Trajectory Regression Model for Indoor Pedestrian Flow Analysis on Billboard Evaluation

Thomas Liebig[1,2]

[1] Fraunhofer IAIS, Schloss Birlinghoven
53757 St. Augustin, Germany
[2] Dept of Computer Sciences III, University of Bonn,
53117 Bonn, Germany
thomas.liebig@iais.fraunhofer.de

Abstract. Over the last few years new measurement technology has revolutionized the performance measurement in outdoor advertising. A handful of pioneer countries trace personal mobility now via GPS devices, which allows for precise performance results of arbitrarily positioned outdoor poster campaigns. However, GPS technology has the drawback that it cannot be applied indoors due to signal loss. In Switzerland and Germany many valuable posters are situated in public buildings such as train stations or shopping malls and their evaluation is of high interest. In this paper we therefore present a new approach for the evaluation of mixed indoor-outdoor campaigns. Our approach consists of a pedestrian movement model denoting quantities and trajectories of the people in restricted spaces. The model is supported by empirical traffic observations and can be integrated into standard trajectory evaluation. Our approach has been implemented for 27 major train stations in Switzerland.

Keywords: pedestrian flow, billboard evaluation, regression, trajectory

1 Introduction

Billboard advertisement is one of the most typical advertising media and still plays an important role in contemporary advertisement mixtures. This is supported by its turnover of 684 million CHF (about 460 million Euro) 2008 in Switzerland and 805 million Euro in Germany [1,2].

Nevertheless, due to the competition with other media (including television, radio and press) and the emergence of digital and online media the market changed rapidly in recent years. To become incorporated by media planners in an advertisement mix, transparent measures are needed for the performance of a campaign. Typical measures are (1) the coverage or reach of a campaign; this is the percentage of persons within a target group defined by socio-demographic attributes that has had contact with a campaign in a certain time interval (often one week). And (2) the number of contacts this group has had. Improved methods for audience measurement have become available in the last years due to technical advances and improved methodology. For example, GPS technology has established itself as a new standard in Switzerland and Germany, greatly improving the possibilities of fine-grained media planning [3,4]. To allow customized advertising, it will become necessary to take account of the time of day.

This paper focuses on the large amount of billboards which are situated at places where no GPS signal is available and other methods of traffic monitoring need to be applied. These places include for example airports and train stations but as well exhibition centers. The performance of these poster locations is highly interesting for media planners because they are considered among the most valuable ones. We will investigate the special case of inner train-station campaigns in Switzerland in this paper, which covers 2,600 poster locations. We are going to compute pedestrian flows for these stations, which utilize evaluation of the campaigns within the biggest 27 Swiss train stations. The challenge we are focusing on is to answer: Where do how many people move and which paths are used?

Tracking pedestrians with cameras, which might produce sufficient empirical data to answer this question and seems to be a perfect choice at the first sight, is often not permitted in train stations because of privacy restrictions. Radio Frequency Identification (RFID) and Bluetooth [5] technology may also be used for tracking, but this becomes expensive and requires an additional infrastructure for deployment of the necessary hardware. Another option for trajectory recording are interviews. However, they are very time consuming and thus expensive.

Our approach bases on relatively inexpensive measurements $m(x)$ of pedestrian quantities $q(x)$ at predefined locations x, that return the number of pedestrians passing by within a fixed time-slice, of course including an additional perturbation $\varepsilon << q(x)$.

$$m(x) = q(x) + \varepsilon$$

We are focusing on pedestrian analysis within train stations which are small, compared to cities or highway road networks. The pedestrian flow model needs to be (1) adjustable by measurements of peoples quantity in order to represent the movement in the public building. Additionally, another requirement is (2) fulfillment of the Kirchhoff-laws at each crossing. This means that the number of incoming equals the number of outgoing people for any closed boundary.

These requirements are challenging to existing micro-scopic simulation models (see the next section) due to their difficult adjustability by empirical data [15]. Our ansatz is based on regression approaches [11-13] that are so far used for quantity estimation. We extend the regression by a micro-scopic, i.e. trajectory-based, pedestrian model in order to end up with quantities and trajectories as required by the billboard evaluation task. The method we present is applicable in practice and was successfully applied to the 27 biggest Swiss train stations as part of an industrial application.

The paper is structured as follows. In section 2 we discuss related work. Next we describe in section 3 our regression model and the necessary assumptions we make according to our application. Section 4 addresses this industrial application: Motivated by reach estimation we estimate pedestrian trajectories and quantities for Zurich central station. We conclude with a section on future work.

2 Related Work

To the best of our knowledge, previous work concerning poster campaign evaluation only deals with outdoor campaigns. For the estimation of traffic flows in general a number of methods and algorithms exist in the literature. Initially being an operation research problem concerned with logistics and transportation issues, with increasing map-sizes it became a problem for spatial data mining. First, there exists a large group of probabilistic micro-simulation models including Monte Carlo methods [6], Markov models [7], cellular automata [8] and multi agent simulation [9]. These models describe individual pedestrians and model their behaviour within the spatial manifold, which could be represented discrete as done by the cellular automata model or continuous as, for example, required by the social force model [14]. The advantage of these microscopic models is that they allow including socio-demographic attributes and thus varying types of movement behaviour out of the box. Furthermore, they directly return individual trajectories, the macro-scopic traffic features (e.g. quantities and densities) become derived afterwards.

Second, there are large-scale macro-scopic algorithms for traffic quantity prediction in (extensive) road networks [10] as well as macro-scopic models that base on gas and fluid-dynamic processes [14]. These models return desired traffic quantities but no individual trajectories. Whereas the scope of the first model [10] is just on prediction of quantities (there called: frequencies) the latter [14] is time-dependent and models multiple aspects of macro-scopic traffic features as functions over space x and time t, e.g. quantities $q(x,t)$, densities $\rho(x,t)$, velocities $v(x,t)$. Based on this notation, figure 1 gives a comparison of the micro-scopic and macro-scopic view on pedestrian movement.

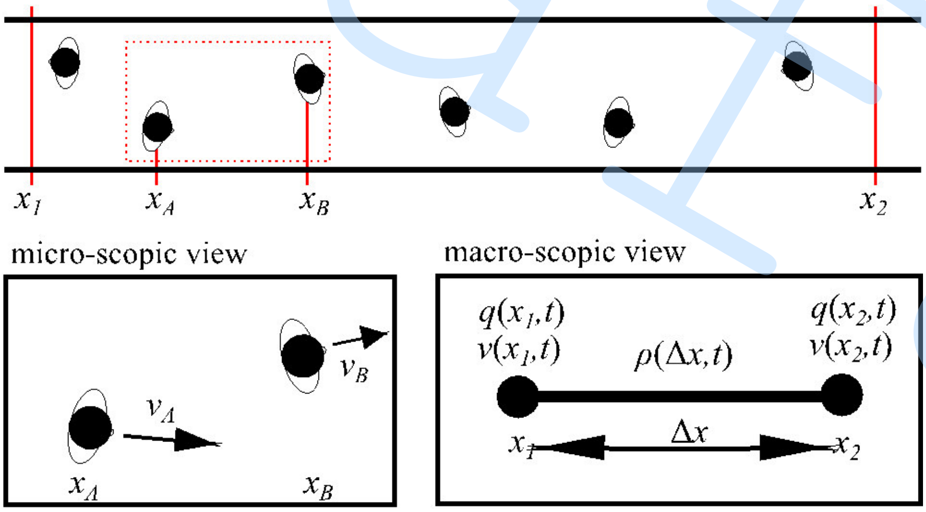

Figure 1: macro-scopic and micro-scopic view on traffic

Each of these models and algorithms makes certain assumptions on the mobility behaviour or trajectory choices, reflecting different aspects of real-world traffic. Thus, model selection depends mainly on the application. Micro-simulation models are useful for evacuation planning and obstacle detection. Macro-scopic approaches are used to plan traffic control systems and to simulate the influence of infrastructural changes.

We are focusing on railway stations which are small, compared to cities or highway road networks. The model needs to be (1) adjustable by measurements of peoples quantity in order to represent current movement in the public building. An additional requirement is (2) fulfillment of the Kirchhoff-laws at each crossing. This means that the number of incoming equals the number of outgoing people for any closed boundary.

These requirements are challenging to micro-scopic models due to their difficult adjustability by empirical data [15]. Our ansatz is based on regression approaches [11-13] that are so far used for quantity estimation. We extend the regression by a micro-scopic, i.e. trajectory based pedestrian model in order to end up with quantities and trajectories as required by the billboard evaluation task.

3 Trajectory Regression Model

Some of the pedestrian models presented in section 2 require detailed representations of the accessible space. Applications for such models are emergency and evacuation planning, capacity analysis or obstacle detection. As we are only interested in quantities and trajectories to identify co-visits of locations, we do not require such a detailed model. Therefore, a directed graph approximation $G=(V,X)$ of the floors, stairways and junctions contains enough information for our task.

Every junction is represented by a vertex v and the connecting floors are represented by pairs of edges x_+ and its opposite direction x_- describing the set $X \subseteq V \times V$. Paths p through the station may then be described by a sequence of edges, starting and ending at an entrance or platform. Quantities $q(x)$ denote the number of people passing a certain edge within a fixed and pre-specified time-interval. The presented model focuses on presence of pedestrians and thus may disregard temporal aspects of movement.

$$q(x) = \int q(x,t)dt$$

Combined with the requirement to hold Kirchhoffs-laws at any junction the modeled time window needs to be chosen appropriate. As nobody lives within public buildings or starts to reside there permanently, the traffic for one day may be considered to hold Kirchhoff: $\sum q(x_+) = \sum q(x_-)$, for all edges x_+,x_- adjacent to any arbitrary (but fixed) vertex $v \in V$.

Empirical recordings of pedestrian presence $m(x)$ denote quantities $q(x)$ plus some noise ε at pre-selected edges $x \in X_M, X_M \subseteq X$. Since the quantity of persons at a position x is given by the number of people walking on each connected path $p \in Paths \subset 2^X$ that passes that location

$$q(x) = \sum q(p), \forall p \in Paths, x \in p,$$

our approach consists of three steps: (1) The enumeration of the set of all valid paths *Paths*, afterwards (2) the computation of path frequencies $q(p), p \in Paths$ such that the error at the measured locations becomes minimized. And, (3), the estimation of the number of people per edge $q(x)$. The three generated output variables $q(X)$, $q(Paths)$ and the set *Paths* are going to hold the desired information to answer the billboard campaign evaluation task (see section 4 for details). By the observation, people prefer routes with the minimal detour most, when walking from a given start to a target [9], the assumption that most trajectories are un-cyclic is reasonable for public buildings. Empirical observation, we made during the application (section 4), also supports this assumption. We consider cyclic paths to be irrelevant for our study. Therefore, the set *Paths* becomes enumerable, utilizing the previous annotation of edges. By application of a characteristic function a_{ij} for a path p_j on the edge-set X, which maps p_j on a binary vector with length $|X|$ and equals one at position i, iff x_i is element of p_j and is zero otherwise; we can rewrite previous function for all edges by a matrix product as follows:

$$A = (a_{ij})_{0 \le j \le |Paths|}^{0 \le i \le |X|} \qquad , a_{ij} = \begin{cases} 1 & \text{if } x_i \in p_j \\ 0 & \text{otherwise} \end{cases}$$

$$q(X) = A^T \times q(Paths) \qquad , q(X) = (q(x_1), \ldots, q(x_{|X|}))^T \text{ and } \quad .$$

$$q(Paths) = ((q(p_1), \ldots, q(x_{|Paths|}))^T$$

The program we solve in step (2) is derived directly from the last equations and given by this formula:

$$q^*(Paths) = \underset{q(Paths) \ge 0}{\arg\min} \left(\left\| {A_M}^T \times q(Paths) - m(X_M) \right\| + \left\| d(Paths) \cdot q(Paths) \right\| \right)$$

Here, the objective function includes an additional term $||d(Paths) \cdot q(Paths)||$, which represents the preference of pedestrians to choose routes from start to goal in a detour d avoiding manner [9]; $d(p)$ is given by the ratio of the length of p divided by the length of the shortest path connecting the same start and target as p. This term not just penalizes detours, but has the advantage to make a specific solution more invariant on the used solver, because it further restricts the set of possible solutions $q^*(Paths)$. The Matrix A_M is constructed similar to previously generated A, except that X is limited to the sensor locations X_M.

The number of people per valid path that traverses the train station G, computed in the previous step, is applied to the equation above, in step (3). This results in edge quantities $q^*(X)$ for all locations X in G covered by valid paths which contain at least one measurement.

$$q^*(X) = A^T \times q^*(Paths) \quad , q^*(X) = (q^*(x_1), ..., q^*(x_{|X|}))^T \text{ and}$$

$$q^*(Paths) = ((q^*(p_1), ..., q^*(x_{|Paths|}))^T$$

The goal of the flow estimation was the calculation of the number of persons per position and a quantity distribution among their chosen paths. Both is given in result by our method and stored in the objects q^* (X), *Paths* and $q^*(Paths)$. The next section discusses its properties. An application to the industrial case is described in section 4.

Properties

The empirical measurements include small deviations. Thus, the combination of these counts in target quantities must lead to contradictions. At junctions, they can be easily recognized in the raw data by violations of Kirchhoff's law, whereby the number of incoming people has to equal the number of the outgoing ones. In general, this problem may arise along multiple junctions. In this case, it is harder to pre-identify the contradictions. Therefore, we require the frequency estimation algorithm to recognise such cases automatically and to eliminate them in the model, if required. Our approach fulfills all of these criteria. The Kirchhoff law holds automatically, because the enumerated path set does: Every single path holds the constraint at any junction that the number of incoming people equals to the number of outgoing. Multiplying with the path quantities, our algorithm increases the number of people on this path, but nevertheless the equilibrium remains fulfilled. If each single route fulfills this constraint, the set of all paths including their final quantities also does. Therefore, in our pedestrian model Kirchhoff's law holds. From this property follows directly an invariance of the method on graph homomorphisms for G. This means, if an edge x becomes divided into two connected parts x_i and x_j the resulting quantities remain the same. Kirchhoff's law ensures that $q(x)=q(x_j)=q(x_i)$. Furthermore, the small perturbances included in the countings are corrected by the least squares regression. Without the need for a pre-analysis of outliers the resulting quantities are chosen such that the differences at the measurement locations are minimal.

One weakness of micro-scopic models quoted in section 2 is their bad adjustability by empirical traffic observations [15]. This requires a re-run of the whole simulation and an adjustment of the parameters [15]. Each simulation run depends in time on the number of pedestrians and the length of the explored time-interval. In contrast, our ansatz scales well and does not depend on number of pedestrians nor the chosen time-interval, but the size of the station and the number of sensors. This property is important in our industrial case, where we made calculations for the 27 major train stations, including some with a daily usage of several hundred thousand passengers per day.

4 Real World Application on Zurich Central Station

The motivation to investigate the question for both (1) pedestrian quantities and (2) trajectories at the same time was given by an industrial use case, namely the performance evaluation of indoor poster campaigns. Thus, we successfully applied the method to the 27 major Swiss train stations. Within this section we describe the process for one example: Zurich central station. Afterwards, a method to combine the calculated model with existing outdoor poster campaign evaluation approaches becomes highlighted.

4.1 Pre-Processing and Empirical Data Collection

The data we base our analysis on are a floorplan image and a few empirical recordings of pedestrian quantities at pre-selected locations. In a first pre-processing step, the floorplan becomes tessellated by a traffic network $G=(V,X)$ representation as follows: Junctions are represented by vertices v which then forms the set V. The connecting floors are represented by two edges $x_+=(v_i,v_j)$ and its reverse $x_-=(v_j,v_i)$ ($x_+, x_- \in X \subseteq V \times V$). Entrances become annotated in this graph by edge attributes.

After doing a pre-study, we concluded that counting the number of people manually at several positions (using a smart phone application for data entry) is the most cost-efficient method for data collection. As noted in the introduction, using video cameras was not feasible because of privacy constraints. To decrease the influence of the day of week on the measurements, we repeated the measurements at three different days. As the number of "sensors" is limited, we had to select locations for counting in advance using the traffic network of the train stations. Therefore we located sensors at the most important junctions and stairways. Figure 2 depicts the measured edges at Zurich central station.

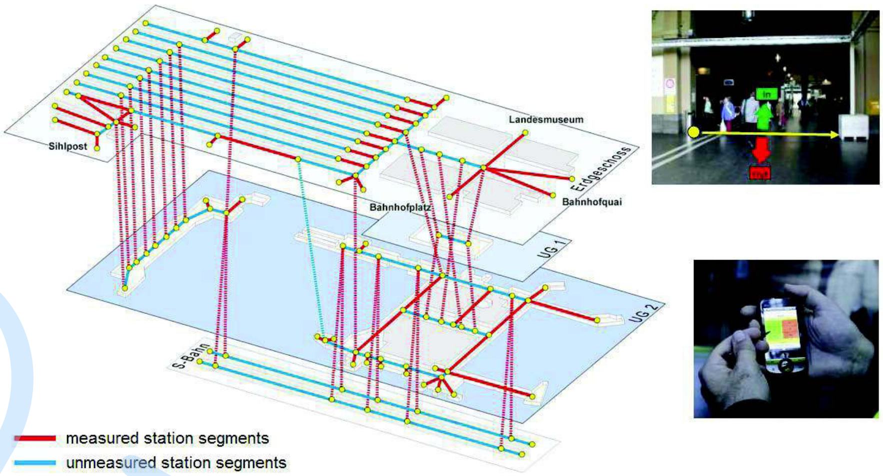

Figure 2: measurement locations in Zurich central station

To assist manual counting and to simplify post-processing of measurements, we developed a smart phone application (figure 1) which records clicks of the surveying person - each click represents the number of pedestrians passing in a specified direction - along with its timestamp. This enables an easy storage of the data in a database. Thus, we know how many people passed at which time into which direction. In an early prototype, we encountered the problem of mixed directions; therefore we added visual hints to the smartphone application as well as to the map. To distinguish directions, the colours red and green are used in our application.

To be able to compare the empirical raw data of pedestrians at measurement locations, e.g. an average number of pedestrians for a complete day or week, post-processing is necessary: after merging the measurements, the counted quantities are weighted and aggregated according to the time interval and day they were taken. As a result, every measured location in the train station has associated with it a number of pedestrians that may be compared against any other location. This is important for ranking locations or tracking segments within the building, which is a first feature of our general mobility model.

4.2 Flow Estimation

For segments where empirical measurements have been taken, the quantities are known. Our task is now to estimate them for the unobserved segments, and to build a pedestrian indoor movement model that is useful for poster and campaign evaluation.

In contrast to other regression models [11-13] that do not give trajectories but just frequencies, we tackle both questions at the same time, using a two stage regression approach. In a first step, we enumerate all plausible routes through the building and collect them in a route set. For example, at the main station in Zurich, there are about

380,000 conceivable routes. Non-plausible routes are eliminated, among them circular routes. Afterwards, we assign frequencies to each route, based on the measurements. The measurements serve as frequency targets in this process. The purpose of this assignment procedure is to find the optimal combination of routes that fulfills all frequency targets. As a result we obtain for every modeled train station (1) a set of routes crossing that station and (2) the number of people walking on each route. Figure 3 gives an example for Zurich central train station. With this information we are able to calculate quantities for each edge $q(x_i)$ in the station by summing over route quantities, no matter whether the edge x_i has been measured empirically or not. This yields the pedestrian movement model based on empirical measurements we aim for. It enables us to denote pedestrian quantities at any location and gives trajectories also at unobserved segments.

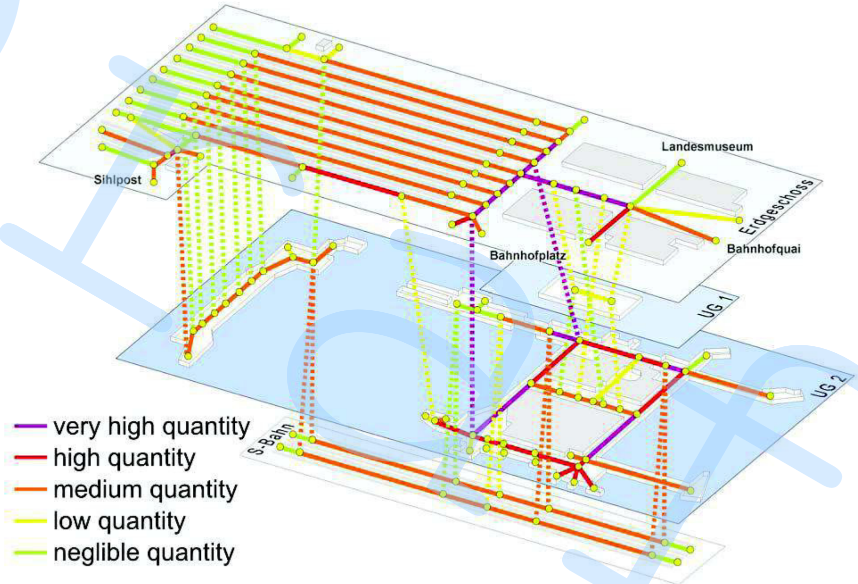

Figure 3: estimated quantities in Zurich central station

4.3 Integration with GPS surveys

In order to apply our pedestrian movement model for the evaluation of mixed indoor-outdoor poster campaigns, we still need to integrate it with GPS mobility data as mentioned in the introduction. This means we have to assign for each GPS person who enters a railway station a corresponding route through the station. To achieve this goal, we define three subsequent steps: (1) visit identification, (2) route assignment and (3) performance evaluation.

In the first step we identify all test persons within the GPS sample visiting a train station. Based on GPS trajectories of over 10,000 test persons recorded over a period of one week per person, we isolate all tracks in the vicinity of a train station using buffers and the spatial join operation "intersect". As GPS signals may be noisy, we apply an individually sized buffer to each of the train station geometries, reflecting its specific local setting. The resulting candidate set, however, contains not only potential rail travelers but also regular pedestrians, car drivers and passengers passing by the station without entering it. We therefore apply a complex multi-level filtering process which identifies the visitors of a train station using, for instance, speed curves, the course of movement and time spent inside the geometric extension of the train station. Knowing all visits to a railway station completes step 1.

Step 2 is the assignment of each visit to one of the routes underlying the pedestrian movement model. The challenge of this task is to find an optimal distribution of personalized routes given the route frequencies. We do this iteratively by drawing routes from the route set and considering the projected weight and socio-demographic information of each test person being assigned to that particular route. At the end of this process each GPS trajectory containing a visit to a train station, as identified in the previous step, has been assigned a route through the corresponding train station.

Finally in step 3 we weight poster contacts and calculate performance measures of mixed in- and outdoor campaigns. Similar to the performance evaluation of outdoor posters, we consider individual visibility criteria at each poster site. Routes passing the visibility area are weighted according to the contact quality, depending e.g. on the viewing direction or clustering of panels. Given weighted contacts for each indoor poster and visiting person, we can estimate total contacts and reach of a mixed indoor-outdoor campaign using the same algorithmic background as for outdoor campaigns. The selection of a campaign and of a target audience determines all relevant (indoor and outdoor) poster contacts and the application of Kaplan-Meier compensates for missing measurement days in the GPS data as described in [4].

5 Conclusion and Future Work

In this paper we developed a method that allows performance measurements for billboards that are placed indoors. We focused on 2,600 poster sites in railway stations as those are being seen as one of the most valuable over all. The challenge results from GPS signal loss inside buildings. Our approach includes the development of a pedestrian model based on empirical data. This mobility information has been integrated with existing GPS mobility data, allowing to infer reach values and weighted contacts. We applied our approach to 27 major Swiss train stations.

Although we showed how to implement a general movement model within the train station which is used for poster campaign evaluation, we do not model time, so far. Our indoor model is as static as the cited outdoor evaluation methods. While for the current needs of outdoor performance measurements a static model is sufficient, in a future world of pervasive advertisement a dynamic model that uses the time dependencies of the measurements, would bring immense benefits. In the future, modeling them with Gaussian Processes enriched with relational data sources, as train schedules or text messages like news tickers, is a promising method. It will enhance our method and allow for more target group specific campaign measurements in the future. In combination with persistent frequency sensors, we aim to model real-time poster evaluations.

Acknowledgments. The authors were partly supported by SPR plus and the European ESS project 217951.

References

1. Stiftung Werbestatistik Schweiz: Werbeaufwand Schweiz, Erhebungsjahr 2008 (Advertising expenditure Switzerland, survey year 2008) (2009)
2. Fachverband Außenwerbung e.V.: Netto-Werbeeinnahmen erfassbarer Werbeträger in Deutschland, 2000-2008 (Net turnover of confirmable advertising media in Gemany, 2000-2008) (2009)
3. May, M., Körner, C., Hecker, D., Pasquier, M., Hofmann, U., Mende, F.: Handling missing values in gps surveys using survival analysis: a gps case study of outdoor advertising. In: ADKDD '09: Proceedings of the Third International Workshop on Data Mining and Audience Intelligence for Advertising, New York, NY, USA, ACM (2009) 78-84
4. Pasquier, M., Hofmann, U., Mende, F.H., May, M., Hecker, D., Körner, C.: Modelling and prospects of the audience measurement for outdoor advertising based on data collection using gps devices (electronic passive measurement system). In: Proceedings of the 8th International Conference on Survey Methods in Transport. (2008)
5. Alt, F., Balz, M., Kristes, S., Shirazi, A.S., Mennenöh, J., Schmidt, A., Schröder, H., Goedicke, M.: Adaptive user profiles in pervasive advertising environments. In: Proceedings of the 3rd European Conference on Ambient Intelligence (AmI'09). Volume 5859 of Lecture Notes in Computer Science., Springer (2009) 276-286
6. Borgers, A., Timmermans, H.J.P.: City centre entry points, store location patterns and pedestrian route choice behaviour: A microlevel simulation model. Socio-Economic Planning Sciences 20(1) (1986) 25-31
7. Yu, G., Hu, J., Zhang, C., Zhuang, L., Song, J.: Short-term traffic flow forecasting based on markov chain model. In: Proceedings of the IEEE Intelligent Vehicles Symposium, 2003., IEEE (2003) 208-212
8. Blue, V.J., Adler, J.L.: Cellular automata microsimulation for modeling bidirectional pedestrian walkways. Transportation Research Part B: Methodological 35(3) (March 2001) 293-312
9. Rindsfüser, G.: Simulation des Aktivitätenplanungsprozesses mit Methoden der Künstlichen Intelligenz. Straßenverkehrstechnik 49(3) (2005) 145-153

10 May, M., Hecker, D., Körner, C., Scheider, S., Schulz, D.: A vector-geometry based spatial knn-algorithm for trac frequency predictions. Data Mining Workshops, International Conference on Data Mining 0 (2008) 442-447

11 Pushkarev, B., Zupan, J.M.: Pedestrian travel demand. Highway Research Record 355(HS-011 999) (1971) 37-53

12 Tanyimboh, T.T., Templeman, A.B.: Calculating maximum entropy Flows in networks. Journal on Operation Research 44(4) (1993) 383-396

13 Yassin-Kassab, A., Templeman, A.B., Tanyimboh, T.T.: Calculating maximum entropy flows in multi-source, multi-demand networks. Journal on Engineering and Optimization 31(6) (1999) 695-729

14.Helbing, D.: Verkehrsdynamik: Neue physikalische Modellierungskonzepte, Springer, 1997

15 Kretz, T, Schreckenberg, M.: The F.A.S.T.-Model. In ACRI, 2006, 712–715