Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 151 (2019) 401–408

Procedia Computer Science

www.elsevier.com/locate/procedia

The 10th International Conference on Ambient Systems, Networks and Technologies (ANT)
April 29 - May 2, 2019, Leuven, Belgium

Bicyclist Route Choice: Data Exploration and Research Project Outline

Luk Knapen[a,b,*], Thomas Koch[c], Elenna Dugundji[a,c]

[a] *VU University, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands*
[b] *UHASSELT, Agoralaan - Gebouw D, 3590 Diepenbeek, Belgium*
[c] *Centrum Wiskunde & Informatica, Science Park 123, 1098 XG Amsterdam, The Netherlands*

Abstract

Microsimulation of travel flows aims to assess the effect of decisions taken by travelers based on personal preferences, time-of-day, properties of the infrastructure and expected or perceived travel flows. *Route choice* represents a particular class of such decisions. Route choice prediction is an essential component of microsimulators. Specification of choice models and estimation of the corresponding parameters based on observations are required in the preparatory stage. Route *choice sets* need to be established for sampling in the simulation stage.

This paper is part of a research project aiming to investigate how *route complexity* can be integrated in the choice process modeling. In particular routes for bikers collected by GPS tracking in the Dutch FietsTelWeek project in 2016 are analyzed.

The data exploration stage and the research project outline are covered. Properties of the publicly available `fietstelweek2016` dataset used for model training are investigated in order to assess their effect on prediction results. In order to achieve the project goal, the research project structure is briefly discussed. It is based on the observation that the number of routes recorded for each OD-pair is too small to observe a frequency distribution for complexity. Hence, complexity data are collected for sub-networks that are similar with respect to particular graph properties.

Keywords: GPS traces; bicycle; route; choice set

Nomenclature			
BFS-LE	Breadth First Search - Link Elimination	BPC	Basic Path Component
CDF	Cumulative Distribution Function	CSN	Consideration SubNetwork
DSCSG	Doubly Stochastic Choice Set Generation	LCP	Limited Complexity Paths set
MNL	Multinomial Logit	PWF	Probability Weight Function
POSDAP	Position Data Processing (ETHZ)	PSL	Path Size Logit
RL+LS	Recursive Logit + Link Size	SPT	Shortest Path Tree

* Corresponding author. Tel.: +32-11-26.91.11 ; fax: +32-11-26.91.99
E-mail address: luk.knapen@uhasselt.be

1877-0509
10.1016/j.procs.2019.04.055

1. Research Context and Problem Statement

Route choice sets are required for two purposes: (i) for parameter *estimation* in route based choice models [2] and (ii) for route *sampling* in microsimulation in order to apply travel demand to the road network. The use of a choice set for model *estimation* can be avoided by applying a recursive logit model. However, in order to *apply* route choice models in stochastic travel simulators, candidate routes need to be generated and compared. Discrete choice models are used to predict the probability for a route in the choice set from socio-demographic properties of the individual, the trip purpose and the actual network state. In this study, *path complexity* is introduced as an additional route attribute along with the classical ones (distance, travel time, number of left turns, road type, scenery, ...).

The *complexity* of a *given* (observed) path in a graph is related to the *minimum* number of least-cost paths that is required to specify that path in the network. It can be interpreted as the (minimum) number of intermediate destinations the traveler may have in mind and that are connected by least-cost subpaths. Note that *complexity* is a graph theoretical property and is not related to geometric properties of the route. *Complexity* is a path attribute which is considered to be important for route choice.

It has been observed by [7] that the distributions for the *path complexity* for sets of *observed* routes on one hand and *predicted* (by the POSDAP DSCSG method) routes on the other hand may drastically differ. Ongoing research shows that the distribution for the path complexity for routes predicted by the POSDAP BFS-LE method also heavily deviates from the observed one. DSCSG seems to heavily overestimate complexity whereas BFS-LE seems to heavily underestimate complexity.

In this study we aim to create choice sets for micro-simulation having a probability weight function (PWF) for path complexity that conforms to the observed reality. Hence, we propose to use *complexity* in the selection criteria when populating choice sets for microsimulation.

Hereafter, probability weight functions (PWF) and the corresponding cumulative distribution functions (CDF) for route complexity will be denoted by f_c and F_c respectively. Additional subscripts and superscripts are used to distinguish particular path sets.

2. State of the Art

Due to lack of space, the literature review is limited and focuses on *recursive logit models* and *route complexity* because they constitute the basis of the proposed research. *Route based* choice models that require a choice set for model estimation are not discussed.

2.1. Route Choice Modeling

Recursive logit (RL) models described by [2], [5] and [8] do not require a choice set for model *estimation*. Conceptually, they are equivalent to MNL models for route choice from an infinite number of alternatives. RL uses link-additive attributes as opposed to route attributes and conceptually applies an MNL at each junction in order to predict the next link. Hence, it can be interpreted as a *link based* choice model. It allows to compute the probability for a any given route.

In [2] the authors formulate a route choice model that does not rely on a prior choice set to estimate its parameters. Instead, a link-pair based choice model (RL , recursive logit) is introduced. The authors prove that the model is equivalent to an MNL for an infinite number of alternatives. The deterministic utility component used to determine the probability to choose the successor link a after having reached link k consists of the *link specific utility* for a and a *value function* term (for the Bellman algorithm), returning the expected value for the maximum utility to draw from the remaining part of the route to the destination (after choosing successor a of k).

[2] also introduces the *link size* (LS) attribute similar to the *path size* attribute in PSL to correct for correlation between overlapping routes. This attribute is determined by first applying the RL method using an initial estimate for the coefficients and then using the resulting link-use probability as a proxy for route overlap. The link flow proxy values need to be computed for each given OD-pair in order to evaluate the route probabilities. These proxy values indicate overlap for routes *for a particular OD-pair*. Solution of a system of linear equations is required for each OD-pair. As soon as the link-flow proxies are known, they can be added as an additional (link-additive) link attribute.

2.2. Path Complexity

The complexity of a given path in a graph is the minimum number of BPC (Basic Path Component) in the decomposition of the path where a BPC is defined as either a *least cost path* or a *non-least cost edge*. A non-least cost edge is an edge e whose vertices are connected by a bypass having a lower cost than the cost to traverse e. Note that a path may have multiple minimum decompositions. In this paper, *distance* is used as the generalized cost.

Formal definitions, the decomposition algorithm and associated mathematical proof and several frequency distributions for (mostly) car trips are presented in [4]. The majority of the trips (0.95) has complexity of at most 5.

[3] introduces the concept of Mental Representation Items (MRI) making use of a layered choice process. The first layer is used to determine a MRI choice set, such as $C_1 = \{avoidCC, aroundCC, throughCC\}$ where CC stands for *city center*. A layer on top of that provides additional details. In order to make the choice set operational, an attribute is assigned to each MRI by calculating the expected maximum utility and taking the sums of the logarithms of all utilities on the path. The MRI are determined interactively by the analyst but are related (although not equivalent) to the intermediate destinations in the decomposed paths.

[7] reports the complexity for 282k routes observed for bicyclists in The Netherlands made available in the `fietstelweek2016` dataset. The observed complexity is significantly larger than the one mentioned above: 0.95 of the routes has complexity of at most 11. This was compared to the complexity for the routes generated by POSDAP DSCSG for the observed set of OD-pairs.

3. Dataset Properties

Using observed GPS traces to analyze route choice behavior requires an elaborate process consisting of (i) trip identification, (ii) map matching and (iii) path decomposition. Although it is not easy to quantify in a formal way the effect of data pre-processing, all steps performed during data cleaning should be rigorously specified in order to allow data quality assessment w.r.t. particular properties. This is because particular pre-processing steps may affect results (e.g. map matching properties are crucial for path complexity analysis). Unfortunately, pre-processing details are not revealed for many publicly available datasets (e.g. the `fietstelweek2016` dataset used in this research which provides routes as link sequences resulting from GPS map-matching).

3.1. Basic Properties

The `fietstelweek2016` dataset contains 282k observed routes for a short survey period (1 week only). This is to be compared to [1] who use a 6-week period to extract *systematic mobility* about visited locations and chosen routes (which, according to the reported results and discussion, seems to be a minimum in order to discover *behavior*).

To achieve anonymization, trip start times are discretized into 1-hour buckets and person identifiers were removed. Furthermore, near the begin and end of the trip a part or random length is removed (*head/tail stripping*). Due to the short survey period and the anonymization (by data manipulation in both space and time) recurrent trips are missing.

Trip purpose was not recorded and it cannot be derived from repeated location visits due to *head-tail stripping* (see Section 3.4).

The associated GIS database considers each link to be bidirectional and it contains many *trivial nodes*. In GIS trivial nodes are used to specify road link geometry and to delimit segments where attribute values change.

Definition 3.1 (Trivial Node). *A* trivial node *corresponds to a vertex having exactly two neighbours.*

Definition 3.2 (String). *A* string *in graph G is a sub-graph consisting of a path in G for which all vertices except the first and last ones are trivial.*

In the `fietstelweek2016` dataset 532 851 out of 903 250 (59.0%) nodes are trivial. The presence of trivial nodes does not affect the complexity of a path but it masks recurrent use of source/destination link pairs.

3.2. OD-Pair problem

Each link in the dataset was mapped to its containing *string* and each string was represented by its *head* link. The use of strings instead of links mitigates the problem of low recurrent use of source/destination link-pairs.

Table 1: The number of unique links resp. strings used as head(tail) in a route. *ReUseFactor* is defined a the quotient *TotalNrOfRoutes/ UniqueLinks* resp. *TotalNrOfRoutes/ UniqueStrings* in each row.

UsedAs	UniqueLinks	ReUseFactor	UniqueStrings	ReUseFactor
head	182 549	1.549	165 169	1.712
tail	189 115	1.495	172 081	1.643

Table 2: OD-pair re-use based on embedding strings.

Schema	nRoutes	nUniqueOD	nReUsedOD	ReUseFactor
`complete`	282 795	276 901	5 894	1.021
`_amsterdam_c`	33 178	29 752	3 426	1.115

The number of unique head and tail links resp. strings in the `fietstelweek2016` dataset were counted: results are summarized in Table 1. The 2-nd and 3-rd columns apply to head/tail links. The 4-th and 5-th columns apply to their respective embedding strings. For head/tail string re-use, fat tail distributions are found. A use frequency of 132 was found whereas 0.91 of the strings is re-used at most 10 times and 0.97 of the strings is re-used at most 15 times. *Link* and *string* re-use were investigated for both *head* and *tail*. Results are shown in Table 1.

OD-pair re-use obviously is much smaller. Embedding strings for the head and tail links were used to count unique OD-pairs. This results in Table 2 which shows that re-use is low and more than half of the re-use occurs for the `_amsterdam_c` facet covering the city center (see Section 4). Low *ReUseFactor* values are caused (i) by the short survey period (1 week) and (ii) by anonymization (see Section 3.4). As a consequence no home-work recurrent travel can be extracted for comparison with e.g. OViN (*Onderzoek Verplaatsingen in Nederland*) by *Centraal Bureau voor de Statistiek*. Finally, even 31k routes for Amsterdam or 282k routes for The Netherlands is not much because of the huge number of possible OD-pairs and routes in the respective road network graphs.

3.3. Missing Link Problem

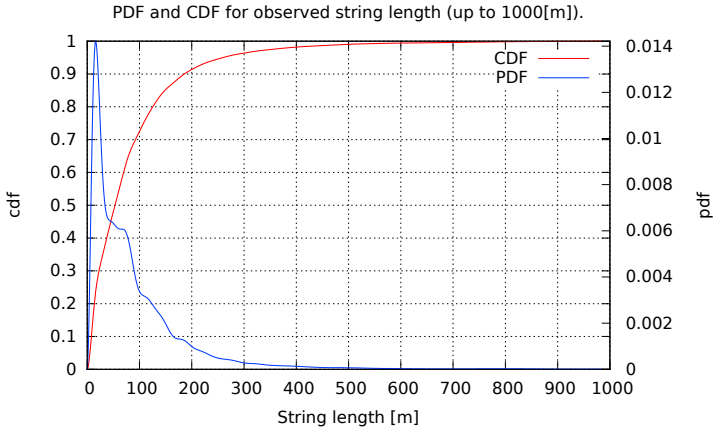
Unexpectedly large string length values were found in some (rural) regions. It was observed that links are missing in the `fietstelweek2016` network. Since no information about the map matching process nor the raw GPS data are available, the effect on the reported link sequences currently cannot be evaluated.

3.4. Effects from Anonymization by Head/Tail Stripping

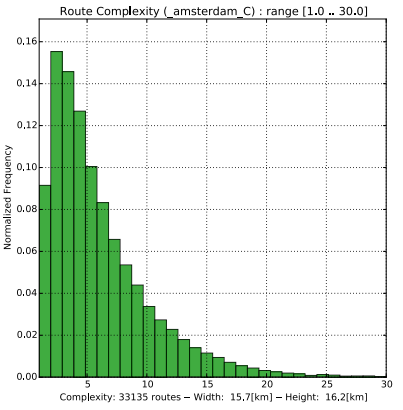
The string length distribution is required to estimate the error induced by stripping parts of random length $L \sim U(0, 400)$ [m] at the head and at the tail of each route. If the probability to strip more than one string is low (i.e. the expected stripped length is much less than the string length), then the correct string may have been identified which means that *head/tail stripping* does not (severely) affect the distribution for the path complexity.

The length distribution was determined for *strings* embedding a link that overlaps the Amsterdam bounding box because after visual inspection it can be provisionally assumed that the missing link problem (see Section 3.3) is less severe in the Amsterdam region than in other regions. Figure 1a shows the distribution (normalized CDF and PDF) for the string length measured along the road (up to 1000[m]). 29 835 of the 30 453 strings (0.9797) are shorter than 400[m] (the value used for head/tail stripping) and 27 765/30 453 (0.9117) are shorter than 200[m] (the expected value for the stripped length). Also, note that 14 522/30 453 (0.4768) of the strings have a length of at most 50[m]. Assume that head/tail stripping uses linear referencing (i.e. exactly the sampled strip length l_s is removed from the head or tail). Let $f(l_s)$ denote the probability to sample the value l_s. Since $l_s \sim U(0, L_s)$, the density function $f(l_s) = \frac{1}{L_s}$. Let $F(x)$ denote the cumulative distribution for the string length (given in Figure 1a). Then the probability to remove at most one string is given by the expected value

$$E[f(l_s) \cdot (1 - F(l_s))] = \int_{l_s=0}^{l_s=L_s} f(l_s) \cdot (1 - F(l_s))\, \partial l_s = \frac{1}{L_s} \cdot \int_{l_s=0}^{l_s=L_s} (1 - F(l_s))\, \partial l_s \tag{1}$$



(a) Distribution for the length (up to 1000[m]) of *strings* embedding links that overlap the Amsterdam bounding box.



(b) Complexity PWF for Facet _amsterdam Facet C, all trips of the day

Fig. 1: Amsterdam Central Facet: Descriptive statistics

where the factor $(1 - F(l_s))$ gives the probability that the head/tail string length is larger than the sampled value l_s. The number of edges in the path is the *walkSize*. The observed ratio walkSize/complexity (after trivial node removal) is 5 to 7 for city centers and 7 to 9 for rural areas. As a consequence, in an urban area, about 6 additional edges increase the complexity by one. Adding/deleting a single string to a given path changes the complexity by at most one. We assume that the number of strings to be added (removed) to increase (decrease) the complexity of a path obeys a Poisson distribution with $\lambda = 6$. The observed PDF ($f_1(x)$) shown in Figure 1a was used to compute the $f_k(x)$ for the sum of k variables having density $f_1(x)$. The probability to change the path complexity by one-side stripping is given by $\overline{p} = \int_{\ell=0}^{\ell=L_S} f^S(\ell) \cdot \Big(\sum_{r\in[1,\overline{r}]} \Big[p_{Pois}(r|\lambda) \cdot \int_{l=0}^{l=\ell} f_r^{SL}(l)\, \partial l \Big] \Big)\, \partial \ell$ where $f^S(\ell)$ is the PDF for the sampled strip length and $f_r^{SL}(l)$ is the PDF for the length of k concatenated strings. For $L_S = 400$ (the value used in `fietstelweek2016`) $\overline{p} = 0.277$ and hence the probability that both-side stripping does not affect complexity $(1 - \overline{p})^2 = 0.523$

4. Preliminary Results

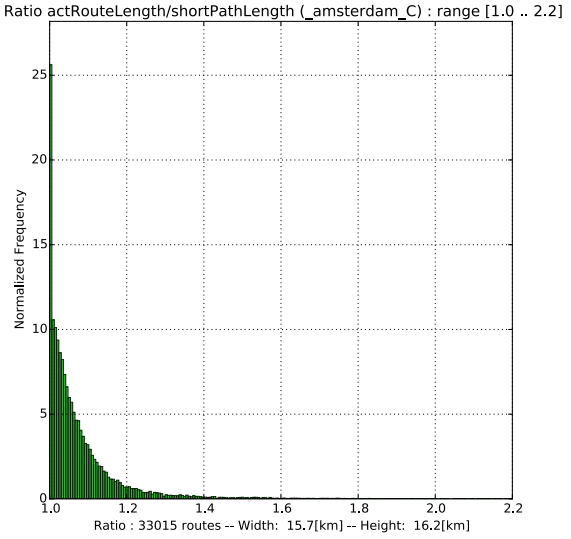
The dataset is divided into 16 regions corresponding to the main cities in The Netherlands. Each region consists of 9 facets. The central facet is the interactively defined bounding box for the city center. The other facets (labeled N,NE,E,SE,S,SW,W,NW) are defined by considering a 10[km] wide band on the four sides of the central one. There are three larger disjoint regions (*zuiden, midden, noorden*) that together cover the entire country.

4.1. Route Length

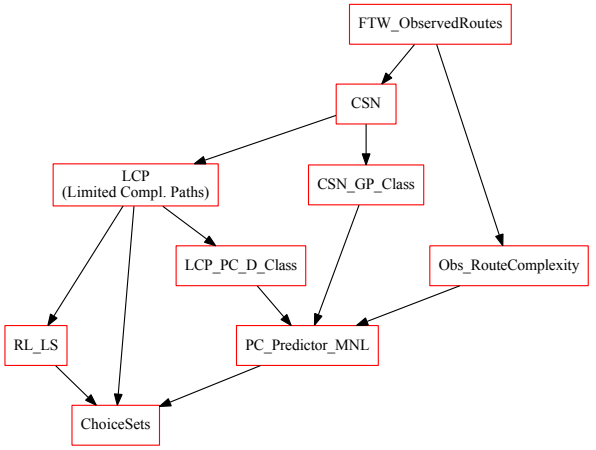
The distribution for the ratio R between the actual route length and the shortest path length seems not to vary much. R barely ever exceeds 1.4 and in many cases $R \leq 1.2$. The case for facet _amsterdam_C is shown in Figure 2a.

4.2. Mean complexity for facets: spatialeffect

The distribution for the complexity differs between facets. First we looked whether a pattern could be found by distinguishing urban and suburban facets (or regions). It turns out that such pattern does not occur although the probability distribution for the average complexity seems to depend on the location. Several potential factors have been investigated to predict the *mean complexity* found for a facet: (i) graph density in the facet (ii) number of network nodes per unit area (kind of geometric density) in the facet (iii) average length for the links in the facet (iv) ratio of



(a) Region _amsterdam Facet C, PWF for Actual-Length/ShortPathLength ratio



(b) Choice Set Generation: Information Flow Overview

Fig. 2: Distance ratio to construct CSN - Components Dependency

actual trip distance to shortest path length (linking O to D). None of these seem to lead to a usable correlation. The facets may be too large, hence insufficiently specific and take irrelevant data into account.

There is a low correlation with the walk *size* (number of links in the route). In [6] the author shows that the correlation with the walk *distance along the road* is near to zero.

4.3. Mean complexity for facets: temporal effect

The trips for each facet were assigned to a group based on the one-hour time slot (bucket) in which the first link of the trip was crossed. The groups are *morningPeak* (07h-09h), *offPeak* (10h-15h) and *eveningPeak* (16h-18h). For each facet-period combination the mean and standard deviation were collected in a table. In many cases, the mean complexity for *offPeak* is the smallest of the three values. For each facet where the mean value for *offPeak* is the smallest of the three values, it is compared to the next larger one. The t-statistic is computed using

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad s_p = \sqrt{\frac{(n_1-1)\cdot s_{X_1}^2 + (n_2-1)\cdot s_{X_2}^2}{n_1+n_2-2}} \tag{2}$$

For the aggregate regions (*zuiden*, *midden*, *noorden*) the mean value for the complexity in the *offPeak* period is *significantly* lower than for the other periods. This also holds for some but not all facets (e.g. $t = -6.764$ for _amsterdam_C). A possible explanation is that this is caused by city centers being crowded by pedestrians during the peak periods.

5. Research Project Setup

Due to the OD-pair problem it is not possible to determine distributions for the complexity of paths for particular OD-pairs in a direct way (even not after aggregating locations using 1[km] grid cells). Therefore, the next project stage is based on classifiers for *consideration sub-networks* (CSN) and *limited complexity path sets* (LCP). An MNL model

is estimated using the observed routes to predict the complexity from the CSN class and LCP class. Finally, the choice set is populated by enumerating routes in the LCP and retaining the ones having a maximal likelihood based on the RL+LS model and the MNL complexity predictor. Details are explained in the subsections. An overview is shown in Figure 2b.

5.1. CSN

For nearly all routes the ratio $\frac{actLength}{shortLength} \leq 1.4$. We define the *consideration subnetwork* for the pair $\langle O, D\rangle$ as the graph $G^C_{O,D}(V^C_{O,D}, E^C_{O,D})$ by selecting the vertex set $V^C_{O,D} \subseteq V^T$ so that

$$\forall v \in V^C_{O,D} : d(O,v) + d(v,D) \leq \alpha \cdot d_S(O,D) = \alpha \cdot shortLength \tag{3}$$

where $d_S(\cdot,\cdot)$ denotes the shortest distance between two vertices and α follows from $\text{Prob}(\frac{actLength}{shortLength} \leq \alpha) = 0.99$ and V^T is the set of road network vertices. The CSN concept is similar (but not equivalent) to the *space-time prism* concept which is based on travel time. Note that a digraph is used to model the road network. The CSN can be constructed from two *shortest path trees*: $SPT_{from}(O)$ starting at the origin and $SPT_{to}(D)$ leading to the destination.

For estimation of RL+LS parameters the CSN shall at least contain every used link in the route. It is possible that the CSN does not contain each link in the observed route. In such cases, the distance along the road for the observed route $d_{obs} > \alpha \cdot d_{short}(O,D)$. We opt to take the value for α sufficiently large and to discard the routes not meeting the threshold condition as outliers. Note that the number of edges in the CSN equals the size of the set of linear equations that needs to be solved for each OD-pair to determine the path probability in RL+LS (which is a reason to keep the value for α small).

The aim is to build a classifier that relates the observed complexity to graph properties of the CSN (average vertex degree $\frac{|E^C_{O,D}|}{|V^C_{O,D}|}$, density $\frac{|E^C_{O,D}|}{|V^C_{O,D}|\cdot(|V^C_{O,D}|-1)}$, the ratio $\frac{\mathcal{D}(G^C_{O,D})}{|V(r)|} = \frac{|V(\overline{p})|}{|V(r)|}$ where $\mathcal{D}(\cdot)$ denotes the graph diameter and $\overline{p}$ denotes the longest shortest path in $G^C_{O,D}$, etc). The classifier is represented by the box labeled *CSN_GP_Class* in Figure 2b. By using the CSN , parts of the road network graph that are irrelevant for the trip are ignored.

5.2. Limited Complexity Paths (LCP)

A recursive algorithm is developed to enumerate the set of limited complexity paths (LCP) for an OD-pair: this is the set of all paths p for which complexity $c(p) \leq \overline{c}$ linking O to D in CSN(O,D). The value $\overline{c}$ will be chosen so that $\overline{c}$ is minimal while $F^{obs}(\overline{c}) \geq p_0$ where p_0 is a given probability e.g. $p_0 = 0.99$. The algorithm will be applied to the CSN of each observed route. This delivers the PWF $f^{lcp}_c(O,D,\overline{c})$. The PWF's are used as follows:

1. The first aim is to find out whether the complexity of the observed route is improbable according to $f^{lcp}_c(O,D,\overline{c})$ (which suggests behaviour plays a significant role) or rather probable in case a random route was chosen (which means that behavioural effects are obscured). Let c^{obs} denote the complexity of the observed route linking O to D. The distributions for the following values over the complete set of observations will be analyzed:
   (a) the occurrence probability for c^{obs} given by $\text{Prob}(c^{obs}) = f^{lcp}(c;O,D,\overline{c})$
   (b) the value $|c^{obs} - \arg\max_{c\in[1,\overline{c}]} f^{lcp}(c;O,D,\overline{c})|$
2. Second, a limited set of typical complexity distributions (PWF) is to be established by clustering based on the first k moments of the $f^{lcp}(c;O,D,\overline{c})$ functions. This classifier is represented by the box labeled *LCP_PC_D_Class* in Figure 2b.

5.3. Complexity Predictor

CSN and LCP categories allow to create collections of *similar* routes for which the complexity distribution can be evaluated. This should solve the problem of insufficient observations for each OD-pair. An MNL model is estimated to establish a PWF for the complexity using the CSN and LCP classes as independent variables.

5.4. Choice Set Generation

Assume that for an OD-pair $\langle o, d \rangle$ a choice set of N_P paths is required in a micro-simulation for sampling. The aim is to use a subset of the LCP (see Section 5.2) as a replacement for BFS-LE and DSCSG to create the required choice set. Take into account that in some cases $|LCP| < N_P$: a particular CSN may contain only a few paths that in addition are of low complexity. The path probabilities estimated by means of the RL+LS method described in [2] and the PWF for the complexity determined by the complexity predictor are combined to determine the path *likelihood*. Let $N = \max(N_P, |LCP|)$; we retain the N paths from LCP having the largest likelihood values. Hence, the resulting choice set takes both link attributes and observed path complexity into account.

The procedure is based on prior evidence from observations that large deviations from the shortest path are never used. For each OD-pair we assign a negative infinite link utility to the links that are certainly not used for a route connecting O to D. In the RL+LS method we set $v_a = -\infty$ for each edge a that is not considered by the traveler in a route for the given OD-pair.
This is equivalent to use the CSN to determine the route probability in the model training stage.
As a consequence, the matrix $\mathbf{M}$ in equation (7) in [2] and equation (8) in [8] has dimension $| E_{CSN(O,D)} |$ where $E_{CSN(O,D)}$ is the set of edges CSN for the pair $\langle O, D \rangle$.

Since CSN are used, the RL+LS method requires the solution of only small sets of linear equations for each observed OD-pair (much smaller than the ones used in [8]). Furthermore, the problem is embarrassingly parallel and hence computation is expected to be feasible.

6. Conclusion

Publicly available observed route data contain few routes for each OD-pair. A method is proposed to classify OD-pairs as sufficiently similar for use in complexity prediction. A research project setup is proposed to generate choice sets for microsimulation from observed routes. A path generation technique is proposed. The likelihood for the paths that populate the choice set is derived from observations and based on link properties and on the path complexity.

Acknowledgements

The research received funding from STAR Cluster, The Netherlands.

References

[1] Bucher, D., Mangili, F., Cellina, F., Bonesana, C., Jonietz, D., Raubal, M., 2019. From location tracking to personalized eco-feedback: A framework for geographic information collection, processing and visualization to promote sustainable mobility behaviors. Travel Behaviour and Society 14, 43 – 56. URL: http://www.sciencedirect.com/science/article/pii/S2214367X18300887, doi:10.1016/j.tbs.2018.09.005.

[2] Fosgerau, M., Frejinger, E., Karlstrom, A., 2013. A link based network route choice model with unrestricted choice set. Transportation Research Part B 56, 70–80. doi:10.1016/j.trb.2013.07.012.

[3] Kazagli, E., Bierlaire, M., Fltterd, G., 2016. Revisiting the route choice problem: A modeling framework based on mental representations. Journal of Choice Modelling 19, 1 – 23. URL: http://www.sciencedirect.com/science/article/pii/S1755534515300518, doi:10.1016/j.jocm.2016.06.001.

[4] Knapen, L., Hartman, I.B.A., Schulz, D., Bellemans, T., Janssens, D., Wets, G., 2016. Determining structural route components from GPS traces. Transportation Research Part B: Methodological 90, 156 – 171. URL: http://www.sciencedirect.com/science/article/pii/S0191261516302296, doi:10.1016/j.trb.2016.04.019.

[5] Mai, T., Fosgerau, M., Frejinger, E., 2015. A nested recursive logit model for route choice analysis. Transportation Research Part B: Methodological 75, 100 – 112. URL: http://www.sciencedirect.com/science/article/pii/S0191261515000582, doi:10.1016/j.trb.2015.03.015.

[6] Wardenier, N., 2017. On Bicycle Choice Set Generation. Master's thesis. UUtrecht. Utrecht, The Netherlands. URL: https://dspace.library.uu.nl/handle/1874/355825.

[7] Wardenier, N., Knapen, L., Koch, T., Dugundji, E., 2019. Improving bicycle route choice set generation using route complexity in GPS traces, in: TRB 2019 Annual Meeting, Transportation Research Board, Washington, D.C.

[8] Zimmermann, M., Mai, T., Frejinger, E., 2017. Bike route choice modeling using GPS data without choice sets of paths. Transportation Research Part C: Emerging Technologies 75, 183 – 196. URL: http://www.sciencedirect.com/science/article/pii/S0968090X16302637, doi:10.1016/j.trc.2016.12.009.