An Internet Survey for an Activity-Based Model

An urban transportation analysis integrated in a G.I.S. environment

Massimiliano Petri, Alessandra Lapucci, Diana Poletti, and Silvana Lombardo
University of Pisa, Department of Civil Engineering

Keywords: Activity-Based Model, G.I.S., Network Analysis, Decision Trees

Abstract: The current research deals with the development of an Activity-Based Multi Agent System fully implemented in a G.I.S. framework and applied to the case study of the historical centre of Pisa. The objective is to create a simulation tool for Pisa population transfers in order to verify how transport demand varies because of interventions on traffic plan (e.g. creating urban areas subject to a toll access for vehicles), or on public transport lines, or on new activities location (e.g. supermarkets, public services etc.). Three different parts of the System have been simultaneously carried out: the first concerns a population sample survey, the second deals with geographical data structuring and the last one, still in elaboration progress, tests the model reliability to extract and implement behavioural rules. The results obtained till now show how the Database itself, containing temporal data about agents activities (extracted by the population sample questionnaire) and urban services the city offers, already represents an important instrument to support decision making process.

1. INTRODUCTION: WHY ACTIVITY-BASED?

The Activity Based micro-simulation has only recently been introduced in transportation model and more in general in the field of territorial planning. The spell of micro-simulation, in short, consists in the possibility of obtaining the behaviour of a complex urban system at meso or macro scale as derived from millions of choices performed by the individuals belonging to the system. This kind of bottom-up approach seems to overcome, from
many points of view, the traditional top-down approach in which some behavioural choice rules of groups of individuals are assumed and formalized and the model is calibrated on the case study data.

Therefore, in our opinion, most of micro-simulation experiments carried out (using Multi Agent Systems, Cellular Automata, etc...) is undermined by the definition of the individual choice rules adopted for the simulations. Indeed, such behavioural rules are often very similar to those adopted in top-down approaches, i.e. behavioural rules known at the meso-scale are used at the micro-scale too. This would allow the generalization of the models structure, but is it correct to attribute the same individual choice rules to people living in Pisa, or in London or in Rome? (Lombardo, Petri, 2005a)

The micro-scale approach can be used only when is available a micro-scale database containing sufficiently detailed information about individuals: this requires direct surveys with appropriately designed questionnaire.

As this database will be very large, usual statistical tools are not adequate to extract the needed behavioural rules; it is, then, necessary to turn to more advanced tools, such as Decision Trees, Bayesian Networks, etc.

Such tools allow the extraction of a set of elements which determine each individual behavioural choice.

Against the time, costs and technical skills involved in such surveys and in their elaboration, the utility of such work is relevant for a number of planning and assessment problems not solved yet.

In the context of territory-transport planning, the activity-based approach is very suitable because it provides a general model framework which is very flexible and able to implement the extracted behavioural rules at every decisional step. Moreover the Activity-Based micro-simulation approach allows to simulate the not-systematic trips (systematic=home-work-home; home-school-home trips) which have been greatly increasing in our cities and are quite adverse to be modelled.

In order to point out the advantages of an activity-based model, some examples will be described that highlight the power of this approach.

First of all, it gives the possibility of considering family members interactions which alter trip types and number. For instance, when a family member takes her/his child to school not being able to pick her/him up later, a second family member has to go and so takes the opportunity to do some shopping on the way back home, further increasing traffic.

Another relevant aspect is related to trips duration and, more generally, to time factor issues. While trip-based models consider time only as a cost, i.e. as an impedance factor, in reality it can play a different role for the compilation of the activities daily diary. For instance, if an individual, who planned to shop after work, arrives late at the office because of traffic congestion, on that day he will be forced to leave the working permises later
and, having little time left, he will decide to shop in a small store close home. Had he not been late that morning, he could have gone to a supermarket, covering a longer distance, but saving money in his purchases.

All these examples show that the only element subject to a substantial variation is the vehicular flow.

2. OBJECTIVES AND METHOD

As a consequence of all above mentioned issues, the present work and the model here presented try to introduce further analysis elements useful to clarify and simplify the complex field of Multi Agent Systems (Waddel, Borning, Noth and others, 2003, Occelli, 2004) : this is mainly due to two innovations:

1. Model implementation inside a GIS platform, able to guarantee an efficient database management and a dynamic display of simulation evolution (Lombardo, Petri, 2005b);
2. Individual (resident or commuters) behavioural rules implementation, i.e choice rules directly extracted from survey. Those rules are formed by an “IF” part which is the condition (hypothesis dependent on individual characteristic) and by a “THEN” part which is the execution (individual choice). For instance IF (age = 30-34 AND income = < 15000 €) THEN (first daily activity = work), or IF (age > 74 AND number of family components = 2 THEN (means of transport = bus).

The final objective of this model is, then, to act as an effective and robust decision support system tool thanks to the possibility to simulate the global effect on urban traffic due to a great number of actions, as the shops opening times variation, the urban vehicular traffic change, the public transportation demand and price fluctuation, the parking toll and localisation changes.

The present model, introducing a greater detail when compared with traditional spatial interaction models, adopts a higher number of variables in the simulation providing, as a consequence, a wider number of operative solutions within urban mobility matters.

Given the complexity of the model project still in progress, the current contribution will mostly highlight: (a) the construction of a “temporal” Urban Information System (U.I.S.) (Burrough, McDonnell, 1997), in ArcView 9.1; (b) the building and distribution of an “ad hoc” questionnaire to a population sample.

In the first part a Geodatabase has been developed: it contains “temporal” information about urban facilities supplied, in order to simulate how the city lives and to display, through a sort of time clock, which services are
activated in each daily timeband. This application represents the city as a “living organism”, therefore as another agent interacting with the whole system (Gilbert, Terna, 1999). In fact the urban system can be considered as a real agent, able to provide different services in each daily timeband and characterized by proper attributes as attractiveness and centrality.

As to the second part it is important to underline that surveys are addressed both to residents and to commuters with the purpose to allow a more realistic view of the urban system complexity.

3. **A CASE STUDY**

The work here presented studies Pisa city center and tries to simulate the single agent motion within it. The model general structure consists of four parts:

1. Construction of the “temporal” Geodatabase representing the territorial information base on which the model has to be founded;
2. Questionnaire building, distribution and collection in order to procure the necessary information on individual peculiarities and on population sample activities;
3. Behavioural rules extraction, from the above mentioned population sample, by Knowledge Discovery tools (Arentze, Timmermans, 2000a). Among them, one of the most appropriate, belonging to Datamining field (Arezzo, Felici, 2003, Lombardo, Pecori, 2004, Lombardo, Pecori, Petri, 2005), is known as Decision Trees Induction and is integrated in WEKA software (Witten, Frank, 2000);
4. Behavioural rules and model implementation on an ArcGIS platform through an “object-oriented” programming and by “on the fly” geoprocessing operation (i.e. influence and service area computation) or network dataset routing.

3.1 **The city as a living organism – a dynamic GeoDatabase**

This is the most demanding phase, from a computational point of view, as it requires the construction of “ad hoc” geographical data and their subsequent insertion in complex storage structures necessary for successive analysis (Arentze, Timmermans, 2000a,b).

The following data have been processed:
1. Pisa city centre road graph;
2. The activities localization (trade, tourism, services, resident, etc) in each street and for each street numbers;

3. Street and square parking areas mapping, specifying car spaces numbers, whether free or paying according to a definite hourly tariff; in this case the nature of “temporal” information consists in detailing such tariffs;

4. Information about each road sweeping times;

The city centre road graph, was employed to build a detailed Network Dataset (ESRI Technical Paper, 2005) containing information about restrictions on road access and detours, one-ways, attributes hierarchy.

The Network features do not remain unalterate during the simulation, but are modified, according to each personal experiences, in a sort of “mental maps”: every mental map reflects the idea each agent has (his cognitive level) of road distance coverage times (Rabino, 2005). At the initial stages, mental maps corresponde to physical and functional network characteristics, as afterwards model dynamics, articulated in three minutes cycles, simulates the individual displacement induced by each agent daily activity.

For each Network road arc, the information about carrying capacity is stored. When this threshold is exceeded, traffic congestion phenomena occur and cause distance coverage time increase only within individual “mental maps” of people involved in this traffic jam, while it does not happen for all the agents unaffected by the congestion.

At the end of each day, every agent updates her/his “perceived network” taking into account also family members experiences: this phase can be considered as the agents training process.

This aspect makes transfers simulation more realistic when compared to traditional models where, under a congestion event, trips coverage times simultaneously vary for each agent, implicitly assuming a global information flow among them.

Once the Network was developed, “locations” feature Datasets were built: they identify both the route destinations, i.e. all available activities and the route origins, i.e. residences, etc. This result has been achieved using Yellow Pages internet Data Base, creating “*.dbf” format tables which were linked to street numbers through the address field, via “table join” GIS operation. (figure 1).

Such Network will allow to built the isochrone for each activity and to compute its closest stops (using command “Find Closest Facility”). The information about each urban activity opening time and about each agent min/average/max permanence time in such activity was considered.

Data from PISAMO (PISA Mobility) official internet website were used to extract city centre parking areas information, gathered in separate tables.
The stored attributes are related to parking spaces available in each road/square, to hourly tariff and its temporal setting.

Figure 1. Activities Tables creation.

The last stored piece of information deals with possible mobility “constraints” due to the real parking availability or to transit hold up over a whole road section caused by midweek markets, road sweeping etc.

All data were introduced in a Geodatabase (Curtin, Noronha, Goodchild and Grisé, 2003, Booth, Shaner, MacDonald and Shancez, 2002) and linked through “relationship classes” in order to dynamically display all available activities (commerce, services, tertiary etc.) and parking supplies throughout the working week and inside each daily time band (figure 2).

Using VBA language an “urban time clock” was designed in order to generate a temporal articulation of all facilities the city offers. The system creates each activity isocrone, automatically displaying available and not available services distinguishing their symbology at each temporal step; this is a first useful result to support decision making process.

Example of an Activity table record
3.2 Population sample questionnaire

For the purpose of reconstructing population movements and performing further analysis, a representative population sample was necessary.

The interviews, conducted by means of a questionnaire, have been carried out both distributing a paper form and building a dedicated website structured in preset interfaces; such double version was chosen to reach different population classes (figure 3 and 4). The questionnaire is articulated in three parts:

Part A: Individual personal data (age class, civil status, occupation, educational qualification, etc.) and family information (number of family members, number of children, cars and driving licences number, etc.);

Part B: Most recent working day activities diary;

Part C: Individual preferences in daily activities planning.
The survey directed to residents in Pisa city centre is structured differently from that addressed to commuters. The former registers both urban and city entry/exit activities, indicating, in the second case, the origin/destination point for this “macro-transfer”. The latter, instead, provides informations about commuters arrival point in Pisa (e.g. train station, long stay parkings, bus station, etc.), their mean of transportation, and about the activities they carry out in the city centre (figure 5).
An Internet Survey for an Activity-Based Model

Data entry was conducted through a VBA (Visual Basic for Application) application, completely integrated in ArcGis, in order to facilitate and speed up an accurate loading of questionnaire results (figure 6).
The area designed to test the prototype of the model is composed by forty Census Sections and contains different activity types, a comprehensive list of transportation means and some “key activities” as university headquarters and the central hospital (figure 7).

The residents’ population sample was singled out by cross-referencing two of the most influential variables available from 2001 National Population Census. Selected variables were sex and educational qualification, because they are able to explain a good part of the information contained in the entire dataset, as proven by the factorial analysis implemented only on personal data (Part A of the form) present in the questionnaire collected up to now. For each variable all classes were intersected and, for each crossing, 10% of the population, responding to those characteristics, was interviewed. Such approach enables the sample to follow the National Population Census probability distribution.
Figure 8 shows how the bivariate analysis has been conducted in a single census section about sex and literate level classes.

Our preliminary intention was to address the questionnaire to tourists as well, however their behaviour cannot be easily integrated in an urban activity based model. Tourists, in fact, initially ignore the urban configuration and visualize their own “mental map” as they visit and walk through the city.

In order to have a model less “factual” that’s to say more capable to anticipate urban system variations, induced by planning policies, the knowledge contribution given by “privileged interwees” was introduced.
The importance of “privileged interwees” resides in the possibility to design a limited size sample able to represent the entire categories distribution of the whole dataset. Asking specific questions to this sub-sample, it is possible to extrapolate agents’ reactions to unforeseen events connected to urban mobility.

Some examples of the asked questions to “privileged interwees” are the following:

1. How long do you spend on activity X on average? What is the activity maximum duration? What is the shortest time spent in case you are in a rush?
2. In case of a delay greater than twenty minutes at the bus stop, would you use an alternative transportation mean or would you rather wait for the bus?
3. How often do you perform activity X?
4. How much are you willing to pay to have direct access to your regular activities in less than five minutes?

With the use of Decision Tree induction methods it is possible to extract information from the answers to such questions and to relate personal features to each target variable linked to every question; this enhances the model forecasting accuracy (Jager, Janssen, 2003). For instance the average time a professional spends for lunch may be shorter than that spent by an housewife (question 1), or a single may have a social activity more intense than a family with two children (question 3).
3.3 Behavioural rules extraction from the population sample

Activity Based models allow the conversion of a daily “activity program” to a daily “activity pattern” where a sequence of activities is completely defined in terms of localisation, type, transportation mean, schedule, duration and possible sharing. This transformation, from a generic list to an ordinate sequence of daily activities, implies a set of decisions: which activity to carry out, where, when, with whom and by which transportation mean.

The agent behavioural rules are extracted from activities personal data and the whole process is implemented through Knowledge Discovery in Databases (KDD) tools (Bonchi, Pecori, 2004).

The instrument initially used was the Decision Tree (Lombardo, Pecori, Petri, 2005, Lombardo, Pecori, 2004) where knowledge is extracted from data in the IF-THEN form; rules structure presents, in fact, an IF part (condition) where all socio-economic attributes and daily activities, related to each individual agent and to her/his family, are taken into account and a THEN part (execution) where the decision tree “target attribute” is considered. The target attribute may be represented by each decisional choice under study (e.g. the possibility to conduct an activity going outside, the type of activity, the transportation mean selection and so on).

The general framework of decision tree induction rule, can be formalized as follows:

\[
\text{IF (Age band = X, and Income = Y, and ... and Total daily activities number = Z)}
\]
\[
\text{THEN (First daily activity = T)} \quad \text{(in relation, for instance, to the target variable called “daily activity to perform”).}
\]

The main difficulty this approach presents is due to the necessity to apply the rules, extracted from the sample, to the whole population. It is possible, indeed, that a combination of conditional attributes, never occurred in the extracted rules (IF part), may come out. In this case, the probability of the considered decision variables to be conjointly present in the population universe would be useful, in others words the problem is to compute a conditional probability distribution. Therefore, we are investigating the possibility to apply the synergy between Decision Trees and Bayesian Networks, able to explain such “variables relation network”.
3.4 Behavioural rules and model implementation in a G.I.S. environment

The behavioural rules extracted and the whole model will be inserted in ArcGis environment through an object oriented programming. This operation will allow the system to perform a real time view of the entire simulation and permit geographical analysis inside the dynamics.

The possibility to keep each computation inside a GIS platform provides the entire instrument to be much more flexible and easily integrable with others techniques: this element represents a significant advantage because avoids computational problems related to the difficulty of exporting data from a software to another and gives the opportunity to easily display simulation results by geographical maps and interfaces.

The complexity of urban planning decision problems requires, in fact, models able not only to simulate agents dynamics, but also to represent results in an clear and transparent form, so that they can be easily communicated to scientists, local authorities, stakeholders etc. This result is surely facilitated by means of a GIS based tool (Zunino, 1998, Longley, Goodchild, Maguire and Rhind, 2001).

4. CONCLUSIONS

The comprehensive framework of the devised activity based mobility model represents a significant step forward in dynamics and phenomena comprehension at urban scale.

The model elaboration, currently in progress, its testing on Pisa case study and the tool implemented till now, already represent a decision support element useful both to locate new possible activities inside city centre and to improve the quality of present services.

The accuracy in the initial stage of the analysis, as database design and implementation, has been highlighted because it has been considered not only the model starting point, but also its foundation and its strong point.

5. REFERENCES

An Internet Survey for an Activity-Based Model


Curtin K., Noronha V., Goodchild M.and Grisé S., 2003 ArcGIS Transportation Data Model (UNETRANS) - UNETRANS Data Model Reference.


ESRI Technical Paper, 2005, Preparing Street Data for Use with the Network Dataset


