道路ネットワークにおける位置情報プライバシを考慮した軌跡データの評価に関する研究

成瀬　真† 高木　駿†† 曹　洋†† 吉川　正俊††

† 京都大学工学部情報学科　〒 606–8501 京都府京都市左京区吉田本町 36-1
†† 京都大学情報学研究科　〒 606–8501 京都府京都市左京区吉田本町 36-1
E-mail: †{naruse,s.takagi}@db.soc.i.kyoto-u.ac.jp, ††{yang,yoshikawa}@i.kyoto-u.ac.jp

あらまし　本論文は，道路ネットワーク上で差分プライバシを保証し曖昧化された経路情報データについて評価を行い，価値付けを行うことを目的とする．価値付けといっても，価値には様々な観点が存在する. 例えば，購入者からすれば，どれほど情報の不確かさ (エントロピー) が減少したかが重要になってくるが，販売者からすれば，実際のところ正解データと予測データがどれほど離れているか (correctness)，プライバシーがどの度合いで保証されているか (差分プライバシにおける ε) などが重要になる．これらの観点から価値を計算し，比較を行った．結果，第 1 の指標である不確かさの減少量は、プライバシパラメータの変化と購買者の効用を上手く結びつけており，購入者目線での経路情報の価値を直感的に表すことが可能となった．第 2 の指標である誤差期待値は，販売者目線でのリスクを算出することができたが，差分プライバシの概念よりこれを決定的に価値に取り入れることはできない．その問題点を解決したのが第 3 の指標であるプライバシ損失であり，これにより販売者目線からの経路情報販売に関するリスクも算定し，価値に含めることが可能となった．

キーワード　経路情報, 差分プライバシ, Geo-Graph 識別不能性, 情報の評価, 情報銀行

1 はじめに

近年，パーソナルデータ市場はますます拡大し，それらのデータを安全かつ正確に価値付けをして売買する必要性は増すばかりである. 売買するといっても，パーソナルデータである以上，そのプライバシは保証されていなければならない. 特に移動の軌跡データは，その特性上，確実にプライバシを保護した上でその価値を正確に定めることは困難である. しかし，人々の移動履歴を表す経路情報データはマーケティング戦略のためにはなくてはならない重要なデータであり，可能な限りそのデータに対して価値付けをしていく必要がある. 先行研究として，連続した位置情報データに対して価値付けをしようと試みた研究が存在するが, 本論文ではそれらの考えを継承し，さらに具現化する．まずは，位置情報を平面上のものではなくグラフ上のものと考える. 位置情報をグラフ上に表現することで，到達不可能な場所や道路の情報・橋などを表現でき，実際の表現に近くなる. 実際，パーソナルデータに対する攻撃者はこれらの知識を元に攻撃を行うので，価値付けをする際にこのような概念を導入することは非常に重要となる．また本論文では，位置情報をグラフ上に表現する際に，Geo-Graph 識別不能性 (Geo-Graph-Indistinguishability) という方法を用いて位置情報データに雑音をかけることにしている. これは紹介する先行研究によって考案された，グラフ上の位置情報に対して差分プライバシーを保証するためのメカニズムである. それらの手法を用いて作成した，道路グラフ上で曖昧化された経路情報データについて，本研究では価値付けを行っていく．

2 関連研究

2.1 Geo 識別不能性

Geo 識別不能性 [4] とは差分プライバシ [5] の考えを元に位置情報に雑音を加え，正確な位置を識別不可能にする技術である．

利用者が位置している可能性のある地点の集合を X としたとき，X 内の全ての点に対する，ある確率的計算機構 K が存在し，X 内の任意の 2 点 x, x' に対して次の式 (1) を満たすとき，K は ε-Geo 識別不能を満足するという．

$$d_p(K(x), K(x')) \leqq \varepsilon d(x, x') \tag{1}$$

この時，$d(x, x')$ は 2 つの点 x, x' のユークリッド距離であり，ε はプライバシパラメータであり，この値が小さいほどプライバシが強く守られている．また，$d_p(K(x), K(x'))$ は異なる 2 つの点 x, x' から計算機構 K によって得られた確率分布 $K(x)$ と $K(x')$ の距離である．Andr‘es ら [4] では，ある集合 S に関する 2 つの確率分布 σ_1, σ_2 の距離を (2) のように定めている．

$$d_p(\sigma_1, \sigma_2) = \sup_{S \subseteq S} \left| \log \frac{\sigma_1(S)}{\sigma_2(S)} \right| \tag{2}$$

K の出力の確率分布が二次元ラプラス分布の場合，式 (2) を満足することが知られている．

2.2 Geo-Graph 識別不能性

Takagi ら [1] は，Geo 識別不能性の概念をグラフ上に導入した．

道路ネットワークを表す無向グラフ $G(V, E)$ が与えられ，各

枝には距離が定義されているとする．ユーザーが位置している可能性のある頂点の集合を V としたとき，任意の $v, v' \in V$ に対して，グラフ上の確率的計算機構 K が存在し，次の式 (3) を満たすとき K は，Geo-Graph 識別不能性を満足するという．

$$d_p(K(v), K(v')) \leqq \varepsilon d_s(v, v') \tag{3}$$

$d_s(v, v')$ はグラフ上での 2 つの頂点 v, v' の最短距離である．この Geo-Graph 識別不能性を満足するものとして，Takagi ら [1] では指数メカニズムを利用したグラフ上のメカニズム GE を以下 (4) のように定義している．

$$GE(v)(w) = \alpha(v) e^{-\frac{\varepsilon}{2} d_s(v,w)} \tag{4}$$

また，$\alpha(v)$ は正規化項であり以下 (5) のように表される．

$$\alpha(v) = \frac{1}{\sum_{w \in V} e^{-\frac{\varepsilon}{2} d_s(v,w)}} \tag{5}$$

本研究では，上述のメカニズム GE でグラフ上の点の曖昧化を行うことを前提とする．

2.3 プライバシに対する様々な尺度

プライバシ保護の尺度として，世の中には様々なものが提案されいるが，Shokri ら [2] はそれらを 3 つの尺度に整理している．

本研究ではそのうち 2 つの指標について扱う．

1 つ目は certainty であり，これはどれだけ確率が偏っているかを表す指標である．確率分布の平均情報量を求めることで certainty を算出することができる．これを経路情報の売買に当てはめると，購入者はこれが小さくなるほど高い効用を得ることとなり，より高い値段で購入することを許容できる．

2 つ目は correctness であり，これは実際に予測がどの程度正解から離れているかを表す指標である．この論文では，ある点が本当の点である確率にその予測が正解からどれほど離れているかを掛けたものの和を取ることで求めている．これを経路情報の売買に当てはめると，販売者はこれが小さくなるほど高いリスクを負うこととなり，より高い値段での販売を望むこととなる．

2.4 時系列データに対するプライバシパラメータの変化

Yang ら [7] は，時系列データに対するプライバシーパラメータの変化の計算手法を示した．一般的に，時系列データでは，1 つの曖昧化したデータを公開することで，他の点に対する情報も与えてしまう．そのため，ある時刻の点をプライバシパラメータ ε を保証して公開しても，他の時刻の点を公開することで実際に保証しているプライバシパラメータは変化してしまう．そのような変化を実際に求めることで，時系列データであっても，どの程度のプライバシ損失が起こっているのかを算出することができる．本研究では，これらの考えを経路情報に拡張した計算手法を提案する．

3 価値算出のための準備

本章では，経路情報についての価値を算出するにあたっての下準備を行う．まず，元データに対して Geo-Graph 識別不能性により曖昧化の処理を行う．次に，Geo-Graph 識別不能性の確率分布により計算した，各時刻の各頂点に対する予測分布を求める．次に，これら操作を隠れマルコフモデルと捉え，複数の公開点が与えられた時に各時刻の各頂点に対する予測分布を求める．また，各変数を表 1 のように定める．

表 1　変数の意味

変数	変数の意味
$G(V, E)$	ある 1 つの道路ネットワーク
V	道路ネットワーク G における頂点集合
E	道路ネットワーク G における道 (重み付きの辺) の集合
$\boldsymbol{l}$	道路ネットワーク G における，ある 1 つの軌跡データの集合
$l^t \in V$	軌跡 l の時刻 t における位置
$r^t \in V$	l^t を Geo-Graph 識別不能性によって曖昧化して公開した位置
$d_s(v_1, v_2)$	2 つの頂点間のグラフ上の最短距離
M_f	道路ネットワーク G における前向きの遷移行列
M_b	道路ネットワーク G における後ろ向きの遷移行列
$P(\boldsymbol{l}^\tau)$	時刻 τ の 1 つの頂点 l^τ に与えられる予測確率ベクトル
$P(\boldsymbol{l}^\tau \mid r^{t_1}, r^{t_2}, ..., r^{t_k})$	時刻 $t_1, t_2, ..., t_k$ の公開点から予測した時刻 τ の確率ベクトル
$P(l_i^\tau \mid r^{\boldsymbol{T}}) \in [0, 1]$	上記ベクトルの第 i 成分であり，時刻集合 $\boldsymbol{T}$ の公開点から予測される分布において l_i^τ に与えられる確率

3.1 軌跡データの各頂点を Geo-Graph 識別不能性によって曖昧化

本節では，軌跡データの各頂点を Geo-Graph 識別不能性より曖昧化することを考える．(4) より，本当の位置が l_j^t であるとき，それが r_i^t に曖昧化される確率は以下のように表される．

$$P(r_i^t | l_j^t) = \alpha(l_j^t) e^{-\frac{\varepsilon}{2} d_s(r_i^t | l_j^t)} \tag{6}$$

逆に，ある公開された点 r_i^t が与えられた時，1 つの点 r_i^t から予測した時刻 t における位置予測の分布 $P(\boldsymbol{l}^\tau | r_i^t)$ を考える (ここでは $\tau = t$)．この時，$P(l_j^\tau | r_i^t)$ は以下のように表される．

$$P(l_j^\tau | r_i^t) = \alpha(r_i^t) e^{-\frac{\varepsilon}{2} d_s(l_j^\tau | r_i^t)} \tag{7}$$

次の節では，この確率分布を利用して他の時刻の分布を求めることを考える．

3.2 他の時刻の点を予測する

この節では，ある時刻の予測確率分布が与えられた時，そこから他の時刻の予測確率分布を求めることを考える．

まず前提として，この経路情報を曖昧化して公開するというモデルは，隠れマルコフモデル (以下，HMM) となっていることを

確認する．存在確率の初期分布を $P(\boldsymbol{\pi}) = P(\boldsymbol{l}^0)$ とすると，時刻 $k(k = 1, 2, ..., T)$ の存在確率は $P(\boldsymbol{l}^k) = P(\boldsymbol{l}^{k-1}) * P(\boldsymbol{l}^k|\boldsymbol{l}^{k-1}) = P(\boldsymbol{l}^{k-1}) * M_f$ と表せるため，これはマルコフ過程である．

一方，実際に公開する点は本当の位置にプライバシメカニズムによって雑音をかけた位置である．この確率は，$P(\boldsymbol{r}^t|l^t)$ で与えられ，これはその時刻に実際に位置している頂点のみで決まる．全ての点を公開した時，これは隠れマルコフ過程となる．このことを図を用いて表すと図 1 のようになる．

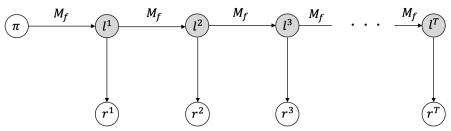

図 1　経路情報の隠れマルコフ過程

次に，一部の点を購入することを考える． 図 1 では，全ての時刻の点について観測が得られていたが，一部の点を購入する場合では一部の観測しか得られない．

ここで，計算しやすくするために，観測がない点についてはマルコフ過程の状態を省略する．省略した場合，各状態の遷移確率は遷移行列 M_f を累乗した行列となる．これを図で表すと図 2 のようになる．これにより，再び隠れマルコフ過程として考えることができる．

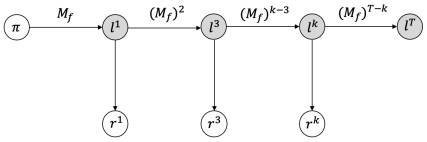

図 2　移動確率が異なる隠れマルコフ過程の例

ここで, Lawrence [6] にある通り, forward-backward アルゴリズムを用いて計算を行う．求めたい分布が $P(\boldsymbol{l}^\tau|r^{t_1}, r^{t_2}, ..., r^{t_k})$ である時，表 2 のように各変数を定義する．

表 2　HMM における変数の意味

変数	変数の意味
$r^{\boldsymbol{T}}$	公開されている全点，$r^{\boldsymbol{T}} = r^{\boldsymbol{b}} \cup r^{\boldsymbol{f}}$
$r^{\boldsymbol{b}}$	時刻 τ 以前の時刻の公開されている点の集合
$r^{\boldsymbol{f}}$	時刻 τ よりも後の時刻の公開されている点の集合
$l^{\boldsymbol{b}}$	$r^{\boldsymbol{b}}$ の真の位置の集合
$l^{\boldsymbol{f}}$	$r^{\boldsymbol{f}}$ の真の位置の集合
r^{b_1}	$r^{\boldsymbol{b}}$ のうち，一番時刻の早い点
r^{b_n}	$r^{\boldsymbol{b}}$ のうち，一番時刻の遅い点，l^τ と等しくなる場合がある
r^{f_1}	$r^{\boldsymbol{f}}$ のうち，一番時刻の早い点
r^{f_m}	$r^{\boldsymbol{f}}$ のうち，一番時刻の遅い点
$r^{b_{1:k}}$	$r^{b_1}, r^{b_2}, ..., r^{b_k}$ の略記

この時，状態を $\boldsymbol{l}^b$，$\boldsymbol{l}^f$，($\boldsymbol{l}^b$ に含まれないのなら)l^τ とし，観測を $r^{\boldsymbol{b}}$，$r^{\boldsymbol{f}}$ とした HMM を考える．これを図で表すと図 3 のようになる．Lawrence [6] では全ての点について観測があるが，ここでは，時刻 τ に観測がない場合も考える．時刻 τ に観測がない場合，後述のように $P(\boldsymbol{l}^\tau|\boldsymbol{l}^{b_n})$ という遷移確率を用いることで目的の確率分布を得ることができる．

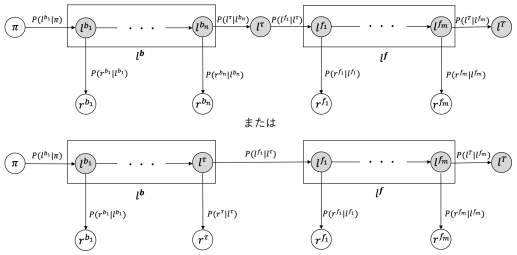

図 3　forward-backward アルゴリズムの俯瞰

この時，$P(\boldsymbol{l}^\tau|r^{\boldsymbol{T}})$ を (8) のように変形する．

$$\begin{aligned} P(\boldsymbol{l}^\tau|r^{\boldsymbol{T}}) &= P(\boldsymbol{l}^\tau|r^{\boldsymbol{b}}, r^{\boldsymbol{f}}) = \frac{P(\boldsymbol{l}^\tau, r^{\boldsymbol{f}}|r^{\boldsymbol{b}})}{P(r^{\boldsymbol{f}}|r^{\boldsymbol{b}})} \\ &\propto P(\boldsymbol{l}^\tau, r^{\boldsymbol{f}}|r^{\boldsymbol{b}}) = P(r^{\boldsymbol{f}}|\boldsymbol{l}^\tau, r^{\boldsymbol{b}})P(\boldsymbol{l}^\tau|r^{\boldsymbol{b}}) \\ &= P(r^{\boldsymbol{f}}|\boldsymbol{l}^\tau)P(\boldsymbol{l}^\tau|r^{\boldsymbol{b}}) \end{aligned} \tag{8}$$

よって，複数点からの予測では，時刻 τ よりも前の時刻の点，後の時刻の点についてそれぞれ計算しその結果の積を求めて正規化すれば良い．片側の点しか存在しない場合には片側の計算結果がそのまま予測結果となる．

ここからは $P(r^{\boldsymbol{f}}|\boldsymbol{l}^\tau)$ と $P(\boldsymbol{l}^\tau|r^{\boldsymbol{b}})$ についてそれぞれ考える．

まずは $P(\boldsymbol{l}^\tau|r^{\boldsymbol{b}})$ について考える．ここでは foward アルゴリズムを用いる．

$P(\boldsymbol{l}^{b_1}|r^{b_1})$ はベイズの定理を用いて以下のように変形できる．

$$\begin{aligned} P(\boldsymbol{l}^{b_1}|r^{b_1}) &= \frac{P(r^{b_1}|\boldsymbol{l}^{b_1})P(\boldsymbol{l}^{b_1})}{P(r^{b_1})} \propto P(r^{b_1}|\boldsymbol{l}^{b_1})P(\boldsymbol{l}^{b_1}) \\ &= P(r^{b_1}|\boldsymbol{l}^{b_1})P(\boldsymbol{l}^{b_1}|\boldsymbol{\pi})P(\boldsymbol{\pi}) \end{aligned} \tag{9}$$

また, $r^{\boldsymbol{b}}$ が 2 つ以上の場合, 図 4 のように $P(\boldsymbol{l}^{b_{k+1}}|r^{b_{1:k+1}})(k = 1, ..., n-1)$ は $P(\boldsymbol{l}^{b_k}|r^{b_{1:k}})$ より再帰的に計算できる．

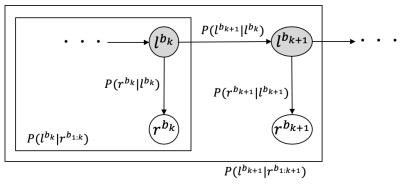

図 4　forward アルゴリズムの再帰的計算

以下のように $P(\boldsymbol{l}^{b_{k+1}}|r^{b_{1:k+1}})$ を変形して $P(\boldsymbol{l}^{b_k}|r^{b_{1:k}})$ との関係を得る．

$$\begin{aligned} P(\boldsymbol{l}^{b_{k+1}}|r^{b_{1:k+1}}) &= \frac{P(\boldsymbol{l}^{b_{k+1}}, r^{b_{k+1}}|r^{b_{1:k}})}{P(r^{b_{k+1}}|r^{b_{1:k}})} \\ \propto P(\boldsymbol{l}^{b_{k+1}}, r^{b_{k+1}}|r^{b_{1:k}}) &= P(r^{b_{k+1}}|\boldsymbol{l}^{b_k+1})P(\boldsymbol{l}^{b_{k+1}}|r^{b_{1:k}}) \\ \propto P(r^{b_{k+1}}|\boldsymbol{l}^{b_k+1})P(\boldsymbol{l}^{b_{k+1}}|\boldsymbol{l}^{b_k})P(\boldsymbol{l}^{b_k}|r^{b_{1:k}}) \end{aligned} \tag{10}$$

l^τ と l^{b_n} が異なる場合，$P(\boldsymbol{l}^{b_n}|r^{b_{1:n}})$ に遷移行列 $P(\boldsymbol{l}^\tau|\boldsymbol{l}^{b_n})$ を乗ずることにより $P(l^\tau|r^{b_{1:n}})$ を求めることができる．結局，$P(l^\tau|r^{\boldsymbol{b}})$ は以下のように計算できる．

$$P(\boldsymbol{l}^\tau|r^{\boldsymbol{b}}) \propto \begin{cases} P(r^{b_1}|\boldsymbol{l}^{b_1})P(\boldsymbol{l}^{b_1}|\boldsymbol{\pi})P(\boldsymbol{\pi}) & (n=1 \wedge l^{b_n}=l^\tau) \\ P(\boldsymbol{l}^\tau|\boldsymbol{l}^{b_1})P(r^{b_1}|\boldsymbol{l}^{b_1})P(\boldsymbol{l}^{b_1}) & (n=1 \wedge l^{b_n} \neq l^\tau) \\ \prod_{k=1}^{n-1}\{P(r^{b_{k+1}}|\boldsymbol{l}^{b_k+1})P(\boldsymbol{l}^{b_{k+1}}|\boldsymbol{l}^{b_k})\}P(\boldsymbol{l}^{b_1}|\boldsymbol{\pi})P(\boldsymbol{\pi}) & (n \geqq 2 \wedge l^{b_n}=l^\tau) \\ P(\boldsymbol{l}^\tau|\boldsymbol{l}^{b_n})\prod_{k=1}^{n-1}\{P(r^{b_{k+1}}|\boldsymbol{l}^{b_k+1})P(\boldsymbol{l}^{b_{k+1}}|\boldsymbol{l}^{b_k})\}P(\boldsymbol{l}^{b_1}) & (n \geqq 2 \wedge l^{b_n} \neq l^\tau) \end{cases} \tag{11}$$

ただし，$P(\boldsymbol{l}^{b_1}) = P(\boldsymbol{l}^{b_1}|\boldsymbol{\pi})P(\boldsymbol{\pi})$

次に $P(r^{\boldsymbol{f}}|\boldsymbol{l}^\tau)$ について考える．ここでは backard アルゴリズムを用いる．まずは，一番未来の $P(\phi|\boldsymbol{l}^{f_m})$ について考える．ここで，Lawrence [6] にもある通り，ここでは，ある観測の集合 $r^{f_{k+1:m}}$ が与えられた時に時刻 f_k の時に $\boldsymbol{l}^{f_k}$ が存在したであろう確率 $P(r^{f_{k+1:m}}|\boldsymbol{l}^{f_k})$ を求めることを目標としている．その初期確率である $P(\phi|l^{f_m})$ は未知なので，given(ここでは一様分布) とみなして計算を行う．即ち，$P(\phi|l^{f_m}) \propto [1\ 1\ ...\ ]^T$ とし，最後に正規化を行うこととする．

まずは $r^{\boldsymbol{f}}$ が 1 つ (m=1) である場合，$P(r^{f_1}|l^\tau)$ を以下のように変形して求める．

$$\begin{aligned} P(r^{f_1}|\boldsymbol{l}^\tau) &= P(r^{f_1}|\boldsymbol{l}^{f_1})P(\boldsymbol{l}^{f_1}|\boldsymbol{l}^\tau) \\ &= P(\phi|\boldsymbol{l}^{f_1})P(r^{f_1}|\boldsymbol{l}^{f_1})P(\boldsymbol{l}^{f_1}|\boldsymbol{l}^\tau) \\ &\propto P(r^{f_1}|\boldsymbol{l}^{f_1})P(\boldsymbol{l}^{f_1}|\boldsymbol{l}^\tau) \end{aligned} \tag{12}$$

$r^{\boldsymbol{f}}$ が 2 点以上の場合，図 5 のように，$P(r^{f_{k:m}}|\boldsymbol{l}^{f_{k-1}})$ を $P(r^{f_{k+1:m}}|\boldsymbol{l}^{f_k})$ から再帰的に求める．

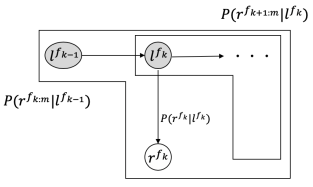

図 5　backward アルゴリズムの再帰的計算

$P(r^{f_{k:m}}|\boldsymbol{l}^{f_{k-1}})$ は以下のように変形して計算を行う．

$$\begin{aligned} P(r^{f_{k:m}}|\boldsymbol{l}^{f_{k-1}}) &= P(r^{f_{k:m}}|\boldsymbol{l}^{f_k})P(\boldsymbol{l}^{f_{k-1}}|\boldsymbol{l}^{f_k}) \\ &= P(r^{f_{k+1:m}}|\boldsymbol{l}^{f_k})P(l^p_{f_k}|\boldsymbol{l}^{f_k})P(\boldsymbol{l}^{f_{k-1}}|\boldsymbol{l}^{f_k}) \end{aligned} \tag{13}$$

よって，以下のように $P(r^{f_{2:m}}|\boldsymbol{l}^{f_1})$ は再帰的に計算できる．

$$P(r^{f_{2:m}}|\boldsymbol{l}^{f_1}) = \prod_{k=2}^{m} P(r^{f_k}|\boldsymbol{l}^{f_k})P(\boldsymbol{l}^{f_{k-1}}|\boldsymbol{l}^{f_k}) \tag{14}$$

以上より，次のように $P(r^{\boldsymbol{f}}|\boldsymbol{l}^\tau)$ を求めることができる．

$$P(r^{\boldsymbol{f}}|l^\tau) \propto \begin{cases} P(r^{f_1}|\boldsymbol{l}^{f_1})P(\boldsymbol{l}^{f_1}|\boldsymbol{l}^\tau) & (m=1) \\ P(\boldsymbol{l}^{f_1}|\boldsymbol{l}^\tau)P(r^{f_1}|\boldsymbol{l}^{f_1})\prod_{k=2}^{m} P(r^{f_k}|\boldsymbol{l}^{f_k})P(\boldsymbol{l}^{f_k}|\boldsymbol{l}^{f_{k-1}}) & (m \geqq 2) \end{cases} \tag{15}$$

以上より，$P(r^{\boldsymbol{f}}|\boldsymbol{l}^\tau)$ と $P(\boldsymbol{l}^\tau|r^{\boldsymbol{b}})$ を求めることができた．正規化するためには Lawrence [6] によると，$P(r^{\boldsymbol{T}}|\boldsymbol{\pi})$ で割れば良い．結局，$P(l_i^\tau|r^{\boldsymbol{T}})$ は以下のように計算できる．

$$P(l_i^\tau|r^{\boldsymbol{T}}) = \frac{P(r^{\boldsymbol{f}}|l_i^\tau)\cdot P(l_i^\tau|r^{\boldsymbol{b}})}{P(r^{\boldsymbol{T}}|\pi)} = \frac{P(r^{\boldsymbol{f}}|l_i^\tau)\cdot P(l_i^\tau|r^{\boldsymbol{b}})}{\prod_{r^k \in r^{\boldsymbol{T}}} P(r^k|\pi)} \tag{16}$$

4　価値の算出

この章では，実際に経路情報データについての評価を行う．ここでは，購入者視点から経路を評価する「不確かさ (entropy)」と販売者視点から経路を評価する「誤差期待値 (correctness)」，「プライバシの損失 (privacy leakage)」を求める．

4.1　不 確 か さ

この節では，価値を算出する指標として「不確かさ (entropy)」を用いる．これは，公開されたデータから計算される確率分布のばらつき度合いを表す指標である．この不確かさが小さいほど，分布にばらつきがあり，予測がしやすいことになる．この指標は 2.3 節で述べたうちの certainty にあたる．

4.1.1　概念の導入

ここでは，指標として各時刻の平均情報量 $H(X)$ を用いる．これは以下 (17) で定義される．

$$H(X) = -\sum_{i=1}^{n} P(x_i)\log P(x_i) \tag{17}$$

これを道路ネットワークの概念上で表現すると，時刻集合 $\boldsymbol{t}$ の公開点から予測した時刻 τ の確率分布の平均情報量 $H(\boldsymbol{l}^\tau|r^{\boldsymbol{t}})$ は以下 (18) で表される．

$$H(\boldsymbol{l}^\tau|r^{\boldsymbol{t}}) = -\sum_{l_i^\tau \in \boldsymbol{l}^\tau} P(l_i^\tau|r^{\boldsymbol{t}})\log P(l_i^\tau|r^{\boldsymbol{t}}) \tag{18}$$

以上で述べた平均情報量は，確率分布が一様分布 (全ての点について予測確率が全く同じ) 時に最大値 $|V|\log|V|$ をとり，1 つの点に予測が定まっている (確率が 1 となっている) 時に最小値 0 をとる．この平均情報量が小さいほど不確かさは減り，位置情報の予測がしやすいことになる．

4.1.2　各時刻の不確かさ

前節で述べた方法により，公開する経路情報から平均情報量を求めることができる．前章で述べた通り，この経路情報を公開するという過程は HMM となっている．一般に，HMM は，観測が増えると不確かさが減少することが知られている．公開されている点から一部の点を用いて予測を行うことも考えられ

るが，ここでは，公開されている点全点より予測した $P(\boldsymbol{l}^\tau|r^{\boldsymbol{t}})$ を用いて各時刻の平均情報量を求めることとなる．

4.1.3 購入者が得られる効用を求める

ここでは，本題である経路情報の購買により購入者がどれほどの効用を得られることができるか，つまり経路を購入することにどれほどの価値があるかを考える．まず，市場にある連続した時系列のデータの集合を U とする．また U のうち，購入者が購入済の (所有している) 点集合を $V \subset U$ とし，新たに購入する点集合を $X \subseteq U$ とする．購入しているデータがない場合は $V = \phi$ となる．この時，U の時刻を $\boldsymbol{T}$，V の時刻を $\boldsymbol{t}_V \subset \boldsymbol{T}$，$X$ の時刻を $\boldsymbol{t}_X \subseteq \boldsymbol{T}$ とし，価格 $p_1(X|V)$ を以下のように定義する．

$$p_1(X|V) = \sum_{\tau \in T} \frac{H(\boldsymbol{l}^\tau|r^{\boldsymbol{t}_V}) - H(\boldsymbol{l}^\tau|r^{\boldsymbol{t}_V}, r^{\boldsymbol{t}_X})}{H(\boldsymbol{l}^\tau|r^{\boldsymbol{t}_V}, r^{\boldsymbol{t}_X})} \tag{19}$$

この式は，H が変化しない場合は 0 となり，販売価格は 0 となる．また，H が購入により元の H の $\frac{1}{n}$ 倍となるとこの式は $n-1$ となる．また，購入後の H が 0 になる場合は無限大に発散する．この関数は，不確かさが多く変化すると購入者は多くの情報を得ることができるという考え，及び不確かさが 0 になると位置情報が特定できてしまうためその場合は価格を無限にする (購入できなくする) という考えから作られている．この指標は，プライバシリスクと価値を上手く結びつけている．各点を曖昧化する際にあるプライバシパラメータを設定していても，経路全体の各頂点の情報量がどれほどになっているかは分からない．しかし，この指標を用いることで直感的にプライバシパラメータと価値を結びつけることができ，保証すべきプライバシパラメータの知識がなくとも情報の売買が可能となる．

4.2 誤差期待値

この節では，価値を算出する指標として「誤差期待値」を用いる．これは，公開されたデータから計算される確率分布と実際の連続位置情報データがどの程度離れていることを表す指標である．この誤差期待値が小さいほど，実際の正解に対しての予測が正しく，販売者にとってはリスクが大きくなることになる．この指標は 2.3 節で述べたうちの correctness にあたる．

4.2.1 概念の導入

ここでは，指標として各時刻の誤差期待値 $E(X)$ を用いる．これは以下 (20) で定義される．

$$E(X) = \sum_{i=1}^{n} P(x_i)||x_i - x_c|| \tag{20}$$

ここで，$||x_i - x_c||$ は予測している確率変数と実際の正解の確率変数とのノルムである．これを道路ネットワークの概念上で表現すると，時刻集合 $\boldsymbol{t}$ の公開点から予測した時刻 τ の確率分布の誤差期待値 $E(\boldsymbol{l}^\tau|r^{\boldsymbol{t}})$ は以下 (21) で表される．

$$E(\boldsymbol{l}^\tau|r^{\boldsymbol{t}}) = \sum_{l_i^\tau \in \boldsymbol{l}^\tau} P(l_i^\tau|r^{\boldsymbol{t}}) d_s(v_i, l^\tau) \tag{21}$$

以上で述べた誤差期待値は，予測が一番誤っている (正解からグラフ上の距離で一番遠い) 点 v_w に対して定まっている (確率が 1 となっている) 時に最大値 $d_s(v_w, l^\tau)$ をとり，予測が正解の点 l^τ に対して定まっている (確率が 1 となっている) 時に最小値 0 をとる．この誤差期待値が小さいほど予測の実際の精度が高く，販売者にとっては情報を売るリスクが大きいこととなる．

4.2.2 各時刻の誤差期待値

ここでは，各時刻に対する予測の誤差期待値を求める．上で述べた誤差期待値を計算するためには実際の正解データを用いるため，購買者はこの誤差期待値を計算することができない．購買者にが計算できるのは予測分布 $P(\boldsymbol{l}^\tau|r^{\boldsymbol{t}})$ のみであるが，ここでは，販売者の目線に立ち正解を知ることができるものとして，時刻 τ の誤差期待値 $E(\boldsymbol{l}^\tau|r^{\boldsymbol{t}})$ を求める．

4.2.3 販売者目線での価値を観る

ここでは，経路情報の購買により販売者がどれほどのリスクを負うこととなるか，つまり経路を販売することに販売者はどれほどの対価を求めるかを考える．U，V，X，T，$\boldsymbol{t}_U$，$\boldsymbol{t}_X$，$\boldsymbol{t}_V$ を 4.1.3 節と同様に定義し，価格 $p_2(X|V)$ を以下のように定義する．

$$p_2(X|V) = \sum_{\tau \in T} \frac{E(\boldsymbol{l}^\tau|r^{\boldsymbol{t}_V}) - E(\boldsymbol{l}^\tau|r^{\boldsymbol{t}_V}, r^{\boldsymbol{t}_X})}{E(\boldsymbol{l}^\tau|r^{\boldsymbol{t}_V}, r^{\boldsymbol{t}_X})} \tag{22}$$

この式は，(19) と同様の考え方のもと作られている．しかし，大きく異なる特徴がある．それは，値がマイナスになり得るということだ．そのケースは予測が悪い方向へ定まる方向に進むことによって起こる．新たな点の購入によって H が更新され，新たな予測が誕生したとする．その予測が高いと予測している点が正解から大きく離れている点ならば E は増加することとなる．このようなケースが多くの時刻で起こることによってこのマイナスの価値は起こり得る．

そもそも，この指標より決定的なアルゴリズムを用いて価格の調整を行うと，経路の曖昧化に対する差分プライバシの保証ができなくなってしまう．そのため，この指標は価格の決定に使うのではなく，その販売に対するフィードバックとして用いるのが適切だろう．

これらの課題の解決のため，次節では予測分布のみから決定できる「プライバシの損失 (privacy leakage)」を考える．

4.3 プライバシパラメータの変化

本節では，購入する頂点の変化によって差分プライバシにおけるプライバシパラメータ ε がどのように変化するのかを考える．これは 2.3 節であげたプライバシの指標にはどれも当てはまらないが，攻撃者からの位置情報のプライバシをどの程度保証できているか重要な指標である．

4.3.1 概念の導入

まずは，経路情報における Geo-Graph 識別不能性の定義を導入する．2.3 節により，時刻を考慮しない場合の Geo-Graph 識別不能性は定義できる．これと，Yang ら [7] の考え方を組み合わせることで，以下のように経路情報のプライバシパラメータを定義する．

定義

実際の経路情報の集合を $\boldsymbol{l}$ とする．また，その経路情報のあ

る 1 つの時刻 τ においての位置を l^τ とし，同じ時刻で位置の異なる頂点を l'^τ とする．

この時，公開する頂点の集合を $r^{\boldsymbol{t}}$ とすると，以下の条件を満たすとき，経路情報で時刻 τ において εGeo-Graph 識別不能性を満たすという．

$$\log \frac{P(r^{\boldsymbol{t}}|l^\tau)}{P(r^{\boldsymbol{t}}|l'^\tau)} \leqq \varepsilon' d_s(l^\tau, l'^\tau) \tag{23}$$

これは，$r^{\boldsymbol{t}} = r^\tau$ の時 1 点である場合の Geo-Graph 識別不能性の定義と一致する．図で表すと図 6 のようになる．公開されている点の集合 $r^{\boldsymbol{t}}$ を見たとき，任意の l^τ, l'^τ について，l^τ からそれぞれの状態を求め $r^{\boldsymbol{t}}$ が出力される確率と，l'^τ からそれぞれの状態を求め $r^{\boldsymbol{t}}$ が出力される確率は有限に抑えられる．

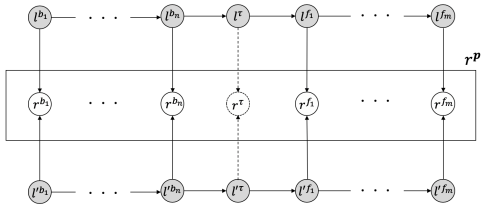

図 6　経路情報の場合の Geo-Graph 識別不能性

ここで，$P(r^{\boldsymbol{t}}|l^\tau)$ について考える．時刻 τ よりも前の公開されている頂点集合を $r^{\boldsymbol{b}}$，τ よりも後の公開されている頂点集合を $r^{\boldsymbol{f}}$ とすると，$P(r^{\boldsymbol{t}}|l^\tau)$ は以下のように変形できる．

$$P(r^{\boldsymbol{t}}|l^\tau) = P(r^{\boldsymbol{b}}|l^\tau)P(r^\tau|l^\tau)P(r^{\boldsymbol{f}}|l^\tau) \tag{24}$$

よって，(23) の左辺は以下のように変形できる．

$$\log \frac{P(r^{\boldsymbol{t}}|l^\tau)}{P(r^{\boldsymbol{t}}|l'^\tau)} = \log \frac{P(r^{\boldsymbol{b}}|l^\tau)}{P(r^{\boldsymbol{b}}|l'^\tau)} + \log \frac{P(r^{\boldsymbol{f}}|l^\tau)}{P(r^{\boldsymbol{f}}|l'^\tau)} + \log \frac{P(r^\tau|l^\tau)}{P(r^\tau|l'^\tau)} \tag{25}$$

ここで，Yang ら [7] の考えを拡張し，これらの式のプライバシに関する上限を以下のように定義する．

定義

$$TPl^\tau = \sup_{r^{\boldsymbol{t}}, l^\tau, l'^\tau} \frac{1}{d_s(l^\tau, l'^\tau)} \log \frac{P(r^{\boldsymbol{t}}|l^\tau)}{P(r^{\boldsymbol{t}}|l'^\tau)} \tag{26}$$

$$FPl^\tau = \sup_{r^{\boldsymbol{f}}, l^\tau, l'^\tau} \frac{1}{d_s(l^\tau, l'^\tau)} \log \frac{P(r^{\boldsymbol{f}}|l^\tau)}{P(r^{\boldsymbol{f}}|l'^\tau)} \tag{27}$$

$$BPl^\tau = \sup_{r^{\boldsymbol{b}}, l^\tau, l'^\tau} \frac{1}{d_s(l^\tau, l'^\tau)} \log \frac{P(r^{\boldsymbol{b}}|l^\tau)}{P(r^{\boldsymbol{b}}|l'^\tau)} \tag{28}$$

$$PL_0^\tau = \sup_{r^\tau, l^\tau, l'^\tau} \frac{1}{d_s(l^\tau, l'^\tau)} \log \frac{P(r^\tau|l^\tau)}{P(r^\tau|l'^\tau)} \tag{29}$$

TPl^τ は全体のプライバシ損失，PL_0^τ は時刻 τ 自身のプライバシ損失，FPl^τ は時刻 τ 以降のプライバシ損失，BPl^τ は時刻 τ 以前のプライバシ損失である．

これらを用いて (25) の上限を表すと以下のようになる．

$$TPl^\tau = FPl^\tau + BPl^\tau + PL_0^\tau \tag{30}$$

よって，全体のプライバシ損失である TPl^τ を知るためには，FPl^τ，BPl^τ，PL_0^τ の 3 つをそれぞれ計算して足し合わせれば良い．

まずは PL_0^τ について考える．これは，Geo-Graph 識別不能性の定義より上限は ε となる．

次に，BPl^τ について考える．3.2.2 節と同様に，$r^{\boldsymbol{b}}$ の個数を n とし (3.2.2 節と異なり $r^{\boldsymbol{b}}$ は r^τ を含まない)，$r^{\boldsymbol{b}} = r^{b_{1:n}}$ と表すと，BPl^τ は以下のように変形できる．

$$BPl^\tau = \sup_{r^{b_{1:n}}, l^\tau, l'^\tau} \frac{1}{d_s(l^\tau, l'^\tau)} \log \frac{\sum_{l^{b_n}} P(r^{b_{1:n}}|l^{b_n})P(l^{b_n}|l^\tau)}{\sum_{l'^{b_n}} P(r^{b_{1:n}}|l'^{b_n})P(l'^{b_n}|l'^\tau)} \tag{31}$$

ここで，$\sup_{r^{b_{1:k}}, l^{b_k}, l'^{b_k}} \frac{1}{d_s(l^{b_k}, l'^{b_k})} \log \frac{P(r^{b_{1:k}}|l^{b_k})}{P(r^{b_{1:k}}|l'_{b_k})}$ を BPl^{b_k} と表すと，BPl^{b_k} は以下のように再帰的に計算できる．

$$\begin{aligned}
&BPl^{b_k} = \\
&\sup_{r^{b_{1:k-1}}, l^{b_k}, l'^{b_k}} \frac{1}{d_s(l^{b_k}, l'^{b_k})} \\
&\log \frac{\sum_{l^{b_{k-1}}} P(r^{b_{1:k-1}}|l^{b_k-1})P(l^{b_k-1}|l^{b_k})}{\sum_{l'^{b_{k-1}}} P(r^{b_{1:k-1}}|l'^{b_k-1})P(l'^{b_k-1}|l'^{b_k})} \\
&+ \sup_{r^{b_k}, l^{b_k}, l'^{b_k}} \frac{1}{d_s(l^{b_k}, l'^{b_k})} \log \frac{P(r^{b_k}|l^{b_k})}{P(r^{b_k}|l'^{b_k})} \\
&= \mathcal{L}^B(BPl^{b_{k-1}}) + PL_0^{b_k}
\end{aligned} \tag{32}$$

$\mathcal{L}^B(\cdot)$ は $BPl^{b_{k-1}}$ と後ろ向きの遷移行列を引数とする関数である．$\mathcal{L}^B(\cdot)$ についての詳細なアルゴリズムは Yang ら [7] を参照すること．また $PL_0^{b_k}$ は ε となる．

次に，FPl^τ について考える．3.2.2 節と同様に，$r^{\boldsymbol{f}}$ の個数を m とし，$r^{\boldsymbol{f}} = r^{f_{1:m}}$ と表すと，FPl^τ は以下のように変形できる．

$$FPl^\tau = \sup_{r^{f_{1:m}}, l^\tau, l'^\tau} \frac{1}{d_s(l^\tau, l'^\tau)} \log \frac{\sum_{l^{f_1}} P(r^{f_{1:m}}|l^{f_1})P(l^{f_1}|l^\tau)}{\sum_{l'^{f_1}} P(r^{f_{1:m}}|l'^{f_1})P(l'^{f_1}|l'^\tau)} \tag{33}$$

ここで，$\sup_{r^{f_{k:m}}, l^\tau, l'^\tau} \frac{1}{d_s(l^k, l'^k)} \log \frac{P(r^{f_{k:m}}|l^{f_k})}{P(r^{f_{k:m}}|l'^{f_k})}$ を FPl^{f_k} と表すと，FPl^{f_k} は以下のように再帰的に計算できる．

$$\begin{aligned}
&FPl^{f_k} = \\
&\sup_{r^{f_{k+1:m}}, l^{f_k}, l'^{f_k}} \frac{1}{d_s(l^{f_k}, l'^{f_k})} \\
&\log \frac{\sum_{l^{f_{k+1}}} P(r^{f_{k+1:m}}|l^{f_{k+1}})P(l^{f_{k+1}}|l^{f_k})}{\sum_{l'^{f_{k+1}}} P(r^{f_{k+1:m}}|l'^{f_{k+1}})P(l'^{f_{k+1}}|l'^{f_k})} \\
&+ \sup_{r^{f_k}, l^{f_k}, l'^{f_k}} \frac{1}{d_s(l^{f_k}, l'^{f_k})} \log \frac{P(r^{f_k}|l^{f_k})}{P(r^{f_k}|l'^{f_k})} \\
&= \mathcal{L}^F(FPl^{f_{k+1}}) + PL_0^{f_k}
\end{aligned} \tag{34}$$

$\mathcal{L}^F(\cdot)$ は $\mathcal{L}^B(\cdot)$ と同様に考えることができる．

これらの計算を行うことで，購入する頂点集合が与えられた時，時刻 τ において実際に保証されているプライバシの度合いを考えることができる．

4.3.2　プライバシ目線での評価の変化を観る

ここでは，前節で求められたプライバシパラメータを価値の評価に使うことを考える．以下のように，購入する頂点の時刻の集合が $\boldsymbol{t}$ である時，時刻 τ のプライバシの損失度合い

$TPl^{\tau}(\boldsymbol{t})$ を定義する．

$$TPl^{\tau}(\boldsymbol{t}) = \sup_{l^{\tau}, l'^{\tau}, r^{\boldsymbol{t}}} \frac{1}{d_s(l^{\tau}, l'^{\tau})} \log \frac{P(r^{\boldsymbol{t}}|l^{\tau})}{P(r^{\boldsymbol{t}}|l'^{\tau})} \tag{35}$$

この時，U，V，X，T，$\boldsymbol{t}_U$，$\boldsymbol{t}_X$，$\boldsymbol{t}_V$ を 4.1.3 節と同様に定義し，価格 $p_3(X|V)$ を以下のように定義する．

$$p_3(X|V) = \sum_{\tau \in T} \{TPl^{\tau}(\boldsymbol{t}_X, \boldsymbol{t}_V) - TPl^{\tau}(\boldsymbol{t}_V)\} \tag{36}$$

この指標は，新たに点を販売することで，元の経路情報に対するプライバシの保証度合いがどれほど大きくなっているかを算出する指標である．(32)(34) より，公開する点が多くなればなるほど各点に対するプライバシパラメータは大きくなることが分かるので，この指標は単調増加となる．また，実際の位置情報を使うことなく計算ができるので，販売者の視点からプライバシロスを考え価値をつけるための良い手法と言える．問題はアルゴリズムである．本論文では $\mathcal{L}^B(\cdot), \mathcal{L}^F(\cdot)$ に対して離散値を入力とする時の効率的なアルゴリズを考案できていない．総当たりでこの値を求めることもできるが，計算量が評価に関するボトルネックとなる．

5 実　　験

この章では，前章までに述べられた 3 つの評価指標のうち，平均情報量と誤差期待値に関して実際にデータを入力として実験を行う．経路情報に対して曖昧化する時の ε を変化させることによって，各評価がどのように変化していくのかを示す．

5.1 実験の準備

実験に先立ち，まずは実在の都市を用いて道路ネットワークを作成した．OpenStreetMap [8] を用いて京都市の一部の道路情報を取得し，そのデータを整形することによって，京都市の一部の道路ネットワークを作成した．移動情報として用いるデータの間隔が約 2 分であることを踏まえ，グラフの重みとして用いる距離は最大でも 150m 程度に収まるように整形した．頂点数は最終的に 536 となった．

次に，緯度経度として記録されている約 700 人の GPS データ [9] を道路ネットワークの範囲内の点において最寄りの頂点へと変換した．こうして，約 500 人，約 20000 点の経路情報が得られた．

この得られた経路情報から遷移行列を作成した．具体的には，各経路情報に対して，ある頂点に存在していたときに次の時刻に存在する頂点を集計し，正規化を行った．集計が十分でなかった頂点に関しては，距離的に近い頂点に対して一様に遷移するとした．

こうして，プログラムへと入力できるデータが作成できた．次節では，実際にプライバシパラメータを設定し価値の変化を可視化していく．

5.2 実験の結果

本実験では，1 つの例を上げて実験の結果を示す．経路情報として，62 の連続した位置情報をもつ経路情報を用いた．この経路情報について，まずは何も点を購入していない状態から，点を 1 つ購入することでどれほどの効用を得られることができるのかを図示する．以下，図中では縦軸は平均情報量を表し，横軸は時刻を表す．また，青線は全く点を購入していない時の平均情報量の曲線を表し，橙線は時刻 30 の点を購入した時の平均情報量の曲線を表す．図 7 では，プライバシパラメータを 1 に設定している．これは，時刻 30 周辺での平均情報量がほぼ 0 となってしまっている．このことは，攻撃者の予測がほぼ 1 点に定まっていることを示している．頂点間の距離に対して ε が大きすぎると，このような事態に陥って適切な価値がつけられなくなる (つけられた価値が非常に大きくなってしまう)．実際，(19) で定義している価値 p_1 は約 2 億 7000 万と後述の 2 例と比べて非常に大きくなっている．図 8 では，プライバシパラメータを 0.001 に設定している．これは，点を購入しても平均情報量の減少量が微々たるものであり，点の購入によりほとんど対価が得られないことを示している．これが起きるのは，プライバシパラメータによる曖昧化が強すぎて，遷移行列による分布変化よりも不確かさを大きく減らすことができていないためである．頂点間の距離に対して ε が小さすぎると，このような事態に陥って適切な価値がつけられなくなる (つけられた価値が非常に小さいかマイナスになってしまう)．実際，価値 p_1 は約 0.11 と後述の 2 例に比べると微々たるものとなってしまっている．

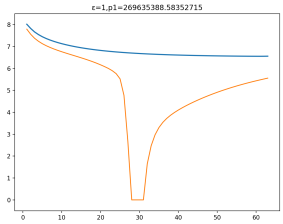

図 7　$\varepsilon = 1$ の平均情報量

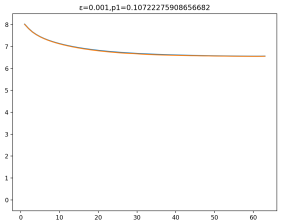

図 8　$\varepsilon = 0.001$ の平均情報量

図 9 ではプライバシパラメータを 0.1，図 10 では，プライバシパラメータを 0.01 に設定している．これらは，先述の 2 つのような事態に陥らずに，適切な範囲の価値をつけられている．実際，p_1 は $\varepsilon = 0.1$ で約 15，$\varepsilon = 0.01$ で約 4.2 となっている．このように，プライバシの保証度合いがしっかりと価値に反映されており，販売のためには適切なパラメータを設定する必要があること確かめられた．

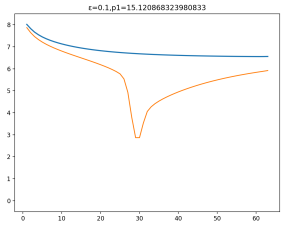

図 9　$\varepsilon = 0.1$ の平均情報量

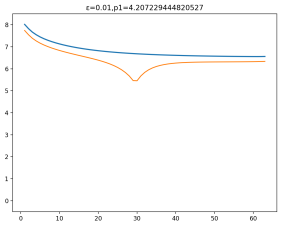

図 10　$\varepsilon = 0.01$ の平均情報量

図 11 と図 12 はさらに 10，20，40，50，60 の点を購入した場合の平均情報量の変化を示している．青線は時刻 30 の点のみを購入した時の平均情報量の曲線を表し，橙線は時刻 10，20，30，40，50，60 の点を購入した時の平均情報量の曲線を表す．点を保有している状態で更に点の購入した時には，遷移行列による情報を曖昧化による情報が更新していくこととなる．図 11 では適度にプライバシパラメータが大きいため更に点を購入することで価値 ($p_1 = 74.946...$) が生まれているが，図 12 ではプライバシパラメータが小さいために複数点の購入では価値 ($p_1 = 8.181...$) があまり生まれていない．また，何も点を購入

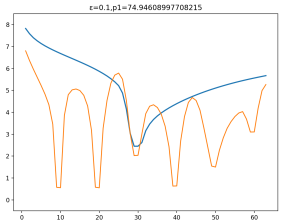

図 11　$\varepsilon = 0.1$ で複数点購入

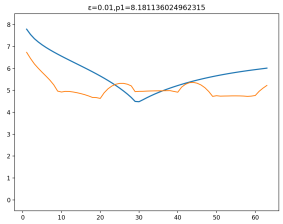

図 12　$\varepsilon = 0.01$ で複数点購入

していない状態から点を 1 つ購入した時の誤差期待値の変化を図示する．図中では縦軸は誤差期待値を表し，横軸は時刻を表す．青線は全く点を購入していない場合の誤差期待値の曲線を表し，橙線は時刻 30 の点を購入した場合の誤差期待値の曲線を表す．図 13 は $\varepsilon = 0.1$ の correctness の変化を表し，図 14 は $\varepsilon = 0.01$ の correctness の変化を表している．時刻 30 周辺では図 14 よりも図 13 の方が誤差が小さくなっているが，それ以外の時刻については一概にそうとも言えない．プライバシパラメータを小さくし強くプライバシを保証しているのにもかかわらず，時刻によってはかえって誤差が大きくなる場合がある．これは，correctness の性質によるものである．プライバシを強く保証した場合には分布に隔たりがなくなり，予測を外すとしてもそこまで誤差期待値が大きくならないためだと考えられる．

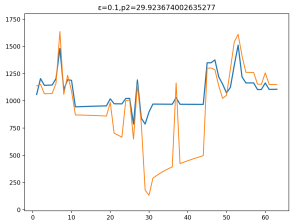

図 13　$\varepsilon = 0.1$ の誤差期待値

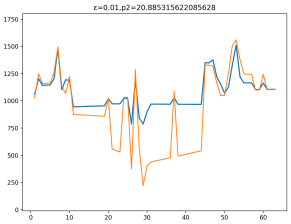

図 14　$\varepsilon = 0.01$ の誤差期待値

6　おわりに

本研究では 2 つの観点，3 つの指標を用いて経路情報の評価について考えた．購入者視点では不確かさが少ない方がより確実な予測ができるため高い対価が払うことになる一方で，販売者視点では実際の正解データとの誤差やプライバシ損失が重要な指標となってくるので，リスクが大きい方が高い対価を求めることとなる．前章の結果を踏まえ，購入者目線での指標では曖昧化時のプライバシ保証の度合いを忠実に表していることが分かった．そのため，この指標から出力される値がどれほどの値かを見ることによって，販売者は自分が適切なプライバシパラメータを設定できているかが分かるようになった．今まで，プライバシパラメータは，数値では与えられても直感的にどのような結果を生むかを可視化する物が少なかった．しかし今回，こうしてプライバシパラメータが与える不確かさの減少量を可視化することで，安全に経路情報を売買するための手法の提案ができた．また，本研究では HMM 以外から得られる知識はないという前提の元で行ったが，経路情報には様々な関連する情報が溢れている．本研究をもとに，更に現実に近い攻撃者モデルを作成し経路情報の評価を行うこと，また，離散データに対するプライバシ損失の効率的な計算手法を確立することが今後の課題である．

7　謝　　辞

本研究は JSPS 科研費基盤研究 (S) No. 17H06099, (A) No. 18H04093, 若手研究 No. 19K20269 の助成を受けたものです．

文　　献

[1] Shun Takagi, Yang Cao, Yasuhito Asano and Masatoshi Yoshikawa: Geo-Graph-Indistinguishability: Protecting Location Privacy for LBS over a Road Network, Conference on Data and Applications Security and Privacy (DBSec'19), Charleston, SC, USA, July 15-17, 2019.

[2] Reza Shokri, George Theodorakopoulos, Jean-Yves Le Boudec, and Jean-Pierre Hubaux: Quantifying Location Privacy, 2011 IEEE Symposium on Security and Privacy, 2011 IEEE, Oakland, California, USA, May 22-25, 2011, pp.247-262 (2011).

[3] Cvrcek D, Kumpost M, Matyas V, and Danezis G: A study on the value of location privacy, Proceedings of the 2006 ACM Workshop on Privacy in the Electronic Society, WPES 2006, Alexandria, VA, USA, October 30, 2006, pp.109-118 (2006).

[4] Andr'es M E, Bordenabe N E, Chatzikokolakis K, and Palamidessi C : Geo-indistinguishability: differential privacy for location-based systems, 2013 ACM SIGSAC Conference on Computer and Communications Security, CCS' 13, Berlin, Germany, November 4-8, 2013, pp.901-914 (2013).

[5] Dwork, C.: Differential privacy: A survey of results, International Conference on Theory and Applications of Models of Computation, Springer, pp.1-19 (2008).

[6] Lawrence R. Rabiner: An introduction to hidden Markov models, IEEE ASSP Magazine, Juang, pp.4 - 15 (1986).

[7] Yang Cao , Masatoshi Yoshikawa , Yonghui Xiao, and Li Xion: Quantifying Differential Privacy in Continuous Data Release Under Temporal Correlations, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 31, NO. 7, JULY 2019, pp.1281-1295 (2019).

[8] Open Street Map (https://www.openstreetmap.org)

[9] Kasahara, Hidekazu, et al. "Business model of mobile service for ensuring students' safety both in disaster and non-disaster situations during school trips." Information and Communication Technologies in Tourism 2014. Springer, Cham, 2013. 101-114.