

Route choice modeling with network-free data

Bierlaire, M., & Frejinger, E. (2008).

Transportation Research Part C: Emerging Technologies, 16(2), 187-198.

理論談話会2019 #1

4月19日

M1 石井 健太

- DDRとは

Domain of Data Relevance = データの有効範囲

→観測したデータが取得されうる範囲

- 本研究の趣旨

マップマッチングのような観測経路を仮定する操作を行わず、
ネットワークフリーデータ

(= GPSなど仮定したネットワークに紐づいていないデータ)

をそのまま用いることのできるモデリングの枠組みを提唱する


- 経路選択モデルにおけるデータに関する問題

- メールや電話を用いた調査

- 電話とGISの組み合わせ

-  × 詳細な経路すべてを記述するのは難しい

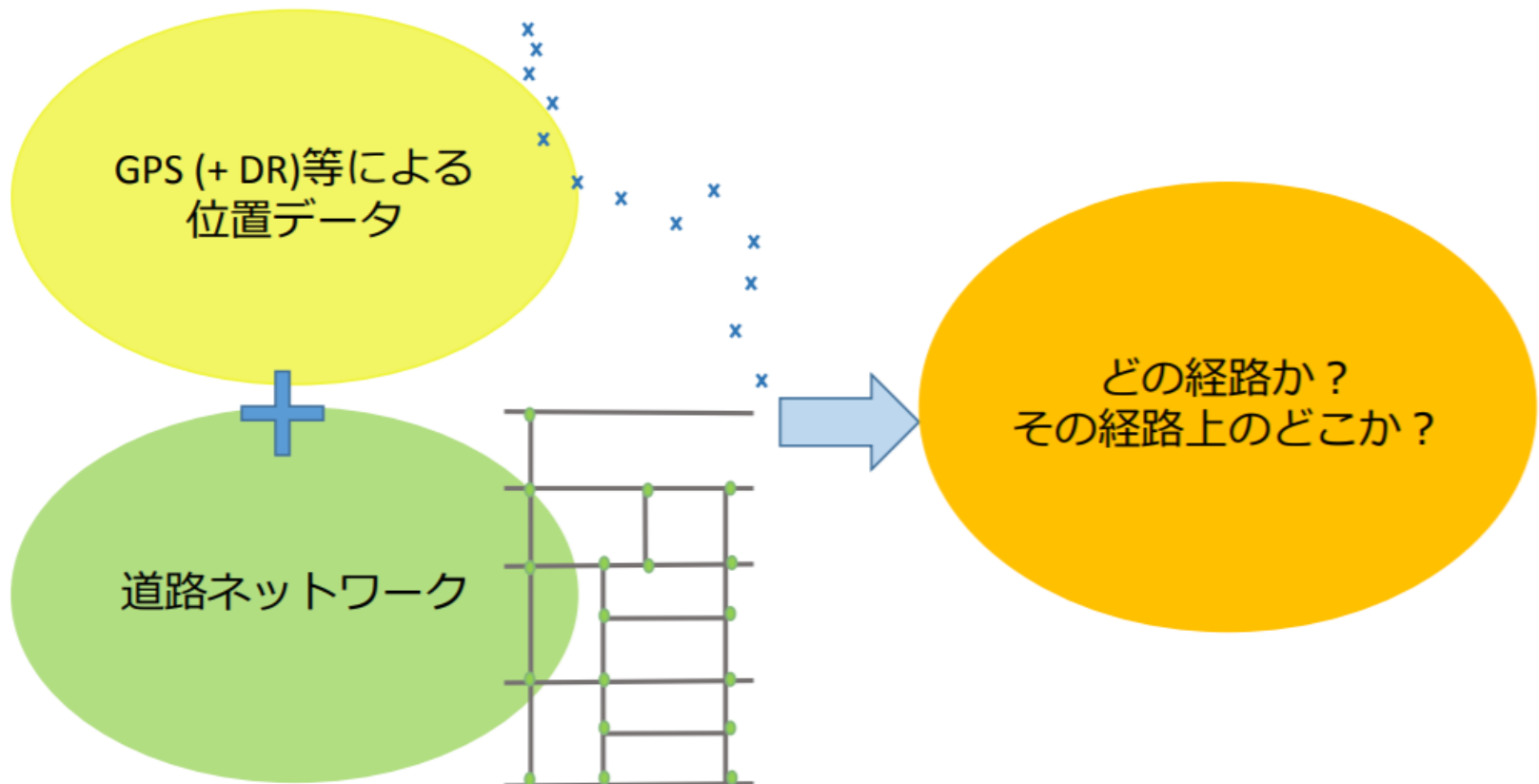
- GPSを用いた受動監視

-  ○ 複数日のトリップデータを自動収集可能
電子形式において直接利用可能

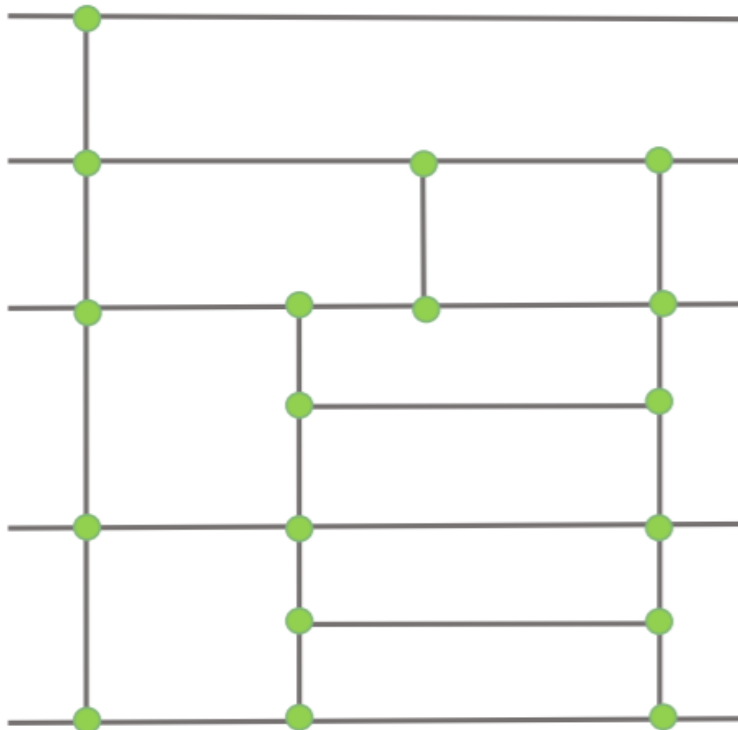
- × 衛星の時計の誤差, 大気状態, 周囲の環境などに精度依存
マップマッチングなどによるデータ加工が必要

マップマッチング (補足)

GPS やデッドレコグニング技術で得られた位置データを用いて、ある時点で移動者がどの経路を利用しているのか、あるいはその経路上のどこにいるのかを特定する技術



道路ネットワーク



凡例



リンク

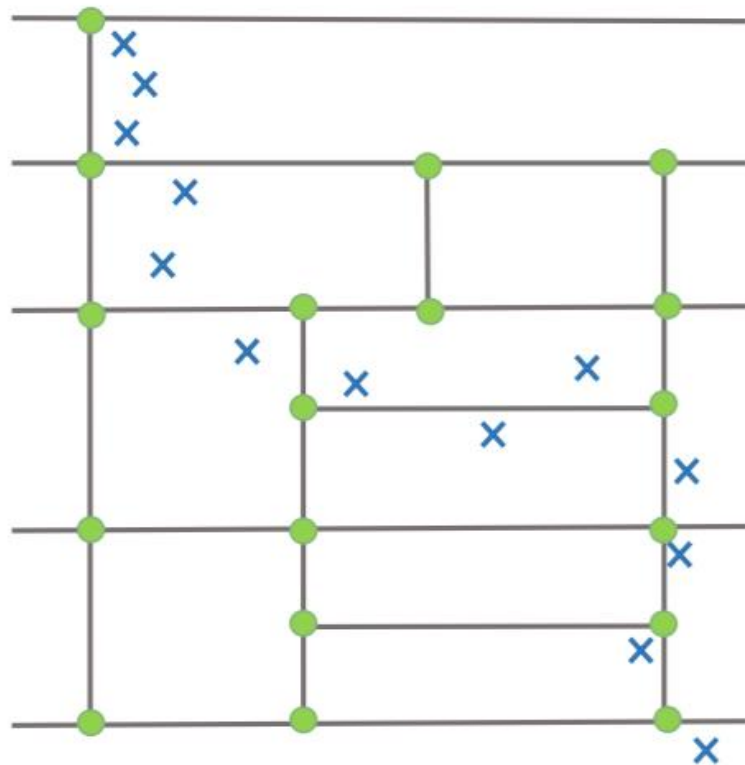


ノード

基本的なアルゴリズム

- STEP1 道路ネットワークを準備
- STEP2 道路ネットワーク上に測位位置データをプロット
- STEP3 位置データとネットワーク上のリンクやノードとの関係を定量化
→通過リンクの特定
- STEP4 リンクごとのパフォーマンス指標 (所要時間や走行速度) の算定

道路ネットワーク

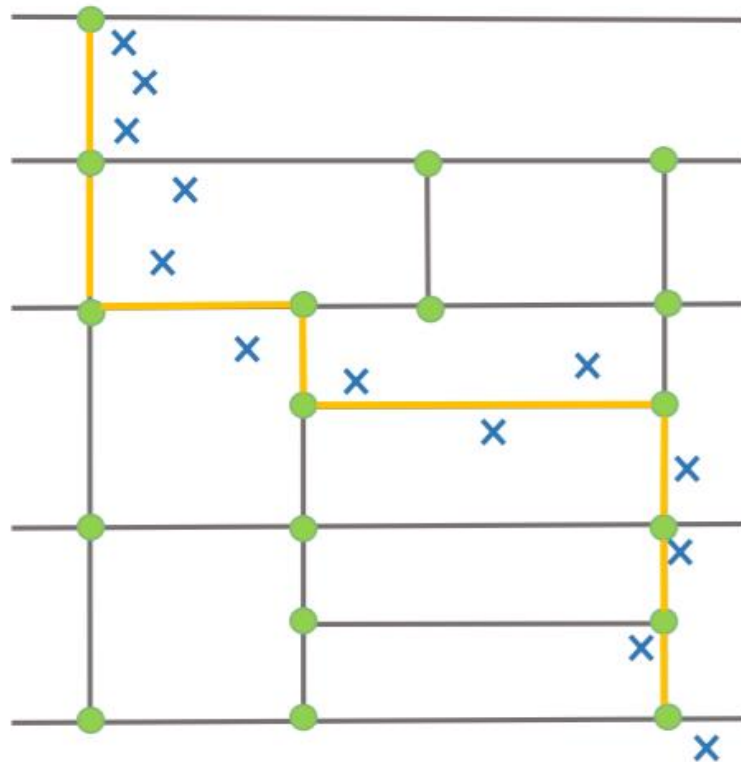


凡例
× 測位点

基本的なアルゴリズム

- STEP1 道路ネットワークを準備
- STEP2 道路ネットワーク上に測位位置データをプロット
- STEP3 位置データとネットワーク上のリンクやノードとの関係を定量化
→通過リンクの特定
- STEP4 リンクごとのパフォーマンス指標 (所要時間や走行速度) の算定

道路ネットワーク



凡例

— 通過リンク

基本的なアルゴリズム

- STEP1 道路ネットワークを準備
- STEP2 道路ネットワーク上に測位位置データをプロット
- STEP3 位置データとネットワーク上のリンクやノードとの関係を定量化
→通過リンクの特定
- STEP4 リンクごとのパフォーマンス指標 (所要時間や走行速度) の算定

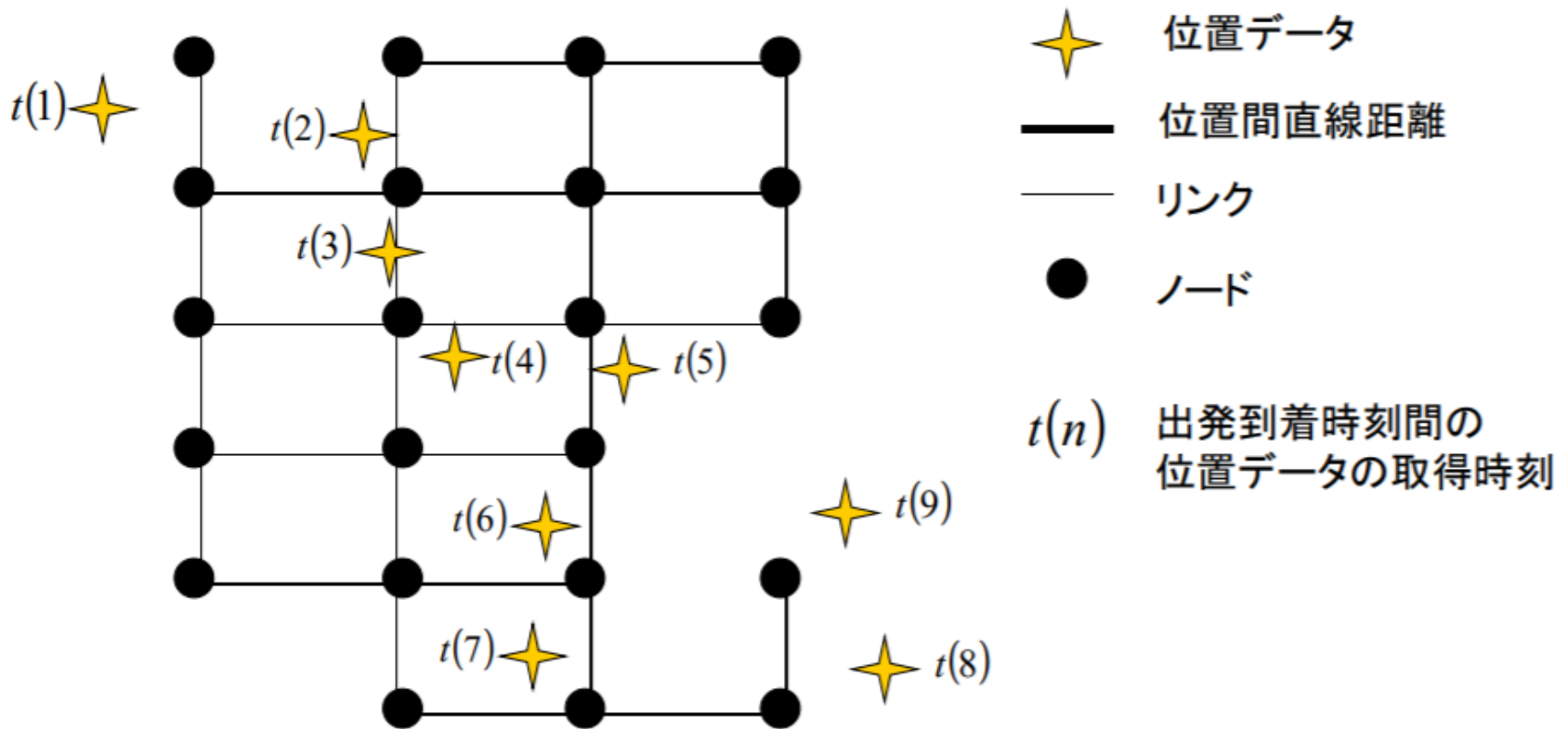
定量化方法や測位補正の方法などによってさまざま手法がある



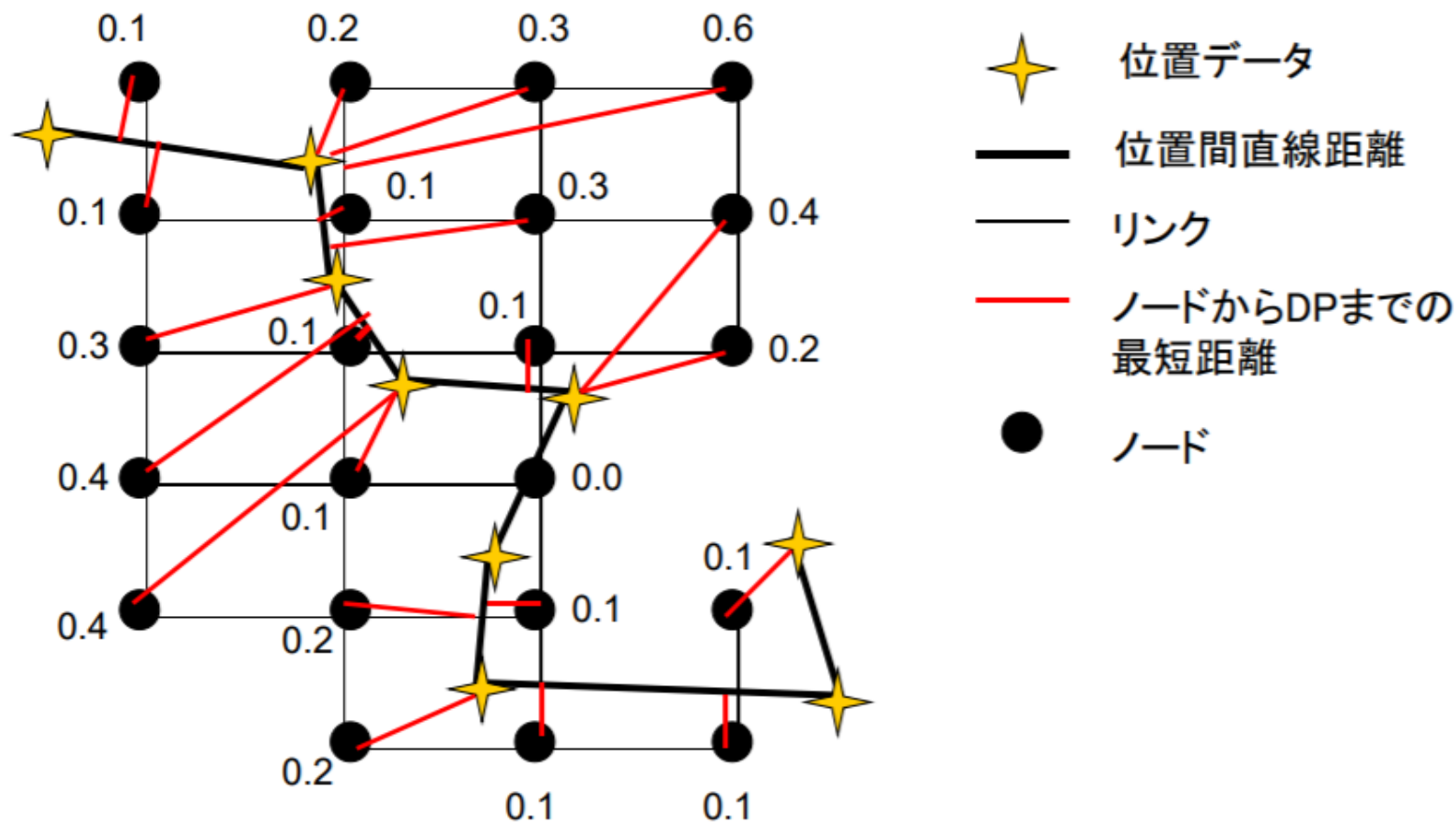
基本的なアルゴリズム

- STEP1 道路ネットワークを準備
- STEP2 道路ネットワーク上に測位位置データをプロット
- STEP3 位置データとネットワーク上のリンクやノードとの関係を定量化
→通過リンクの特定
- STEP4 リンクごとのパフォーマンス指標 (所要時間や走行速度) の算定

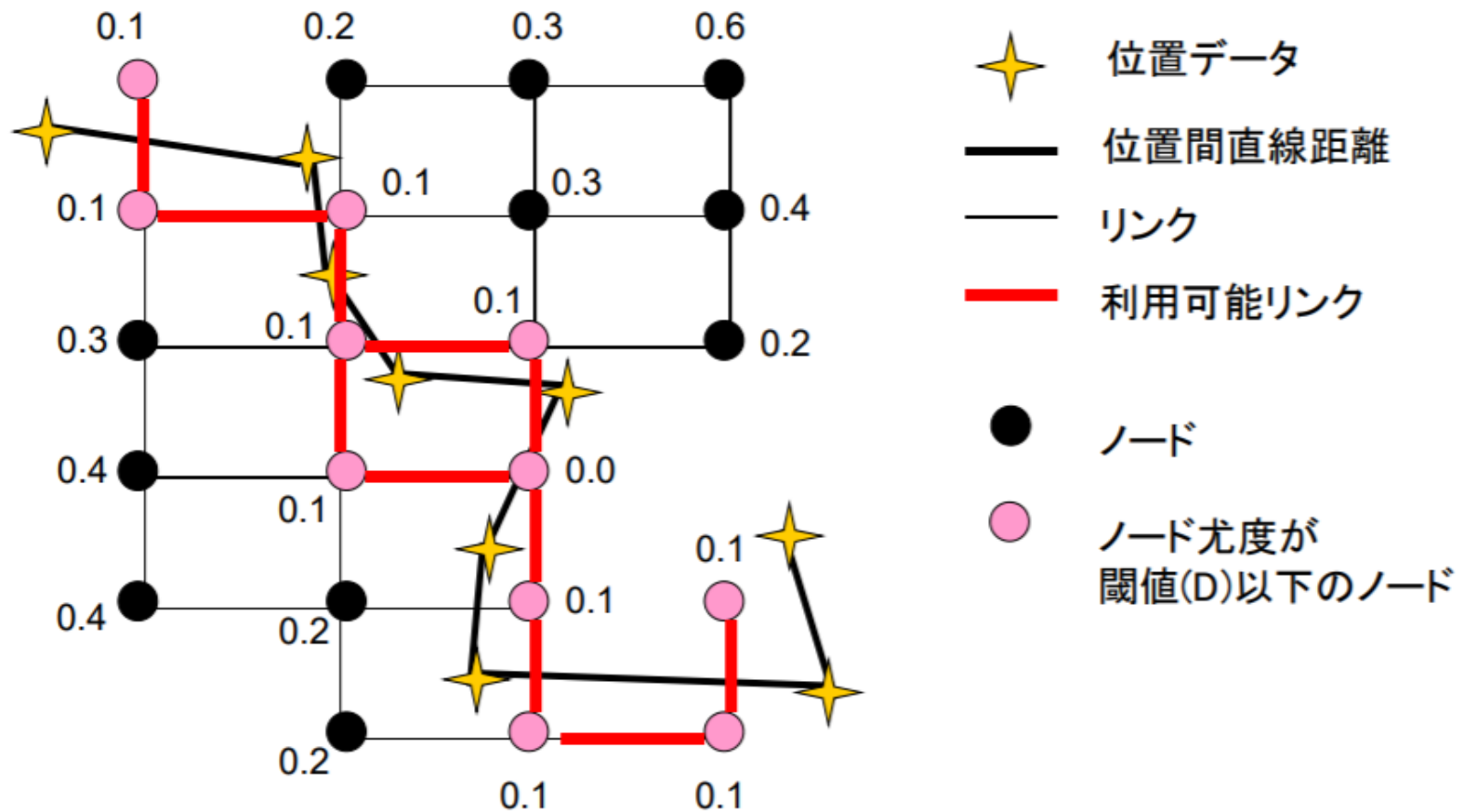
トリップデータの抽出



ノード尤度の算出









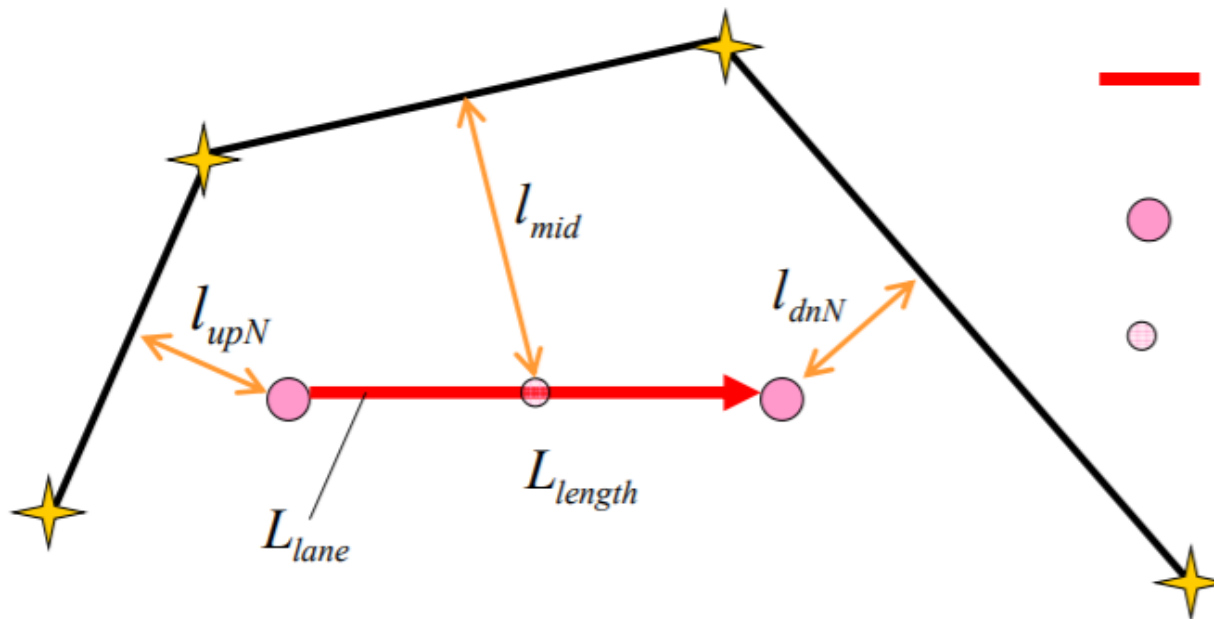
利用可能リンクの抽出

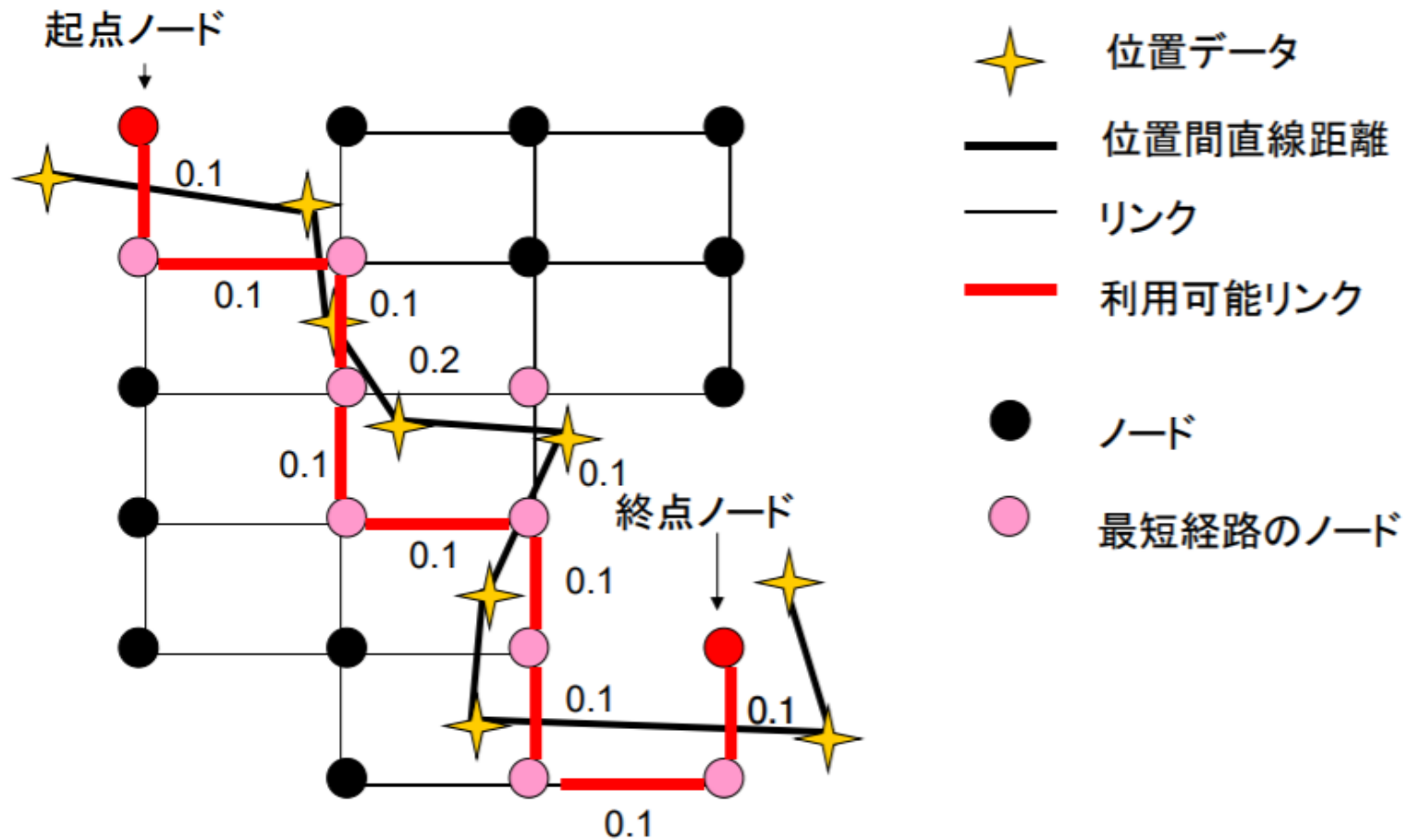


リンク尤度の算出

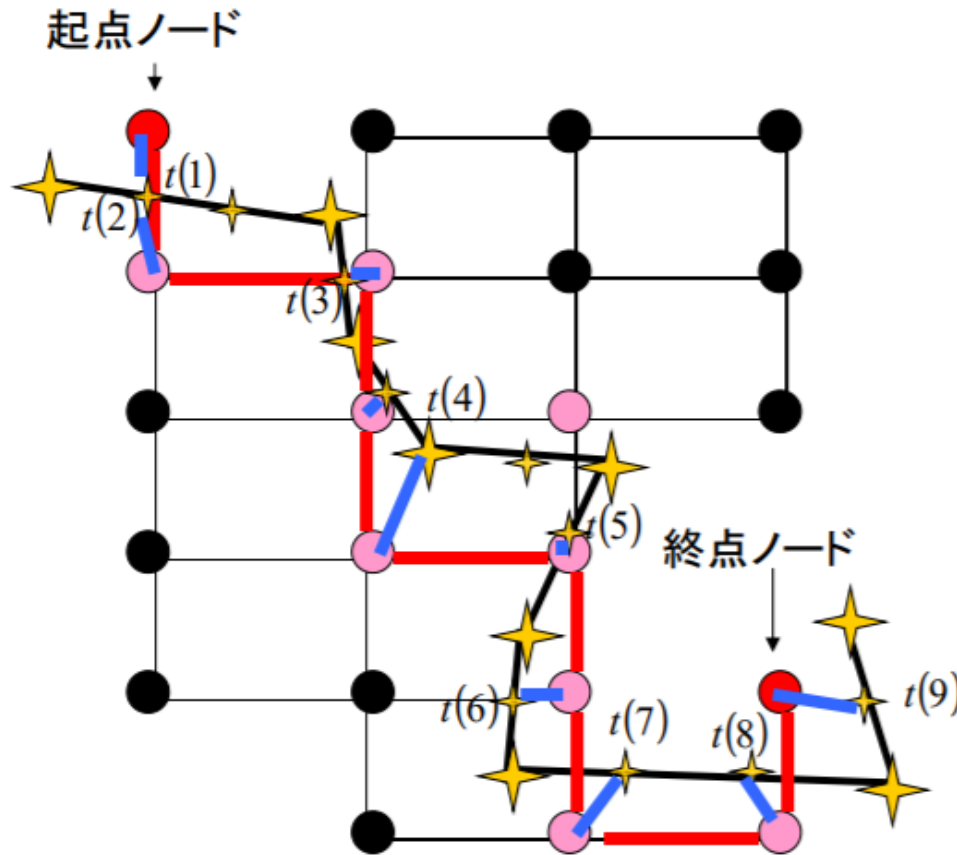
$$LL = (l_{upN} + l_{dnN} + a \times l_{mid}) \times \frac{L_{length}}{b \times L_{lane}}$$

-  位置データ
-  位置間直線距離
-  リンク
-  利用可能リンク
-  ノード
-  リンクの中点





リンク旅行時間の算出



★ 1秒毎の補完位置データ

★ 位置データ

— 位置間直線距離

— リンク

— 利用可能リンク

● ノード

● 最短経路のノード

$t(n)$ 出発到着時刻間の
位置データの取得時刻

- 推定に利用できるデータの特性のまとめ

- ネットワークに対応した選択経路データはほぼ得ることができない
- データには測定誤差が存在するが、既存の経路とのマッチング手法は経路を一意に決定する
- データの加工自体が手間

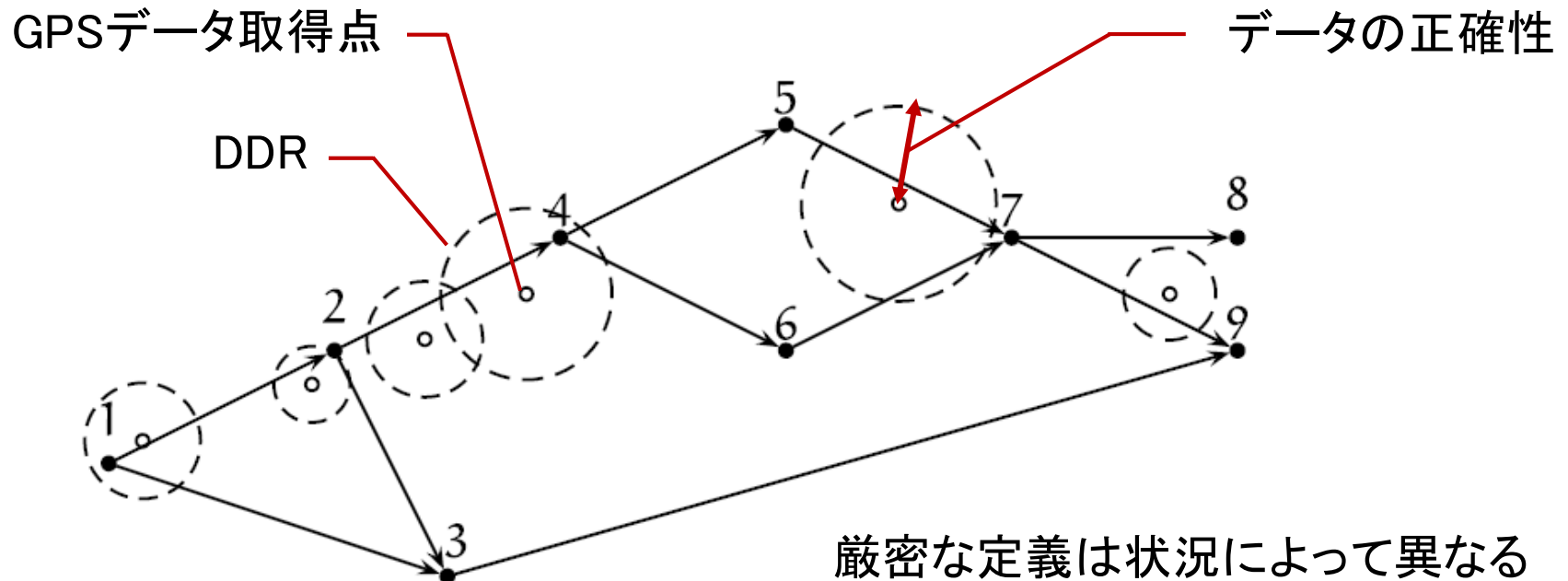


データ操作せずネットワークベースのモデルに
ネットワークフリーデータを調和させるモデルの枠組みを構築する

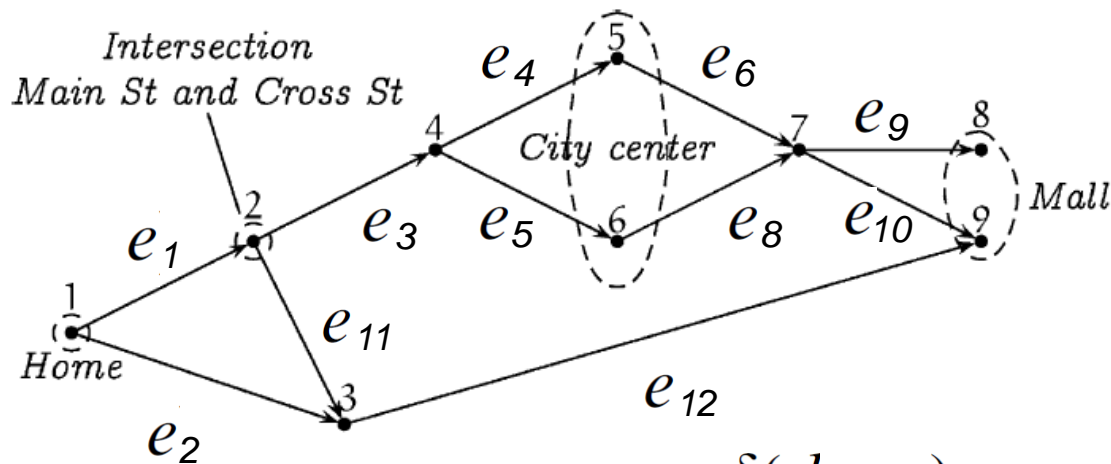
- DDR (Domain of Data Relevance)

定義：データが関連性を持っている物理的な空間

➤ GPSデータの例



- 指標関数 $\delta(d, e)$ where $d : \text{DDR}$
 $e : \text{ネットワーク要素}$
 - 各データに関連するネットワークをリスト化する
 - ネットワーク要素がデータのDDR と関連があれば 1
そうでなければ 0
 - 報告データの例



$$\delta(d_1, e_1) = 1, \quad \delta(d_2, e_2) = 0$$

- 目的

経路選択モデル $P(p|\mathcal{C}_n(s);\beta)$ の未知パラメータ β を推定する

$\mathcal{C}_n(s)$: ODペア s を結ぶ経路集合 (選択肢集合)

p : $\mathcal{C}_n(s)$ 中の1つの経路

関連あるODペアの集合 \mathcal{S}_i (\mathcal{S} : ネットワーク内すべてのODペアの集合)

$$\mathcal{S}_i = \{s \in \mathcal{S} | \delta(d_1, s_o) \delta(d_k, s_d) = 1\}$$



空集合の場合, DDRの定義を変更しなければならない

- \mathcal{S}_i が与えられた上で個人 n の観測 i を再現する確率

$$P_n(i|\mathcal{S}_i) = \sum_{s \in \mathcal{S}_i} P_n(s|\mathcal{S}_i) \sum_{p \in \mathcal{C}_s} P_n(i|p) P_n(p|\mathcal{C}_s; \beta)$$

$P_n(s|\mathcal{S}_i)$: ODペアの集合 \mathcal{S}_i が与えられたとき, 実際のODペアが s である確率

$P_n(i|p)$: 観測 i において実際の経路が p である確率を与える計測方程式

$P_n(p|\mathcal{C}_s; \beta)$: 経路選択モデル



$P_n(i|\mathcal{S}_i)$ を最大化する

計測方程式は以下の条件を満たすとき0より大きい値をとる

- 経路の中に個々の DDR に関連するリンクが少なくとも1つはある, つまり全ての $m = 1, \dots, k$ に対して, $\delta(d_m, l_q) = 1$ のような $q \quad 1 \leq q \leq P$ が存在する
- 一連の報告された場所が経路のリンクの順序と一致している, つまり全ての $m_1 \leq m_2$ に対して $\delta(d_m, l_{q_1}) = 1 \cap \delta(d_m, l_{q_2}) = 1$ のとき $q_1 \leq q_2$ である.

➤ 報告トリップの場合

経路が報告された場所をすべて通っているかいないかで0, 1

➤ GPSデータの場合

例) i と p の距離の関数 $D(d, p)$ を用いて定義する

$$D(d, p) = \min_{l \in A_{pd}} \Delta(d, l) \quad \text{where } \Delta(d, l) : \text{データ}d\text{からリンク}l\text{の最近地点へのユーグリッド距離}$$

$$A_{pd} = \{l \in l_1, \dots, l_P \mid \delta(d, l) = 1\}$$

$P_n(i|p)$ が $D(d, p)$ によってどう定義されるかはケース依存

$|\mathcal{S}_i| > 1$ となる観測 i が1つ以上存在する場合

➡ $P_n(s|\mathcal{S}_i)$ に対応するモデルを定義する必要がある

例) すべてのODペアに等しく確率を割り当てる (最も簡単な例)

$$P_n(s|\mathcal{S}_i) = \frac{1}{|\mathcal{S}_i|} \quad \forall s \in \mathcal{S}_i$$

追加情報が利用可能な場合,

自宅や勤務地が含まれているODペアに高い確率を割り当てる
といった操作も考えられる

厳密な定義は**問題依存**

どのケースにおいても以下の2点に注意が必要

1. データ d のDDRは空であってははいけない

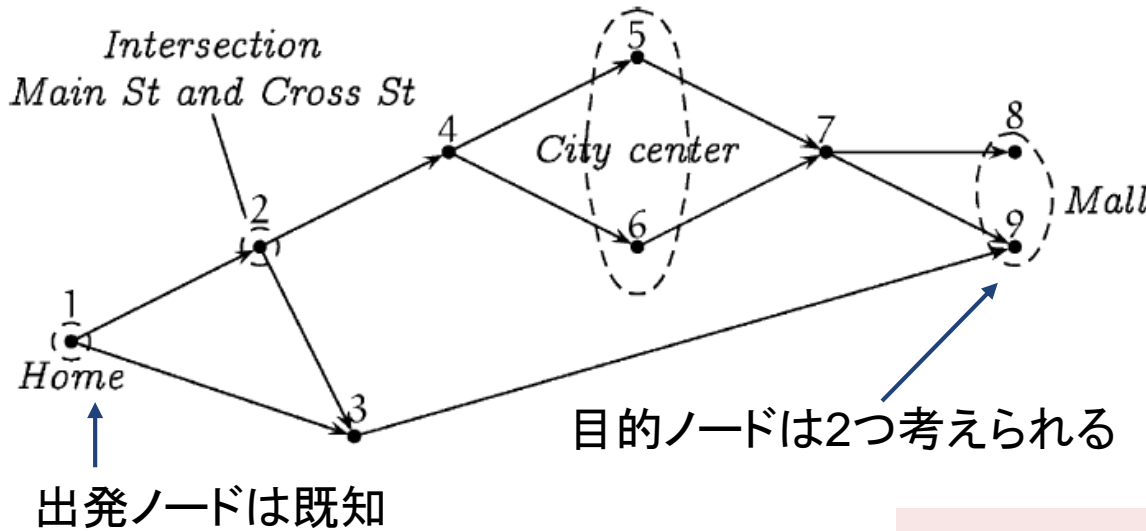
$\delta(d, e) = 0 \forall e$ のとき, DDRの定義が適切ではない

 再定義する必要がある (多くの場合サイズを大きくする)

2. DDRが関係するネットワーク要素はできるだけ少なくする

$P_n(i|\mathcal{S}_i)$ の定義式の和の部分が指数関数的に増加してしまい, モデルを特定化できない

例 (1) 報告トリップ



観測 i に関するODペアの集合

$$\mathcal{S}_i = \{(1, 8), (1, 9)\}$$

(s_1 と s_2 と呼ぶ)

追加情報が存在しないため

$$P(s_1 | \mathcal{S}_i) = P(s_2 | \mathcal{S}_i) = \frac{1}{2}$$

$$\begin{aligned} P_n(i | \mathcal{S}_i) &= \sum_{s \in \mathcal{S}_i} P_n(s | \mathcal{S}_i) \sum_{p \in \mathcal{C}_s} P_n(i | p) P_n(p | \mathcal{C}_s; \beta) \\ &= \frac{1}{2} [P(p_1 | \mathcal{C}(s_1); \beta) + P(p_2 | \mathcal{C}(s_1); \beta)] \\ &\quad + \frac{1}{2} [P(p_3 | \mathcal{C}(s_2); \beta) + P(p_4 | \mathcal{C}(s_2); \beta)] \end{aligned}$$

$$\mathcal{C}(s_1) = \{(1, 2, 4, 5, 7, 8), (1, 2, 4, 6, 7, 8)\}$$

p_1 と p_2 と記す

$$\mathcal{C}(s_2) = \{(1, 2, 4, 5, 7, 9), (1, 2, 4, 6, 7, 9), (1, 2, 3, 9), (1, 3, 9)\}$$

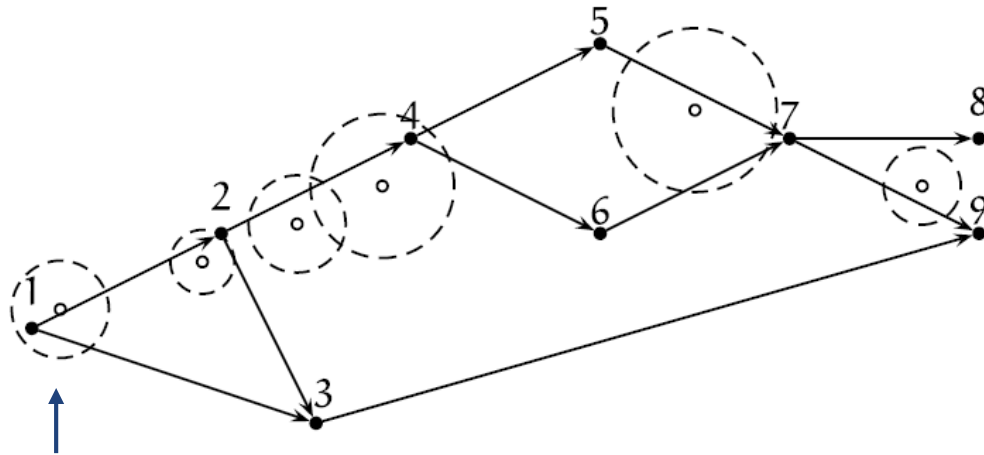
それぞれ p_3, \dots, p_6 と記す

観測と経路が対応している

$$P(i | p_1) = P(i | p_2) = 1 \quad P(i | p_3) = P(i | p_4) = 1$$

観測と経路が対応していない

$$P(i | p_5) = P(i | p_6) = 0$$



出発ノードに相当するノードが1つ

$$P_n(i|\mathcal{S}_i) = \sum_{s \in \mathcal{S}_i} P_n(s|\mathcal{S}_i) \sum_{p \in \mathcal{C}_s} P_n(i|p) P_n(p|\mathcal{C}_s; \beta)$$

$$\mathcal{C}(s) = \{(1, 2, 4, 5, 7, 9), (1, 2, 4, 6, 7, 9), (1, 2, 3, 9), (1, 3, 9)\}$$

$$P(i|p_3) = P(i|p_4) = 0 \quad p_1, \dots, p_4 \text{ と記す}$$

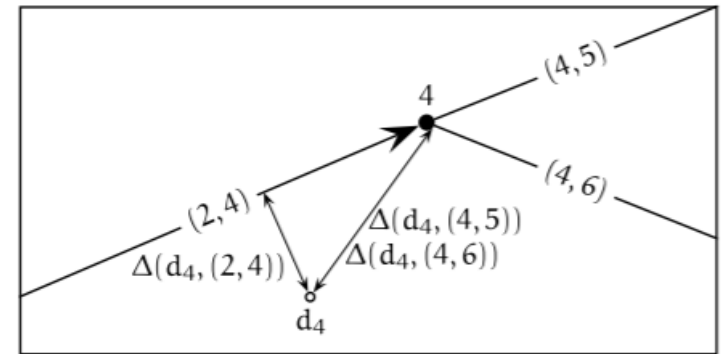
$P(i|p_1)$ と $P(i|p_2)$ はデータ取得場所と経路との距離の関数で表される

$$\Delta(d_4, (4, 5)) = \Delta(d_4, (4, 6)) > \Delta(d_4, (2, 4)) \quad d_4 \text{ と } p_1 \text{ の距離は } \Delta(d_4, (2, 4))$$

最後のデータのDDRに関連するノードなし

このDDRが交わるリンクのシンクノードを考える

➡ $\mathcal{S}_i = \{(1, 9)\}$



- **適用対象: スイスで行われた電話調査**
長距離トリップを対象とした調査
出発地・目的地と共に, 最大3つ途中で通過した場所を記述
- **DDRは報告された場所の郵便番号のエリア範囲と定義**
(エリアの大きさは場所によって異なる, 平均3.8km²)
- **用いるネットワークは39411の一方向リンクと
14841のノードを含む**
(特に都心部において細街路が簡略化されている)
- **選択肢集合**
ランダムに2つのODペアを選び, 各ODペアに等しく確率を割り当てている
780の観測が利用可能

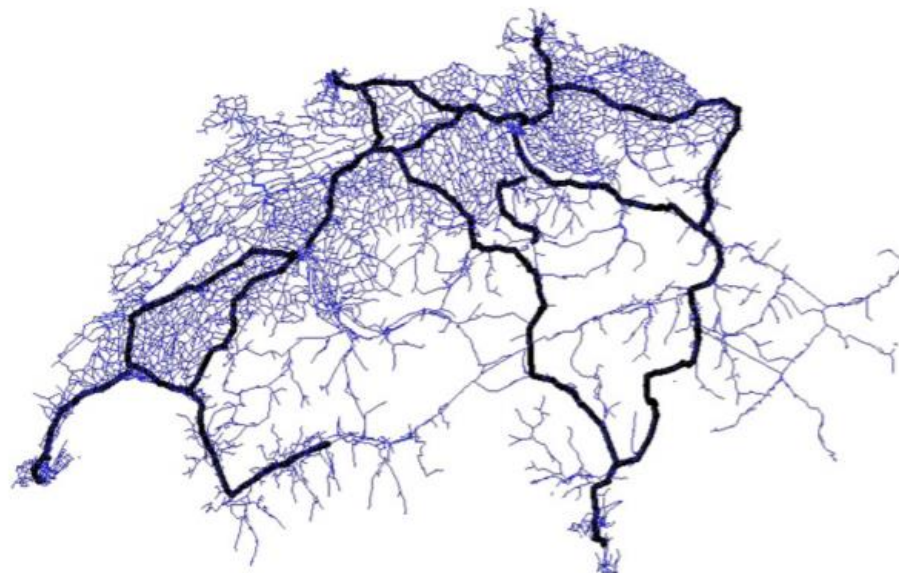
- 2つの異なる経路選択モデル $P_n(p|\mathcal{C}_s;\beta)$ を推定した

Path Size Logit (PSL) モデル (Ben-Akiva and Ramming, 1998; Ben-Akiva and Bierlaire, 1999)

Subnetwork モデル (Frejinger and Bierlaire, 2007)



観測されたデータの例



使用したスイス国内ネットワーク

通常のロジット型配分では経路間のリンクの重複が問題となる



下の2つの経路の相関が考慮されず
平等に配分されてしまう



経路の重複による魅力低下を考慮

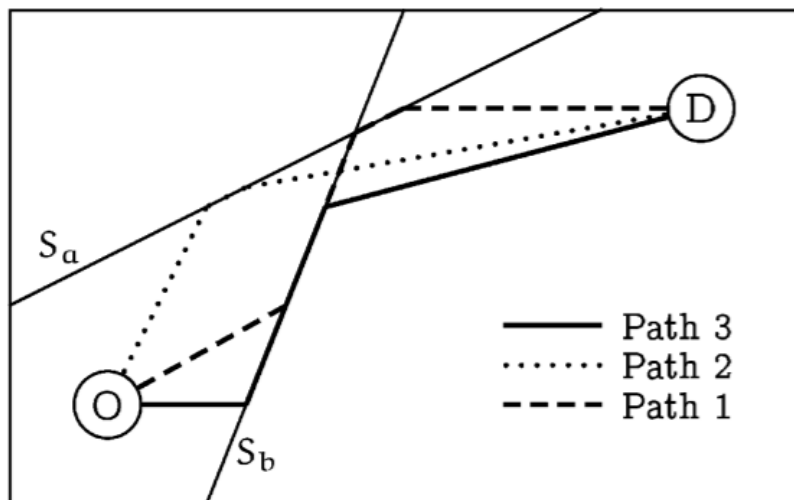
$$U_{ni} = V_{ni} + \frac{1}{\mu} \ln PS_{ni} + \varepsilon_{ni}$$

$$PS_{ni} = \sum_{a \in \Gamma_i} \frac{l_a}{L_i} \frac{1}{\sum_{j \in C_n} \delta_{aj}}$$

Path-Size 修正項
(重複により魅力減少)

Γ_i	経路 i の全てのリンク集合
l_a	リンク a のリンク長
L_i	経路 i の経路長
$\sum_{j \in C_n} \delta_{aj}$	選択肢集合 C_n 内で いくつの経路がリンク a を利用しているか

- モデルを複雑にすることなく, 経路重複構造を表現
- ネットワークの骨格となるリンク以外を削除し, 部分グラフを構築する
物理的な重複に関係なく, 経路がリンクを共有した場合には相関していると仮定
- Mixed logit の一種であるError Component Logit を用いて記述
- PSL, MNL, 通常のECと比較した結果, 既往モデル同等の尤度比が得られ, 十分な適合度を持つことを確認



$$U_1 = \beta^T X_1 + \sqrt{l_{1a}} \sigma_a \xi_a + \sqrt{l_{1b}} \sigma_b \xi_b + \varepsilon_1$$

$$U_2 = \beta^T X_2 + \sqrt{l_{2a}} \sigma_a \xi_a + \varepsilon_2$$

$$U_3 = \beta^T X_3 + \sqrt{l_{3a}} \sigma_a \xi_a + \varepsilon_3$$

β^T : 未知パラメータ X_n : 説明変数ベクトル ε_n : 誤差項
 F_n : Error Componentの要素行列 ($J \times Q$)
 T : 分散共分散行列 ($Q \times Q$) l : subnetworkにおける経路重複長
 J : 経路数 Q : Subnetwork数 ξ : 標準正規乱数 $N(0,1)$

各モデルの推定の適合度や係数

Table 3
Estimation results (continued)

Coefficient	PSL	Subnetwork
ln(Path size) based on free-flow time	1.04	1.10
<i>Scaled estimate</i>	<i>1.04</i>	<i>1.05</i>
(Rob. standard error) Rob. <i>T</i> -test	(0.134) 7.81	(0.141) 7.78
Covariance		0.217
<i>Scaled estimate</i>		<i>0.205</i>
(Rob. standard error) Rob. <i>T</i> -test		(0.0543) 4.00
Number of simulation draws	–	1000
Number of parameters	11	12
Final log-likelihood	–1164.850	–1161.472
Adjusted rho square	0.145	0.147

Sample size: 780, Null log-likelihood: –1375.851.

BIOGEME (Bierlaire, 2003, 2007) has been used for all model estimations.

- 経路の大きさ: 推定値は正であり, 理論 (Frejinger and Bierlaire, 2007) と一致している
- 共分散推定値: 0から有意に離れている.
 高速道路を用いた経路間に有意に相関があることを示している
- 両モデルの比較:
 Subnetwork モデルはPSLモデルに比べ十分に良い適合度
 これは Frejinger and Bierlaire, 2007 での知見と一致している
 $(\chi^2_{0.05,1} = 3.84$ に比べて尤度比検定は 6.756)

- データ操作なしにネットワークフリーデータとネットワークモデルを調和させるようなモデリングの枠組みを提案
- DDRはマップマッチングのような経路を仮定する手法を用いず、あいまいさを残したままモデリングしている
- 経路選択モデルは既存のモデルを用いて推定可能