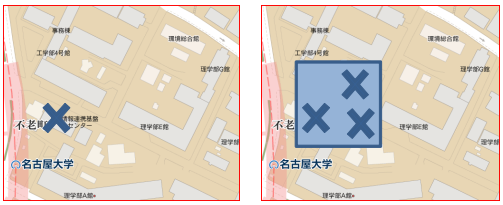
D-019

位置情報サービスのためのプライバシーを考慮した匿名化処理について

眞野 将徳 †　　石川 佳治 †† †
† 名古屋大学 情報科学研究科　†† 名古屋大学 情報基盤センター

1　まえがき

今日，GPS などを用いた位置情報サービスが私たちの生活に浸透しつつある．しかし，これらのサービスには利用履歴のログをサービスの管理者や第三者に覗かれると，個人が特定されたり，自宅，職場の場所が知られてしまうなどの危険がある．これを防ぐために，既存研究の多くでは，ユーザの位置情報に基づき，k 人と区別がつかないようにする k-匿名性 (k-anonymity) [2] を保持する匿名領域生成によりユーザのプライバシーを保護している [1]．このとき，ユーザは位置情報のみを持つ点データとして扱われる．k-匿名化について図 1 に示す．図 1(a) のように匿名化をせずにサーバに問合せすると，ユーザの正確な位置情報までわかってしまう．これに対し，図 1(b) では，k=3 で匿名化をおこなっている．このように匿名化することで，サーバには 3 人のユーザが含まれる矩形領域しか送信されず，ユーザの正確な位置情報はわからなくなる．たとえ，矩形領域内の 3 人のユーザの位置が何らかの情報でわかったとしても，その 3 人のうちだれが問合せをしたのかまでは絞り込むことはできない．



(a)　(b)

図 1: k-匿名化

しかし，このような前提に対して，位置情報サービスの中にはユーザの位置情報だけでなく，その他の情報も利用するサービスも考えられる．このようなサービスのひとつに広告配信が考えられる．広告配信では，広告主が広告料を支払い広告を配信する．ここでたとえば，60 代男性に対して 20 代女性向けの化粧品の広告を出してもその広告に対する効果は低い．企業にとって，支払う広告料に対して高い広告効果を得るために，広告に対象となるユーザの性別，年齢などを設定し，それと合ったユーザに対して広告を送信する仕組みが望ましい．また，広告の受け手であるユーザにとっても，そのような属性考慮した広告は，欲しいと思えるものがくる可能性が高く便利なものとなる．

Privacy-awared Anonymization for Location-based Information Services
Masanori Mano†, Yoshiharu Ishikawa†† †
† Graduate School of Information Science, Nagoya University
†† Infomation Technology Center, Nagoya University

[3] では，このような広告配信システムに対して，一致度という概念を導入し，プロファイルが近いユーザを選んで匿名化する手法を提案した．これに対し，本論文では，価値関数を導入し，実際に配信される広告に即した匿名化処理手法を提案する．

2　手法の概要

本論文の提案手法は次のようなシステムを想定している．システム概念図を図 2 に示す．本手法の対象となるシステムでは，各ユーザと各広告との間に，信頼できる第三者であるサーバを導入する．本論文ではこれを matchmaker と呼ぶ．matchmaker の役割は，ユーザの要求に応じて匿名化処理をおこない，そのユーザのプロファイルにあった広告を配信することである．

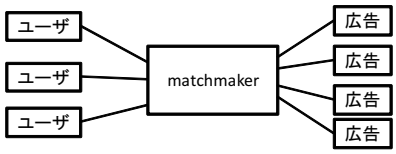


図 2: システム概念図

本論文で提案する手法では，ユーザ，matchmaker，広告がそれぞれプロファイルを持ち，それを活用することで適切な匿名化処理および広告配信をおこなう．以下にそれぞれのプロファイルについて説明する．

2.1　ユーザのプロファイル

ユーザは位置情報以外の個人に関する情報として，性別，年齢などの属性をプロファイルとして登録する．ユーザのプロファイルの例を図 3 に示す．

k	性別	年齢	住所
4	男	23	愛知県名古屋市千種区
3	女	25	愛知県豊田市

図 3: ユーザのプロファイル例

属性には数値属性 (例：年齢) と名義属性 (例：性別，住所) の 2 種類がある．名義属性については概念階層が定義されているものとする．たとえば，「住所」という属性を考えた場合，「愛知県名古屋市」や「愛知県豊田市」の上位に「愛知県」が位置する．また，匿名化のために k-匿名性の条件 k をプロファイルとしてもつ．matchmaker は，この k の値に応じて k 人と区別がつかないような匿名化処理をおこなう．

2.2　matchmaker のプロファイル

matchmaker には事前知識として，そのエリアにおける属性のドメインごとの属性値の存在確率が保持されている（たとえば，そのエリアが学生街ならば大学

生ぐらいの年齢のユーザが他の年齢のユーザよりも多く存在するなど）．名義属性については，先述したとおり概念階層が定義されるが，概念階層の階層の上位の存在確率は下位の存在確率の分布によって定まる．

2.3 広告のプロファイル

本研究では，広告主側から見た広告の効果を，匿名化処理における候補選択基準の一つとして用いる．広告費を負担する広告主には，ユーザに提示された広告の（予測）効果に応じて課金がなされると考える．そのため，広告の効果の高い匿名化処理を実現することが広告主および matchmaker の利益につながる．

広告には，属性ごとにその広告が各属性値 (例：年齢=25) をどの程度有益と評価するかをあらわす価値関数 (utility function) U を設定する．U はドメイン全体での積分値が 1 であるという条件を満たしており，確率のように扱うことができる．そのため，ある広告にとって，あるユーザがどれほど価値があるか計算するためには，各属性値について U を用いて価値を求め，それらの積 (同時分布) を求めればよい．U の例を図 4 に示す．

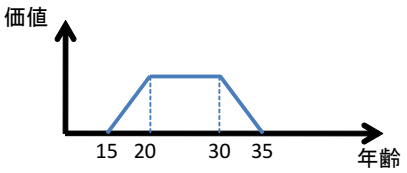


図 4: 価値関数の例

この例は，ある広告の年齢に関する U であり，20 から 30 歳のユーザ，およびそのまわりの年齢のユーザに対して広告を配信したいと考えていることがわかる．

本論文では，この U，存在確率の生成方法についてはこれ以上踏み込まない．

3 匿名化処理と広告配信処理

3.1 用語について

本論文では，説明のため次のような記号を用いる．

- c: k-匿名化処理結果の候補
- C: 候補集合 ($c \in C$)
- $r(c)$: c に対応する空間領域 (匿名領域)
- $U(c)$: c の広告効果 (価値)

3.2 匿名化処理と広告配信の流れ

matchmaker では，次のようにして匿名化処理および広告配信処理をおこなう．

1. ユーザ u の広告要求により，u を起点として匿名化処理を開始する．
2. ユーザ u のプロファイル k にしたがい，近傍のユーザ k 人を含むような匿名化の候補の集合 C を生成する．
3. C の各候補 c について，$U(c)$ を計算する．$U(c)$ の算出法はこの後述べる．
4. 最終的に，$U(c)$ の高い広告 n 件を，価値の高い順に順列をつけて配信する．

3.3 $U(c)$ の算出法について

ユーザのプライバシー保護の観点から，c の価値 $U(c)$ の算出に u の属性を直接利用することは問題がある．なぜなら，u の属性が知られると，他の外部情報と合わせることで，u について属性以上のことが知られてしまうおそれがあるからである．そこで提案手法では，背景知識であるユーザの属性値の分布確率を用いて，$U(c)$ を推定する．

ある広告の $U(c)$ の例を図 5 に示す．この例では，k=6 で匿名化しようとしたときに，匿名領域内のユーザの年齢の最大値が 22，最小値が 20 であったとする．まず，matchmaker のプロファイルとして用意されているこのエリアのユーザの年齢の分布確率 (図 5(a)) を参照すると，20 から 22 歳のユーザの推定分布の比は 1:2:3 であることがわかる．次に，20 から 22 歳の価値を調べる (図 5(b))．これらふたつを組み合わせ，領域内の 6 人ユーザの年齢分布は，前提知識の分布確率にしたがうと仮定し，その価値を平均したものがこの広告の年齢に関する c の価値となる (図 5(c))．

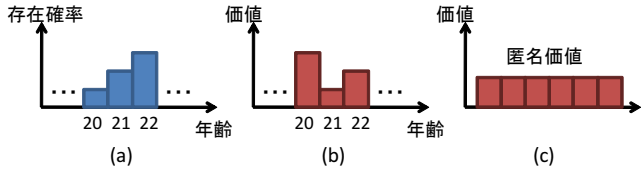


図 5: $U(c)$ 算出の例

他の属性についても同様に計算することで，$U(c)$ を求めることができる．

4 まとめと今後の課題

本論文では，広告配信サービスにおいて，ユーザのプロファイルを考慮し，また価値関数を用いることで，匿名性を保持しながらも，広告主の要求に答えることのできる手法を提案した．今後はシステムの実装と実験評価をおこなう予定である．

謝辞

本研究の一部は，科学研究費（21013023, 22300034）および内閣府「最先端研究開発支援プログラム」の助成による．

参考文献

[1] M. F. Mokbel, C.-Y. Chow, and W. G. Aref. The new casper: query processing for location services without compromising privacy. In *Proc. of VLDB*, pp. 763–774, 2006.

[2] P. Samarati. Protecting respondents' identities in microdata release. *IEEE Transactions on Knowledge and Data Engineering*, 13(6):1010–1027, 2001.

[3] 眞野 将徳, 石川 佳治. プライバシーを考慮した位置情報サービスの実現について. 情報処理学会創立 50 周年（第 72 回）全国大会, 6ZP-3, 2010.