Modeling Location decisions of retailers with an agent-based approach

Francesco Ciari
Modeling Location decisions of retailers with an agent-based approach

Francesco Ciari
IVT, ETH Zurich
ETH Hönggerberg, CH-8093
Zurich

Phone: +41-44-633 71 65
ciari@ivt.baug.ethz.ch

May 2011

Abstract

The paper reports about a project, where an agent-based module of location decisions of retailers has been implemented into the agent-based traffic simulator MATSim-T. The retailer model focuses on location choice of retailers. Retailers implement a simple strategies in order to maximize the number of their customers. The paper discusses first the conceptual background of the model. The results obtained using a large scale scenario, the urban agglomeration of Zurich, and where retailers are applying different location strategies, are presented and thoroughly discussed.

Key Words: retail planning; location choice; microsimulation

Preferred citation style: Ciari, F. (2011) Modeling Location decisions of retailers with an agent-based approach
1 Introduction

Models of travel and land use are a category of models aimed to comprehensive describe urban systems. Since a system is “a regularly interacting or interdependent group of items forming a unified whole”, those items, through their interactions, are all contributing to determine the behavior of the system. This means that, in principle, the description of the system should imply the description of all these items. However, models of travel and land use tend to have rich descriptions of the travelers, but the other actors of the urban system are normally abstracted into market clearing mechanisms. This is no surprise, since it allows keeping the models relatively simple, but intrinsically prevents the models to capture the dynamic interactions among all the actors of the urban system. In this paper it is reported about an effort aimed to overcome this problem. It is presented a model where not only travelers, but also retailers are represented at the individual level. This has been made building on an existing agent-based micro-simulation tool, MATSim (Balmer, 2007). It uses the concept of the Evolutionary Algorithm (EA) in order to generate consistent daily activity schedules for each individual (agent) of a population and travel times on the network. In the standard implementation only individuals are described using the agent paradigm. Adding retailers in the modeling system is a first step to a much richer description of the urban system in this simulation toolkit in the future. It is also a first effort in order to integrate the supply side choices, here for retailing, into such a system. The implemented model opens a wide range of new policy evaluation applications for the simulation tool. The behavior of the different retailers is made explicit by developing a specific retailer agent. This agent can address the location choice problem of a retailer. The reason for this choice is related with the substantial importance that location has for retail stores. It is commonly accepted that location is the most important factor for the success of a retail store, and sometimes is even acknowledged as the only one (Zentes et al., 2007). Being so, it is understandable that many different techniques have been created for the optimal location of retail shops. In recent years, the use of agent-based modeling in spatial issues has become more popular and location problems are among those targeted by this stream of research. The popularity of agent-based systems in itself is a logical consequence of the increased calculation power of computers, but for the application to retail markets the availability of more and more precise data, plays a crucial role. The introduction of new data collection technologies such as EPOS (electronic point-of-sale) systems and the
increasing popularity of store cards simplify the task to obtain detailed information on consumers for retailers (Nakaya et al., 2007). However, the use of the agent paradigm to model retail markets is usually limited to the agent-based modeling of customers and not retailers. Agent-based implementations of retailers are rare so far and models were both retailers and customers are modeled as agent are even fewer.

The remainder of this paper is composed of four sections. Section two provides a short overview on two streams of the literature which are relevant for this work. The first part gives an overview on combined travel and land use models. The second shortly review some models dealing with the representation of retail markets using agent-based approaches. In section three, it is reported in more details on the implementation of the retail agent. Section four discusses the results obtained running the simulation on a large scale scenario representing the metropolitan area of Zurich, showing that the strategies implemented enable retail agents to obtain sensible improvements in terms of number of customers. Section five review some of the possible future enhancements of the presented model.
2 Relevant Literature

In this short literature review papers dealing with traffic agent-based simulations have been are purposely omitted. Readers interested in a deeper insight on this methodology in general can refer for example to Salvini and Miller, (2005), Vovsha et al. (2002), TRAMSIM (2008).

The stream of work which tries to integrate land use with transport models has already a relatively long tradition; some older works of that kind are DRAM/EMPAL (Putman 1983) and TRANUS (De la Barra et al. 1984). Recently land use-transport models have moved from aggregate models of various types to discrete choice logit models and more recently toward disaggregated, activity based models. Recent examples are UrbanSim (Waddell 2000, 2007) ILUMASS (Wegener 1999, Beckmann et al., 2007) and the work of Arentze and Timmermans (2000, 2007). In the case of ILUMASS for example, where all actors of the different markets (land, work, retail) were represented as agents, the project ended without matching the initial objectives. The simulation tool, especially the transport module, was too heavy. A simulation of the Dortmund region, the final goal of the project, was never performed and only small test scenarios were run. A complete report on its development, its achievements and its failure can be found in Wagner and Wegener (2007). UrbanSim applications have not experienced such problems, but a fully agent-based implementation is not yet realized and not the actual goal of the efforts. Arentze and Timmermans (2005) presented a multi-agent model of consumer behaviour which, besides the structural attributes of and the distance to the store, also includes opening hours as part of the institutional context of the shopping destination choice. Later (Arentze and Timmermans, 2007), they provide probably the only fully implemented example of an integrated land use- and transport model where both, retailers and customers, are modelled as agents. For the customers’ side the model makes use of ALBATROSS (Arentze et al., 2000) an activity scheduling model which generates a day plan for each agent of the simulation. The supply side is made up of facilities which are classified according to different demand types and sectors. Moreover, it distinguishes among three structurally different facilities: elementary (only one activity is possible), mixed (the sum of several different elementary facilities in the same place) and higher level (several elementary facilities organized in a “higher level” structure, allowing also activities not possible in elementary facilities). For each demand sector an agent controls the development of the facilities net-
work, with sub-agents controlling the development of a certain type of facility in the sector. The agents seek a location for the facility using a catchment area analysis. Once facilities are located the simulation allocates the demand and supply agents are allowed to rearrange their facility. A test case, where a mid-size town is simulated, shows that the model is able to reproduce reasonably well location patterns of suppliers and a sensitivity test shows that the model correctly react to modification of main parameters.

Some work approaching the modeling of retail markets using multi-agent systems has been already attempted. However, the two sides of retail markets, supply and demand, are usually not both represented as agents, typically only consumers are. In this vein of work, Happendal et al. (2006) use a hybrid approach were the petrol market is represented by combining an agent-based approach for the supply side and a spatial interaction model for the consumer side. Individual petrol stations are represented as agent-objects and supplied with knowledge of their initial starting price, production costs and the prices of stations within their neighbourhood. The location of an outlet is considered as fixed and retailers are allowed to react to the sales only by adjusting their price to competitors’ prices, which are considered known. The problem of the location is not directly tackled but the model successfully reproduces consumer’s spatial choices observed in real markets and also the profitability of single retail outlets in the long term. In the market used as a case study in this paper the product sold can be considered homogeneous, which is an advantage, but also a limitation. The lack of complex trade-offs typical of many other retail markets is of course a considerable simplification, but the representation of a composite retail market, where an entire array of products would be available to the customers, would imply a substantial modification of the model. Schenk et al. (2006) use a micro-simulation approach to model the shopping behaviour of inhabitants of an entire region of northern Sweden. Also in this case the agents of the simulation are only modelling the demand side, here at the household level. Nevertheless, the model is accurate and the supply side is modelled with a high spatial resolution. In particular it is one of the few examples were prices are taken into account in order to characterise a retail store. The family agent is described with the socio-demographic attributes of its components and by some family specific attributes such as size and income. Agents can select the stores by evaluating a bundle of attributes such as distance from home and work, location in the agglomeration, price, assortment, quality etc. The simulation is implemented using the SeSAm multi-agent simulation shell (www.simsesam.de), and the model is tested comparing the calculated with the simulated turnovers of the stores. Another example of application of multi-agent simulations to retail markets is van Leeuwen et al. (2007). A discussion about possible applications
of micro-simulations in spatial analysis is followed by a small scale simulation example. However, the simulation exercise is conducted “by hand” and, again, only the consumer side, in this case at the household level, is described with agents. An exception can be found in Lombardo et al. (2004), where both sides of a retail market have agent-based descriptions. The multi-agent system is integrated in a GIS. The aim of the consumers is to reach the stores with minimal generalized costs. The aim of retailers is to maximize their profit. Shortcomings of this work are the summary description of both agents (basically within an agent type they are completely undifferentiated) and the simple way in which the environment is depicted (only eight macro-zones with few links connecting them in which 80,000 consumers and 12,000 retailers are allowed to evolve).


3 Modeling framework

As a first step to a fully agent-based representation of the simulated world a new agent type for retailers is introduced in the model system. Their goal is to optimize the location of their shops, in order to maximize the number of customers. By this, the model should be able to capture the mutual interdependency of retailers' location choices and individuals' shopping location choices. In the future, other typical choices of retailers, such as price policies will be modeled. All the attributes and the functionality of retailer agents are specified in a newly created module which has been embedded in the simulation toolkit MATSim.

3.1 MATSim

MATSim is a fast, dynamic, agent-based and activity-based microscopic transport modeling toolkit. The basic idea is to let a synthetic population of agents act in a virtual world. The synthetic population reflects census data while the virtual world reflects the infrastructure such as road network, land use, and the available transport services and activity possibilities. Each agent has its daily activity plan, which describes the chain of activities that it needs to perform in the virtual world. Each agent tries to perform optimally according to a utility function that defines what is useful for an agent. One virtual day is iteratively simulated. From iteration to iteration a predefined amount of agents are allowed to change some of their daily decisions to get a higher utility. The iterative process continues as long as the overall score of the population increases. The equilibrium reached represents what real individuals do in the real world.

More technically, MATSim is a toolkit composed of different modules. Each module is responsible for one part of the whole process. A module can have an underlying model (e.g. the traffic simulation, the mode choice, etc.) and can work together with, but also independently from, other modules. In this sense, MATSim can be seen as a comprehensive, flexible, framework, which simulates the daily life of persons and produces travel demand as a side product.
Each agent has socio-demographic attributes like age, gender, occupation, home location, car availability, etc. His plan contains information on the daily activities, like where and when those activities will be performed, and which mode of transport will be used to reach the different locations. The underlying activity-chain is assigned to each agent according to its socio-demographic attributes. The plans are executed simultaneously during the traffic flow simulation. Several plans for each agent are retained, given a score, and compared. The plans with the highest scores are kept, and used to create new plans based on the agent’s previous experiences. In order to improve their score the agents can vary their departure time, transport modes and routes. The system iterates between plan generation and traffic flow simulation until a relaxed state is reached.

MATSim’s most prominent application is a simulation of the travel behavior of the entire Swiss population, where 7.5 millions of agents are simulated, and about 2.3 million individuals are travelling by car on a network with 882,000 links. Additional information on MATSim can be found in Meister et al., (2010), Rieser (2010).

3.2 Retailer agents

The scope of the work presented here is to introduce the location behavior of retailers in the framework. In the simulated world activities may be performed in different places called facilities. Each facility is an entity with following attributes: type, location, capacity, opening time, closing time. In a single facility one or more activities of different types can be performed (home, leisure, education, shop, work). Each activity type of a facility has a capacity, which defines the maximum number of agents which are permitted to perform a given activity in this facility at the same time. The focus here lies on shop facilities, interpreted as a retail store. In the current standard scenarios used to run MATSim shop are undifferentiated, no particular types of shops are distinguished. The retailer agent is represented as the decision maker having the control on a certain number of shop facilities. The retailer agent, in general, does not necessarily represent an individual (i.e. the owner of the shop) but for example it might also be the board of a retail chain. According to most of definitions of agents in the artificial intelligence literature (see for example Ferber, 1999), this entity can be provided with attributes, knowledge, one or multiple objectives, a strategy to pursue the objective(s), a methodology to implement its strategy and a group of allowed choices. The retailer agents in-
roduced here are relatively simple, but still satisfying this definition. A retailer agent controls some of the shop facilities of the simulation scenario, which will have certain location and opening times in turn. The knowledge of a retail agent is in principle of two types; knowledge about customers and knowledge about competitors. The knowledge about customers will be limited to the number of primary activities of individual agents (in the MATSim framework home, work and education activities are primary activities) happening in a determined area. However, the retail agent knows how many customers have shopped in one of his stores after each iteration. The retailer is also able to see the location of competitors’ shops. The choice of a new location is made taking into account such information. Another information which can be easily added to the retailer’s knowledge is the land use regulation. Only land suitable for commercial use will be allowed as a new location. The objective for a retailer agent is the maximization of the number of customers. This is a relatively rough proxy for the maximization of revenues, but since the expenditure for shop activities is not yet modeled, this is the only way to model this at the moment. The methodological level is the way the retailer will effectuate the choice, in this case the location choice. The market support analysis will be the technique used by retailer agents. This technique is simple but still used in practice by retailers (Birkin et al., 2002; Rogers, 2007). The caption area of the store is estimated along with the population in it. An inventory of potential competitors of the store in that area is also compiled. Using this information, potential locations are evaluated and the best possible are chosen. In this first stage of the implementation of retailer agents, the only choice dimension available is location. Other choices will be allowed in the future work, such as changing price level or opening times.

3.3 The retailers relocation module

The retail relocation module implements a model of the retail market. This model is conceived to seek a stable equilibrium in the facility location/facility usage interaction. This is similar to the model defined by Timmermans and Arentze (2007), while other integrated land use-transport models are seeking a time path. The optimization of the supply side is separated from that of the demand side. The retailer module is not part MATSim’s main loop, in which the demand is optimized. More precisely, the demand for a given scenario is estimated and then the supply side optimization tool is feed with this demand. With this input, retailer agents try to improve the location of their shops. This produces a new scenario, which is giv-
en back to the optimization tool. This process can be performed once, or iterated as many times as one likes, either until a relaxed state or a fixed number of iterations are reached (Fig. 1).

Figure 1  Loop for the retailers module

3.4 The relocation process

The main idea behind the relocation process is that retailer agents try to find a constellation of shops which is more convenient for potential customers than the actual one. They will have the possibility to relocate their shops at one, or more, specific point of the iterative simulation process. The available links are fixed before the simulation starts and are updated during the relocation process according to retail agents’ moves. Retailers are allowed to relocate their shops sequentially, which means first retailer 1, than retailer 2, and so on until the last retailer has relocated its shops. The relocation of shops is controlled by a specific genetic algorithm. As already mentioned, this is inspired by a location methodology called marked support ana-
ysis. The algorithm seeks to find, for a given retailer, the constellation with maximum accessibili-
ity for potential customers, given the initial constellation and a set of free links. More precisely, for a
given, hypothetical, constellation of shops is drawn a caption area with a 3 km radius. For all potential
customers – the individual agents having the corresponding shop activity in their daily plan – having a primary activity within the caption area, the generalized travel cost to the reach the shop is computed. An average is made for each of the caption areas and those averages are summed up. This sum is the value which is minimized by the algorithm. In a more formal way we can write:

\[
\min \sum_{i} \left( \sum_{j} \frac{C_{ij}}{m} \right)
\]

Where \( n \) is the number of shops of the considered retailer, \( m \) the number of potential custom-
ers for the shop \( i \) at the given location and \( c_{ij} \) is the generalized cost of travel for the individu-
al agent \( j \) to travel to location \( i \). The fact that the generalized cost of travel is taken into ac-
count, is an important point. This is not only more precise than using simple distances, but it introduces in the optimization process time dependency (generalized cost are not constant during the day) and policy sensitiveness (for example road pricing measures can influence the cost of one trip between two given points).
4 Test Case: the metro area of Zurich (Switzerland)

4.1 The simulation scenario

The set up of a simulation scenario is a time consuming task, involving the integration of different data sets. The description of this process is beyond the scope of this paper. More information can be found in (Ciari et al., 2007). The scenario used here is a “Greater Zürich” scenario. It is a subset of the Swiss scenario, and covers an area of about 2800 km$^2$, obtained by drawing a 30 km circle around the “Bellevue” place in the centre of Zurich. This scenario is built with geo-coded data from the year 2000 population census (individuals, households, commuting matrices), the year 2000 census of workplaces (facilities by type and capacity) and the national travel survey for the year 2005 (477 types of activity chains, 9429 types of activity chains classified by duration; eight classes of agents by age and work status are distinguished). The study area has approximately 1 million inhabitants. Moreover, the scenario contains all agents that have plans with at least one activity within the area and all agents crossing the area during their travel. Transit traffic through the country is included based on relevant border survey data.

In the scenario two retailers, with respectively 29 and 17 shops, are represented. The initial locations of these shops are taken from real store locations of two real Swiss retailers, leaders in the grocery market. The locations which are considered free for the retailers to move in have been taken among the locations which didn’t host already a shop facility in the current initial scenario. Candidate locations have been selected randomly, but added to the free-for-relocation links only if some requirements were fulfilled. For each candidate location, an hypothetical caption area around it has been considered. Only link with a number of potential customers above a fixed, arbitrarily defined, threshold within the caption area could be taken. Moreover, also the ratio obtained dividing the number of potential customers for the available shop capacity in the area (computed as the sum of the capacities of all the shops in the area) needed to be above another given threshold. The first of these two condition tries to ensure that only “good” candidates are taken for the optimization process, with the goal of reducing convergence time. The second is a very simple way to tale into account concurrence and
avoid cannibalism (divert customers from its own shops placing another one too close), since this ratio is a measure of the potential still to be exploited in the area. Instead of this process the software is also able to take a list with the relevant free locations and their coordinates as input. This is also a simple way to eventually represent land use restrictions in the scenario.

The road network model has more than 236,000 directed links and more than 73,000 nodes. It is obtained from the Teleatlas navigation network. The number of facilities for out-of-home activities is 373,155. The transport modes allowed are: car, public transport, bicycle, and walk. For computational reasons the simulation is run on a 10% sample of this scenario, which means that 161,810 agents are actually simulated. The network capacity is also scaled (each link’s capacity is set to 10% of the original capacity) in order to have realistic traffic flows on the network links. In the 10% sample, the number of agents crossing the study area while transiting Switzerland is 5’791, linked to 880 home facilities outside Switzerland.

As previously mentioned, in standard versions, MATSim scenarios allow five different types of activities, and shop is one of them. In this particular case two different types of shop activities has been considered, grocery and non-grocery shop. It has been assumed that all retailers’ shops are grocery shops. The other shops allowing grocery shop activity have been chosen randomly, with the condition that the capacity of all grocery shops had to sum up to 20% of the total shop capacity (grocery plus non-grocery). This has been made because in the 10% scenario the number of facilities is not scaled down and some experiments showed without doing so the effects of the optimization were not large enough.

With the computer used for the simulation, a shared-memory machine of the type Sun Fire X4600 M2 with 8 dual-core CPUs and 128 GB RAM, the 10% sample scenario takes about 10 hours of computing time (using 3 cores and 40G RAM) for 50 iterations, which is enough to reach an equilibrium with the settings used.

### 4.2 Results and discussion

The simulation has been repeatedly run with different settings, each time for 300 iterations. The goal of the experiments was to show that the relocation of the shops brings an advantage in terms of number of customers for the retail agents. As already mentioned, it is possible to
run the relocation module more than once, but in the experiments described here the relocation module has been run only once. The reason to let the simulation run for 300 iterations, which is much higher than the 50 iterations which are usually necessary to obtain an equilibrium, has a specific reason. In fact, each run is composed of three phases. In the first, the simulation runs until equilibrium is reached, which means for 50 iterations. Equilibrium means that the average score for the agents is constant. We assume that at this point also travel time between two points is constant. This is important because the travel time for potential customers (or generalized travel cost) is what retailers seek to minimize through the relocation of their shops. The relocation of shops is the second phase and it goes as described in paragraph 3.4, outside the main MATSim loop. In the third phase we want the system to reach a second equilibrium point, but this time not in terms of average score of the agents, or not only in this terms, but also in terms of the number of customers in retailers’ shops. For this, we don’t have, so far, any general rule. However, after a number of experiments, it has been observed that after the relocation of the shops, between 200 and 300 additional iterations are necessary to reach a stable point in that sense. In figure (2) is shown a comparison between two simulation runs. In one run retail agents were not allowed to relocate their shops, in the second they were allowed to relocate them after 50 iterations.

Figure 2  Optimized move vs. No move
This basically means that we are measuring the performance of our optimization comparing the number of customers visiting the relocated shops against the number of customers visiting the shops in their actual locations. The relocated shops provide a better performance and the increase in the number of customers is about 10%. Note that in this figure the sum of customers of both retail agents of this scenario is considered. In fact, the two retail agents are optimizing their network of shops separately, the one after the other. A more detailed look at the effects of the optimization on each of them is provided in Table (1).

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Summary of the simulation’s results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Retailer 1</td>
</tr>
<tr>
<td>Number of Shops</td>
<td>29</td>
</tr>
<tr>
<td>Shops Moved</td>
<td>12</td>
</tr>
<tr>
<td>Shops Moved %</td>
<td>41%</td>
</tr>
<tr>
<td>Customers (Move)</td>
<td>3541</td>
</tr>
<tr>
<td>Customers (No move)</td>
<td>3199</td>
</tr>
<tr>
<td>Increase %</td>
<td>10.6%</td>
</tr>
</tbody>
</table>
As it can be observed, the number of customers is increasing for both of them, which means that also the retail agent which is allowed to move second still have enough room to optimize its locations. However, it is also interesting to note that not so many shops for each retail agents have been moved, respectively 12 and 6. A possible explanation is that we are comparing the results with the actual locations of the shops which are likely to be already “near-optimal”. Indeed, other experiments showed that compared with a random constellation of shops (before to start the simulation, retail agents’ shops have been all moved randomly picking up the new locations among a set of randomly chosen links) the increase was considerably larger. This also means that actual locations perform much better than random locations, which was to be expected, since they have been, supposedly, carefully chosen by the retailers. In this sense, an increase of about 10%, compared to actual locations, seems a good performance.

4.3 Conclusions

The results that we have showed are consistent with the expectations. Retailer agents are able to increment the number of their customers in the simulation by relocating some of their shop facilities. Since the scenario used for the experiments is a realistic one, it seems safe to affirm that this methodology could be used in the real world by retailers in order to optimize the locations of their shops. This is particularly true for retailers which are approaching a new market and, thus, are confronted with the problem of finding the optimal location for a given number of shops. Nevertheless, the model should be tested thoroughly, observing the sensitivity of the results to the different parameters of the optimization algorithm. In fact, there is no guarantee that an optimum has been reached, and other constellation of locations might perform even better than those found.
5 Summary and future work

This paper reported on work which aims to overcome some of the limits which integrated traffic and land use models traditionally have. Those models tend to have a rich description of one of the two sides, the demand or the supply, paired with some market clearing mechanism representing the other. In most of the cases is the traveler which is described more precisely. The work presented here is a first step toward an improved modeling tool in which all the actors of the urban system are represented with richness of details and at the individual level. This has been done building on an already existing traffic simulation tool called MATSim. The agent paradigm, already implemented in this tool for the representation of individuals, the demand side, has been extended to retailers, which are part of the supply side, through the creation of specific retailer agents. Those new retail agents are able to collect information on the behavior of individual agents, and can implement a strategy in order to reach their goals. In the specific case they try to maximize the number of customers by relocating their shops. The results for a test scenario, the metro area of the Swiss city of Zurich, show that retail agents in the simulation are able to increase the number of their customers by relocating their shops.

The extension of the agent-based approach to the supply side proposed here is still at an early stage of development. Nevertheless, it has been shown that this technique has the potential to become a powerful instrument in the hands of both location planners and policy makers. For location planners the emphasis lies on the location optimization aspect, and in this sense the tool can be further refined, at least in two ways. The first is seeking a better calibration for the parameters of the genetic algorithm lying behind the optimization process, which can significantly enhance the results (in terms of improvement of the number of customers). The second is making the model further realistic, for example introducing other typical choice dimensions of retailers, such as price policies, which are at the moment neglected and introducing, in an automatic fashion, the evaluation of the trade-off between the cost of shops’ moving against the estimated increased turnaround.

For policy makers the most interesting aspect is to have a tool which is able to predict the global system outcome of a given policy. In this sense, the most important aspect lacking in
the current model is probably the competitive bid process among the different retailers for the available locations. The modeling of the legislative context, even if in a really simple fashion, and not used in this experiment, is already possible.

Furthermore, a validation of the model has been performed only in terms of traffic counts among the network. However, a specific validation regarding the number of customers shopping in one store could be easily performed as soon as the relevant data will be available. In this sense, it can be expected that the use of a 100% scenario will help in order to have more reliable results. Finally, the supply is composed by various actors of different nature – planners, public administrators, legislators, etc. – and an effort should be made also in the addition of further agent types to represent also these actors, and their underlying optimization process, at the individual level.
6 References


