理論談話会#2 2025/05/12

A sequential transit network design algorithm with optimal learning under correlated beliefs

Gyugeun Yoon, Joseph Y.J. Chow. (2024).
Transportation Research Part E: Logistics and Transportation Review, vol.191, 103707.

M1 門坂 佑真

Abstract

「逐次的な輸送ネットワーク設計 ＋ 最適学習 → AI駆動型アルゴリズム」の提案

- 現在GPS等により様々なデータにアクセスできるようになった
- しかしモビリティシステム導入時、需要の推定が難しい
- 需要との不整合を防ぐためにルートを徐々に拡張する状況を考える
- 多腕バンディット、知識勾配、相関信念付き知識勾配の3つの学習方針を比較

Novelty, Usefulness, Reliability

新規性

- セグメントレベルの拡張を考え、情報を逐次的に得る**SSTNDP問題**を提示
- **最適学習システム**（MAB, KG, KGCB）を導入

有用性

- 十分な需要パターンデータがない地域にも適用可能
- 需要の不確実性を交通ネットワーク設計に反映可能
- プロジェクト予算を時間的に分割し、順次実行する段階的拡大に効果的

信頼性

- 3つの異なる学習方策 (MAB, KG, KGCB) を2つのシナリオ (人工グリッドネットワーク, NYC) で比較
- 相関信念ベースのKGCBがSSTND問題で優位であることを確認

1.Introduction

2. Literature review

3. Proposed Methodology

4. Numerical experiments

5. Conclusions

1. Introduction

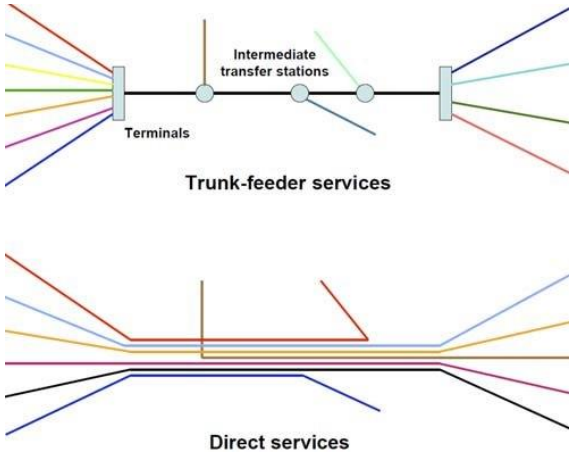
従来モビリティサービスのルート設計

- 調査データを収集 → ルート設計全体の需要予測に利用
- 旅行者が既存交通手段との比較を熟知しているとすれば予測可能

↓

新興モビリティサービス（e.g. modular autonomous vehicle, feeder-trunk route）のルート設計

- **旅行者がシステムの特徴を認識していない → 仮定の質問は信頼性に欠ける可能性**
- **調査は5～10年単位で行われるのに対し、より早いスパンで展開される**
- **運行中にデータを蓄積可能（路線自体がデータ収集センサー）**
- **需要に適合させつように逐次的にルート設計を実施可能**



“transit service route design with optimal learning” を考える！

1. Introduction

最適学習 (optimal learning)

- 後続の決定の不確実性が低くなるように逐次決定が行われる = 強化学習の一種
- 類似性から**逐次的なモビリティルート設計に使える**
- Yoon & Chiw (2020)はバンディットを使用 → **ルート拡張から得られる知識が独立としている**
- Frazier et al. (2008)は知識勾配を使用 → **共分散行列の大きさが問題となる可能性**

貢献

- **方法論的貢献：セグメント拡張レベルでの相関確率変数を用いた最適学習を統合したシステム設計**
  - データの流れ、出入力、経路設計ルゴリズムと最適学習手法の統合
  - 観測されたOD需要の共分散をセグメントレベルの共分散に関連づける
- **実証的な貢献：3つの異なる学習方策を2つのシナリオで比較**
  - 学習方策：MAB, KG, KGCB (ベンチマークとしてGreedyとCR参照方策も)
  - シナリオ：人工グリッドネットワーク, NYC

1. Introduction

2.Literature review

3. Proposed Methodology

4. Numerical experiments

5. Conclusions

2-1. Mobility service route design with fixed routes

2.1.1. Conventional approaches with static demand

従来の考え方

- 静的な需要に基づいて目的関数 (e.g.総旅行時間、総乗客数) を最大化する固定経路の集合を確立する
- ルートが変更されることはない ← 需要が一定ならOK

具体的なルート生成方法

- ネットワーク特性に基づいてルートを構築
- あらかじめ設定された実行可能なルートの中から選択
- 例えば：ルート構築ヒューリスティックス, 遺伝的アルゴリズム, 列生成アルゴリズム, 適応近傍探索メタヒューリスティクス

需要変動下での経路設計 (最近の発展)

- 利用者数が提供されるサービスに依存する
- 需要行列とモデルが既知で、乗車率を推定可能であるという仮定を置いている

2-1. Mobility service route design with fixed routes

2.1.2. Sequential design

従来の考え方

- **進化モデル**：周辺環境の変化に対応するようにネットワークを修正
- **繰り返しネットワーク設計**：順次利用可能な予算に対して1タイムステップに一段階ずつネットワークを拡張
- どちらも各時間ステップでは決定論的な需要を前提

動的需要（最近の発展）

- 二段階の確率的プログラムやロバスト最適化を用いたルート設計研究
- 段階的な開発を扱った研究もいくつかある（：不確実性軽減のために建設→次路線の決定→建設…）
- 近年はADP (Approximate dynamic programming) とRL (Reinforcement learning) が用いられる
  - ADP：不確実性の分布が意思決定プロセス全体を通じて既知と仮定
  - RL：分布の信念に不確実性を含み、exploitationとexplorationの間の意思決定を管理
- ルートの事前列挙や「ネットワーク設計＋頻度設定」にRLを適用した例はある

需要の最適学習へのRLの適用
セグメントレベルの決定を伴う逐次経路設計問題の検討
} **文献なし！**

2-2. Reinforcement learning in sequential planning for transportation systems

強化学習

- 一連の行動と観測を通じて環境に関する知識を更新していく方法
- 探索して知識を蓄積した後、現在の報酬と将来の期待値の組み合わせに基づいて選択

Examples of Learning Techniques Used in Transportation Domain.

Subject	Approach	Information (I) and Action (A)
Intercity route planning (Zolfpour-Arokhlo et al., 2014)	Q-value based dynamic programming	I: travel time A: route choice
Sequential reliable route selection (Zhou et al., 2019)	Multi-armed bandit	I: generalized travel time, reliability A: route choice
Delivery vehicle allocation (Huang et al., 2019)	Knowledge gradient	I: operational cost curve A: vehicle allocation to regions
Demand management of EV charging stations (Römer et al., 2019)	Contextual bandit	I: load on grid A: charging price adjustment
Stochastic online shortest path routing (Zhu and Modiano, 2018)	Combinatorial bandit	I: end-to-end delay A: path choice
Dynamic electric vehicle routing (Basso et al., 2022)	Q-learning-based Safe RL	I: cumulative energy cost and risk of failure A: choice of node to visit
Bus network design and frequency setting (Yoo et al., 2023)	Q-value based learning	I: passenger demand A: route expansion

2-3. Research gap summary

既存手法

- ネットワークの動的手法は存在 ← **需要が決定論的**
- 不確実な需要へのネットワーク設計も存在 ← **逐次のネットワーク拡張や動的需要ではない**

逐次トランジットネットワーク設計と最適学習方策の共通点

- セグメント延長が行動、カバーされる追加需要が報酬に対応
- 観測は期待報酬推定の精度を向上させるために使用
- 現在の時間ステップでの行動は信念を再形成して将来行動に影響を与える

本研究の位置付け

RLをトランジットネットワーク設計問題に統合することで...

- **現在利用可能な情報のみに基づく設計と更新された情報から導かれる設計のギャップを縮める**
- **良質なデータを収集するためのアプローチを提供**

1. Introduction

2. Literature review

3.Proposed Methodology

4. Numerical experiments

5. Conclusions

3-1. Preliminaries: Learning policy

変数一覧

Notation	Description	Notation	Description
π	Learning policy	G	Network
N	Set of nodes	n	Node
A	Set of bidirectional segments	a	Segment
T	Time horizon	t	Extension time step
M	Number of segment extension trials	$\mathbb{N}(\bullet, \bullet)$	Normal distribution with mean and covariance
a^t	Segment that can be chosen at t	A^t	Set of a^t
$a^{t,m}$	a^t chosen at segment extension trial m	N^t	Set of nodes corresponding to A^t
$n^{t,m}$	Node corresponding to $a^{t,m}$	$v_{a^t}^{m,\pi}$	Value of a^t after m trials under π
K	Number of candidate routes	R_k	k-th route
$x_{i,j}$	True OD demand between i and j	X^t	Total OD demand coverage at t
X_{a^t}	Sum of all available $x_{i,j}$'s if appending a^t	$\Delta X_{a^{t,M}}$	Sum of additional $x_{i,j}$'s expected to be covered by appending $a^{t,M}$
$a^{t(q)}$	q-th segment of A^t	D^q	Set of possible $x_{i,j}$'s if integrating $a^{t(q)}$
$\boldsymbol{\theta}^D$	Vector of $x_{i,j}$'s	Σ^D	True covariance matrix of $x_{i,j}$'s
$\boldsymbol{\theta}^A$	Vector of $X_{a^{t(q)}}$	Σ^A	Covariance matrix of $X_{a^{t(q)}}$
$\boldsymbol{\theta}_t^A$	Vector of $X_{a^{t(q)}}$ at t	Σ_t^A	Covariance matrix of $X_{a^{t(q)}}$ at t
$\widehat{\boldsymbol{\theta}}_0^D$	Initial prior of $\boldsymbol{\theta}^D$	$\widehat{\Sigma}_0^D$	Initial prior of Σ^D
$\widehat{\boldsymbol{\theta}}_t^D$	Updated prior of $\boldsymbol{\theta}^D$ at t	$\widehat{\Sigma}_t^D$	Updated prior of Σ^D at t
$\widehat{\boldsymbol{\theta}}_{t,0}^A$	Initial prior of $\boldsymbol{\theta}^D$ for segment choice trials at t	$\widehat{\Sigma}_{t,0}^A$	Initial prior of Σ^D for segment choice trials at t
$\widehat{\boldsymbol{\theta}}_{t,m}^A$	Updated prior of $\boldsymbol{\theta}^D$ for segment choice trial m at t	$\widehat{\Sigma}_{t,m}^A$	Updated prior of Σ^D for segment choice trial m at t
$\widehat{x}_{i,j}$	Observed $x_{i,j}$	$\widehat{X}_{a^t}$	Observed X_{a^t}
$\widetilde{\sigma}_{a^t}^{2,m}$	Change in estimated variance of $\widehat{X}_{a^t}$ at m	$\widehat{\sigma}_W^2$	Variance of observations
$\widetilde{\boldsymbol{\sigma}}(a^{t,m})$	Change in $\widehat{\Sigma}_{t,m}^A$ after observing $a^{t,m}$	$W_{a^{t,m}}^{m+1}$	Observation of the measurement of $a^{t,m}$
$\widehat{\boldsymbol{\theta}}_{t,m}^A(a^{t,m})$	Belief of the measurement of $a^{t,m}$	$\boldsymbol{e}_{a^{t,m}}$	Unit vector with 1 indicating $a^{t,m}$
Var^m	Variance after m observations	ε^{m+1}	Random error
κ	Number of time steps in MAB	n_{a^t}	Number of a^t chosen
$\boldsymbol{\beta}^n$	Vector of precision of $x_{i,j}$'s after n-th update	$\boldsymbol{\theta}^n$	Vector of belief on mean of $x_{i,j}$'s after n-th update
$\boldsymbol{\beta}^W$	Vector of precision of observations	$\boldsymbol{W}^{n+1}$	Vector of observed flows
Ω	Flow observation indicator matrix	ω	Vector indicating observations by 0 and 1
σ_ε^2	Observation error per flow	L	Maximum route length
P	Number of pilots	L_P	Minimum pilot route length
Q_P	Number of observations per pilot	$\boldsymbol{H}$	Diagonal matrix with $\sigma_{i,j}$'s
$\boldsymbol{B}$	Correlation matrix		

3-1. Preliminaries: Learning policy

3.1.1. Multi-armed bandit (MAB)

各試行で明らかになる報酬から生じる**後悔ρ_nを最小化**するように**独立した**選択肢 (arms) から選ぶ手法

||

可能な最大限の報酬と得られた報酬の差

$$\rho_n = \max_{i=1,\cdots,K} \sum_{t=1}^{n} Y_{i,t} - \sum_{t=1}^{n} Y_{I_t,t}$$

I_t：タイムステップtで選んだ選択肢
$Y_{I_t,t}$：観測された報酬
$Y_{i,t}$：未知の分布に従うと仮定された選択肢iの報酬

しかし未知分布のため「最大限の報酬」がわからない

Upper Confidence Bound (UCB)を用いる

||

$O(n)$より低いオーダーで後悔の総量増加を制限

cを選んだときの期待価値 $$\nu_c^{MAB,n} = \mathbb{E}(Y_{I,t}) + \sqrt{\frac{2\log n}{n_i}}$$

exploitation　exploration

n：総試行回数
n_i：選択肢iの試行回数

3-1. Preliminaries: Learning policy

3.1.2. Knowledge gradient (KG)

その後の意思決定を改善するために得られる潜在的な知識を考慮して行動を逐次選択することを考える

測定後の状態における価値の期待改善量を最大化する手法

cを選んだときの期待価値　$\nu_c^{KG,n} = \mathbb{E}[V^{n+1}(S^{n+1}(c) - V^n(S^n)|S^n)]$

S^n：n回目の測定後の状態
$S^{n+1}(c)$：n回目の測定後cを選んだときの状態
V：そのときの価値

最適選択肢　$C^{KG,n} = arg\max_{c\epsilon\mathbb{C}} \nu_c^{KG,n}$

- **UCBよりも「次に得られる情報の価値」を直接計算できる**
- **依然各選択肢は独立と仮定**

3-1. Preliminaries: Learning policy

3.1.2. Knowledge gradient with correlated belief (KGCB)

選択肢相関を考慮したKG

cを選んだときの期待価値　$\nu_c^{KGCB,n} = \mathbb{E}[\max_i(\mu_i^n + \tilde{\sigma}_i(c^n)Z^{n+1})|S^n = s, c^n = c]$

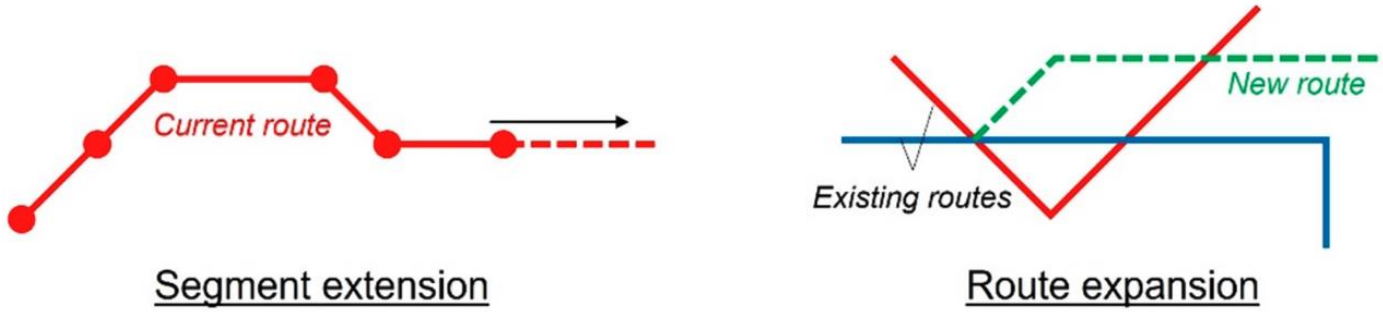
μ_i^n：n回目における測定値の平均についての信念
$\tilde{\sigma}_i(c^n)$：n回目に選んだcについての信念の変化 ← 相関行列Σ^nが影響
Z^{n+1}：観測と信念の違いに関するランダム変数

- **選択肢の相関を考慮可能**
- **「Aを測定する価値」＝「Aの評価改善の価値」＋「Aと相関のあるBやCの評価も同時に改善する価値」**

3-2. Problem statement

2つの拡張

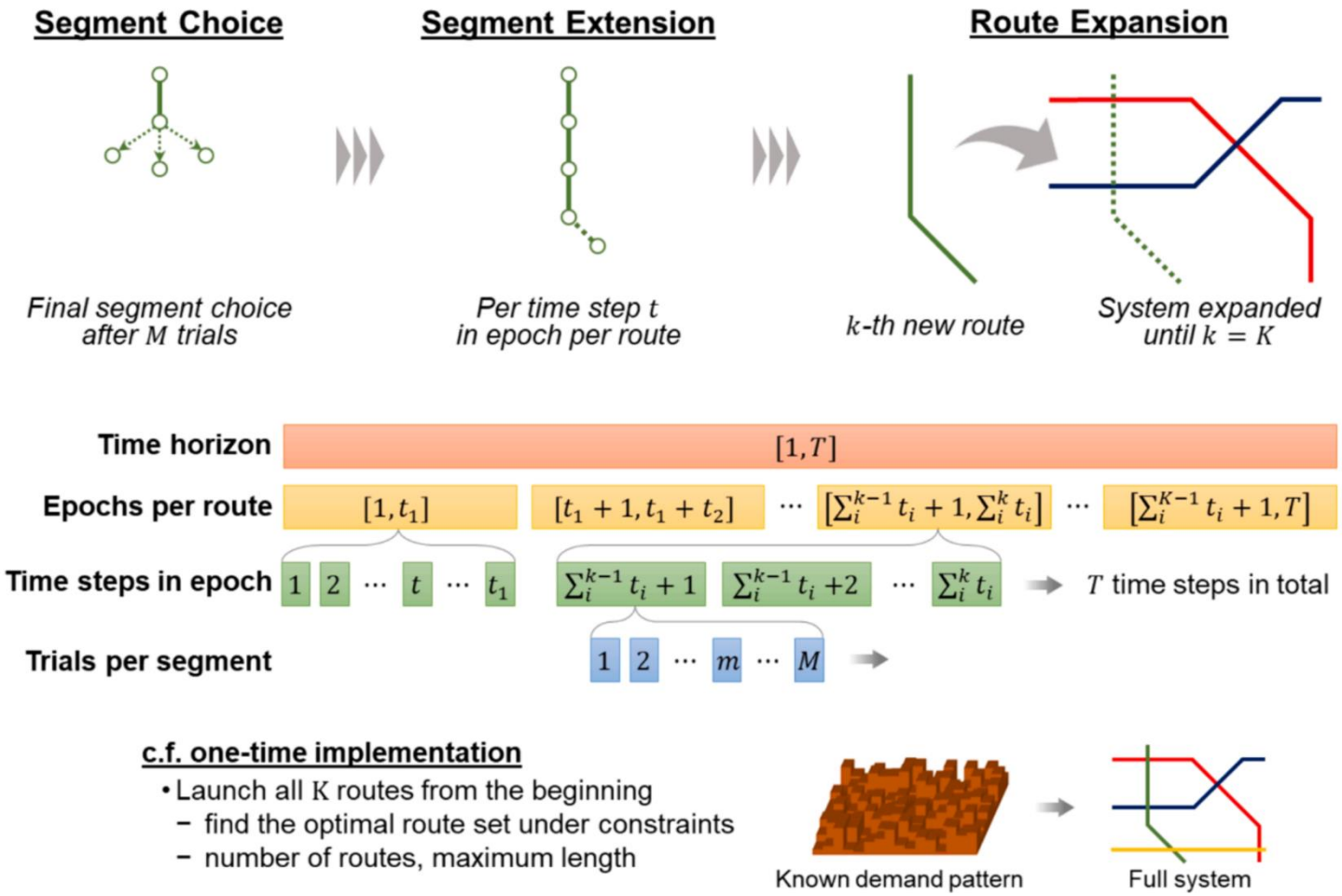
- **Segment extension** : 現在のルートにセグメントを追加して長さを長くする
- **Route extension** : 既存のルート集合に新しいルートを追加
- 区間追加のたびにそこを利用するユーザーから新しい情報を得る
- ルートの最大長さと数は予算の利用可能性から計算される



適応先

- バスはそのまま使える
- 鉄道の場合事前にバスのような代替手段による運行を行う or 嗜好調査により、需要パターンを把握する必要がある
- オンデマンド交通は決定係数がサービス地域ゾーンとなる

3-2. Problem statement



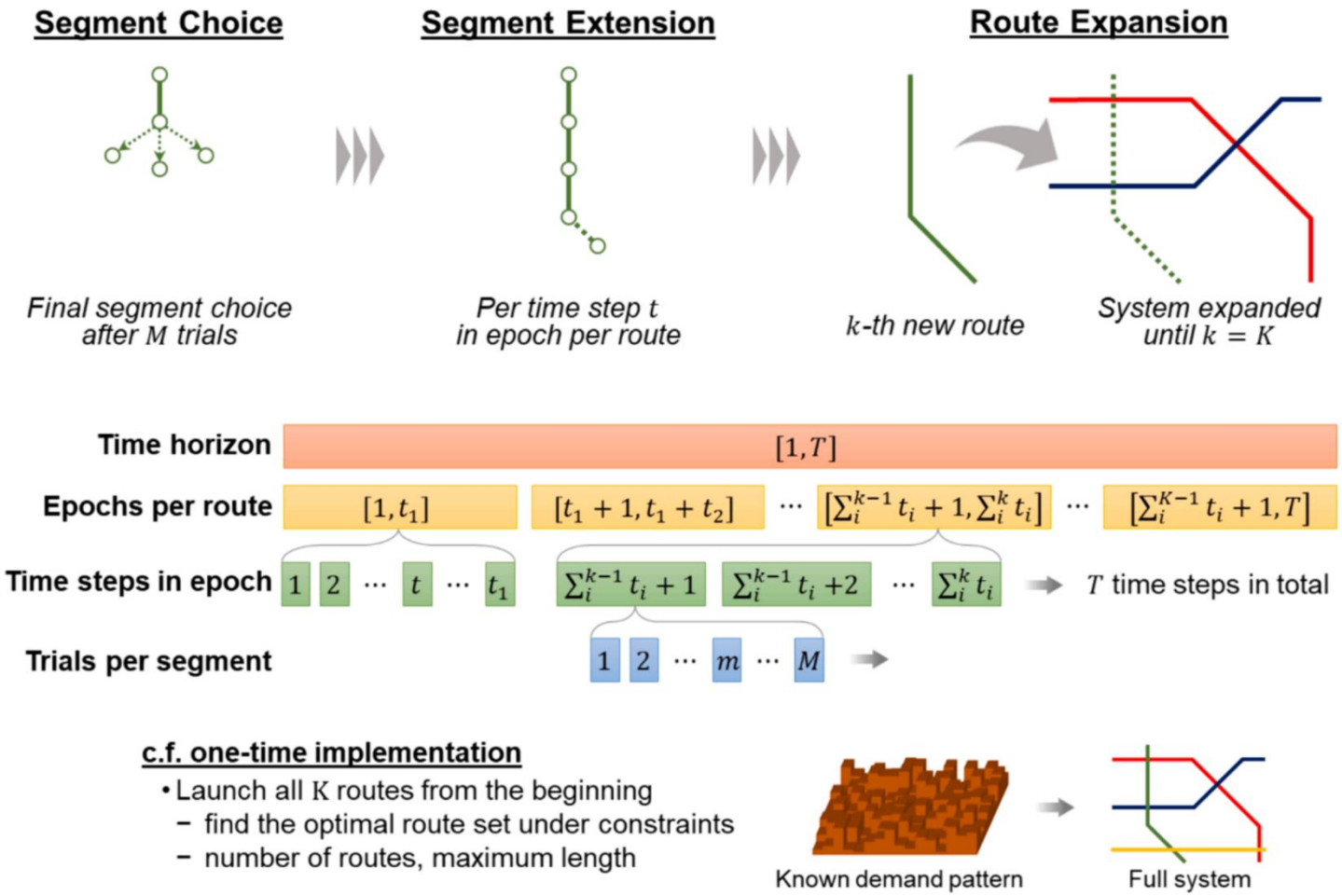
問題設定1：拡張

- K本のルートをネットワーク$G(N, A)$上に時間Tだけかけて作る
- ネットワークはノード$n \in N$と両方向のセグメント$a \in A$からなる
- セグメントaは(p, q)のエンドからなる
- セグメント拡張時間ステップtの中で、M回の試行を行って情報を得たのちに(各試行mで$a^{t,m}$を選択)、ルートR^kにセグメント集合A^tの中から$a^{t,M}$を追加
- 計画者は方策πを使用
- セグメント選択は次式に従う

$$a^{t,m} = \underset{a^t \in A^t}{\operatorname{argmax}}\, \nu_{a^t}^{m-1,\pi}$$

The sequential segment-level transit network design problem (SSTNDP)

3-2. Problem statement



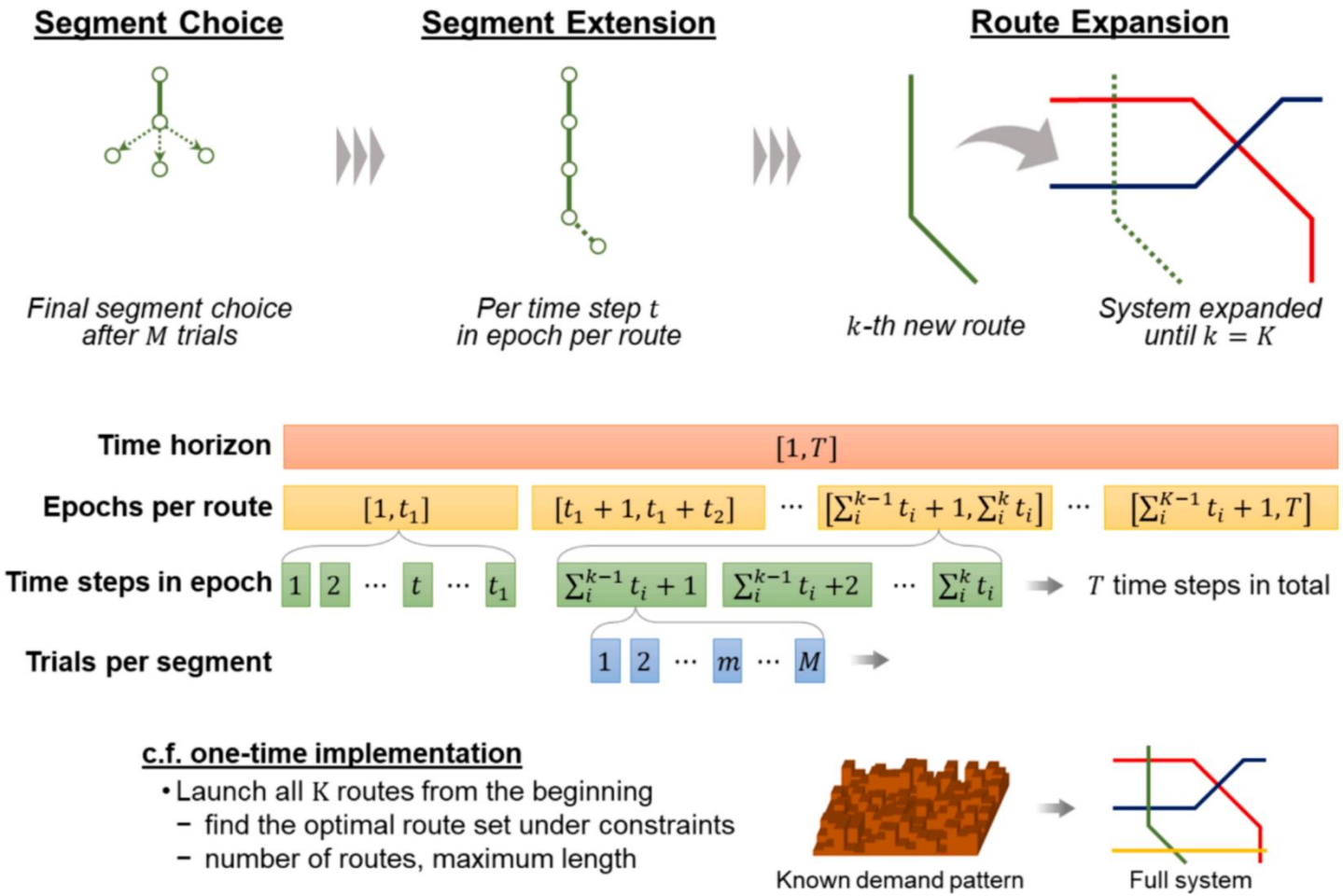
問題設定2：OD需要

- セグメントの価値はOD需要$x_{i,j}$（双方向）に依る

→累積価値の最大化問題

- ルートが同じノードを通っている場合乗り換え可能
- 異なるODペアの需要に相関がある場合もある
- $x_{i,j}$は$\mathbb{N}(\boldsymbol{\theta}^D, \Sigma^D)$の多変量正規分布に従う
- 乗客の割り当てには**all or nothing配分**を採用
  - 容量なし
  - 現実はより複雑
  - 学習メカニズムとネットワークの関係を捉えるためにここは簡単に

3-2. Problem statement



問題設定2：OD需要

- $t+1$時点での需要カバレッジは

$$X^{t+1} = X^t + \Delta X_{a^{t,M}}$$
$$= X^{t-1} + \Delta X_{a^{t-1,M}} + \Delta X_{a^{t,M}}$$
$$= \cdots = \sum_t \Delta X_{a^{t,M}}$$

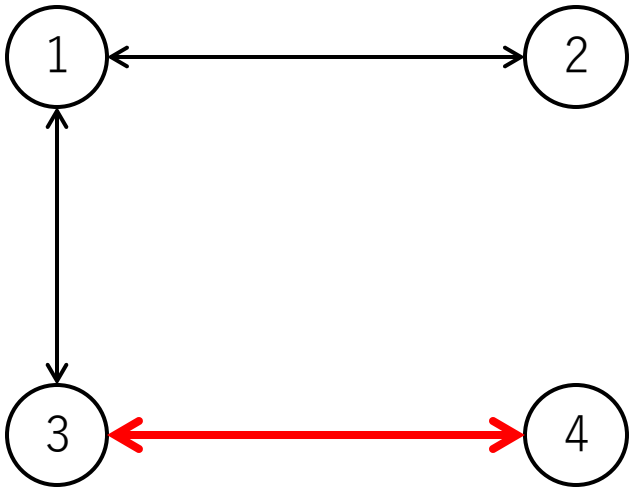
- A^tからq番目のセグメント$a^{t(q)}$追加時のカバーできるODをD^qとすると、その時の需要カバレッジは

$$X_{a^{t(q)}} = \sum_{(i,j)}^{D^q} x_{i,j}$$
$$\mathrm{cov}\left(X_{a^{t(q_1)}}, X_{a^{t(q_2)}}\right) = \sum_{(i,j)}^{D^q} \sum_{(r,s)}^{D^q} \mathrm{cov}(x_{i,j}, x_{r,s})$$

- 直接$\hat{X}_{a^{t(q)}}$(全体)は観測できない → $\hat{x}_{i,j}$(各OD)を観測

3-2. Problem statement

具体例



追加したセグメント： $a^{t(q)} = (3,4)$

カバーできるODペア： $D^q = \{(1,2),(1,3),(2,3),(1,4),(2,4),(3,4)\}$

カバレッジ： $X_{a^{t(q)}} = x_{1,2} + x_{1,3} + x_{2,3} + x_{1,4} + x_{2,4} + x_{3,4}$

- 分散（この価値の不確かさ）σ^2はこれら6つのODペア需要の組み合わせの共分散を足し合わせたもの
- $X_{a^{t(q)}}$のベクトルθ^Aと共分散行列Σ^Aも時間変動する

3-3. Learning scheme design

3.3.1. Design of a prior

- 学習方策は選択のたびに更新される事前知識に依存する
- 確率的過程においては**選択とその報酬の関係に関する事前知識**が必要

真値

- $(\boldsymbol{\theta}^D, \Sigma^D)$：OD需要
- $(\boldsymbol{\theta}_t^A, \Sigma_t^A)$：セグメントレベルの需要

} 不明→事前知識を代わりに置く必要

事前知識

- $(\hat{\boldsymbol{\theta}}_0^D, \hat{\Sigma}_0^D)$：$(\boldsymbol{\theta}^D, \Sigma^D)$の初期信念 ← pilot operation or 外部情報源
- $(\hat{\boldsymbol{\theta}}_t^D, \hat{\Sigma}_t^D)$：各時間ステップ$t$で更新された信念

- $(\hat{\boldsymbol{\theta}}_{t,0}^A, \hat{\Sigma}_{t,0}^A)$：ODレベルの事前知識から導出される$(\boldsymbol{\theta}_t^A, \Sigma_t^A)$の初期信念
- $(\hat{\boldsymbol{\theta}}_{t,m}^A, \hat{\Sigma}_{t,m}^A)$：各試行$m$で更新された信念 ← 観測された需要$\hat{\boldsymbol{x}}_{ij}$

3-3. Learning scheme design

3.3.2. Evaluation and choice

MABにおいて

a^tを選ぶことの価値　$$\nu_{a^t}^{m,MAB} = \hat{X}_{a^t} + \sqrt{\frac{2\log\kappa}{n_{a^t}}}$$

κ：タイムステップの数
n_{a^t}：a^tが選ばれた回数

3-3. Learning scheme design

3.3.2. Evaluation and choice

KGにおいて

a^tを選ぶことの価値　$\nu_{a^t}^{m,KG} = \tilde{\sigma}_{a^t}^{m} f(\zeta_{a^t}^{m})$

$\tilde{\sigma}_{a^t}^{m}$：$\hat{X}_{a^t}$の推定分散の変化量
$\zeta_{a^t}^{m}$：a^t選択の正規化された影響

直接観測不可能

- $\hat{\sigma}_{a^t}^{2,m}$：$\hat{X}_{a^t}$の推定分散
- $\tilde{\sigma}_W^2$：観測の分散

$$\tilde{\sigma}_{a^t}^{2,m} = \frac{\hat{\sigma}_{a^t}^{2,m}}{1 + \tilde{\sigma}_W^2 / \hat{\sigma}_{a^t}^{2,m}}$$ と計算される

Proof.

Suppose that a precision β is the inverse of σ^2. In Bayesian methods with Gaussian variables, $\beta^{m+1} = \beta^m + \beta^W$. Then,

$$\tilde{\sigma}_{a^t}^{2,m} = \hat{\sigma}_{a^t}^{2,m} - \hat{\sigma}_{a^t}^{2,m+1}$$

$$= \hat{\sigma}_{a^t}^{2,m} - \left(\frac{1}{\hat{\sigma}_{a^t}^{2,m}} + \frac{1}{\hat{\sigma}_{a^t}^{W}} \right)^{-1}$$

$$= \hat{\sigma}_{a^t}^{2,m} - \frac{\hat{\sigma}_{a^t}^{2,m} \hat{\sigma}_{a^t}^{W}}{\hat{\sigma}_{a^t}^{2,m} + \hat{\sigma}_{a^t}^{W}}$$

$$= \frac{\hat{\sigma}_{a^t}^{4,m}}{\hat{\sigma}_{a^t}^{2,m} + \hat{\sigma}_{a^t}^{W}}$$

3-3. Learning scheme design

3.3.2. Evaluation and choice

KGにおいて

a^tを選ぶことの価値 $\nu_{a^t}^{m,KG} = \tilde{\sigma}_{a^t}^{m} f(\zeta_{a^t}^{m})$

$\tilde{\sigma}_{a^t}^{m}$：$\hat{X}_{a^t}$の推定分散の変化量
$\zeta_{a^t}^{m}$：a^t選択の正規化された影響

$$\zeta_{a^t}^{m} = -\left|\frac{\hat{X}_{a^t} - \max_{a^{t\prime} \neq a^t} \hat{X}_{a^t}}{\tilde{\sigma}_{a^t}^{m}}\right|$$

$$f(w) = w\Phi(w) + \phi(w)$$

$\phi(w) = \int_{-\infty}^{w} \phi(s)ds$：標準正規分布の確率密度関数 (PDF)

$\Phi(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}}$：標準正規分布の累積分布関数 (CDF)

3-3. Learning scheme design

3.3.2. Evaluation and choice

KGCBにおいて

全ての選択肢の価値信念　$\hat{\boldsymbol{\theta}}^{A}_{t,m+1} = \hat{\boldsymbol{\theta}}^{A}_{t,m} + \tilde{\boldsymbol{\sigma}}(a^{t,m})Z^{m+1}$

$$\tilde{\boldsymbol{\sigma}}(a^{t,m}) = \frac{\hat{\Sigma}^{A}_{t,m}\boldsymbol{e}_{a^{t,m}}}{\sqrt{\hat{\Sigma}^{A}_{t,m}(a^{t,m},a^{t,m}) + \hat{\sigma}^{2}_{W}}}$$

$\hat{\Sigma}^{A}_{t,m}$：相関行列が影響

Proof.

Suppose that $\boldsymbol{B}^m$ is the precision matrix equal to $\left(\hat{\Sigma}^m\right)^{-1}$. For simplicity, $\hat{\Sigma}^m = \hat{\Sigma}^A_{t,m}$ and $x^m = a^{t,m}$. In Bayesian methods with correlated Gaussian variables, $\boldsymbol{B}^{m+1} = \left(\boldsymbol{B}^m + \beta^W e_{x^m}(e_{x^m})^T\right)$, similar to Proposition 1, where e_{x^m} is a unit column vector. Meanwhile, the Sherman-Morrison formula in Eq. (18) can be used to update $\boldsymbol{B}^n$ since a matrix P is invertible (Powell and Ryzhov, 2012b).

$$\left[P + uu^T\right]^{-1} = P^{-1} - \frac{P^{-1}uu^TP^{-1}}{1+u^TP^{-1}u} \tag{18}$$

Through some modification, $\frac{1}{\beta^W}\boldsymbol{B}^{m+1} = \left(\frac{1}{\beta^W}\boldsymbol{B}^m + e_{x^m}(e_{x^m})^T\right)$. Then, Eq. (17) is converted as follows:

$$\left(\frac{1}{\beta^W}\boldsymbol{B}^{m+1}\right)^{-1} = \left(\frac{1}{\beta^W}\boldsymbol{B}^m\right)^{-1} - \frac{\left(\frac{1}{\beta^W}\boldsymbol{B}^m\right)^{-1}e_{x^m}(e_{x^m})^T\left(\frac{1}{\beta^W}\boldsymbol{B}^m\right)^{-1}}{1+(e_{x^m})^T\left(\frac{1}{\beta^W}\boldsymbol{B}^m\right)^{-1}e_{x^m}}$$

$$\beta^W\hat{\Sigma}^{m+1} = \beta^W\hat{\Sigma}^m - \frac{\beta^W\hat{\Sigma}^m e_{x^m}(e_{x^m})^T\beta^W\hat{\Sigma}^m}{1+(e_{x^m})^T\beta^W\hat{\Sigma}^m e_{x^m}}$$

$$\hat{\Sigma}^{m+1} = \hat{\Sigma}^m - \frac{\Sigma^n e_{x^m}(e_{x^m})^T\hat{\Sigma}^m}{1/\beta^W+(e_{x^m})^T\hat{\Sigma}^m e_{x^n}}$$

$$\hat{\Sigma}^m - \hat{\Sigma}^{m+1} = \frac{\hat{\Sigma}^m e_{x^m}(e_{x^m})^T\hat{\Sigma}^m}{\hat{\sigma}^2_W+(e_{x^m})^T\hat{\Sigma}^m e_{x^m}} = \tilde{\Sigma}^m$$

Therefore, $\tilde{\boldsymbol{\sigma}}(x^m)$, a square root of $\tilde{\Sigma}(x^m)$, can be expressed as follows:

$$\tilde{\boldsymbol{\sigma}}(x^m) = \frac{\hat{\Sigma}^m\boldsymbol{e}_{x^m}}{\sqrt{\hat{\Sigma}^m(x^m,x^m)+\hat{\sigma}^2_W}}$$

3-3. Learning scheme design

3.3.2. Evaluation and choice

KGCBにおいて

全ての選択肢の価値信念　$\widehat{\boldsymbol{\theta}}^{A}_{t,m+1} = \widehat{\boldsymbol{\theta}}^{A}_{t,m} + \tilde{\boldsymbol{\sigma}}(a^{t,m})Z^{m+1}$

$\tilde{\boldsymbol{\sigma}}(a^{t,m})$：$a^{t,m}$を選択時の更新された$\hat{\Sigma}^{A}_{t,m}$の変化
Z^{m+1}：観測と信念の違いに関するランダム変数

$$Z^{m+1} = \frac{\boldsymbol{W}^{m+1}_{a^{t,m}} - \widehat{\boldsymbol{\theta}}^{A}_{t,m}(a^{t,m})}{\sqrt{\mathrm{Var}^{m}(\boldsymbol{W}^{m+1}_{a^{t,m}} - \widehat{\boldsymbol{\theta}}^{A}_{t,m}(a^{t,m}))}}$$

$\boldsymbol{W}^{m+1}_{a^{t,m}}$：観測されたフロー

$\widehat{\boldsymbol{\theta}}^{A}_{t,m}(a^{t,m})$：$a^{t,m}$を試行するときの信念

$$\mathrm{Var}^{m}\left(\boldsymbol{W}^{m+1}_{a^{t,m}} - \widehat{\boldsymbol{\theta}}^{A}_{t,m}(a^{t,m})\right) = \mathrm{Var}^{m}\left(\boldsymbol{W}^{m+1}_{a^{t,m}}\right)$$

$$= \mathrm{Var}^{m}\left(\widehat{\boldsymbol{\theta}}^{A}_{t,m}(a^{t,m}) + \varepsilon^{m+1}\right)$$

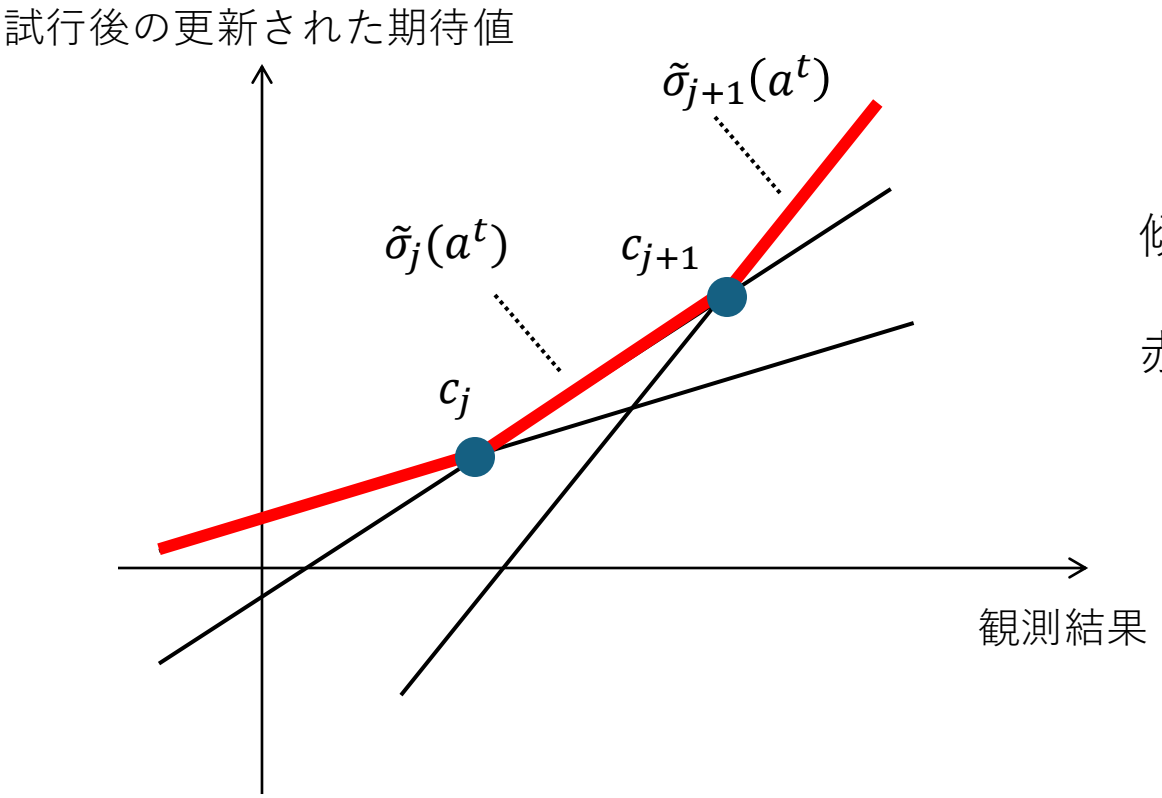
$$= \hat{\Sigma}^{A}_{t,m}(a^{t,m}, a^{t,m}) + \hat{\sigma}^{2}_{W}$$

3-3. Learning scheme design

3.3.2. Evaluation and choice

KGCBにおいて

a^tを選ぶことの価値　$$\nu_{a^t}^{m,KGCB} = \sum_{j=1}^{J} (\tilde{\sigma}_{j+1}(a^t) - \tilde{\sigma}_j(a^t)) f(-|c_j|)$$



傾きが急：試行で評価が大きく変わる、関連性の高い選択肢

赤線を選ぶための式

3-3. Learning scheme design

3.3.3. Observation and knowledge update

相関のないフロー（ベイズ事前分布更新手順を用いる）

$$\boldsymbol{\theta}^{n+1} = \frac{\boldsymbol{\beta}^n \boldsymbol{\theta}^n + \boldsymbol{\beta}^W \boldsymbol{W}^{n+1}}{\boldsymbol{\beta}^n + \boldsymbol{\beta}^W}$$

$$\boldsymbol{\beta}^{n+1} = \boldsymbol{\beta}^n + \boldsymbol{\beta}^W$$

$\boldsymbol{\beta}^n$：需要の正確性(分散の逆数)についての現在の信念
$\boldsymbol{\theta}^n$：需要の平均についての信念
$\boldsymbol{\beta}^W$：観測の正確性(分散の逆数)
$\boldsymbol{W}^{n+1}$：観測されたフロー

相関のあるフロー（ベイズ事前分布更新手順を用いる）

$$\Omega = Diag(\omega/\sigma_\varepsilon^2)$$

$$\boldsymbol{\theta}^{n+1} = (\Omega + \Sigma^{n-1})^{-1}(\boldsymbol{W}^{n+1^T}\Omega + \boldsymbol{\theta}^{n^T}\Sigma^{n-1})^T$$

$$\Sigma^{n+1} = (\Omega + \Sigma^{n-1})^{-1}$$

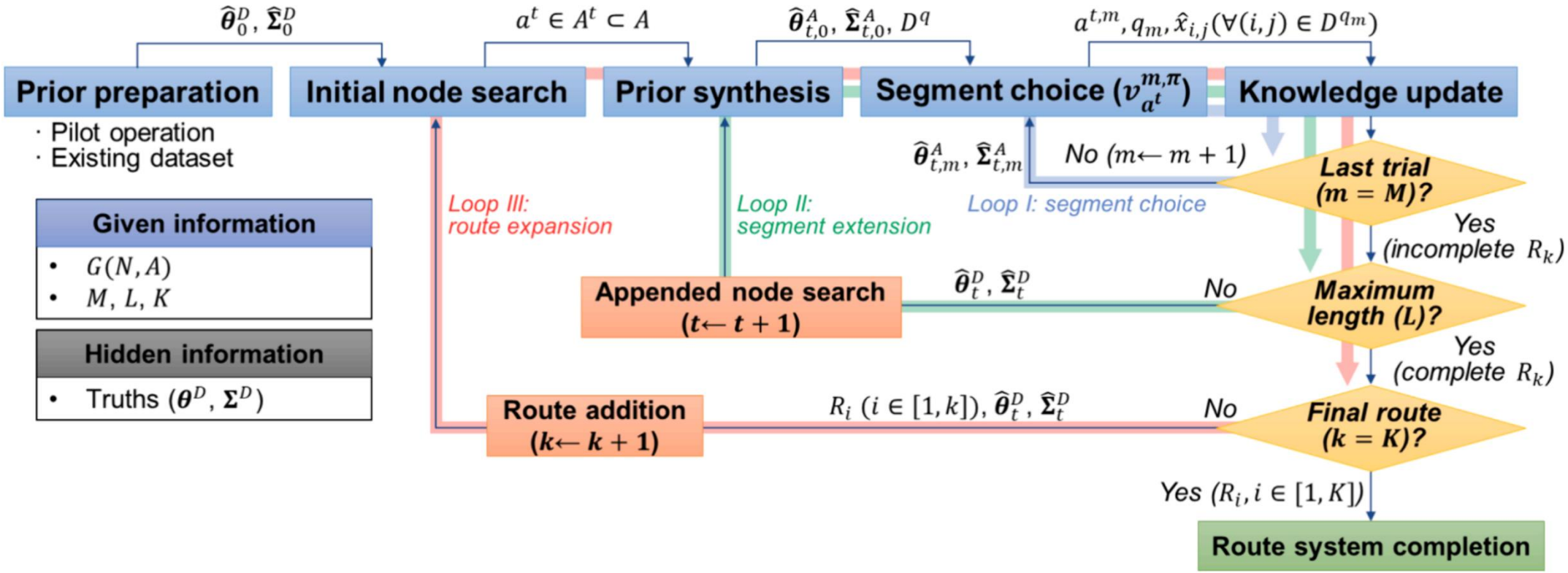
Ω：フローが観測されたかどうかを示す行列

3-4. Proposed AI-based sequential segment-level transit network design algorit

ルートデザインの最適学習との違い

- 同じ選択肢集合で評価は繰り返されない（∵拡張される）
- 観測と操作の数が限られる
- 報酬推定のために拡張ごとにODフロー情報の集約をする必要性

提案アルゴリズム



3-4. Proposed AI-based sequential segment-level transit network design algorit

提案アルゴリズム

1. Prepare network $G(N, A)$ and node pair flows .$\{x_{i,j}|i, j \in N\}$

2. Input simulation settings: number of routes required K, maximum route length L, number of pilots P, minimum pilot route length L_P, number of observations per pilot Q_p, number of observations per extension M, learning policy π, time step $t(=0)$.

2.1 If available, import priors from existing knowledge ($\widehat{\boldsymbol{\theta}}_0^D$, $\widehat{\Sigma}_0^D$).

3 Conduct pilots.

For $p = 1$ to P **do**.

3.1. Randomly choose two nodes (i, j) as terminals.

3.2. Operate a route between (i, j), longer than L_P.

3.3. Observe Q_p operations on chosen route.

3.4. From observed $x_{i,j}$, create $\widehat{\boldsymbol{\theta}}_0^D$ and $\widehat{\Sigma}_0^D$ if not existent. Otherwise, update $\widehat{\boldsymbol{\theta}}_t^D$ and $\widehat{\Sigma}_t^D$. $t = t + 1$.

End For.

4. Implement route expansion based on segment-level extension

For $k = 1$ to K **do**.

4.1 Designate an initial node as a starting terminal of R_k.

While $|R_k| < L$ **do**.

4.2 Identify available segments $a^t \in A^t$ from both ends of R_k.

For $q = 1$ to $\left|A^t\right|$ **do**.

4.2.1 Identify $x_{i,j}$ covered by the route system after $a^{t(q)}$ appended and create D^q.

End For.

4.3 Aggregate $\widehat{\boldsymbol{\theta}}_t^D$ and $\widehat{\Sigma}_t^D$ of $x_{i,j}$ to segment-level ($\widehat{\boldsymbol{\theta}}_{t,0}^A$, $\widehat{\Sigma}_{t,0}^A$) using transit assignment. Set $m = 0$, $\widehat{\boldsymbol{\theta}}_{t,m}^A = \widehat{\boldsymbol{\theta}}_{t,0}^A$, and $\widehat{\Sigma}_{t,m}^A = \widehat{\Sigma}_{t,0}^A$.

For $m = 1$ to M **do**.

4.3.1. Estimate $v_{a^t}^{m-1,\pi}$ of segments using $\widehat{\boldsymbol{\theta}}_{t,m-1}^A$ and $\widehat{\Sigma}_{t,m-1}^A$ based on π.

4.3.2. Determine $a^{t,m}$ to observe that satisfies: $a^{t,m} = \underset{a^t \in A^t}{\operatorname{argmax}}\, v_{a^t}^{m-1,\pi}$.

4.3.3. Observe $x_{i,j}$ corresponding to $a^{t,m}$.

4.3.4. $\widehat{\boldsymbol{\theta}}_{t,m}^A = \widehat{\boldsymbol{\theta}}_{t,m-1}^A$ and $\widehat{\Sigma}_{t,m}^A = \widehat{\Sigma}_{t,m-1}^A$. Update $\widehat{\boldsymbol{\theta}}_{t,m}^A$ and $\widehat{\Sigma}_{t,m}^A$ using observations.

End For.

4.4 Append $a^{t,M}$ to R_k according to Eq. (6). $t = t + 1$.

End While.

End For.

3-4. Proposed AI-based sequential segment-level transit network design algorit

注記事項

- アルゴリズムの複雑性は
  - ルート数K、ルート長さL、試行回数Mにより決定
  - ネットワークの大きさには強く影響しない
  - 運用の時間枠もスケーラビリティに関わる
- 方策ごとの計算時間は
  - KGCBが大きい
  - 高頻度でルートを変更することは非現実的なため問題ではない

1. Introduction

2. Literature review

3. Proposed Methodology

4.Numerical experiments

5. Conclusions

4-1. Simple grid networks

4.1.1. Problem and network illustrations

- 5×5のグリッドネットワーク
- 下図から真実の情報と観測された情報の間に知識のギャップがあることがわかる

計画

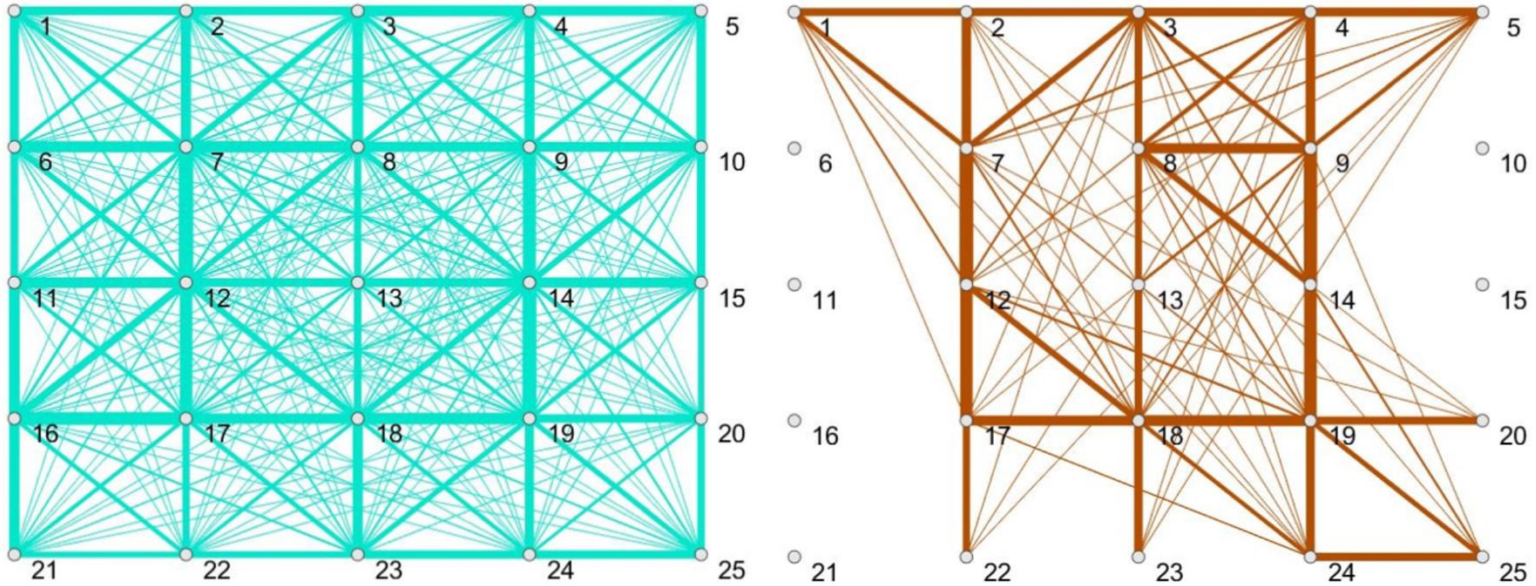
(1) 最終形は4ノードの長さのルート3本

(2) 5つのパイロットサービスを10期間で実施

(3) 198のノードペアが相関あり

(4) 乗り換えは1度のみ

(5) $\sigma_{i,j}$は$x_{i,j}$の10~30%



(a) True flow pattern　(b) Flows observed by pilot service

4-1. Simple grid networks

4.1.2. Experiment inputs

真値

- $x_{i,j}$は以下の数に比例
  - ノードiのトリップ生産定数
  - ノードjのトリップ誘発定数
  - 距離の2乗の逆数
- 共分散行列$\Sigma^D = \boldsymbol{HBH}$（$\boldsymbol{H}$を対角行列、$\boldsymbol{B}$を相関行列）

事前知識

- フローが観測されたとき事前平均と事前標準偏差が計算される

4-1. Simple grid networks

4.1.3. Results

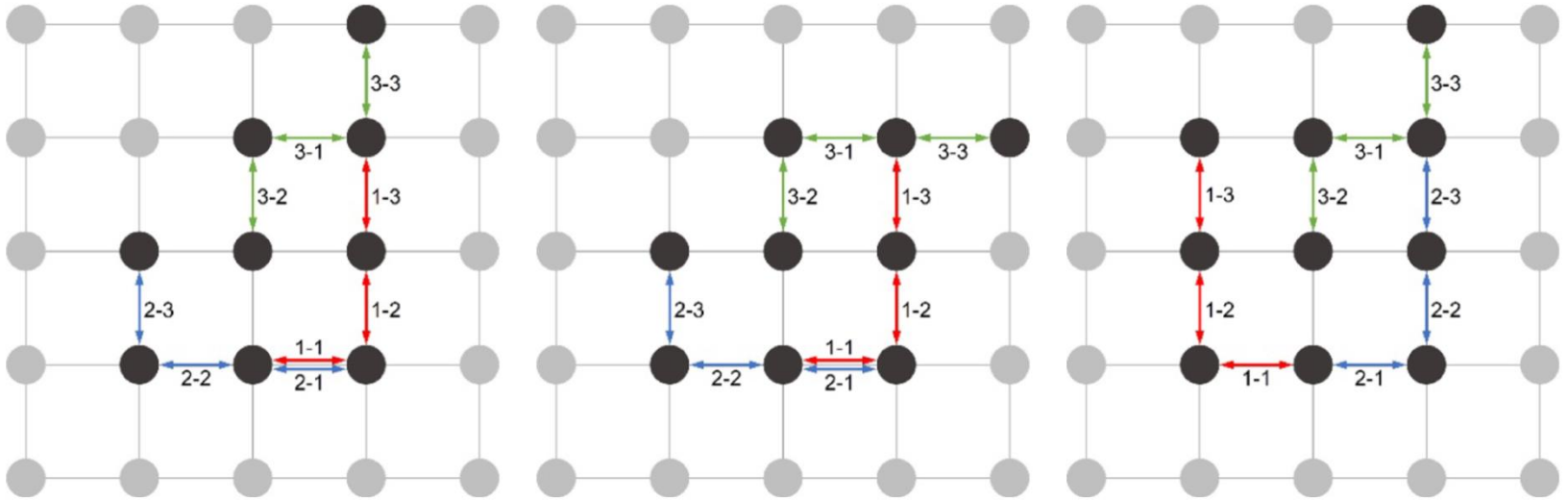
- MABとKGはリンクの重複がある

← **現在のサービス状況に関わらず、残りの未サービス状況を考えて選択**するから

- KGCBもある程度出力は似ている

← 与えられた事前分布 (特に平均値) に依存する方策の評価過程の類似性を示唆

- **KGCBは平均需要カバレッジが高く、標準偏差も低い**（安定して結果が良い）



(a) MAB
(Mean: 22.0%,
Standard Deviation: 2.2%)

(b) KG
(Mean: 22.8%,
Standard Deviation: 2.7%)

(c) KGCB
(Mean: 24.3%,
Standard Deviation: 0.9%)

4-2. NYC PUMA-based network

4.2.1. Problem and network illustrations

- New York City Public Use Microdata Areas (NYC PUMAs)に基づいたネットワーク
- 55のノードと123の双方向リンク

計画

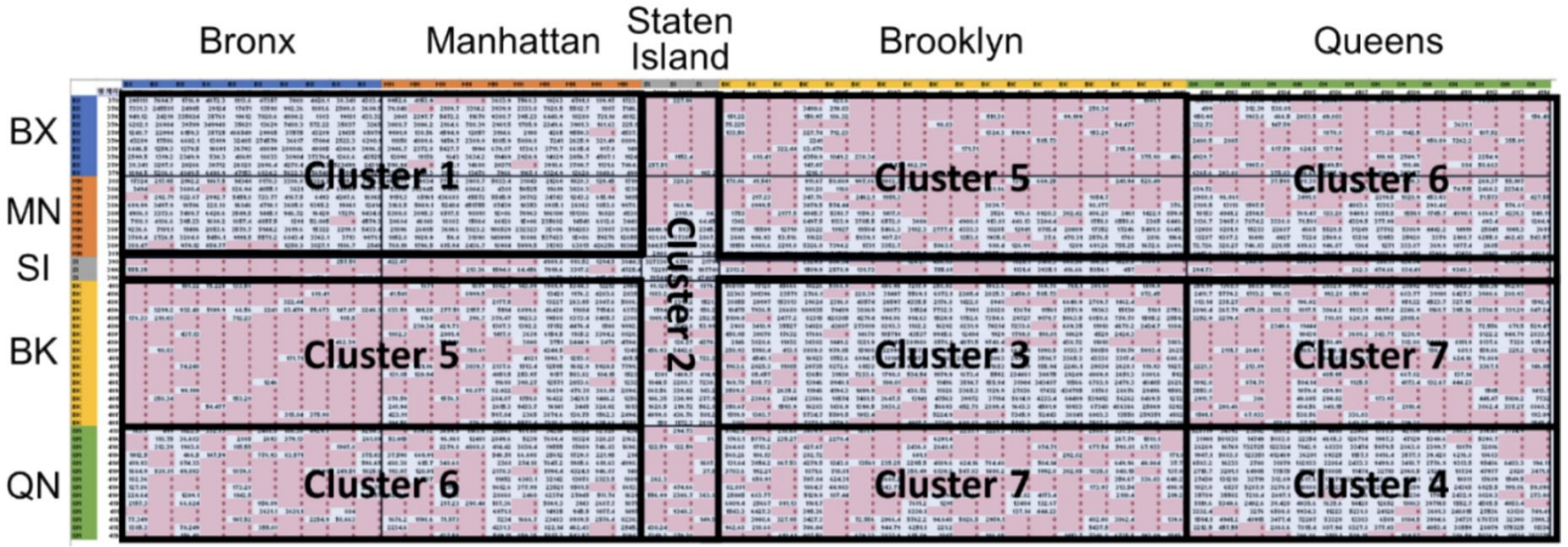
(1) 最終形は8ノードの長さのルート5本
(2) 30のパイロットサービス (最小6ノード) を10期間で実施
(3) 300のノードペアが相関あり
(4) 乗り換えは1度のみ
(5) セグメント延長時は同じコスト

4-2. NYC PUMA-based network

4.2.1. Problem and network illustrations

クラスター

- 5つの行政区がある … Manhattan, Brooklyn, Staten Island, Queens, and Bronx
- 相関が高いODペアをグループ化するクラスターを導入
  - 同じ行政区ペアの場合は相互に相関
  - マンハッタンとブロンクスは、その高い接続性から“megaborough”に統合

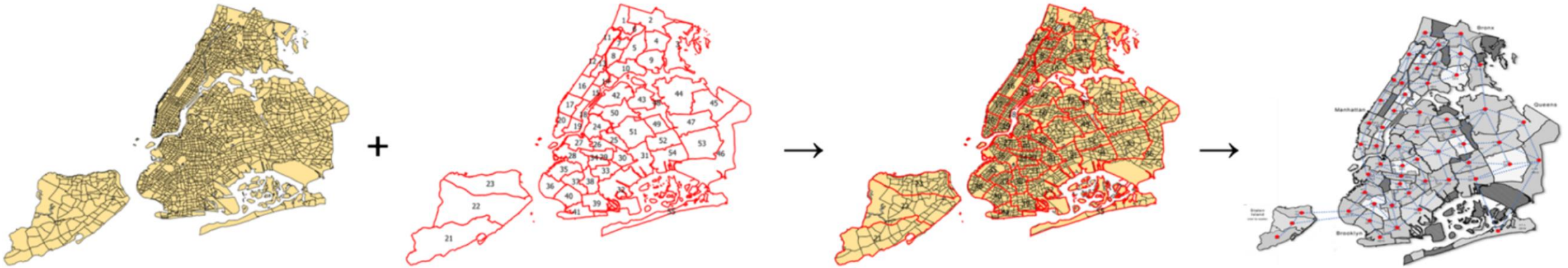


4-2. NYC PUMA-based network

4.2.2. Simulation inputs

- 初期事前平均にはRegional Household Travel Survey (RHTS) を用いる
  - 下図に示すように、調査結果はPUMAレベルに集約
  - RHTSは単一の観測値 → 分散や共分散は導出不可
  - 標準偏差は◯%と設定し、共分散はパイロット運行から導出

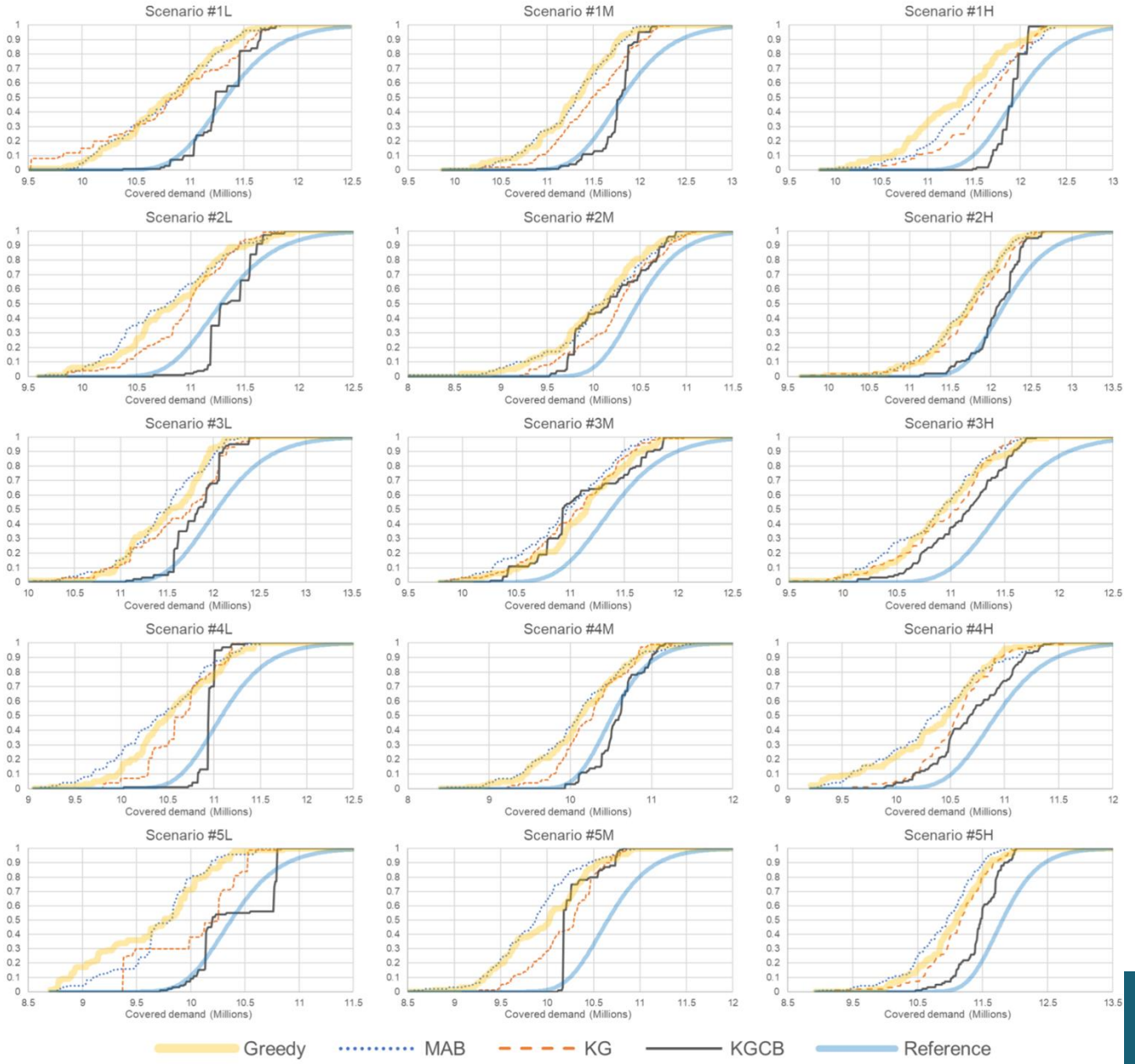
- 時間帯ごとの観測フローの変動の大きさを考える：Low、Middle、High
  - 下限値は同等であるが、上限値は各レベルで異なる



4-2. NYC PUMA-based network

4.2.3. Results

- 15シナリオを考える
  - 3つのフロー変動レベル
  - 5つの需要パターン
- 100回シミュレーションした平均を示す
- グラフは右下にあるほど良い
- 黄線は**探索を考慮しないgreedy方策**
- 青透明線は**CR参照方策**
  - 全ての策をランダムサンプリング
  - 非現実的な視点
  - オンライン学習ではない

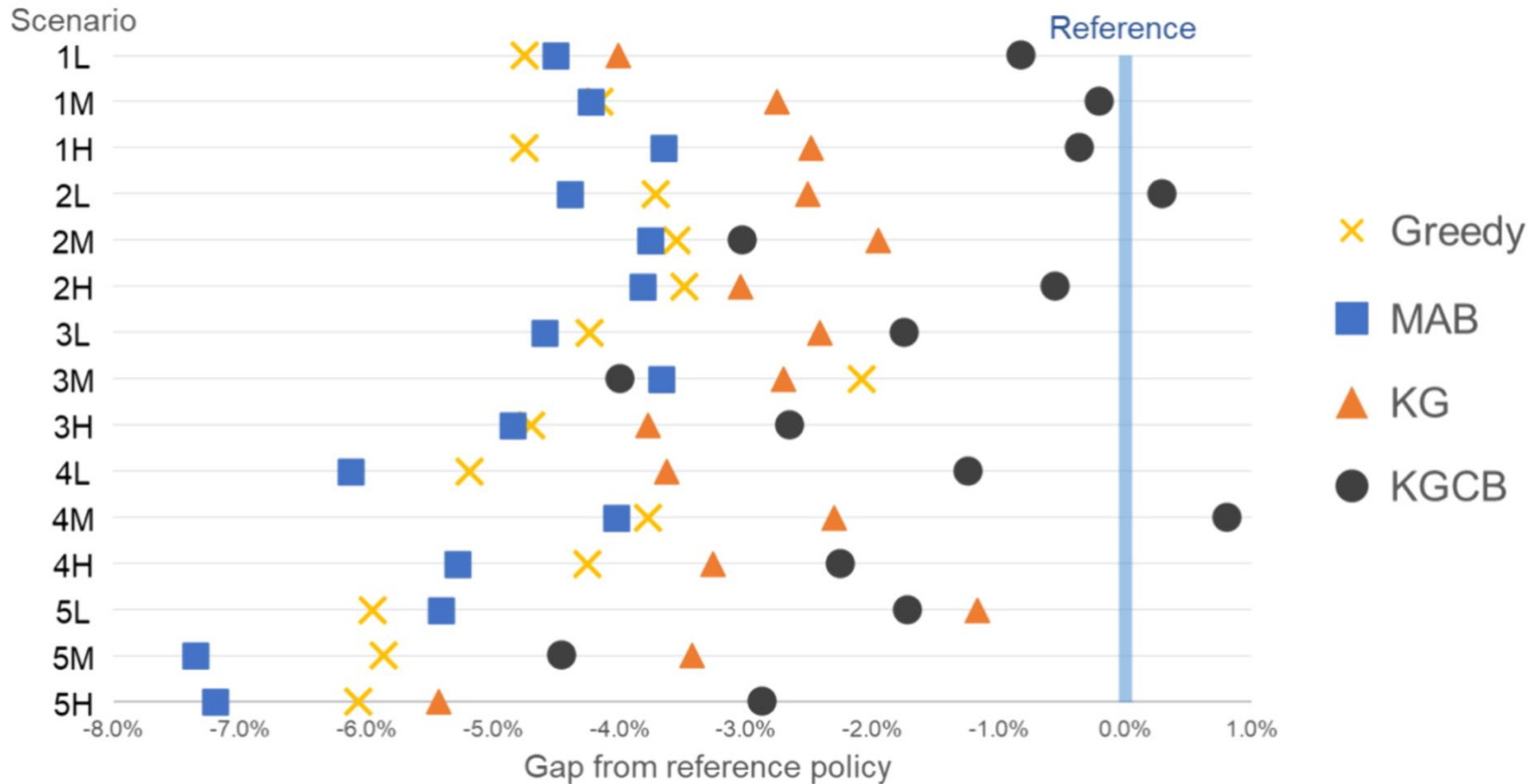


4-2. NYC PUMA-based network

4.2.3. Results

- 方策ごとの50パーセンタイルを示す
  - 一部greedyより劣る例もある
  - **KGCBはCR参照を超えることもある**

→ 需要間の相関を考慮することは価値がある



4-2. NYC PUMA-based network

4.2.3. Results

- t検定を用いたカバレッジの分布の違いの統計的有意性
  - MABとKGはgreedyより劣る場合もある
  - **KGCBはMABに対して15中14シナリオで有意**
  - **KGCBはKGに対して15中13シナリオで有意**

Scenario	Descriptive statistics (mean & std.dev in million)			Two-tail *t*-test result (*p*-value)		
	MAB	KG	KGCB	MAB vs KG	MAB vs KGCB	KG vs KGCB
#1L	10.78 (0.49)	10.80 (0.64)	11.29 (0.26)	*0.854*	***	***
#1M	11.27 (0.45)	11.48 (0.43)	11.76 (0.22)	0.001	***	***
#1H	11.48 (0.55)	11.59 (0.45)	11.90 (0.14)	*0.133*	***	***
#2L	10.78 (0.52)	10.95 (0.42)	11.37 (0.21)	0.001	***	***
#2M	10.05 (0.58)	10.24 (0.46)	10.18 (0.41)	0.011	*0.057*	*0.376*
#2H	11.68 (0.48)	11.75 (0.52)	12.08 (0.31)	*0.314*	***	***
#3L	11.46 (0.46)	11.61 (0.52)	11.83 (0.27)	0.024	***	0.000
#3M	10.94 (0.46)	11.04 (0.44)	11.10 (0.45)	*0.151*	0.016	*0.318*
#3H	10.85 (0.46)	10.93 (0.44)	11.12 (0.38)	*0.209*	***	0.001
#4L	10.41 (0.54)	10.64 (0.35)	10.95 (0.12)	0.000	***	***
#4M	10.10 (0.58)	10.25 (0.41)	10.59 (0.29)	0.028	***	***
#4H	10.35 (0.52)	10.57 (0.34)	10.70 (0.37)	0.001	***	0.007
#5L	9.73 (0.40)	10.03 (0.45)	10.40 (0.35)	***	***	***
#5M	9.85 (0.46)	10.20 (0.37)	10.30 (0.20)	***	***	0.022
#5H	10.87 (0.55)	11.12 (0.50)	11.46 (0.33)	0.001	***	***

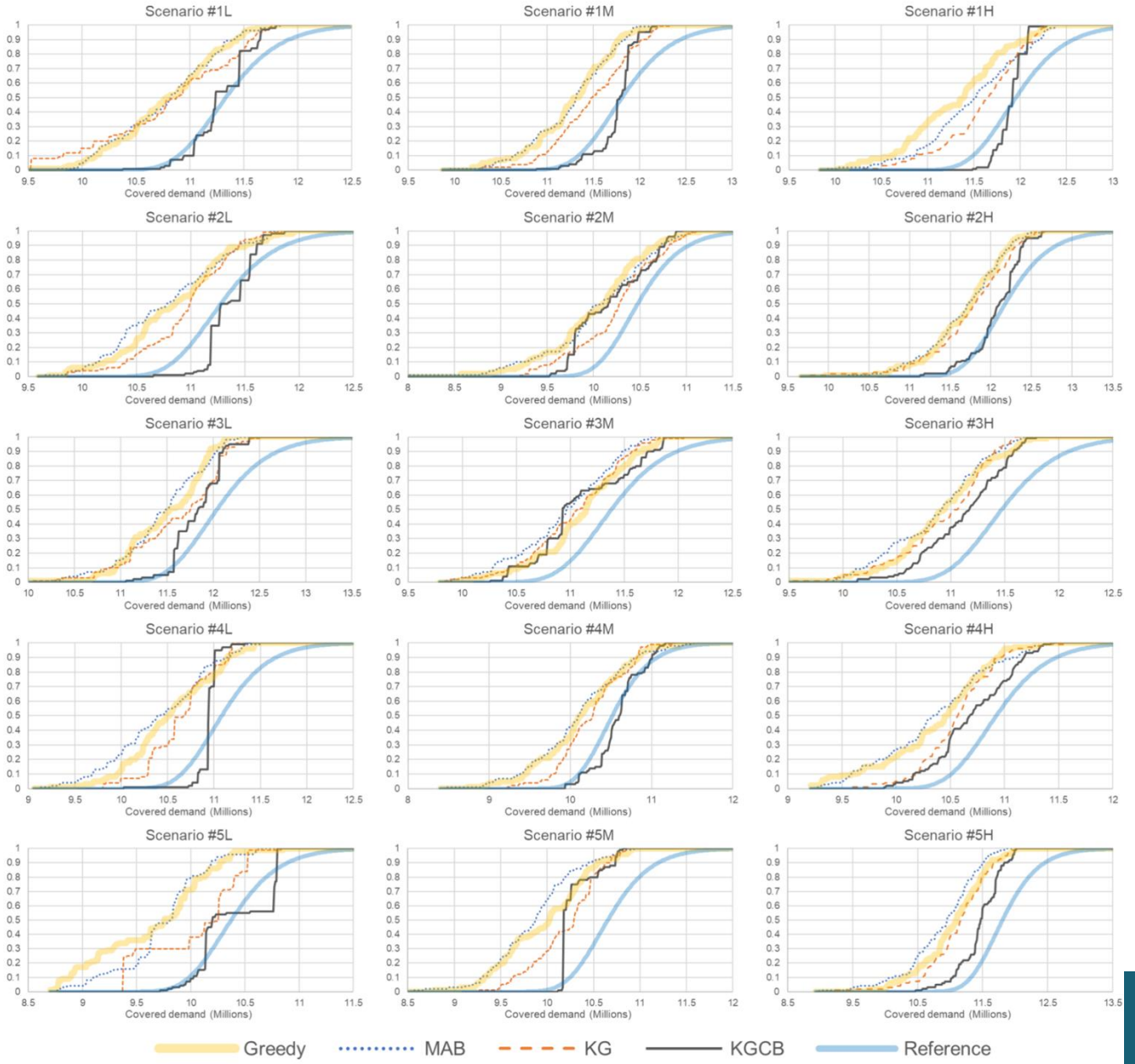
Note: Underlined and italic numbers mean statistically insignificant *p*-values with $\alpha = 0.05$. *** are ones with very strong significance.

4-2. NYC PUMA-based network

4.2.3. Results

- **KGCBは他2つの方策よりも良い**
- フローと需要の影響
  - フロー変動に大きな影響は受けない

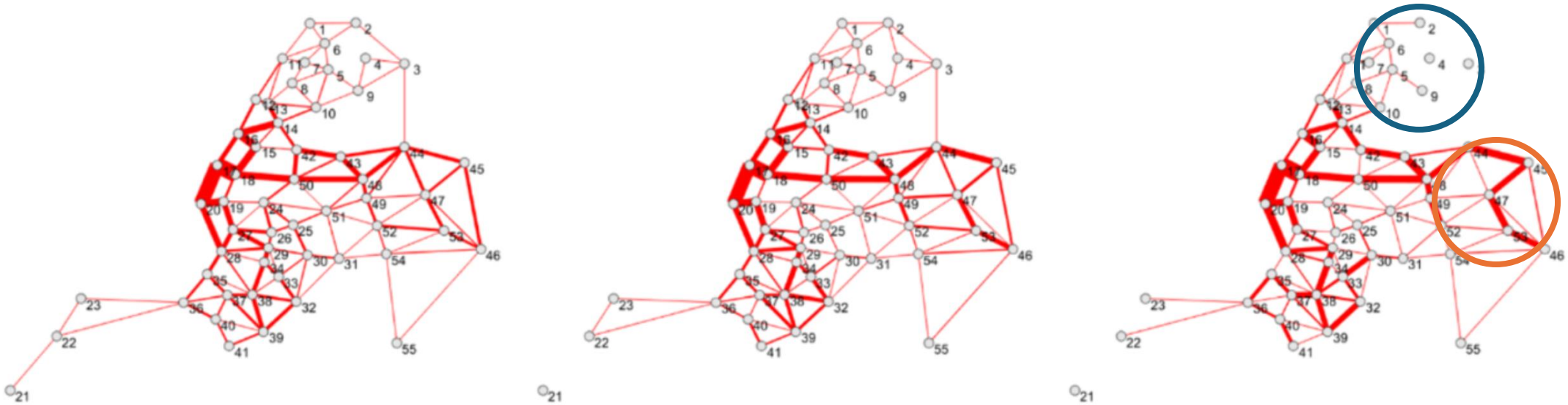
  ← 仮説に反する
  - 需要そのものが結果に影響与える（？）



4-2. NYC PUMA-based network

4.2.3. Results

- 100回のシミュレーションの間に行われた選択の数に比例して重み付けされたリンクを持つネットワーク
  - ほぼ類似
  - KGCBはクイーンズ地区により重点を置き、ブロンクス地区にはあまり重点を置かない傾向
  - MABやKGよりも多くの情報を使用するため、 **KGCBは未知を探索する機会が減少**



MAB (120 links)　　KG (120 links)　　KGCB (107 links)

1. Introduction

2. Literature review

3. Proposed Methodology

4. Numerical experiments

5.Conclusions

5. Conclusions

貢献

- セグメントレベルの拡張を考え、情報を逐次的に得る**SSTNDP問題**を提示
- **最適学習**システム（MAB, KG, KGCB）を導入
- 3つの方策をグリッドネットワークとNYC PUMAを用いた数値実験し、 KGCBが優れていることを確認

本研究が考えきれていない問題

- 制限のない路線間の乗り換え回数
- セグメントごとに異なる延長費用
- 更新頻度に関する制限（オペレータも利用者もシステム適応に時間を要する）
- より現実な配分メカニズム

Future work

- 初期調査やパイロット調査を用いた事前分布のより良い設計
- ゾーンベースのネットワーク構造
- バス性能と交通渋滞の統合
- 時間経過とともに成長・縮小するネットワーク

所感

- ネットワーク設計に馴染みがなさすぎて既往研究は特にへえの連続
- 選択肢相関の嬉しさは結局何だったのか定量的な分析があると良かった（指標導入など）
- 表における変数の多さで疲れるのに後に出てくる全てを網羅しているわけではないのが腹立つ

- VRPとかは研究室でも結構やられているが、世間のイメージする「交通」ぽい研究
  - 1度固定するとルートの変更はなし
  - ハード整備コストを直接的に考えられる余地がある（社基らしい）
- 強化学習と探索行動・回遊のアナロジー
- 観光経路に見立てると…？