternatives (qualitative) to models of different nature.

In [15] authors present a modified floor field cellular automata model to simulate pedestrian evacuation. The world is discretized in cells, and geometrical inertia and social-force [17] features are considered to model agents. Both types of attributes can be modelled by means of FCA (the third one by means of thresholds), and the model may be simulated using the agency model proposed in this paper (this is a future aim).

In [10] two models of pedestrian mobility were presented, one is CA-based but with continuous treatment of space, and the other is based on ODEs using navigation on geodesics as in CA models. Authors show how these models are similar, and suggest that their idea can be a bridge between the two classes of models. As they assert, the gap between discrete and continuous models can be closed only if new formulations are designed in order to state the correctness (equivalence) among models. Soundness of our model devised here is a first step to state the utility of our proposal in presenting CA (qualitative) simulations as an interpretation in the discrete field of continuous models. Our approach allows to preserve in time *mental models* (namely Luxenburger basis) among cells, that is to say, knowledge bases are not location-dependent.

In [9] a model, which does not need origin–destination trip matrices (that is to say, prefixed exit cells), is presented. The scenario description and global simulation parameters in a particular case (International Fairs) is developed and used for simulation. The graph-based modelling provides a macroscopic simulation that could be an excellent starting point to design Agent-Based Models representing individual behaviour.

9 Conclusions and Future Work

The work presented is based on a general hybrid approach to phenomenological reconstruction of Complex Systems (CS), using FCA as main tool for conceptual data mining (see [2]). In [3], the idea was applied to a classic CA (Conway's game of Life). The approach presented here specifies and implements a general method for movable agents. The key advantage of this methodology is that the observer can select (computable, qualitative) attributes in order to understand (and model) pedestrian behaviour. The selection can comprise any feature on both pedestrians and streets. From this selection, our method provides a qualitative model. Therefore our approach uses two well know Artificial Intelligence tools: concept mining and agent-based simulation. In this way the model provides a solution for simulations based on agents' behaviour by using qualitative agent reasoning. It is a realistic model which produces sound results in other cases of Complex Systems [3,6,11].

The model is easily extensible; it can be improved in any moment by adding a number of new features (as for example, digital information received or generated by pedestrians, digital footprints, real time decisions on traffic and flows, etc.).

With respect to the attribute sets, the one with detailed potentials performs better on deterministic scenarios. However, in real world scenarios it is expected to deal with uncertainty. In experiments with non-deterministic scenarios, the attribute set with simplified potentials showed to be more robust.

Likewise, the Luxenburger basis showed to perform better than the Stem basis in non-deterministic scenarios, as rules confidence deals with uncertainty, capturing predominant behaviours from pedestrian dynamics. However, for a detailed analysis of pedestrian flow, in order to detect and study anomalies in the scenario, it seems to be more interesting the use of the Stem basis. Finally, it is possible to consider *fuzzy attributes* instead of qualitative ones in order to obtain a more precise and flexible knowledge base.

Since our model works with knowledge-based agents, it could be extended, by joining knowledge from distinct agents, in order to work in the problem of collaborative localization [18]. Stem (and Luxenburger) basis of different agents can be re-interpreted in order to fuse their knowledge base by conciliation (as in [6]) or by considering a common (global) Knowledge [4]. In general terms, the integration of different knowledge elements within FCA can pose some issues on the extensibility and scalability of the model, which could be solved by means of discretization methods, multi-valuation and scales from FCA [12]. Likewise, it is feasible to consider other neighbourhoods following a similar modelling strategy.

Acknowledgment. Work partially supported by TIN2013- 41086-P (Spanish Ministry of Economy and Competitiveness), co-financed with FEDER funds.

References

[1] Aranda-Corral, G.A., Borrego Díaz, J., Galán Páez, J.: Confidence-based reasoning with local temporal formal contexts. In: Cabestany, J., Rojas, I., Joya, G. (eds.) IWANN 2011. LNCS, vol. 6692, pp. 461–468. Springer, Heidelberg (2011). doi:10.1007/978-3-642-21498-1_58

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

[3] Aranda-Corral, G.A., Borrego-Díaz, J., Galán-Páez, J.: Qualitative reasoning on complex systems from observations. In: Pan, J.-S., Polycarpou, M.M., Woźniak, M., Carvalho, A.C.P.L.F., Quintián, H., Corchado, E. (eds.) HAIS 2013. LNCS (LNAI), vol. 8073, pp. 202–211. Springer, Heidelberg (2013). doi:10.1007/978-3-642-40846-5_21

[4] Aranda-Corral, G.A., Borrego-Díaz, J., Galán-Páez, J.: Simulating language dynamics by means of concept reasoning. In: Caro, G.A., Theraulaz, G. (eds.) BIONETICS 2012. LNICSSITE, vol. 134, pp. 296–311. Springer, Cham (2014). doi:10.1007/978-3-319-06944-9_21

[5] Aranda-Corral, G.A., Borrego-Diaz, J., Galan-Paez, J.: Scale-free structure in concept lattices associated to complex systems. In: 2012 International Conference on Complex Systems (ICCS), pp. 1–6 (2012)

[6] Aranda-Corral, G.A., Borrego-Díaz, J., Giráldez-Cru, J.: Agent-mediated shared conceptualizations in tagging services. Multimedia Tools Appl. **65**(1), 5–28 (2013)
[7] Armstrong, W.W.: Dependency structures of data base relationships. In: IFIP Congress, pp. 580–583 (1974)
[8] Balcázar, J.L., Garriga, G.C.: Horn axiomatizations for sequential data. Theor. Comput. Sci. **371**(3), 247–264 (2007)
[9] Canca, D., Zarzo, A., Algaba, E., Barrena, E.: Macroscopic attraction-based simulation of pedestrian mobility: a dynamic individual route-choice approach. Eur. J. Oper. Res. **231**(2), 428–442 (2013)
[10] Dietrich, F., Köster, G., Seitz, M., von Sivers, I.: Bridging the gap: from cellular automata to differential equation models for pedestrian dynamics. J. Comput. Sci. **5**(5), 841–846 (2014)
[11] Galán-Páez, J., Borrego-Díaz, J., Aranda-Corral, G.A.: Modeling lexicon emergence as concept emergence in networks. In: Abraham, A., Wegrzyn-Wolska, K., Hassanien, A.E., Snasel, V., Alimi, A.M. (eds.) AECIA 2015. AISC, vol. 427, pp. 409–420. Springer, Cham (2016). doi:10.1007/978-3-319-29504-6_39
[12] Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations, 1st edn. Springer, Secaucus (1997)
[13] Genesereth, M.R., Nilsson, N.J.: Logical Foundations of Artificial Intelligence. Morgan Kaufmann Publishers Inc., San Francisco (1987)
[14] Guigues, J.L., Duquenne, V.: Familles minimales d'implications informatives résultant d'un tableau de données binaires. Mathématiques et Sciences Humaines **95**, 5–18 (1986)
[15] Guo, R.-Y., Huang, H.-J.: A modified floor field cellular automata model for pedestrian evacuation simulation. J. Phys. A Math. Theor. **41**(38), 385104 (2008)
[16] Helbing, D., Johansson, A.: Pedestrian, crowd, and evacuation dynamics. In: Meyers, R.A. (ed.) Encyclopedia of Complexity and Systems Science, pp. 1–28 (2014)
[17] Helbing, D., Molnár, P.: Social force model for pedestrian dynamics. Phys. Rev. E **51**, 4282–4286 (1995)
[18] Kampis, G., Kantelhardt, J.W., Kloch, K., Lukowicz, P.: Analytical and simulation models for collaborative localization. J. Comput. Sci. **6**, 1–10 (2015)
[19] Luxenburger, M.: Implications partielles dans un contexte. Mathématiques Informatique et Sciences Humaines **29**, 35–55 (1991)
[20] Parsons, B., Vidal, J.M., Huynh, N., Snyder, R.: Automatic generation of agent behavior models from raw observational data. In: Grimaldo, F., Norling, E. (eds.) MABS 2014. LNCS (LNAI), vol. 9002, pp. 121–132. Springer, Cham (2015). doi:10.1007/978-3-319-14627-0_9
[21] Vizzari, G., Bandini, S.: Studying pedestrian and crowd dynamics through integrated analysis and synthesis. IEEE Intell. Syst. **28**(5), 56–60 (2013)