Vol.2015-DPS-165 No.6
2015/12/10

位置プライバシ保護のための嗜好の保存度合いを考慮したダミー生成手法の検討

水野 聖也[1]　原 隆浩[1]　Xing Xie[2]

概要：近年注目を集めている位置情報サービスでは，自身の位置に基づいた有用なサービスが受けられる一方，ユーザの位置情報をサービスプロバイダに送信するという性質上，プライバシに関する問題が指摘されている．この問題に対し，筆者らの研究グループでは，先行研究において，ユーザの訪問場所の履歴から頻出するパターンを抽出し，それに従ってダミーを遷移させることでユーザとダミーの識別を困難にし，ユーザの訪問場所の傾向を知っている攻撃者に対して効果的にユーザの位置情報を匿名化する手法を提案した．しかし，この手法では，ダミーがユーザの行動履歴の頻出パターンに従って動作しているため，それらが観測されることによりユーザの大まかな嗜好が露見してしまう危険性がある．一方，こういった嗜好情報は公開することでサービスのパーソナライズに利用することができるという側面もあるため，嗜好をどの程度公開するかはユーザの要求に応じて制御できることが望ましい．そこで本稿では，ユーザの訪問場所のカテゴリシーケンスのサポート (支持度) を嗜好と定義し，この嗜好情報の保存度合いをユーザの要求に応じて制御可能なダミー生成手法について検討する．

1. はじめに

GPS を搭載した端末の普及に伴い，位置情報サービスが数多く展開されるようになった．位置情報サービスでは，ユーザが自身の位置をサービスプロバイダに通知することによりその位置に対応したサービスが受けられるが，送信した位置情報の管理はサービスプロバイダに委ねられるため，サービスプロバイダが攻撃を受けたり，サービスプロバイダ自身によるデータの売買によって第三者に位置情報が露見し，ユーザの位置プライバシが侵害される危険性が指摘されている．

こうした位置情報サービス利用におけるユーザの位置プライバシの保護を目的とした研究は数多く行われており，その中の一つとして，ダミーの位置情報を用いたユーザの位置曖昧化手法がある．この手法では，ユーザがサービスプロバイダに位置情報を送信する際に，ユーザの実際の位置情報と共に複数のダミーの位置情報を送信する．これにより，送信された情報の中から実際のユーザの位置を一意に特定することが困難になり，ユーザの位置が曖昧化される．筆者らの研究グループでは，先行研究において，ユーザの訪問履歴中で頻出する遷移パターンに従ってダミーを移動させることで，ユーザとダミーの識別を困難にし，ユーザの訪問場所の傾向を知っている攻撃者に対して効果的にユーザの位置情報を曖昧化する手法を提案した [11]．しかし，この先行研究では，ダミーがユーザの行動履歴の頻出パターンに従って動作しているため，それらが観測されることでユーザの大まかな嗜好情報が露見してしまう．一方，嗜好情報には，スポットの推薦等，位置情報サービスのパーソナライズに利用可能であるという側面もあるため，一概に曖昧化すればよいとはいえず，ユーザの要求に応じて制御できることが望ましい．

本稿では，チェックインサービスのように，ユーザが実際にある場所を訪問した際にサービスを利用し，次の場所へ移動するというモデルを想定する．その下で，訪問場所のカテゴリシーケンスのサポート (支持度) をユーザの嗜好情報と定義し，この嗜好情報の保存度合いをユーザの要求に応じて制御可能なダミー生成手法を提案する．提案手法では，サービスプロバイダから観測可能なユーザとダミーの位置情報系列から嗜好を解釈する方法を一意に定義し，その方法に基づいて構築される嗜好情報と実際のユーザの嗜好情報の差分を計算する．その差分を低減できるようなシーケンスに従うダミーをユーザの嗜好の保存度に対する要求に応じた数生成し，その他のダミーにはユーザの嗜好と異なる動作をさせることで，ユーザの嗜好の保存度合い

[1] 大阪大学 大学院情報科学研究科
Graduate School of Information Science and Technology, Osaka University
[2] マイクロソフトリサーチアジア
Microsoft Research Asia

を制御する．

ダミーの生成の際には，ダミーが広範囲に分布するように位置を決定することで，ユーザの位置が広範囲に曖昧になるようにし，さらにユーザ及びダミー間で同時刻に同じ場所に訪問でさせるようにすることで，経路の交差を発生させ，一時的にユーザの位置が特定された場合にも連続的にサービス利用が追跡されないようにする．

以下では，2 章で関連研究を説明し，3 章で嗜好の保存度合いを制御可能なダミー生成手法について述べ，最後に 4 章で本稿のまとめと今後の課題について述べる．

2. 関連研究

本章では，ユーザの位置プライバシ保護を目的とした代表的なアプローチを述べる．

文献 [1][4] では，ユーザの実際の位置と異なる代替地点を計算し，その位置情報を用いてサービス利用を行う手法を提案している．この手法では，実際のユーザの位置情報を送信しないため，ユーザの実際の位置情報は保護されるが，代替地点として実際のサービス利用の地点から離れた地点を用いてしまうと，ユーザの位置に無関係なサービスが提供されてしまうため，サービスの質の低下を招く．そのため，ユーザのおおよその位置は推測しやすく，十分な位置曖昧性を確保しつつサービスの質を維持するのは難しい．

文献 [2][9] では，Mix Zone と呼ばれる全ユーザに対してサービスを禁止する領域を設定し，その領域に同時に入ったユーザ間で ID を入れ替えることで長期的なサービス利用の追跡を防止する手法を提案している．文献 [3][6] で提案されている手法では，匿名化サーバと呼ばれるミドルウェアを導入し，ユーザは匿名化サーバに位置情報を送信する．匿名化サーバは，受信した位置情報に基づき，k 人以上のユーザが存在するような領域を算出し，その領域をクエリとしてサービスプロバイダに送信する．これにより，ユーザの位置情報を k 匿名化することが可能となる．上記の手法は，いずれも第三者サーバが完全に信頼できるという想定のもとに成り立っているため，実環境で用いるのは困難である．

文献 [5][7][8][10] では，自身の位置情報とともに架空の位置情報であるダミーをクエリに付与して送信することで，ユーザの実際の位置の特定を困難にする手法が提案されている．この手法では，ユーザはサービスプロバイダに送信した全ての位置情報に対する応答を受信するが，その中から自身の位置に基づくもののみを選択することにより，自身の位置が特定されることを防止しつつ，自身の実際の位置に基づいたサービスを受けることができる．また，ダミーの生成は全てユーザの端末上で行うことができるため，第三者サーバの存在を必要としない．そのため，本稿では，ダミーを用いた手法を採用する．

筆者らの研究グループでは，先行研究において，ユーザの訪問履歴中で頻出するパターンに従ってダミーを遷移させることでユーザとダミーの判別を困難にし，ユーザの訪問場所の傾向を知っている攻撃者に対して効果的にユーザの位置情報を曖昧化する手法を提案した [11]．

しかし，この手法では，ユーザの嗜好の保存性については考慮されていないため，ダミー及びユーザの動きから観測できる嗜好情報にユーザ自身の嗜好情報との乖離が生じてしまう可能性があり，サービスのパーソナライズ等に用いる場合にその質を低下させる要因となる．また，ダミーがユーザの行動履歴中の頻出パターンに従っているため，それらが観測されることで，ユーザの大まかな嗜好が露見してしまう危険性がある．それに対し提案手法では，位置プライバシ保護に関する要求を考慮しつつも，ユーザが高い嗜好の保存度を要求した場合，ダミー及びユーザの動きから観測できる嗜好情報とユーザ自身の嗜好情報の差分を計算し，その差分を低減するようにシーケンスを選択することで，ユーザの嗜好情報をより正確に保存することができる．さらに，ユーザが低い嗜好の保存度を要求した場合にも，ユーザの嗜好に従わないダミーを混入することで嗜好を保護することが可能である．

3. 嗜好情報の保存度合いを制御可能なダミー生成手法

本章では，まず想定環境を説明した後，実環境でダミーを生成する際に考慮すべき制約とユーザの位置プライバシに関する要求について述べる．次に，本稿におけるユーザの嗜好とその保存度の定義について述べ，最後に，それらの定義に従って，プライバシ要求を満たしつつユーザの嗜好情報の保存度を制御するためのダミー生成手法について述べる．

3.1 想定環境

本稿では，チェックインサービスのように，ユーザがある場所を訪問した際に位置情報サービスを利用するというモデルを想定する．ユーザは各訪問場所を訪問した際に，自身の位置情報とダミーの位置情報をサービスプロバイダに送信し，位置情報サービスを利用する．各訪問場所では最小 T_m 秒から最大 T_M 秒までの間停止し，訪問場所間は最短路を通って停止地点間を移動するものとしている．さらに本稿では，ユーザの訪問場所，及びその到着時刻が事前に取得できるものと想定している．このような想定は，実環境では必ずしも妥当ではないが，ユーザの過去の行動履歴から予測するなど，ある程度の精度で予測可能な場合も多い．予測精度が低い場合の対応については，今後の課題と考え，本稿では対象としないが，先行研究 [10] によると，到着時刻に関しては，ある程度のずれが生じても，軽微な修正によって対応できることが明らかになっている．

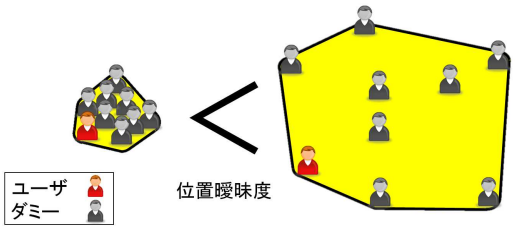

図 1　匿名領域

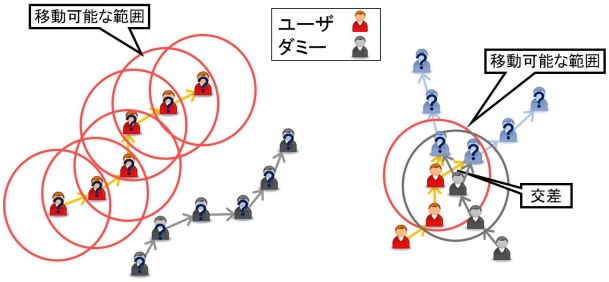

図 2　追跡可能性

ユーザが所有するモバイル端末上のシステムは，地図情報とユーザの訪問履歴を保持しており，ユーザやダミーが通っても不自然でない道路，訪問場所，訪問場所のカテゴリ，及びユーザの嗜好情報をすべて把握しているものとする．

3.2　実環境でダミーを生成する際に考慮すべき条件

短時間にサービス利用が繰り返される場合，サービス利用間のダミーの位置関係において，その時間における地理的到達可能性を考慮する必要がある．例えば，あるユーザが一度サービスを利用してから三分後に再度サービスを利用する場合を考える．この際，新しいサービス利用において，ダミーを直前のサービス利用のどの地点からも三分以内に到達できない箇所に設置してしまうと，その位置情報がダミーであると容易に特定されてしまう．提案手法では，地図情報に基づき訪問地点間の道のりを計算することで，各サービス利用において，直前のサービス利用時に生成したダミーの位置から到達可能な位置に次のダミーを生成することを保証する．

3.3　位置プライバシに関する要求

3.3.1　匿名領域

ユーザの位置プライバシを保護するためには，ユーザとダミーの位置情報の中からユーザのものを一意に特定できないようにすることだけでなく，どの程度の大きさの領域に位置情報が曖昧化されているかも重要である．例えば図 1 で，左のようにダミーをユーザの周りに密集して配置した場合，ユーザの特定はできないが，ユーザ及びダミーの存在範囲が小さいため，ユーザが存在する可能性のある領域が小さい範囲に絞り込めてしまう．そのため右の図のようにダミーを広範囲に分散して配置し，位置曖昧度を大きくする必要がある．本稿では，Lu ら [7] の定義に基づき，ユーザとダミーで形成される凸包領域を匿名領域と定義し，この大きさが定常的にユーザの要求する大きさを達成できるようにダミーを配置する．

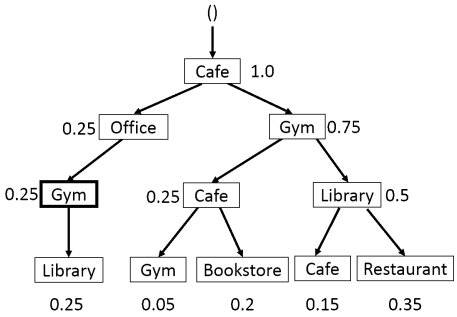

図 3　ユーザの嗜好を表す木

3.3.2　追跡可能性

短時間で連続してサービスを利用する場合，追跡可能性に対する配慮も必要である．例えば図 2 で，左のように，あるサービス利用時刻におけるエンティティ (ユーザもしくはダミーを指す) の位置から次のサービス利用時刻において到達可能な範囲に対応する位置情報が 1 つしかない場合，前後のサービス利用の対応関係が一意に定まり，特定のエンティティのサービス利用が追跡できてしまう．この性質を追跡可能性と呼ぶ．追跡可能性が高い場合，何らかの原因で一度ユーザが特定された場合に，前述の対応付けによって前後のサービス利用もユーザのものであると特定されてしまい，より多くの情報の流出を招く．そこで提案手法では，右の図のようにエンティティ間に，同時刻に同じ場所を訪問させる (共有地点を設定する) ことにより，追跡可能性を低下させるようにする．

3.4　嗜好の定義

本稿では，ユーザの嗜好として，各訪問場所に付与されている意味情報であるカテゴリのシーケンスを用い，それを図 3 のように木構造で表現する．この木構造では各ノードがシーケンスに対応し，例えば図中の太線で囲った Gym のノードは $Cafe \rightarrow Office \rightarrow Gym$ という遷移に対応する．各ノードに付与されている数値はサポート値を表し，その値は発生したトラジェクトリの総数 N_{traj} とシーケンス S の発生度数 $n(S)$ を用いて次式で計算される．

$$Sup(S) = \frac{n(S)}{N_{traj}} \qquad (1)$$

3.5　嗜好の保存度の定義

ダミーを用いた位置情報サービス利用の仕組みは図 4 のようになっており，ユーザは実際の位置情報に加え，ダミーの位置情報を複数送信する．そのため，サービスプロバイ

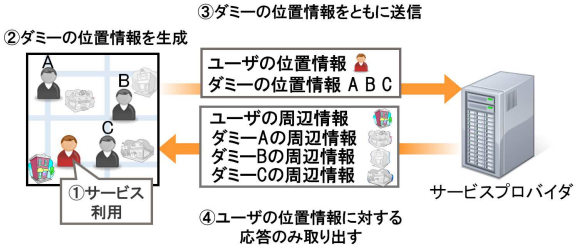

図 4　ダミーを伴う位置情報サービス利用

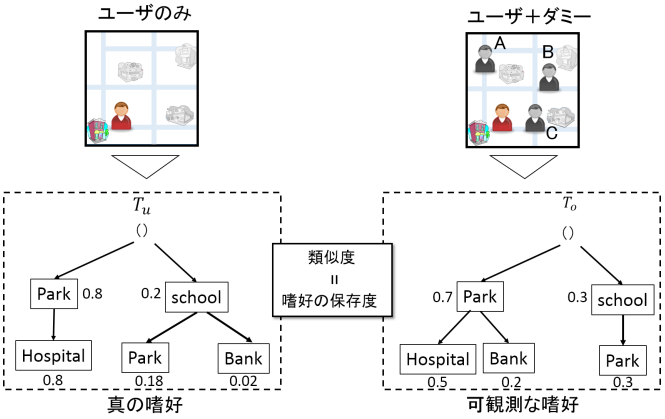

図 5　真の嗜好と可観測な嗜好

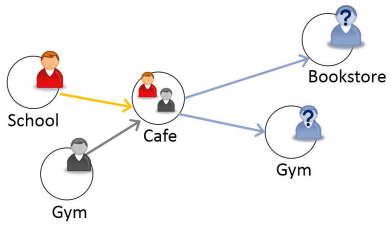

図 6　カテゴリシーケンスが一意に特定できない例

ダ，及び攻撃者が観測できるのは，ダミーと，ユーザが混合した位置情報系列となる．サービスプロバイダは，この情報を観測することで，ユーザの嗜好を取得し，攻撃者も同様にその情報に基づいて，ユーザの特定を試みる．以降では，このユーザとダミーの混合された位置情報系列から構築される嗜好を可観測な嗜好と呼び，T_o で表す．これに対して，ユーザのみの行動に基づく嗜好をユーザの真の嗜好と呼び，以降 T_u と表記する．ダミー生成時に，真の嗜好におけるサポート値を特に考慮しない場合，この 2 つの木構造の乖離は大きくなり，嗜好は保護されるが，パーソナライズ等に用いることは難しくなる．そこで，本稿ではこの点を考慮し，ユーザの嗜好の保存度，すなわちこの二つの嗜好の木構造をどの程度類似させるかをユーザの要求に応じて制御できるダミー手法を提案する．

ここで，シーケンス S の T_u におけるサポート値 $Sup_u(S)$，T_o におけるサポート値 $Sup_o(S)$，T_u, T_o に含まれるシーケンスの集合をそれぞれ $SSet_u = \{S|S \in T_u\}, SSet_o = \{S|S \in T_o\}$ とし，嗜好の保存度 $Sim(T_u, T_o)$ を次式で定義する．

$$Sim(T_u, T_o) = 1 - \sum_{S \in \{SSet_u \cup SSet_o\}} \frac{|Sup_o(S) - Sup_u(S)|}{|SSet_u \cup SSet_o|} \quad (2)$$

ただし，一方の木にしか含まれないシーケンスはそのサポート値を 0 として扱う．

サービスプロバイダが観測したユーザとダミーの位置情報系列から T_o を更新する際に，交差によって，各エンティティのカテゴリシーケンスが一意に決定できない場合が生じる．その際には，とりうるシーケンス両方にシーケンスの発生度数を等配分することによって T_o の更新を行うものと定める．図 6 の例では，2 番目の訪問地点である $Cafe$ において地点の共有が発生したことで，1 番目の School でのサービス利用に対応する 3 番目のサービス利用が Bookstore と Gym のどちらであったか一意に対応付けられない．そのため，取りうるシーケンス $School \rightarrow Cafe \rightarrow Bookstore$ 及び，$School \rightarrow Cafe \rightarrow Gym$，に，度数 1 を等配分し，それぞれ 0.5 ずつ発生度数を加算しサポート値を更新するものとする．

3.6　嗜好の保存度を制御可能なダミー生成手法

提案手法では，ユーザの行動プラン $trajectory_0$，要求ダミー数 n，要求匿名領域 A_r，ユーザの真の嗜好 T_u，可観測な嗜好 T_o，及び嗜好の要求保存度 $\alpha = [0, 1]$ を入力として用いる．T_o に関しては，各エンティティの交差の判定基準を共有していれば，ユーザ側でもシミュレートすることが可能なので，これも同様にダミー生成の際に用いられるものとしている．

まず，入力に基づき生成するダミーを $\lfloor \alpha n \rfloor$ 個と，$\lceil (1-\alpha)n \rceil$ 個の 2 つのグループに分割する．以降では，前者をグループ A，後者をグループ B とする．グループ A のダミーは，ユーザの嗜好に従って動作させ，ユーザの嗜好を知っている攻撃者に対する有効な交差の設定と，ユーザの嗜好の保存性の確保を行う．グループ B のダミーは，ユーザの嗜好を考慮せず，全体の交差や匿名領域を調整するように動作させる．これにより，ユーザの嗜好と異なるダミーが要求に応じて混入され，ユーザの嗜好が曖昧化される．各ダミーはユーザがサービス利用を行う各時刻にどの訪問場所に存在するかを決定することにより，1 つずつ行動を確定させていく．最初のダミーはユーザの行動プランのみを考慮して決定し，以降の i 番目のダミーは $i-1$ 番目までで生成された全てのダミー及びユーザの経路を考慮して決定する．ダミーの生成はグループ A→ グループ B の順序で行う．以下でそれぞれのダミーの生成方法について述べる．

3.6.1　グループ A のダミーの生成手順

グループ A のダミーは，ユーザの嗜好を知っている攻撃

Algorithm 1 グループ A のダミーの生成手順

Require: ユーザの行動プラン $traj_0$, ダミー数 n_A, 要求匿名領域 A_r, 真の嗜好 T_u, 可観測な嗜好 T_o

```
1: //確定したエンティティのシーケンスによる更新をシミュレート
2: T_cpに T_oをコピー
3: trajs ← {traj_0}                ▷ 確定したエンティティの経路リスト
4: while 生成したダミー数が n_A 未満 do
5:     //シーケンス S と共有地点 SP の組 < S, SP > のスコア
6:     seqScoreSet ← {}
7:     for S ∈ SSet_u do
8:         for 取りうる全ての組 < S, SP > do
9:             //シーケンス S と共有地点の組 SP のスコアリング
10:            Score_s ← calcSScore(< S, SP >, T_u, T_o, trajs)
11:            seqScoreSet.push({< S, SP >, Score_s})
12:        end for
13:    end for
14:    //実際に到達可能な経路の作成
15:    trajectoryScoreSet ← {}
16:    for {< S, SP >, Score_s} ∈ seqScoreSet do
17:        //到達可能な経路を生成
18:        //要求匿名領域達成の観点からスコアリング
19:        (traj, score_traj) ← createTraj(trajs, < S, SP >)
20:        trajectoryScoreSet.push((traj, score_traj))
21:    end for
22:    //スコアが最大となる経路を確定させる
23:    (traj_max, score_max) ← max(trajectoryScoreSet)
24:    trajs.push(traj_max)
25:    T_cp.update(traj_max)
26: end while
```

者に対し効果的に追跡可能性を低下させるための交差の設定と，ユーザの嗜好を保存するためのシーケンスの生成を行う．本稿では，ユーザの嗜好を知っている攻撃者は可観測な嗜好 T_o のに基づいてユーザの実際の位置を推測するものと想定する．このとき，ユーザの追跡可能性を効果的に低下させるには，ユーザの嗜好に従って移動するユーザらしいダミーで交差を発生させる必要がある．一方，ユーザの嗜好情報をより正確なものにするためには，T_u と T_o の類似性が高まるようにシーケンスを発生させる必要がある．そこでまず，真の嗜好 T_u に含まれるすべてのシーケンス集合 $SSet_u$ 中の各シーケンスに対し上記 2 つの観点からスコアリングを行い，発生させるカテゴリシーケンスの優先度を決定する．次に，決定した優先度に基づき，到達可能性，及び匿名領域に対する要求を考慮して実際の経路を決定する．これを n_A 回繰り返すことによってグループ A のダミーの行動を決定する．Algorithm1 にグループ A のダミー生成全体の手順を示す．

以降では，カテゴリシーケンスのスコアリング方法と，それを基に実際に経路を作成する手順について詳述する．

3.6.1.1 カテゴリシーケンスのスコアリング

生成するカテゴリシーケンスの優先度を決定するため，前述した交差の効果と嗜好の保存性の観点から $SSet_u$ 中の各シーケンスと，それに対して設定可能な共有地点の組 $< S, SP >$ にスコアリングを行う (Algorithm1 10 行目)．生成するダミーの従うシーケンスには，嗜好の保存度を向上させ，かつ効果的に追跡可能性を低下させられるものを選択することが望ましい．そこで，カテゴリシーケンスと，そのシーケンス S において設定する共有地点 SP の組 $< S, SP >$ を次のスコア $Score_s$ で評価する．

$$Score_s(< S, SP >) = \beta Score_{pref}(S) + (1-\beta) Score_{cross}(< S, SP >) \quad (3)$$

ここで，$Score_{pref}$ は嗜好の保存度に関するスコア，$Score_{cross}$ は交差の効果に関するスコアである．β は上記の重要度の比率を決定する定数で $[0, 1]$ の値をとる．本稿ではこの値について特に規定しないが，今後のシミュレーション実験によって適切な値を求めるか，ユーザの入力として決定することを検討している．

ここで，(3) 式の $Score_{pref}$ と $Score_{cross}$ の内容についてそれぞれ詳細に検討する．まず $Score_{pref}$ について，可観測な嗜好 T_o の T_u に対する差分を小さくするようにシーケンスを決定することで，サポート値の正確性を確保することを考える．そのためには，T_u におけるサポート値に対し，T_o でのサポート値が低くなっているシーケンスを優先的に選択する必要がある．そこで，$Score_{pref}$ を次式で定義する．

$$Score_{pref}(S) = Sup_u(S) - Sup_o(S) \quad (4)$$

続いて S_{cross} について考える．ユーザの嗜好を知っている攻撃者に対し，効果的に追跡可能性を低下させるには，交差前後のサブシーケンスの組み合わせによって形成されるシーケンスがいずれもユーザらしいシーケンスである必要がある．そのため，共有地点は共通のプレフィックス上に設定する．ここで，プレフィックスとはカテゴリシーケンスの先頭を含むサブシーケンスのことを指し，例えば図 7 において，$id = 1$ のエンティティと $id = 2$ のエンティティ間で共通のプレフィックスは $B \to E \to F$ である．このように共有地点を設定することで，2 つのエンティティが取りうるシーケンスが，$B \to E \to F \to C$ と $B \to E \to F \to A$ で，いずれも T_u に基づいたユーザらしい遷移となり効果的に追跡可能性を低減できる．また，嗜好の保存の観点からも，このように共通のプレフィックス上で交差を設定することで，3.5 節における交差時の嗜好の更新方法の定義により，実際に発生した両シーケンスが 1 回ずつ発生したと計上され，意図しないシーケンスが嗜好として計上されてしまうことを防止することができる．交差を設定する時刻について，ある時間帯に交差が集中するとその時間帯にはエンティティ間の距離が小さくなるため，匿名領域が小さくなる可能性が高い．そのため，設定される共有地点数は，各時刻で均一に設定されることが望ましい．ここで，共有地点 SP を地点を共有させる対象のエンティティ e_i と時刻 t の組 $< e_i, t >$，$n_{share}(t)$ を時刻 t

	id = 0	id = 1	id= 2	共有地点数
t = 0	Cafe 共有	Cafe	Cafe	1
t = 1	Gym	Gym	Gym	0
t = 2	Cafe	Library	Library	0
t = 3	Bookstore	Cafe	Restaurant	0

図 7 カテゴリシーケンスの決定

に設定済みの共有地点数，e_i と共有地点を有するエンティティの集合を SES_i，及び e_i が従うカテゴリシーケンスを S_{e_i} として，交差の効果を表すスコア S_{cross} を次式で定義する．

$$Score_{cross}(< S, < e_i, t >>) = \frac{\delta}{1 + n_{share}(t)} \times \frac{Sup_o(S)}{1 + \sum_{e_k \in SES_i} Sup_o(S_{e_k})} \quad (5)$$

ここで，δ は SP が共通のプレフィックス上で設定されている場合に 1，そうでない場合に 0 となる変数である．上式の定義により，設定済み共有地点数が少ない時刻に，ユーザらしいシーケンスで，それまでにユーザらしい動きをするエンティティと共有地点が設定されていないエンティティに対し交差を設定する場合に S_{cross} は高い値となる．図 3 の木構造を T_o として用い，図 7 の $id = 2$ のダミーのシーケンス $S = Cafe \rightarrow Gym \rightarrow Library \rightarrow Restaurant$ で $id = 0$ のエンティティと共有地点を設定する場合，(5) 式において $\delta = 1$ となるのは，$SP =< e_0, 0 >, < e_0, 1 >$ のときで，それぞれスコアは $Score_{cross}(S, < e_0, 0 >) = \frac{1}{1+1} * \frac{0.35}{1+0.15} \approx 0.15$，$Score_{cross}(S, < e_0, 1 >) = \frac{1}{1+0} * \frac{0.35}{1+0.15} \approx 0.30$ となる．

3.6.2 経路の決定方法

3.6.1 節で計算したカテゴリシーケンスと共有地点の組 $< S, SP >$ それぞれに対し，先行研究 [11] の匿名領域確保のための訪問場所の決定方法を適用することで，実際に地理的に到達可能，かつ匿名領域に対する要求を考慮した経路を生成する．最終的なダミーの経路は，それらそれぞれに対し，スコア $Score_s(< S, SP >)$ と実際に生成された経路の匿名領域に関する要求の達成度合いに基づきスコアリングを行い，最良のものを選出することで決定する．

具体的にはまず，各ダミーに対し個別に目標とする匿名領域の大きさ (設定匿名領域) を設定する．この設定匿名領域の大きさは後に生成されるダミーほど大きくなるように設定し，最後のダミーには要求匿名領域の大きさとなるように設定する．すなわち i 番目のダミーに対する設定匿名領域 $SA_r(i)$ は，$SA_r(i) = f(i), f(n) = A_r$ となるような単調増加関数 $f(i)$ に基づき決定する．この手順はグループ B のダミーについても同様に行う．なお，具体的な $f(i)$ については，本稿では言及せず，シミュレーション実験の結果に基づいて検討する予定である．この設定匿名領域を用

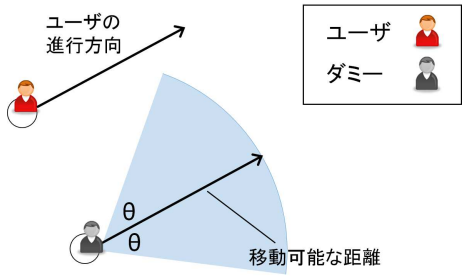

図 8 ユーザの進行方向を考慮したダミーの移動

いて，共有地点設定時刻を起点に，それ以前と以降のサービス時刻におけるダミーの位置を順に決定していく．時刻 t におけるダミーの位置 P_t は，時刻の $t-1$ における位置 P_{t-1} に基づいて次の手順で決定する．まず初めに，P_{t-1} から時刻 t までの間に到達可能な範囲にある訪問場所のうちカテゴリの制約を満たすものを取得する．訪問場所は，その中から選択されるが，このとき，生成済みダミーとユーザによって形成される匿名領域の面積を計算し，生成中のダミーの設定匿名領域の大きさと比較する．この面積が生成中ダミーの設定匿名領域より小さければ，生成中ダミーを含めた匿名領域の大きさが設定匿名領域の大きさに最も近づく地点を訪問場所として選択する．逆にこの面積が，既に設定匿名領域より大きい場合は，ユーザの進行方向を考慮し，図 8 のように，ユーザの進行方向ベクトルに対する角度が θ 以内の場所を選択する．これは，ユーザの進行方向に従ってダミーを移動させることにより，ユーザが孤立して移動し，ユーザであると特定されやすくなることを防ぐためである．ここでは，共有地点でのサービス利用以降の時刻における位置の決定方法について述べたが，これを共有地点を設定した時刻以前の時刻にも同様の方法で適用可能である．

上記の手順により，到達可能性の制約を満たし，かつ匿名領域に関する要求を考慮した経路を $Score_s$ を算出した全ての $< S, SP >$ について生成し，それぞれの経路 $traj_{<S,SP>}$ に対し，以下のスコアを計算する．

$$Score_{traj}(traj_{<S,SP>}) = Score_s(< S, SP >) \times \frac{n_{achieve}(traj_{<S,SP>})}{|traj_{<S,SP>}|} \quad (6)$$

ここで，$n_{achieve}(traj_{<S,SP>})$，$|traj_{<S,SP>}|$ はそれぞれ，$traj_{<S,SP>}$ において，設定匿名領域を達成できた回数，及びサービスの利用回数を表す．上記のスコアが最良となるものを，生成中ダミーの移動経路として決定する．

3.7 グループ B のダミーの生成

グループ B のダミーでは，訪問地点のカテゴリを考慮せず，ユーザの嗜好と異なる経路を生成する．これにより，ユーザの嗜好が曖昧化される．さらに，設定されている共

有地点数が少ないエンティティとの間に共有地点を設けることで，ユーザの嗜好を知らない攻撃者に対して，追跡可能性をさらに低下させる．

具体的にはまず，共有地点が設定されているエンティティ数が最小の時刻 t_{min} と設定されている共有地点数が最小のエンティティ id_{min} を取得し，そこに共有地点を設定する．この共有地点をもとに，カテゴリの制約を除いてグループ A のダミーと同様の方法で到達可能な点から設定匿名領域に基づいて，訪問場所を選択していくことで，経路を決定する．

3.8 まとめと今後の課題

本稿では，位置プライバシ保護のためのダミー生成において，パーソナライズのために嗜好のみ公開したいユーザと，嗜好も保護したいユーザ両方の要求を満たすため，訪問場所のカテゴリシーケンスをユーザの嗜好と定義し，その保存度を制御可能なダミー生成手法を提案した．提案手法では，可観測な嗜好と，ユーザ自身の行動のみに基づく真の嗜好の差分をシミュレートし，要求に応じてその差分を小さくするようにダミーを生成することで，嗜好の保存度を制御する．その際，サービス利用の追跡可能性も考慮し，嗜好を知っている攻撃者に対しても効果的に追跡可能性を低下させられるよう，カテゴリシーケンスと共有地点を選択する．

今後は，本稿で述べたモデルに従って，追跡可能性に関するプライバシの尺度を検討し，FourSquare のチェックイン履歴のデータセットと，実際の地図情報を用いてシミュレーション評価を行うことで，プライバシ保護と嗜好の制御可能性，両方の観点から提案手法の有効性を確認する予定である．

謝辞 本研究の一部は，日立財団研究助成「倉田奨励金」の研究助成によるものである．ここに記して謝意を示す．

参考文献

[1] Ardagna, C. A., Cremonini, M., Damiani, E., di Vimercati, S. and Samarati, P.: Location Privacy Protection Through Obfuscation-based Techniques, *In Proc. CODASPY*, pp. 47–60 (2007).
[2] Beresford, A. R. and Stajano, F.: Mix Zones: User Privacy in Location-aware Services, *In Proc. PerCom*, pp. 127–131 (2004).
[3] Gruteser, M. and Grunwald, D.: Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking, *In Proc. MobiSys*, pp. 31–42 (2003).
[4] Gutscher, A.: Coordinate transformation - a solution for the privacy problem of location based services?, *In Proc. IPDPS*, pp. 7 pp.– (2006).
[5] Kido, H., Yanagisawa, Y. and Satoh, T.: An anonymous communication technique using dummies for location-based services, *In Proc. ICPS*, pp. 88–97 (2005).
[6] Lee, B., Oh, J., Yu, H. and Kim, J.: Protecting Location Privacy Using Location Semantics, *In Proc. SIGKDD*, KDD '11, pp. 1289–1297 (2011).
[7] Lu, H., Jensen, C. S. and Yiu, M. L.: PAD: Privacy-area Aware, Dummy-based Location Privacy in Mobile Services, *In Proc. MobiDE*, MobiDE '08, pp. 16–23 (2008).
[8] Niu, B., Zhang, Z., Li, X. and Li, H.: Privacy-area aware dummy generation algorithms for Location-Based Services, *In Proc. ICC*, pp. 957–962 (2014).
[9] Palanisamy, B. and Liu, L.: MobiMix: Protecting location privacy with mix-zones over road networks, *In Proc. ICDE*, pp. 494–505 (2011).
[10] 加藤 諒，原 隆浩，Xie, X.，岩田麻佑，西尾章治郎：ユーザの行動プランの変更を考慮したダミーによるユーザ位置曖昧化手法，*In Proc. DEIM* (2015).
[11] 加藤 諒，松野有弥，原 隆浩，荒瀬由紀，Xie, X.，西尾章治郎：ユーザの訪問場所の傾向を考慮したダミーによるユーザ位置曖昧化手法，*In Proc. DICOMO*, Vol. 2014, pp. 1174–1181 (2014).