Title	有用性保持を目的とした多次元データとしての経路情報匿名化
Sub Title	
Author	佐久間, 結子(Nishikawa, Akihito) Tran, Thai P.(Nishi, Hiroaki) 岩井, 智夢 西川, 誠人 西, 宏章
Publisher	慶應義塾大学AI・高度プログラミングコンソーシアム
Publication year	2023
Jtitle	AICカンファレンス予稿集 (2023.) ,p.55- 55
JaLC DOI	
Abstract	経路情報には自宅や職場といったユーザの潜在的な情報が多く含まれており，データ提供時に匿名化によって保護する必要がある．経路情報を直接匿名化すると現実では通らない経路を匿名化結果として場合がある．そこで時系列データ向け深層学習モデルで経路情報を学習し，潜在空間上に射影した上で匿名化する．潜在空間上には複雑な意味的情報が含まれているため，経路の属性を保持しつつ現実的な匿名化を実現できる．
Notes	会議名 : AICカンファレンス2023 開催地 : 慶應義塾大学日吉キャンパス 日時 : 2023年3月4日 第3章既発表セッション要旨 既発表要旨-4
Genre	Conference Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO11003001-20230304-0055

有用性保持を目的とした多次元データとしての経路情報匿名化

佐久間結子[1], Thai P. Tran[1], 岩井智夢[1], 西川誠人[1], 西宏章[2]
[1]慶應義塾大学理工学研究科
[2]慶應義塾大学理工学部システムデザイン工学科

Abstract: 経路情報には自宅や職場といったユーザの潜在的な情報が多く含まれており，データ提供時に匿名化によって保護する必要がある．経路情報を直接匿名化すると現実では通らない経路を匿名化結果として場合がある．そこで時系列データ向け深層学習モデルで経路情報を学習し，潜在空間上に射影した上で匿名化する．潜在空間上には複雑な意味的情報が含まれているため，経路の属性を保持しつつ現実的な匿名化を実現できる．

Keywords: trajectory anonymization, differential privacy, Seq2Seq, latent space

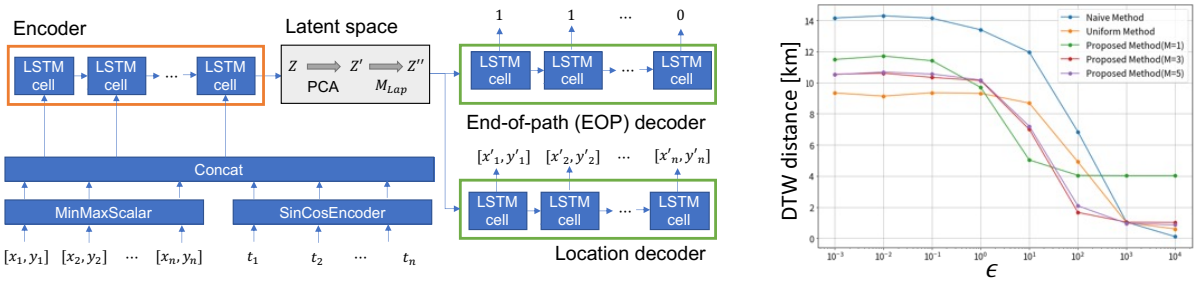

Fig. 1 (a) Proposed Seq2Seq model, (b) Dynamic time warping distance (DTW) results for the proposed method

1. 研究背景・目的

スマートフォン等により蓄積される経路情報は，渋滞予測や地図サービスなど利用される．しかし，経路情報には自宅や職場の場所といった個人情報が含まれているため公開時には匿名化する必要がある．先行研究では，差分プライバシを用いて経路情報に直接ノイズを付加する方法が提案されている [1] [2]. しかし生データに直接ノイズを付加することによる情報損失は大きく，現実に通らない経路を出力してしまうことでデータの有用性を著しく損なってしまう．さらに可変長な経路情報への適用は考慮されていない．そこで本研究では有用性を保持しつつ可変長データにも対応できる経路情報匿名化手法を提案する．

2. 提案方法

現実的な経路を匿名化結果として出力するために，オートエンコーダ型の深層学習モデルを用いて入力データを低次元・固定長の潜在空間に射影した上で，潜在空間上で差分プライバシにより匿名化する．(Fig.1 (a))

2.1. 深層学習モデル

時系列データによく用いられる Seq2Seq を採用した．緯度経度の位置情報を MinMaxScaler，時間情報を SinCosEncoder で標準化した．Encoder は入力を潜在空間に射影する．Location decoder は緯度経度情報を再構成し，End-of-path decoder は可変長の出力に対応するため，経路の存在をブール値で出力する．

2.2. 差分プライバシによる匿名化

潜在変数をラプラスメカニズムに基づくϵ-差分プライバシで匿名化する．敏感度をΔ, プライバシバジェット（ノイズ）をϵとするとクエリqに対する匿名化結果yは Equation 1 で表現できる．

$$r \sim Lap\left(0, \frac{\Delta}{\epsilon}\right), y = q + r \quad (1)$$

匿名化結果の有用性を向上させるために，principle component analysis(PCA)による潜在変数の各次元の重要度に応じてプライバシバジェットを分配する．

3. 結果と考察

実験には Agoop 社により 2017 年 1 月から 8 月に収集された GPS データを用い，さいたま市の 2×2 km の範囲を対象とした．比較対象の naïve method は入力データに直接ノイズを付加した．Uniform method は潜在変数の全ての次元に均一なノイズを付加した．Fig. 2(b)に異なるϵに対する匿名化前後の dynamic time warping(DTW)距離の結果を示す．ϵが小さいほど匿名化の度合いが大きい．Mは PCA の要素数である．$M = 3$のときに提案手法は比較手法よりも小さい DTW 距離を示している．すなわち，匿名化による情報損失の度合いが小さいことがわかる．

参考文献

[1] R Assam *et al.*, 2012. Differential private trajectory protection of moving objects. In Proceedings of the 3rd ACM SIGSPATIAL International Workshop on GeoStreaming, 68–77.

[2] K. Al-Hussaeni *et al.*, 2018. SafePath: Differentially-private publishing of passenger trajectories in transportation systems. Computer Networks, 143, 126-139.