Route choice modeling with network-free data

Bierlaire, M. and Frejinger, E.,
Transportation Research PartC 16, pp. 187-198, 2008

2008年10月02日
論文ゼミ #10
M1 山川佳洋

BehaviorInNetworks
都市生活行動学研究室

目次

- 背景
- 既存の研究
- DDRの提案
- DDRを用いた定式化
- ケーススタディ
- 結論

背景

経路選択モデルの問題点

経路選択モデルは構築や推定が難しい

- ■選択肢集合の同定
- ■選択肢間の相関
- ■データの取得や扱い方

本論文はデータに関する問題に焦点を当てる

既存のデータ取得方法(1)

- メール，電話を用いた調査
- GISと電話などを組み合わせる

方法

- 経路をセグメントごとに報告させる
- 長距離トリップにおいて通過した中間地点を報告

短所

- 詳細な全てのトリップを記述することは不可能
- 抜け落ちデータは既知の２点間の最短経路

既存のデータ取得方法(2)

□GPSを用いた受動監視

長所

- 複数日のトリップデータを自動的に集められる
- 電子形式なので直接利用可能

短所

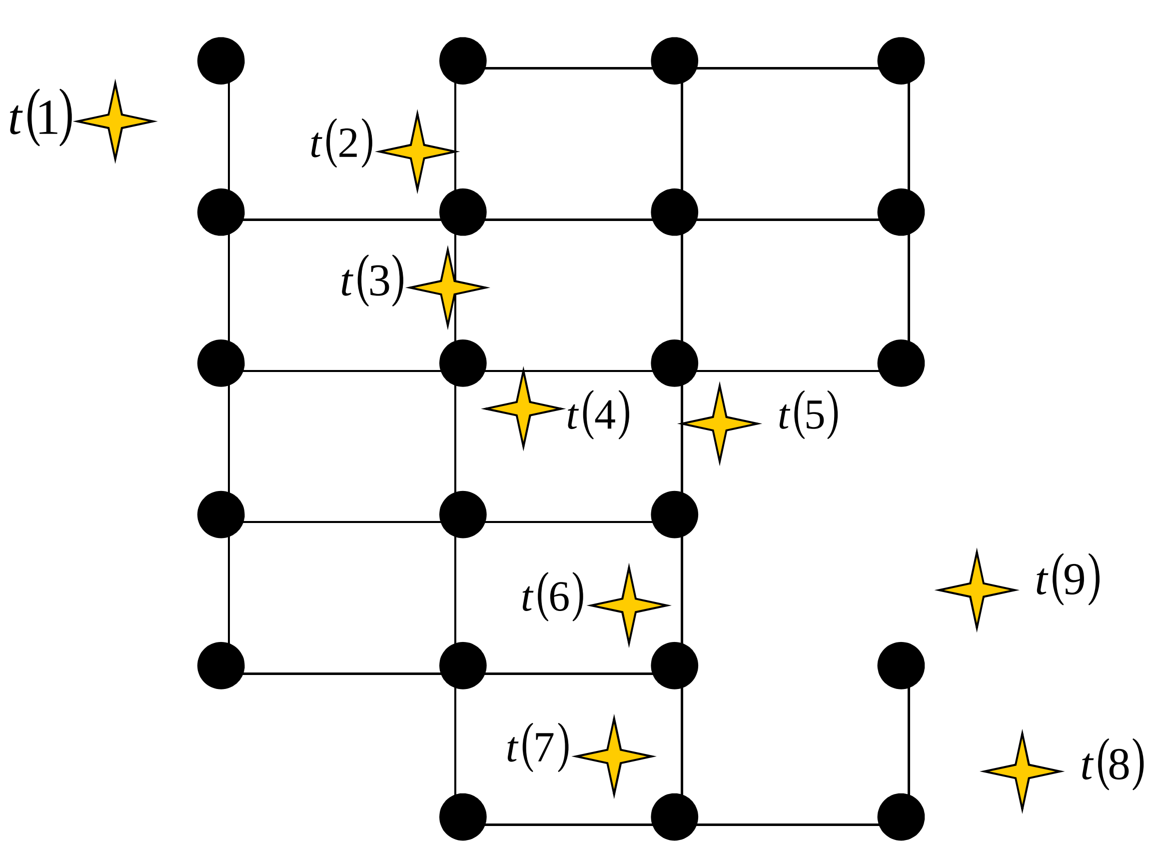
- 衛星の時計の誤差，受信機のノイズエラー，種類
- 衛星の数，大気状態，周囲の環境に影響を受ける
- マップマッチングなどのデータ加工が必要

→バイアスや誤差につながる
　マップマッチングは計算時間でしか性能を評価できない

マップマッチング(1)

補足

トリップデータの抽出



位置データ

位置間直線距離

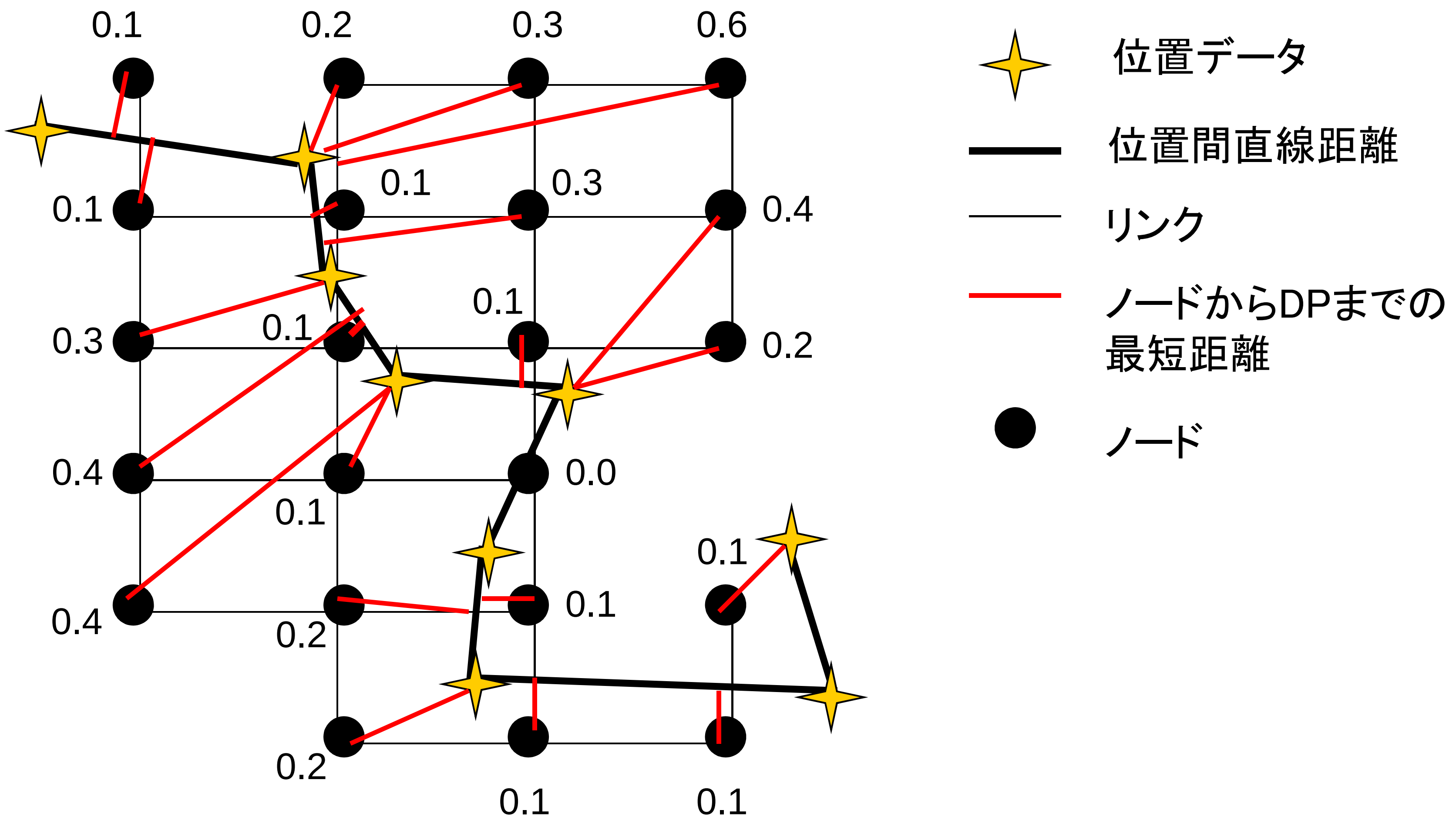
リンク

ノード

$t(n)$ 出発到着時刻間の位置データの取得時刻

マップマッチング(2)

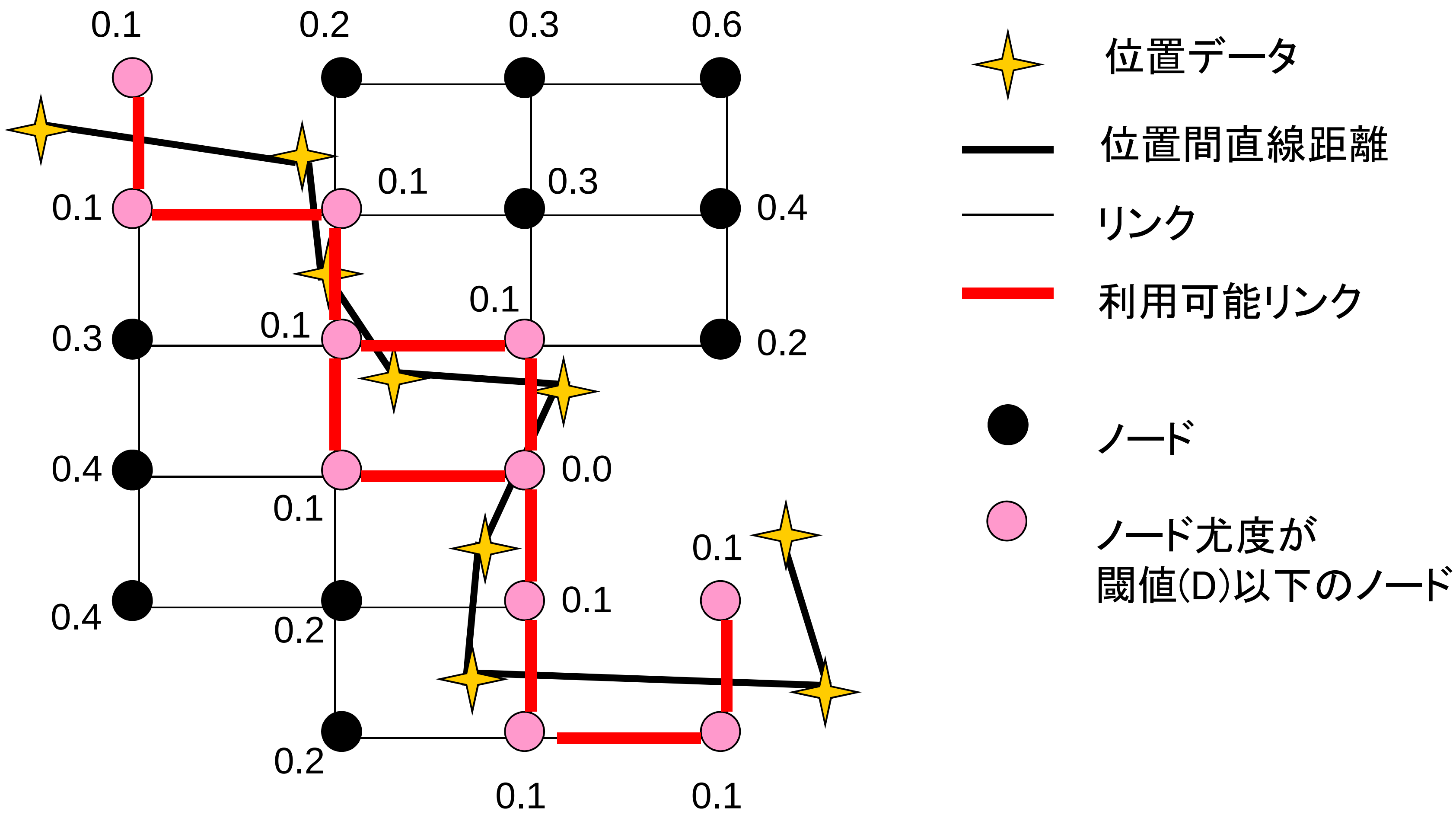
ノード尤度の算出



マップマッチング(3)

補足

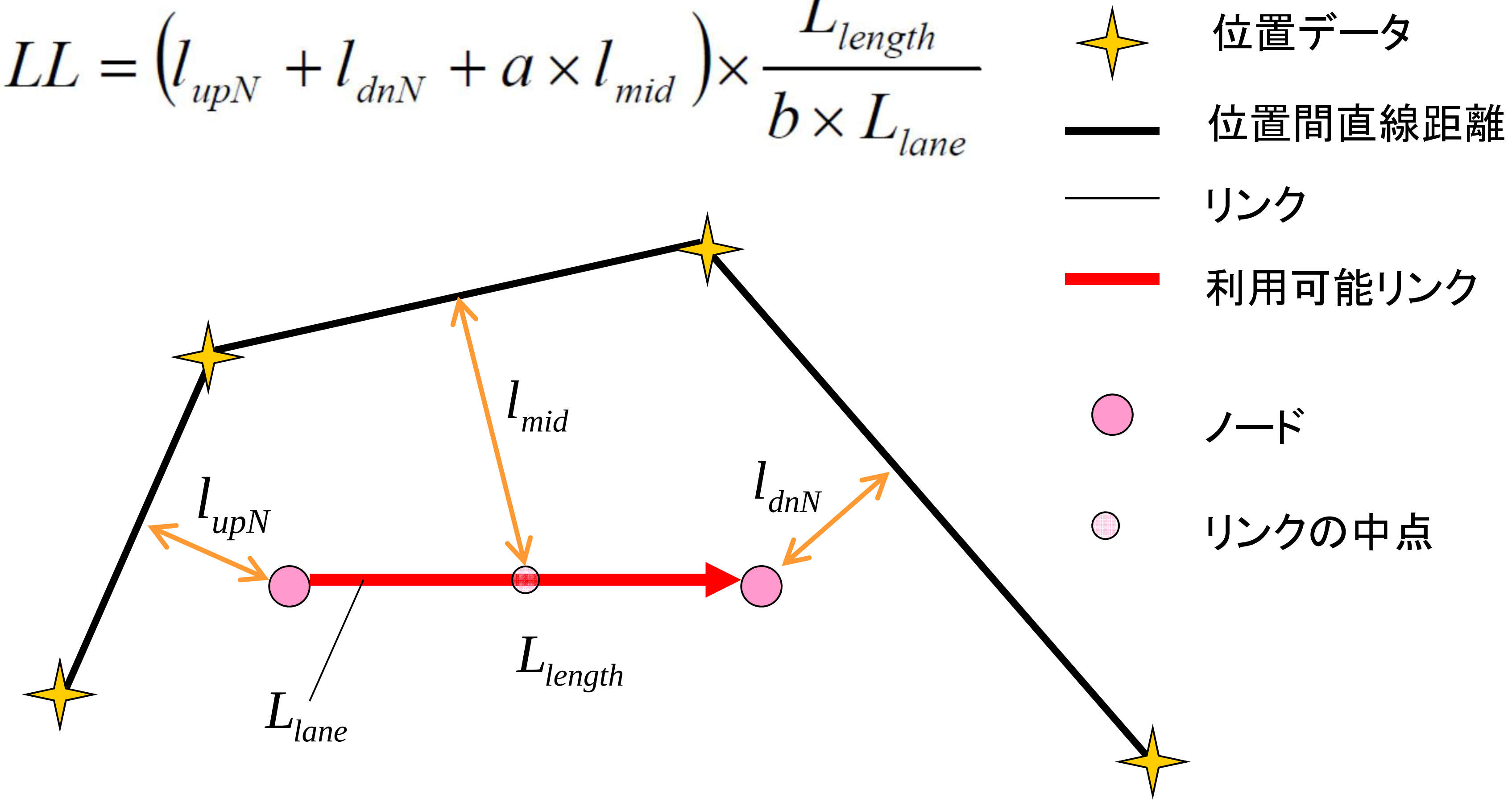
利用可能リンクの抽出



マップマッチング(4)

リンク尤度の算出

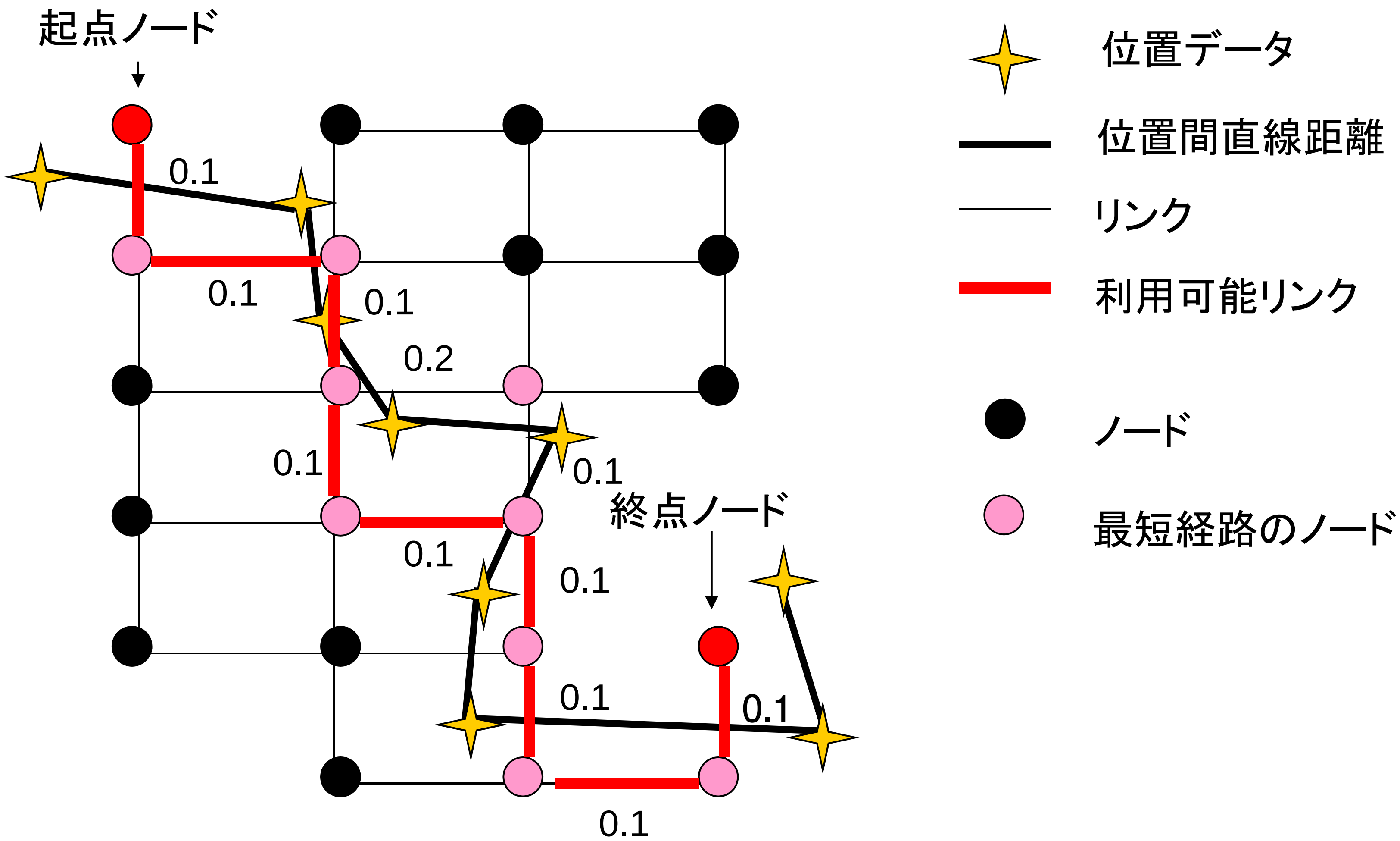
$$LL = \left(l_{upN} + l_{dnN} + a \times l_{mid}\right) \times \frac{L_{length}}{b \times L_{lane}}$$



マップマッチング(5)

補足

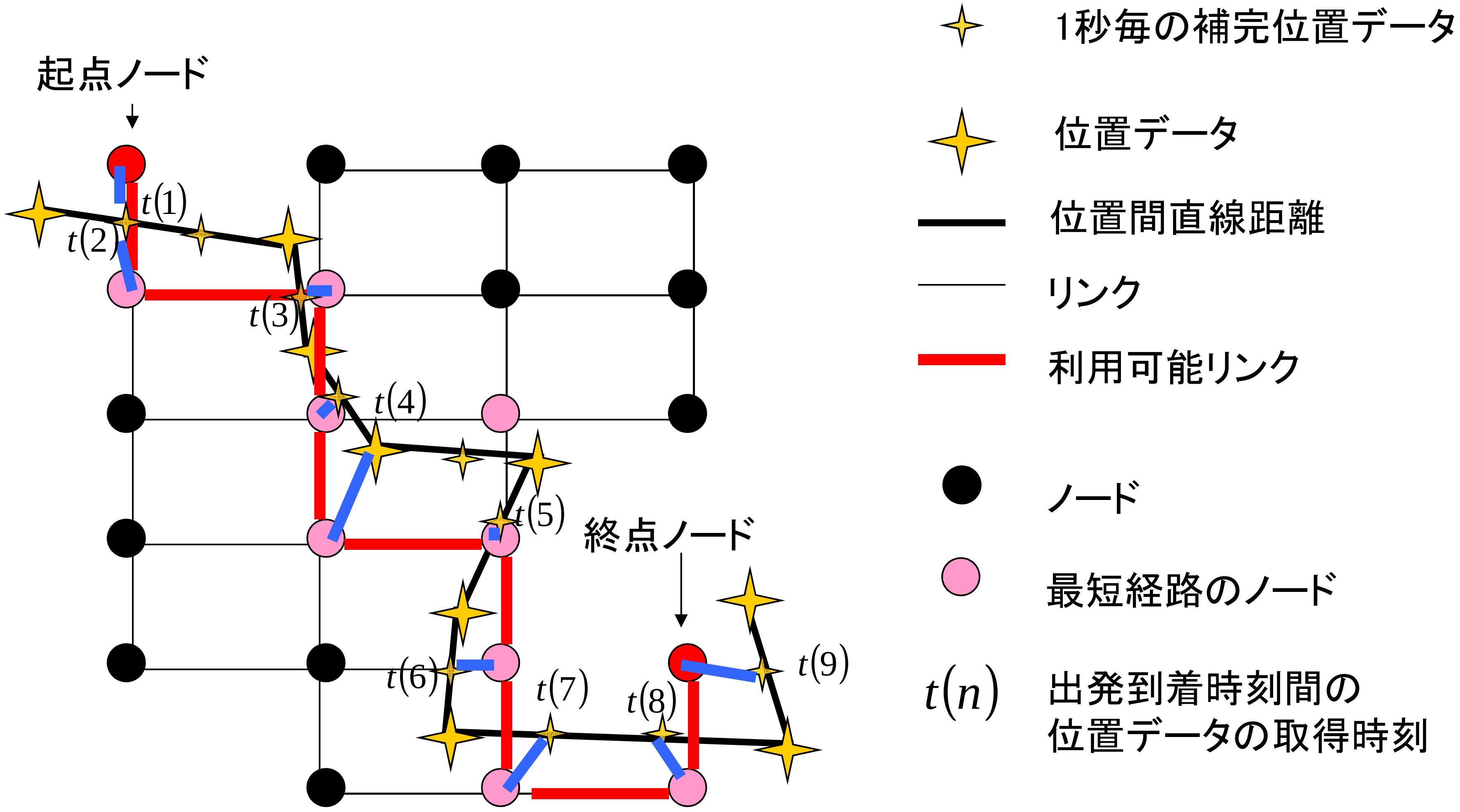
OD決定と利用可能リンクでの最短経路探索



マップマッチング(6)

補足

リンク旅行時間の算出



新たなモデリング手法の提案

ネットワークに対応した正確な選択経路データを得ることは不可能

マップマッチングなどのデータ操作を行うことなく,
ネットワークフリーデータ(ネットワークと直接つながりを持たないデータ)と
ネットワークモデルをうまく調和させモデリングする新たな手法論を提案する

特徴

- 1つの観測について対応するいくつかの経路を許容している
- あいまいなデータに関して1つの経路を特定するための仮定を必要としない

DDRの定義(1)

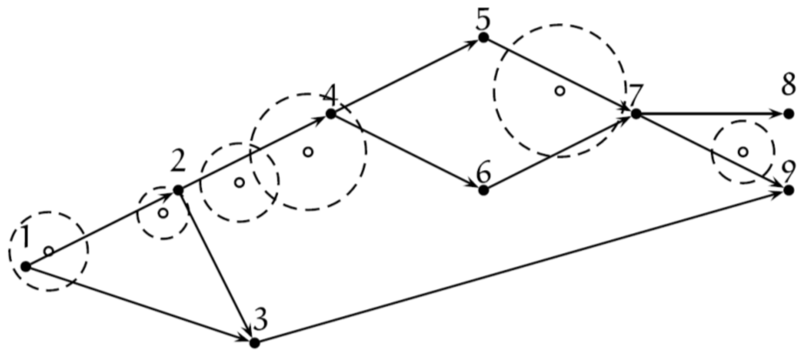
DDR (Domain of Data Relevance)

データの有効範囲，データが関連をもつ領域

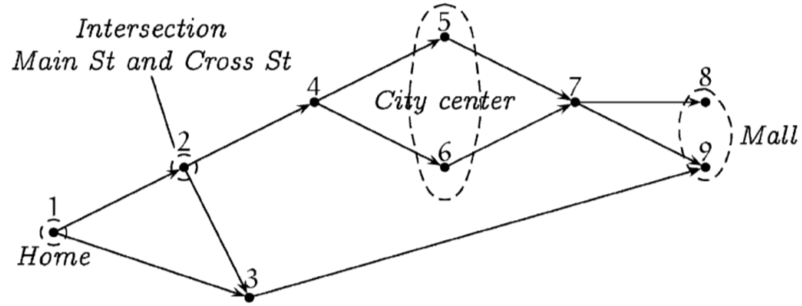
ネットワークフリーデータとネットワークを結びつけるもの

- 厳密な定義は状況によって異なる
- DDRの大きさはデータの正確さ(GPSなら誤差楕円など)
- データを不確かなものとしてモデリング

GPSデータの例



報告データの例



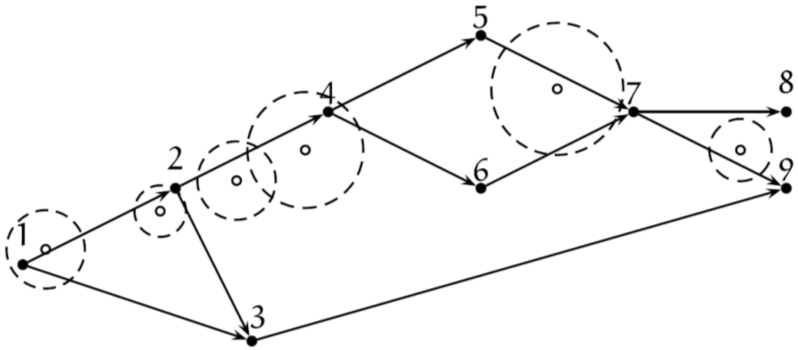
点線がDDRを表す

DDRの定義(2)

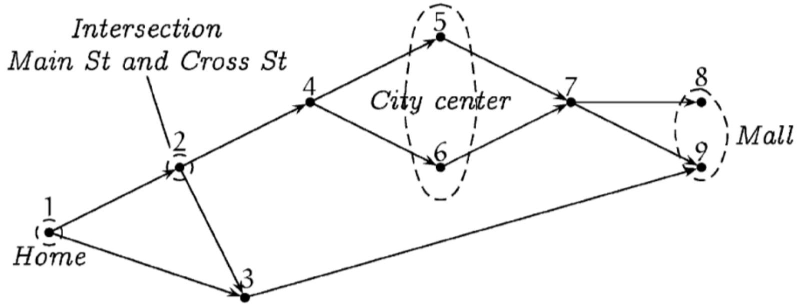
指標関数 $\delta(d,e)$

- 個々のデータとネットワークの要素を関連付ける
- ネットワーク要素 e がDDRと関連していれば1
そうでなければ0
- 個々のデータについて関連するネットワークの
要素のリストを作成

GPSデータの例



報告データの例



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

DDRを用いたモデルの定式化(1)

□経路選択モデル $P(p|\mathscr{C}_n(s);\beta)$

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

p　$\mathscr{C}_n(s)$ の中の1つの経路

□ODペア

$\mathscr{S}$　ネットワーク内の全てのODペア

$\mathscr{S}_i$　観測 i に関連する全てのODペア

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

$\mathscr{S}_i$ は空ではなく, 空の場合はDDRの定義を変更する

DDRを用いたモデルの定式化(2)

■集合 $\mathscr{S}_i$ において観測 i が再現される確率

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

$P_n(s|\mathscr{S}_i)$ ODペアの集合 $\mathscr{S}_i$ の中で実際の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

$P(p|\mathscr{C}_n(s);\beta)$ 経路選択モデル

$P_n(i|\mathscr{S}_i)$ を最大化する

計測方程式 $P_n(i|p)$

計測方程式は観測 i がリンク $(l_1, \dots, l_p)$

から成る経路 p に一致しているとき0以上の値をとる

- 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 $q \quad 1 \le q \le P$ 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 $m_1 \le m_2$ に対して $\delta(d_{m_1}, l_{q_1})=1 \cap \delta(d_{m_2}, l_{q_2})=1$ のとき $q_1 \le q_2$ 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，真の経路 p によって観測 i が生み出される確率を
i と p の距離で定義する

$\Delta(d,l)$ データ d からリンク l の最近地点への距離

$D(d,p)=min_{l\in A_{pd}}\Delta(d,l)$ データ d から経路 p の距離

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

$D(i,p)$ 観測 i と経路 p の距離，$D(d,p)$ の平均や合計で定義

$P_n(i|p)$ $D(i,p)$ の分布によって定義する

$P_n(s|\mathscr{S}_i)$

$|\mathscr{S}_i| > 1$ のとき $P_n(s|\mathscr{S}_i)$ に対応するモデルを定義する

最も簡単なもので全てのODペアに等しい確率を割り当てたもの

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

追加情報が利用可能であれば，自宅や勤務地が
含まれるODペアに高い確率を割り当てることもできる

DDR補足

- DDRの厳密な定義は問題依存なのでケースごとに以下の2点に注意して定義する

1. データ d のDDRは空ではいけない

   $\delta(d, e) = 0 \forall e$ のとき

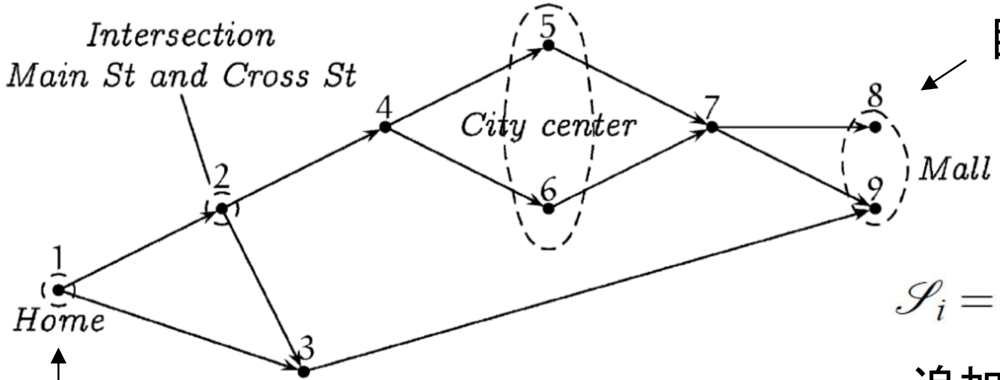
   DDR定義しなおす(基本的にはサイズを大きくする)

2. DDRが大きすぎるといけない

   大きすぎると定義式の和の部分が指数関数的に増えてしまい，モデルを特定化できない(計算できない)

例1(報告トリップ)

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



目的ノードは2つ考えられる

出発ノードは既知

$\mathscr{S}_i = \{(1,8),(1,9)\}$ (s_1 と s_2 と呼ぶ).

追加情報はないので

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

$\mathscr{C}(s_1) = \{(1,2,4,5,7,8),(1,2,4,6,7,8)\}$ p_1 と p_2 と記す.

$P(i|p_1) = P(i|p_2) = 1$

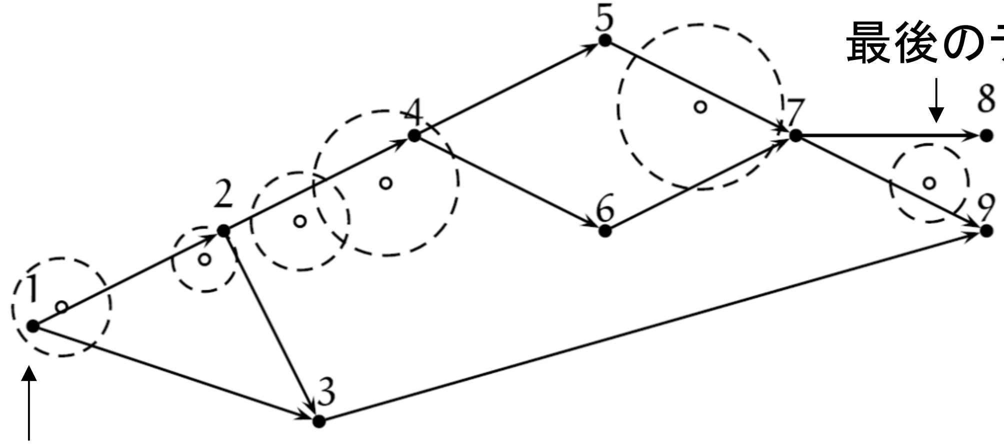
$\mathscr{C}(s_2) = \{(1,2,4,5,7,9),(1,2,4,6,7,9),(1,2,3,9),(1,3,9)\}$ $p_3,\ldots,p_6$ と記す

$P(i|p_3) = P(i|p_4) = 1 \quad P(i|p_5) = P(i|p_6) = 0$

$$P(i|\mathscr{S}_i) = \frac{1}{2}[P(p_1|\mathscr{C}(s_1);\beta) + P(p_2|\mathscr{C}(s_1);\beta)] + \frac{1}{2}[P(p_3|\mathscr{C}(s_2);\beta) + P(p_4|\mathscr{C}(s_2);\beta)]$$

例2(GPSデータ)

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

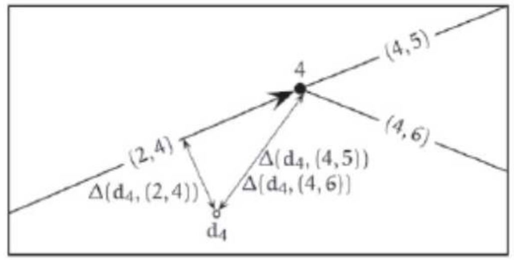


最後のデータのDDRに関連するノードがない

このDDRが交わるリンクのシンクノードを考える

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

出発ノードに相当するノードが1つ



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

$P(i|p_3)=P(i|p_4)=0$ 　　　　$p_1,\ldots,p_4$ と記す

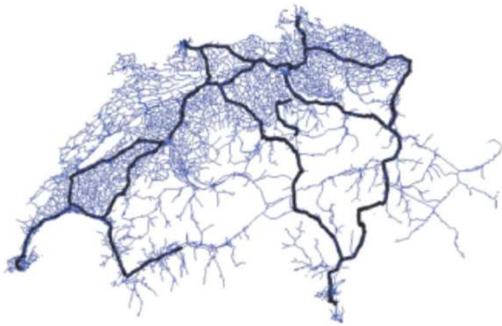
$P(i|p_1)$ と $P(i|p_2)$ はデータ取得場所と経路との距離の関数で表される

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

$$P(i|s)=P(i|p_1)P(p_1|\mathscr{C}(s);\beta)+P(i|p_2)P(p_2|\mathscr{C}(s);\beta)$$

ケーススタディ

- スイスにおいて行われた調査
  長距離トリップにおいて出発地と目的地，さらに最大3つまで途中に通過した都市を回答
- DDRは報告された場所の郵便番号がカバーするエリア
- 使用するネットワークは39411の一方向リンク14841のノード(22737リンク，8107ノード，MPP2004)
- 計算上全てのODペアを考慮せず，ランダムに2つ選んでいる
- 選択肢集合は link elimination approach を用いて各ODに対し45本生成
- 経路選択モデルには Path Size Logit と Subnetwork モデルを使用



おまけ1

Path Size Logit モデル (Ben-Akiva and Lerman, 1985)

$$U_{in} = V_{in} + \beta_{\mathrm{PS}} \ln \mathrm{PS}_{in} + \varepsilon_{in}$$

$$\mathrm{PS}_{in} = \sum_{a \in \Gamma_i} \frac{l_a}{L_i} \frac{1}{\sum_{j \in C_n} \delta_{aj}}$$ (他の定義もある)

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

おまけ2

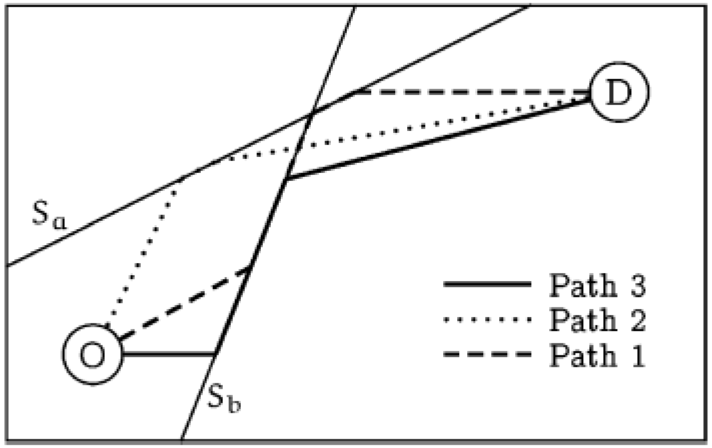
Subnetwork モデル (Bierlaire and Frejinger, 2007)

$$\mathbf{U}_n = \beta^{\mathrm{T}}\mathbf{X}_n + \mathbf{F}_n\mathbf{T}\boldsymbol{\zeta}_n + \boldsymbol{v}_n$$

$$U_1 = \beta^{\mathrm{T}}X_1 + \sqrt{l_{1a}}\sigma_a\zeta_a + \sqrt{l_{1b}}\sigma_b\zeta_b + v_1$$

$$U_2 = \beta^{\mathrm{T}}X_2 + \sqrt{l_{2a}}\sigma_a\zeta_a + v_2,$$

$$U_3 = \beta^{\mathrm{T}}X_3 + \sqrt{l_{3b}}\sigma_b\zeta_b + v_3,$$



結論

- 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はあいまいなデータに対して経路を仮定するのではなく，あいまいさを持たせたままモデリングをしている
- 経路選択モデルは既存のモデルを用いて推定可能
- データ操作の手間がはぶけるのでモデル推定が簡単になる