

JSAI2021, 2D1-OS-6-03

緯度経度情報追加による 行動履歴分散表現の高精度化

佐賀健志 田中宏季 中村哲

理化学研究所 革新知能統合研究センター 観光情報解析チーム
奈良先端科学技術大学院大学 知能コミュニケーション研究室

2021/6/11

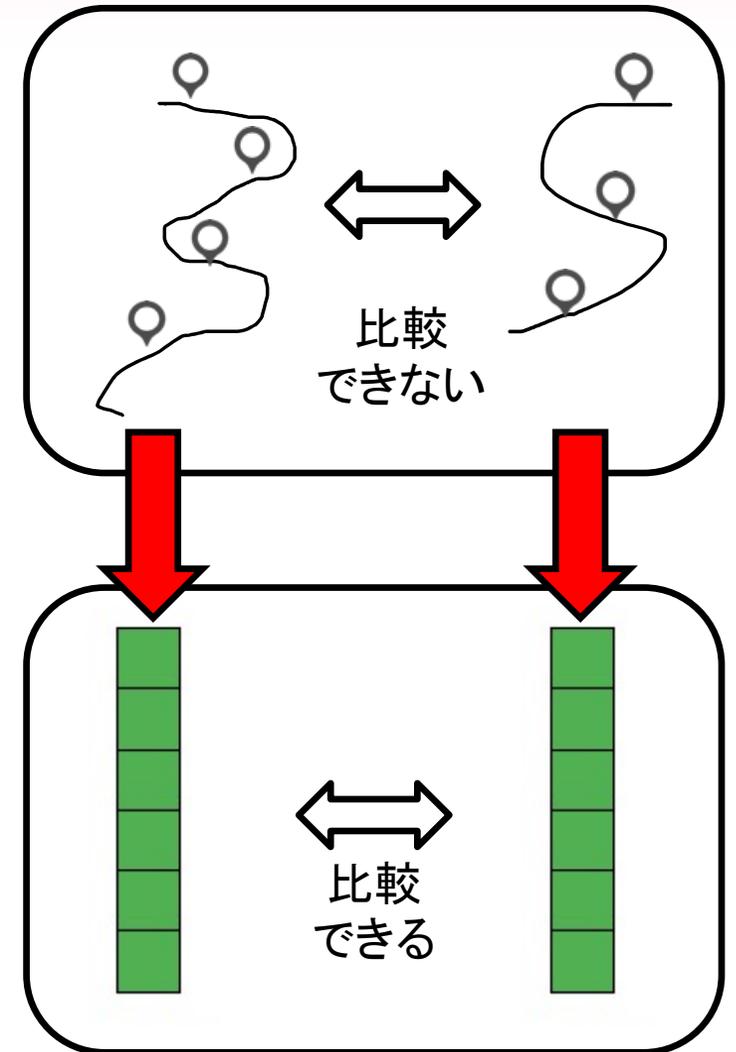
Nakamura SRG

Augmented Human Communication (AHC) Laboratory

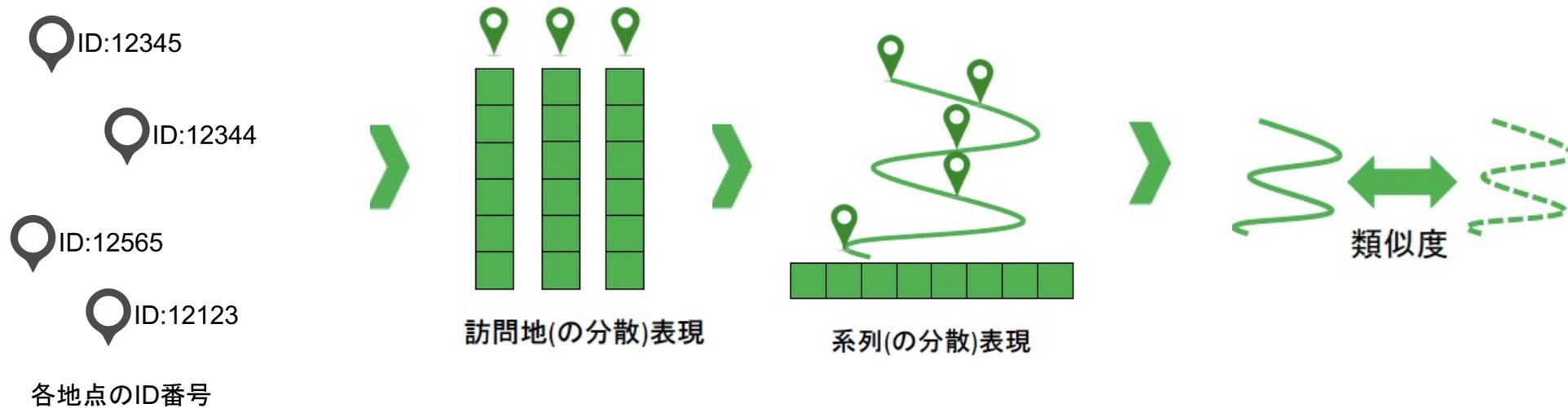
NAIST

- ▶ **移動経路の傾向分析**を行うことで効果的な観光戦略や商品開発等に活用
 - 問題点: データ量が膨大かつ多様で分析が難しい

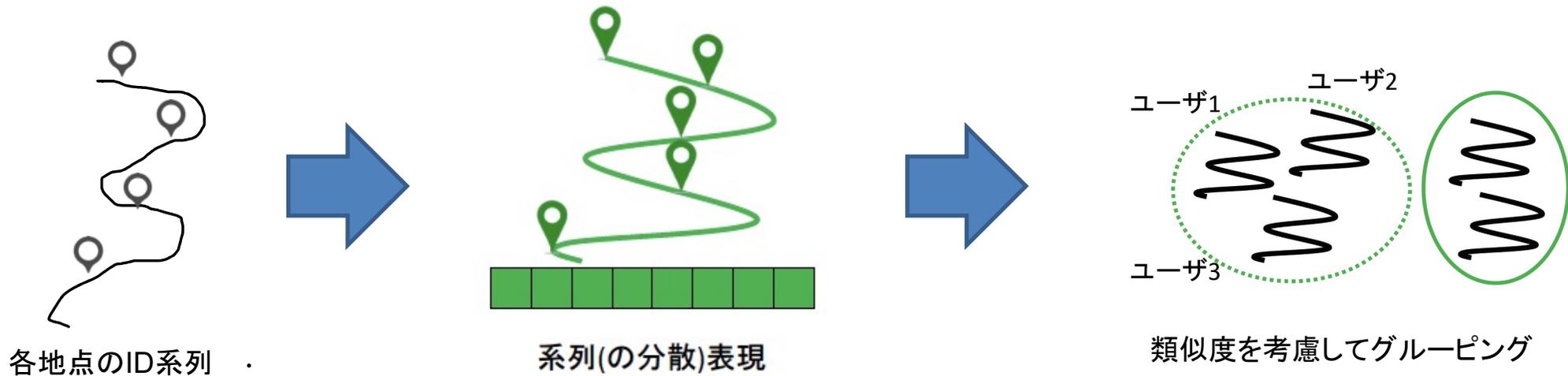
- ▶ 経路の長さや通過箇所がバラバラ
 - 経路の意味を固定長のベクトル化: **分散表現**



- ▶ 分散表現を用いた移動系列の類似度による人流解析 [Crivellari+ 2019]
 - 行動近接性(移動系列の意味的な近さ)を考慮
 - Mesh2Vec (Word2Vecを人の移動系列に適用)



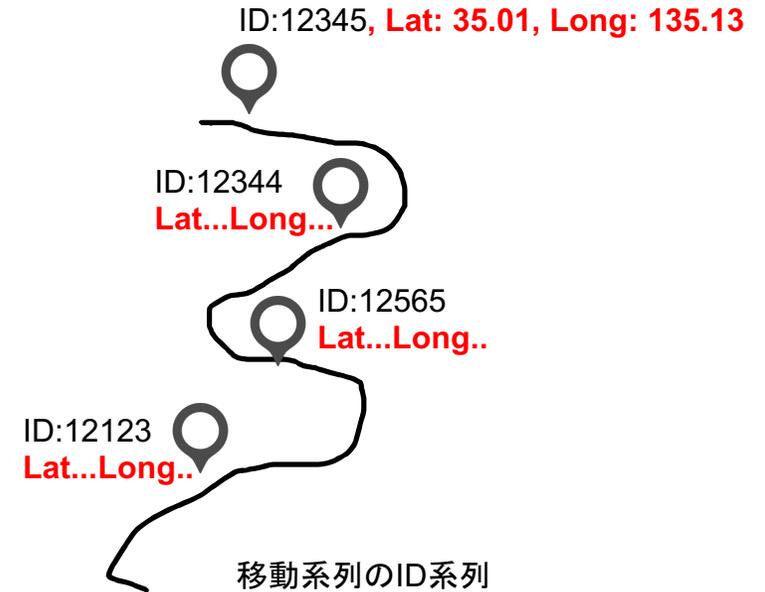
- ▶ BiLSTM分散表現と階層的クラスタリングによる人流解析 [田中ら 2020]
 - BiLSTMを用いることで移動系列のコンテキストを考慮した分散表現を獲得
 - 行動予測ではなく行動分析を目的としているため順路の順方向・逆方向を区別しない
 - 階層的クラスタリングを用いることで類似した系列をグルーピング
 - 系列の分散表現 -> 階層的クラスタリング



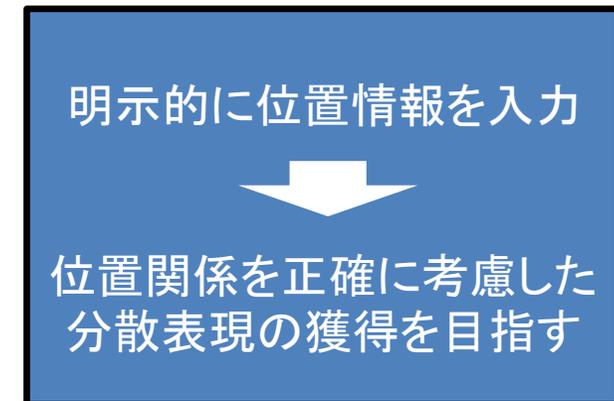
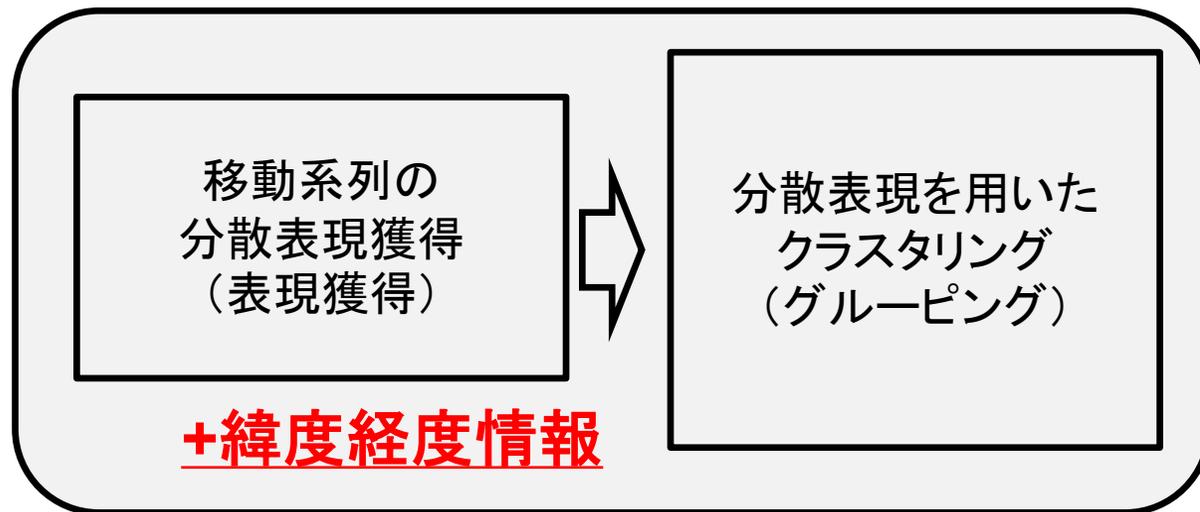
- ▶ 先行研究
 - 入力として離散的なMesh-IDのみを使用
- ▶ 先行研究の課題
 - 各地点の位置(緯度経度)情報が明示的に示されていない

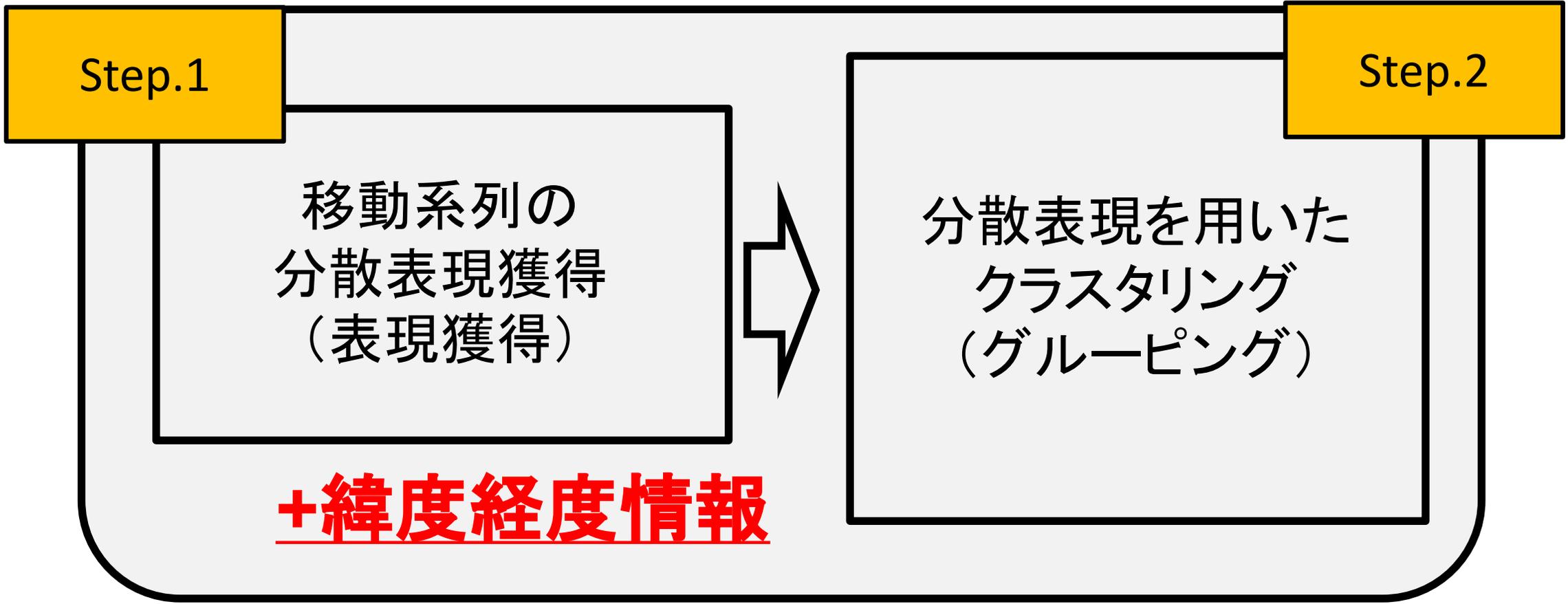


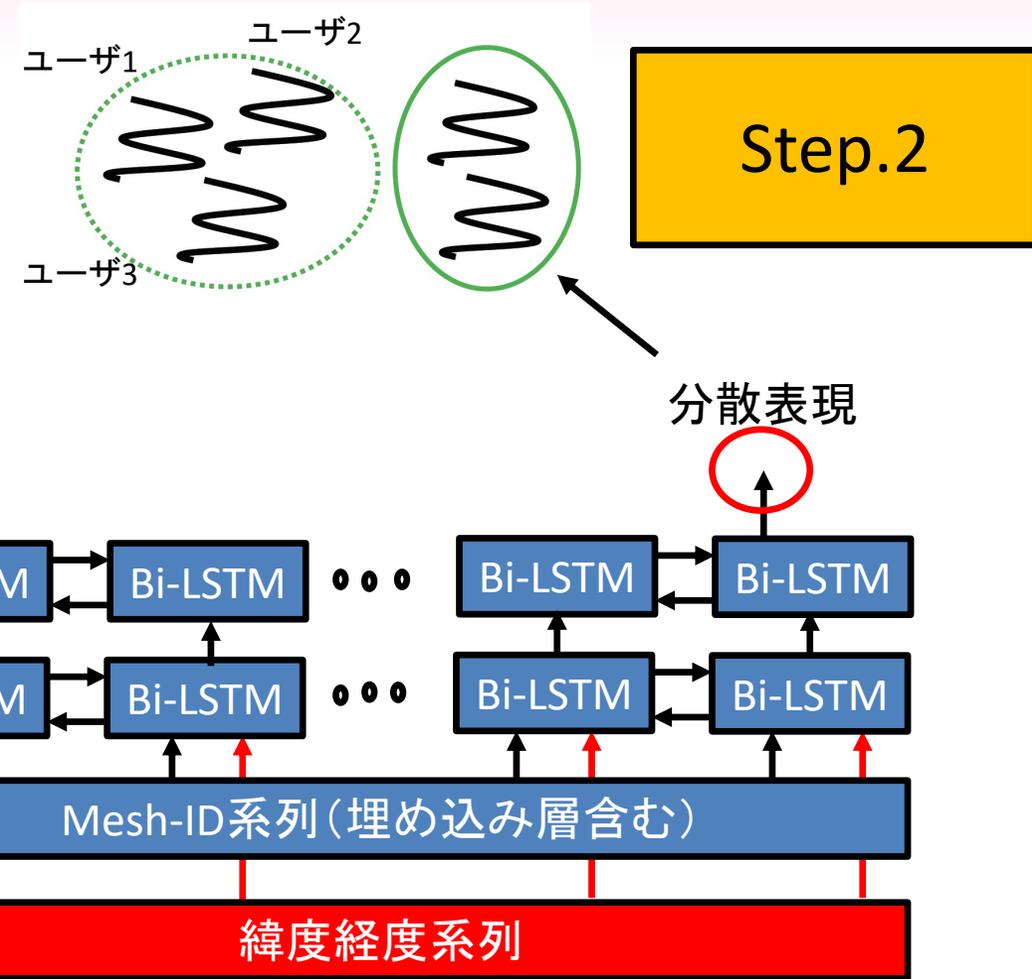
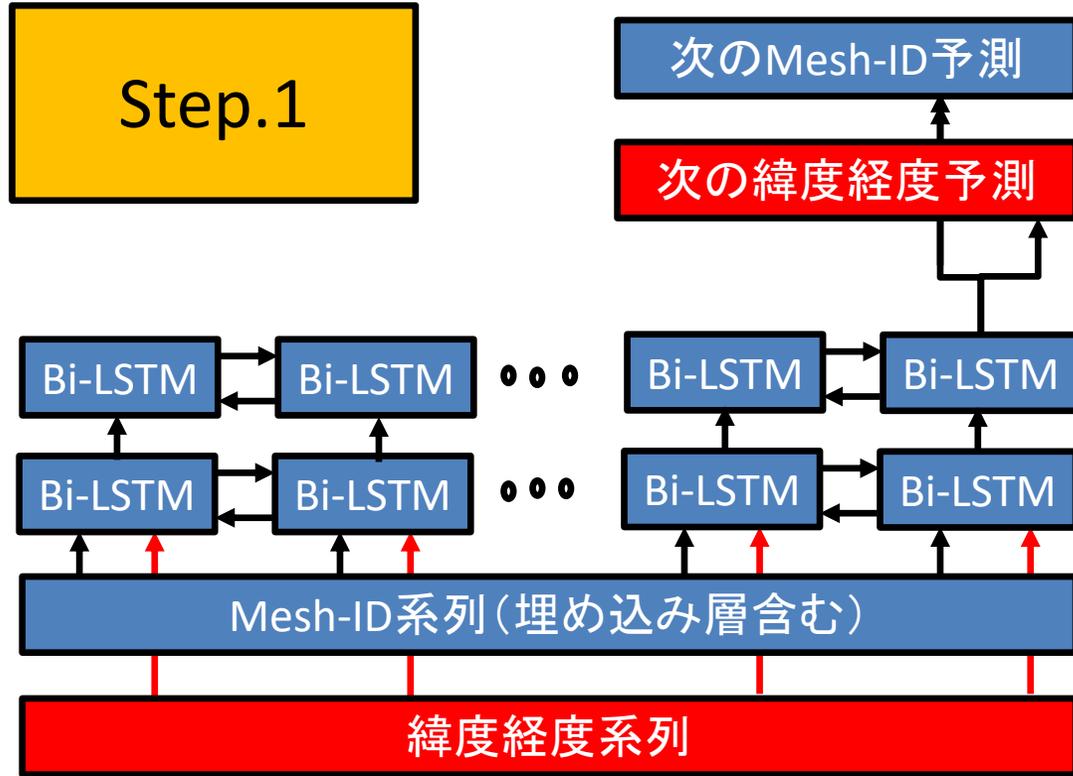
離散的なMesh-IDと連続的な緯度経度情報に基づいた分散表現の獲得
及び階層的クラスタリングを通じた妥当性の検討

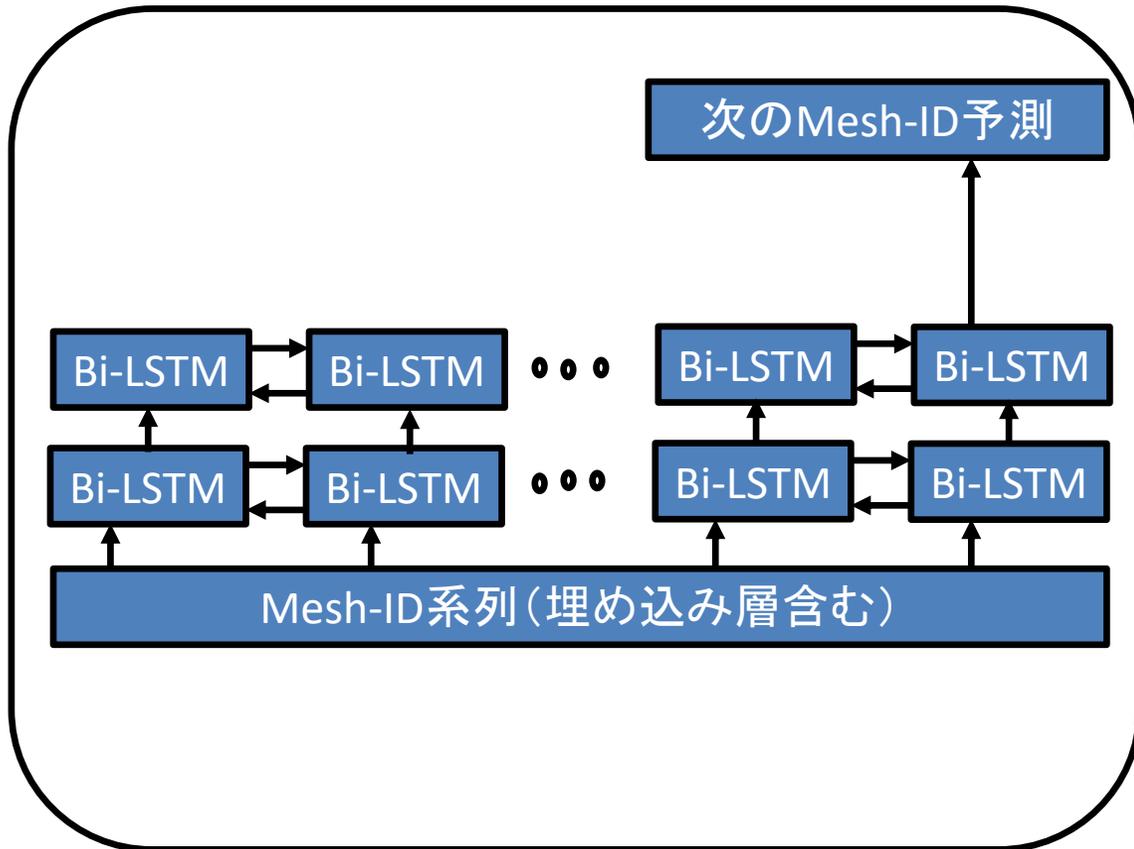


- ▶ 目的: 時系列分散表現に基づく人流クラスタリング
- ▶ 提案:
 - 入力として位置情報を追加したBi-LSTMによる移動系列の分散表現の獲得
 - 階層的クラスタリングを用いた入力への位置情報追加の妥当性検討

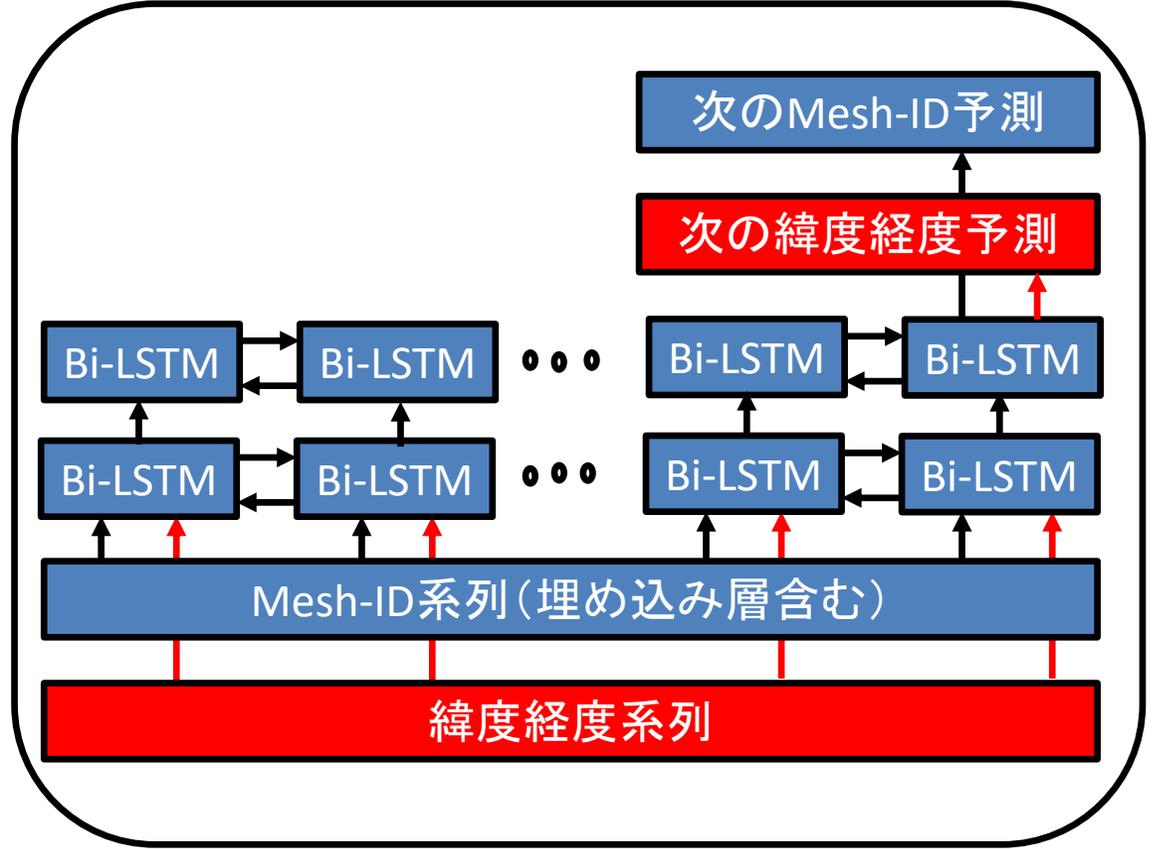




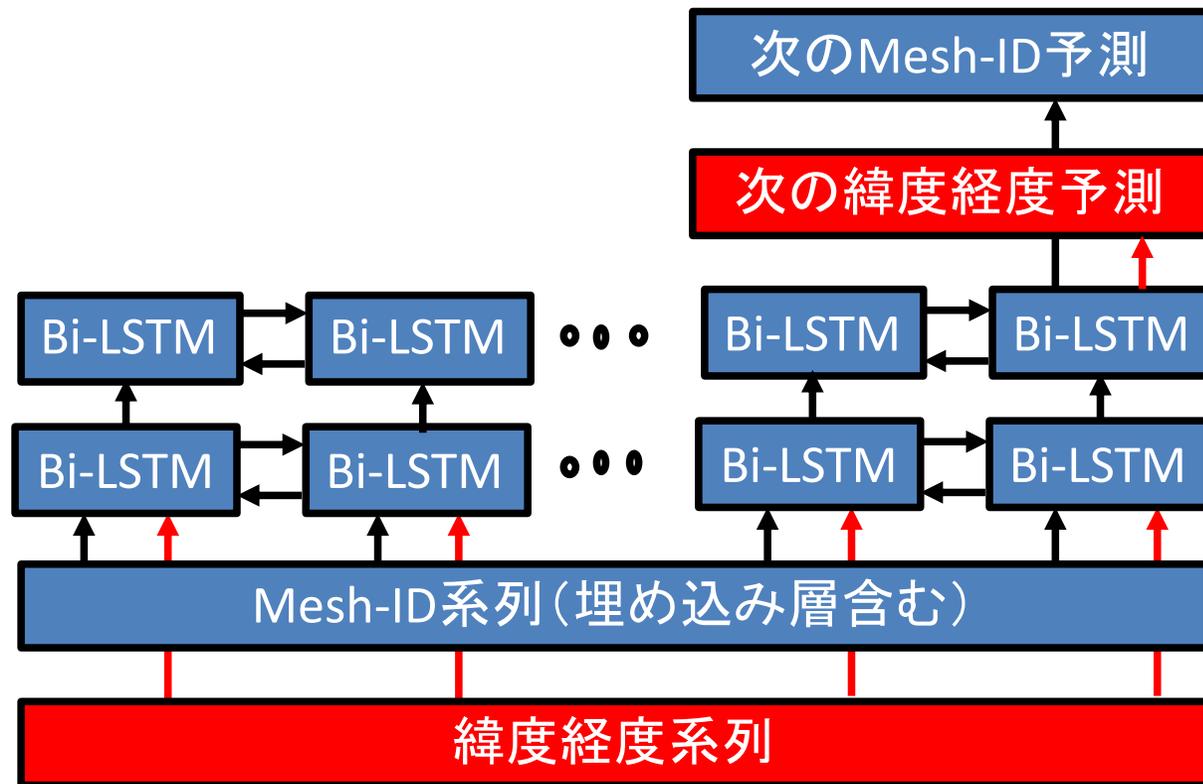




ベースラインモデル



提案モデル



予測対象:

- 最終Mesh-ID
- 最終緯度(Lat)
- 最終経度(Long)

損失関数(最適化対象):

$$L_{total} = \alpha * L_{mesh} + \beta * (L_{lat} + L_{long})$$

L_{mesh} : 交差エントロピー損失

$L_{\{Lat, Long\}}$: 二乗平均誤差

α, β : 定数(1:1)

▶ 階層的クラスタリング

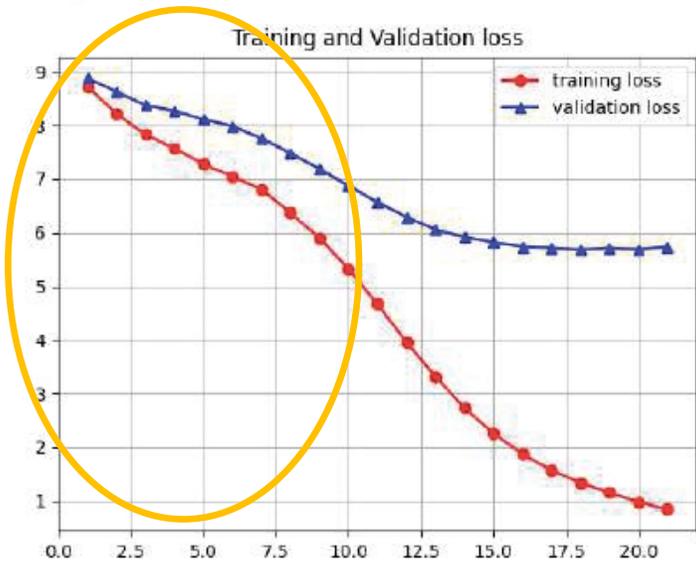
- クラスタ決定法 : ウォード法
- 距離指標 : コサイン距離
- クラスタ距離閾値 : 1.2

▶ データセット (提供: Agoop Corp)

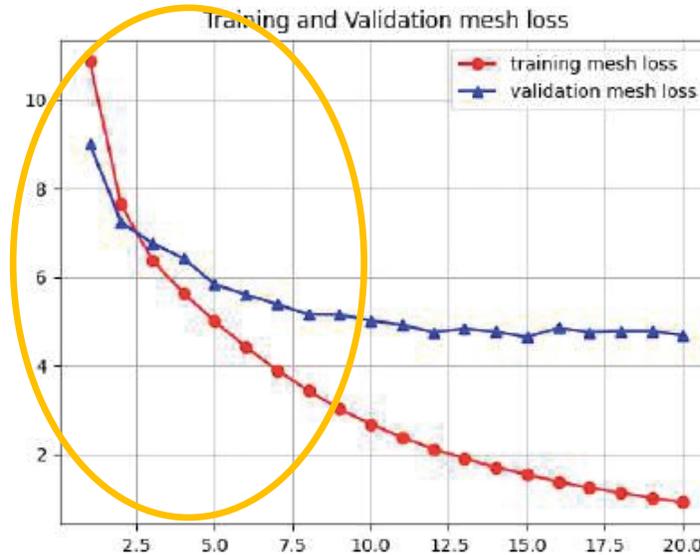
- 対象地域 : 東京都
- メッシュ幅 : 100m
- 総メッシュ数 (面積より算出) : 219,396 (うち **81,897** が対象期間に含まれる)
- 対象期間 : 7-10月, 2020年
- BiLSTM学習対象ユーザ数 : 8,768,203
- クラスタリング対象ユーザ数 : 24

モデル	Accuracy
Mesh-ID のみのモデル	0.32
緯度経度を加えたモデル	<u>0.34</u>

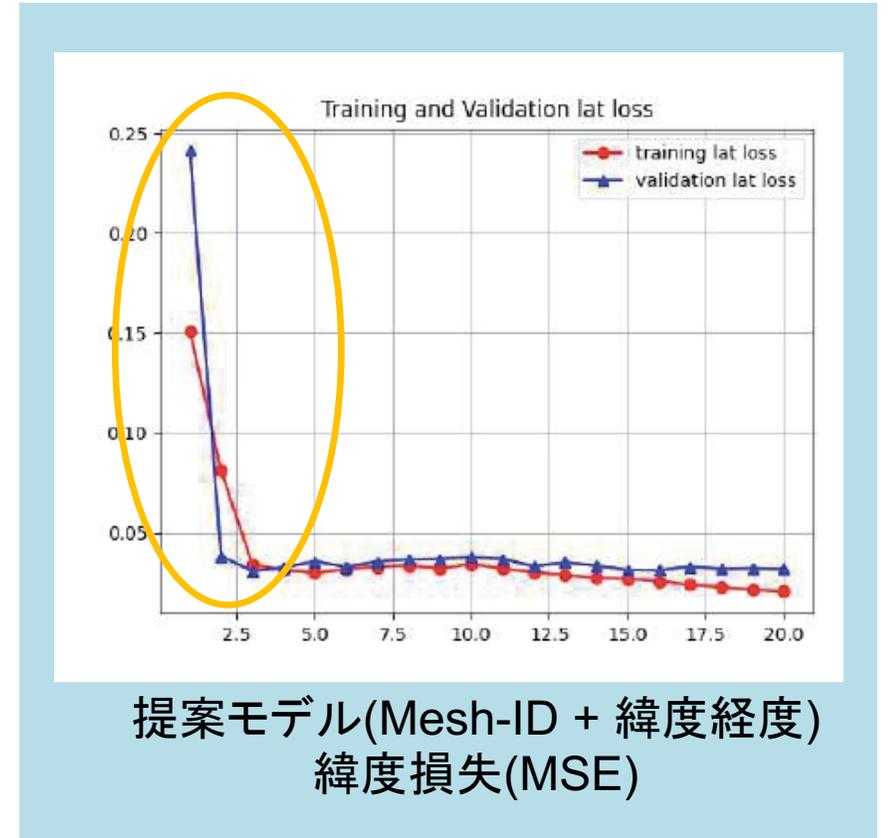
- Top-1 Accuracy
- 予測対象クラス数81,897



ベースラインモデル(Mesh-IDのみ)
Mesh-ID交差エントロピー損失

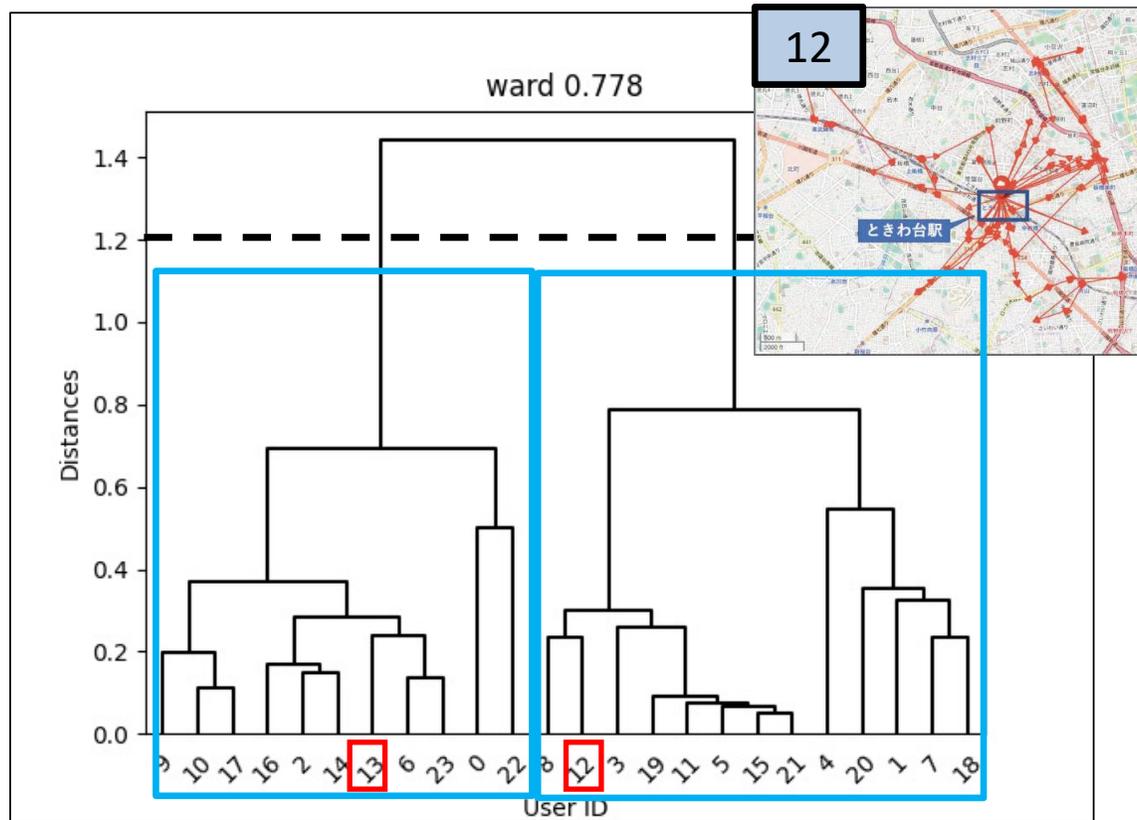


提案モデル(Mesh-ID + 緯度経度)
Mesh-ID交差エントロピー損失

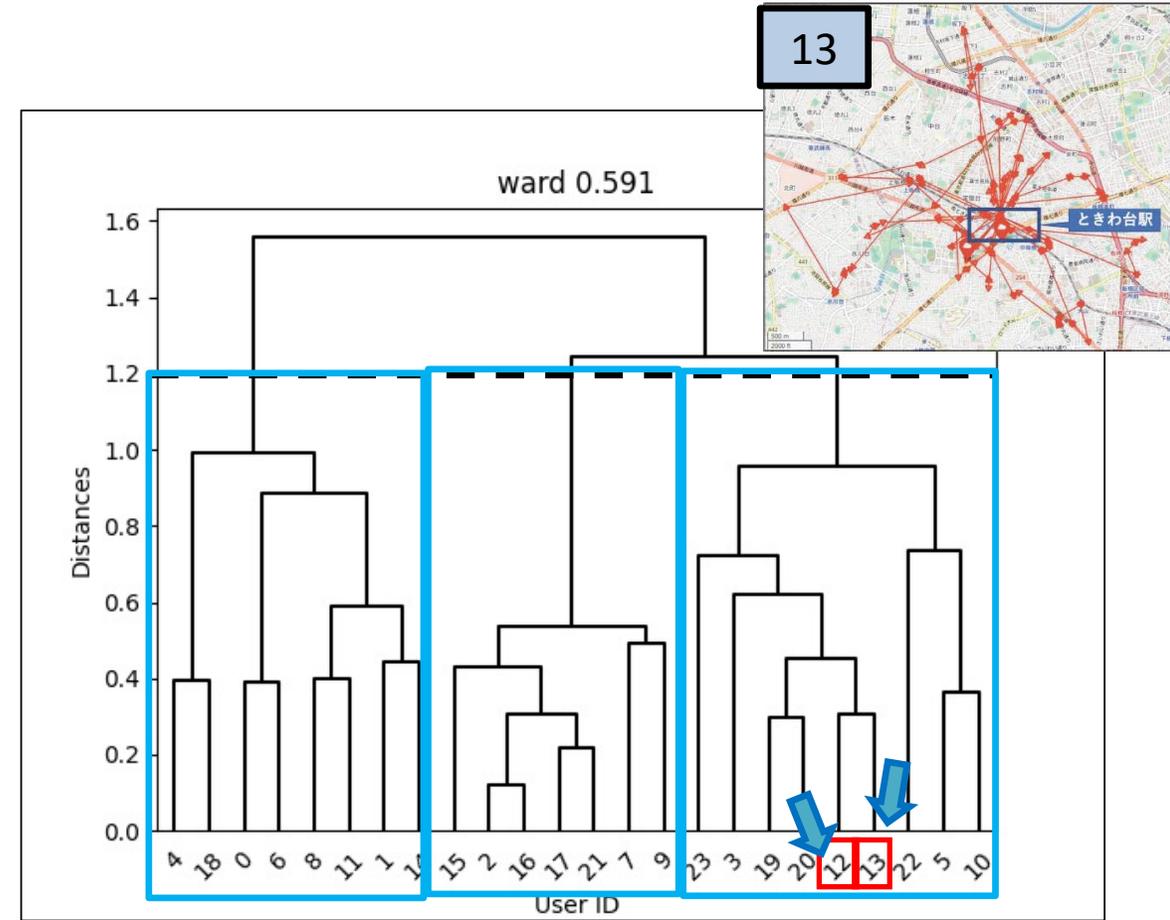


提案モデル(Mesh-ID + 緯度経度)
緯度損失(MSE)

緯度経度情報をモデルの入力に加えることで位置情報を考慮した分散表現が獲得でき、クラスタリング結果も妥当なものになった



従来手法



提案手法

まとめ

- ▶ Bi-LSTMの入力としてMesh-IDと位置情報(緯度経度)を与えることで正確に位置情報を考慮した分散表現を生成し、クラスタリング結果の妥当性を示した

今後の課題

- ▶ クラスタリング結果の解釈手法の検討
- ▶ 入力データのフィルタリング形式の検討

付録

クラスタリング結果ビューア

■ 検索対象
 読みデータ設定ファイル C:/Users/tapas/Dropbox (NAIST AHC)/work/Tokyo_part/TIS_demo2020/src/読みデータ設定.json 参照..

■ 検索条件
 ユーザ クラスタ

+	クラスタNo.	期間	時間帯	近傍	色(第n近傍)
-	1			2	1 2 3

クラスタNo. 10の中心ID-1のID番号
 ユーザID: 10
 クラスタNo.: 10
 経度: 139.7344
 緯度: 35.6812
 日時: 2020-02-19 23:07:00

- ▶ 解析対象データ(提供:Agoop Corp)
 - 場所 : 東京都
 - 期間 : 7-10/2020
 - 位置精度 : < 100m
 - 最小系列長 : 30
 - Subcategory : departure/arrival
 - 仕様の関係からiOS端末のみ対象

- ▶ 機械学習モデル : Bi-LSTM
- ▶ クラスタリング : 凝集型階層的クラスタリング

- 似た(カテゴリの)場所に訪問する行動はその行動の意味が近いという考え方[Crivellari+ 2019]



COM1-COM2間およびCOM1-COM3間の距離は同一だが、TRACE1とTRACE2はその行動において、同じ関心があるといえる

●階層的クラスタリングの評価指標

- 1に近いほどクラスター間距離とクラスター内距離の歪みが小さい

$$c = \frac{\sum_{i < j} (Y_{ij} - y)(Z_{ij} - z)}{\sqrt{\sum_{i < j} (Y_{ij} - y)^2 \sum_{i < j} (Z_{ij} - z)^2}}$$

Y_{ij} : Y に含まれるオブジェクト*i*とオブジェクト*j*の距離

Z_{ij} : Z から得られるオブジェクト*i*とオブジェクト*j*のコーフェン距離

y, z : Y, Z の平均

- 文章の距離を測る為に用いられる指標
 - 2つの観光行動ベクトルの距離が計算可能

$$\text{cos_similality} = \cos(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}| |\vec{y}|} = \frac{\sum_{k=1}^n x_k y_k}{\sqrt{\sum_{k=1}^n x_k^2} \sqrt{\sum_{k=1}^n y_k^2}}$$

●階層的クラスタリングのクラスタ併合方法

- ・クラスタ間の距離関数

$$D(C_1, C_2) = E(C_1 \cup C_2) - E(C_1) - E(C_2)$$

$$E(C_1) = \sum_{x_1 \in C_1} (D(x, c_i))^2$$

$$E(C_1 \cup C_2)$$

: 併合後のC1とC2の重心からの距離の二乗和

$$E(C_1)$$

: クラスタC1の重心からの距離の二乗和

