社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
DE2011-35 (2011-8)

位置情報サービスにおける属性を考慮したプライバシ保護について

眞野　将徳† 　石川　佳治††,†,†††

† 名古屋大学大学院情報科学研究科
†† 名古屋大学情報基盤センター
††† 国立情報学研究所
E-mail: †mano@db.itc.nagoya-u.ac.jp, ††y-ishikawa@nagoya-u.jp

あらまし　近年，GPS 機能を有するモバイル端末の普及や無線通信網の発展により，測位された位置情報に基づいて情報を提供する位置に基づくサービスが普及している．位置に基づくサービスにはプライバシに関わる問題が存在するため，ユーザのプライバシを保護しようとする研究が多くなされている．しかし，これら既存手法の多くは位置情報のみを考慮しており，ユーザの性別・年齢といった属性情報を併用するようなサービスには対応できない．本論文では，このようなユーザの属性情報を併用するサービスを対象とした匿名化手法を提案する．

キーワード　位置に基づくサービス，プライバシ，匿名化

Masanori MANO† and Yoshiharu ISHIKAWA††,†,†††

† Graduate School of Information Science, Nagoya University
†† Information Technology Center, Nagoya University
††† National Institute of Informatics
E-mail: †mano@db.itc.nagoya-u.ac.jp, ††y-ishikawa@nagoya-u.jp

Abstract　In recent years, location-based services have become popular due to the progress of GPS devices and mobile communication technology. Since a location-based service has a risk of privacy, many approaches have been proposed to tackle to protect users' privacy. However, exisiting approaches only consider only users' location information and cannot cope with the services which use attribute information such as age and addres of a user. In this paper, we propose an anonymization method for location-based services that consider attribute information.

Key words　Location-based services, privacy, anonymization

1. 研究の背景と動機

1.1 研究の背景

近年，GPS 機能を有するモバイル端末の普及や無線通信網の発展により，測位された位置情報に基づいて近隣の店舗の情報などのユーザに有益な情報を提供する，**位置に基づくサービス**（location-based services）が普及している．位置に基づくサービスは便利ではあるものの，プライバシに関わる問題が存在する．サービスを利用するためにはユーザの位置を送信する必要があるが，詳細な位置情報を送信してしまうと，悪意を持った攻撃者である可能性があるサービス提供者にユーザがどこにいるか知られてしまう．たとえば自宅でこのようなサービスを利用すると，その場所と住所録とを照合させることで，ユーザを特定することもできてしまう．これを防ぐため，**位置の匿名化**（location anonymization）と呼ばれる位置情報を曖昧にする手法を用いて，ユーザのプライバシを保護しようとする研究が多くなされている [1]．位置情報をあまりに曖昧化してしまうと適切なサービスを受け入れられなくなる可能性があるため，適度な匿名化が求められる．

1.2 モバイル広告配信サービス

位置に基づくサービスのうち，位置情報のみを用いるものについては，先に述べた位置の匿名化がユーザのプライバシ保護に有効である．しかし，位置に基づくサービスには位置情報のみでなく，ユーザの性別・年齢といった**属性情報**も併用するようなサービスも考えられる．ここでは，そのようなサービスの例として**モバイル広告配信サービス**を考える．

ここで想定するモバイル広告とは，広告ごとにそれを配信するエリアや時間帯の指定を可能とするもので，たとえばある店舗のタイムセール情報を店舗周辺のモバイルユーザに配信するために用いられる．さらにここでは，広告配信のためにユーザの属性を利用することを想定する．たとえば，女性向け化粧品のモバイル広告の場合，対象は女性であり，想定する年齢層も存在する．モバイルユーザの属性を考慮して，条件に合致するユーザに配信できれば高い広告効果を得られると考えられる．

本稿で想定するモバイル広告配信サービスの構成を図1に示す．ここで重要な役割を果たすのが**仲介業者**である．仲介業者は信頼できるサードパーティであり，各ユーザの属性情報を**プロファイル**（profile）として管理している．また，各広告主から提供されるモバイル広告の情報（対象とするユーザの属性に関する情報，配信時間など）も管理する．後述のように，仲介業者は匿名化処理も担当する．モバイルユーザが仲介業者に対してモバイル広告の要求を発行すると，仲介業者はそのユーザの情報を広告主に提示し，広告主が配信を希望した広告の中から適切なものを選びユーザに配信する．このようなモデルは，ユーザにとっては割引やクーポンのような形でのメリットがあり，広告主にとっては想定に合った，その場に実在するユーザに広告配信できる点でメリットがある．

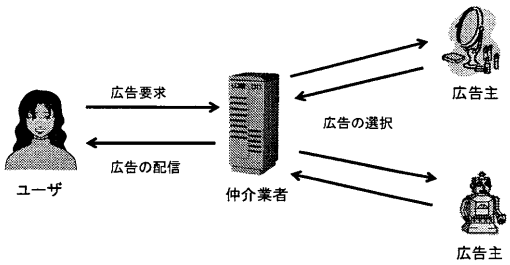

図1　モバイル広告配信サービスの構成
Fig. 1　Organization of mobile advertisement service

1.3　匿名性の問題

ここで問題となるのは，広告主は必ずしも信頼できず，**攻撃者**となりうる可能性があるということである．ユーザの正確な位置が広告主に通知される場合には，攻撃者である広告主は，自身の広告エリアを観測することにより，ユーザを特定することができる．このような状況に対しては，既存の位置の匿名化に関する技術が適用できるが，ユーザの属性も考慮する場合には以下の例で示すような問題が発生する．

ユーザが図2のように存在しているとする．また，それぞれの属性情報を図3に示す．

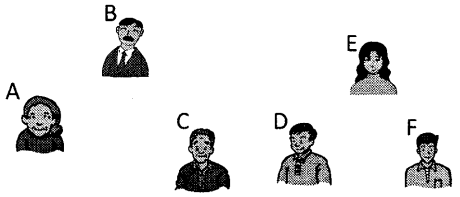

図2　ユーザの分布
Fig. 2　User locations

ユーザDが広告の要求をしたとする．k人のユーザを位置情報に基づいてグループ化する***k*匿名化**（k-anonymization）のアプローチを用いた場合，位置情報のみが利用されるため，近傍のユーザC，Eとグループ化されて，図4の匿名化1のように匿名化される（ここでは$k=3$とする）．ここでユーザの属性情報をそのまま提供する場合，広告主には，匿名化1の領域中に，それぞれ（男，63，岐阜），（男，52，愛知），（女，26，

ユーザ	性別	年齢	住所
A	女	70	愛知
B	男	48	長野
C	男	63	岐阜
D	男	52	愛知
E	女	26	三重
F	男	23	岐阜

図3　ユーザの属性
Fig. 3　User attributes

三重）という属性を有する3名のユーザが存在し，そのうちの1名が広告配信を希望していることが通知される．攻撃者は，匿名化1の領域を観察することにより，女性ユーザを容易に特定できる．男性ユーザについても年齢に開きがあるので，63歳の方のユーザがCであると特定できる可能性が高い．よって，3名のユーザの属性情報がほぼ明らかになってしまう．

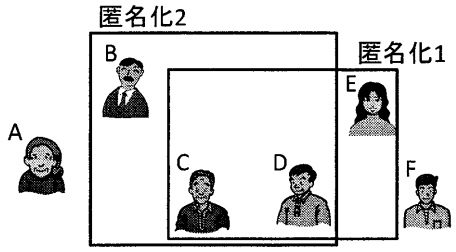

図4　匿名化のグループ
Fig. 4　Anonymized groups

次に，図4の匿名化2を考えてみる．この場合，全員が男性であることから，性別によるユーザ識別は行えない．また，BとDの年齢が近いことから，両者を年齢情報のみで見分けることは困難である．このため，匿名化1に比べると，属性の観点からはより適切な匿名化になっているといえる．ただし，匿名化された領域のサイズはより大きくなっているため，よりピンポイントで広告を配信したいという広告主の要求に関してはマイナスとなる可能性がある．

1.4　研究の目的

本研究では，上で述べたように，位置情報のみでなく属性情報も考慮する匿名化手法の実現を目的とする．位置情報に関する匿名化のみならず，属性に関する匿名化も行う点に特徴がある．この場合，その属性が観測によりどの程度判定できるかということが重要なポイントとなる．また，上記の例では述べていないが，広告主にどの程度まで詳細な属性情報を提供してよいかはユーザによって異なりうる．個々のユーザの属性に関するプライバシのポリシーを匿名化処理に反映できることが望まれる．本研究で提案する手法では，これらの点を考慮して匿名化を実現する．

2.　関連研究

2.1　位置に基づくサービスのための匿名化手法

位置に基づくサービスのために用いられる匿名化のアプローチとして主流であるのが，**空間クローキング**（spatial cloaking）に基づく手法である．空間クローキングでは，ユーザの位置を

包含するような**秘匿領域**（cloaked region）を構築し匿名化を行う．[2]〜[5] で提案された手法では，秘匿領域を矩形形状で表現し，一般の匿名化に関する研究でしばしば用いられる**k匿名化**（k-anonymization）[6] の考え方を用いている．サービスを要求したユーザ以外の $k-1$ 人のユーザを含む領域を秘匿情報として用いることで，サービスを要求したユーザが k 人のうちの誰かが特定できないという点で，匿名性が保障される．

秘匿領域としては矩形領域以外のアプローチをとる手法もある．空間クローキングのためのシステム構成にもいくつかのアプローチがあるが，*location anonymizer* と呼ばれる，全ユーザの位置情報を把握し，匿名化処理を実行する信頼できるサードパーティを想定することが一般的である．本研究では，仲介業者がこの役割を果たす．

これまで提案された位置情報の匿名化のアプローチは，本研究で対象とする，位置情報に加え属性情報も考慮した匿名化処理には対応することができず，新たな技術開発が求められている．この種の匿名化手法に関する提案としては [7] がある．この手法では，ユーザの属性の値に応じてそれぞれ属性のベクトルを作成する．匿名化処理では，近傍にいるユーザの中から，ベクトルの距離ができるだけ近いユーザを探索することで匿名化をおこなう．ただし，この手法では，攻撃者からの属性の観測されやすさを考慮しておらず，必要以上に属性を一般化することにより，サービスの質に悪影響を与えるおそれがある．

2.2 プライバシ保護に関する属性の観点

属性に関する匿名化は，一般的なデータベース出版（database publishing）における匿名化 [8] との関連が深い．データベース出版における匿名化の議論では，しばしば属性が以下の 3 つに分類される．

- **機密属性**（sensitive attribute）：プライバシに深く関わり，個人との結びつきを秘匿したい属性（例：病名）
- **識別子**（identifier）：氏名，住所など個人を直接特定できる属性
- **準識別子**（quasi-identifier）：個人を直接特定はできないが，他の情報との組合せによって個人の特定に寄与しうる属性

匿名化処理では，識別子である属性は取り除き，準識別子の組合せがユーザの特定に結びつかないように，属性の汎化などの処理を適用する．

これに対し，本研究が対象とする位置に基づくサービスで用いる属性については，匿名化に果たす各属性のとらえ方が異なる．まず，データベース出版の場合とは異なり，真に秘匿すべき「病名」などの機密属性は一般には存在しない．また，たとえば医療に関するデータの出版では，「年齢」，「性別」，「郵便番号」という準識別子がわかると，別のデータ（例：選挙人名簿）と照らし合わせて個人が特定されてしまうといった説明がなされる．しかし，位置に基づくサービスの場合，サービス対象のエリアに訪れる可能性がある者の情報をあらかじめ得ることは困難であり，別のデータと照らし合わせて個人が特定される可能性は小さい．問題となるのは，実際にサービス対象のエリアが観測されて，位置情報と属性情報をたよりにユーザが特定されてしまうことである．以上の理由から，本研究においては，属性の扱いを見直す必要がある．

属性のとらえ方に関する関連研究の一つとして，ソーシャルネットワークにおけるプライバシ保護のため，属性の扱いに着目した [9] がある．この論文では，プライバシに関する属性のとらえ方の基準として，以下の 2 つを考えている．

- **機密性**（sensitivity）：その属性の情報がどれだけプライバシ侵害につながりやすいかという基準である．たとえば，「住所」は自宅の特定につながり非常に機密性が高いといえるが，「出身地」はそれほどプライバシに影響は与えず，機密性は「住所」ほどは高くない．[9] では，各属性の機密度（機密性の度合い）はユーザには依存せず，属性ごとに個別の値をとるものとしている．ただし，Facebook は実名登録であるため「本名」の機密度は低いが，他の SNS では高い値を取りうるなど，システムや応用によってとらえ方が異なるとしている．
- **可視性**（visibility）：ユーザがどれだけ詳細な属性値を公開してよいかという基準であり，ユーザごとに異なる値を持つ．たとえば，「生年月日」という属性を考えてみると，全て公開，月と日だけ公開，非公開など，ユーザの公開のポリシーは異なる．この場合，全て公開しているユーザは可視性を高く設定し，非公開のユーザは低く設定しているといえる．

これら 2 つの基準を位置に基づくサービスでも用いようとすると，機密性はそのまま利用できるが，可視性については検討の余地がある．ある属性値を利用者が隠したいと望んだとしても，属性によっては，攻撃者がそのユーザを観察すればその値が明らかになってしまうものがある．すなわち，位置に基づくサービスにおいては，ユーザがどれだけその属性を公開してもよいかということよりも，外から観察したときにその属性値がどれだけ推測できるかが重要となる．本研究では，これを**可観測性**（observability）と呼ぶ．これについては後ほど詳しく述べる．

2.3 個人の嗜好を反映したプライバシ保護

本研究が想定する状況においては，どのユーザがどの属性をどの程度公開してもよいかは個人に依存する部分が大きいため，個人の嗜好を反映したプライバシ保護が必要となる．一般の匿名化において問題となるのは，機密属性をどの程度詳細に提供してよいかということに，個人の嗜好を反映させることができない点である．たとえば「病気」という機密属性を考えてみると，ガンの人はそのことを隠したいと思うかもしれないが，風邪の人は別にそのことを知られても構わないと考えているかもしれない．[10] は，このような状況を対象にしたプライバシ保護手法であり，機密属性について属性値のタキソノミ（階層構造）を前もって構築しておく．そしてユーザはその階層構造のどのレベルまで詳細化してよいかを指定することで，プライバシ保護のレベルをコントロールすることができる．本研究では，個人の嗜好を反映するこのようなアプローチを発展させて用いる．

3. 属性とプロファイル

3.1 プライバシ保護のための属性の基準

2.2 で述べたように，位置に基づくサービスでは，ユーザがそこにまさしく存在するため，攻撃者がその地点を観測すれば，サービスを要求したユーザがそこに存在しているかわかってしまう．そこで，本研究では位置に基づくサービスにおけるプライ

バシ保護のための新たな基準として，**可観測性**（observability）を導入する．

- **可観測性**（observability）：ユーザを外部から観察したときに，その属性をどれだけ推測しやすいかを示す．たとえば，「性別」は観察すればその推測は容易であるのに対し，「出身地」は観測のみで判定するのは困難である．

本研究では，プライバシに関わる属性の性質について以下のようなアプローチをとる．

- 機密性について：各属性がどの程度機密であるは，ユーザごとに異なると考える．属性値をどの程度詳細化してよいかは，ユーザのプロファイルにおいて記述されるものとする．匿名化アルゴリズムは，機密性に関するユーザの要求を反映した匿名化処理を実現する．
- 可観測性について：各属性についての可観測性は，システムにおいて固定された値をとるものとする．「性別」といった属性が最大となり，「出身地」は小さい値をとる．
- 可視性について：可視性については本研究では直接的には扱わないものとする．本研究では，各ユーザによる各属性の機密性に関する指定とその属性の可観測性に応じて，匿名化のアルゴリズムにより，そのユーザのその属性の可視性の度合いが決められる．たとえば，あるユーザが性別の情報を提供したくなければ，プロファイルにおけるその記述が反映され，匿名化の結果としてユーザの性別が判別しにくいように属性情報が加工される．

3.2 属性のタキソノミとその利用

次に，匿名化過程における汎化（generalization）処理で用いる属性の**タキソノミ**（taxonomy）について述べる．各属性について，あらかじめシステムによって定められた階層的なタキソノミが設定されているとする．例として，年齢のタキソノミを図5に示す．anyは根ノードであり全ての値を含み，根ノードから葉ノードへ進むにつれて値が詳細化されていく．タキソノミは例に示した年齢以外にも，性別，ZIPコード，職業などあらゆる属性について作成されると考えるが，本論文ではその作り方については踏み込まない．

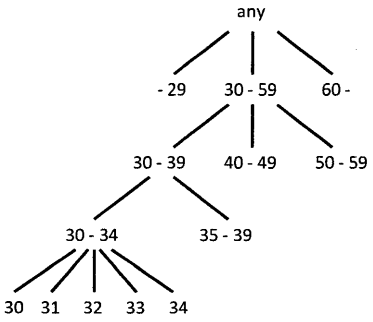

図5　年齢のタキソノミ
Fig. 5　Taxonomy for Age attribute

各ユーザは各属性におけるタキソノミの階層構造において，自身の属性をどの程度まで詳細化して提供してよいかを，**ガーディングノード**（guarding node）により指定する．ガーディングノードにより個人の嗜好に合ったプライバシ制御をするアイデアは[10]にある．たとえば，あるユーザの年齢が33歳であったとする．そのユーザが，自身が30代であることまでは情報として提供してよいなら，ノード[30–39]をガーディングノードとして指定する．ガーディングノードがより根ノードに近づけば近づくほど，その属性に対する匿名化の要求が高くなることを意味する．ユーザにとっては，ガーディングノードのレベルを葉レベルに近くするほど，より個人に特化された広告等のサービスを享受することができることから，ガーディングノードの設定は，プライバシの質とサービスの質にどれだけ重みを置くかのトレードオフとなる．

3.3 プロファイル

ユーザの属性情報およびプライバシ制御に関する情報は**プロファイル**（profile）に記述される．プロファイルの例を図6に示す．実際には各行がユーザのプロファイルに該当する．プロファイルでは，各属性について，真の属性値（例：$A_{\text{属性名}}$）とガーディングノード（例：$G_{\text{属性名}}$）のペアが記述される．

id	$A_{\text{性別}}$	$G_{\text{性別}}$	$A_{\text{年齢}}$	$G_{\text{年齢}}$	$A_{\text{出身地}}$	$G_{\text{出身地}}$
1	男	男	23	[20–24]	名古屋	名古屋
2	女	女	22	[20–29]	横浜	神奈川
3	男	any	18	[15–19]	岐阜	岐阜

図6　プロファイルの記述例
Fig. 6　Example of profile description

プロファイルは，匿名化処理を行う信頼あるサードパーティにより管理され，匿名化処理で用いられる．1.2で説明したモバイル広告配信サービスにおいては，仲介業者が管理することになる．

4. 匿名化処理の概要

4.1 基本的なアイデア

匿名化処理では，要求を出したユーザに対し，位置が近く属性が似通っているユーザをグループ化することが基本となる．これによりユーザの区別がつかなくなりプライバシを保護できるとともに，位置と属性の情報が過度に汎化されることによるサービスの質の低下も防ぐことができる．ただし，すべての属性を同じように扱うべきではなく，その属性の機密性や可観測性に応じて属性の扱いを変えることによって，プライバシとサービスの質をできるだけ維持することが求められる．

3.1で導入した可観測性は，情報を一般化するときのガイドラインとして用いることができる．図7では，匿名化処理における2通りのグループ化の例を示している．左側の図では，可観測性が高い「性別」の値が異なるユーザが同じグループに含まれている．グループに含まれるユーザのプロファイル情報がそのまま提示された場合，観察により，攻撃者は容易にどちらのプロファイルがどのユーザに対応するかを特定できる．特定されないようにするためには，それぞれの性別の属性値をanyなどと汎化しなければならない．一方，右側の図では，可観測性が低い「職業」の値が異なるユーザが同じグループに含まれている．ユーザの職業を観察のみで推測することは困難であるため，どのユーザがどのプロファイルに対応するかの特定は難しい．そのため，「職業」については，これ以上一般化しなくて

も十分にユーザのプライバシを守れるといえる．

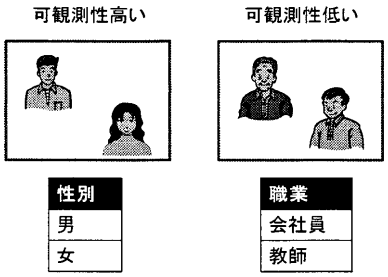

図 7　匿名化における属性の可観測性の影響
Fig. 7　Effects of attribute observability for anonymization

図 8 では同じように「性別」と「職業」について考えているが，この例では，片方のユーザの属性値が，そのユーザのガーディングノードの要求に基づいて汎化されている．職業については，先ほどと同様，観察しても推測することは難しいので，ユーザのプライバシを保護できている．それに対し性別の方は，片方のユーザが any であるが，領域内のユーザは両方男性であるため，属性値が any であるにも関わらず，このユーザが男性であることはわかってしまう．たとえ 2 人のユーザの属性値を共に any に一般化したとしても，観察すれば 2 人とも男性であることは攻撃者に知られてしまう．もし，男性であるということを知られたくないのであれば，女性ユーザと一緒にグループ化しなければならない．以上の 2 つの例を用いて示したように，匿名化する際にはその可観測性に応じて属性の一般化をする必要がある．

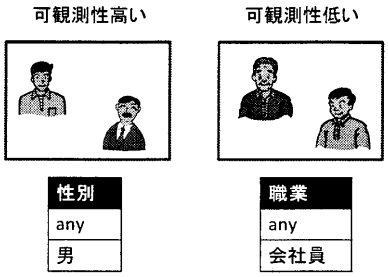

図 8　汎化された属性値の影響
Fig. 8　Effects of generalized attribute values

5. 匿名化の評価基準

5.1 可観測性に基づく評価

図 6 において，ユーザ 1 がサービス要求を出し，ユーザ 1〜3 に対し匿名化のグループが構築されたとする．3 人の誰が要求を出したかわからないという意味では $k = 3$ の匿名性が実現できている．ただし，ユーザのプロファイル情報をどのように提示するかについては多くの選択肢がある．

図 9 の左は，ユーザプロファイルの情報をそのまま提示するものである．この例は匿名性の点でかなり問題がある．なぜなら，攻撃者は対象領域に男性 2 名，女性 1 名が存在することを観測できるため，2 番目のエントリが観測できた唯一の女性に対応することがわかる．つまり，100%の確率でユーザ 2 を特定できる．残り 2 名の男性については，ユーザ 1, 3 のどちらに対応するかは完全には判断できないが，18 歳と 23 歳という年齢を考慮すると，かなり高い確度での対応付けができる．

汎化なし

id	性別	年齢	出身地
1	男	[20–24]	名古屋
2	女	[20–29]	神奈川
3	any	[15–19]	岐阜

最大限の汎化

id	性別	年齢	出身地
1	any	[15–29]	名古屋
2	any	[15–29]	神奈川
3	any	[15–29]	岐阜

図 9　プロファイル情報の候補（その 1）
Fig. 9　Candidates of presented profile information (1)

一方，図 9 の右図のように，観測可能なすべての属性について，タキソノミ階層の最上位のエントリを用いて最大限に汎化することも考えられる．なお，観測できない「出身地」については汎化の必要がない．この場合，あるユーザとそのエントリが対応づけられる確率は 1/3 となり，3 名による匿名化のもとでは最大限の匿名化が実現できる．しかし，プロファイル情報はあまりに一般化されすぎており，有用な情報ではない．

次に，図 10 に示す中間的な汎化の例を示す．この例について，各ユーザのプロファイルが特定される状況を考える．まず，1 番目のエントリが正しく特定されるのは，攻撃者がエントリの並びを正しく判定したとき（1-2-3），もしくは 2 番目と 3 番目を間違って認識したが 1 番目は正しく特定できたとき（1-3-2）の 2 つの場合（可能世界）である．

id	性別	年齢	出身地
1	男	[20–29]	名古屋
2	any	[20–29]	神奈川
3	any	[15–29]	岐阜

図 10　プロファイル情報の候補（その 2）
Fig. 10　Candidates of presented profile information (2)

実際の確率の推定は，確率的データベースの問合せ評価 [11] と似たアプローチが必要となる．まず，ある与えられた年齢のユーザを観測したとき，その推定年齢とその確率（信頼度）を返すような確率質量関数（probability mass function）が与えられているとする．たとえば 23 歳のユーザを観測したとき，その年齢を 25 歳と判定する確率が 0.1 といったものである．これをもとにモンテカルロ法により確率を求める．具体的な評価手法は今後の課題としたい．

同様に，2 番目のエントリが正しく認識されるのは 1-2-3 および 3-2-1 の場合で，3 番目のエントリが正しく認識されるのは 1-2-3 の場合のみである．1 番目のエントリの性別が男性と限定されているため，取りうる可能世界に制約が生じ，2-1-3 という場合は発生しない．

このようなアプローチに基づき，各エントリが特定される確率が推定される．それぞれの確率が，各ユーザによりあらかじめ指定された特定確率の上限の閾値を超えないなら，そのプロファイル情報の候補は妥当な候補の 1 つとなる．このアプローチの利点としては，k 匿名化で用いられる k のような値ではなく，ユーザが与える特定確率の閾値で匿名化が制御できる点に

ある．すなわち，ユーザの閾値を満たすのであれば，任意の人数でグループ化してよいことになる．

5.2 可視性に基づく評価

別の観点として，サービス提供者（例：広告主）の立場からの評価が考えられる．サービス提供者にとっては，提示されたプロファイル情報がいかにユーザの生の情報に近いかが重要となる．たとえば，図9では，左に示したものが最も生の情報に近く有用性が高く，右に示したものが最も匿名化され有用性が低いといえる．この考え方は前節で述べた可観測性に基づく評価とちょうど逆のようにも思えるが，観点が異なる．たとえば，生の情報により近い匿名化を実現するには，機密性に関する制約を緩めて属性値の可視性を高めているユーザをグループ化に用いることがより重要となる．

この問題に対しては，[10] で提案されたアプローチが活用できると考える．この論文では，個人の嗜好を反映した匿名化手法を提案しており，本研究でも活用しているタキソノミに基づく匿名化処理のアイデアを示している．重要な概念として，プライバシの**侵害確率**（breach probability）の計算法が提案されている．たとえば，図9左については，1番目のエントリは年齢が [20–24] とあるため，特に年齢の分布に関する情報がなければ，23歳と特定されるのは 1/5 の確率である．一方，3番目のエントリについては，同様に属性値の分布に関する情報がなければ，性別で2通り，年齢で5通りの選択肢があり，1/10 の確率となる．

本研究では，このアプローチを本研究に対し拡張して用いたいと考えている．詳細については今後の検討課題とする．

5.3 他の評価の観点

他の評価の観点としては，一つには匿名化後の秘匿領域（cloaked region）のサイズがある．似た属性を持つユーザを組み合わせることが本研究の基本的な方針であるが，そのようなユーザが近隣にいない場合には，やや距離があるユーザを含めて秘匿領域が広くなることを犠牲にするか，もしくはこれまでの評価の尺度の方を重視するか，トレードオフをとる必要がある．

また，サービス提供者とのマッチングの度合いも一つの観点となる．たとえば，広告対象のエリアが比較的小さい代わりにより詳細なユーザ情報を求めるような広告主と，広告対象エリアが広いがユーザの詳細な情報はさほど求めない広告主では，それぞれの意向に沿った匿名化処理を行うことで，より適切なサービスの提供が可能となると考えられる．また，「30代の女性をターゲットとした広告」のような指定されたターゲットの情報を考慮して匿名化処理を進めることも考えられる．このようなアプローチについては，我々の先行研究でもアイデアを示したが [12]，今回の提案内容に応じてより洗練させることが必要となる．

6. まとめと今後の課題

本研究では，位置情報だけでなく属性情報も併用するような位置に基づくサービスを対象とした新たなプライバシ保護手法を提案した．この手法は，属性に対して機密性，可観測性という2つの基準を設けることで，位置に基づくサービスに特有のプライバシの脅威に対して有効となることを目指している．特に，ユーザの属性が外部からの観測により推測されてしまうか否かという観点は，位置に基づくサービスのための匿名化処理において新たなものであると考えている．

今後の課題としては，まず，5. で述べた属性の評価基準の詳細化を行いたい．確率的なアプローチを用いることで，匿名化の結果として生成されサービス提供者に提示されるプロファイル情報がどの程度詳細であり，また，匿名性を保持することができるかが評価できるため，有効な評価尺度の開発を図りたい．また，アルゴリズムの開発も大きな課題である．まず，評価における確率の計算には，場合によっては多大なコストが生じる．位置に基づくサービスを支援するためには，リアルタイムの匿名化処理が発生する．データベース出版に比べて1回の処理で匿名化しなければいけないデータ量はわずかではあるが，瞬時に応答するためにはアルゴリズムの実装技術の開発や，場合によっては近似解の導出など，より効率的な手法を開発する必要がある．また，実験評価により手法の有効性を示すなどについてもさらに検討を進めていく必要がある．

謝　辞

本研究の一部は，内閣府最先端研究開発プロジェクト（FIRST）および科学研究費（22300034）の助成による．

文　献

[1] L. Liu, “Privacy and location anonymization in location-based services,” SIGSPATIAL Special, vol.1, no.2, pp.15–22, 2009.
[2] M.F. Mokbel, C.-Y. Chow, and W.G. Aref, “The New Casper: Query processing for location services without compromising privacy,” Proc. VLDB, pp.763–774, 2006.
[3] P. Kalnis, G. Ghinita, K. Mouratidis, and D. Papadias, “Preventing location-based identity inference in anonymous spatial queries,” IEEE TKDE, vol.19, no.12, pp.1719–1733, 2007.
[4] M. Gruteser and D. Grunwald, “Anonymous usage of location-based services through spatial and temporal cloaking,” Proc. MobiSys, pp.31–42, 2003.
[5] C.-Y. Chow, M.F. Mokbel, and X. Liu, “A peer-to-peer spatial cloaking algorithm for anonymous location-based service,” Proc. ACM GIS, pp.171–178, 2006.
[6] P. Samarati, “Protecting respondents’ identities in microdata release,” IEEE TKDE, vol.13, no.6, pp.1010–1027, 2001.
[7] H. Shin, V. Atluri, and J. Vaidya, “A profile anonymization model for privacy in a personalized location based service environment,” Proc. MDM, pp.73–80, 2008.
[8] B.C.M. Fung, K. Wang, R. Chen, and P.S. Yu, “Privacy-preserving data publishing: A survey of recent developments,” ACM Computing Surveys, vol.42, no.4, 2010.
[9] K. Liu and E. Terzi, “A framework for computing the privacy scores of users in online social networks,” Proc. ICDM, pp.288–297, 2009.
[10] X. Xiao and Y. Tao, “Personalized privacy preservation,” Proc. ACM SIGMOD, pp.229–240, 2006.
[11] N. Dalvi and D. Suciu, “Efficient query evaluation on probabilistic databases,” Proc. VLDB, pp.864–875, 2004.
[12] M. Mano and Y. Ishikawa, “Anonymizing user location and profile information for privacy-aware mobile services,” Proc. of ACM SIGSPATIAL International Workshop on Location Based Social Networks (LBSN 2010), pp.68–75, 2010.