Some main findings from passive data
Displacement

as unit of analysis
• Both power laws and exponential laws have been found for displacement patterns

• Goal-oriented and random movements result in the same distribution law and it is the underlying street network that matters!

Gonzalez et al., 2008; Jiang et al., 2009; Kang et al., 2012
• Theoretical predictability limit: 0.93

• Practical predictability: above 0.8

Song et al., 2010; Lu et al., 2013
model development
- behavior choices
Discrete choice models

• Properties of a discrete choice

  • answering the question of which one

  • a finite and feasible choice set
Random utility theory

\[ Pr[U_i] = Pr[U_i \geq U_j, \forall i, j \in C] \]
\[ = Pr[V_i + \epsilon_i \geq V_j + \epsilon_j] \]
\[ = Pr[\epsilon_j - \epsilon_i \leq V_i - V_j] \]

- \( V \) is the systematic utility,
- \( \epsilon \) is the random error term
Logit model

\[ Pr[i] = \frac{e^{V_i}}{\sum_j e^{V_j}} \]
Example choices

- bus
- walking
- bicycle
- auto
from single to multiple choices

- bus & store 1
- bus & store 2
- auto and store 1
- auto and store 2
From single to multiple trips

- Go home and no shopping
- Go shopping on the way home
- Go home first then go shopping
Limitations

\[ U_i = V_i + \epsilon_i \]

- random utility is independent and identically distributed (IID) with a type 1 extreme value distribution
- no random taste variation
- proportional substitution

\[ \frac{P^1_{ni}}{P^1_{nk}} = \frac{P^0_{ni}}{P^0_{nk}} \]

Bhat, 1997, 2000; Train 2009
Proportional substitution

<table>
<thead>
<tr>
<th></th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>large gas cars</td>
<td>0.66</td>
<td>0.6</td>
</tr>
<tr>
<td>small gas cars</td>
<td>0.33</td>
<td>0.3</td>
</tr>
<tr>
<td>EV cars</td>
<td>0.01</td>
<td>0.10</td>
</tr>
</tbody>
</table>
Heteroscedastic models

\[ U_i = V_i + \epsilon_i, \text{ where } \epsilon_i \sim G(0, \theta_i) \]

\[ Pr(i) = Pr(U_i > U_j), \forall j \neq i, j \in C \]

\[ = Pr(\epsilon_j \leq V_i - V_j + \epsilon_i), \forall j \neq i, j \in C \]

\[ = \int_{\epsilon_i=-\infty}^{\epsilon_i=+\infty} \sum_{j \in C, j \neq i} \Lambda\left[\frac{V_i - V_j + \epsilon_i}{\theta_i}\right] \frac{1}{\theta_i} \lambda\left(\frac{\epsilon_i}{\theta_i}\right) d\epsilon_i \]

where \( \lambda(t) = e^{-t}e^{-e^{-t}} \) and \( \Lambda(t) = e^{-e^{-t}} \)

Bhat, 2000
GEV models

Let \( Y_i = \exp(V_i); \ G = G(Y_1, Y_2, \ldots Y_J); G_i = \partial G / \partial Y_i \)

if \( G \) satisfies certain conditions, then \( P_i = \frac{Y_i G_i}{G} \)

is the choice probability for a discrete choice model that is consistent with utility maximization

1. \( G \geq 0 \) for all possible values of \( Y_j \forall j \).
2. \( G \) is homogeneous of degree one.
3. \( G \rightarrow \infty \) as \( Y_j \rightarrow \infty \) for any \( j \)
4. The cross partial derivative of \( G \) change signs in a certain way. That is, \( G_i \geq 0 \) for all \( i \), \( G_{ij} = \partial G_i / \partial Y_j \leq 0 \) for all \( j \neq i \), \( G_{ijk} = \partial G_{ij} / \partial Y_k \geq 0 \) for any distinct \( i,j,k \), so on for higher cross-partials

Train, 2009
An example GEV model for dest. choices

\[
G(y_1, y_2, \ldots, y_c) = \sum_{i=1}^{c-1} \sum_{j=i+1}^c \left[ (k_{i,ij} y_i)^{1/(1-\rho_{ij})} + (k_{j,ij} y_j)^{1/(1-\rho_{ij})} \right]^{1-\rho_{ij}}
\]

Where,

- \( y_k \geq 0 \) for all \( i \),
- \( k_{i,ij} \) is weighting parameter for alternative \( i \) in nest \( ij \).
\( \rho_{ij} \) and \( k_{i,ij} \)

- \( \rho_{ij} \) is the similarity index, measuring the degree of correlation between alternatives \( i \) and \( j \) in nest \( ij \)
- \( k_{i,ij} \) is the weighting parameter, measuring the extent alternative \( i \) belongs to nest \( ij \)
Assumed correlation structure

A

nest AB

B

nest BC

nest CD

C

D

nest AC

nest AD

nest BD
The probability function for household $n$ is:

$$P_n(j) = \frac{\sum_{i=1, i \neq a}^{c} (k_{j,ij} V_{n,j})^{1/1-\rho_{ij}} \left[ (k_{j,ij} V_{n,j})^{1/1-\rho_{ij}} + (k_{i,ij} V_{n,i})^{1/1-\rho_{ij}} \right]^{\rho_{ij}} - 1}{\sum_{i=1}^{c-1} \sum_{j=i+1}^{c} \left[ (k_{j,ij} V_{n,j})^{1/\rho_{ij}} + (k_{i,ij} V_{n,i})^{1/\rho_{ij}} \right]^{\rho_{ij}}}$$

Where,

$$\rho_{ij} = \exp(-\lambda \times \text{dist}_{ij}^{0.2})$$, if dist$_{ij} \leq 3.5$ miles

$$\rho_{ij} = 0$$, if dist$_{ij} > 3.5$ miles

$$k_{i,ij} = \frac{l(\rho_{ij})}{\sum_{m=1, m \neq i}^{c} l(\rho_{im})}$$
mixed logit model

\[ P_{ni} = \int \left( \frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}} \right) f(\beta) d\beta \]

\[ \frac{\Delta P_{ni}/P_{ni}}{\Delta x^m_{nj}/x^m_{nj}} = -x^m_{nj} \int \beta^m L_{nj}(\beta) \left[ \frac{L_{ni}(\beta)}{P_{ni}} \right] f(\beta) d\beta \]

\[ L_{ni}(\beta) = \frac{e^{V_{ni}(\beta)}}{\sum_{j=1}^J e^{V_{nj}(\beta)}} \]
mixed logit model

error component model

\[ U_{ni} = \gamma' y_{ni} + \zeta_{ni} = \gamma' y_{ni} + \mu' z_{ni} + \epsilon_{ni} \]

random coefficient model

\[ U_{ni} = \beta'_n x_{ni} + \epsilon_{ni} \]
Hazard model

\[ F(t) = Pr[T < t] \]

\[ f(t) = \frac{dF(t)}{d(t)} = \lim_{dt \to 0} \frac{Pr(t \leq T < t + dt)}{dt} \]

\[ S(t) = Pr[T \geq t] = 1 - F(t) \]

\[ h(t) = \frac{f(t)}{S(t)} = \lim_{dt \to 0} \frac{Pr(t \leq T < t + dt | T \geq t)}{dt} \]
Activity-based models

- Types of underlying models
  - Feasibility based
  - Feasibility and behavior based
  - behavior based
- Building an activity/travel pattern
  - simultaneous
  - sequential
BEHAVIORAL FACTORS
Motivation

• Understanding spatial behaviors
  - factors
  - underlying behavioral mechanisms
• Policy evaluation
• *Finding behavioral triggers*
The usual factors

*Most often*

- socio-demographics
- the built environment
- trip/alternative related factors

*Less often*

- attitudes
- feelings
Accessibility

• Conventional measures (relative or integral)
  - uses a single reference location
  - enumerates all possible destinations
  - uses an impedance function to capture the effect of distance decay

• Space-time accessibility measures
Travel behaviors

self-selection

Built environment
Travel behaviors

Built environment

Attitudes
Travel behaviors

Attitudes

Wang and Chen, 2012
Attitudinal statements

For each question, response on a scale: strongly agree, agree, neutral, disagree, strongly disagree, no idea.

- The price of oil should be increased to reduce congestion and pollution
- More public transportation is necessary, even if it means additional taxes
- Ecology is a threat to minorities and small companies.
- People and employment are more important than the environment.
- I feel concerned by the global warming.
- Decisions must be taken to reduce the greenhouse gas emission.
Explanatory Variables

Utility

Latent variables

\[ X^*_k = \sum j \beta_j x_j \]

\[ I_i = \lambda_i + \sum k L_{ik} X^*_k \]

Choice

Indicators
Mathematicians and travel behavior researchers in partnership to answer *policy-relevant questions*
Validation, validation, validation
Movement patterns
We need models that describe our activity/travel patterns evolve over time and a framework that tie the models together!
A few other suggestions

• Plot probability distributions under different circumstances (e.g., context, population…)

• Scramble with random movements too

• Get temporal: look at changes over time

• Downscale: look at metrics at the individual level
Model development
Choice set

Can we leverage the mobile phone dataset and knowledge on space-time prisms to generate more realistic choice sets for individuals?
Destination choices

• a choice problem that does not correspond to the use of standard discrete choice models

  • complex substitution patterns

  • choice set formation
Spatial set and mental map
Comparing models

- To uncover source of disagreement
- To identify critical data that are needed to resolving these differences and solidify our understanding of a phenomenon
The need for simple, analytical models

- Need to think more about the fundamental mechanisms that govern the observed travel patterns
- Connect to data-driven computational approaches to investigate the impact of the various complex features of real systems on the basic properties
Behavior factors
Understanding behaviors

- Identifying additional factors (e.g., social network factors, though it is not a new discovery)
- Uncovering decision rules applied
- Understanding experience accumulated, feelings generated and attitudes formed
Policy evaluation
Some policy-relevant questions

• If density is increased, will people use bus or walk more often?

• What incentives shall be provided for people to reduce driving?

• How are emerging transportation services (e.g., Uber) changing people’s travel patterns?

• If we build a light rail, will people use it?

• How to move toward reducing auto ownership and moving to more compact locations?

• .....

114
Cannot be answered with cross-sectional data!
What about those passive mobile phone data?

- If we can correctly guess:
  - essential elements about one’s travel pattern
  - other attributes such as socio-demographics
  - behavioral triggers (what, when and where)
- data length is sufficiently long
We still need:

- Conceptual frameworks that describe behaviors
- Carefully designed experiments
- Hypotheses
- ...
- ...
- ...
Prediction, Policy and Behavior Triggers

• Let us talk about meaningful locations!

• Move beyond just location prediction or even trip-related prediction

• Learn individual preferences over time

• Be innovative on the alternatives offered — Think Marketing

• Find the low-hanging fruit

• Together, contribute to both individual and societal welfare!
Why not cross-sectional data?

• no notion of (ir)regularity of individual behaviors

• infrequent trips (e.g., long-distance trips) are not captured

• not possible to distinguish between intra- and inter-personal variability

• if behavior changes are inferred, we assume that they are:

  • instantaneous, symmetric, and stationary
What can be learned from a panel survey?

- many subjects, multiple time points (explicit time dimension)

- best suited for policy analysis (e.g., how a change in the system will affect travel behaviors?)

- causal analysis (since causes and effects are observed)

- potential to improve model forecasts (due to temporal and interpersonal aspects)

- can explicitly measure and model asymmetric effects, response lags/leads, and state dependence etc.

- explicitly recognize the “multiple equilibria” phenomenon (Mahmassani, 1990)