hEART 2017 - 6th Symposium of the European Association for Research in Transportation
September 12-14, 2017, Faculty of Civil and Environmental Engineering, Haifa, Israel

Pedestrian multi-class speed-density relationship: evaluation of integrated and sequential approach

Marija Nikolić, Michel Bierlaire, Iliya Markov, Romain Konde Olivier Bondo
Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering
École Polytechnique Fédérale de Lausanne, Switzerland
Email: {marija.nikolic, michel.bierlaire, iliya.markov, romain.konde}@epfl.ch

March, 2017

1 Introduction

The focus of this paper is the modeling of speed-density relationship for pedestrian movements using the potential of detailed tracking observations (Alahi et al., 2014). This relationship is in the literature usually specified under the assumption that the traffic system is at equilibrium, that is stationary and homogenous (Jabari et al., 2014). The empirical analysis we have performed rules out the use of a unique equilibrium relationship, due to a high scatter in the data. We attribute the observed scatter to the violation of the homogeneity assumption, and specify a multi-class model of the speed-density relationship. We propose the integrated approach and more data-driven sequential approach to the specification of the model, and evaluate their performance using real data.

2 Modeling framework

Let (v_i, k_i, X_i) be a triplet representing the speed v_i, the density k_i and the vector of observable characteristics X_i (e.g. age, trip purpose, etc.) associated with individual i. We assume that the population is partitioned into J classes of pedestrians. In this framework, the individual speeds are random variables. The distribution of v_i is characterized by its probability density function conditional on the density k_i experienced by individual i, and the class j

$$f_j(v_i|k_i, j; \theta_j(k_i)), \tag{1}$$

where $\theta_j(k_i)$ are parameters. We refer to this distribution as a class-specific model (CSM). To identify the classes of pedestrians we consider two different approaches, detailed below.

The first approach refers to the grouping of similar pedestrian trajectories using machine learning techniques (e.g. clustering). The similarity is determined using different measures, suitable in the context of pedestrian movements:

- feature-based similarity measures (e.g. based on trip purpose, traversed distance, etc.),
- shape-based similarity measures (e.g. L_p distance, Dynamic Time Warping Distance, EDIT Distance, Longest Common Subsequence, etc.).

We refer to this approach as the sequential approach, since it involves two stages: the grouping stage and the model estimation stage. In the second stage a separate speed-density model (1) is to be estimated for each discovered group.

In the second approach we assume that the heterogeneity may come from multiple factors, and that the class of each individual is unknown a priori. Based on this assumptions we propose a class membership model (CMM) (Walker and Li, 2007). The CMM provides the probability that a pedestrian i, characterized by her socio-economic characteristics X_i, belongs to class j

$$\Pr(j|X_i;\beta_j), \tag{2}$$

where β_j are parameters. The specification of the CMM is based on a fitness function, that is a continuous variable measuring how much individual i fits into class j

$$U_{i,j} = V_{i,j} + \varepsilon_{i,j} = CSC_j + \beta_j X_i + \varepsilon_{i,j}, \tag{3}$$

where CSC_j and β_j are unknown parameters to be estimated from data, and $\varepsilon_{i,j}$ is a random term. A specific distribution assumption for $\varepsilon_{i,j}$ leads to a specific probability model. The exact specification of $V_{i,j}$, and in particular the exact list of characteristics involved in X_i, is application dependent. In this approach the segmentation of the population and the movement behavior are modeled simultaneously by combining the CSM and the CMM as follows

$$f(v_i|k_i, X_i;\theta_j(k_i),\beta_j) = \sum_{j=1}^{J} \underbrace{f_j(v_i|k_i, j;\theta_j(k_i))}_{\text{CSM}} \underbrace{\Pr(j|X_i;\beta_j)}_{\text{CMM}}. \tag{4}$$

Therefore, we term it the integrated approach. For more details on the approach, we refer to Nikolić et al. (2017).

3 Applying the framework

The performance of the approaches presented in Section 2 is evaluated based on the data collected in the train station in Lausanne. The data set contains (i) individual pedestrian trajectories collected using a large-scale network of smart sensors during five days in February, 2013 (Alahi et al., 2014), (ii) infrastructure data, that is detailed plans containing the locations and dimensions of all relevant parts of the monitored system, and

(iii) the train timetable for the period under study. Based on available data we extract traffic condition measurements, that is individual speed and density values, as well as the characteristics of pedestrians, such as the pedestrian type (arriving, departing, transferring and non-passenger), the traversed origin-destination (OD) distance, the period when a pedestrian is observed (peak and off-peak), and the time to departure. Traffic condition data are measured using Voronoi-based method (Nikolić et al., 2016).

We assume the existence of two classes, denoted as C_1 and C_2, in order to keep the model parsimonious and to avoid potential over-fitting. We assume that the speed (1) follows Rayleigh distribution in each class in both approaches. The choice of this distribution is motivated by its properties (continuous distribution that is defined on the positive support) that are in accordance with the physical properties of the speed. The mean of the class-specific distributions is modeled as linear speed-density relationship.

In the first stage of the sequential approach, pedestrian trajectories are clustered based on the OD distance by using the k-means algorithm (Tan et al., 2005). The average OD distances of the identified clusters are 58.536 and 22.408 meters in C_1, respectively C_2. We thus term C_1 the class of "pedestrians associated with longer OD distances" and C_2 the class of "pedestrians associated with shorter OD distances". In the second stage a separate speed-density model is estimated for each cluster, using maximum likelihood procedure. Figure 1a shows that the estimated average movement behavior is not significantly different across the classes: pedestrians associated with longer OD distances (C_1) are characterized by slightly lower walking speeds, compared to pedestrians with shorter OD distances (C_2), and the sensitivity to congestion is similar in both classes.

In the integrated approach we assume that the error term in (3) is i.i.d. type 1 Extreme Value (EV1(0,1)) across classes and individuals. This assumption yields the binary logit CMM. The explanatory variables constituting the deterministic parts of the fitness function are the pedestrian type, the OD distance, the period and the time to departure. Figure 1b shows that the C_1 class pedestrians are characterized with a higher free-flow walking speed and significantly lower sensitivity to congestion, compared to pedestrians belonging to C_2 class. We thus call C_1 the class of "pedestrians less sensitive to congestion" and C_2 the class of "pedestrians more sensitive to congestion". The results of the integrated model also reveal what are the underlying factors leading to different movement behavior. For instance, they suggest that departing and transferring passengers are more likely to be in class C_1, compared to arriving passengers. This is expected, given that departing and transferring passengers are usually associated with higher time pressure that drives their more "aggressive" behavior. Also, when the time to departure and the OD distance increase, pedestrians are more likely to belong to C_2 class. Note that other integrated models with higher number of classes are estimated and compared using statistical tests. The results of this comparison show that the two-class model is superior for this case study, in terms of the fit and the behavioral interpretation of the results.

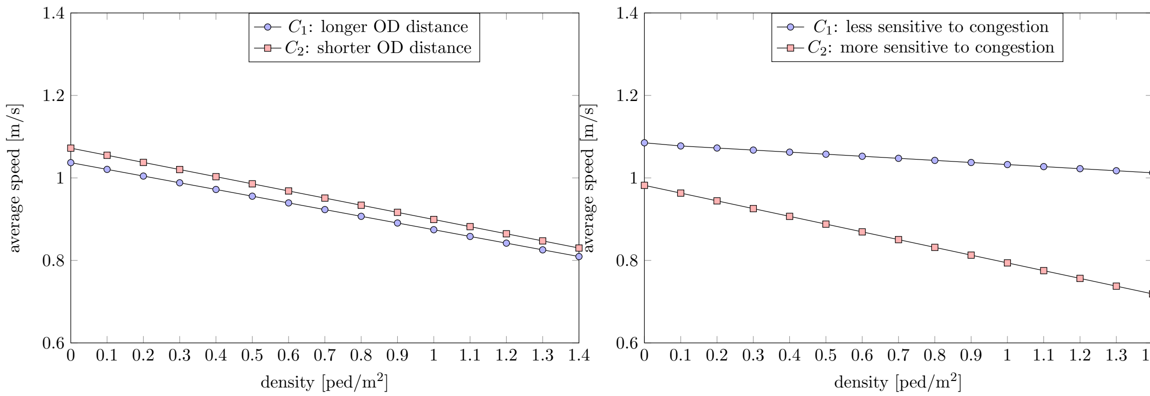
The distributions of the speed observations classified using sequential and integrated approach are shown in Figure 2, respectively Figure 3. The distributions are shown for different Level of Service (LoS) (Fruin, 1971). It can be observed that C_1 and C_2 are characterized with similar speed distributions for all LoS in the sequential approach. In the integrated approach, C_1 is characterized by a distribution of speed values that is shifted towards higher values, for all LoS, compared to the observations from C_2 class.

We also calculate the statistics related to the cohesion and separation of the identified classes, as defined in Tan et al. (2005). The cohesion shows how closely related are observations in a cluster, while the separation measures how distinct a cluster is from other clusters. Given that the integrated approach represents probabilistic model-based clustering analysis, the statistics are computed using simulation. The cohesion is similar in both approaches (around 0.242), while the separation favors the integrated approach (0.258), over the sequential one (0.243).

The presented analysis suggests that the integrated approach is able to better capture the underlying population heterogeneity and its impact on pedestrian movement behavior, than the sequential approach. However, we emphasize that the specifications proposed here are merely one possibility. The framework described in Section 2 is general and allows for different specifications to be tested, which is our immediate next step.

4 Conclusion

The aim of this study is to examine and explain the impact of pedestrian heterogeneity on their walking speed. Motivated by the analysis of real-world data, we propose the multi-class speed-density relationship derived using the sequential and integrated approach. The preliminary analysis indicates that the observed pattern can be explained by the segmentation in the population. More satisfactory performance is achieved in the case of the integrated approach, which might be attributed to the fact that it (i) avoids any measurement error that may exist in the sequential approach, and (ii) directly uses the behavior of interest to define the segmentation in population, which is not exploited by the sequential approach. However, the presented analysis is the starting point in modeling and various specifications will be evaluated in the paper. We will consider other feature-based, and also shape-based similarity measures in the first stage of the sequential approach. Also, we will examine the performance of different class-specific formulations and different number of classes in both approaches. Finally, the evaluation of the approaches based on additional criteria, and on the comparison between the predicted and empirical distributions will be performed.



(a) Sequential approach

(b) Integrated approach

Figure 1: Class-specific speed-density relationships

References

Alahi, A., Bierlaire, M. and Vandergheynst, P. (2014). Robust real-time pedestrians detection in urban environments with a network of low resolution cameras, *Transportation Research Part C: Emerging Technologies* **39**: 113–128.

Fruin, J. J. (1971). Designing for pedestrians: A level-of-service concept, number 355, Highway Research Board, Washington, DC, pp. 1–15.

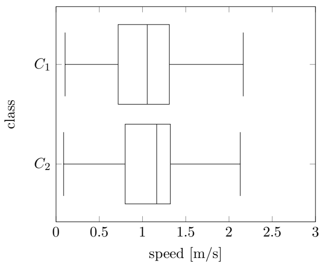
Jabari, S. E., Zheng, J. and Liu, H. X. (2014). A probabilistic stationary speed–density relation based on Newell's simplified car-following model, *Transportation Research Part B: Methodological* **68**: 205–223.

Nikolić, M., Bierlaire, M., de Lapparent, M. and Scarinci, R. (2017). Multi-class speed-density relationship for pedestrian traffic, *Technical Report TRANSP-OR 170115*, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne.

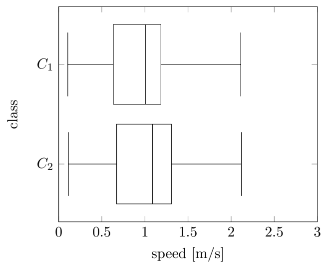
Nikolić, M., Bierlaire, M., Farooq, B. and de Lapparent, M. (2016). Probabilistic speed-density relationship for pedestrian traffic, *Transportation Research Part B: Methodological* **89**: 58–81.

Tan, P.-N., Steinbach, M. and Kumar, V. (2005). *Introduction to Data Mining, (First Edition)*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

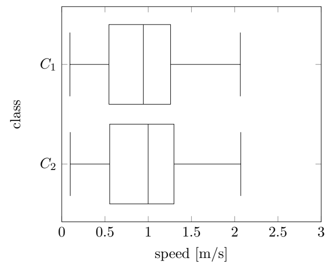
Walker, J. L. and Li, J. (2007). Latent lifestyle preferences and household location decisions, *Journal of Geographical Systems* **9**(1): 77–101.



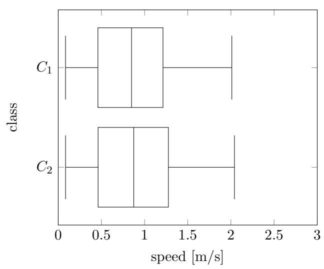
(a) LoS A



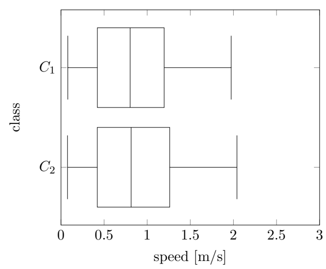
(b) LoS B



(c) LoS C

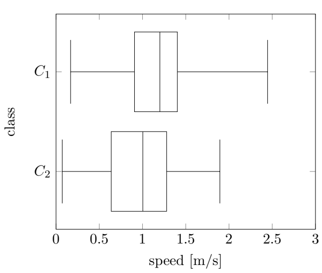


(d) LoS D

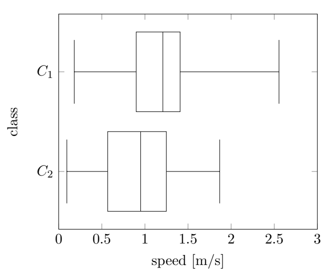


(e) LoS E

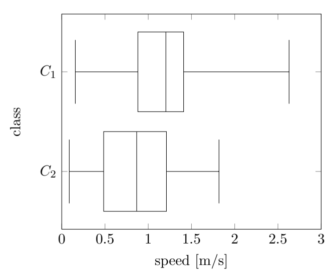
Figure 2: Sequential approach: speed distribution for different classes and different LoS



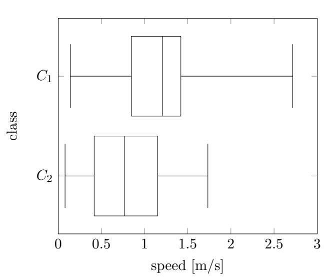
(a) LoS A



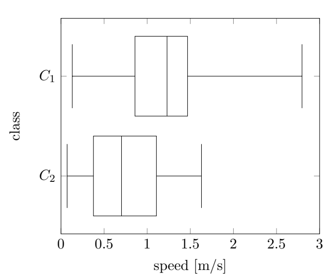
(b) LoS B



(c) LoS C



(d) LoS D



(e) LoS E

Figure 3: Integrated approach: speed distribution for different classes and different LoS