L-014

誤差を含む属性値のための柔軟な匿名データ収集
Anonymized Data Collection for Attributes Containing Errors

清 雄一 †
Yuichi Sei

大須賀 昭彦 ‡
Akihiko Ohsuga

1. はじめに

ユビキタスコンピューティング技術やセンシング技術の発展により，様々なユーザ情報を収集しマイニングを行う研究が多数提案されている [7, 13]．しかし，ユーザ情報には個人を特定できる情報が含まれることがあり，プライバシが漏洩するリスクがある．

一方，各ユーザの属性分布を統計的に把握することが目的である場合は，各ユーザの情報を正確に収集する必要はない．ユーザの属性データをカテゴリ化し，一定の確率でユーザが**属さない**カテゴリからランダムに選択されたカテゴリ情報を収集する Negative Survey という手法が提案されている [19, 6, 10, 9]．たとえば，センシング結果として取り得るユーザの属性値の範囲が 0 から 99 であるとする．0 から 9 までをカテゴリ C_1，10 から 19 までをカテゴリ C_2 のように表現する．あるユーザ k の属性値が 37 である場合，カテゴリは C_4 である．このとき，k のカテゴリとしてある確率 p で C_4 をサーバへ報告するが，確率 $1-p$ で C_4 以外の任意のカテゴリをサーバへ報告する．サーバへ報告された情報からは，ユーザが属すカテゴリを正確に判定できないため，プライバシが一定レベルで保護される．しかし多くのユーザから情報を収集することで，各カテゴリに属すユーザ数の概算値を統計的に計算することができる．Negative Survey には，プライバシ漏洩リスクと，各カテゴリに属すユーザ数の計算誤差との間に，トレードオフの関係がある．

従来の Negative Survey では，2.1 で定義するプライバシ漏洩リスク R を全ユーザで共通に設定する必要があった．そのため，ユーザによっては R を緩和しても良いと考えていたとしても，最もレベルが厳しいユーザに合わせて共通の R を設定する必要がある．したがって，必要以上にプライバシを保護し，結果として各カテゴリに属すユーザ数の計算誤差が大きくなってしまう．

また，既存研究では各ユーザ属性は全て誤差無く計測できるとの前提を置いている．しかし実際には，センシング等によって計測されたユーザ属性には誤差が含まれる場合がある．本研究では，ユーザ属性を計測する際に，その精度情報も取得できる環境を想定する．

本論文の構成を示す．2 章では，本論文が対象とするスコープ，想定モデル及び，Negative Survey の基本アルゴリズムを述べる．3 章では，ユーザデータの匿名化技術に関する既存研究について記述する．4 章において，本論文が提案する手法を記述する．5 章では，提案手法のプライバシ，推測精度及び計算量について解析を行う．6 章では，提案手法と既存手法の比較を，数学的解析及びシミュレーションによって実施する．7 章において，ユーザ属性の取得精度及びプライバシ漏洩リスクの設定値に関する考察を示す．8 章で将来課題と本論文のまとめを記す．

†電気通信大学, The University of Electro-Communications

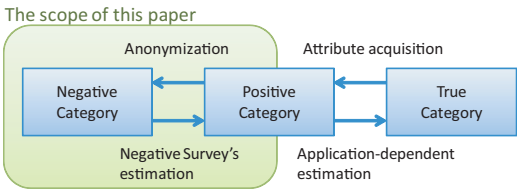


図 1: 本論文のスコープ

2. 背景

2.1. 研究のスコープ

本研究のスコープを図 1 に示す．ここで，利用する語句は以下の通りである．

- **True Category** ユーザ属性の真のカテゴリ．ユーザ自身のみが知っている情報．
- **Positive Category** センサネットワーク等で計測されたユーザ属性のカテゴリ．誤差を含み得る．
- **Negative Category** Positive Category を基に算出した，サーバへ報告するカテゴリ．

センサネットワーク等により，True Category を計測し，その計測結果が Positive Category である．本研究では，精度情報と共に Positive Category が得られた前提で議論を進める．Positive Category を匿名化してサーバに収集し（“Anonymization”），その結果から，各 Positive Category に属すユーザ数をサーバ上で推測する（“Negative Survey’s estimation”）までを本論文のスコープとする．True Category から Positive Category を計測する手法及び Positive Category から True Category を推測する手法は，アプリケーション依存の手法であるため本論文の対象外とする．

ユーザ属性のカテゴリ数を F とする．また，真のユーザ属性（True Category）の計測精度（“Attribute acquisition” における精度）を $\alpha\,(1/F \leq \alpha \leq 1)$ とする．$\alpha = 1$ のとき，計測されたユーザ属性（Positive Category）と True Category は確実に一致する．$\alpha = 1/F$ のとき，どのカテゴリに属すか全く計測できない状態である．ユーザ属性の計測精度向上は既存研究で取り組まれており，本研究の対象外とする．本論文では，α を正確に取得できるとする．α 自体に誤差がある場合は，7 章で考察する．

2.2. 想定モデル

本研究が想定しているサービスモデル及び攻撃モデルについて定義する．

2.2.1. サービスモデル

既存研究と同様のサービスモデルを想定する．ユーザの属性をセンサネットワーク等で計測し，取得されたデータをサーバへ送信する．公共施設等に設置されたセンサネットワークや，自宅で行う医療システムにおいて計測された医療データ等が考えられる．これらの情報を基に，ユーザに関する情報の大まかな分布（あらかじめ設定された各カテゴリに属するユーザの人数）を把握することを目的としたサービスを想定する．取得するユーザの情報は，Public Health[4] におけるユーザの年齢，性別，人種，体重や病名，匿名交通モニタリングにおける自動車の速度 [17] 等が考えられる．

2.2.2. 攻撃モデル

サーバは semi-honest であることを想定する [16]．これは，サーバはプロトコルから逸脱したことは行わないが，受信したデータから各ユーザの真のカテゴリを推測しようとするという攻撃モデルである．本論文の目的は，各ユーザの True Category を匿名化した上で，各 True Category に属すユーザ数を高精度に推測することである．

また，ユーザ属性計測機器自体への攻撃による情報抽出，センサノード間のネットワーク通信の盗聴，悪意を持ったノードの投入による情報抽出等の攻撃も想定される [15]．悪意を持つノードの検知手法 [20] 等の既存研究を用いて防ぐこととし，このような攻撃は本論文の対象外とする．

2.3.Negative Survey の基本アルゴリズム

Negative Survey では，Positive Category が C_i であるとき，確率 $p_{i,i}$ で Negative Category として C_i を選択し，確率 $p_{i,j}$ で Negative Category として C_j を選択する．通常の Negative Survey では，$p_{i,i}$ を全ての $i = 1, \ldots, F$ において同一の値 p に設定し，$i \neq j$ の $p_{i,j}$ を全て $(1-p)/(F-1)$ に設定する．このような確率行列 $M(p)$ を次式のように定義する．

$$M(p) = \begin{pmatrix} p & \frac{1-p}{F-1} & \frac{1-p}{F-1} & \cdots \\ \frac{1-p}{F-1} & p & \frac{1-p}{F-1} & \cdots \\ \frac{1-p}{F-1} & \frac{1-p}{F-1} & p & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{1}$$

サーバに Negative Category を報告したユーザ数を N とし，報告された各 Negative Category の数を $Y = \{Y_1, \ldots, Y_F\}$ とする．各 Positive Category に属すユーザ数の推測値 $A = \{A_1, \ldots, A_F\}$ は

$$A = M(p)^{-1} Y^{\tau} \tag{2}$$

で計算することができる．ここで，$M(p)^{-1}$ は行列 $M(p)$ の逆行列，Y^{τ} は行列 Y の転置行列を表す．

確率行列 $M(p)$ を利用した際の推測誤差として RMSD（Root Mean Square Deviation）を用い，E_p で表す．既存手法によっては平方根を取らないものや，カテゴリ数での除算有無の違いはあるが，本質的には同じ評価指標が用いられている．元々の Positive Category が C_i である割合を $P(X = C_i)$ とする．また，サーバに報告された Negative Category を基に推測された Positive Category が C_i である割合を $P(A = C_i)$ とする．この差を各カテゴリについて計算し，その平均値を取る．Negative Category が C_i である割合を $P(Y = C_i)$ とすると，E_p は以下の式の平方根で表すことができる [11].

$$\begin{aligned} E_p^2 &= \frac{1}{F^2} E(P(A = C_i) - P(X = C_i))^2 \\ &= \frac{1}{F^2} \sum_{i_1=1}^{F} \left[\sum_{i_2=1}^{F} \left(M(p)^{-1}_{i_1,i_2} \right)^2 Var\left(\frac{N_{i_2}}{N} \right) \right. \\ &\quad \left. + \sum_{\substack{i_2,i_3 \\ i_2 \neq i_3}}^{F} 2 \cdot M(p)^{-1}_{i_1,i_2} M(p)^{-1}_{i_1,i_3} Cov\left(\frac{N_{i_2}}{N}, \frac{N_{i_3}}{N} \right) \right], \end{aligned} \tag{3}$$

where

$$\begin{aligned} Var(\frac{N_i}{N}) &= \frac{1}{N} \cdot P(Y = C_i)\left(1 - P(Y = C_i)\right), \\ Cov(\frac{N_i}{N}, \frac{N_j}{N}) &= -\frac{1}{N} \cdot P(Y = C_i) P(Y = C_j) \end{aligned}$$

ここで，$M(p)^{-1}_{i,j}$ は，行列 $M(p)$ の逆行列 $M(p)^{-1}$ における i 行 j 列目の要素を表す．

2.4. 貢献

既存の Negative Survey では，全データに対して同一の確率行列を利用するが，本研究ではユーザのプライバシ設定やデータ精度によって変更する．確率行列が統一されていない状況は，これまで既存研究では想定されておらず，既存研究の手法をそのまま利用することはできない．

また，3 章で示すように既存の Negative Survey 手法には，攻撃者が，あるカテゴリに属さないユーザ群を抽出可能であるという問題がある．本研究は，このような問題にも対応する．

3. 関連研究

本章では，Negative Survey を含め，ユーザデータの匿名化手法について既存研究を記す．

3.1.Negative Survey の拡張

Horey らは Negative Survey をセンサネットワーク分野に適用させた [10]．式 1 において $p = 0$ と設定している．確率行列として $M(0)$ を利用する手法を **Straight 手法**と呼ぶ．サーバは各ユーザについて Negative Category の情報しか判断できないため，プライバシ保護レベルは高い．しかし，各 Positive Category に属すユーザ数を高精度で推測するためには数多くのサンプルを必要とする．

Straight 手法よりも少ないサンプル数で高精度の推測が可能な手法が，Xie らによって提案された [19]．以降で，この手法を **GSN 手法**と呼ぶ．GSN 手法では，式 1 の確率行列は用いず，各行の値を正規分布に基づいて大きく変動させる．多くの値が限りなく 0 に近い値となるため，攻撃者は任意のユーザの Negative Category から，

当該ユーザがほぼ確実に属さない Positive Category を複数特定できるという問題がある．

Groat らは，ユーザ属性が複数存在する場合に高精度でカテゴリ分布を推測可能な手法を提案している [9]．以降で，この手法を **MDA 手法**と呼ぶ．MDA は，ユーザ属性が一つしか無い場合においても，それを分割し多次元化することによって適用可能である．MDA では，$i \neq j$ におけるいくつかの $p_{i,j}$ の値を 0 に設定する．つまり，複数のカテゴリについて Negative Category として報告する確率を 0 にしている．

これらの手法はいずれも，あるユーザの Negative Category を取得した攻撃者が，当該ユーザの Positive Category を一定範囲に限定できてしまう．したがって，攻撃者はあるカテゴリに属さないユーザ群を確実に抽出可能であるという問題がある．たとえば，6 章において利用する評価対象のデータでは，多くのユーザの人種は白人である．このとき，ほぼ確実に白人でないユーザ群が抽出できてしまう．また，ユーザ属性の誤差や，各ユーザのプライバシ漏洩リスクを柔軟に変更することが考慮されていない．6 章において，これらの手法と比較する．

また，位置情報の匿名化に特化した手法が Forrest らによって提案されている [6] が，本研究と匿名対象データが異なるため，比較対象外とする．

3.2. 匿名化及び Differential privacy

個人が特定できないようにプライバシ情報を含むデータを匿名化することを目的とし，k-匿名性 [8][12] や，l-多様性 [14][18] 等の指標に基づく匿名化手法が提案されている．これらの匿名化手法の多くは，サーバにユーザの真の情報が格納されているという前提を置いており，サーバも信頼できない環境でユーザのプライバシを保護するという本研究の目標は達成できない．

Differential privacy[3] は，データにノイズを加えてクエリに答える手法である．信頼できるサーバの存在を前提としており，ユーザがどのサーバにも真の情報を提供したくないという要望には答えられない．

3.3. ランダムアプローチ

Negative Category 手法に類似した手法として，ランダムアプローチ [2][5] がある．ユーザの属性値にランダムな値を加えてサーバへ報告し，サーバ側で真の値を推測する．しかし，各ユーザは真の値から大きく外れることのないランダム値を加えることになるため，サーバへ報告された値から，各ユーザにおける真の値の取り得る範囲が明らかになるという問題が指摘されている [3]．

4. 提案手法のアルゴリズム

ユビキタスコンピューティング等のサービスが，ユーザ k の属すカテゴリを計測し，計測結果を Positive Category とする．計測精度 α_k と，許容するプライバシ漏洩リスク R_k に基づいて p_k を決定する．p_k を基に式 1 で表される確率行列を構築し，ユーザ k の Negative Category を算出する．その結果をデータ収集サーバへ送信する．多くのユーザから送信を受けたサーバは，各 Positive Category に属すユーザ数を推測する．

プライバシ漏洩リスク R_k の定義を 4.1 に，p_k の算出方法を 4.2 に，収集された Negative Category 及び各ユーザの p_k から Positive Category のデータ分布を推測する方法を 4.3 において記述する．

4.1. プライバシ漏洩リスク

攻撃者は各ユーザの Negative Category から True Category を推測する．このとき，あるユーザ k について，任意のカテゴリ C_i が True Category の候補から除外されることを防ぐ必要がある．あるユーザの Negative Category から推測される True Category が，カテゴリ C_i である確率を T_i で表す．このとき，$\min_i(T_i)$ を大きい値にする確率行列を構築することで，この課題に対応する．

カテゴリ数を F とすると，$\min_i(T_i)$ は 0 から $1/F$ までの値を取り得る．攻撃者に全く情報を与えない場合，$\min_i(T_i) = 1/F$ となる確率行列を利用することになるが，この場合はサーバ側でも何も情報を得られない．$\min_i(T_i)$ がその最大値 $1/F$ からどの程度離れているかを表す指標として，$\min_i(T_i) = (1 - R_k)/F$ を満たすプライバシ漏洩リスク R_k を定義する．R_k は 0 から 1 までの値を取り，値が小さいほどプライバシ保護レベルは高い．また，

$$\widehat{R_k} = \frac{1 - R_k}{F} \tag{4}$$

を満たす $\widehat{R_k}$ を定義し，以降で利用する．

4.2. プライバシ漏洩リスク及び精度に基づいた確率行列の構築

式 1 で表される確率行列の p の値を，ユーザごとに変更する．4.2.1 では，ユーザが設定するプライバシ漏洩リスク R_k を満たすための p_k が取り得る値の範囲を特定し，4.2.2 では取り得る値の範囲の中で最も推測精度を高くすることができる p_k の値を一意に特定する．

4.2.1. 各ユーザのプライバシ漏洩リスク R_k を満たす p_k の値の範囲算出

ユーザ k について，True Category と Negative Category が一致している確率 P_{match} は，

$$P_{match} = \alpha_k \cdot p_k + (1 - \alpha_k) \cdot \frac{1 - p_k}{F - 1}$$

であり，True Category と Negative Category が一致していない確率は，

$$P_{unmatch} = \alpha_k \cdot (1 - p_k) + (1 - \alpha_k) \cdot \left(1 - \frac{1 - p_k}{F - 1}\right)$$

で表される．$P_{match} + P_{unmatch} = 1$ の関係がある．

したがってプライバシ漏洩リスク R_k が与えられた場合，$\widehat{R_k} \leq P_{match}$ 及び $\widehat{R_k} \leq P_{unmatch}/(F-1)$ を満たす必要がある．これを満たすための p_k は P_{match}，$P_{unmatch}$ より，次式で表すことができる．

$$\begin{cases} 0 \leq p_k \leq 1 & (\widehat{R_k} \leq \frac{1-\alpha_k}{F-1}) \\ \frac{\alpha_k - 1 + (F-1)\widehat{R_k}}{\alpha_k F - 1} \leq p_k \leq \frac{\alpha_k - 2 + F - (F-1)^2 \widehat{R_k}}{\alpha_k F - 1} & (otherwise) \end{cases} \tag{5}$$

4.2.2.RMSD を最小化する p_k の値算出

式 5 の範囲で最も RMSD を小さくできる p_k の値を決定する．式 3 において，$p = p_k$，$N = 1$ とし，p_k について偏微分すると，

$$\frac{(F-1)^3(F+1)}{F^2(p_kF-1)^3}\sqrt{\frac{F(p_kF-1)^2}{(F-1)(F^2+2p_k-F(1+p_k^2)-1)}}$$

となる．したがって，$p_k < 1/F$ までの範囲では，p_k の値が大きくなるほど RMSD は増加し，$1/F < p_k$ の範囲では，p_k の値が大きくなるほど RMSD は減少する．また，

$$E_{1/F+\delta} = E_{1/F-\delta} = \sqrt{-\frac{(F-1)(F(1+(1+\delta^2-F)F)-1)}{\delta^2F^6}} \quad (6)$$

であるため，$p_k = 1/F$ で対称な関数となる．以上より，R_k，α，F から，式 5 を満たす範囲の中で，RMSD を最小化する p_k は以下のように定めることができる．

$$p_k = \begin{cases} 1 & (\widehat{R_k} \leq \frac{1-\alpha_k}{F-1}) \\ \frac{\alpha_k-2+F-(F-1)^2\widehat{R_k}}{\alpha_kF-1} & (otherwise) \end{cases}$$

4.3.Negative Category から Positive Category の推測

各ユーザ k の確率行列を M_k とする．各カテゴリ C_i について，Negative Category として報告された数を Y_i，Negative Category を基に Positive Category として推測される数を A_i とする．

p_k が同一であるユーザをそれぞれグループ化する．あるグループ $p_k = \hat{p}$ のユーザ数を $S_{\hat{p}}$ とし，当該ユーザ群を抽出する．抽出されたユーザ群に対し，各カテゴリ C_i について Negative Category として報告された数を $Y_{\hat{p},i}$，Positive Category として推測される数を $A_{\hat{p},i}$ とする．

ユーザ k の Positive Category が C_i の場合，$1-p$ の確率で C_i 以外のカテゴリが Negative Category として選択される．また，ユーザの Positive Category が $C_j\,(j \neq i)$ の場合，$1-(1-\hat{p})/(F-1)$ の確率で C_i 以外のカテゴリが Negative Category として選択される．したがって，下記式が成立する．

$$Y_{\hat{p},i} = S_{\hat{p}} - A_{\hat{p},i}\cdot(1-\hat{p}) - \sum_{\substack{j=1 \\ j\neq i}}^{F} A_{\hat{p},j}\cdot\left(1-\frac{1-\hat{p}}{F-1}\right) \quad (7)$$

式 7 を各 $Y_{\hat{p},1},\ldots,Y_{\hat{p},F}$ に対して構築することにより，F 元一次連立方程式が構築される．この連立方程式を解くことにより，$A_{\hat{p},1},\ldots,A_{\hat{p},F}$ が求められる．この処理を共通の $\hat{p}$ ごとに実施する．

最後に，p_k が共通な各ユーザ集合の Positive Category の推測値である $A_{\hat{p},1},\ldots,A_{\hat{p},F}$ に対し，その誤差に応じて加重平均を取ることで A_i を算出する．ここで，$p_k = \hat{p}$ であるユーザ集合における RMSD を $E_{\hat{p}}$ とする．$E_{\hat{p}}$ は，式 3 の計算対象を $p_k = \hat{p}$ であるユーザ集合に限定することで算出される．当該ユーザ集合において，サーバに報告された Negative Category を基に推測された Positive Category が，C_i である割合を $P(A_{\hat{p}} = C_i)$ とする．

このとき $E_{\hat{p}}$ は当該ユーザ集合の各 $P(X = C_i)$ と $P(A_{\hat{p}} = C_i)$ の差における標準偏差の平均値とみなすことができる．ここで，$P(X = C_i)$ と $P(A_{\hat{p}} = C_i)$ の誤差の確率密度関数が正規分布に従うと仮定すると，各 C_i について一組の $A_{\hat{p},i}$ の測定値が得られる確率は次式で表される．

$$\left(\prod_{\hat{p}}\frac{1}{\sqrt{2\pi}E_{\hat{p}}}\right)\exp\left(-\sum_{\hat{p}}\frac{(P(X=C_i)-P(A_{\hat{p}}=C_i))^2}{2E_{\hat{p}}^2}\right) \quad (8)$$

上記式を最大化する $P(X = C_i)$ が最も確からしい $P(X = C_i)$ と言える．これは，式 8 の $\sum_{\hat{p}}\frac{(P(X=C_i)-P(A_{\hat{p}}=C_i))^2}{2E_{\hat{p}}^2}$ を最小化することによって達成される．この式を $P(X = C_i)$ で微分すると

$$\sum_{\hat{p}}\frac{1}{E_{\hat{p}}^2}(P(X=C_i)-P(A_{\hat{p}}=C_i)) \quad (9)$$

となる．したがって，$P(X = C_i)$ が式 9= 0 を満たすとき，式 8 は最大化される．このときの $P(X = C_i)$ は以下の式で表される．

$$P(X=C_i) = \frac{\sum_{\hat{p}}(1/E_{\hat{p}}^2)P(A_{\hat{p}}=C_i)}{\sum_{\hat{p}}1/E_{\hat{p}}^2} \quad (10)$$

ここで，$P(A_{\hat{p}} = C_i) = A_{\hat{p},i}/S_{\hat{p}}$ である．また，A_i は $N\cdot P(X = C_i)$ で計算することができるため，目的の A_i は以下の式で表される．

$$A_i = N\cdot\frac{\sum_{\hat{p}}(1/E_{\hat{p}}^2)(A_{\hat{p},i}/S_{\hat{p}})}{\sum_{\hat{p}}1/E_{\hat{p}}^2} \quad (11)$$

ここで，$E_{\hat{p}}$ は式 3 に基づいて計算することもできるが，次のように簡潔に表現される数式で計算することができる．式 1 より，$p_k = \hat{p}$ のユーザのみに注目すると，

$$M(\hat{p})^{-1} = \begin{pmatrix} \frac{\hat{p}+F-2}{F\cdot\hat{p}-1} & \frac{\hat{p}-1}{F\cdot\hat{p}-1} & \cdots \\ \frac{\hat{p}-1}{F\cdot\hat{p}-1} & \frac{\hat{p}+F-2}{F\cdot\hat{p}-1} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \quad (12)$$

である．$P(Y = C_i) \approx P(Y = C_j)$ と仮定すると，式 3，12 より，$E_{\hat{p}}$ は次の簡約化された式で表すことができる．

$$E_{\hat{p}} = \sqrt{\frac{(F-1)(F^2+2p_k-F(1+p_k^2)-1)}{F^3S_{\hat{p}}(p_kF-1)^2}} \quad (13)$$

このように，式 7，11，13 より，各 Positive Category に属すユーザ数の推測値 A_i が計算される．

5. 提案手法の数学的解析

提案手法によって推測された各 Positive Category に属すユーザ数と，本来のユーザ数との誤差を解析する．

また，提案手法における計算量を導出する．

5.1. 誤差

$p_k = \hat{p}$ であるユーザ群における推測誤差は，式 13 で表されるとおりである．これを，p_k が異なる全てのユーザに拡張するため，誤差伝播の法則を利用する．誤差伝播の法則によると，$y = f(x_1, x_2, \ldots, x_n)$ で表される関数 y があり，各 x_i の分散が $\sigma_{x_i}^2$ で表されるとき，y の分散 σ_y^2 は次の式で表される．

$$\sigma_y^2 = \sum_{i=1}^{n} \left(\frac{\partial y}{\partial x_i} \right)^2 \sigma_{x_i}^2 \tag{14}$$

提案手法では，y に相当するものが $P(X = C_i)$，つまり A_i/N である．また，各 $\hat{p}$ を $\hat{p}_1, \hat{p}_2, \ldots$ とすると，$x_1, x_2, \ldots$ に相当するものが $P(A_{\hat{p}_1} = C_i), P(A_{\hat{p}_2} = C_i), \ldots$，つまり $A_{\hat{p}_1,i}/S_{\hat{p}_1}, A_{\hat{p}_2,i}/S_{\hat{p}_2}, \ldots$ である．したがって式 11，14 より，提案手法における RMSD は以下の式で表される．

$$E = \sqrt{\sum_{\hat{p}} \left(\frac{\partial \frac{A_i}{N}}{\partial \frac{A_{\hat{p},i}}{S_{\hat{p}}}} \right)^2 E_{\hat{p}}^2} = \sqrt{\frac{1}{\sum_{\hat{p}} 1/E_{\hat{p}}^2}} \tag{15}$$

5.2. 計算量

式 7 は，変数が F 個の連立一次方程式を解くことを表している．この計算量は，反復法のガウス＝ザイデル法を用いた場合，反復回数を m とすると $O(mF^2)$ である．p_k が異なるユーザ集合が N_u 組あるとすると，式 7 を用いた連立一次方程式の求解は，N_u 回実施する必要があるため，計算量は $O(mF^2N_u)$ となる．

また式 11 の計算は F 回実施する必要があるため，計算量は $O(FN_u)$ である．したがって，最終的な計算量は $O(mF^2N_u)$ で表される．

Straight 手法は $O(F)$ で計算することが可能である [19]．Straight 手法と比較すると提案手法の計算量のオーダーは大きい．しかし，十分現実時間内で計算可能であることを，6 章において示す．

6. 評価

式 15 を利用した推測誤差における評価及び，実際に True Category から Positive Category，Negative Category を生成して推測誤差を計算するシミュレーション評価を行った．比較対象は 3 章に記載した Straight，GSN，MDA 及び，以下で定義する **Variable 手法**である．

Straight，GSN，MDA は 3 章に示したとおり，ユーザ k の Negative Category から，確実にユーザ k の Positive Category でないカテゴリを特定できるという問題がある．多くのユーザが属すカテゴリが除外できてしまうと，ユーザ k の Positive Category がマイナーなカテゴリであることが判明してしまう．したがってこれら 3 手法は，本研究の目的をそもそも満たすことができないが，参考として比較対象とした．Straight は，あるユーザの Negative Category から，そのユーザの Positive Category でないカテゴリを確実に 1 つ特定できる．ここで，Positive Category の候補として除外できるカテゴリ数の割合を示す指標 s を導入する．Straight はカテゴリ数を F とすると $s = 1/F$ となる．GSN も厳密には $s = 1/F$ であるが，“ほぼ” 確実にユーザの Positive Category でないカテゴリを複数特定できる．99.999%以上の確率で特定できるものについてもカウントすることとする．MDA 及び GSN は，パラメータによって s の値が変動するが，いずれも $s \geq 1/F$ である．

本研究の目的を満たすことができる簡易的な手法として，Variable 手法を定義する．これは，確率行列として $M((1 - R_k)/F)$ を利用する手法である．提案手法，Variable 手法いずれも，$s = 0$ である．

デフォルト値として，ユーザ数 $N = 1,000$，カテゴリ数 $F = 50$，True Category から Positive Category を計測する精度 $\alpha_k = 0.8$，プライバシ漏洩リスク $R_k = 0.05$ と設定した．以下で特に明記しない場合は，これらの設定値を用いた．R_k の設定値に関する考察は，7 章で記す．

6.1. 数学的解析による評価

まず，True Category から Positive Category を計測する精度 α_k を全ユーザ共通で 0.05 から 1 まで変動させた結果を図 2 に示す．Straight，MDA，GSN については，上述した s の値も参考として示している．Straight は，あるユーザの Negative Category の情報から，そのユーザの Positive Category の候補として 1 つのカテゴリを確実に除外できるため，$s = 1/50 = 0.02$ である．MDA は，50 のカテゴリを 2 次元化することで Straight 手法よりも RMSD を削減しているが，s の値が 0.28 と大きい．言い換えると，あるユーザの Negative Category の情報から，そのユーザの Positive Category の候補として 28%，つまり 14 のカテゴリを除外できてしまう．MDA はカテゴリをさらに多次元化することによって RMSD を削減可能であるが，s の値がさらに増大する．GSN についても同様である．

Straight，MDA，GSN はこのような問題を許容しているにも関わらず，$s = 0$ を実現する提案手法がこれら既存手法と比べてさらに RMSD を削減できていることが分かる．$\alpha_k = 0.8$ の場合は 30% 以上，$\alpha_k = 1$ の場合でも 14% 以上の削減ができている．Variable 手法と比較すると，95% 以上 RMSD を削減している．

プライバシ漏洩リスク R_k を全ユーザ共通で 0.05 から 0.5 まで変動させた結果を図 3 に示す．Variable 手法と比較すると，提案手法は RMSD を大きく削減できている．Straight，MDA，GSN と比較した場合においても，同等以下まで RMSD を削減している．

次に，ユーザ数 N を 100 から 10,000 まで変化させて評価した結果を図 4 に示す．各手法ともに，$\sqrt{N}$ に比例して RMSD が削減されることが分かる．

カテゴリ数 F を 20 から 100 まで変化させて評価した結果を図 5 に示す．MDA 及び GSN に関しては，s を任意の値に設定することができないため，$0.2 < s < 0.3$

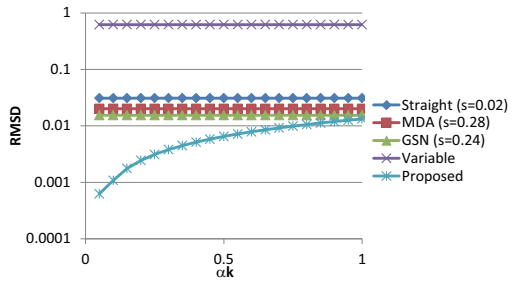


図 2: α_k を変化させたときの RMSD

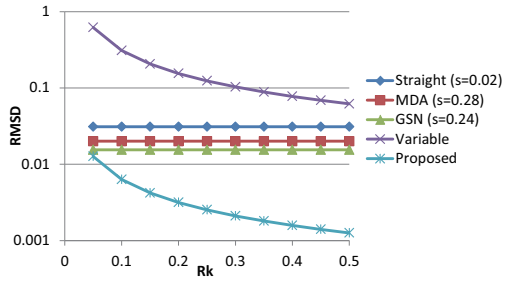


図 3: R_k を変化させたときの RMSD

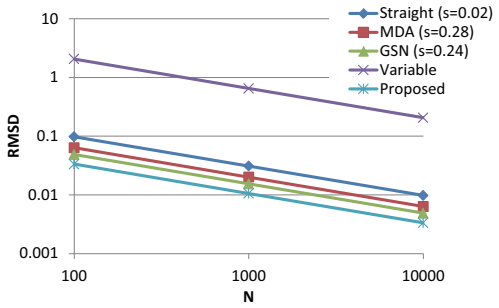


図 4: N を変化させたときの RMSD

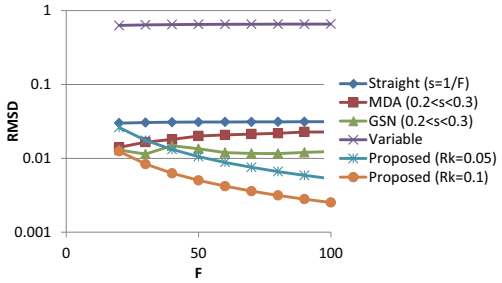


図 5: F を変化させたときの RMSD

の範囲の中で設定した．既存手法はカテゴリ数が増加しても RMSD はほとんど変わらないが，提案手法ではカテゴリ数に応じて適切な確率行列を構築しているため，RMSD を削減できている．$R_k = 0.05$ のとき，F の値が小さいと MDA や GSN の RMSD は提案手法よりも小さいが，$R_k = 0.1$ に緩和すると，提案手法のほうが RMSD を削減できることが分かる．

次に，R_k 及び α_k を全ユーザ共通で同時に変化させた結果を図 6(a) に示す．Straight，MDA，GSN の結果は図に載せていないが，R_k や α_k の値に関わらずそれぞれ 0.031，0.02，0.015 であった．また，Variable の結果は，提案手法よりも常に 10 倍以上の RMSD 値であった．範囲内全ての R_k 及び α_k について，提案手法の RMSD が比較対象の手法よりも下回っていることが分かる．

6.2. シミュレーション評価

数学的解析結果との評価誤差及び，各ユーザのプライバシ漏洩リスクや属性取得精度が異なる場合における実データを用いたシミュレーション結果を示す．

6.2.1. 数学的解析結果との評価誤差

式 13，15 で表される RMSD を求める式を導出するために，論文中で述べたとおりいくつかの仮定を置いている．これらの仮定が妥当であることを示すために，6.1 で行った評価結果と，シミュレーション結果が概ね一致することを示す．

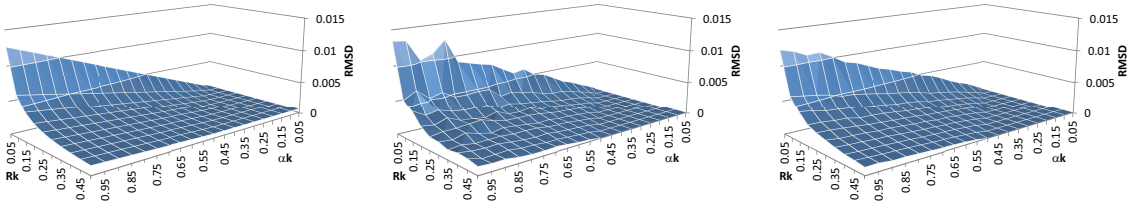
1,000 人のユーザを生成し，それぞれ True Category を割り当てた．このとき，各カテゴリに属すユーザ数がランダムになるように割り当てを行った．その後，設定した α_k に基づいて Positive Category を導出し，R_k に基づいて Negative Category を導出した．全ての Negative Category を導出した後，各 Positive Category に属すユーザ数を推測し，RMSD を算出した．シミュレーション回数を S とし，RMSD の平均値を算出した．6.1 において相当する評価結果は図 6(a) である．

$S = 1$, $S = 10$ としたシミュレーション結果を，それぞれ図 6(b)，図 6(c) に示す．$S = 1$ の場合においても数学的解析による評価結果と概ね一致しており，シミュレーション回数が増えるほど，数学的解析による評価結果に近づくことが分かる．

6.2.2. 実データを用いたシミュレーション評価

各ユーザの R_k や α_k をそれぞれ異なる値に設定した評価を，UCI の Adult データセット [1] を用いて行った．本データセットは，匿名化手法における研究分野で広く利用されている [12]，[14]．評価においては，Adult データセットの属性のうち，年齢，人種の 2 属性を利用した．年齢は 10 代，20 代のようにカテゴリ化して利用した．Adult データセットには 10 代から 90 代まで存在し，人種は White, Black, Amer-Indian-Eskimo, Asian-Pac-Islander, Other の 5 種類であった．したがって計 45 の組合せがある．データ総数は 32,561 件であった．

年齢及び人種の組合せにおけるデータ分布を，図 7 に示す．このデータに対して，確率 α_k で元のデータを Positive Category とし，$1 - \alpha_k$ の確率で異なるデータをランダムに選択し，Positive Category とした．



(a) Mathematical result　(b) Simlation result ($S = 1$)　(c) Simlation result ($S = 10$)

図 6: R_k と α_k の両方を変化させたときの RMSD（数学的解析及びシミュレーション）

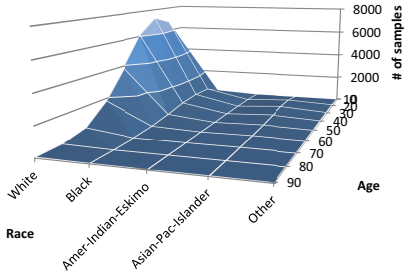


図 7: Adult データセットの属性分布

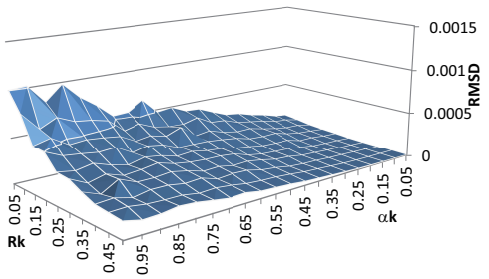


図 8: Adult データセットを用いたシミュレーション

推測誤差を測定した結果を図 8 に示す．R_k 及び α_k いずれも，平均値を図に示すとおりに設定し，標準偏差を 0.1 とした正規分布に基づいて値を設定した．ここで，正規分布は無限の値まで取り得るが，α_k については $1/F$ から 1 までを，R_k については 0.01 から 1 までを閾値として 0.05 単位で設定した．Straight($s = 1/45$)，MDA($s = 13/45$)，GSN($s = 13/45$)，Variable 手法に関しては結果を載せていないが，それぞれ約 0.0054，0.0030，0.0018，0.0068 であった．設定した R_k，α_k の値に関わらず，提案手法が比較対象の手法よりも RMSD を削減できていることが分かる．

6.2.3. 計算時間

提案手法において，収集された Negative Category のデータから，Positive Category のデータ分布を求めるために必要な時間を計測した．表 1 に実験環境を示す．

表 1: 実験環境

OS	Windows 7 Professional 64 bit
CPU	Intel Xeon CPU X5675 @ 3.07GHz 3.06 GHz
RAM	12.0 GB

上述した Adult データセットにおいて，収集された Negative Category のデータから，各 Positive Category に属すユーザ数を求めるために要した時間は，平均 1,517ms であった．この結果より，計算時間は十分実用に耐え得ると判断することができる．

7. 考察

7.1. ユーザ属性の計測精度 α における誤差

本論文では，True Category から Positive Category を取得する際の精度 α_k を正しく取得できるという前提を置いている．しかし実環境においては，この精度情報にも誤差が含まれ得る．本論文ではスコープ外とした，Positive Category のデータ分布から True Category のデータ分布を推測する際には，α_k の値をできるだけ正確に取得できるほうが良いと考えられる．一方，攻撃者の視点に立っても，α_k の値に誤差が無いほうが，各ユーザの True Category をできるだけ正確に判断することが可能である．したがって，α_k の取得精度に誤差がある場合においても，ユーザが設定する R_k は本論文で提案する手法に従うことで守ることができる．

7.2. プライバシ漏洩リスク R_k の設定値

攻撃者に全く情報を与えない場合，$R_k = 0$ と設定することになるが，この場合はサーバ側でも何も情報を得ることができない．$R_k = 0$ のとき，任意のカテゴリ C_i がユーザの True Category である確率は，カテゴリ数を 50 とすると 0.02 である．この確率を 5% だけ変動させる（つまり $R_k = 0.05$ に設定する）ことを許容すると，任意のカテゴリ C_i がユーザの True Category である確率の最小値は，0.019 である．このようにわずかだけ確率を変動させることで，提案手法は十分に RMSD を削減することができている．

また，厚生労働省における会議資料 [21] によると，匿名化された診療報酬明細書データベースには，“公表さ

れる成果物において，患者等の数が 10 未満になる集計単位が含まれてはならない” と記載されている．これを，患者の属性を 1/10 以上の確率で特定されてはならない，と言い換えることができるとする．本研究では，ユーザがあるカテゴリに属す確率の最大値は

$$1 - \frac{1 - R_k}{F} \cdot (F - 1) \tag{16}$$

で表すことができる．この値が 1/10 以下であれば良いと考えると，$R_k = 0.05$，$F = 50$ のとき，式 16 の値は 1/14.5 となり，この制約を十分満たしている．

これらより，通常は $R_k = 0.05$ 程度に設定しておき，よりプライバシ情報を提供してよいと考えるユーザは，R_k を 0.05 以上の任意の値に設定すると良いと考えられる．

8. おわりに

本論文では，取得されたユーザ属性に誤差がある場合や，各ユーザでプライバシ漏洩リスクを変更したい場合に対応する，新しい Negative Survey 手法を提案した．既存の Negative Survey では，全ユーザでプライバシ漏洩リスクを同一に設定する必要があったが，提案手法を用いることにより，柔軟なプライバシ漏洩リスクの設定が可能となった．

また，既存手法はユーザの属性候補から，ある一定範囲を攻撃者が除外できるという問題があった．提案手法はプライバシ漏洩リスクを定義し，この問題に対応した上で，デフォルト値を設定した場合の推測誤差を，従来の手法と比較して 30% から 90% 以上削減した．

将来課題として，Positive Category から True Category への推測も総合的に考慮した Negative Survey アルゴリズムの提案を行う必要がある．また，離散化されたカテゴリではなく，連続値で表現されるユーザ属性を匿名化する手法に拡張することを考えている．

参考文献

[1] : UCI Machine Learning Repository: Adult Data Set, `http://archive.ics.uci.edu/ml/datasets/Adult`.

[2] Agrawal, R. and Srikant, R.: Privacy-Preserving DataMining, *Proc. ACM SIGMOD*, pp. 439–450 (2000).

[3] Chen, R. and Reznichenko, A.: Towards statistical queries over distributed private user data, *Proc. USENIX NSDI*, pp. 169–182 (2012).

[4] Corburn, J.: Confronting the challenges in reconnecting urban planning and public health, *American journal of public health*, Vol. 94, No. 4, pp. 541–546 (2004).

[5] Evfimievski, A., Gehrke, J. and Srikant, R.: Limiting privacy breaches in privacy preserving data mining, *Proc. ACM PODS*, pp. 211–222 (2003).

[6] Forrest, S. and Groat, M.: Reconstructing Spatial Distributions from Anonymized Locations, *Proc. IEEE ICDEW*, No. Section II, pp. 243–250 (2012).

[7] Froehlich, J., Larson, E., Gupta, S., Cohn, G., Reynolds, M. and Patel, S.: Disaggregated End-Use Energy Sensing for the Smart Grid, *IEEE Pervasive Computing*, Vol. 10, No. 1, pp. 28–39 (2011).

[8] Gkoulalas-Divanis, A., Kalnis, P. and Verykios, V. S.: Providing K-Anonymity in location based services, *ACM SIGKDD Explorations Newsletter*, Vol. 12, No. 1, p. 3 (2010).

[9] Groat, M. M., Edwards, B., Horey, J., He, W. and Forrest, S.: Enhancing privacy in participatory sensing applications with multidimensional data, *Proc. IEEE PerCom*, pp. 144–152 (2012).

[10] Horey, J., Groat, M. M., Forrest, S. and Esponda, F.: Anonymous Data Collection in Sensor Networks, *Proc. MobiQuitous*, IEEE, pp. 1–8 (2007).

[11] Huang, Z. and Du, W.: OptRR: Optimizing Randomized Response Schemes for Privacy-Preserving Data Mining, *Proc. IEEE ICDE*, pp. 705–714 (2008).

[12] LeFevre, K., DeWitt, D. and Ramakrishnan, R.: Incognito: Efficient full-domain k-anonymity, *Proc. ACM SIGMOD*, pp. 49–60 (2005).

[13] Li, Z., Li, M., Wang, J. and Cao, Z.: Ubiquitous data collection for mobile users in wireless sensor networks, *Proc. IEEE INFOCOM*, pp. 2246–2254 (2011).

[14] Machanavajjhala, A., Kifer, D., Gehrke, J. and Venkitasubramaniam, M.: l-diversity: Privacy beyond k-anonymity, *ACM TKDD*, Vol. 1, No. 1, pp. 3–es (2007).

[15] Platon, E. and Sei, Y.: Security software engineering in wireless sensor networks, *Progress in Informatics*, Vol. -, No. 5, pp. 49–64 (2008).

[16] Santana de Oliveira, A., Kerschbaum, F., Lim, H. W. and Yu, S.-Y.: Privacy-Preserving Techniques and System for Streaming Databases, *IEEE PASSAT*, pp. 728–733 (2012).

[17] Sharp, C., Schaffert, S., Woo, A., Sastry, N., Karlof, C., Sastry, S. and Culler, D.: Design and implementation of a sensor network system for vehicle tracking and autonomous interception, *Proc. Second European Workshop on Wireless Sensor Networks*, IEEE, pp. 93–107 (2005).

[18] Shi, P., Xiong, L. and Fung, B.: Anonymizing data with quasi-sensitive attribute values, *Proc. ACM CIKM*, pp. 1389–1392 (2010).

[19] Xie, H., Kulik, L. and Tanin, E.: Privacy-aware collection of aggregate spatial data, *Data & Knowledge Engineering*, Vol. 70, No. 6, pp. 576–595 (2011).

[20] Ye, F., Yang, H. and Liu, Z.: Catching ”Moles” in Sensor Networks, *Proc. IEEE ICDCS*, pp. 69–69 (2007).

[21] 厚生労働省：医療機関等における個人情報保護のあり方に関する検討会資料「安全に匿名化等がされた状態について」 (2012).