修士論文

参加型センシングにおける
プライバシー保護手法

Privacy-Preserving Participatory Sensing

2014年2月6日

指導教員　瀬崎 薫　教授

東京大学大学院情報理工学系研究科
電子情報学専攻　48-126401

青木 俊介

内容梗概

我々の生活する社会では，様々な事象を把握するために，多彩な方法で調査を行う．しかしながら，経済的なコストや調査員の労力の限界があるため，必ずしも現実と一致した質の高い情報を取得できているとは言えない．このような問題を背景に，世界中で普及しつつあるモバイルデバイスをセンサとして活用する参加型センシングの実現が様々な分野で期待されている．参加型センシングでは，膨大な数の一般ユーザからセンサデータを収集し，詳細な情報を取得することで，既存の調査法では得ることのできない知識・情報を明らかにする．センサデバイスの数は，参加するユーザの数に比例するため，より多くのユーザが参加型センシングに貢献することのできる環境を構築することが重要である．しかし，モバイルデバイスからのセンサデータには，ユーザのコンテキストに関するプライバシー情報が含まれるため，プライバシー情報の侵害・漏洩が非常に大きな問題となる．これらをふまえ，本研究では参加型センシングを実現するプライバシー保護手法を提案し，一般ユーザが安心して参加型センシングに参加する環境を構築する．本研究では，実社会で用いる参加型センシングを想定し，それぞれのに最適なプライバシー保護手法を設計する．

本研究ではまず，都市環境のための参加型都市環境センシングを実装し，本システムを社会で活用するためのプライバシー保護手法を提案する．本提案手法では，都市環境センシングで扱うデータの特徴を考慮し，位置情報・環境情報を細かいデータ区分で扱うことのできる手法を開発する．本手法の実現によって，位置情報・環境情報を同時に収集することが可能となり，モバイルデバイスを用いた住生活環境の調査や道路モニタリングを実現することができる．

続いて，二者択一型選択肢を持つ多次元データを扱うためのプライバシー保護手法を提案する．本提案手法では，既存の Perturbation 技術を組み合わせて，二者択一型選択肢を持つ多次元データの取り扱いを可能にする．本手法によって，性別や疾患の有無などのプロファイル情報を扱う参加型センシングや，閾値判定情報を扱うイベント検知のための参加型センシングを実現できる．

次に，医療調査用参加型センシングシステムを実現するプライバシー保護手法を提案する．本提案手法では，2 段階のプライバシー保護処理を施すことによって，時間依存性の低い医療情報の取り扱いを可能にする．加えて，センシングで取得したデータとプライバシー保護処理済データの関係に注目し，中央サーバ上におけるデータ処理に必要となる計算量を大幅に削減する．本手法を用いることによって，慢性疾患と環境要因の関係を明らかにする統計的な情報を長期間取得することのできるインフラとして参加型センシングを活用することができる．

最後に，参加型センシングで取得した情報を一般社会に公開した際に，悪意ある攻撃者がこの情報を利用することを防ぐために，取得した情報そのものをプライバシー保護に用いるプライバシー保護機構を提案する．本機構では，取得した情報に応じてプライバシー保護処理に用いる確率分布を変化させることによって，データ収集から利活用・公開まで一貫したプライバシー保護を実現し，参加型センシングで取得した情報を一般社会に還元することを可能とする．

本研究により，個々の一般ユーザのプライバシー情報を侵害することなく，様々な現象を解き明かす情報を取得する参加型センシングを社会で活用することができる．プライバシー情報の侵害・漏洩に煩わされることなく，参加型センシングを活用することのできる環境を構築することによって，一般ユーザの持つモバイルデバイスを新たな社会インフラとして活用し，医療・健康分野や都市工学・環境学の分野で，人類にとって新たな知・発見をもたらすことができると確信している．

目 次

第 1 章　序論 1
1.1　はじめに 1
1.2　本論文の構成 3

第 2 章　関連研究 5
2.1　概要 5
2.2　モバイルデバイスを用いたセンシング技術の概要 5
2.2.1　参加型センシング 5
2.2.2　ユーザコンテキストに関するセンシング 8
2.3　データマイニングのためのプライバシー保護技術 10
2.3.1　暗号技術を用いたプライバシー保護手法 10
2.3.2　k-匿名性を用いたプライバシー保護手法 10
2.3.3　差分プライバシー (Differential Privacy) 11
2.3.4　Perturbation 技術を用いたプライバシー保護手法 12
2.4　本章のまとめ 17

第 3 章　モバイルデバイスを用いた参加型センシングとプライバシー情報 19
3.1　参加型センシングにおけるプライバシー情報の侵害 19
3.2　事前実験と技術的な課題 22

第 4 章　参加型都市環境センシングとプライバシー保護 27
4.1　参加型都市環境センシング 27
4.1.1　参加型都市環境センシングの概要 27
4.1.2　システムの実装と歩行センシング実験 28
4.1.3　データ校正 31
4.1.4　センサデータの可視化と考察 36
4.1.5　効率的なシステム運用のためのユーザ分析 38
4.1.6　自然環境調査への拡張 39
4.2　参加型都市環境センシングのためのプライバシー保護手法 44
4.2.1　提案手法: Selective Negative Surveys 45
4.2.2　評価 48
4.2.3　提案手法に関する考察 52
4.3　本章のまとめ 53

第 5 章　プロファイル情報を対象とした参加型センシングとプライバシー保護 55
5.1　二者択一型選択肢を含む多次元データを扱う参加型センシングシステムの概要 . . 55
5.2　プロファイル情報を対象とした参加型センシングシステムのためのプライバシー保護手法 56
5.2.1　提案手法: MNS-RRT アルゴリズム 56

5.2.2 評価 61
5.3 本章のまとめ 65

第 6 章 医療調査用参加型センシングとプライバシー保護 68
6.1 医療調査用参加型センシングの概要 68
6.2 医療調査用参加型センシングのためのプライバシー保護手法 69
6.2.1 一次元データに対する提案手法: Duplicated Perturbation 70
6.2.2 多次元データに対する提案手法: Duplicated Perturbation 74
6.2.3 評価 76
6.3 本章のまとめ 78

第 7 章 参加型センシングのデータ公開とプライバシー保護 80
7.1 参加型センシングのデータ公開の概要 80
7.2 公開情報や背景知識に対するプライバシー保護機構 Data-centric Perturbation Infrastructure 81
7.2.1 Data-centric Perturbation Infrastructure のシステム概要 82
7.2.2 提案手法: Empirical Negative Surveys 83
7.3 本章のまとめ 86

第 8 章 プライバシー保護に関する考察 87
8.1 技術的な課題に関する考察 87
8.2 データ信頼性とプライバシーレベルに関する考察 88

第 9 章 結論 90
9.1 本研究の主たる貢献 90
9.2 今後の課題と展望 91

謝辞 92

図 目 次

1.1 スマートフォン普及の割合 (2013)[2]. 2
1.2 参加型センシングシステムの概要図. 4

2.1 Haze Watch プロジェクトで設計したシステムの概要 [17]. 6
2.2 BikeNet のシステムの概要 [17]. 6
2.3 sBike のシステムの概要 [22]. 7
2.4 sBike で取得したデータの例 [22]. 7
2.5 S. P. Tarzia ら研究 [26] では，背景音をフィンガープリントのように用いて屋内位置を推定する. 8
2.6 J. R. Kwapisz らの提案するモバイルシステムでは，3 軸加速度のみでユーザの行動認識を実現する [8]. 9
2.7 E. De Cristofaro らの提案する暗号技術を用いたプライバシー保護機構 PEPSI[34]. 11
2.8 L. Sweeney が指摘した linked attack[37]. 11
2.9 差分プライバシーのシステム概要 [42]. 12
2.10 R. K. Ganti らの PoolView のシステム概要 [52]. 13
2.11 Negative Surveys を用いた時，回答者は他人に知られたくないセンシティブ情報を報告する必要がない [64]. 16
2.12 取得したオリジナルデータに基づく環境騒音地図及び MNS の手法を用いた時の環境騒音地図 [67]. 17

3.1 二酸化炭素濃度センサの一日の挙動 [12, 69]. 20
3.2 既存手法 MNS[59, 67] を用いた時のプライバシー保護処理の様子．評価指標としては，式 (3.1) のカルバックライブラー情報量 D_{KL} を用いた. 22
3.3 センサデータとして $\vec{x} = <x_1 = 2, x_2 = 1>$ を取得した時の様子．既存手法 MNS[59, 67] を用いた時，モバイルデバイス上で白セルの中から等確率で 1 つのセルを選択し，サーバへと報告する．このため α_2 のフィールド情報に関して，オリジナルデータの推定を容易に行うことができる. 23
3.4 攻撃者が公開情報や背景知識を利用すると，より容易にプライバシー情報を取得できる [64]. 26
3.5 オリジナルデータが一様分布である時 (左図) と偏りのある分布である時 (右) の Negative Surveys[58] の処理の様子. 26

4.1 国立環境研究所が公開している全国自動車交通騒音マップ (東京都 世田谷区)[73]. 28
4.2 国立環境研究所が公開している全国自動車交通騒音マップ (熊本県 熊本市)[73]. . 29
4.3 都市環境センシングアプリケーションのインターフェース. 30
4.4 東京都 世田谷区における既存の環境情報取得地点 (左図) と地価サンプル地点 (右図). 30
4.5 k-means 法によって 60 に分割したセンシング領域. 31
4.6 各エリアのポイント数. 32
4.7 歩行センシングの様子. 32

4.8 4π 無響室. 33
4.9 端末間校正実験の様子. 33
4.10 音圧レベルと各端末の等価騒音レベルの関係. 34
4.11 Audacity[81] による取得データのボリュームレベルの推移の様子. 35
4.12 相対的な雑音発生レベルの導出によるユーザ間校正手法の様子. 35
4.13 歩行センシング実験で取得したセンシングポイント. 36
4.14 あるユーザの足跡に沿った等価騒音レベルの様子. 36
4.15 世田谷区内全域の等価騒音レベルの様子. 37
4.16 歩行センシング実験で取得した等価騒音レベルの様子. 37
4.17 各ユーザのタスク量 (センシングエリアの数). 39
4.18 提案するインセンティブモデルの例. 39
4.19 野鳥の音声自動識別システムの概要. 40
4.20 既存の野鳥生態調査法に関するイメージ図. 40
4.21 Cicada Fingerprinting System の概要図．異種のセミの鳴き声を用いて位置補正を行う. 41
4.22 Cicada Fingerprinting System の評価実験を行ったエリア．周辺の地図を図 4.23 に示す. 41
4.23 Cicada Fingerprinting System の評価実験を行ったエリアの地図. 42
4.24 地点 A で取得した音声の周波数分析の結果. 42
4.25 地点 B で取得した音声の周波数分析の結果. 43
4.26 夏期にセミの鳴き声を取得した 4ヶ所の様子. 44
4.27 図 4.26 の 4ヶ所で取得したセミの鳴き声を含む音声データを周波数スペクトル解析した結果. 44
4.28 8×13 の 2 次元データを取得する時の提案手法 SNS での処理の様子．$\{x_1 = 4, x_2 = 4\}$ をセンシングした時，白セルの中から 1 つのセルをランダムに選んで伝送する. . 46
4.29 ユードム社の無線型二酸化炭素測定器 (左) と Android 端末 Galaxy S3(右). . . . 49
4.30 既存手法 MNS 及び提案手法 SNS を用いた時の二酸化炭素濃度のデータ分布の様子. 50
4.31 環境情報確率分布を用いた時のデータ整合率 $\mathbf{C}$ の推移. 54
4.32 正規分布を用いた時のデータ整合率 $\mathbf{C}$ の推移. 54

5.1 センシングで取得したデータが $\vec{x} = <x_1 = 2, x_2 = 1>$ である時，MNS-RRT アルゴリズムにおけるプライバシー保護処理で生成されるデータの生成確率は式 (5.3) で示す $\mathbf{\Psi_4}$ の様になる. 58
5.2 天候観測用参加型センシングのデータに対して MNS-RRT アルゴリズムを適用した時の様子．95% の入力データが正確な情報であると想定した．またパラメータ p の値は 0.2 とした. 62
5.3 4 つのエリアに紐づく形で，3 段階の建物損壊レベル，2 段階の放射線レベルを参加型センシングで取得した時の様子. 64
5.4 6×2 の 2 次元データ，9×2 の 2 次元データ，4×3×2 の 3 次元データ，3×3×3×2 の 4 次元データに対して，それぞれ MNS-RRT アルゴリズムを施した時のデータ信頼性に関する結果．5 回の処理を施し，その平均値を記す．本提案手法では，統計処理を用いるため，取得データ数が少ない場合にはデータ信頼性は低下する. . . 66

5.5 6×2 の 2 次元データ，9×2 の 2 次元データ，$4\times3\times2$ の 3 次元データ，$3\times3\times3\times2$ の 4 次元データに対して，それぞれ MNS-RRT アルゴリズムを施した時のプライバシー保護に関する評価結果．パラメータ p の値が 0.5 に近付くほど，より強固にプライバシー情報を保護することができる． 67

6.1 モバイルデバイスを用いた医療調査用参加型センシングの概要図 [98]. 69
6.2 一次元データに対する Duplicated Perturbation のプロトコル. 71
6.3 Duplicated Perturbation による処理を 6×4 の 2 次元データに施した時の様子．センシングしたオリジナルデータ (Original data) が $\langle b, 2\rangle$ であった時，灰色セルの中から 1 つのセルが等確率に，一次摂動化データ (First-order perturbed data) として選ばれる．また，同様の処理で一次摂動化データから二次摂動化データ (Second-order perturbed data) が選択される． 75
6.4 m-Health による慢性疾患調査システムに提案手法 Duplicated Perturbation を用いた時の様子．4×3 の 2 次元データを扱った時の，オリジナルデータのデータ分布，一次摂動データのデータ分布，二次摂動データのデータ分布，再構築データのデータ分布を示す． 78

7.1 既存のプライバシー保護システムの概要．モバイルデバイス上のプライバシー保護処理には，一定の確率分布関数を用いるため，常に同様の処理を施す． 81
7.2 フィードバック型プライバシー保護機構 Data-centric Perturbation Infrastructure の概要．プライバシー保護処理に用いる確率分布は，取得した統計的なデータ分布そのものに基づいたものとなる． 82
7.3 オリジナルデータ (Common Surveys)，既存手法におけるプライバシー保護処理済データ (Negative Surveys[58])，提案手法におけるプライバシー保護処理済データ (Empirical Negative Surveys) の比較図 [64]. 84

表 目 次

4.1 6×7 の 2 次元データを用いた時のデータ信頼性とプライバシーに関する評価. . . 49
4.2 環境情報確率分布を用いた時のデータ信頼性とプライバシーに関する評価. 51
4.3 正規分布を用いた時のデータ信頼性とプライバシーに関する評価. 52
4.4 センサデータとプライバシー保護処理済データの例. 53

5.1 プロファイル情報とセンサデータを同時に収集する多次元データの例. 56
5.2 図 5.2 のシミュレーションにおいて，パラメータ p の値を変えた時の結果．10 回の処理を行い，その平均値を示した. 63

6.1 図 6.4 で行ったシミュレーション実験においてユーザ数を変化させた時の評価. . 77
6.2 提案手法 Duplicated Perturbation を様々な多次元データに適用した時の様子. . . 79

第1章　序論

1.1　はじめに

我々の生活する社会では，様々な事象を把握するために，無作為抽出を伴う統計的社会調査が幅広く活用されている．しかし，無作為抽出を伴う既存の調査法では，統計の母集団に偏りが生じる恐れがあるのは明らかである．また，調査員の労力や調査に費やすことのできる経済的コストには限界があるため，必ずしも現実社会と一致した質の高い情報を取得できるとは言えない．例えば，精神疾患や慢性疾患の原因調査のためには，長期間にわたって様々な情報を頻繁に調査する必要があるが，調査参加者と調査実行者の労力を考慮すると，実現は極めて難しい．このため，現状の調査では精神疾患や慢性疾患の多くの原因は判明しておらず，一般的な傾向のみが社会の共有知となっている．また，住生活環境の調査や不動産価値の評価のために，都市環境の調査や人の流れの調査 (パーソントリップ調査) が行われているが，現状の調査では特定の日時に，限られた地点に調査員を配置した調査法となっているため，調査日時以外の情報は取得できていないのが現状である．

このように，実社会の情報を適切に調査できていない問題を解決するために，世界中で急速に普及しているモバイルデバイスをセンサとして活用する参加型センシング (Participatory Sensing, Community Sensing)[1] の実現が期待されている．本論文では，一般ユーザが持ち歩くことのできる腕時計型，眼鏡型，靴型のウェアラブルデバイス及びスマートフォン，タブレット端末などを総称して，モバイルデバイスと呼ぶこととする．特にスマートフォンは図 1.1[2] に示すように，地域に依ることなく世界中で普及しており，現在もこの普及率は上昇し続けている [2].

参加型センシングでは，図 1.2 に示すように，一般ユーザが所有し，普段の生活で持ち歩くモバイルデバイスに搭載された GPS 受信機や加速度センサ，照度センサ，マイク等を用いて周囲の情報を取得する．このため，既存のセンサネットワークとは大きく異なり，センサ設置などの初期投資コストがほとんど必要無いという利点がある．また，参加型センシングでは，モバイルデバイスで取得したセンサ情報やユーザの入力情報を収集することによって，大規模かつ詳細な統計的な情報を集積することが可能となる [3]．この大規模かつ詳細な統計的な情報は，現状の社会調査では決して取得できない質・量の情報であるため，データ解析を行うことによって，人類にとって新たな知見や仮説を導き出し，社会にとって大きな貢献を生み出すことが期待できる．特に，医療・健康分野での利活用は既に米国 国立衛生研究所 (National Institutes of Health) の Genes and Environment Initiative が中心となって検討されており，現在の診療・調査では原因究明が難しい精神疾患・慢性疾患の遺伝的要因・環境的要因を明らかにすることが期待されている [4, 5].

この参加型センシングは，固定型センサや埋め込みセンサを用いるセンサネットワークとは大きく異なり，一般ユーザの持つモバイルデバイスをセンシングに用いる．一般ユーザの持つモバイルデバイスは，ユーザ自身と近い距離を保つことが A. K. Dey らの研究 [6] により明らかになっており，またモバイルデバイスに搭載されたセンサ群を用いて所有者である一般ユーザの行動・コンテキスト・感情を認識・推定する研究 [7, 8] やライフログ取得に関する研究 [9] が進んでいることを考慮すると，参加型センシングで取得するセンサ情報からのプライバシー情報の侵害・漏洩が非常に重要な問題となる [10].

一方で，一般ユーザが持ち歩くモバイルデバイスの数には限りがあることは明らかであるため，

参加型センシングシステムで活用するセンサの数は，センシングに参加する一般ユーザの数に依存すると言える．このため，より質の高い統計的な情報を収集し，信頼性の高い社会調査を実現するためには，より多くの一般ユーザをセンシングに巻き込む必要がある．

これらを考慮すると，実社会で参加型センシングシステムを運用し，社会的な調査を行うためには，適切なプライバシー保護手法を確立し，一般ユーザが安心してセンシングに参加できる土壌を構築することが非常に重要である [10].

また，参加型センシングシステムを運用するアプリケーション提供者にとって，社会的な統計を取得することが目的であり，一般ユーザのプライバシー情報を取得・保持する必要は無い．このため，センシングシステムのアプリケーション提供者側が，個々のプライバシー情報を保持する危険性を有すること無く，統計的な情報を取得するシステムの実現が望ましい [11].

これらをふまえて本研究では，都市環境調査や医療調査などの分野を対象に，参加型センシングで目的とする社会調査の概要と取り扱うデータの特性・想定するデータの利用方法について明らかにした後に，一般ユーザが自身のプライバシー情報を他者に開示することの無い手法を提案する．本提案手法では，一般ユーザの持つモバイルデバイス上で，全てのセンサデータに対してプライバシー保護処理を施すことによって，プライバシー情報を秘匿化する．但し，スマートフォンをはじめとするモバイルデバイスに搭載されている計算資源・電源は限りのあるものであるため，プライバシー保護処理のための計算量を抑えた設計とする．本研究を通して，参加型センシングの安心・安全な運用を実現し，世界中に普及しつつあるモバイルデバイスをインフラとして活用することのできる社会を実現することを目指す.

図 1.1: スマートフォン普及の割合 (2013)[2].

1.2　本論文の構成

本論文の構成を以下に示す.

まず，第 2 章では，本研究に関連する，モバイルデバイスを用いたセンシング技術及びプライバシー保護技術に関する先行研究について紹介し，参加型センシングのプライバシー保護として適切なプライバシー保護技術について検討する.

続いて，第 3 章では，モバイルデバイスを用いた参加型センシングの全体像を述べ，Android 端末を用いた事前実験を行うことによって既存のプライバシー保護技術の課題について明らかにする.

第 4 章では，モバイルデバイスを用いた参加型都市環境センシングを開発・実証実験を行い，取得データの特徴を明らかにした上で，プライバシー保護手法 Selective Negative Surveys を提案する.

続けて，第 5 章では，プロファイル情報を対象とした参加型センシングを想定したプライバシー保護手法 MNS-RRT アルゴリズムを提案し，シミュレーションによってプライバシーとデータ信頼性から評価を行う.

第 6 章では，モバイルデバイスによる医療調査用参加型センシングを実現するプライバシー保護手法 Duplicated Perturbation を提案し，シミュレーションによる評価を示す.

第 7 章では，参加型センシングで取得したデータの公開方法とデータ利用について言及した上で，フィードバック型プライバシー保護機構 Data-centric Perturbation Infrastructure を提案する.

次に，第 8 章では，本論文で提案した手法を参加型センシングに用いた時の，センサデータとプライバシー保護されたデータの関係について考察を行い，また各アプリケーションで求められるプライバシーレベルとデータ信頼性に関する考察・議論を行う.

最後に，第 9 章にて全体をまとめた上で，今後の課題を明らかにする.

図 1.2: 参加型センシングシステムの概要図.

第2章　関連研究

2.1　概要

多種のセンサを搭載したモバイルデバイスの急速な普及に伴って，一般ユーザの持つモバイルデバイスをセンサとして用いる参加型センシングの利活用が，公衆衛生や交通調査，都市環境調査など様々な分野で期待されている．一方で，モバイルデバイスは一般ユーザの日常生活と密接な関係を築くため，モバイルデバイスで取得したセンサデータからのプライバシー情報の侵害は，参加型センシング実現の大きな課題である [10].

このような背景を踏まえて，本章ではまず参加型センシングの適用例とシステムの概要を示す．次に，モバイルデバイスから取得したセンサデータを用いた行動認識技術やユーザコンテキスト推定技術を示すことによって，センサデータからのプライバシー情報の侵害という危険について述べる．最後に，参加型センシングを実社会で実現するために非常に重要となるプライバシー保護技術について，既存の手法である暗号技術・k-匿名性・差分プライバシー・Perturbation 技術を紹介する．

2.2　モバイルデバイスを用いたセンシング技術の概要

様々なセンサを搭載したスマートフォンやウェアラブルデバイスの普及を背景に，一般のユーザが所有するモバイルデバイスをセンサとして用いるモバイルセンシングシステムに関する技術が盛んに研究されている [12]．このモバイルセンシングシステムは，ユーザからの情報を統合し，統計的なデータ分布の取得を目指す参加型センシング [13] と，ユーザコンテキストを取得するセンシング [14] の 2 つに分けることができる．

2.2.1　参加型センシング

参加型センシングでは，多数のユーザのモバイルデバイスをセンサとして用いることによって，広範囲・細粒度のデータを収集し，統計的なデータ分布を取得することを目指す [13]．参加型センシングにおけるセンサ密度は，ユーザの数に比例するため，特に人口密度の高い都市部で有用であると考えられ，都市環境調査 [15, 16, 17] や交通状況の調査 [18, 19]，公衆衛生 [20, 21] の分野での研究が進んでいる．参加型センシングのシステムでは，図 1.2 に示すようにセンシング参加者の合意に基づいて入力されたプロファイル情報とセンサで取得したセンサデータを収集し，中央サーバで統計処理を行う．

N. Maisonneuve らは都市における住生活環境調査を目的として，NoiseTube[15, 16] を設計・開発した．NoiseTube では，Nokia N95 に搭載されたマイク及び GPS 受信機を用いて，人間の生活環境における環境騒音情報を取得し，統計処理を行うことで環境騒音地図を作成する．また，図 2.1 に示す Haze Watch プロジェクト [17] では，モバイルデバイスに大気センサを外部センサとして搭載することによって，都市における二酸化炭素濃度やオゾン濃度などを測定する大気汚染調査システムを提案した．これらの都市環境調査は従来，特定の測定地点で，特定の測定時間に，特定の専用機器を用いていたために計測に多大な労力を要した．モバイルデバイスに搭載されたセンサ

図 2.1: Haze Watch プロジェクトで設計したシステムの概要 [17].

図 2.2: BikeNet のシステムの概要 [17].

を用いると，従来の専用機器を用いた場合に比べて測定データの信頼性は低下するが，動的にセンシングすることによって物理的，時間的に広範囲の情報を取得することが可能となる．

また，モバイルデバイスを用いた道路交通モニタリングを目的とする参加型センシングも盛んに研究されている [18, 19]．P. Mohan らは，モバイルデバイスに搭載したマイク・加速度センサ・GPS 受信機を活用し，多数のユーザからクラクション音や急ブレーキ，衝撃の発生地点を収集することによって，リアルタイムに交通状況や路面状況を共有する交通モニタリングシステム Nericell を提案している [18]．更に，S. Eisenman らは，多数のセンサをユーザ及び自転車に装着し，自転車に乗っている人からのセンサデータを収集することによって，サイクリングに最適な道や周囲の大気汚染度などを同時に明らかにする BikeNet を設計した [19]．このシステム概要を図 2.2 に示す．BikeNet では，自転車の速度，ペダルの回転数，機種磁方位に加えて，周囲の二酸化炭素濃度，環境騒音レベル，車両通行量などに関する情報を取得する．斉藤らは，Android 端末を用いて自転車走行状態収集・共有機構 sBike を提案している [22]．このシステム概要を図 2.3 に示す．また，sBike で得ることのできた走行状態に関する情報を，地図上に表示したものを図 2.4 に示す．自転車の走行状態を共有・可視化することで，障害物や人ごみの多い道が明らかにできる他，信号によ

図 2.3: sBike のシステムの概要 [22].

図 2.4: sBike で取得したデータの例 [22].

る停止時間の長い交差点を直感的に把握することが可能となる．

このように，人間の周囲の環境・状況を収集・共有する参加型センシングの他に，人間そのものをセンシングし，情報を収集・共有する参加型センシングも既存研究では数多く開発されている．特に，スマートフォンやウェアラブルデバイスは所有者であるユーザと多くの時間を同じ空間で過ごすことから，モバイルデバイスを用いた医療 (m-Health) の実現が期待され，公衆衛生や医療研究の分野での活用方法が盛んに研究されている [23, 24]．この例として，M. Annavaram らはウェアラブルデバイスを活用することによる小児肥満症の患者の行動・生活習慣モニタリングシステムを考案している [20]．また，E. Kanjo らは，外部センサをモバイルデバイスに搭載することによって，喘息を慢性的に患う一般ユーザの喘息発生地点を地理的なデータとして収集し，喘息の発作と周囲の大気汚染度を計測するシステムを設計・開発した [21]．医療分野で用いる参加型センシング

では，特定の疾患を持つ患者からの情報を収集し，統計的な傾向を明らかにすることによって，疾患と環境要因・遺伝要因の関係を明らかにすることを目的とする．

以上で述べたように，参加型センシングは多数のユーザから情報を収集し，統計的なデータ分布を明らかにすることが目的である．また，位置情報と環境情報の相関性や，ユーザのプロファイル情報と加速度センサ情報の相関性を明らかにすることが目的であるため，多次元データの持つ複数フィールド情報の相関性を取得することが重要である．

2.2.2 ユーザコンテキストに関するセンシング

統計的な情報を取得することを目的とする参加型センシングの研究が盛んに行われている一方で，モバイルデバイスに搭載された 3 軸加速度センサや照度センサ，マイク等を用いて，ユーザのコンテキストや感情，精神状態を推定する技術も急速な発展を遂げている．

特に，GPS 受信機からのセンサデータに頼ることなく，ユーザの位置情報を推定する技術の発展は著しく [25, 27, 28]，M. Azizyan らはモバイルデバイスに搭載された Wi-Fi 受信機，マイク，照度センサ，加速度センサ，カメラを用いてユーザの位置を推定する手法を提案している [25]．また，S. P. Tarzia ら [26] は，マイクから取得した音声データの背景音のみを用いた屋内測位手法を提案している．S. P. Tarzia らの手法では，図 2.5 に示すように，取得した音声データの背景音に

図 2.5: S. P. Tarzia ら研究 [26] では，背景音をフィンガープリントのように用いて屋内位置を推定する．

対して周波数スペクトル解析を行い，周波数と各成分の音圧を Wi-Fi フィンガープリントの様に扱うことによって，ユーザの屋内位置の推定を実現する．

また，加速度センサを用いた行動認識技術も盛んに研究されている [7, 8]．図 2.6 に示すように，J. R. Kwapisz らは，スマートフォンの 3 軸加速度センサのみを用いて，ユーザの歩行・走行・着席・階段の上り下り等の状態を認識するモバイルシステムを提案し，29 人のユーザによる実証実験を行い，その有効性を示した [8]．

更に，モバイルデバイスを用いてユーザの感情や精神状態を推定する研究も盛んになされている [29, 30, 31]．H. Lu らは，マイクを用いて取得したユーザ自身の声から，ユーザのストレスレベルを推定するシステムを設計・開発している [29]．本研究では，実際にユーザがストレスを感じる状況を用意し，実証実験を行っている．また，R. LiKamWa らは，スマートフォンの利用状況とユーザからの入力データを蓄積することによって，経験的にユーザの感情を推定・検知する MoodScope システムを設計・開発している [30]．

以上で述べたように，モバイルデバイスに搭載されたセンサを用いたユーザコンテキスト推定技術は著しく発展しているため，モバイルデバイスで取得できるセンサデータには多彩なプライバシー情報が含まれていると言える．

図 2.6: J. R. Kwapisz らの提案するモバイルシステムでは，3 軸加速度のみでユーザの行動認識を実現する [8]．

2.3 データマイニングのためのプライバシー保護技術

データを取得するセンサの普及やデータベース技術の発展に伴い，取得したデータを有効に活用しつつ，ユーザのプライバシー情報を保護する技術の研究が盛んになされている [32]．本節では，取得データを有効に活用するデータマイニングを実現するための，プライバシー保護技術について，暗号技術・k-匿名性・差分プライバシー・Perturbation 技術に分けて述べ，参加型センシングにおけるプライバシー情報保護を実現する適切な手法を探る．

2.3.1 暗号技術を用いたプライバシー保護手法

ユーザのプライバシー情報保護の最も基本的な技術として，暗号技術を用いた手法が挙げられる [10, 33]．暗号を用いた手法は，公開鍵暗号方式と共通鍵暗号方式の 2 つに大きく分けることができる．いずれの方式も，データ送信側でデータに対して暗号化処理を施し，データ受信側で復号化を行うものであり，データ通信時の盗聴からプライバシー情報を保護することができる．暗号技術を用いた場合，データ受信側は復号化処理によって，データ加工前の情報を手にすることができる．

E. De Cristofaro らは，参加型センシングにおけるプライバシー保護を実現するために，暗号技術を用いたプライバシー保護機構 PEPSI[34] を設計し，Nokia N900 を用いてシステムの実装を行った．プライバシー保護機構 PEPSI では，電子証明書申請者の審査・登録・確認を行う登録機関 (Registration Authority) を設けることによって，第三者へのユーザのプライバシー情報の漏洩を防ぐ．また，アプリケーション提供者として honest-but-curious モデル [36] を想定しており，センシングを行う一般ユーザ及びクエリ問い合わせ者の両者のプライバシーを保護することを目的としている．

また，G. Drosatos らは，都市環境を目的とした参加型センシングシステム NoiseTube[15] においてプライバシー情報の保護を行うために，暗号を用いるシステム [33] を提案した．本提案システムでは，実際の都市環境センシングシステム [15] への実装を行っているが，暗号化処理のための計算量が非常に大きいため，モバイルデバイス上で多くの消費電力を要すると指摘している．また，参加型センシングでは膨大な数のセンシング参加者全員がデータ送信側に該当するため，暗号鍵の管理も大きな問題として挙げられる [10]．

2.3.2 k-匿名性を用いたプライバシー保護手法

k-匿名性 (k-anonymity) は，公開されたデータにおいて同じデータを持つ人数が k 人存在することを保証することによって，他の $k-1$ 人との識別を不可能にする手法 [37, 38] である．

L. Sweeney[37] は，氏名と住所が削除された上で一般に公開された医療情報と，選挙投票者名簿を付き合わせる linked attack によって，個人のセンシティブ情報を容易に特定することができることを示し，公開情報におけるプライバシー情報の保護を保証するために，k-匿名性の概念を提案した．k-匿名化手法は，データベース上のデータに対するプライバシー保護手法であり，各フィールド情報の抽象度を調整することによって，悪意ある攻撃者が同定できるユーザを個人ではなく，k 人のグループとする手法である．k-匿名性を満たすデータベースとするためには，情報に対して，丸め処理や一般化処理を行う．

また，この k-匿名性の概念を拡張し，悪意ある攻撃者がデータベース上の情報について背景知識を持っていることを考慮した l-多様性という概念が A. Machanavajjhala らによって提案されている [39]．

k-匿名性はプライバシーレベルの設定を直感的に捉えることのできる手法であるため，プライバシー保護技術として様々なアプリケーションで注目されており，位置情報の秘匿化 [40] や無線センサネットワークで取得したデータのプライバシー保護 [41] などで活用されている．

2.3.3 差分プライバシー (Differential Privacy)

C. Dwork らは，図 2.9[42] に示すような対話型データベースにおけるプライバシー情報の保護を目的として，差分プライバシー (Differential Privacy)[43] の概念を提案した [44, 45]．差分プライバシーは，個々のレコードからだけでなく，これらから計算によって出力される統計的なデータからもプライバシー情報の侵害が起こるという出力プライバシーの問題に対する手法である [43]．

図 2.7: E. De Cristofaro らの提案する暗号技術を用いたプライバシー保護機構 PEPSI[34]．

図 2.8: L. Sweeney が指摘した linked attack[37]．

図 2.9: 差分プライバシーのシステム概要 [42].

差分プライバシーのモデルでは，データ利用者がデータデースに対して様々なクエリを問い合わせることによって，個々のレコードの持つプライバシー情報を取得する攻撃を想定する．この攻撃に対抗するために，Differential Privacy では，データ利用者からのクエリに対して，データベースの真の出力値にラプラスノイズを加えて出力を行う．これによって，データベースの出力値からのプライバシー情報の侵害を防ぐ．差分プライバシーはデータベースの持つ統計的な性質を利用した概念であり，あるレコードの有無によるデータベース出力値の差をある一定値以下に保証することで，個々のレコードが有するプライバシー情報の保護を担保する．

差分プライバシーを利用したデータベースクエリ言語として，Microsoft 社は Privacy Integrated Queries(PINQ)[46] がプロトタイプとして一般公開している．また，差分プライバシーの概念を他分野のプライバシー保護に拡張した研究も盛んに行われており，F. McSherry らは，ネットワーク解析での利活用に関する研究を行っている [47]．また，H. Haddadi らは一般ユーザのプライバシー情報の価値に関する研究などを，差分プライバシーの概念を利用して行っている [48].

2.3.4 Perturbation 技術を用いたプライバシー保護手法

Perturbation 技術 (データ摂動法) とは，取得データに対してノイズ付加を行うことによって，個々の情報が有するプライバシー情報を秘匿化する手法である．データベース上の情報に対するプライバシー保護データマイニングのための研究 [49] が盛んに行われている一方，既存のプライバシー保護処理の脆弱性に関する研究 [50, 51] も盛んになされている．

以下では，Perturbation 技術を用いたプライバシー保護手法 [52, 55] と Perturbation の基礎技術であり，S. L. Warner が 1965 年に提案した Randomized Response (ランダム回答法)[56, 57] とその拡張手法 Negative Surveys (否定調査法)[58, 59] について紹介する．

単一センサデータに対するプライバシー保護手法

R. K. Ganti らは，Perturbation 技術を参加型センシングに拡張し，プライバシー保護を実現する PoolView[52, 53] を提案・設計し，ユーザの体重変化のモニタリングや加速度センサ情報の推移

図 2.10: R. K. Ganti らの PoolView のシステム概要 [52].

を収集するシステムのプライバシー保護機構として実装を行っている．このシステム概要を図 2.10 に示す．本提案システムではセンシングに参加するユーザが計算機を利用することを前提としており，ユーザ側でプライバシー保護処理を行う．PoolView では，プライバシー保護のために API を配布し，情報を取得するユーザ側でノイズ付加を行い，情報を中央サーバに集約する．このノイズ付加されたデータに対し，統計処理を施すことによって，統計的なデータ分布を明らかにする．また N. Pham らは PoolView を拡張し，複数フィールド情報を同時に収集するシステムを設計・開発した [54]．しかしながら N. Pham らのシステムでは，各フィールド情報に対して個別にプライバシー保護処理を行うため，多次元データの持つ複数フィールド情報の相関性を保存することはできず，各フィールド情報の統計的なデータ分布のみを取得する．このため，位置情報と加速度センサ情報を同時に収集する実アプリケーションへの実装は行っていない．

PoolView では，ノイズの値を決める確率分布はアプリケーションを管理するサーバ側が計算し，センシングに参加するユーザと共有する．このため，センシングに参加するユーザの中に悪意ある攻撃者が存在した場合，取得データとノイズ付加後のデータを比較することによって，ノイズを決める確率分布が容易に推定できる [55]．F. Zhang らはこの脆弱性を解決するために，各ユーザごとにノイズの確率分布を異なるものにする手法 PESP[55] を提案している．

これらのプライバシー保護に関する研究では，取得データに対してユーザ側でプライバシー保護処理を行うため，他者にプライバシー情報を提供する必要が無い．一方で，N. Pham らの研究 [54] からも明らかな様に，既存手法では多次元データの持つ複数フィールド情報間の相関性を保存することができないため，実アプリケーションでは活用できていない．更に，PoolView のシステム及び PESP の手法では，ユーザ側には豊富な計算資源があることを想定しているため，プライバシー保護処理の計算量が莫大なものとなっているという特徴がある．

Randomized Response / ランダム回答法

Perturbation 技術の代表的な手法として，1965 年に S. L. Warner が提案した Randomized Response (ランダム回答法)[56] が挙げられる．Randomized Response は，他人に知られたくないセンシティブ情報を各回答者が明かすことなく，回答者全体の統計的な情報を得るための手法であり，薬物使用者数や脱税者数など，通常の社会調査では明らかにすることのできない情報を把握するために考案された．以下に Randomized Response の処理について記す．

Randomized Response では，社会調査の集計者と n 人の回答者を想定し，あるグループ A に属しているか否かを調査する．但し，あるグループ A に属しているという情報は，薬物使用・脱税など，他人に知られたくないセンシティブな情報であるとする．この時，調査対象である回答者は以下の 2 グループに分けることができる．

1. グループ A に属する．(グループ A の条件を満たす．)
2. グループ A に属さない．(グループ A の条件を満たさない．)

この時，真にグループ A に属する割合を π とし，属さない割合を $1-\pi$ とする．S. L. Warner の Randomized Response では，ユーザが乱数発生器を用いて，真の回答を集計者に回答するかを決定する．この真の回答を行う確率を p_0 とした時，偽の回答を $1-p_0$ の確率で行う．但し，$p_0 \neq 0.5$ を満たす．

この時，集計者が受け取る n の回答のうち，$\left\{n_1 = p_0\pi + (1-p_0)(1-\pi) \times n\right\}$ の数の回答は真と報告されたものである．また，回答者は集計者に，真の回答を行ったのか偽の回答を行ったのかは報告しない．このため，個々の回答者が真の回答を行ったかどうかは集計者はわからない．

この集計の後，以下の式 (2.1) に示す処理により，グループ A に属する人の割合 π_{true} を明らかにすることができる．この統計処理によって，回答者の個々のプライバシー情報を保護しながら，統計的なデータ分布を明らかにすることができる．

$$\pi_{true} = \frac{n_1/n - (1-p_0)}{2p_0 - 1} \tag{2.1}$$

S. L. Warner の Randomized Response[56] は，二者択一型選択肢に対する手法であったが，その後，単一回答法 (多肢選択型の質問) に拡張され，カテゴリデータを対象としたプライバシー保護手法 [57] として確立された．また，データマイニングに関する分野 [60, 61] だけでなく，Location Based Service を対象とした位置情報のプライバシー保護に関する研究 [62] など様々なアプリケーションで Randomized Response を基盤としたプライバシー保護に関する研究が行われている．

単一回答法に対する Randomized Response では，真の値を示すカテゴリデータを，一定確率で異なるカテゴリデータに遷移させて報告することによって，プライバシー情報の保護を実現する．選択肢 α で表されるカテゴリデータを扱う際，プライバシー保護処理は式 (2.2) に示す行列 M で表すことができる．行列 $\boldsymbol{M}$ は $\alpha \times \alpha$ の正方行列である．真の値であるオリジナルデータは，確率 p で真の値のまま集計者に報告される．また，オリジナルデータとは異なる $\alpha - 1$ 個のカテゴリに対して，それぞれ確率 $\frac{1-p}{\alpha-1}$ で遷移し，集計者に報告される．

$$\boldsymbol{M} = \begin{pmatrix} p & \frac{1-p}{\alpha-1} & \cdots & \frac{1-p}{\alpha-1} \\ \frac{1-p}{\alpha-1} & p & \frac{1-p}{\alpha-1} & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1-p}{\alpha-1} & \cdots & \cdots & p \end{pmatrix} \tag{2.2}$$

例えば，選択肢 a，b，c，d から 1 つを選ぶアンケート形式の調査で Randomized Response ($p = 0.1$) を用いる時，回答者は該当する選択肢を用いて報告する選択肢を選ぶ．この時，回答者が選択肢 a に当てはまる時には 0.1 の確率で a と回答する．また，それぞれ 0.3 の確率で，b，c，d を集計者に回答する．集計者側では，受け取った回答から回答者個々の真の回答を知ることができないが，再構築処理を行うことによって統計的なデータ分布を得ることができる．

集計者が受け取ったデータの統計的なデータ分布を，行列で $\hat{\boldsymbol{Y}}$ と表し，真の値を集約した時の統計的なデータ分布を行列で $\hat{\boldsymbol{A}}$ と表す時，再構築処理は式 (2.3) の様に表すことができる．但し，$\hat{\boldsymbol{Y}} = (Y_1, \cdots, Y_\alpha)^\tau$，$\hat{\boldsymbol{A}} = (A_1, \cdots, A_\alpha)^\tau$ を満たし，Y_i は，i 番目のカテゴリを示すデータの総数を意味するものとする．再構築処理の式 (2.3) から明らかなように，データ集計者は個々のデータの真の値を知ることなく，統計的なデータ分布を知ることができるため，ユーザのプライバシー保護を実現することができる．

$$\hat{\boldsymbol{A}} = \boldsymbol{M}^{-1}\hat{\boldsymbol{Y}} \tag{2.3}$$

Z. Huang らは，Randomized Response の手法をデータマイニングに用いる研究において，複数フィールドを持つ多次元データに対する Randomized Response に言及している [61]．各フィールド情報に対して，Randomized Response の手法を適用することは容易であるが，その場合には複数フィールド間の相関性が損なわれてしまう．つまり，Randomized Response を用いた場合，位置情報に環境情報を紐づける形や加速度センサ情報に医療情報を紐づける形でプライバシー保護処理を施すことはできない．このため，既存の Randomized Response では複数フィールド情報を持つ多次元データのプライバシー保護を適切に行うことはできない．

Negative Surveys / 否定調査法

Randomized Response の手法を拡張する形で，F. Esponda らは Negative Surveys (否定調査法) の手法を提案した [58, 63]．この Negative Surveys では，式 (2.2) における p の値を 0 として，プライバシー保護処理を行う．つまり，取得したカテゴリデータの示す真の値とは常に異なる値を集計者に報告し，集計者は統計的なデータ分布のみを再構築処理によって知ることができる．この様子を図 2.11 に示す．Negative Surveys では，回答者は他人に知られたくないセンシティブ情報を報告することがないため，より安心して情報を提供することができる [58]．また，Randomized Response と比較して，プライバシー保護のための処理が簡潔であり，計算量も少ない [63]．

例えば，選択肢 a，b，c，d から 1 つを選ぶアンケート形式の調査で Negative Surveys を用いる時，回答者は該当する選択肢とは異なる選択肢を等確率に選んで報告する．つまり回答者が選択肢 c に当てはまる時には a，b，d の中から等確率に 1 つの選択肢を選んで報告する．集計者側では，再構築処理を行うことによって統計的なデータ分布を得ることができる．集計者は，それぞれの回答者が該当する真の値を知ることはできないため，プライバシー情報の保護が約束される．選択肢の数を α，取得データ数を N とする時，集計者側における再構築処理は以下の式 (2.4) の様に定義される．この時，$A(i)$ は再構築した時のデータ分布における i 番目の選択肢のデータ総数，$Y(i)$ はプライバシー保護処理済データのデータ分布における i 番目の選択肢のデータ総数を示す．

$$\forall i \mid A(i) = N - (\alpha - 1) \cdot Y(i) \tag{2.4}$$

図 2.11: Negative Surveys を用いた時，回答者は他人に知られたくないセンシティブ情報を報告する必要がない [64].

Negative Surveys では，等確率で真の値とは異なるカテゴリを選択する．H. Xie らは，再構築処理後のデータ信頼性を向上させるために，プライバシー保護処理にガウス分布を用いる Gaussian Negative Surveys[65] を提案している．特に，地理空間情報のように隣り合うカテゴリに相関性がある場合は，ガウス分布を用いることによって，再構築処理後のデータ信頼性を向上させることができる [65].

また，J. Horey ら [66] は，固定的なセンサネットワークを対象として，Negative Surveys のデータ処理によって得たセンサデータを伝送するプロトコルを示した．Negative Surveys によるプライバシー保護で得たデータサイズは，センシングによって得たデータのサイズと等しくなるため，センサネットワークのトラフィック量を抑える形で各々のセンサノードに対してプライバシー保護処理を行うことができる．

M. M. Groat らの研究 [59, 67] では，この Negative Surveys を拡張して，複数フィールド情報を持つ多次元データに対するプライバシー保護手法 Multidimensional Negative Surveys を提案している．本手法では，取得した多次元データの各次元において，それぞれ真の値とは異なる値を等確率で選んでプライバシー保護処理済データとする．これによって，複数のフィールド情報の相関性を保持しながら，ユーザのプライバシー情報を保護することが可能となる．集計者側の再構築処理を式 (2.5)，(2.6) に示す．この再構築処理は，式 (2.4) を拡張したものである．

$$\forall \vec{x} \mid A(\vec{x}) = N + \sum_{k=1}^{D} (-1)^k \cdot \Gamma(\vec{x}, k) \tag{2.5}$$

この $\Gamma(\vec{x}, k)$ は，以下に示す処理である．

$$\Gamma(\vec{x}, k) = \sum_{\substack{d \in \\ B(\{1,2,3,\cdots,D\},k)}} \left(\left[\prod_{j \in d} (\alpha_j - 1) \right] \cdot \sum_{\substack{\vec{y} s.t. \\ y_i \in \vec{x} \\ \forall i \in d}} Y(\vec{y}) \right) \tag{2.6}$$

ここで，多次元データの値は $\vec{x}$ のように，ベクトル形式で表すこととし，再構築処理した統計的なデータには $\vec{x}$ の値を示すデータが $A(\vec{x})$ 個あるとする．また，D はデータの持つ次元の数を表し，$B(\{1, \cdots, D\}, k)$ は $\{1, \cdots, D\}$ の中から k 個の数を選ぶ時の組み合わせを表すものとする．この Multidimensional Negative Surveys をモバイルセンシングに実装し，環境騒音地図として構築したものを図 2.12 に示す．図 2.12 では，地図に対して格子状に騒音レベルを重ね合わせて表示しており，騒音レベルは赤と緑で表している．騒音レベルの高い地域は明るい赤で，低い地域は緑で示されている．また，黄で示される線は道路を表す．

この図 2.12 では，センシングによって取得したオリジナルデータに基づく環境騒音地図及び，

Multidimensional Negative Surveys の手法を用いた時の環境騒音地図が示されているが，統計処理を用いているため，データ間に相違のあるエリアが数多く存在する．

2.4 本章のまとめ

本章では，参加型センシングのシステム概要について述べた後に，参加型センシングを実現するために非常に重要な技術であるプライバシー情報保護手法を紹介した．本節では，これらをまとめると共に，既存手法の問題点について明らかにし，今後の課題を示す．

まず，第 2.2.1 項では，都市環境調査・交通モニタリング・公衆衛生などを想定した，既存の参加型センシングシステムを紹介し，参加型センシングの目的が統計的なデータ分布の取得であることを明らかにした．これらのシステムでは，モバイルデバイスに搭載されたセンサをセンシングに用いるが，一般のモバイルデバイスが持つ電源・計算資源は限りあるものである．また，複数センサ情報を同時に収集し，互いの情報の相関を明らかにすることが目的である．

第 2.2.2 項では，一般ユーザの持つモバイルデバイスに搭載されたセンサからのデータを用いたユーザコンテキスト推定技術について述べ，モバイルデバイスで取得したセンサデータがユーザのプライバシー情報を多く含んでいることを明らかにした．モバイルデバイスで取得できるセンサデータは位置情報以外にも，様々なユーザコンテキストに関するプライバシー情報を含むため，参加型センシング実現のためにはプライバシー保護技術の開発は非常に重要な課題である．但し，参加型センシングの特徴を考慮すると，多次元データの持つ複数フィールド間の相関性を保存することのできるプライバシー保護手法を活用する必要がある．また，モバイルデバイスにおけるデータ処理は，計算量の低い設計とする必要がある．

次に，第 2.3.1 項から第 2.3.4 項にかけて，既存のプライバシー保護技術を紹介し，それぞれの特徴を述べた．

第 2.3.1 項では暗号技術について紹介した．参加型センシングに暗号技術を用いる場合，一般ユーザの持つモバイルデバイス上で常に暗号化処理を行う必要があるため，計算資源に限りのあるモバイルデバイスを用いる参加型センシングには適していないと言える．また，暗号技術を用いた

図 2.12: 取得したオリジナルデータに基づく環境騒音地図及び MNS の手法を用いた時の環境騒音地図 [67].

場合，悪意ある攻撃者からの盗聴に対するプライバシー保護を実現するが，一方で復号化後にはプライバシー情報が明らかになってしまう．このため，アプリケーション提供者がユーザのプライバシー情報を容易に取得できるという問題点がある．

また，第 2.3.2 項や第 2.3.3 項では，k-匿名性や差分プライバシーの手法について紹介したが，これらは取得したデータをデータベース上に集約した後に行うプライバシー保護手法である．これらの手法を参加型センシングで用いる場合，一般ユーザは無条件にプライバシー情報をアプリケーション提供者に公開する必要があるため，多くのユーザの貢献を前提とする参加型センシングのプライバシー保護手法としては適切ではない．

第 2.3.4 項では，取得データに対してユーザ側でノイズを加える Perturbation 技術を用いたプライバシー保護システム PoolView 及び PESP と基礎技術である Randomized Response，Negative Surveys について述べた．参加型センシングでは，第 2.2.1 項で挙げたように，複数フィールド情報を有する多次元データを取得する必要があるが，既存のプライバシー保護システムでは，この多次元データを適切に扱うことができないのが現状である．また，Perturbation の基礎技術である Randomized Response でも同様に，多次元データの相関性を維持しながらデータを収集することはできない．一方で，Negative Surveys では多次元データの持つ複数フィールド間の相関を維持しながらプライバシー保護を実現する手法が既存研究で示されている．

これらをふまえて，本論文では Perturbation 技術を拡張して参加型センシングにおけるプライバシー保護を実現する．このため，まず第 3 章では，既存の Perturbation 技術を参加型センシングのプライバシー保護に用いる事前実験を行った上で，技術的課題を示す．第 4 章，第 5 章，第 6 章，第 7 章では，第 3 章で述べた技術的課題に対し，それぞれ想定する参加型センシングのアプリケーションの概要と目的を明らかにした上で，適切なプライバシー保護手法を提案する．

第3章　モバイルデバイスを用いた参加型センシングとプライバシー情報

第 2 章では，モバイルデバイスのセンシング技術について述べ，参加型センシングの実現のためにはプライバシー保護技術の開発が非常に重要であることを明らかにした．

そこで本章では，参加型センシングにおけるプライバシー情報の侵害の問題についてアプリケーション提供者とユーザの立場からそれぞれ考察し，本研究が対象とする参加型センシングのプライバシー保護を実現する具体的な要件について明らかにする．また，既存の Perturbation 技術を Android SDK を用いて実装し，事前実験を行うことによって，参加型センシングのプライバシー保護を実現する技術のうち未解決な問題を挙げ，本研究で取り組むべき課題を明確にする．

3.1　参加型センシングにおけるプライバシー情報の侵害

参加型センシングの目的は，多数のユーザから情報を収集し，統計的なデータ分布を明らかにすることである．一方で，モバイルデバイスで取得したセンサデータに基づく行動認識技術・位置情報推定技術が著しい発展を遂げていることから，GPS 受信機で取得した位置情報のみならず，他のセンサから得た情報もユーザのプライバシー情報を含むことが明らかである．ここで，事前実験として，無線型二酸化炭素測定器 (C2D-W02TR) を一般ユーザが一日持ち歩いたときの取得データの様子を図 3.1 に示す．二酸化炭素センサから取得したデータによって，測定者の周囲の人の有無や密度が推測できる他，測定者が屋内にいるか屋外にいるかを計測することができる．図 3.1 に示す二酸化炭素センサのデータから明らかなように，センサの示す値だけでなく，データの取得日時や値の推移も大きな情報を持つため，悪意ある攻撃者が背景知識を併用してプライバシー情報を特定する危険性もあると言える．

また，参加型センシングでは多様なセンサからデータを取得し，多次元データとしてセンサデータを扱う．このため，1 種類のセンサが取得するユーザのプライバシー情報に関する情報量が小さくとも，複数のセンサ情報を取得する際にはプライバシー情報に関する情報量が大きくなることに注意しなければならない．また，各々のセンサデータがユーザのプライバシー情報に直結せずとも，あるセンサデータの値の組み合わせが特異となり，ユーザのプライバシー情報の特定に結びつく可能性も考えられる．このため，位置情報などのプライバシーに直結する情報以外のセンサデータに対しても，プライバシー保護処理を施すことが望ましい．

以下では，参加型センシングにおけるプライバシー情報の取り扱いについて，アプリケーション提供者とユーザの立場から考察をすることによって，プライバシー保護を実現するための具体的な要件を明らかにする．

アプリケーション提供者から見たプライバシーの問題

都市環境調査や公衆衛生，道路交通モニタリングなどを目的に参加型センシングを実社会で活用する場合，その管理と運用を担う機関が必要不可欠であり，民間企業や国立機関がこのアプリケー

ション提供者に該当する．参加型センシングでは，ユーザ全体の統計的なデータ分布を取得することを目的としているため，ユーザ個々人のプライバシー情報はアプリケーション提供者にとって不必要な情報であることが明らかである．このため，アプリケーション提供者が本来不要であるプライバシー情報を保持・蓄積することは，ユーザのプライバシー情報の侵害・漏洩という余計な危険性のみを背負うことを意味する．

プライバシー情報を取得することに対して，一般ユーザに対価を支払う研究 [70] 等もなされているが，莫大な数のユーザをセンシング参加者とする参加型センシングを実現する場合には，この対価も莫大なものとなってしまう．また，対価を支払う形態を取ると，参加型センシングに参加するユーザと取得することのできるデータに偏りが生まれることも考えられるため，プライバシー情報の取得に対して対価を支払うことは望ましくない．

これらをふまえると，より多くのアプリケーション提供者が参加型センシングシステムを提供・活用する土壌を構築するためには，センサデータに対してユーザ側でプライバシー保護処理を施し，プライバシー情報侵害という危険を可能な限り減らすことが望ましい．

また，本研究ではアプリケーション提供者及びサーバ管理者として，V. Božović らが示した honest-but-curious モデル [36] を想定する．V. Božović らの honest-but-curious モデル [36] では，アプリケーション提供者及びサーバ管理者は，定められたプロトコルに従ってデータ処理を行う．しかしながら同時に，プライバシー情報やセンシティブ情報が取得できると判断した場合は，興味に従ってユーザのプライバシー情報の取得を試みる．

アプリケーション提供者が安全に参加型センシングを運用するためには，データ伝送中の盗聴によるプライバシー情報の漏洩・流出を防ぐことはもちろん，アプリケーション提供者及びサーバ管理者がユーザ個々のプライバシー情報を保持しない環境を構築することが非常に重要である [3]．

図 3.1: 二酸化炭素濃度センサの一日の挙動 [12, 69]．

ユーザから見たプライバシーの問題

従来の固定型センサを用いたセンサネットワークとは異なり，参加型センシングでは一般ユーザの持ち歩くモバイルデバイスをセンサとして活用する．参加型センシングにおけるセンサデバイスの数は，参加するユーザの数に比例するため，より質の高い統計的な情報を取得・構築するためには，より多くのユーザが参加型センシングに参加することのできる土壌を実現することが重要である．

しかし第 2 章で指摘したように，モバイルデバイスからのセンサデータには，ユーザの様々なプライバシー情報が含まれるため，プライバシー情報の侵害・漏洩が非常に大きな問題となる．参加型センシングにおいて，プライバシー情報の侵害・漏洩という危険性が残る場合，ユーザをセンシングに巻き込むことは難しい．

これらをふまえると，一般ユーザのプライバシー情報を他者に提供することなくセンシングに参加・貢献できるプライバシー保護技術を開発することが非常に重要である．

以上で述べたように，参加型センシングにおけるプライバシー情報の取り扱いの問題は，アプリケーション提供者及びユーザの立場から深刻な問題である．一般ユーザのプライバシー情報の侵害・漏洩の危険性がある場合，実社会で参加型センシングを用いることは難しい．このため，多くの企業や組織が積極的に参加型センシングを活用・提供し，同時に多くの一般ユーザが参加型センシングに参加・貢献するためには，ユーザのプライバシー情報をアプリケーション提供者が取得・保持する必要のないプライバシー保護技術を開発することが重要となる．

これらをふまえて，本研究では一般ユーザがセンシングに用いるモバイルデバイス上で，プライバシー情報を処理する手法を提案する．本提案手法では，プライバシー保護処理のために Perturbation 技術を用いることによって，モバイルデバイス上の計算量を抑えた設計とする．

3.2 事前実験と技術的な課題

本節では，既存の Perturbation 技術をモバイルデバイスを対象に実装し，参加型センシングでの利用を想定した事前実験を行うことによって，モバイルデバイス上で行うプライバシー保護処理を実現するための技術的課題を明らかにする．但し，第 3.1 節で述べたように，モバイルデバイスの持つ計算資源・電源には限りがあるため，モバイルデバイス上のプライバシー保護処理は計算量を抑えた設計とする．

まず，既存手法である Multidimentional Negative Surveys (MNS)[59, 67] を Android OS 搭載のスマートフォンに実装し，データ信頼性に関する評価実験を行った．ここでは，3×6 と 9×16 の 2 次元データ，$3 \times 3 \times 2$ と $9 \times 4 \times 4$ の 3 次元データ，$3 \times 3 \times 2 \times 2 \times 2 \times 2$ の 6 次元データを擬似的にスマートフォン上で生成し，それぞれ Multidimentional Negative Surveys によるプライバシー保護処理を行う．但し，“3×6 の 2 次元データ” は，選択肢 3 のフィールド情報と選択肢 6 のフィールド情報を有する多次元データを意味するものとする．また，擬似的に生成するデータの分布は一様分布をとるものとする．評価指標としては，式 (3.1) の様なカルバックライブラー情報量 D_{KL}[89] を用いた．D_{KL} は，真の離散確率分布 X と他の離散確率分布 A に関する一致性を表し，この値が 0 に近い値となるほど X と A が近い確率分布をとる [89]．

$$D_{KL}(X||A) = \sum_i X(i) \log \frac{X(i)}{A(i)} \tag{3.1}$$

評価では，参加型センシングで取得するオリジナルデータの分布を X とし，伝送の後に中央サーバで再構築したデータ分布を A とする．10 回の評価実験を行った時の，D_{KL} の平均の値を図 3.2 に示す．この時，データ総数 N は $10,000$，$10,000$，$20,000$，$30,000$，$40,000$，$100,000$ とする．この結果から明らかなように，それぞれの次元の持つ選択肢の数が大きな時，つまり連続値により

図 3.2: 既存手法 MNS[59, 67] を用いた時のプライバシー保護処理の様子．評価指標としては，式 (3.1) のカルバックライブラー情報量 D_{KL} を用いた．

図 3.3: センサデータとして $\vec{x} = <x_1 = 2, x_2 = 1>$ を取得した時の様子．既存手法 MNS[59, 67] を用いた時，モバイルデバイス上で白セルの中から等確率で 1 つのセルを選択し，サーバへと報告する．このため α_2 のフィールド情報に関して，オリジナルデータの推定を容易に行うことができる．

近い形でデータ取得を行った時，オリジナルデータと再構築データの統計的なデータ分布は大きく乖離し，データの信頼性が大きく低下する．これは，プライバシー保護処理に用いる疑似乱数が，真の乱数とは異なり，デバイス上の計算によって作られたものであることが原因であると考えられる．このため，既存手法の Multidimentional Negative Surveys では，データカテゴリ数が多い時，つまり各次元のデータ区分を細かく扱う時，正確にデータの取得を行うことはできないと言える．都市環境調査や交通モニタリングなどの調査を行う場合，環境騒音レベルや位置情報などを詳細なデータとして取得する必要があるため，データ区分の細かい多次元データを取得する際のプライバシー保護は非常に重要な技術課題であると言える．

また，既存手法 Negative Surveys 及び Multidimensional Negative Surveys では，二者択一型選択肢を有する多次元データを扱うことができないという技術的な課題がある．これは，Negative Surveys[58] によるプライバシー保護処理において，オリジナルデータとプライバシー保護処理済データのカテゴリ値が互いに相反する関係になり，常に異なるカテゴリ値をとるためである．例として，4×2 の 2 次元データ ($\alpha_1 = 4$，$\alpha_2 = 2$) に対する Multidimensional Negative Surveys の処理の様子を図 3.3 に示す．図 3.3 では，センサデータとして $\vec{x} = <x_1 = 2, x_2 = 1>$ を取得した時の様子を示す．既存の Perturbation 技術である Multidimensional Negative Surveys によるプライバシー保護処理済データは，図 3.3 における白セルの中から 1 つが等確率で選ばれる．しかしながら，α_2 の次元のデータに関しては，プライバシー保護処理済データから容易にオリジナルデータを推定することができる．つまり，センサデータとして $\vec{x} = <x_1 = 2, x_2 = 1>$ を取得した時，伝送されるプライバシー保護処理済データは常に $y_2 = 2$ を満たす．このため，盗聴者やアプリケーション提供者がプライバシー保護処理済データを取得できた場合，確実にユーザのオリジナルデータを取得できてしまう．つまり，適切なプライバシー保護処理を施すことができない．モバイルデバイスを用いた参加型センシングでは，イベント検知や閾値判定に関する情報を扱う場合に二者択一型の質問を行う機会が頻繁にあり，また性別などのプロファイル情報を収集する際にも二者択一型選択肢は多く用いられる．このため，二者択一型選択肢を含む多次元データのためのプライバシー保護手法の確立は重要な技術課題である．

更なる課題として，既存の Perturbation 技術は，取得データに対して逐一処理を施すため，長期間にわたる盗聴による蓄積攻撃からのプライバシー情報の保護を実現できていない．特に，ヘル

スケアや医療調査，公衆衛生の分野で参加型センシングを活用する際，収集するデータは時間依存性の低い，準静的な情報 (quasi-static data) となることが多く，常にデータの値が動くわけではない [4]．このため，長期間のデータを取得し続ける蓄積攻撃からのプライバシー情報の侵害は，医療分野で用いる参加型センシングの実現にとっては大きな脅威となる．

最後に，既存の Perturbation 技術では，公開情報や背景知識を用いた攻撃を考慮していないという課題が挙げられる．図 3.4 に示すように，実社会で悪意ある攻撃者が他人のプライバシー情報の取得を試みる時，出来る限りの手を尽くすと考えられるため，盗聴と並行して，政府機関や民間企業からの公開情報や背景知識，社会の共有知を活用すると想定しなければならない．既存の Perturbation 技術である Negative Surveys[58] を用いて，a，b，c の 3 つの選択肢のアンケートを行った時の様子を図 3.5 に示す．a，b，c に該当するデータの数が一様であった時，公開情報・背景知識はほとんど攻撃者の手助けをすることは無いが，データ分布に偏りがある場合，盗聴によって取得したプライバシー保護済データからのプライバシー情報特定を助けることになる．例として，プライバシー保護処理済データが b の値を示している場合，高い確率でオリジナルデータの値は a である．このように，統計的なデータ分布に関する大まかな傾向や背景が社会で共有されている場合，これらの情報が悪意ある攻撃者を手助けすることが明らかである．また，参加型センシングでは一般ユーザのモバイルデバイスを用いてデータを取得し，一般ユーザがデータを活用できる形で提供することを前提としている．このため，参加型センシングで取得した情報も，公開情報・社会の共有知の質を高めるために用いられる．これらのことを考慮すると，参加型センシングで取得したデータを一般社会で活用する際には，公開情報・背景知識を用いたプライバシー情報特定攻撃に備える必要がある．

これらを総括すると，アプリケーション提供者・ユーザの双方が安心して参加型センシングを活用するためには，

- データ区分の細かい多次元データのためのプライバシー保護
- 二者択一型選択肢を含む多次元データのためのプライバシー保護
- 長期間にわたる蓄積攻撃に対するプライバシー保護
- 公開情報や背景知識を利用した攻撃に対するプライバシー保護

の 4 点を実現する必要がある．本論文では，これら 4 つの技術的課題を解決する手法をそれぞれ示すことによって，より多くのアプリケーション提供者及びユーザが参加型センシングを活用する社会環境を構築することを目指す．

まず，第 4 章では，都市環境調査のための参加型センシングを実現するプライバシー保護手法 Selective Negative Surveys を提案する．本提案手法では，既存手法の Multidimensional Negative Surveys を拡張し，到達先パラメータを設定したプライバシー保護処理を行うことによって，データ区分の細かい多次元データのプライバシー保護処理を実現する．本手法の実現によって，データ信頼性を確保しながら，センサデータを連続値に近い形で処理することが可能となる．

次に，第 5 章では，既存の Perturbation 技術である Randomized Response と Negative Surveys を組み合わせることによって，二者択一型選択肢を含む多次元データのプライバシー保護を実現する MNS-RRT アルゴリズムを提案する．本提案手法の実現によって，Yes/No での回答が必要なイベント検知のための参加型センシングや，二者択一型選択肢となるプロファイル情報を収集する参加型センシングのプライバシー情報を保護することが可能となる．

第 6 章では，医療調査分野で用いる参加型センシングを想定し，準静的な情報 (quasi-static data) を扱う Duplicated Perturbation を提案することによって，長期間の盗聴による蓄積攻撃からのプライバシー保護を実現する．本章では，Duplicated Perturbation の処理を一次元データ・多次元データに分けて提案する．更に，Duplicated Perturbation ではオリジナルデータとプライバシー保護処理済データの関係に注目することによって，サーバ上でのデータ再構築処理のための計算量を削減し，リアルタイム処理を目指す.

最後に，第 7 章では，参加型センシングで取得したデータを一般社会に公開することを想定し，プライバシー保護機構 Data-centric Perturbation Infrastructure を提案し，その処理に用いるプライバシー保護手法 Empirical Negative Surveys を提案する．本機構 Data-centric Perturbation Infrastructure では，参加型センシングで取得したデータそのものをプライバシー保護処理に用いることで，より強固なプライバシー保護を実現し，データ公開が悪意ある攻撃者を手助けすることの無い参加型センシングシステムを実現する.

本研究を通して，アプリケーション提供者及び一般ユーザがプライバシー情報の侵害を懸念すること無く，参加型センシングとモバイルデバイスを社会インフラとして活用することのできる社会の実現を目指す.

図 3.4: 攻撃者が公開情報や背景知識を利用すると，より容易にプライバシー情報を取得できる [64].

図 3.5: オリジナルデータが一様分布である時 (左図) と偏りのある分布である時 (右) の Negative Surveys[58] の処理の様子.

第4章　参加型都市環境センシングとプライバシー保護

第 3 章では，参加型センシングのプライバシー情報の取り扱いの問題について述べ，既存の Perturbation 技術では連続的な多次元データを扱うことができないことを明らかにした．そこで本章では，連続的な多次元データを収集するアプリケーションとして，まず住生活環境調査のための参加型都市環境センシングシステムを実装し，実証実験を行うことによってセンサデータの特徴を明らかにする．本システムでは，一般ユーザの持つモバイルデバイスをセンサとして用い，環境騒音レベル・日照度・二酸化炭素濃度・臭気レベルを同時に収集し，都市地域における住生活環境の調査を行う [71]．また，参加型都市環境センシングシステムで取得する多次元データの特徴をふまえた上で，本システムの実現のために重要な課題となるプライバシー情報の保護のための手法 Selective Negative Surveys を提案する．本提案手法 Selective Negative Surveys では，到達先パラメータを導入することによって，データ区分の細かいセンサデータの取得を実現する．

4.1　参加型都市環境センシング

本節では，参加型都市環境センシングの概要を示した後に，実際のモバイルデバイスを用いたセンシングシステムの実装の様子を示す．更に，本システムを用いて行った東京都 世田谷区全域における歩行センシング実験で取得したデータを示し，本センシングシステムの実用性に関しての考察及びセンシング実験に参加したユーザ分析について記す．また，取得したデータを解析することによって，都市の住生活環境のみならず，動物・昆虫などの自然環境から発せられる情報を有効に活用するモバイルセンシングシステムの実現が示唆された．この自然環境情報を活用した生態調査システムの概要と，自然環境情報による位置情報補正技術について示す．なお，参加型センシングは人口密度の高い都市圏で活用されることを想定するため，本章では都市圏に生息するハシブトガラスとセミに注目し，システムに関する考察を記す．

4.1.1　参加型都市環境センシングの概要

住居の密集する都市地域においては，騒音レベル・日照度・大気汚染度・放射線量などの環境情報は，住生活環境の観点から非常に重要な指標である [15]．これらの情報は従来，JIS 規格 [72] に従って，選別された時間・地点で，特定の専用機器を用いて取得し，環境情報地図として公開されていた．このため，データの収集には多大な労力を要し，また経済的コストも莫大であった．また環境情報は時々刻々と変化をするため，本来はより広域の測定点において様々な時刻に観測する必要がある．

ここで，既存の環境情報地図の例として，国立環境研究所が公開している全国自動車交通騒音マップ [73] を図 4.1 及び図 4.2 に示す．図 4.1 及び図 4.2 は交通道路における騒音値を地図上に示したものであるが，調査の粒度が非常に粗いことがわかる．また，東京都 世田谷区を表す図 4.1 の様な都心部では，位置によって環境騒音値が異なるが，一方で熊本県 熊本市街地を表す図 4.2 からは，位置による環境騒音値の違いがほとんど分からない．

固定センサ群を用いたセンサネットワークの活用も考えられるが，人が多く居住する地域では固定センサ設置のための制約が厳しく，経済的コストも非常に高くなるため，現実的ではない．これらのことを考慮すると，データを空間的・時間的に詳細に収集し，かつセンサ設置コストがほとんど必要ない参加型センシングは，都市圏における住生活環境調査に最適なアプリケーションであると言える [74, 75].

参加型都市環境センシングでは，図 1.2 に示したように，一般ユーザの持ち歩くモバイルデバイスをセンサとして，環境情報と位置情報を収集し，都市の環境情報地図を構築することを目指す．都市環境調査に参加型センシングを活用することによって，より実態に即した信頼性の高い環境情報地図を構築することが期待されたおり，都市における不動産価値の再評価 [74] や住民の精神的ストレスの測定 [76, 77] への利用が期待されている．

参加型都市環境センシングで用いるスマートフォンなどのモバイルデバイスには，GPS 受信機をはじめとして，加速度センサ，マイク，照度センサ，大気センサが搭載されている．参加型都市環境センシングでは，これらのセンサ群から同時にセンサデータを収集することによって，複数のフィールド情報を持つ多次元データの統計を構築する．この際，第 3 章で言及したように，フィールド情報の相関が重要な情報であり，位置情報と騒音情報の関係や位置情報と日照度の関係を収集することが，システムの目的となる．

4.1.2　システムの実装と歩行センシング実験

スマートフォンを用いた参加型都市環境センシングシステムを構築するために，Android SDK の API を用いたアプリケーションを開発した．このセンシングアプリケーションのインターフェースを図 4.3 に示す．本アプリケーションを用いることによって，位置情報と共に日照度，騒音レベル，高度，3 軸加速度情報を同時に取得することができる．また，外部センサを用いて二酸化炭素濃度し，ユーザからの入力によって臭気レベルも同時に取得する．取得するセンサデータの中でも

図 4.1: 国立環境研究所が公開している全国自動車交通騒音マップ (東京都 世田谷区)[73].

特に，マイク音声情報はビット数 16[bits]，サンプリング周波数 44,100[Hz] で取得し，音声情報の Wave データから等価騒音レベルの導出には，ISO[78] 標準に基づいた処理を行う NoiseTube の手法 [75] を用いた．センサデータは一定時間ごとに CSV ファイルに記録し，また，連続的にマイク音声を録音する．

次に，この実装したセンシングシステムを用いて，総面積約 60[km^2] の東京都 世田谷区内全域の環境情報を取得することにより，参加型都市環境センシングで取得するデータの特徴を明らかにする．

ここで，東京都 世田谷区における，既存の JIS 規格 [72] に基づいた騒音情報の取得地点と，地価サンプル地点を図 4.4 に示す．不動産の市場価値は，この地価サンプル地点に基づいて決定されている．しかしながら，図 4.4 からも明らかなように，地価サンプル地点の数に対して，既存の騒音情報の取得地点が少ないことが明らかである．

本歩行センシング実験では，この地価のサンプル地点である 1,713 地点に加えて，工場の所在する 869 地点の計 2,582 地点の計測を行うことによって，より実世界に近い，信頼性の高い環境情報地図の構築を目指す．

歩行センシング実験は，計 4 日間行うが，世田谷区内全域を効率良く調査するために，本実験を行う前に，センシング地点 2,582 をクラスタリングによって分割を行った．各ユーザは，このセンシング実験でクラスタを選択し，センシング実験を行う．本研究では，このクラスタリングのために k-means 法 [79] を用いて，センシング地域を 60 の小地域に分割した．この k-means 法の具体的な手順は以下の通りである．

まず，それぞれのノード (センシング地点) が，$k(=60)$ 個のクラスタのうち，ランダムに 1 つのクラスタに属しているものとする．次に，それぞれのクラスタごとの重心 V_j (j = 1, 2, ⋯, k) を，クラスタに属する全てのノードの座標から導出する．その後，全てのノードの所属クラスタを，重心がノードに最も近いクラスタへ変更する．この重心の導出，クラスタの変更を繰り返し，

図 4.2: 国立環境研究所が公開している全国自動車交通騒音マップ (熊本県 熊本市)[73].

図 4.3: 都市環境センシングアプリケーションのインターフェース.

図 4.4: 東京都 世田谷区における既存の環境情報取得地点 (左図) と地価サンプル地点 (右図).

全てのクラスタの重心が変化しなくなった時点でクラスタリングを終了する．最終的なクラスタリングが，センシングのための小地域に対応する．この k-means 法を用いて，世田谷区を 60 の小地域に分割したものを図 4.5 に示す．k-means 法を用いてクラスタリングを行った場合，各クラスタに属するポイント数は，一様にはならない．この 60 のエリアごとのポイント数を図 4.17 に示す．本センシング実験では，このエリアごとのポイント数を各ユーザに提示し，自由にセンシングエリアを選択する権利を与えた．

本センシング実験では，世田谷区内全域の騒音情報を取得するために，計 4 日間，のべ 40 人のユーザによるセンシング実験を行った．本実験では，歩行時のセンシングの高さとスマートフォンの状況がほぼ一定となるよう，図 4.7 に示すようにアームバンドを用いてスマートフォンを固定し

図 4.5: k-means 法によって 60 に分割したセンシング領域.

た．また，ユーザの歩行時には連続的にデータ取得を行うが，同時に，取得する情報の信頼性を確保するために，測定地点の 2,582 箇所では 10 秒間の停止をして測定を行う．測定に用いるスマートフォンとしては，Android OS を搭載した複数の機種 (Galaxy S3，Galaxy Nexus，Nexus One) を用いた．

4.1.3 データ校正

従来のセンサネットワークと，モバイルデバイスを用いた参加型センシングの大きな違いとして，センサ精度のばらつきが挙げられる．センサネットワークでは，高価かつ高精度の固定型センサによって情報を取得する．一方で，参加型センシングでは，幅広いユーザが様々なデバイスを用いてセンシングに参加する．また，固定型センサでは純粋なセンサデータのみを取得するが，その一方でモバイルデバイスに搭載されたセンサにはユーザの動きや行動からの雑音も含まれる．この

図 4.6: 各エリアのポイント数.

図 4.7: 歩行センシングの様子.

ため，端末間のセンサの校正及びユーザ間のセンサの校正は，信頼性の高いデータ取得のためには非常に重要である．そこで本研究では，特にモバイルデバイスに搭載されたマイクと環境騒音情報を対象とした校正手法を示す.

端末間校正

本歩行センシング実験では，スマートフォンに搭載されたマイクを用いて音声情報を取得し，等価騒音レベルを導出する．このスマートフォンは，製造工程の関係上，同一機種の端末であっても，異なる種類のセンサが搭載されている可能性がある [80]．このため，機種間の校正だけでなく，端末ごとの校正を行う必要がある.

適切なマイク校正を行うために，図 4.8 のような 4π 無響室 (有効容積 3.8m×4.8m×3.8m，浮構造，内壁 80cm 厚吸音楔) において，校正実験を行った．この実験の様子を図 4.9 に示す．

本校正実験では，音源からピンクノイズを発し続け，音圧の大きさを変化させる．このピンクノイズの音を，JIS 規格を満たすリオン社製の騒音計 NL-32 と，Android 端末 (Galaxy S3，Galaxy Nexus，Nexus One) によって測定を行い，音源の音圧レベルと測定機器の示す等価騒音レベルの関係を導く．この音圧レベルと等価騒音レベルの関係を図 4.10 に示す．図 4.10 における音源の値は騒音計 NL-32 で計測・表示した値を示す．

歩行センシング実験で環境騒音を測定した後に，端末ごとに観測した等価騒音レベルを図 4.10 の関係グラフに沿って校正を行うことによって，より信頼性の高い騒音情報を取得することが可能となる．

図 4.8: 4π 無響室.

図 4.9: 端末間校正実験の様子.

図 4.10: 音圧レベルと各端末の等価騒音レベルの関係.

ユーザ間校正

参加型都市環境センシングでは，ユーザが都市空間を動き回る中で環境情報を取得するため，ユーザ自身の動きがセンサデータに影響を与える．特に，マイクによって取得した音声情報には，図 4.11 に示す様にユーザ自身の足音や服の擦れる音など，様々な種類の雑音が含まれる．マイクから取得した音声情報に対して，周波数スペクトル解析をリアルタイムで行い，音の発生源を分類する研究等 [26] も幅広く行われているが，計算資源と電源の限られたモバイルデバイス上で，常時フーリエ変換を行うことは非常に難しい．また，周波数フィルタリングによる雑音除去は音声情報処理の分野では盛んに研究されているが [26]，環境騒音そのものが様々な周波数を含んでいるため，フィルタリングによる除去は，データ信頼性の観点から望ましくない．また，ユーザ固有の雑音は今回のセンシングに参加した 40 人のユーザ全てで観測され，ユーザごとに異なる大きさの雑音が観測された．このユーザ固有の雑音からの影響を除去するために，参加型センシングでは複数センサが同じエリアをセンシングするという特性を利用して，相対的雑音発生レベルをユーザごとに導出し，ユーザ間の雑音発生レベルを一様化するユーザ間校正手法を提案する [82].

この校正手法の様子を図 4.12 に示す．本手法では，同じエリアをセンシングしたユーザ間で取得した環境騒音情報の差異を基にして，相対的な雑音発生レベルを導出する．この相対雑音発生レベルを用いて，ユーザの取得データ全体の校正を行う．ここで，あるユーザ A とユーザ B が同じエリアをセンシングし，等価騒音レベルの平均値がユーザ A の方が大きかった時の処理の様子を式 (4.1), (4.2) に示す．本校正手法では，等価騒音レベルの平均値が高いユーザのみデータ校正を行う．

$$\forall k \mid N_k'^A = N_k^A - |N_{average}^A - N_{average}^B| \tag{4.1}$$

図 4.11: Audacity[81] による取得データのボリュームレベルの推移の様子.

図 4.12: 相対的な雑音発生レベルの導出によるユーザ間校正手法の様子.

$$\forall k \mid N_k'^B = N_k^B \tag{4.2}$$

この時，N_k^A はユーザ A が，ある地点 k で実際に取得した騒音値を示す．また，$N_{average}^A$ と $N_{average}^B$ は，それぞれユーザ A とユーザ B があるエリアにおいて取得した騒音値の平均値を示す．また，$N_k'^A$ は本校正手法を施した後の騒音値を示す．

更に，JIS 規格に沿った従来の測定方法 [72] で取得された特定の騒音情報を基準として用いることによって，他の測定地域においてもユーザの歩行時の雑音を含まない，より信頼性の高い騒音情報を取得することが可能となる．

4.1.4　センサデータの可視化と考察

本歩行センシング実験では，ユーザの歩行足跡に沿って，スマートフォンのセンサからの環境情報を連続的に取得した．このセンシングした地点の一覧を地図上に表示したものを図 4.13 に示す．また，あるユーザに注目し，歩行足跡に沿った等価騒音レベルを地図上に表示したものを図に示す．ここで，等価騒音レベルの高い地点のデータは赤で，低い位置のデータは青で記す．交通量の多い交差点付近や，鉄道の駅周辺では等価騒音レベルが高い値を示していることが明らかである．また，世田谷区内全域 60[km^2] の環境騒音情報を統合処理し，地図上に表したものを図 4.15 に示す．図 4.15 から明らかなように，住宅地として人気である地域の住生活環境が必ずしも良いもの

図 4.13: 歩行センシング実験で取得したセンシングポイント．

図 4.14: あるユーザの足跡に沿った等価騒音レベルの様子．

ではないと言える．更に，図 4.16 には，同様に歩行センシング実験で取得した等価騒音レベルを示す．この図 4.16 から明らかなように，国道沿いや東京都道沿い，交差点付近では，その周囲と比較して高い騒音レベルを示す．これは，自動車や人の通行量が多く，環境騒音が大きいためと考えられる．

以上の結果より，モバイルデバイスを用いた参加型都市環境センシングが，実際の都市調査手法として有効であることが明らかである．

図 4.15: 世田谷区内全域の等価騒音レベルの様子．

図 4.16: 歩行センシング実験で取得した等価騒音レベルの様子．

但し，音声データの解析を行うことで，モバイルデバイスを用いた参加型センシングを実社会で活用する際に発生する新たな問題点が確認された．

本実験では，ユーザの持ち歩くスマートフォンから情報を取得した．取得した Wave データ分析したところ，ユーザがその地点を歩行・通過することによって発生する騒音が存在することが確認された．具体的には，繁華街や商店街では，ユーザの通過時に勧誘・声掛けがあり，等価騒音レベルの上昇が確認された．また，住宅街では，番犬の鳴き声や警察官からの声かけがあり，同様に等価騒音レベルの上昇が確認された．これらの音声情報は，フーリエ変換による周波数スペクトル解析を行うことによって，ある程度クラスタリングすることができる．しかしながら，スマートフォン上の計算資源・電源は限られたものとなるため，計算量の大きな高速フーリエ変換をリアルタイムにかけ，判定を行うことは難しい．

従来の固定的なセンサを用いたセンサネットワークによる観測や，JIS 規格に基づく騒音調査法では，上記のようなユーザの通過に伴う騒音発生は起こらない．今回は歩行実験後に，騒音レベルの高い値を示す地域の音声 Wave データを切り出し，確認作業を行った．しかし，スマートフォンを用いたセンシングを実社会で活用するためにも，より労力・コストのかからない判定手法の開発が重要であることが明らかであり，参加型都市環境センシングの実現のためには今後の重要な課題である．

また本システムは，都市環境に関する詳細な情報を取得することを目的としている．このため，センサデータを取得したモバイルデバイス上でプライバシー保護処理を行う場合，この詳細な情報を保持する必要がある．

4.1.5 効率的なシステム運用のためのユーザ分析

参加型センシングを運用するためには，多くのユーザを巻き込む必要があるため，報酬を与えることを前提としたインセンティブ付与に関する研究も数多く行われている [70, 83]．しかしながら，既存の研究ではインセンティブ付与の数理モデル化とシミュレーションを用いた研究に留まり，実際の参加型都市環境センシングに適用した例はない．

このため，本項では歩行センシング実験におけるユーザの行動を分析することによって，より効率的な参加型センシングの運用方法とインセンティブ付与の手法を探る．

まず，本センシング歩行実験を行うにあたり，図 4.5 に示す様に世田谷区全域のセンシング地点を 60 の小地域に分割し，センシングに参加するユーザ自身が歩行する地域を選択する形をとった．この時，取得すべきセンシング地点は事前に決定していたため，センシングのための歩行ルートを決定する問題は巡回セールスマン問題に帰着する．このため，ルート決定の計算量と社会における参加型センシング運用時の実用性を考慮して，歩行ルートは事前に決定しなかった．この時，ユーザごとのタスク量 (センシングエリアの数) は図 4.17 のようになった．

図 4.17 から明らかなように，多くのユーザが 3 以下のエリア数のみをセンシングしている一方で，少数のユーザが多大な貢献を行っている．この現象は，既存のオープンソースプロジェクト [84] と同様であり，少数のユーザが参加型センシングを支えていることを示す．このため，多大な貢献を示す少数のユーザは，センシングによる報酬ではなく，センシングに参加すること自体に対して魅力を感じていると示唆されるため，インセンティブ報酬の全体額を抑えながら，多くの情報を取得するためには，以下の 2 つの条件を満たすインセンティブ付与モデルを用いるべきである．

- あるユーザが受け取るインセンティブ報酬の合計は常に単調増加

図 4.17: 各ユーザのタスク量 (センシングエリアの数).

図 4.18: 提案するインセンティブモデルの例.

- ある一定のタスク γ 消化後はタスクあたりの報酬額は低下

この 2 点の条件を満たすインセンティブ報酬モデルとして，式 (4.3) を示す．ここで，タスクの量を t，α 及び β を一定タスク量 γ を決定する変数とする．また，インセンティブ付与による報酬額の合計を I とする．

$$I = (t+1)^{\frac{1}{\beta}} \times \log_e(t+1+\alpha) \tag{4.3}$$

また，この式 (4.3) に表すインセンティブ報酬モデルの例を図 4.18 に示す．

本提案手法は，インセンティブ付与による影響を最大限にするための設計である．この目的のために，より報酬によってセンシングへの貢献度が変化すると考えられる一般のユーザをセンシングに巻き込むインセンティブ報酬モデルを構築した．

4.1.6 自然環境調査への拡張

従来の参加型環境センシングは，人が居住する都市の住生活環境や人への影響を評価するために用いられてきた．しかし，今回の歩行センシング実験で取得した音声データを解析した結果，動物

図 4.19: 野鳥の音声自動識別システムの概要.

図 4.20: 既存の野鳥生態調査法に関するイメージ図.

や昆虫等の自然環境から発せられる情報をモバイルセンシングシステムで扱うことができる可能性が示唆された．ここで，参加型センシングは人口密度の高い都市圏で活用することを想定しているため，特に都市圏を生息地をするハシブトガラスとセミに注目し，自然環境情報の利活用の可能性について述べる．

モバイルデバイスを用いたハシブトガラス検知システム

今回の歩行センシング実験で取得した音声データ (Wave データ) を解析したところ，ハシブトガラスの鳴き声が確認できた．ハシブトガラスの鳴き声は $1,200$[Hz] 付近の成分を多く含むことが明らかにされており，一方で都市圏における環境騒音の多くは $1,200$[Hz] よりも低い音である．ま

図 4.21: Cicada Fingerprinting System の概要図．異種のセミの鳴き声を用いて位置補正を行う．

た，ハシブトガラスの音声自動識別の研究 [85] も進んでいることから，図 4.19 に示すような，モバイルデバイスを活用した参加型ハシブトガラス検知システムを構築できると考えられる．

従来の野生動物の調査は，図 4.20 に示すように，専門的な知識や器具を持つフィールド調査員の貢献に大きく依存しており，このため多くの調査員を必要とする広域に及ぶ多地点で同時に生態調査を行うことは困難であった．しかしながら，図 4.19 のようなシステムを実用化することによって，一般ユーザを生態調査に巻き込むことが可能となる．本システムを活用することによって，野鳥の観測を多地点で同時に行うことが可能になる．この特性を用いて，鳴き声の特徴が個体差であるのか，地域差であるのかを明確に示すことが可能となり，従来の研究では実証が難しかったハシブトガラスの地域に基づく方言の存在を確認することが期待できる．また本システムを用いることによって，都市における人と人工物 (工場，電車，飛行機，自動車等) の関係だけでなく，都市環境の中で生活する人と動物の関係に関する調査を，低コストで実現できる可能性が示唆された．

図 4.22: Cicada Fingerprinting System の評価実験を行ったエリア．周辺の地図を図 4.23 に示す．

セミの鳴き声を用いたモバイルセンシングシステム

都市圏に生息する生物の代表的な例として，ハシブトガラスの他にセミが挙げられる [86]．セミは一年の中で，限られた時間のみ活動するが，東京やタイなどのアジア地域では非常に多くの数が生息し，種によって鳴き声が異なる．また北米では，七年や十三年など長期間のサイクルで大発生することから，周期ゼミとして知られている [87, 88]．

まず，第 4.1.2 項で実装したセンシングアプリケーションを活用することによって，自然環境情報を Wi-Fi フィンガープリントの様に活用した位置補正システム Cicada Fingerprinting System を実現することができる．これは，都市圏におけるセミの生息地域と行動範囲が限定的なものとなることに依拠する [27]．また，日本やタイにおいては，同地域に複数のセミが生息し，各々の鳴き声が異なるという特徴を利用する [86]．このセミの鳴き声を利用した位置補正システム Cicada

図 4.23: Cicada Fingerprinting System の評価実験を行ったエリアの地図.

図 4.24: 地点 A で取得した音声の周波数分析の結果.

Fingerprinting System の概要図を図 4.21 に示す．この Cicada Fingerprinting System の実現によって，人工的なインフラを設置・利用することなく位置情報を取得することを可能とする．

この Cicada Fingerprinting System の評価実験として，セミの活動が活発となる夏期に，図 4.22 に示す東京大学駒場リサーチキャンパス内で簡易実験を行った．この図 4.22 の地図を図 4.23 に示す．約 12m 離れた二地点 A と B で 15 秒間の音声データを取得したときの様子を図 4.24 と図 4.24 に示す．この簡易実験を行ったエリアでは，GPS 信号を遮る屋根と壁があるために，GPS 情報のみでは正確な位置情報を取得することができない．しかしながら，評価実験の結果を示す図 4.24 及び図 4.25 から明らかなように，セミの鳴く木から近い地点では，高周波数の音を取得している．

また，追加実験として，図 4.26 に示すように，都市部の公園・道路・駐車場・東京大学キャンパス内の計 4ヶ所でそれぞれ 3 秒間の音声データを取得した．これらの 4ヶ所は，それぞれ 300m しか離れていない．本実験において取得した音声データに対して，周波数スペクトル解析を行った様子を図 4.27 に示す．この実験結果から明らかなように，植物の生い茂る公園やキャンパス内では，特に高周波数帯の音圧が大きくなる．一方，木の少ない道路などで取得できる音声データには高周波数帯の音圧はほとんど含まれない．

また，このセミの鳴き声の音量レベルを測定することによって，従来評価することが困難であった都市圏における “自然環境の豊かさ” を明らかにすることができる可能性が示唆された．従来，都市における “自然環境の豊かさ” の調査はフィールド調査員の主観的な評価によってなされており，都市計画や緑地計画に用いられる重要な指標となっていた．本提案手法では，セミの鳴き声分布を冬期・夏期に分けて取得することによって，“自然環境の豊かさ” を明らかにすることができる．上記の実験結果からも明らかなように，人工物の多い道路上や建造物の中においては，環境騒音レベルの季節間変位はほとんど見られない．一方で，自然環境が豊かで，植物の生い茂る地域では，セミの声によって環境騒音レベルが大きく変位する．今回の歩行センシング実験は住生活環境を目的としているため，冬期に行ったが，セミの活動が活発になる夏期に同様のデータを取得し，比較することによって，都市における緑地・公園の位置，また都市における “自然環境の豊かさ” を調査・可視化することが可能であると言える．

上記で示した自然環境情報に関する参加型センシングを実現することによって，生態調査や都市設計の分野に役立てることができるだけでなく，一般ユーザが普段の生活をおくる中で自然環境の存在を身近に感じることができる．

図 4.25: 地点 B で取得した音声の周波数分析の結果.

図 4.26: 夏期にセミの鳴き声を取得した 4ヶ所の様子.

図 4.27: 図 4.26 の 4ヶ所で取得したセミの鳴き声を含む音声データを周波数スペクトル解析した結果.

4.2　参加型都市環境センシングのためのプライバシー保護手法

第 4.1 節では，モバイルデバイスを用いた参加型都市環境センシングシステムを実装し，実際に歩行センシング実験を行うことでアプリケーションの可能性と取得するデータの特徴を明らかにした．特に，参加型都市環境センシングでは，連続値に近い形で位置情報及び環境情報を取得する．このため第 3 章で指摘したように，取得した統計のデータ信頼性が大きく低下してしまう．そこ

で，本節では細粒度のデータを扱う都市環境センシングを想定し，既存手法を拡張したプライバシー保護手法 Selective Negative Surveys を提案し，シミュレーション実験を通してプライバシーとデータ信頼性の観点から評価を行う．

4.2.1 提案手法: Selective Negative Surveys

参加型環境センシングで環境情報を取得する場合，データ区分の細かいセンサデータを取得する．このため，第 3 章で指摘し，図 3.2 に示す事前実験の結果からも明らかなように，既存の Perturbation 技術では適切に扱うことができない．そこで，この環境情報をデータ区分の細かい形で取得するための手法として，Selective Negative Surveys (SNS) を提案する．提案手法 SNS では，取得したセンサデータに対して，モバイルデバイス上でプライバシー保護処理を行い，プライバシー保護処理済データを生成する．このプライバシー保護処理済データを中央サーバに伝送し，サーバはこのプライバシー保護処理済データを収集した後に，再構築処理を施して，統計的なデータ分布を取得する．

ここで，多次元データの次元を D，各次元の持つ選択肢の数を $\alpha_1, \alpha_2, \cdots, \alpha_D$ と定義する．また，センシングで取得したオリジナルデータを $\vec{x}$，プライバシー保護処理済のデータを $\vec{y}$ と表し，それぞれの次元における選択肢を $\vec{x} = \langle x_1, x_2, \cdots, x_D \rangle$，$\vec{y} = \langle y_1, y_2, \cdots, y_D \rangle$ と表すこととする．

また，本提案手法では到達先パラメータ $\beta_1, \beta_2, \cdots, \beta_D$ を設定し，プライバシー保護処理に用いる．この到達先パラメータを用いることによって，細かいデータ区分で表す情報の取得を可能なものとする．また，この到達先パラメータを各次元で調整することによって，それぞれの次元の持つ選択肢数とプライバシーレベルの要求に応じたプライバシー保護処理を実現する．

以下では，提案する Selective Negative Surveys(SNS) の処理をモバイルデバイス上のプライバシー保護処理，中央サーバ上の再構築処理に分けて示す．

モバイルデバイス上の処理

提案手法 Selective Negative Surveys では，一般ユーザの持つモバイルデバイスで，都市環境情報を取得し，カテゴリ化した後に，プライバシー保護処理を行う．このプライバシー保護処理済データを中央サーバに報告する．但し，モバイルデバイスは有する計算資源・電源は限りあるものであるため，プライバシー保護処理は簡潔な処理を用い，計算量を削減する．

本提案手法における，オリジナルデータのフィールド i における値 x_i と，プライバシー保護処理済データにおけるフィールド i の値 y_i の関係を式 (4.4) に示す．つまり，取得したデータ x_i の値に応じて，y_i の取りうる値が変化する．本提案手法では，到達先パラメータ β_i を設定することによって，生成されるプライバシー保護処理データ y_i の取りうるカテゴリ値を限定し，これによって計算機による擬似乱数が与える影響を抑える．また，常に $x_i \neq y_i$ を満たすため，ユーザは取得したオリジナルデータを報告する必要がない．

$$y_i < x_i - \beta_i \ , \quad x_i + \beta_i < y_i \tag{4.4}$$

モバイルデバイス上の処理を具体的に示すために，8×13 の 2 次元データを参加型センシングで収集すると想定する．但し，8×13 の 2 次元データは選択肢 8 と選択肢 13 の 2 つのフィールド情報を有することを意味する．また，y_i を決定する到達先パラメータを $\{\beta_1 = 1, \beta_2 = 2\}$ とする．

この時，モバイルデバイスに搭載されたセンサで，$\{x_1 = 4, x_2 = 4\}$ のセンサデータを取得した時の様子を図 4.28 に示す．本提案手法 Selective Negative Surveys のプライバシー保護処理では，取得したオリジナルデータ $\vec{x}$ とアプリケーション提供者が設定した到達先パラメータ β_1，β_2 から，プライバシー保護処理済データ $\vec{y}$ を決定する．図 4.28 の場合では，40 個の白セルの中から等確率に 1 つのセルを選び，プライバシー保護処理済データとする．これは，常に $\{x_1 = 4, x_2 = 4\}$ とは異なるデータをサーバに報告することを意味する．モバイルデバイス上の処理は常に式 (4.4) の関係に従い，プライバシー保護処理済データ $\vec{y}$ を生成する．

本提案手法では，各々のユーザは，センシングしたデータと異なるデータを報告することとなる．このため，データ伝送中に悪意ある攻撃者が盗聴によってデータ取得に成功しても，ユーザのプライバシー情報特定はできない．また，アプリケーション提供者 (サーバ側) が個々のユーザがセンシングで得た真の値を取得することはできないので，プライバシー情報を特定することはできない．つまり，本提案手法ではセンシングに参加するユーザが自身のプライバシー情報を他者に提供する必要がないため，プライバシーの保護が保証される．また，本提案手法のプライバシー保護処理は，非常に簡潔な手法であり，計算量も小さなものとなるため，計算資源の乏しいモバイルデバイスでも十分に実用可能な手法である．

中央サーバ上の処理

提案手法 Selective Negative Surveys において，中央サーバは各モバイルデバイス上でプライバシー保護処理を施したデータを収集する．これらのプライバシー保護処理済データを用いて統計的なデータ分布の再構築処理を行う．

プライバシー保護処理済データの離散分布を $Y(\vec{x})$，再構築データの離散分布を $A(\vec{x})$ と表す．また，$A'(\vec{x})$ を式 (4.5) の様に定義する．

図 4.28: 8×13 の 2 次元データを取得する時の提案手法 SNS での処理の様子．$\{x_1 = 4, x_2 = 4\}$ をセンシングした時，白セルの中から 1 つのセルをランダムに選んで伝送する．

$$A'(\vec{x}) = \sum_{\substack{\vec{y}s.t. \\ x_i-\beta_i \leq y_i \leq x_i+\beta_i \\ \forall i \in d}} A(\vec{y}) \tag{4.5}$$

$A'(\vec{x})$ は，$A(\vec{x})$ を中心として隣接する，$(1+2\beta_1) \times (1+2\beta_2) \times \cdots \times (1+2\beta_D)$ 個のカテゴリに属するデータの総計を指す．この時，$A'(\vec{x})$ と $Y(\vec{x})$ の関係は，式 (4.6)，(4.7) の様に表すことができる．この関係式は，式 (2.4) を D 次元の多次元データへ拡張したものであり，理論的には常に統計的なデータ分布を正確に再構築することができる．$B\big(\{1,\cdots,D\},k\big)$ は $\{1,\cdots,D\}$ の中から k 個の数を選んだ時の組み合わせを表す．

$$\forall \vec{x} \mid A'(\vec{x}) = N + \sum_{k=1}^{D} (-1)^k \cdot \Gamma(\vec{x},k) \tag{4.6}$$

$$\Gamma(\vec{x},k) = \sum_{\substack{d \in \\ B(\{1,2,3,\cdots,D\},k)}} \left(\left[\prod_{j \in d} (\alpha_j - 1 - 2\beta_j) \right] \cdot \sum_{\substack{\vec{y}s.t. \\ y_i \in \vec{x} \\ \forall i \in d}} Y(\vec{y}) \right) \tag{4.7}$$

式 (4.5)，(4.6)，(4.7) を用いることによって，プライバシー保護処理済データの統計的なデータ分布から，実際にセンシングで取得したオリジナルデータの統計的なデータ分布を再構築することができる．

図 4.28 を例にとって，A' と A の関係を考えると，$A'(4,4) = A(3,2)+A(3,3)+A(3,4)+A(3,5)+A(3,6)+A(4,2)+A(4,3)+A(4,4)+A(4,5)+A(4,6)+A(5,2)+A(5,3)+A(5,4)+A(5,5)+A(5,6)$ と表すことができる．$\beta_1 = 1$，$\beta_2 = 2$ であるため，$A'(4,4)$ は図 4.28 において，$\big\{3 \leq x_1 \leq 5, 2 \leq x_2 \leq 6\big\}$ を満たす 15 個のカテゴリに属するデータの総計を表すこととなる．式 (4.6)，(4.7) から，$A'(\vec{x})$ を導いた後に，連立方程式を解くことによって，再構築データのデータ分布 $A(\vec{x})$ を求めることができる．

本提案手法を用いることによって，ユーザの各々のプライバシー情報を秘匿化しながら，多次元データの統計的なデータ分布を構築することが可能である．また，到達先パラメータ $\beta_1, \beta_2, \cdots, \beta_D$ の値を変化させることによって，アプリケーションの要求するプライバシーレベルとデータ信頼性に応じた処理を行うことが可能である．

4.2.2 評価

本提案手法 Selective Negative Surveys (SNS) の評価のために実データを用いたシミュレーションを行った．まず，二酸化炭素濃度の地域差に関するデータを用いて，プライバシー保護とデータの信頼性の観点から評価を行った．次に，環境情報の確率分布をモデル化したデータと，正規分布の確率分布をとるデータを用意し，データ構造の異なる多次元データに対して，本提案手法のプライバシー保護処理を行った．ここで，データ構造は，データの次元数と各次元が持つ選択肢の数を指すものとする．

プライバシー保護の評価を行うために，悪意ある攻撃者がプライバシー保護処理済データを盗聴によって取得した際に，センシングで取得したオリジナルデータを正確に推定できる確率を用いる．このため，プライバシーの評価指標を $Privacy$ と定義し，式 (4.8)，(4.9) のように表す．$Privacy$ は，最大事後確率推定法 [61] を基礎とした指標である．但し，攻撃者はユーザに関しての事前知識は全く持っていないものとする．

この $Privacy$ の値が 0 に近いほど，攻撃者がセンサデータを推定できる確率が低い．

$$Privacy = \sum_{\substack{\Upsilon \in Y(\vec{x}) \\ \forall \vec{x}}} P\left(\Upsilon | \widehat{X_\chi}\right) \cdot P(\widehat{X_\chi}) \tag{4.8}$$

$$\widehat{X_\chi} = \arg\max_{\substack{\chi \in X(\vec{x}) \\ \forall \vec{x}}} P(\chi | Y) \tag{4.9}$$

また，データ信頼性のための評価指標として，式 (3.1) で定義したカルバックライブラー情報量 D_{KL}[89] と，データ整合率 $\mathbf{C}$ を用いる．このデータ整合率 $\mathbf{C}$ を式 (4.10) に示す．

$$\mathbf{C} = 100 \times \sum_i min(X(i), A(i)) \tag{4.10}$$

二酸化炭素濃度センシングへの適用

本シミュレーションでは，図 4.29 に示すようなユードム社の無線型二酸化炭素測定器 (C2D-W02TR) を用いて実際の二酸化炭素濃度を測定し，これを用いて 6 つの地点における二酸化炭素濃度に関する擬似データを用意した．この擬似データに対して既存手法である Multidimensional Negative Surveys(MNS) と本提案手法 Selective Negative Surveys(SNS) を用いた場合に取得したデータ分布の信頼性とプライバシー保護について評価を行う．本シミュレーションでは，6 の地点 (α_1) の二酸化炭素濃度を測定する．二酸化炭素濃度のデータ区分 (α_2) は 7 つと設定する．この時，扱うデータは 6×7 の 2 次元データとなる．二酸化炭素濃度のデータ分布は，地点 1，2，3 においては二項分布をとり，地点 4 においては一様分布，地点 5 においてはバースト状のデータ分布をとるものとする．地点 6 のデータとして，実際に自然環境で取得した二酸化炭素濃度の確率分布を用いる．

本シミュレーションは，都市地域で環境情報を取得することを想定しており，Bertaud らの研究 [90] で示された都市モデルを参考にしている．Bertaud らの都市モデル [90] より，それぞれの都市地域においては，10000 人のセンシング参加者を確保することが可能であることから，あるセンシング地点において 10000，6 地点の合計で 60000 のデータを取得すると想定する．この取得した環境情報に対して，既存手法 MNS 及び提案手法 SNS($\beta_1 = 0$ かつ $\beta_2 = 1$, $\beta_1 = 0$ かつ $\beta_2 = 2$) を用

表 4.1: 6×7 の 2 次元データを用いた時のデータ信頼性とプライバシーに関する評価.

Data sets: 6×7, $D = 2$, $N = 60000$			
	データ整合率 $\mathbf{C}$	D_{KL}	$Privacy$
Negative Surveys	79.6	0.4256	0.03057
SNS($\beta_1 = 0$,$\beta_2 = 1$)	86.5	0.2932	0.05312
SNS($\beta_1 = 0$,$\beta_2 = 2$)	90.6	0.1154	0.1241

いてシミュレーション実験を行った．センシングしたオリジナルデータのデータ分布と，MNS 及び SNS の処理を施した時の再構築データのデータ分布を図 4.30 に示す.

図 4.30 の実験結果において，既存手法 MNS によるカルバックライブラー情報量 D_{KL}[89] は 0.4331，$Privacy$ は 0.03198 となる．一方で，提案手法 SNS($\beta_1 = 0$, $\beta_2 = 1$) の処理では，D_{KL} は 0.2798，$Privacy$ は 0.05242 となる．また，SNS($\beta_1 = 0$, $\beta_2 = 2$) では D_{KL} は 0.1039，$Privacy$ は 0.1391 となった．これらの結果より，提案手法 SNS を用いた再構築処理の方が，既存手法 MNS を用いた時よりも，センシングデータにより近い形で，統計的なデータ分布を再構築できることが明らかである．つまり，データ信頼性が向上している．また，同様のデータ分布において，それぞれの処理を 10 回行い，データ整合率 $\mathbf{C}$ とカルバックライブラー情報量 D_{KL} 及び $Privacy$ の平均したものを表 4.1 に示す.

この表 4.1 に示す結果から，SNS($\beta_1 = 0$, $\beta_2 = 1$) の処理では，既存手法 MNS におけるデータ整合率 $\mathbf{C}$ と比べて 7 %ほど改善していることが明らかである．また，SNS($\beta_1 = 0$, $\beta_2 = 2$) では，既存手法 MNS と比べて 11 %ほど改善したことがわかる．この一方で，$Privacy$ はそれぞれ 0.02, 0.09 増大している．このことから，β の値を大きくすると，プライバシー保護処理済データからセンシングデータがより推定しやすくなることが明らかである.

図 4.29: ユードム社の無線型二酸化炭素測定器 (左) と Android 端末 Galaxy S3(右).

図 4.30: 既存手法 MNS 及び提案手法 SNS を用いた時の二酸化炭素濃度のデータ分布の様子.

本提案手法における，データ信頼性とプライバシー保護は，トレードオフの関係となる．到達先パラメータを調整することにより，センシングに参加するユーザの数やデータ構造 (次元数，各次元の持つ選択肢の数) に合わせたプライバシー保護処理を行う事が可能となる．

環境情報地図

本提案手法は，6×7 の 2 次元データだけでなく，様々なデータ構造 (次元数，各次元の持つ選択肢の数) を持つ多次元データに対して用いることができる．本評価実験では，環境情報をモデル化したデータ分布と，正規分布をとるデータ分布を用いる．環境情報をモデル化したデータ分布を取得するために，実際に二酸化炭素センサ及び湿度センサを用いて環境情報を収集し，モデル化を行ったものを環境情報確率分布関数とする．この環境情報確率分布関数を $\Psi(i)$ として式 (4.11) に示す．この時，平均値を μ とする．また，正規分布関数を式 (4.12) に示す．

$$\Psi(i) = \begin{cases} \frac{1}{4} \times N & (\mu - 1 \leq i \leq \mu + 1) \\ \frac{1}{8} \times N & (i = \mu - 2, \mu + 2) \\ 0 & (\text{otherwise}) \end{cases} \tag{4.11}$$

$$N(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{\sigma}} \varepsilon(\mu) = \frac{1}{\mu} \tag{4.12}$$

環境情報確率分布関数 $\Psi(i)$ 及び正規分布関数をとるデータを用いて，様々なデータ構造の多次元データに対する MNS 及び提案手法 SNS の評価を行う．$D = 2$ (14×17 の 2 次元データ)，$D = 3$

表 4.2: 環境情報確率分布を用いた時のデータ信頼性とプライバシーに関する評価.

Data sets: 14×17, $D = 2$, $N = 30000$			
	データ整合率 $\mathbf{C}$	D_{KL}	$Privacy$
Negative Surveys	53.8	0.5489	0.004982
SNS($\beta_1 = 2$,$\beta_2 = 5$)	60.3	0.5501	0.01859
SNS($\beta_1 = 5$,$\beta_2 = 6$)	76.8	0.4287	0.07211
Data sets: $8 \times 8 \times 9$, $D = 3$, $N = 30000$			
	データ整合率 $\mathbf{C}$	D_{KL}	$Privacy$
Negative Surveys	54.9	0.5382	0.003162
SNS($\beta_1 = 2$,$\beta_2 = 2$,$\beta_3 = 3$)	68.6	0.4713	0.05645
SNS($\beta_1 = 3$,$\beta_2 = 3$,$\beta_3 = 3$)	93.9	0.0853	0.1942
Data sets: $8 \times 8 \times 8 \times 8 \times 9 \times 9$, $D = 6$, $N = 30000$			
	データ整合率 $\mathbf{C}$	D_{KL}	$Privacy$
Negative Surveys	51.1	0.5539	0.00003551
SNS($\beta_1 = 2$,$\beta_2 = 2$,$\beta_3 = 2$,$\beta_4 = 2$,$\beta_5 = 2$,$\beta_6 = 3$)	51.3	0.5186	0.001242
SNS($\beta_1 = 3$,$\beta_2 = 3$,$\beta_3 = 3$,$\beta_4 = 3$,$\beta_5 = 3$,$\beta_6 = 4$)	78.9	0.3898	0.1206

($8 \times 8 \times 9$ の 3 次元データ)，$D = 6$ ($8 \times 8 \times 8 \times 8 \times 9 \times 9$ の 6 次元データ)，に対して MNS 及び提案手法 SNS を用いた時のオリジナルデータと再構築データのデータ整合率 $\mathbf{C}$ について，環境情報確率分布を用いた時の様子を図 4.31 に，また正規分布を用いた時の様子を図 4.32 に示す．それぞれ 10 回の処理を行い，その平均の値を記す．また，それぞれの多次元データにおけるデータ信頼性とプライバシーに関する評価を表 4.2 と表 4.3 に示す．評価指標 $Privacy$ は，式 (4.8)，(4.9) のものを用いているため，データ数 N には依存しないことが明らかである．このため，表 4.2 と表 4.3 には，$N = 30000$ の時の評価値のみを示す．これらの結果より，環境情報確率分布のデータ及び正規分布のデータの両方において，提案手法 SNS の処理がデータ信頼性の観点で，既存手法よりも優れていることが明らかである．

表 4.3: 正規分布を用いた時のデータ信頼性とプライバシーに関する評価.

Data sets: 14×17, $D = 2$, $N = 30000$			
	データ整合率 $\mathbf{C}$	D_{KL}	$Privacy$
Negative Surveys	54.9	0.5129	0.005492
SNS($\beta_1 = 2$,$\beta_2 = 5$)	58.8	0.4923	0.02011
SNS($\beta_1 = 5$,$\beta_2 = 6$)	77.3	0.4888	0.07857
Data sets: $8 \times 8 \times 9$, $D = 3$, $N = 30000$			
	データ整合率 $\mathbf{C}$	D_{KL}	$Privacy$
Negative Surveys	50.3	0.5521	0.004209
SNS($\beta_1 = 2$,$\beta_2 = 2$,$\beta_3 = 3$)	69.2	0.4503	0.05998
SNS($\beta_1 = 3$,$\beta_2 = 3$,$\beta_3 = 3$)	92.4	0.0701	0.1974
Data sets: $8 \times 8 \times 8 \times 8 \times 9 \times 9$, $D = 6$, $N = 30000$			
	データ整合率 $\mathbf{C}$	D_{KL}	$Privacy$
Negative Surveys	51.9	0.5391	0.00001874
SNS($\beta_1 = 2$,$\beta_2 = 2$,$\beta_3 = 2$,$\beta_4 = 2$,$\beta_5 = 2$,$\beta_6 = 3$)	59.2	0.4924	0.001078
SNS($\beta_1 = 3$,$\beta_2 = 3$,$\beta_3 = 3$,$\beta_4 = 3$,$\beta_5 = 3$,$\beta_6 = 4$)	76.3	0.4201	0.09884

4.2.3 提案手法に関する考察

図 4.31，図 4.32 の結果から明らかなように，取得するデータの持つ各フィールドのカテゴリ数が大きくなると，データの信頼性は低下する．一方で，データ数 N，つまりセンシングに参加するユーザの数が多い時は，データの信頼性が向上する．これらのことを考慮し，センシングに参加するユーザの数が少ない時には，到達先パラメータを大きな値に設定し，参加するユーザの数が多い時には到達先パラメータを小さな値に設定することによって，アプリケーションの本来の目的であるデータの信頼性を常に確保することが可能となる．特に，参加型環境センシングのアプリケーション提供開始時には，センシングに参加するユーザの数が少ないことが考えられる．このことを考慮し，アプリケーション提供開始時には到達先パラメータを大きく設定し，参加ユーザ数が増えるにつれて，到達先パラメータの値を小さく変化させることによって，一定の質のサービスを提供することができる．

また，参加型環境センシングで位置情報・二酸化炭素濃度情報の 2 種類を取得するアプリケーションを想定し，あるユーザが 1 分間隔で，センサデータを取得した時の例を表 4.4 に示す．データのカテゴリは上述のシミュレーション実験と同様のものとし，到達先パラメータはそれぞれ $\beta_1 = 0$, $\beta_2 = 1$ とする．表 4.4 に示した例から明らかなように，仮に悪意ある攻撃者が盗聴によってプライバシー保護処理済データを取得した場合にも，センシング参加者の取得したオリジナルのセンサデータを正確に推定することは不可能である．

また，悪意ある攻撃者がセンシング参加者のセンサデータを推定する手法として，プライバシー保護処理済データをある一定時間 Δt 連続して取得する蓄積攻撃が考えられる．時間変化をしないデータであれば，$(\alpha_1 - 2\beta_1 - 1) \times (\alpha_2 - 2\beta_2 - 1) \times \cdots \times (\alpha_D - 2\beta_D - 1)$ の量のデータを連続的に取得する蓄積攻撃によって，オリジナルデータを推定することが可能となる．しかし，本研究で

表 4.4: センサデータとプライバシー保護処理済データの例.

センサデータ		プライバシー保護処理済データ	
地区	CO_2[ppm]	地区	CO_2[ppm]
1	700〜800	2	500〜600
1	600〜700	5	1000〜
2	500〜600	5	900〜1000
2	600〜700	1	〜500

対象とする，音量や二酸化炭素濃度などの環境情報は，時々刻々と変化をするデータである．本提案手法ではユーザ側でプライバシー保護処理を行うため，攻撃者がセンサデータの変化が起こる時点を知る事ができない．このため，蓄積攻撃によってオリジナルのセンサデータを推定することは不可能であると言える．

4.3 本章のまとめ

本章では，まずモバイルデバイスを用いた参加型都市環境センシングシステムを実装し，東京都世田谷区全域において歩行センシング実験を行い，都市環境に関するデータを収集した．本実験で取得した都市環境に関するデータを地図上に可視化することによって，参加型センシングで取得したデータが都市環境調査に有効であることを示した．更に，取得したセンサデータを解析することによって，参加型センシングを，人間の生活のための都市環境調査だけでなく，自然環境調査にも活用することができる可能性が示唆された．

また，参加型都市環境センシングで取得するデータの特徴を考慮した上で，既存の Perturbation 技術を拡張し，プライバシー保護手法 Selective Negative Surveys を提案した．本提案手法では，既存手法の Negative Surveys が計算機を用いて生成する擬似乱数の影響を強く受け，データ信頼性が大きく低下する問題を解決するために，到達先パラメータを設定し，擬似乱数による影響を抑えた．これによって，より連続値に近い，データ区分の細かいカテゴリデータを参加型センシングで扱うことも可能となった．このため，本提案手法 Selective Negative Surveys を用いることによって，一般のユーザ及びアプリケーション提供者が安心して参加型都市環境センシングを用いて統計的なデータ分布を取得することができる土壌を確立したと言える．

本章で取り組んだ参加型都市環境センシングを実現することによって，一般ユーザの持つモバイルデバイスを都市環境調査に活用することができる．この取り組みによって，データの信頼性の高い環境騒音地図や粒子分布図，環境汚染地図など様々な環境情報地図を同時に構築する手法を確立した．

図 4.31: 環境情報確率分布を用いた時のデータ整合率 **C** の推移.

図 4.32: 正規分布を用いた時のデータ整合率 **C** の推移.

第5章　プロファイル情報を対象とした参加型センシングとプライバシー保護

第 3.2 節に示したように，既存の Perturbation 技術では，二者択一型選択肢を含む多次元データを扱うことができないという深刻な技術的な課題がある．モバイルデバイスを用いた参加型センシングでは，性別などのプロファイル情報を収集する際に二者択一型選択肢を頻繁に用いるため，二者択一型選択肢を含む多次元データのためのプライバシー保護手法の確立は非常に重要である．また，Yes/No での回答が求められるイベント検知用の参加型センシングでも同様に二者択一型選択肢を含む多次元データを扱う．

これらをふまえて本章では，ある次元の選択肢の数が 2 つとなる多次元データのプライバシー保護を実現するために，Randomized Resonse と Negative Surveys を組み合わせる MNS-RRT アルゴリズムを提案し，シミュレーションを通してその有効性を示す．

5.1　二者択一型選択肢を含む多次元データを扱う参加型センシングシステムの概要

まず本提案手法が想定する，二者択一型選択肢を含む多次元データを扱う参加型センシングシステムの概要と具体例について示すことによって，本章で対象とする参加型センシングシステムを明らかにする．

ヘルスケア分野や企業の市場調査のための参加型センシング [1] では，ユーザの性別や特定疾患の有無など二者択一型で表すプロファイル情報を扱う機会が頻繁にある．ここで，表 5.1 にプロファイル情報とセンサデータを同時に収集する参加型センシングで扱う多次元データの例を示す．表 5.1 に示すように，プロファイル情報や閾値判定では選択肢が 2 つのみであるフィールド情報を扱う．更に，モバイルデバイスに搭載されたセンサを用いた地震検知や交通事故のリアルタイム調査，放射線量のモニタリングなど，イベント検知のための参加型センシングシステムの研究が幅広く進んでおり，これらのシステムでも Yes/No で報告を行うため二者択一型選択肢を含む多次元データを扱う．

しかしながら，二者択一型選択肢の情報をアプリケーション提供者に報告することによって，ユーザの位置情報特定を助けたり，二者択一型選択肢の回答そのものがプライバシー情報となる場合がある．特に，ユーザ自身の情報を示すプロファイル情報を扱う際や，放射線量の閾値判定などの際には，プライバシー情報保護の観点から情報の取り扱いは非常に難しいため，これらの参加型センシングシステムを実社会で活用していくためには，二者択一型選択肢を含む多次元データに対するプライバシー保護手法の開発が非常に重要となる．

第 2 章では，既存の Perturbation 技術を紹介し，この二者択一型選択肢を含む多次元データのプライバシー保護を可能とする手法が存在せず，大きな課題となっている点について述べた．Perturbation 技術として広く用いられている Randomized Response では，そもそも多次元データの持つ複数フィールド情報の相関性を保持したまま，プライバシー保護処理を行うことはできない [61]．また，Negative Surveys の手法では，オリジナルデータとプライバシー保護処理済データが常に相反した関係となるため，二者択一型選択肢のプライバシー保護に用いることはできない．

以下では，既存の Perturbation 技術で課題となっている二者択一型選択肢を含む多次元データに対するプライバシー保護を実現するために，MNS-RRT アルゴリズムを提案する．

5.2 プロファイル情報を対象とした参加型センシングシステムのためのプライバシー保護手法

本節では，二者択一型選択肢を含む多次元データを扱う参加型センシングにおけるプライバシー保護を実現する MNS-RRT アルゴリズムを提案し，シミュレーション実験を通してデータ信頼性とプライバシー保護の観点から評価を行う．

5.2.1 提案手法: MNS-RRT アルゴリズム

二者択一型選択肢を含む多次元データに対するプライバシー保護を実現するための手法として，MNS-RRT アルゴリズムを提案する．本提案手法は，第 2 章で紹介した Randomized Response を Negative Surveys に組み合わせたものであり，二者択一型選択肢のフィールド情報に対しては Randomized Response によるプライバシー保護処理を，その他のフィールド情報に対しては Negative Surveys のプライバシー保護処理を行う．つまり，取得したセンサデータに対して，モバイルデバイス上でプライバシー保護処理を行い，プライバシー保護処理済データを生成する．このプライバシー保護処理済データを中央サーバに伝送し，サーバはこのプライバシー保護処理済データを収集した後に，再構築処理を施して，統計的なデータ分布を取得する．

ここで，第 4 章と同様に，多次元データの次元を D，各次元の持つ選択肢の数を $\alpha_1, \alpha_2, \cdots, \alpha_D$ と定義し，また α_i は式 (5.1) のように各次元が有する選択肢数の降順とする．

$$\alpha_1 \geqq \alpha_2 \geqq \cdots \geqq \alpha_D \tag{5.1}$$

以下では，提案する MNS-RRT アルゴリズムの処理をモバイルデバイス上のプライバシー保護処理，中央サーバ上の再構築処理に分けて示す．

表 5.1: プロファイル情報とセンサデータを同時に収集する多次元データの例.

User ID	Time	Location	Sex	Generation	Condition	Moving (Y/N)	Volume
14	17:15:35	(0.77, 1.68)	Male	40-50	Lumbago	Yes	55.58
14	17:16:35	(0.79, 1.94)	Male	40-50	Lumbago	Yes	59.74
14	17:17:35	(0.79, 1.94)	Male	40-50	Lumbago	No	60.64
15	17:15:46	(1.24, 1.68)	Female	30-40	Good	No	68.87
15	17:17:46	(1.24, 1.68)	Female	30-40	Good	Yes	62.24
17	17:19:35	(2.31, 0.14)	Male	70-80	Dementia	Yes	41.28
17	17:15:35	(2.25, 0.11)	Male	70-80	Dementia	Yes	45.19
17	17:15:35	(2.31, 0.14)	Male	70-80	Dementia	No	46.25
17	17:15:35	(2.31, 0.14)	Male	70-80	Dementia	No	48.34
20	18:20:04	(0.98, 1.91)	Male	30-40	Good	Yes	81.88
20	18:25:04	(1.47, 1.32)	Male	30-40	Good	No	79.76

モバイルデバイス上の処理

提案手法 MNS-RRT アルゴリズムは，プロファイル情報を扱う参加型センシングへの実装を想定しており，モバイルデバイス上では二者択一型選択肢を含む多次元データを取得し，プライバシー保護処理を行った後に，プライバシー保護処理済データを中央サーバに報告する．但し，モバイルデバイスは有する計算資源・電源は限定されるため，プライバシー保護処理は簡潔な処理を用い，計算量の削減を目指す．

センシングで取得したオリジナルデータを $\vec{x}$，プライバシー保護処理済のデータを $\vec{y}$ と表し，それぞれの次元における選択肢を $\vec{x} = \langle x_1, x_2, \cdots, x_D \rangle$，$\vec{y} = \langle y_1, y_2, \cdots, y_D \rangle$ と表すこととする．また，二者択一型選択肢のデータ次元は式 (5.1) の関係より，α_D となるため，$\alpha_D = 2$ を常に満たす．この時，プライバシー保護処理済データの遷移確率は，式 (5.2) に示す $\mathbf{\Psi}_{\alpha_1,\cdots,\alpha_{D-1}}$ に従うものとする．

$$\mathbf{\Psi}_{\alpha_1,\cdots,\alpha_{D-1}} = \begin{cases} 0 & (y_k = x_k) \\ p \times \prod_{k=1}^{D-1} \frac{1}{\alpha_k - 1} & (y_k \neq x_k \vee y_D = x_D) \\ (1-p) \times \prod_{k=1}^{D-1} \frac{1}{\alpha_k - 1} & (y_k \neq x_k \vee y_D \neq x_D) \end{cases} \tag{5.2}$$

ここで，パラメータ p は常に $0 < p < 1$ $(p \neq 0.5)$ を満たすものと定義する．パラメータ p は，α_D の次元の選択肢のプライバシー保護処理済データの生成確率を表す．このパラメータ p の値は，アプリケーション提供者が事前に設定することができるパラメータである．

つまり，3 つ以上の選択肢を持つフィールド情報に対しては，既存の Negative Surveys に従ってプライバシー保護処理を施し，同時に 2 つのみの選択肢を持つフィールド情報に対しては，Randomized Response を施す．これによって，多次元データの複数フィールド間のデータ相関性を保ったまま，プライバシー保護処理を行うことが可能となる．

ここで，4 つの選択肢を持つフィールド情報と 2 つの選択肢を持つフィールド情報を持つ 2 次元データ $(\alpha_1 = 4,\ \alpha_2 = 2)$ を例にとって，MNS-RRT アルゴリズムによるプライバシー保護処理の様子を記す．センシングで取得したデータが，$\vec{x} = <x_1 = 2, x_2 = 1>$ であった時，プライバシー保護済データの生成確率は図 5.1 の様になる．この時，x_1 の値は Negative Surveys の処理に従って，y_1 の値を決定する．また，x_2 の値は Randomized Response に従って，y_2 の値を決定する．この時の，プライバシー保護済データの生成確率は，式 (5.3) のようになる．

$$\mathbf{\Psi_4} = \begin{cases} 0 & (\text{ある灰色セルの生成確率}) \\ p \times \frac{1}{3} & (\text{ある黒色セルの生成確率}) \\ (1-p) \times \frac{1}{3} & (\text{ある白色セルの生成確率}) \end{cases} \tag{5.3}$$

この MNS-RRT アルゴリズムの処理では，プライバシー保護処理済データとオリジナルデータが常に相反した関係となるわけではない．二者択一型選択肢である次元において，プライバシー保護処理済データとオリジナルデータが同じ値となりうる．このため，悪意ある攻撃者がプライバシー保護処理済データを盗聴によって取得した場合も，正確にオリジナルデータを推定することは不可能である．また，たとえ悪意ある攻撃者がパラメータ p の値を知っていたとしても，確実に推定することは不可能であるため，センサデータを報告するユーザのプライバシーの保護を実現したと言える．また，本提案手法のモバイルデバイス上の処理は，非常に簡潔な手法であり，計算量も小さなものとなる．このため，計算資源の乏しいモバイルデバイスでも処理可能な手法である．

図 5.1: センシングで取得したデータが $\vec{x} = <x_1 = 2, x_2 = 1>$ である時，MNS-RRT アルゴリズムにおけるプライバシー保護処理で生成されるデータの生成確率は式 (5.3) で示す $\boldsymbol{\Psi_4}$ の様になる．

中央サーバ上の処理

提案手法 MNS-RRT アルゴリズムにおいて，中央サーバでは各モバイルデバイス上でプライバシー保護処理を施したデータを収集し，統計的なデータ分布の再構築処理を行う．ここで，センシングによって取得したオリジナルデータの統計的なデータ分布を $\hat{\boldsymbol{X}}$，プライバシー保護処理されたデータの統計的なデータ分布を $\hat{\boldsymbol{Y}}$，サーバで再構築された統計的なデータ分布を $\hat{\boldsymbol{A}}$ とする．$\hat{\boldsymbol{X}}$ と $\hat{\boldsymbol{Y}}$ の関係は，式 (5.4) に示す行列 $\hat{\boldsymbol{M}}_{\alpha_1,\cdots,\alpha_D}$ で表すことができる．この行列 $\hat{\boldsymbol{M}}_{\alpha_1,\cdots,\alpha_D}$ は，各モバイルデバイス上のプライバシー保護処理を巨視的に捉えたものである．

$$\hat{\boldsymbol{Y}} = \hat{\boldsymbol{M}}_{\alpha_1,\cdots,\alpha_D} \cdot \hat{\boldsymbol{X}} \tag{5.4}$$

つまり，サーバ上でオリジナルデータの統計的なデータ分布を導き出すためには，式 (5.5) に示すように逆行列 $\hat{\boldsymbol{M}}^{-1}_{\alpha_1,\cdots,\alpha_D}$ を用いれば良い．逆行列 $\hat{\boldsymbol{M}}^{-1}_{\alpha_1,\cdots,\alpha_D}$ 及びプライバシー保護処理済データの統計的なデータ分布を $\hat{\boldsymbol{Y}}$ から，統計的なデータ分布 $\hat{\boldsymbol{A}}$ を導くことができる．

$$\hat{\boldsymbol{M}}^{-1}_{\alpha_1,\cdots,\alpha_D} \cdot \hat{\boldsymbol{Y}} = \hat{\boldsymbol{A}} \tag{5.5}$$

以下では，まず各モバイルデバイス上のプライバシー保護処理を統計的に表す $\hat{\boldsymbol{M}}_{\alpha_1,\cdots,\alpha_D}$ を導き，その規則性と再帰性を利用して逆行列 $\hat{\boldsymbol{M}}^{-1}_{\alpha_1,\cdots,\alpha_D}$ を導出し，センシングによって取得したデータの統計的なデータ分布の再構築処理方法を示す．

$\hat{\boldsymbol{M}}_{\alpha_1,\cdots,\alpha_D}$ は，以下の式 (5.6) に示す通り，再帰的に表すことが可能である．これは，$\prod_{m=k-1}^{D} \alpha_k \times \prod_{m=k-1}^{D} \alpha_k$ の正方行列で表すことができる．この時，対角成分は $\mathbf{0}$ となり，非対角成分は $\hat{\boldsymbol{M}}_{\alpha_k,\cdots,\alpha_D}$ となる．

$$\hat{\boldsymbol{M}}_{\alpha_{k-1},\cdots,\alpha_D} = \begin{pmatrix} \mathbf{0} & \hat{\boldsymbol{M}}_{\alpha_k,\cdots,\alpha_D} & \cdots & \hat{\boldsymbol{M}}_{\alpha_k,\cdots,\alpha_D} \\ \hat{\boldsymbol{M}}_{\alpha_k,\cdots,\alpha_D} & \mathbf{0} & \cdots & \hat{\boldsymbol{M}}_{\alpha_k,\cdots,\alpha_D} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\boldsymbol{M}}_{\alpha_k,\cdots,\alpha_D} & \cdots & \cdots & \mathbf{0} \end{pmatrix} \tag{5.6}$$

また，$\hat{\boldsymbol{M}}_{\alpha_{D-1},\alpha_D}$ は式 (5.7) のように表すことができる．この行列 $\hat{\boldsymbol{M}}_{\alpha_{D-1},\alpha_D}$ の対角成分は $\mathbf{0}$，非対角成分は $\hat{\boldsymbol{N}}_{\alpha_1,\cdots,\alpha_D}$ となる．

$$\hat{\boldsymbol{M}}_{\alpha_{D-1},\alpha_D} = \begin{pmatrix} \mathbf{0} & \hat{\boldsymbol{N}}_{\alpha_1,\cdots,\alpha_D} & \cdots & \hat{\boldsymbol{N}}_{\alpha_1,\cdots,\alpha_D} \\ \hat{\boldsymbol{N}}_{\alpha_1,\cdots,\alpha_D} & \mathbf{0} & \cdots & \hat{\boldsymbol{N}}_{\alpha_1,\cdots,\alpha_D} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\boldsymbol{N}}_{\alpha_1,\cdots,\alpha_D} & \cdots & \cdots & \mathbf{0} \end{pmatrix} \tag{5.7}$$

ここで，$\hat{\boldsymbol{N}}_{\alpha_1,\cdots,\alpha_D}$ は式 (5.8) に示すように 2×2 の正方行列である．

$$\hat{\boldsymbol{N}}_{\alpha_1,\cdots,\alpha_D} = \frac{1}{\alpha_1 - 1} \cdots \frac{1}{\alpha_{D-1} - 1} \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix} \tag{5.8}$$

ここで，常にパラメータ p の条件より，$p \neq 0.5$ を満たす．このため，式 (5.8) より逆行列 $\hat{\boldsymbol{N}}^{-1}_{\alpha_1,\cdots,\alpha_D}$ は常に存在することが明らかである．この逆行列 $\hat{\boldsymbol{N}}^{-1}_{\alpha_1,\cdots,\alpha_D}$ を用いて，データ再構築に必要な $\hat{\boldsymbol{M}}^{-1}_{\alpha_1,\cdots,\alpha_D}$ を導く．

式 (5.6)，(5.7) では，前述の通り $\hat{\boldsymbol{M}}_{\alpha_1,\cdots,\alpha_D}$ と $\hat{\boldsymbol{M}}_{\alpha_{D-1},\alpha_D}$ の対角成分は全て $\mathbf{0}$ であり，非対角成分はそれぞれ $\hat{\boldsymbol{M}}_{\alpha_k,\cdots,\alpha_D}$ と $\hat{\boldsymbol{N}}_{\alpha_1,\cdots,\alpha_D}$ である．これらの特徴と，$\hat{\boldsymbol{N}}^{-1}_{\alpha_1,\cdots,\alpha_D}$ から，式 (5.9) のように逆行列 $\hat{\boldsymbol{M}}^{-1}_{\alpha_{D-1},\alpha_D}$ を求めることができる．

$$\hat{\boldsymbol{M}}^{-1}_{\alpha_{D-1},\alpha_D} = \frac{1}{\alpha_{D-1} - 1} \cdot \tag{5.9}$$
$$\begin{pmatrix} -(\alpha_{D-1}-2)\hat{\boldsymbol{N}}^{-1}_{\alpha_1,\cdots,\alpha_D} & \hat{\boldsymbol{N}}^{-1}_{\alpha_1,\cdots,\alpha_D} & \cdots & \hat{\boldsymbol{N}}^{-1}_{\alpha_1,\cdots,\alpha_D} \\ \hat{\boldsymbol{N}}^{-1}_{\alpha_1,\cdots,\alpha_D} & -(\alpha_{D-1}-2)\hat{\boldsymbol{N}}^{-1}_{\alpha_1,\cdots,\alpha_D} & \cdots & \hat{\boldsymbol{N}}^{-1}_{\alpha_1,\cdots,\alpha_D} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\boldsymbol{N}}^{-1}_{\alpha_1,\cdots,\alpha_D} & \cdots & \cdots & -(\alpha_{D-1}-2)\hat{\boldsymbol{N}}^{-1}_{\alpha_1,\cdots,\alpha_D} \end{pmatrix}$$

また，$\hat{\boldsymbol{M}}^{-1}_{\alpha_{k-1},\cdots,\alpha_D}$ は式 (5.10) のように表すことができる．

$$\hat{\boldsymbol{M}}^{-1}_{\alpha_{k-1},\cdots,\alpha_D} = \frac{1}{\alpha_{k-1} - 1} \cdot \tag{5.10}$$
$$\begin{pmatrix} -(\alpha_{k-1}-2)\hat{\boldsymbol{M}}^{-1}_{\alpha_k,\cdots,\alpha_D} & \hat{\boldsymbol{M}}^{-1}_{\alpha_k,\cdots,\alpha_D} & \cdots & \hat{\boldsymbol{M}}^{-1}_{\alpha_k,\cdots,\alpha_D} \\ \hat{\boldsymbol{M}}^{-1}_{\alpha_k,\cdots,\alpha_D} & -(\alpha_{k-1}-2)\hat{\boldsymbol{M}}^{-1}_{\alpha_k,\cdots,\alpha_D} & \cdots & \hat{\boldsymbol{M}}^{-1}_{\alpha_k,\cdots,\alpha_D} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\boldsymbol{M}}^{-1}_{\alpha_k,\cdots,\alpha_D} & \cdots & \cdots & -(\alpha_{k-1}-2)\hat{\boldsymbol{M}}^{-1}_{\alpha_k,\cdots,\alpha_D} \end{pmatrix}$$

式 (5.9)，(5.10) を用いて，再帰的に $\hat{\boldsymbol{M}}^{-1}_{\alpha_1,\cdots,\alpha_D}$ を導くことができる．この $\hat{\boldsymbol{M}}^{-1}_{\alpha_1,\cdots,\alpha_D}$ とプライバシー保護処理済データの統計的なデータ分布 $\hat{\boldsymbol{Y}}$ を用いることによって，統計的なデータ分布 $\hat{\boldsymbol{A}}$ を導くことができる．この統計的なデータ分布 $\hat{\boldsymbol{A}}$ は，理論的には常にオリジナルデータの統計的なデータ分布 $\hat{\boldsymbol{X}}$ と等しいデータ分布となる．

ここで本提案手法のサーバ側の処理の具体例として，4×2 の 2 次元データを用いて，データ再構築処理の様子を示す．この時，モバイルデバイス上のプライバシー保護処理を示す $\hat{\boldsymbol{M}}_{4,2}$ と $\boldsymbol{N}_{4,2}$ は，次のように表すことができる．

$$\hat{\boldsymbol{M}}_{4,2} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{N_{4,2}} & \boldsymbol{N_{4,2}} & \boldsymbol{N_{4,2}} \\ \boldsymbol{N_{4,2}} & \boldsymbol{0} & \boldsymbol{N_{4,2}} & \boldsymbol{N_{4,2}} \\ \boldsymbol{N_{4,2}} & \boldsymbol{N_{4,2}} & \boldsymbol{0} & \boldsymbol{N_{4,2}} \\ \boldsymbol{N_{4,2}} & \boldsymbol{N_{4,2}} & \boldsymbol{N_{4,2}} & \boldsymbol{0} \end{pmatrix} \tag{5.11}$$

$$\boldsymbol{N_{4,2}} = \frac{1}{3} \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix} \tag{5.12}$$

式 (5.11) と (5.12) から，$\hat{\boldsymbol{M}}_{4,2}$ の逆行列 $\hat{\boldsymbol{M}}_{4,2}^{-1}$ は式 (5.13) のようになる．また，逆行列 $\boldsymbol{N_{4,2}^{-1}}$ は式 (5.14) のようになる．

$$\hat{\boldsymbol{M}}_{4,2}^{-1} = \frac{1}{3} \begin{pmatrix} \boldsymbol{-2N_{4,2}^{-1}} & \boldsymbol{N_{4,2}^{-1}} & \boldsymbol{N_{4,2}^{-1}} & \boldsymbol{N_{4,2}^{-1}} \\ \boldsymbol{N_{4,2}^{-1}} & \boldsymbol{-2N_{4,2}^{-1}} & \boldsymbol{N_{4,2}^{-1}} & \boldsymbol{N_{4,2}^{-1}} \\ \boldsymbol{N_{4,2}^{-1}} & \boldsymbol{N_{4,2}^{-1}} & \boldsymbol{-2N_{4,2}^{-1}} & \boldsymbol{N_{4,2}^{-1}} \\ \boldsymbol{N_{4,2}^{-1}} & \boldsymbol{N_{4,2}^{-1}} & \boldsymbol{N_{4,2}^{-1}} & \boldsymbol{-2N_{4,2}^{-1}} \end{pmatrix} \tag{5.13}$$

$$\boldsymbol{N_{4,2}^{-1}} = \frac{3}{2p-1} \begin{pmatrix} p & p-1 \\ p-1 & p \end{pmatrix} \tag{5.14}$$

つまり，4×2 の 2 次元データに対して MNS-RRT アルゴリズムを用いた場合，式 (5.13)，(5.14) を用いて，データの再構築処理を行うことができる．この再構築処理は，一見すると計算量が多い処理に見える．しかしながら，式 (5.13) に示す $\hat{\boldsymbol{M}}_{4,2}^{-1}$ が有する成分は 4 つのみであるため，常に同様の計算を繰り返すのみである．このため，再構築処理のための計算量は低く抑えることができる．

以上で述べた処理が，提案手法 MNS-RRT アルゴリズムにおける再構築処理である．これらの再構築処理は行列を用いて表しているが，再帰的な処理であり，また各行列の要素が数種類に限られるため，非常に簡潔な計算で構成されている．また，扱う多次元データの有するフィールド情報の数及びデータ区分，パラメータ p が不変である限り，サーバ上の処理は常に同様の計算を繰り返すのみであり，再構築処理のための行列 $\hat{\boldsymbol{M}}_{\alpha_1,\cdots,\alpha_D}^{-1}$ は不変である．このため，再構築処理に必要な行列演算を行うのは，サービス開始時またはパラメータ p の変更時のみであり，計算量は低く抑えてあると言える．

5.2.2 評価

本項では，プライバシー保護の観点とデータの信頼性の観点から提案手法 MNS-RRT アルゴリズムの評価を行う．

まず，本評価に用いるジェンソンシャノン情報量 JSD(Jensen-Shannon Divergence)[93] について述べる．本評価指標は，式 (3.1) で示したカルバックライブラー情報量 D_{KL} を基にした指標であり，式 (5.15)，(5.16) に示すように用いることができる．この時，ジェンソンシャノン情報量 JSD[93] は 2 つの離散確率分布 X と A の一致性を示し，カルバックライブラー情報量 D_{KL} に対称性を持たせ，改良した指標である．

$$JSD(X||A) = \frac{1}{2}D_{KL}(X||M) + \frac{1}{2}D_{KL}(A||M) \tag{5.15}$$

$$M = \frac{1}{2}(X + A) \tag{5.16}$$

また，式 (4.10) に示すデータ整合率 $\mathbf{C}$ も共に用いてデータ信頼性を評価することによって，より質の高い評価を行うことを目指す．また，プライバシーの評価指標としては，式 (4.8)，(4.9) に示す $Privacy$ を用いる．この $Privacy$ は，最大事後推定確率を基にした指標である．

以下では，二者択一型選択肢を含む多次元データを用いる参加型センシングを具体的に想定し，シミュレーション評価を行う．また，より一般的な有効性を示すために，様々な多次元データを用いた評価を行う．

天候観測用参加型センシングへの適用

本シミュレーション実験では，天候観測のためのモバイル参加型センシングを想定する．既存のイベント検知のための参加型センシング [91, 92] のアプリケーションと同様に，ユーザは位置情報に紐づいた形で，Yes/No 形式の回答をサーバに報告する．この時，イベントの存在自体が位置情報へと結びつく可能性があるため，プライバシー保護処理を施す必要がある．本評価ではまず，60,000 人のユーザがセンシングに参加すると想定する．また，6 つのエリアを調査することを前提とし，各エリアから等しい数のユーザが参加するものとする．また，人為的な入力操作のミスを想定して，各エリアからの 95% のデータは正確な情報であり，残りの 5% は誤った情報であるとする．エリア 1，2，4，5，6 では降雨しておらず，エリア 3 のみで降雨している時にユーザが入力するデータを Original Data，MNS-RRT アルゴリズムによってプライバシー保護処理されたデータを Disguised Data，サーバ上で統計処理によって再構築されたデータ分布を Reconstructed Data として，図 5.2 に示す．この時，パラメータ p を 0.2 と設定する．図 5.2 の結果より明らかなように，再構築したデータ分布とオリジナルデータのデータ分布は類似したものとなり，ジェンソンシャノン情報量 JSD の値も 0.003101 となる．また，パラメータ p の値を 0.2，0.25，0.3 とした時のデータ信頼性に関する評価を表 5.2 に示す．

既存の Perturbation 技術では，二者択一型選択肢を含む多次元データを取り扱うことが出来なかったが，本提案手法では Randomized Response と Negative Surveys を組み合わせることによって，これを可能とした．また，ユーザが安心してセンシングに参加できる環境を構築することによって，イベント検知の様な，ユーザからの積極的な貢献が必要なアプリケーションにおいても参加型センシングシステムを運用することが可能となる．

図 5.2: 天候観測用参加型センシングのデータに対して MNS-RRT アルゴリズムを適用した時の様子．95% の入力データが正確な情報であると想定した．またパラメータ p の値は 0.2 とした．

震災被害地域の調査への適用

現段階で普及しているモバイルデバイスには，放射線レベルを測定するためのセンサはほとんど搭載されていない．しかしながら，米国政府をはじめとして，様々な機関がモバイルデバイスを用いた放射線モニタリングの活用を進めている [5]．既存の固定センサを用いた放射線量モニタリング用センサネットワークは，センサ設置・運用が困難である都市地域や放射線量の高い地域で活用することが難しく，また固定センサ導入のための初期コストも莫大なものとなる．また，放射線量は不動産価値や企業戦略などに大きな影響を持つため，第三者によって故意に値を改竄される可能性が大いに考えられる [5]．このため，一般ユーザの持つモバイルデバイスを用いた参加型センシングを放射線モニタリングに活用し，正確かつ信頼のできるデータを取得することが望ましい．

本評価実験では，実際の震災による被害状況に関するデータ [94, 95] を用いて，震災被害地域の調査のための参加型センシングを想定したシミュレーションを行う．このセンシングシステムでは，位置情報と共に，放射線レベルと建物損壊レベルを報告するものと想定する．放射線レベルは，そのユーザの健康状況に直結する重要なセンシティブ情報であり，守るべきプライバシー情報に該当

する．また，建物損壊レベルは地域全体で見た時に，空間的な相関性があるため，ユーザの位置情報及び重要なセンシティブ情報である放射線レベルの推定に結びつく可能性がある．このため，本シミュレーション実験では守るべきプライバシー情報として扱う．

本システムでは，建物損壊レベルを No damage，Light damage，Heavy damage の 3 段階で評価することとし，また放射線レベルは，ある閾値を超えている値か (*above*)，閾値より低い値か (*below*) という 2 段階で評価するものとする．また，エリアごとのユーザ，つまりセンシング参加者は震災被害地域の人口を参考に [94]，それぞれ $21,580$，$8,448$，$15,868$，$11,408$ とする．また，プライバシー保護処理に用いるパラメータ p は 0.25 と設定する．

この時，ユーザがセンシングによって取得するデータを Original Data，MNS-RRT アルゴリズムによってプライバシー保護処理されたデータを Disguised Data，サーバ上で統計処理によって再構築されたデータ分布を Reconstructed Data とした様子を図 5.3 に示す．データ信頼性を示す評価指標であるジェンソンシャノン情報量 JSD 及びデータ整合率 $\mathbf{C}$ はそれぞれ，0.004784，93.73% となる. また，同様の処理を 10 回行った時のジェンソンシャノン情報量 JSD の平均値は 0.012401 であった.

表 5.2: 図 5.2 のシミュレーションにおいて，パラメータ p の値を変えた時の結果．10 回の処理を行い，その平均値を示した.

Data sets: 6×2, $D = 2$ ($p = 0.2$)		
Participants	JSD	Concordance rate $\mathbf{C}(\%)$
$15,000$	0.010355	94.55
$20,000$	0.009031	94.42
$30,000$	0.004776	95.62
$60,000$	0.003101	96.77
Data sets: 6×2, $D = 2$ ($p = 0.25$)		
Participants	JSD	Concordance rate $\mathbf{C}(\%)$
$15,000$	0.017538	89.75
$20,000$	0.018541	90.12
$30,000$	0.012544	90.62
$60,000$	0.008741	93.17
Data sets: 6×2, $D = 2$ ($p = 0.3$)		
Participants	JSD	Concordance rate $\mathbf{C}(\%)$
$15,000$	0.025810	83.54
$20,000$	0.023899	85.20
$30,000$	0.023447	89.75
$60,000$	0.012241	91.19

図 5.3: 4 つのエリアに紐づく形で，3 段階の建物損壊レベル，2 段階の放射線レベルを参加型センシングで取得した時の様子.

様々なデータ構造における評価

上記のシミュレーション実験では 6×2 の 2 次元データ及び $4 \times 3 \times 2$ の 3 次元データを扱った．本シミュレーションでは，提案手法 MNS-RRT アルゴリズムがより様々な多次元データに適用可能であることを示す．このため，本シミュレーション実験では，6×2，9×2 の 2 次元データ，$4 \times 3 \times 2$ の 3 次元データ，$3 \times 3 \times 3 \times 2$ の 4 次元データに対して MNS-RRT アルゴリズムによるプライバシー保護処理を施し，プライバシー保護とデータ信頼性の両側面から評価を行う．

センシングに参加する人口数は 15,000，20,000，30,000，60,000 と変化させ，その際のデータ信頼性をジェンソンシャノン情報量 JSD を用いて評価したものを図 5.4 に示す．また，それぞれの多次元データに対して，パラメータ p の値を 0.2，0.25，0.3 と変化させた評価を示す．いずれの多次元データにおいても，パラメータ p の値が小さいほど，データ信頼性が高くなることが明らかである．また，図 5.4 から明らかなように，MNS-RRT アルゴリズムでは，取得データの次元の数，各次元の持つ選択肢の数に関係無く，正確に統計的なデータ分布を取得することが可能である．本提案手法では，サーバ上の再構築処理において統計処理を用いているため，参加するユーザ

数が減少した時にはデータ信頼性が損なわれていることが明らかである．

また，式 (4.8)，(4.9) を用いてプライバシーに関する評価を行った結果を図 5.5 に示す．このプライバシーの評価では，悪意ある攻撃者がプライバシー保護処理済データを盗聴によって取得した場合を想定するが，たとえパラメータ p の値が攻撃者に漏洩したとしても，攻撃者はオリジナルデータを推定できない．このため本評価では，よりプライバシー情報の漏洩の危険が大きい場合を想定して，悪意ある攻撃者がパラメータ p の値を知っているものとする．図 5.5 に示す結果から明らかなように，パラメータ p が 0.5 に近付くほど，プライバシー保護処理済データの持つ情報量は小さくなるため，攻撃者がオリジナルデータを推定できる可能性も小さくなる．つまり本提案手法では，p が 0.5 に近い値であるほど，より強固にプライバシー保護がなされていると言える．

以上で述べたことから，このパラメータ p の値は，プライバシーレベルとデータ信頼性のバランスを決定する重要なパラメータであることがわかる．パラメータ p の値が 0.5 から遠ざかるほど，再構築される統計的なデータ分布のデータ信頼性が高くなる．一方で，パラメータ p の値が 0.5 に近付くほど，ユーザから提供される情報量が小さくなるため，より高いプライバシーレベルを保つことが可能となる．また，データ信頼性を決定する要因として，センシングに参加するユーザの数 N が重要であるといえ，ユーザの数 N が大きいほど，統計としてのデータ信頼性が向上する．これは，大数の法則からも明らかである．

これらを考慮すると，参加型センシングの本来の目的である統計的なデータ分布を取得するために，ユーザの数 N が小さな値である時は，パラメータ p の値を小さなものとし，データ信頼性を確保することが望ましい．また，センシングに参加するユーザの数 N が増えるにつれて，p の値を 0.5 に近づける．これによって，参加型センシングのサービスが実社会で大きな存在となり，社会的責任を担うに連れて，強力なプライバシー保護を実現するシステムを実現することができる．

5.3 本章のまとめ

本章では，まず二者択一型選択肢を含む多次元データを扱う参加型センシングの例として，プロファイル情報を収集する参加型センシングやイベント検知のための参加型センシングを紹介した．ユーザ自身の情報を示すプロファイル情報を扱う際や，放射線量の閾値判定などの際には，プライバシー情報保護の観点から情報の取り扱いが非常に難しいことについて言及し，また既存の Perturbation 技術では二者択一型選択肢を含む多次元データを適切に扱うことができないことを指摘し，この二者択一型選択肢を含む多次元データに対するプライバシー保護手法の開発の重要性を明らかにした．

また，プロファイル情報を収集する参加型センシングやイベント検知のための参加型センシングで扱う多次元データの特徴を考慮した上で，既存の Perturbation 技術を拡張し，プライバシー保護手法 MNS-RRT アルゴリズムを提案した．本提案手法 MNS-RRT アルゴリズムは，Randomized Resonse と Negative Surveys を組み合わせた手法であり，計算資源の乏しいモバイルデバイス上でプライバシー保護を行うことも可能な手法である．本手法を用いることによって，二者択一型選択肢を含む多次元データのプライバシー保護を実現でき，一般のユーザ及びアプリケーション提供者が安心して参加型センシングでプロファイル情報を収集することのできる土壌を確立したと言える．

図 5.4: 6×2 の 2 次元データ，9×2 の 2 次元データ，$4 \times 3 \times 2$ の 3 次元データ，$3 \times 3 \times 3 \times 2$ の 4 次元データに対して，それぞれ MNS-RRT アルゴリズムを施した時のデータ信頼性に関する結果．5 回の処理を施し，その平均値を記す．本提案手法では，統計処理を用いるため，取得データ数が少ない場合にはデータ信頼性は低下する．

図 5.5: 6×2 の 2 次元データ，9×2 の 2 次元データ，$4 \times 3 \times 2$ の 3 次元データ，$3 \times 3 \times 3 \times 2$ の 4 次元データに対して，それぞれ MNS-RRT アルゴリズムを施した時のプライバシー保護に関する評価結果．パラメータ p の値が 0.5 に近付くほど，より強固にプライバシー情報を保護することができる．

第6章　医療調査用参加型センシングとプライバシー保護

既存の Perturbation 技術では，取得したセンサデータに対して逐一プライバシー保護処理を施し，長期間に渡る盗聴による蓄積攻撃からのプライバシー情報の侵害の危険性を考慮していなかった．このため，第 2 章で指摘したように，悪意ある攻撃者による蓄積攻撃は，参加型センシング実現のために大きな脅威である．特に，医療調査やヘルスケア，公衆衛生の分野では，時間依存性の低い，準静的な情報 (quasi-static data) を扱うことが多く，常に取得データの値が変化するわけではない [4, 98]．このため，長期間のデータを取得し続ける蓄積攻撃によるプライバシー情報の侵害は，医療用参加型センシングの実現にとっては大きな課題となる．

この蓄積攻撃からのプライバシー保護を実現するために，本章ではプライバシー保護手法 Duplicated Perturbation を提案する．また本提案手法では，センシングによって得たオリジナルデータとプライバシー保護処理済データの関係に注目することによって，中央サーバ上におけるデータ再構築処理に必要となる計算量を削減する．最後に，本提案手法をプライバシーとデータ信頼性の観点から評価を行う．

6.1　医療調査用参加型センシングの概要

ユーザの日常生活と密接な関わりを持つスマートフォンやウェアラブルデバイスの普及に伴って，これらモバイルデバイスを用いた医療 m-Health [23] の活用が期待されている．このモバイルデバイスを用いた医療調査用参加型センシングの概要図を図 6.1 に示す．一般ユーザは，モバイルデバイスを用いてライフログデータやプロファイル情報をアプリケーション提供者に報告し，これによってアプリケーション提供者は全体の統計的なデータ分布を明らかにすることができる．アプリケーション提供者は，収集した統計的なデータ分布を用いて，疾患と環境的な要因の関係性などを明らかにすることができる．但し，モバイルデバイスに搭載されたセンサで取得できるセンサデータは，ライフログデータやプロファイル情報も含んだものとなるため，ユーザの守るべきプライバシー情報に該当すると言える．このため，ユーザの所有するモバイルデバイス上でプライバシー保護処理を行い，プライバシー情報をユーザ自身以外が取得できないシステムとすることが望ましい．

m-Health は，特に慢性疾患や精神疾患の原因解明に役立てられると期待できる．本来，慢性疾患や精神疾患の診療・調査は病院や医療施設では難しく，このため米国 国立衛生研究所 (National Institutes of Health) の Genes and Environment Initiative を中心として，m-Health を活用して日常生活における人間の活動量・食習慣と疾患の関連を研究するプロジェクトが行われている [4]．また，伝染病の伝搬モデルの構築を目的として，公衆衛生の分野でも m-Health の利用が期待されている [23]．

これらの医療調査用参加型センシングでは，普遍的なセンシティブ情報である医療情報 (medical information) を活用するため，プライバシー情報の適切な利用が非常に重要となる．また，医療調査用参加型センシングでは，疾患と環境的要因・遺伝的要因の関係を明らかとすることを目的としているため，個々人のプライバシー情報ではなく，全体の統計的なデータ分布を明らかにすることを目指す．この医療情報の特徴として，時間依存性の低い準静的データ (quasi-static data) である

図 6.1: モバイルデバイスを用いた医療調査用参加型センシングの概要図 [98].

ことが挙げられる [4]．センシングに参加するユーザの疾患の状態や健康状態は，都市環境情報の特徴とは大きく異なり，長い期間を経て推移するものである．また，ユーザの日常生活は，昼に活動し，夜に睡眠を取るなど，ほとんどの場合は周期的である．

このため，医療調査用参加型センシングにとって，長期間に渡る盗聴によってデータ収集を行う蓄積攻撃は大きな脅威であり，公衆衛生・医療調査の分野で参加型センシングを有効に活用するためにも，プライバシー情報の保護手法の開発は非常に重要となる．

以下では，悪意ある攻撃者の長期間に渡る盗聴によるプライバシー情報の侵害を防ぐための Duplicated Perturbation を提案する．

6.2 医療調査用参加型センシングのためのプライバシー保護手法

第 2 章では，既存の Perturbation 技術を参加型センシングに活用するためには，悪意ある攻撃者からの蓄積攻撃が大きな課題であることを指摘した．また，Multidimensional Negative Surveys や第 4 章で提案した Selective Negative Surveys，第 5 章で提案した MNS-RRT アルゴリズムでは，オリジナルデータに応じて，プライバシー保護処理済データの取りうるカテゴリ値を限定する．このため，特に時間依存性の低い準静的データ (quasi-static data) を扱う際には，プライバシー保護処理済データからオリジナルデータの推定が容易になるという脆弱性がある．

そこで本提案手法では，2 段階のプライバシー保護処理を施すことによって，プライバシー保護

処理済データとオリジナルデータの直接的な関係を解消し，蓄積攻撃からのプライバシー情報の保護を実現する．以下では，提案手法 Duplicated Perturbation を一次元データに適用する際の処理と，多次元データに適用する際の処理に分けて示す．

6.2.1 一次元データに対する提案手法: Duplicated Perturbation

本提案手法では，センシングを行う各モバイルデバイス上で 2 段階のプライバシー保護処理を行い，このプライバシー保護処理済データを中央サーバに伝送する．中央サーバでは，このプライバシー保護処理済データを収集した後に，統計的なデータ分布の再構築処理を行う．但し，本提案手法は計算資源・電源に限りのあるモバイルデバイスを用いた参加型センシングに適用することを想定しているため，モバイルデバイス上の処理に必要となる計算量は低く抑える設計とする．

まず，一次元データに対する Duplicated Perturbation のプロトコルを図 6.2 に示す．データ総数を N とし，選択肢の数を α，あるセンサデータを x，一次摂動化データを y，二次摂動化データを z とする．これら x，y，z はカテゴリ値で表すものとする．また，パラメータ q と離散分布 $\Phi = \{\phi_1, \phi_2, \cdots, \phi_\alpha\}$ を設定する．パラメータ q の値及び離散分布 Φ はユーザのモバイルデバイス及びアプリケーション提供者が共有するものとする．但し，離散分布 Φ は式 (6.1) を満たすものとする．

$$\sum_{k=1}^{\alpha} \phi_k = N \tag{6.1}$$

以下に，モバイルデバイス上におけるプライバシー保護処理と中央サーバにおける再構築処理に分けて提案手法である Duplicated Perturbation を一次元データに用いる時の詳細を記す．

モバイルデバイス上の処理

本提案手法では，モバイルデバイスに搭載したセンサからのデータやユーザの入力情報を取得し，2 段階のプライバシー保護処理を施した後に，プライバシー保護処理済データを中央サーバに報告する．このモバイルデバイスにおける処理を 3 つに分けて記す．

1. 各種センサやユーザの入力によって，オリジナルデータ x を取得する．

2. オリジナルデータ x から，一次摂動化データ y を生成する．空間疑似データの生成は，パラメータ q を用いた離散分布 Γ に沿うものとする．この Γ を式 (6.2) に示す．$2 \leq x \leq \alpha - 1$ を満たすとき，$1 - 2q$ の確率で $y = x$ を満たす一次摂動化データ y が生成される．また，$1 = x$ と $\alpha = x$ の時，$1 - q$ の確率で $y = x$ となる．但し，常に $1 \leq y \leq \alpha$ を満たす．

$$\begin{cases} \Gamma_{2 \leq x \leq \alpha - 1}(x' = x) = 1 - 2q \\ \Gamma_{x=1,\alpha}(x' = x) = 1 - q \\ \Gamma(x' = x \pm 1) = q \end{cases} \tag{6.2}$$

つまり，一次摂動化データ y の値は，オリジナルデータ x の値と等しくなるか，隣接するカテゴリ値となる．

3. 次に，二次摂動化データは常に $y \neq z$ を満たすよう選択する．この時，式 (6.1) に示す離散分布 Φ に従って処理を行う．

   つまり，一次摂動化データの値 $y = k$ であった時，$\{\phi_1, \cdots, \phi_{k-1}, 0, \phi_{k+1}, \cdots, \phi_\alpha\}$ の分布に従ってプライバシー保護処理を行い，二次摂動化データ z を決定する．モバイルデバイスは，この二次摂動化データ z をアプリケーション提供者に報告する．

この二次摂動化データは，オリジナルデータと一定確率で同じ値となる．このため，悪意ある攻撃者が盗聴によって二次摂動化データを取得しても，容易にオリジナルデータを推定することはできない．また，一次摂動化データとオリジナルデータは隣接または一致したカテゴリ値となる．これは，疾患レベルを表すフィールド情報などでは，隣接したカテゴリ同士が強い相関をもつ特性を利用し，アプリケーション提供者側での統計処理における誤差を小さく抑えるためである．

中央サーバ上の処理

中央サーバでは，各モバイルデバイス上でプライバシー保護処理によって生成したデータ，二次摂動化データを収集する．

二次摂動化データを集計した統計的なデータ分布を $\boldsymbol{Z} = \{Z_1, Z_2, Z_3, \cdots, Z_{\alpha-1}, Z_\alpha\}$ とする．ここで，Z_k とは二次摂動化データの統計における，カテゴリ k 二該当するデータの数を指す．また，同様に再構築処理によって得た一次摂動化データの統計的なデータ分布を $\boldsymbol{Y} = \{Y_1, Y_2, Y_3, \cdots, Y_{\alpha-1}, Y_\alpha\}$，再構築処理によって得たオリジナルのセンサデータの統計的なデータ分布を $\boldsymbol{A} = \{A_1, A_2, A_3, \cdots, A_{\alpha-1}, A_\alpha\}$ とする．

図 6.2: 一次元データに対する Duplicated Perturbation のプロトコル.

この統計的なデータ分布の再構築処理を以下に 3 段階に分けて記す.

1. 二次摂動化データを各カテゴリごとに集計し，統計的なデータ分布 $\boldsymbol{Z}$ を構築する.
2. 二次摂動化データのデータ分布 $\boldsymbol{Z}$ から一次摂動化データの統計的なデータ分布 $\boldsymbol{Y}$ を明らかにする．このためにまず，$\boldsymbol{Z}$ と $\boldsymbol{Y}$ の関係を以下の行列で表す．この時，行列 $\boldsymbol{L}$ は二次摂動化データのデータ分布 $\boldsymbol{Z}$ と一次摂動化データの統計的なデータ分布 $\boldsymbol{Y}$ の関係を表す.

$$\begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ \vdots \\ Z_\alpha \end{pmatrix} = \boldsymbol{L} \cdot \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_\alpha \end{pmatrix} \tag{6.3}$$

この式 (6.3) に表す関係は，各モバイルデバイス上の処理を巨視的に捉えたものである．このため，行列 $\boldsymbol{L}$ はモバイルデバイス上のデータ処理によって表すことができる.

$$\begin{aligned}
\boldsymbol{L} &= \begin{pmatrix} 0 & \frac{\phi_1}{N-\phi_2} & \frac{\phi_1}{N-\phi_3} & \cdots & \frac{\phi_1}{N-\phi_\alpha} \\ \frac{\phi_2}{N-\phi_1} & 0 & \frac{\phi_2}{N-\phi_3} & \cdots & \frac{\phi_2}{N-\phi_\alpha} \\ \frac{\phi_3}{N-\phi_1} & \frac{\phi_3}{N-\phi_2} & 0 & \cdots & \frac{\phi_3}{N-\phi_\alpha} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\phi_\alpha}{N-\phi_1} & \frac{\phi_\alpha}{N-\phi_2} & \cdots & \frac{\phi_\alpha}{N-\phi_{\alpha-1}} & 0 \end{pmatrix} \\
&= \begin{pmatrix} \phi_1 & 0 & \dots & 0 \\ 0 & \phi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \phi_\alpha \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & \dots & 1 \\ 1 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 0 \end{pmatrix} \\
&\quad \cdot \begin{pmatrix} \frac{1}{N-\phi_1} & 0 & \dots & 0 \\ 0 & \frac{1}{N-\phi_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{N-\phi_\alpha} \end{pmatrix}
\end{aligned} \tag{6.4}$$

この時，$\boldsymbol{L}$ を構成する行列の特徴から，常に逆行列 $\boldsymbol{L}^{-1}$ が存在する．逆行列 $\boldsymbol{L}^{-1}$ は以下の様に表すことができる.

$$\begin{aligned}
\boldsymbol{L}^{-1} &= \frac{1}{\alpha-1} \cdot \begin{pmatrix} N-\phi_1 & 0 & \dots & 0 \\ 0 & N-\phi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & N-\phi_\alpha \end{pmatrix} \\
&\cdot \begin{pmatrix} 2-\alpha & 1 & \dots & 1 \\ 1 & 2-\alpha & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 2-\alpha \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\phi_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\phi_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\phi_\alpha} \end{pmatrix}
\end{aligned} \tag{6.5}$$

逆行列 $\boldsymbol{L}^{-1}$ と式 (6.3) に示す二次摂動化データのデータ分布 $\boldsymbol{Z}$ と一次摂動化データのデータ分布 $\boldsymbol{Y}$ の関係式から，一次摂動化データのデータ分布 $\boldsymbol{Y}$ を明らかにすることができる．この数式処理を以下の式 (6.6) に示す．

$$\boldsymbol{Y} = \boldsymbol{L}^{-1} \cdot \boldsymbol{Z} \tag{6.6}$$

逆行列 $\boldsymbol{L}^{-1}$ の要素は，カテゴリ数 α，離散分布 $\Phi = \{\phi_1,\ \phi_2,\ \cdots,\ \phi_\alpha\}$，データ数 N で構成される．このため，取得できるデータ数が大きく変化しない限り逆行列 $\boldsymbol{L}^{-1}$ を計算する必要は無いと言える．

3. 一次摂動化データのデータ分布 $\boldsymbol{Y}$ に関する連立方程式を解くことによって，取得したセンサデータの統計的なデータ分布 $\boldsymbol{A}$ を明らかにすることができる．

$$\begin{cases} Y_1 = A_1 \times (1-p) + A_2 \times p \\ Y_2 = A_2 \times (1-2p) + A_1 \times p + A_3 \times p \\ \vdots \\ Y_k = A_k \times (1-2p) + A_{k-1} \times p + A_{k+1} \times p \\ \vdots \\ Y_\alpha = A_\alpha \times (1-p) + A_{\alpha-1} \times p \end{cases} \tag{6.7}$$

以上が一次元データに対する Duplicated Perturbation のプライバシー保護処理である．中央サーバを管理するアプリケーション提供者は，一般ユーザの各々のセンサデータの値を明らかにすることなく，統計的なデータを得ることができる．

6.2.2 多次元データに対する提案手法: Duplicated Perturbation

複数フィールド情報を有する多次元データに対して 2 段階のプライバシー保護処理を行う Duplicated Perturbation の手法を示す.

多次元データの次元を D，各次元の持つ選択肢の数を $\alpha_1, \alpha_2, \cdots, \alpha_D$ と定義する．また，センシングしたオリジナルデータを $\vec{x} = \langle x_1, x_2, \cdots, x_D \rangle$，一次摂動化データを $\vec{y} = \langle y_1, y_2, \cdots, y_D \rangle$，二次摂動化データを $\vec{z} = \langle z_1, z_2, \cdots, z_D \rangle$ と表すものとする．本提案手法では，この一次摂動化データを $\vec{y}$，また二次摂動化データを $\vec{z}$，と設定することによって，オリジナルデータ $\vec{x}$ と二次摂動化データ $\vec{z}$ の取りうる値が相反するという関係を解消する．また，これらを集計したものをそれぞれ $\boldsymbol{X}$，$\boldsymbol{Y}$，$\boldsymbol{Z}$ として表し，オリジナルデータ，一次摂動化データ，二次摂動化データの統計的なデータ分布を示すものとする．また，サーバ上の処理によって，再構築するオリジナルデータのデータ分布を $\boldsymbol{A}$ と表すものとする．

以下に，モバイルデバイス上におけるプライバシー保護処理と中央サーバにおける再構築処理に分けて提案手法 Duplicated Perturbation の詳細を記す.

モバイルデバイス上の処理

本提案手法では，モバイルデバイスに搭載したセンサからのデータやユーザの入力情報を取得し，2 段階のプライバシー保護処理を施した後に，プライバシー保護処理済データを中央サーバに報告する.

センサデバイスとなる各モバイルデバイスでは，まずオリジナルデータ $\vec{x}$ を取得する．次に，オリジナルデータ $\vec{x}$ は Multidimensional Negative Surveys と同様の処理によって，一次摂動化データ $\vec{y}$ となる．この時，一次摂動化データの次元 i の値 y_i は，オリジナルデータの次元 i の値 x_i とは異なるカテゴリの中から 1 つのカテゴリが等確率で選ばれる．つまり，常に $x_i \neq y_i$ の関係を満たす．次元 i のカテゴリ数が α_i である時，それぞれのカテゴリ値へと遷移する遷移確率は $\frac{1}{\alpha_i - 1}$ と表すことができる．更に 2 段階目の処理では，一次摂動化データ $\vec{y}$ に対して，Multidimensional Negative Surveys のモバイルデバイス上の処理と同様の処理を施すことによって，二次摂動化データ $\vec{z}$ を生成する.

本提案手法では，この二次摂動化データ $\vec{z}$ を，プライバシー保護処理済データとして，中央サーバに伝送する．ここで，$r^{(\vec{x})}(\vec{z})$ は，プライバシー保護処理によって，あるオリジナルデータ $\vec{x}$ がある二次摂動化データ $\vec{z}$ となる遷移確率を示すものと定義する.

本手法において，センシングによって取得したオリジナルデータの次元 i の値 x_i と，二次摂動化データの次元 i の値 z_i はある一定確率で同じ値となる．このため，時間依存性の低い準静的データ (quasi-static data) を扱う場合にも，悪意ある攻撃者が蓄積攻撃によってオリジナルデータを推定することはできない.

ここで，図 6.3 のように 6×4 の 2 次元データに対する処理を例にとり，オリジナルデータとして，$\langle$b，2$\rangle$ を取得した際の遷移確率を示す．但し，6×4 の 2 次元データは，選択肢 6 のフィールド情報と選択肢 4 のフィールド情報を持つ多次元データである．本提案手法で k 次元のデータを扱う際には，遷移確率は計 2^k 通りで表すことができる．つまり，図 6.3 の様な 2 次元データを扱う際には，遷移確率は 4 通りの値をとる．また $\frac{1}{15}$ の確率で，二次摂動化データ (Second-order perturbed data) はオリジナルデータ (Original data) と同じ値となる.

図 6.3: Duplicated Perturbation による処理を 6×4 の 2 次元データに施した時の様子．センシングしたオリジナルデータ (Original data) が $\langle$b, 2$\rangle$ であった時，灰色セルの中から 1 つのセルが等確率に，一次摂動化データ (First-order perturbed data) として選ばれる．また，同様の処理で一次摂動化データから二次摂動化データ (Second-order perturbed data) が選択される．

中央サーバ上の処理

中央サーバでは，各モバイルデバイス上でプライバシー保護処理によって生成したデータ，二次摂動化データを収集する．この二次摂動化データを用いて，オリジナルデータのデータ分布を再構築し，$\boldsymbol{A}$ を明らかにする．

各モバイルデバイス上の処理は，Multidimensional Negative Surveys と同一の処理を 2 回繰り返したものであるため，同様に Multidimensional Negative Surveys のデータ再構築処理を 2 回繰り返せすことによって，オリジナルデータのデータ分布を再構築することができる．つまり，式 (2.5)，(2.6) に示す再構築処理を 2 回繰り返すことによって，再構築データ分布 $\boldsymbol{A}$ を導くことができる．

しかしながら，この再構築処理は既存の Multidimensional Negative Surveys の 2 倍の計算量が必要となる．以下では，オリジナルデータと二次摂動化データの関係を利用して，再構築処理の計算量を削減する．

再構築処理のために必要となる計算量を削減するために，あるオリジナルデータ $\vec{x}$ がある二次摂動化データ $\vec{z}$ となる遷移確率 $r^{(\vec{x})}(\vec{z})$ を用いる．式 (2.5)，(2.6) の 2 回の繰り返し処理をまとめた処理を式 (6.8) と (6.9) に示す．

$$\forall \vec{x} \mid \boldsymbol{A}(\vec{x}) = \left(N + \sum_{k=1}^{D} (-1)^k \cdot \Delta(\vec{x}, k) \right) \times \left(\prod_{j=1}^{D} (2 - \alpha_j) \right) \tag{6.8}$$

$$\Delta(\vec{x}, k) = \sum_{\substack{d \in \\ B(\{1,2,3,\cdots,D\},k)}} \left(\left[\frac{\prod_{j \notin d} (\alpha_j - 1)}{\prod_{j=1}^{D} (\alpha_j - 1)} \right] \sum_{\substack{\vec{z}\, s.t. \\ z_i \in \vec{x} \\ \forall i \notin d}} \frac{1}{r^{(a,b,c)}(\vec{z})} \sum_{\substack{\vec{z}\, s.t. \\ z_i \in \vec{x} \\ \forall i \in d}} \boldsymbol{Z}(\vec{z}) \right) \tag{6.9}$$

具体例を示すために，$D = 3$ の多次元データを対象とした再構築処理を式 (6.10) に示す．ここで，$A(a, b, c)$ は，再構築処理されたデータ分布のうち，それぞれの次元においての値が a，b，c となるデータの総数を示すものとする．

$$
\begin{aligned}
A(a,b,c) =& (N - \frac{1}{(\alpha_2-1)(\alpha_3-1)} \sum_{\substack{\vec{z}\,s.t.\\ z_2=b,\\ z_3=c}} \frac{1}{r^{(a,b,c)}(\vec{z})} \sum_{\substack{\vec{z}\,s.t.\\ z_1=a}} Z(\vec{z}) \\
&- \frac{1}{(\alpha_3-1)(\alpha_1-1)} \sum_{\substack{\vec{z}\,s.t.\\ z_3=c, z_1=a}} \frac{1}{r^{(a,b,c)}(\vec{z})} \sum_{\substack{\vec{z}\,s.t.\\ z_2=b}} Z(\vec{z}) \\
&- \frac{1}{(\alpha_1-1)(\alpha_2-1)} \sum_{\substack{\vec{z}\,s.t.\\ z_1=a, z_2=b}} \frac{1}{r^{(a,b,c)}(\vec{z})} \sum_{\substack{\vec{z}\,s.t.\\ z_3=c}} Z(\vec{z}) \\
&+ \frac{1}{(\alpha_1-1)} \sum_{\substack{\vec{z}\,s.t.\\ z_1=a}} \frac{1}{r^{(a,b,c)}(\vec{z})} \sum_{\substack{\vec{z}\,s.t.\\ z_2=b, z_3=c}} Z(\vec{z}) \\
&+ \frac{1}{(\alpha_2-1)} \sum_{\substack{\vec{z}\,s.t.\\ z_2=b}} \frac{1}{r^{(a,b,c)}(\vec{z})} \sum_{\substack{\vec{z}\,s.t.\\ z_3=c, z_1=a}} Z(\vec{z}) \\
&+ \frac{1}{(\alpha_3-1)} \sum_{\substack{\vec{z}\,s.t.\\ z_3=c}} \frac{1}{r^{(a,b,c)}(\vec{z})} \sum_{\substack{\vec{z}\,s.t.\\ z_1=a, z_2=b}} Z(\vec{z}) \\
&- Z(a,b,c) \sum_{\substack{\vec{z}\,s.t.\\ z_1\neq a,\\ z_2\neq b,\\ z_3\neq c}} \frac{1}{r^{(a,b,c)}(\vec{z})}) \cdot (2-\alpha_1)(2-\alpha_2)(2-\alpha_3)
\end{aligned}
\tag{6.10}
$$

この合成した再構築処理を用いることによって，中央サーバはより低い計算量でデータ再構築処理を行うことができる．アプリケーションを提供する中央サーバには，計算資源・電源が豊富にあることが考えられるが，よりリアルタイムに近い形でセンシング結果を明らかにするためにも，この計算量の削減は非常に重要である．

本提案手法で示した Duplicated Perturbation を用いることによって，ユーザは盗聴者及びサーバ管理者に対してプライバシー情報を提供することなく，参加型センシングの本来の目的である統計的なデータ分布を取得することが可能となる．

6.2.3 評価

本項では，提案手法 Duplicated Perturbation を用いた時の，プライバシー保護とデータの信頼性に関する評価をシミュレーションを用いて行う．

まずプライバシーの保護に関する評価としては，式 (4.8)，(4.9) に示す $Privacy$ を用いる．この指標は，最大事後確率推定法を基にした評価指標であり，プライバシー保護処理済データとして伝送された二次摂動化データが悪意ある攻撃者の盗聴によって漏洩した場合に，オリジナルデータが推定される確率を示す．また，データ信頼性に関する評価のために，式 (5.15)，(5.16) に示すジェンソンシャノン情報量 JSD を用いる．

慢性疾患の調査システム

本評価実験では，m-Health による慢性疾患調査を想定したシミュレーション実験を行う [23],[5, 21]．アプリケーションの目的としては，ユーザの日常生活を調査することによって，日常的な生活習慣

と慢性疾患の関係を明らかにすることである．慢性疾患の調査には，長時間のモニタリングを行う必要があるため，原因が遺伝的であるのか日常生活の習慣に依るものであるのか，現状で明確に証明することは難しい [5].

本シミュレーションでは慢性疾患の中でも，代謝症候 (Metabolic syndrome) 及び糖尿病 (Diabetes mellitus) に焦点をあてる．まず，ユーザの健康状態を ‘代謝症候かつ糖尿病’(Metabolic syndrome & Diabetes mellitus)，‘代謝症候のみ’(only Metabolic syndrome)，‘糖尿病のみ’(only Diabetes mellitus)，‘健康’(No symptoms) の 4 つのカテゴリとして扱う．また，三軸加速度センサから一日の運動量を算出し，‘運動不足’(Non-active)，‘普通’(Moderate)，‘充分な運動量’(Active) の 3 レベルに振り分ける．つまり，ここでは 4×3 の 2 次元データを扱う．ここで，参加するユーザの数 N は $10,000$ 人とし，それぞれの健康状態のユーザが $2,500$ 人参加すると想定する．ここで，糖尿病と日常の運動量の関係は薄く，また ‘代謝症候かつ糖尿’ の患者は ‘運動不足’ の人が多いと仮定するオリジナルデータを用意する．この時のオリジナルデータのデータ分布 (Original data)，一次摂動データのデータ分布 (First-order perturbed data)，二次摂動データのデータ分布 (Second-order perturbed data)，再構築データのデータ分布 (Reconstructed data) を図 6.4 に示す．この結果から明らかなように，オリジナルデータのデータ分布と再構築データのデータ分布は似たものとなる．また，センシングに参加するユーザの数 N を変えた時の，データ信頼性の指標であるジェンソンシャノン情報量 JSD 及び，プライバシーの評価指標である $Privacy$ を表 6.1 に示す.

本シミュレーションは，参加型センシングを用いた医療アプリケーションの 1 つに過ぎないが，長期的な医療調査や公衆衛生の分野でモバイルセンシングが注目されていることを考慮すると，より多くのアプリケーションが実社会で用いられる可能性があると言える [23].

様々な多次元データにおける評価

本提案手法 Duplicated Perturbation は，4×3 の 2 次元データだけでなく，様々な多次元データに適用することができる．これを示すために，本評価実験では 5×7 の 2 次元データ，4×12 の 2 次元データ，$3 \times 4 \times 4$ の 3 次元データ，$2 \times 2 \times 3 \times 4$ の 4 次元データに対してプライバシー保護処理を行い，本提案手法における挙動を確認する．

センシングに参加するユーザ数 N を $10,000$，$20,000$，$30,000$，$40,000$ に変化させた時の結果を表 6.2 に示す．評価指標である JSD の値の 10 回の処理を行った時の平均と $Privacy$ の値を記す．また，これらのシミュレーションではオリジナルデータは一様分布をとるものとした．本提案手法は統計処理を用いているため，センシングに参加するユーザの数が増えるにつれて，データ信頼性が向上することが分かる．

表 6.1: 図 6.4 で行ったシミュレーション実験においてユーザ数を変化させた時の評価.

Data sets: 3×4, $D = 2$		
Number of data	Jensen-Shannon divergences	$Privacy$
$100,000$	0.00921	0.1105
$200,000$	0.00498	0.1182
$300,000$	0.00138	0.1293
$400,000$	0.00108	0.1428

図 6.4: m-Health による慢性疾患調査システムに提案手法 Duplicated Perturbation を用いた時の様子．4×3 の 2 次元データを扱った時の，オリジナルデータのデータ分布，一次摂動データのデータ分布，二次摂動データのデータ分布，再構築データのデータ分布を示す．

本提案手法 Duplicated Perturbation では，プライバシー保護処理済データはオリジナルデータに決定されず，そのためプライバシー保護処理済データからオリジナルデータを絞り込むことができない．このため，悪意ある攻撃者が伝送中のデータもしくはサーバ上のデータを取得しても，オリジナルデータを推定することは困難であり，ユーザのプライバシー情報が他者によって明らかにされる可能性がないため，ユーザのプライバシー情報を保護することができると言える．

6.3 本章のまとめ

本章ではまず，慢性疾患や精神疾患の原因解明のために期待されている医療調査用参加型センシングや m-Health の概要を紹介し，医療調査用参加型センシングのアプリケーションで扱うデータの特徴について明らかにした．また，医療調査用参加型センシングを安心して活用することのできる社会を実現するためのプライバシー保護手法 Duplicated Perturbation を提案し，一次元データ・多次元データに分けて述べた．

本手法を用いることによって，悪意ある攻撃者が長期間に渡って盗聴を行う蓄積攻撃からのプライバシー情報の侵害を防ぐことが可能となる．これによって，医療分野で頻繁に活用する，準静的な情報 (quasi-static data) のプライバシー保護を実現した．また，センシングによって得たオリジ

ナルデータとプライバシー保護処理済データの関係に注目することによって，中央サーバ上におけるデータ再構築処理に必要となる計算量を大幅に削減することに成功した．

本章で提案した Duplicated Perturbation を用いることによって，プライバシー情報の侵害に余計なコストを費やすことなく，医療調査用参加型センシングを実社会で活用することができる．本提案手法を実装した医療調査用参加型センシングを用いることによって，原因究明の難しい慢性疾患や精神疾患の調査・研究の発展に寄与できると考えられる．

表 6.2: 提案手法 Duplicated Perturbation を様々な多次元データに適用した時の様子.

Participants	Jensen-Shannon divergences	$Privacy$
Data sets: 5×7, $D = 2$		
$100,000$	0.1981	0.03084
$200,000$	0.1092	0.03823
$300,000$	0.08615	0.03891
$400,000$	0.04292	0.03972
Data sets: 4×12, $D = 2$		
$100,000$	0.3981	0.02439
$200,000$	0.1091	0.02736
$300,000$	0.1292	0.02511
$400,000$	0.09281	0.02974
Data sets: $3 \times 4 \times 4$, $D = 3$		
$100,000$	0.2091	0.05103
$200,000$	0.09282	0.05291
$300,000$	0.07498	0.05283
$400,000$	0.03893	0.05739
Data sets: $2 \times 2 \times 3 \times 4$, $D = 4$		
$100,000$	0.008921	0.09812
$200,000$	0.003822	0.1035
$300,000$	0.002091	0.1492
$400,000$	0.002278	0.1387

第7章　参加型センシングのデータ公開とプライバシー保護

第2章及び第3章では，取得データを一般社会で活用する際には，公開情報や背景知識を用いたプライバシー情報特定攻撃に備える必要があることが既存研究で指摘されていることに言及した．このため，参加型センシングを実社会で活用するためには，政府機関・民間企業からの公開情報や背景知識を用いた攻撃から，プライバシー情報を守る手法の開発が重要である．更に，参加型センシングでは一般ユーザのモバイルデバイスを用いてデータを取得し，一般ユーザがデータを活用できる形で提供することを前提としている．このため，参加型センシングで取得した統計的な情報も公開情報の質を高めるために用いられる．

そこで本章では，まず公開情報や背景知識を用いた攻撃について明らかにする．また，参加型センシングで取得したデータが具体的にどのような形でユーザに提示されるのかを示す．更にこれらをふまえて，公開情報や背景知識に対するプライバシー保護機構 Data-centric Perturbation Infrastructure を提案し，そのために必要なプライバシー保護手法 Empirical Negative Surveys を提案する．Data-centric Perturbation Infrastructure では，参加型センシングで取得したデータそのものをプライバシー保護処理に用いることで，より強固なプライバシー保護を実現する．

7.1　参加型センシングのデータ公開の概要

まず本提案手法が想定する，公開情報や背景知識を用いた攻撃を明らかにすることで，本章で取り組むべき課題を明確にする．

政府機関・民間企業など多数の組織が，ユーザのプライバシー情報に関するデータを持ち，場合によってはユーザの特定ができない形で統計的なデータを一般に公開している．しかしながら，複数の組織から公開された情報をつきあわせることによって，個々のユーザを特定することができることが L. Sweeney[37] によって明らかにされている．

同様に，参加型センシングにおいて Perturbation 技術を用いてプライバシー保護処理を行う場合にも，この公開情報を用いた攻撃を想定しなければならない．特に，種々のモバイルデバイスとアプリケーションが普及していることから，一般ユーザのプライバシー情報は Web 上に偏在しているのが現状であり，これらを悪意ある攻撃者が活用し，参加型センシングのデータからプライバシー情報を取得することは大いに考えられる．

第4章，第5章，第6章で提案した手法では，ユーザの持つモバイルデバイス上でセンサデータに対してプライバシー保護処理を施す．しかしながら，悪意ある攻撃者は，ユーザのプライバシー情報特定のためには出来る限りの手を尽くすと考えられるため，盗聴によって取得したプライバシー保護処理済データに加えて，背景知識・公開情報・社会における共有知を駆使することは明らかである．

更に重大なこととして，参加型センシングで取得したデータは，一般ユーザのために用いられるということが挙げられる．参加型都市環境センシングや道路交通用参加型センシングでは，取得したデータを統計処理し，一般ユーザと共有することを目的とする．また，医療調査用参加型センシングでは，取得したデータを解析することによって，一般ユーザにとって有益な情報を取得するこ

とを目指す．つまり，一般ユーザ及び悪意ある攻撃者は，参加型センシングで取得された統計データを何らかの形で知ることができ，これはより質の良い公開情報となり，悪意ある攻撃者の攻撃を手助けすることとなる．このため，センシングを行う一般ユーザとアプリケーション提供者が安全・安心に参加型センシングを活用していくためには，この公開情報・背景知識・社会の共有知を用いた攻撃を想定したプライバシー保護機構の開発が非常に重要である．

以下では，公開情報・背景知識を用いた攻撃に対抗するためのプライバシー保護機構 Data-centric Perturbation Infrastructure を提案する．また，プライバシー保護機構においてより強固なプライバシー保護を実現する Empirical Negative Surveys の手法を提案する．プライバシー保護機構 Data-centric Perturbation Infrastructure では，参加型センシングで取得した統計的なデータ分布を一般ユーザにも共有することで，悪意ある攻撃者が背景知識・公開情報・社会における共有知を攻撃に活用することを防ぐ．

7.2 公開情報や背景知識に対するプライバシー保護機構 Data-centric Perturbation Infrastructure

Perturbation 技術を参加型センシングのプライバシー保護に用いる時のシステム概要を図 7.1 に示す．第 4 章，第 5 章，第 6 章で提案したプライバシー保護手法では，図 7.1 のように，モバイルデバイス上でセンサデータ (Sensed Data) に対してプライバシー保護処理を施し，その後，プライバシー保護処理済データ (Perturbed Data) を中央サーバに伝送する．このモバイルデバイス上のプライバシー保護処理には，一定の確率分布関数を用いるため，常に同様の処理を施す．

一方で，本プライバシー保護機構 Data-centric Perturbation Infrastructure では，各ユーザから取得した統計的なデータ分布に応じたプライバシー保護処理を施す．このため中央サーバは，各ユーザの統計的なデータ分布を集計した後に，この統計的なデータ分布に基づく確率分布関数を各モバイルデバイスにフィードバックする．本章では，このフィードバック型プライバシー保護機構 Data-centric Perturbation Infrastructure について述べた後に，本プライバシー保護機構のプライバシー保護処理に関する提案手法 Empirical Negative Surveys について述べる．

図 7.1: 既存のプライバシー保護システムの概要．モバイルデバイス上のプライバシー保護処理には，一定の確率分布関数を用いるため，常に同様の処理を施す．

図 7.2: フィードバック型プライバシー保護機構 Data-centric Perturbation Infrastructure の概要．プライバシー保護処理に用いる確率分布は，取得した統計的なデータ分布そのものに基づいたものとなる．

7.2.1 Data-centric Perturbation Infrastructure のシステム概要

本プライバシー保護機構 Data-centric Perturbation Infrastructure では，各ユーザから取得した統計的なデータ分布に応じて，各モバイルデバイスにおけるプライバシー保護処理に用いる確率分布関数を変化させる．このため，取得した統計的なデータ分布を各ユーザにフィードバックすることによって，プライバシー保護処理に用いる確率分布関数を決定する．このフィードバック型プライバシー保護機構 Data-centric Perturbation Infrastructure のシステム概要を図 7.2 に示す．以下にシステムの処理について述べる．

まず，フィードバック型プライバシー保護機構 Data-centric Perturbation Infrastructure におけるモバイルデバイス上のプライバシー保護処理のための確率分布を $\boldsymbol{P}(t)$ とする．モバイルデバイス上におけるプライバシー保護処理はこの確率分布 $\boldsymbol{P}(t)$ に従う．但し，確率分布 $\boldsymbol{P}(t)$ は時間 t の関数である．また同様に，時間 t の関数として，中央サーバ上で再構築処理によって得ることのできる統計的なデータ分布を $\boldsymbol{S}_t$ とする．この統計的なデータ分布 $\boldsymbol{S}_t$ は一定時間 Δt ごとに集計されるものとする．

この時，本プライバシー保護機構における全てのプライバシー保護処理は $\boldsymbol{S}_{t-\Delta t}$ を基に生成される確率分布 $\boldsymbol{P}(t)$ に沿った処理とする．つまりプライバシー保護処理に用いる確率分布は，取得した統計的なデータ分布そのものに基づいたものとなる．本プライバシー保護機構では，取得した統計的なデータ分布を効率的に用いることによって，攻撃者の公開情報・背景知識を用いた攻撃を想定した強固なプライバシー保護を実現する．

本手法では，モバイルデバイス上のプライバシー保護処理において，サーバから共有された離散確率分布 $\boldsymbol{P}(t)$ を用いる．適切な離散確率分布 $\boldsymbol{P}(t)$ を用いたプライバシー保護処理を実現する手法として，以下に Empirical Negative Surveys[64] を提案する．Empirical Negative Surveys は既存の Perturbation 技術を基礎とした手法である．

7.2.2 提案手法: Empirical Negative Surveys

本項では，フィードバック型プライバシー保護機構 Data-centric Perturbation Infrastructure におけるプライバシー保護処理部に関する手法 Empirical Negative Surveys について示す．本プライバシー保護機構では，各ユーザの持つモバイルデバイス上でセンサデータ x を離散値として取得した後に，プライバシー保護処理を行い，プライバシー保護処理済データ y を生成する．このプライバシー保護処理済データ y を中央サーバに伝送する．中央サーバでは，各ユーザからプライバシー保護処理済データ y を収集し，統計的な再構築処理を行い，データ分布を明らかにする．

ここで，オリジナルデータ x，プライバシー保護処理済データ y は α 個のデータ区分となる離散値 (カテゴリ値) として扱うものとし，値が i である時，それぞれ $x(i)$，$y(i)$ と表すものとする．センシングによって取得するデータ x の統計的なデータ分布を $\boldsymbol{X} = (X_1, \cdots, X_\alpha)^\tau$ と表すものとする．また，モバイルデバイス上でプライバシー保護処理を行ったデータ y の統計的なデータ分布を $\boldsymbol{Y} = (Y_1, \cdots, Y_\alpha)^\tau$ とする．更に，中央サーバにおける再構築処理によって集計したデータ分布を $\boldsymbol{A} = (A_1, \cdots, A_\alpha)^\tau$ と表すものとする．本提案手法では，各々のプライバシー情報を秘匿化しながら，オリジナルデータのデータ分布 $\boldsymbol{X}$ と理論的には同じ確率分布である $\boldsymbol{A}$ を構築する．

以下に，モバイルデバイス上におけるプライバシー保護処理と中央サーバにおける再構築処理に分けて提案手法である Empirical Negative Surveys の詳細を記す．

モバイルデバイス上の処理

本提案手法 Empirical Negative Surveys は，フィードバック型プライバシー保護機構 Data-centric Perturbation Infrastructure で用いる処理であり，モバイルデバイス上のプライバシー保護には，時刻 $t - \Delta t$ に取得した統計的なデータ分布 $\boldsymbol{S}_{t-\Delta t}$ を基に生成される確率分布 $\boldsymbol{P}(t)$ を用いる．

この確率分布 $\boldsymbol{P}(t)$ は，α のカテゴリで表される離散確率分布とし，$\boldsymbol{P}(t) = \{P_1^t : P_2^t : P_3^t : \cdots P_{\alpha-1}^t : P_\alpha^t\}$ と表すものとする．但し，$1 \leq k \leq \alpha$ を満たす全ての P_k^t は，$0 \leq P_k^t \leq 1$ を満たすものとする．また，式 (7.1) の関係式を常に満たすものとする．

$$\sum_{k=1}^{\alpha} P_k^t = 1 \tag{7.1}$$

提案手法 Empirical Negative Surveys では，この離散確率分布 $\boldsymbol{P}(t)$ に応じた処理を行い，プライバシー保護処理済データ y を生成する．但し，常に $x \neq y$ の関係を満たすものとする．つまり，センシングによって得た値が $x(k)$ であった時，プライバシー保護処理済データ y は $\{P_1^t : P_2^t : \cdots P_{k-1}^t : 0 : P_{k+1}^t : \cdots P_{\alpha-1}^t : P_\alpha^t\}$ の離散確率分布に従って遷移する．例として $k \neq i$ の時に，センシングによって得た値が $x(k)$ がプライバシー保護処理によって $y(i)$ になる遷移確率は，$\frac{P_i^t}{1-P_k^t}$ である．本処理において，プライバシー保護処理済データ y は，k 以外の $\alpha - 1$ 個のカテゴリから 1 つを選ぶこととなる．

以上が提案手法 Empirical Negative Surveys におけるモバイルデバイス上の処理である．この提案手法 Empirical Negative Surveys において生成されるプライバシー保護処理済データを既存手法 [58] と比較したものを図 7.3 に示す．選択肢が A，B，C，D の 4 つあり，センシングによって得た値が C であった時，既存手法 [58] では一定確率で A，B，D の中から 1 つの選択肢を選んで中央サーバに報告する．対して，提案手法 Empirical Negative Surveys では中央サーバから共有された離散確率分布 $\boldsymbol{P}(t)$ に従って，プライバシー保護処理を行う．

図 7.3: オリジナルデータ (Common Surveys)，既存手法におけるプライバシー保護処理済データ (Negative Surveys[58])，提案手法におけるプライバシー保護処理済データ (Empirical Negative Surveys) の比較図 [64].

各ユーザは，プライバシー保護処理済データ y を中央サーバに報告することになり，プライバシー保護処理済データ y は常に $x \neq y$ の条件の中生成されるので，ユーザはセンシングによって得た値を他者に提供する必要が無い．また，悪意ある攻撃者が盗聴によって伝送中のデータを取得しても，ユーザのプライバシー情報を推定することはできない．

また，本提案手法はモバイルデバイスを用いた参加型センシングでの利用を想定しているが，オリジナルデータ x，プライバシー保護処理済データ y のデータ容量は同じものとなるため，通信による消費電力は非常に低いものとなる．また，上記のモバイルデバイス上の処理は，非常に簡潔な手法であり，計算量も小さなものとなる．このため，計算資源の乏しいモバイルデバイスに最適な処理であると言える．

中央サーバの処理

本提案手法 Empirical Negative Surveys において，中央サーバでは各モバイルデバイス上でプライバシー保護処理を施したデータ y を集計した後に，統計的なデータ分布の再構築処理を行う．

まず，各モバイルデバイスにおけるプライバシー保護処理を統計的に捉えると，プライバシー保護処理済データ y の統計的データ分布 $\boldsymbol{Y} = (Y_1, \cdots, Y_\alpha)^\tau$ と $\boldsymbol{X} = (X_1, \cdots, X_\alpha)^\tau$ の関係式は式 (7.2) の様に表すことができる．この関係式は，モバイルデバイス上のプライバシー保護より明らかである．

$$
\begin{aligned}
Y_k &= X_1 \times \frac{P_k^t}{1 - P_1^t} + X_2 \times \frac{P_k^t}{1 - P_2^t} + \cdots X_\alpha \times \frac{P_k^t}{1 - P_\alpha^t} \\
&= \sum_{l \neq k} \left\{ X_l \times \frac{P_k^t}{1 - P_l^t} \right\} \qquad (7.2)
\end{aligned}
$$

この式 (7.2) の関係式を，$k = 1$ から $k = \alpha$ まで統合し，行列の処理として表したものを式 (7.3) に記す．この式 (7.3) は，プライバシー保護処理済データ y の統計的データ分布 $\boldsymbol{Y}$ とオリジナルデータ x の統計的なデータ分布 $\boldsymbol{X}$ を表したものであり，行列 $\boldsymbol{\Omega}$ はプライバシー保護処理による統計的なデータ分布の変化を巨視的に表したものである．

$$
\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_\alpha \end{pmatrix} = \mathbf{\Omega} \cdot \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_\alpha \end{pmatrix} \tag{7.3}
$$

この行列 $\mathbf{\Omega}$ は，$\alpha \times \alpha$ の正方行列となり，逆行列 $\mathbf{\Omega}^{-1}$ を用いることによって，プライバシー保護処理済データ y の統計的なデータ分布 $\boldsymbol{Y}$ からの再構築処理が可能になる．

以下で，モバイルデバイス上におけるプライバシー保護処理から，行列 $\mathbf{\Omega}$ を明らかにする．

$$
\begin{aligned}
\mathbf{\Omega} &= \begin{pmatrix} 0 & \frac{P_1^t}{1-P_2^t} & \frac{P_1^t}{1-P_3^t} & \cdots & \frac{P_1^t}{1-P_\alpha^t} \\ \frac{P_2^t}{1-P_1^t} & 0 & \frac{P_2^t}{1-P_3^t} & \cdots & \frac{P_2^t}{1-P_\alpha^t} \\ \frac{P_3^t}{1-P_1^t} & \frac{P_3^t}{1-P_2^t} & 0 & \cdots & \frac{P_3^t}{1-P_\alpha^t} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{P_\alpha^t}{1-P_1^t} & \frac{P_\alpha^t}{1-P_2^t} & \cdots & \frac{P_\alpha^t}{1-P_{\alpha-1}{}^t} & 0 \end{pmatrix} \\
&= \begin{pmatrix} P_1^t & 0 & \dots & 0 \\ 0 & P_2^t & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P_\alpha^t \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & \dots & 1 \\ 1 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 0 \end{pmatrix} \\
&\quad \cdot \begin{pmatrix} \frac{1}{1-P_1^t} & 0 & \dots & 0 \\ 0 & \frac{1}{1-P_2^t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{1-P_\alpha^t} \end{pmatrix}
\end{aligned} \tag{7.4}
$$

式 (7.4) から，行列 $\mathbf{\Omega}$ を分解した時の 3 つの行列は特徴的なものとなることがわかる．これら 3 つの行列は，対角成分・非対角成分に特徴を持ち，それぞれ容易に逆行列を示すことができるため，常に逆行列 $\mathbf{\Omega}^{-1}$ が存在することは明らかである．取得したデータの統計的なデータ分布を明らかにするために，以下の式 (7.5) に逆行列 $\mathbf{\Omega}^{-1}$ を示す．

$$
\begin{aligned}
\mathbf{\Omega}^{-1} &= \frac{1}{\alpha - 1} \cdot \begin{pmatrix} 1-P_1^t & 0 & \dots & 0 \\ 0 & 1-P_2^t & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1-P_\alpha^t \end{pmatrix} \\
&\quad \cdot \begin{pmatrix} 2-\alpha & 1 & \dots & 1 \\ 1 & 2-\alpha & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 2-\alpha \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{P_1^t} & 0 & \dots & 0 \\ 0 & \frac{1}{P_2^t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{P_\alpha^t} \end{pmatrix}
\end{aligned}
$$

α データのカテゴリ数を表す変数であり，常に 2 以上であることは明らかであるため，式 (7.5) より逆行列 $\mathbf{\Omega}^{-1}$ は常に存在する．

この逆行列 $\boldsymbol{\Omega}^{-1}$ を，式 (7.3) の関係式に用いることによって，式 (7.5) に示す様に，プライバシー保護処理済データの統計的なデータ分布 $\boldsymbol{Y}$ から，データ再構築処理を行うことができる．本提案手法において，再構築できるデータ分布 $\boldsymbol{A}$ は理論上，オリジナルデータのデータ分布 $\boldsymbol{X}$ と等しいものとなる．

$$\boldsymbol{A} = \boldsymbol{\Omega}^{-1} \cdot \boldsymbol{Y} \tag{7.5}$$

以上で述べた処理が，提案手法 Empirical Negative Surveys におけるデータ再構築処理である．再構築処理のための逆行列 $\boldsymbol{\Omega}^{-1}$ を構成する各行列の非対角成分は 0 または 1 となるため，非常に簡潔な処理でデータ再構築を行うことができる．

本章で提案したフィードバック型プライバシー機構では，この Empirical Negative Surveys に従った処理を行うことによって，攻撃者が用いることのできる背景知識・公開情報よりも質の高い統計的なデータ分布を各ユーザに共有する．これによって，常に悪意ある攻撃者よりもセンシングに参加するユーザが優位な立場となるため，悪意ある攻撃者は盗聴及び背景知識・公開情報を並行して用いても，ユーザのプライバシー情報を正確に推定することはできない．

7.3 本章のまとめ

本章では，まず悪意ある攻撃者が公開情報・背景知識・社会の共有知などを駆使して攻撃を行うことを明らかにし，既存の参加型センシングのプライバシー保護手法がこれらの攻撃を想定していないことについて言及した．更に，これらをふまえて，公開情報や背景知識に対するプライバシー保護機構 Data-centric Perturbation Infrastructure を提案し，そのために必要なプライバシー保護手法 Empirical Negative Surveys について記した．本プライバシー保護機構では，参加型センシングで得た統計的なデータ分布そのものをユーザに共有し，プライバシー保護に活用する．

このフィードバック型のプライバシー保護機構を実現することによって，ユーザは常に悪意ある攻撃者よりも多くの情報量を手にすることができるため，公開情報・背景知識を用いた攻撃に対抗することができる．

本章で設計したプライバシー保護機構 Data-centric Perturbation Infrastructure を実現することによって，参加型センシングで取得した統計的なデータ分布や新たな知識・傾向・情報を一般社会に公開することが可能となる．本提案システムによって，一般ユーザによる，一般ユーザのためのセンシング機構を実現する手法を確立した．

第8章　プライバシー保護に関する考察

本研究では，第 4 章，第 5 章，第 6 章にてそれぞれ想定する参加型センシングシステムを実現するプライバシー保護手法を提案し，シミュレーションによる評価を行った．また，第 7 章では参加型センシングで取得した情報を適切に一般社会に公開するためのプライバシー保護機構を提案した．

本章では，これら本研究で提案した手法・システムの考察を行い，本研究が第 3 章で示した技術的課題を解決したことを示す．

8.1　技術的な課題に関する考察

本論文では，まず参加型センシングのシステム概要と目的について明確にし，またモバイルデバイスによるセンシング技術を紹介することによって，参加型センシングの実現のためには適切なプライバシー保護技術の開発が非常に重要であることを述べた．

第 3 章では，既存のプライバシー技術の技術的な限界を示した．この技術的な限界は，具体的に以下に示す 4 点を実現することを目指した．

- データ区分の細かい多次元データのためのプライバシー保護
- 二者択一型選択肢を含む多次元データのためのプライバシー保護
- 長期間にわたる蓄積攻撃に対するプライバシー保護
- 公開情報や背景知識を利用した攻撃に対するプライバシー保護

まず，第 4 章では，Android 端末を用いた参加型都市環境センシングシステムを実装し，このシステムで扱うデータの特徴として，データ区分を細かく扱う必要があることを明らかにした．この課題を解決するために，第 4 章では既存の Negative Surveys に到達先パラメータを設定するプライバシー保護手法 Selective Negative Surveys を提案し，データ区分の細かい多次元データのプライバシー保護処理を実現した．本手法の実現によって，環境情報や位置情報などデータ区分の細かい情報のデータ信頼性を確保しながら，個々のデータが持つプライバシー情報を保護することができた．これによって，一般ユーザがモバイルデバイスを街中で持ち歩く中で，一般ユーザを煩わせることなくセンシングを行うことが可能となる．

次に，第 5 章では，イベント検知のための参加型センシングやプロファイル情報を扱う参加型センシングの中で，二者択一型選択肢を持つ多次元データを扱うことを明らかにした．この課題を解決するために，第 5 章では Randomized Response と Negative Surveys を組み合わせた MNS-RRT アルゴリズムを提案した．本手法の実現によって，性別に関するプロファイル情報や閾値判定を示す情報である，二者択一型選択肢を持つ多次元データのプライバシー情報を保護しながら収集する手法を確立した．これによって，プロファイル情報を扱う必要のあるマーケット戦略策定のための調査や，放射線等に関するハザードマップを参加型センシングとモバイルデバイスを用いて構築することが可能となる．

第 6 章では，医療調査用の参加型センシングシステムや m-Health システムについて明らかにした後に，これらのシステムで収集・活用するデータが時間依存性の低い準静的なデータ (quasi-static data) であることについて言及し，長期間に及ぶ盗聴による蓄積攻撃に対して大きな脆弱性があることを明らかにした．これに対抗するために，第 6 章では，Duplicated Perturbation の手法を提案した．Duplicated Perturbation は，複数のフィールド情報を持つ多次元データに対して，フィールド間の相関を保存しながら Randomized Response を行うことができる．本手法の実現によって，時間依存性の低い準静的なデータについて，個々の情報が持つプライバシー情報を保護しながら収集する手法を確立した．提案手法 Duplicated Perturbation を活用することによって，精神疾患や慢性疾患など，従来調査が難しい疾患の環境的要因を明らかにすることができることが期待される．また本手法では，中央サーバにおける再構築処理の計算量削減にも取り組んでいる．本手法を，データベース上に蓄積されたデータに対して施すこともできるため，プライバシー保護データマイニング (Privacy-Preserving Data Mining) へと拡張できる可能性もあると言える．

最後に，第 7 章では，参加型センシングで取得・共有したデータを一般社会に公開することを想定し，フィードバック型プライバシー保護機構 Data-centric Perturbation Infrastructure を提案し，その機構で行う処理 Empirical Negative Surveys を提案した．本機構 Data-centric Perturbation Infrastructure では，参加型センシングで取得したデータ分布そのものをプライバシー保護処理に用いることで，より強固なプライバシー保護を実現する．本手法は，ユーザの持つモバイルデバイスとアプリケーション提供者の中央サーバが協調的にプライバシー情報の保護処理を行う非常に画期的な手法であり，理論的にも正確な統計的なデータ分布を取得することができる．しかしながら，本手法では各モバイルデバイスがサーバ側と通信するデータサイズが増大し，またモバイルデバイスにおけるプライバシー保護処理に要する計算量及び消費電力も大きなものとなる．このため，実社会で本機構を用いるためには，全体の計算量を削減する必要がある．

以上で示したように，本論文では第章で言及した技術的課題についてそれぞれ手法・システムを提案・評価することによって，問題の解決を行った．プライバシー情報の扱いは，技術的な観点だけでなく，社会学・法学の観点から非常に複雑な問題であり，プライバシー情報の価値を一般ユーザの反応から決定することを試みる研究もなされている [100, 101]．しかしながら，参加型センシングでは一般ユーザの持つモバイルデバイスをセンサとして用いるため，適切なプライバシー保護処理を施す手段が無い場合，多くのユーザはセンシングに参加しないと言える．仮に，一定数のユーザがセンシングに参加したとしても，その場合は取得データに偏りが生じることが考えられる．このため，より大規模なユーザが参加型センシングを活用する社会を実現するためには，適切なプライバシー保護が非常に重要である．

8.2 データ信頼性とプライバシーレベルに関する考察

データ利活用する際のプライバシー情報の取り扱いに関する議論において，データ信頼性とプライバシーレベルの間に存在するトレードオフは非常に重要な問題である．Perturbation 技術 (データ摂動) を用いる際や k-匿名化手法の一般化・丸め処理において，プライバシーレベルを高めるとデータの信頼性が損なわれる．一方で，データの信頼性を高めるために，個々のレコード情報がもつ情報量を高めると，プライバシーレベルが下がる．

本研究で示した参加型センシングのプライバシー保護処理において，理論的には統計的なデータ分布を完全に再構築することができる．つまり，十分に大きなデータ量 N を取得することができる時，データ信頼性を高めることができる．

これらを考慮すると，実社会で参加型センシングシステムを活用する場合，全体のデータ量 N，つまりセンシングに参加する一般ユーザの数と社会における普及率に対応する形で，データ信頼性とプライバシーレベルのバランスを決定すべきである．具体的には，E. M. Rogers の提唱した Diffusion of Innovation[102] を参考にする．

まず，参加型センシングシステムの提供直後には，社会における認知度が低いために収集できるデータ N とユーザ数が少ないことが考えられる．このため，参加型センシング本来の目的である統計的なデータ分布を取得・提供するために，まずはデータ信頼性を優先する．次に，参加型センシングが社会に普及し，収集することのできるデータ量が増えるに連れて，プライバシーレベルを高める．これによって，常に信頼性の高いデータを取得し，良質なサービスを提供し続けることができる．

特に，本研究で示した Selective Negative Surveys 及びプライバシー保護機構 Data-centric Perturbation Infrastructure では，データ信頼性とプライバシーレベルをアプリケーション提供者が調整することのできる設計となっている．

また，本研究で示した手法は多次元データを対象としたプライバシー保護手法であるが，多次元データは次元数と各次元のカテゴリ数を調整することが可能である [59]．つまり，4×12 の 2 次元データは $4 \times 3 \times 4$ の 3 次元データ，$2 \times 2 \times 3 \times 4$ の 4 次元データ，48 カテゴリを持つ 1 次元データとして扱うことが可能である．この際，本研究で示した手法では統計処理を施すため，次元数が小さく次元あたりのカテゴリ数が大きいほどデータ信頼性が低下する．また，次元数を大きくし次元あたりのカテゴリ数を小さくすると，データ信頼性が向上し，プライバシーレベルが低下する．つまり，盗聴などによってプライバシー保護済データが取得された場合，プライバシー情報の特定がより容易になる．また，データの次元数と各次元のカテゴリ数を調整する場合，各モバイルデバイス上でセンシングの後に施すプライバシー保護処理に必要な計算量が大きなものとなる．

第9章　結論

9.1　本研究の主たる貢献

一般ユーザの所有するモバイルデバイスをセンサとして活用する参加型センシングを実現するためには，ユーザがプライバシー情報の侵害を懸念すること無く，センシングに参加することのできる環境を構築することが非常に重要である．また，多くのアプリケーション提供者が参加型センシングシステムを社会の中で運用していくためにも，プライバシー情報の漏洩という危険の無い参加型センシングシステムを設計することは非常に重要である．

このため，本研究では参加型センシングにおいて，モバイルデバイスの計算資源・電源が限られたものとなるという特徴を考慮した上で，計算量の低いプライバシー保護処理を各デバイス上で行い，プライバシー情報の含まれていない情報を中央サーバに伝送する手法を提案した．本手法の実現によって，ユーザは他者にプライバシー情報を提供する必要が無くなり，また中央サーバは個々のユーザのプライバシー情報を保持する必要無く，統計的な情報を取得することが可能となる．

このような計算量の低いプライバシー保護処理を実現するために，第 2 章では，プライバシー情報の保護手法とモバイルデバイスを用いた参加型センシングについて整理した上で，Perturbation の手法が参加型センシングに用いるためのプライバシー保護処理として適していることを明らかにした．次に，第 3 章では，既存の Perturbation の手法を Android 端末に実装し，攻撃者が盗聴・蓄積攻撃を行うことを想定した実証実験を行うことによって，参加型センシングで想定するアプリケーションに適用する際の技術的課題を示した．第 4 章では参加型都市環境センシングシステムを開発し，東京都 世田谷区全域でセンシング実験を行った上で，参加型都市環境センシングで取得するデータ区分の細かいセンサデータを想定したプライバシー保護手法 Selective Negative Surveys を提案し，プライバシーとデータ信頼性の両側面から評価を行った．第 5 章では，プロファイル情報を対象とした参加型センシングのプライバシー保護を実現するために，Randomized Response と Negative Surveys を組み合わせてプライバシー保護処理を行う MNS-RRT アルゴリズムを提案し，シミュレーションによって評価を行った．提案手法 MNS-RRT アルゴリズムによって，二者択一型選択肢のデータを扱うことが可能となり，イベント検知のための参加型センシングにおけるプライバシー保護も実現した．第 6 章では，医療調査や公衆衛生のための参加型センシングをアプリケーションとして想定した上で，悪意ある攻撃者からの蓄積攻撃によるプライバシー情報の侵害・漏洩を防ぐためのプライバシー保護手法 Duplicated Perturbation を提案し，シミュレーション評価を行った．第 7 章では，悪意ある攻撃者が公開情報・背景知識を用いて攻撃をしてくることについて言及した上で，参加型センシングで取得したデータを一般公開するためのフィードバック型プライバシー保護機構 Data-centric Perturbation Infrastructure を提案した．本提案機構では，取得した統計的なデータ分布そのものを，プライバシー保護処理に活用し，より強固なプライバシー情報保護を実現した．また，第 8 章では，各アプリケーションで求められるプライバシーレベルとデータ信頼性，センシング参加者の数について言及した上で，参加型センシングシステムを実社会で運用する際のプライバシー保護に関する議論を記した．

本研究は，各デバイス上で取得したセンサデータのプライバシー保護処理を行い，プライバシー情報の含まれていない情報を中央サーバに伝送することによって，参加型センシングで扱う多次元データの統計的性質のみを安全に取得する初めての研究である．本論文は，モバイルデバイスを用

いた参加型センシングにおいて，ユーザがプライバシー情報の侵害・漏洩を懸念することなくセンシングに参加する環境を構築した点で学術的な貢献がある．

9.2 今後の課題と展望

本研究で取り扱った参加型センシングとプライバシー情報の保護に関して，今後の展望を述べる．提案手法では，Perturbation 技術を活用することによって，参加型センシングにおけるプライバシー情報の保護を技術的に実現したが，センサデータに含まれるプライバシー情報の経済的価値や情報量，一般ユーザのプライバシー情報に対する価値観・プライバシー情報の利活用に対する反応に関する議論は行っていない．参加型センシングで取得するデータの信頼性とシステム最適化のためにも，このプライバシー情報の価値を実社会の中で明らかにする必要がある．この目的のため，一般社会の中で参加型センシングを実運用し，プライバシー情報の経済的価値を明らかにすることは今後の重要な課題である．

また，第 7 章で提案・設計したフィードバック型プライバシー保護機構 Data-centric Perturbation Infrastructure は，理論的にはプライバシー保護を強固に行うことができるが，モバイルデバイス上の消費電力・計算量の評価は行っていない．データの取得から公開・利活用に至るまで一貫してプライバシー情報を守るためには，フィードバック型プライバシー保護機構 Data-centric Perturbation Infrastructure の計算量に関する評価は今後の課題である．

本論文で提案した手法と，これらの課題の解決によって，一般ユーザを煩わせることなく，世界中で普及しつつあるモバイルデバイスをセンサとして活用する参加型センシングを実現できる．このモバイルデバイスを用いた参加型センシングが，社会調査や医療研究など様々な場面で活用され，人類の発展を根底から支えるインフラとなることを確信している．

謝辞

本論文は，筆者が東京大学大学院 情報理工学系研究科 電子情報学専攻 修士課程に在籍中に取り組んだ研究をまとめたものであり，本研究の遂行を通して情報技術の奥深さを知ることができただけでなく，学問・研究の面白さに触れることができました．まず，本研究の基礎を築き上げて来た偉大な先人達に敬意を表し，また感謝致します．

本研究の遂行及び本論文の執筆にあたり，研究に対する助言だけでなく，学問に対する姿勢・思考方法・取り組み方など様々な面でご指導くださった 瀬崎薫 教授に，心より感謝致します．また，研究の立ち上げ時から様々なサポートをしてくださった東京電機大学 未来科学部 岩井将行 准教授と，連日深夜まで，時には朝まで研究に関する議論にお付き合い下さった東京大学 空間情報科学研究センター 小林博樹 助教に，深く感謝致します．加えて，半年間の間，様々な研究に関する議論にお付き合いくださった東京大学 生産技術研究所 伊藤昌毅 助教と研究のみならず日常生活や進路に関しても様々なご助言をくださり，常にサポートしていただいた東京大学 空間情報科学研究センター 奥野淳也 研究員にも，深く感謝致します．

また，研究を進めるにあたり，必要であった音声データを提供してくださいました総合研究大学院大学 学融合推進センター 塚原直樹 助教と慶應義塾大学大学院 社会学研究科 近藤紀子 博士研究員，また修士課程在籍中に Visiting Student として滞在させていただき，非常に有意義な議論を持つことができた ETH Zurich の Institute of Theoretical Computer Science と，研究遂行にあたり実験にご協力くださった東京電機大学 岩井研究室の方々と University of Oulu の Teemu Leppanen 氏にも謝意を表したいと思います．

更に，常に研究に対する姿勢を示してくださった，東京大学 瀬崎研究室 博士課程の Asif Hossain Khan 氏，党聡維 氏，劉広文 氏，Niu Hao 氏，Dunstan Matekenya 氏，また同期として楽しく濃い時間を共に過ごした木田裕一朗 氏，Jose Alvarez Pablo 氏，中川慶次郎 氏，高い研究意欲を持ち，多くの刺激を与えてくれた後輩の 稲葉瞳 氏，坂本敬太 氏，中村直人 氏，加藤宗肖 氏，国内外での研究活動を支えてくださった秘書の 松本夏穂 氏，内藤潤 氏に，感謝の言葉を述べさせていただきたいと思います．

最後に，現在に至るまで育ててくれた両親に深く感謝し，本論文を閉じたいと思います．

発表文献 / Publications

原著論文 / Journals

[1] 青木俊介，岩井将行，瀬崎薫，“参加型環境センシングを用いた統計情報構築のための プライバシー保護手法，” 電子情報通信学会 論文誌 B，vol. J97-B，no. 01，pp. 41-50，2014 年 1 月.

[2] S. Aoki and K. Sezaki, “Negative Surveys with Randomized Response Techniques for Privacy-aware Participatory Sensing, ” IEICE Transactions on Communications, vol. E97-B, No. 04, 2014. (in Publish)

国際会議 / International Conferences

[3] S. Aoki, M. Iwai, and K. Sezaki, “Limited Negative Surveys: Privacy-Preserving Participatory Sensing, ” The 1st IEEE International Conference on Cloud Networking (IEEE CLOUDNET), pp. 158-160, Paris, France, Nov 2012. (**Best Short Paper Candidate**)

[4] S. Aoki, M. Iwai, and K. Sezaki, “Perturbation for Privacy-Preserving Participatory Sensing on Mobile, ” IEEE International Conference for Internet Technology and Secured Transactions (IEEE ICITST), London, UK, Dec 2012. (Poster)

[5] S. Aoki, H. Kobayashi, M. Iwai, and K. Sezaki, “Perturbation with General Tendency for Mobile Community Sensing, ” The 2nd IEEE International Conference on Mobile Services (IEEE MS), pp. 23-30, CA, USA, Jun 2013.

[6] S. Aoki, M. Iwai, and K. Sezaki, “Privacy-aware Community Sensing using Randomized Response, ” IEEE 8th International Workshop on Security, Trust and Privacy (IEEE STPSA), pp. 127-132, Kyoto, Japan, Jul 2013.

[7] M. Iwai, H. Shigeta, G. Liu, S. Aoki, and K. Sezaki, “A System for Large-scale Participatory Environmental Noise Sensing, ” International Conference on Instrumentation, Control, Information Technology and System Integration, SICE Annual Conference, Nagoya, Japan, Sep 2013.

[8] S. Aoki, H. Kobayashi, and K. Sezaki, “Cicada Fingerprinting System: From Artificial to Sustainable, ” The International Conference on Mobile and Ubiquitous Multimedia (MUM 2013), Lule, Sweden, Dec 2013.

[9] S. Aoki and K. Sezaki, “Privacy-Preserving Community Sensing for Medical Research with Duplicated Perturbation, ” The IEEE International Conference on Communications (IEEE ICC), Sydney, Australia, Jun 2014. (in Publish)

国内会議 / Domestic Conferences

[10] 青木俊介，岩井将行，瀬崎薫，“多次元データに対する Negative-Survey を用いたプライバシー保護手法の検討，” 電子情報通信学会ソサイエティ大会，B-19-5，2012 年 9 月.

[11] 青木俊介，岩井将行，瀬崎薫，“参加型センシングのための携帯端末上でのプライバシー保護の検討，” 電子情報通信学会 ヒューマンプローブ研究会 公開シンポジウム，P14，2012 年 11 月．(Poster)

[12] 青木俊介，岩井将行，瀬崎薫，“個人情報保護を考慮した空間情報構築手法の検討，” 電子情報通信学会 技術研究報告，vol. 112，no. 430，IA2012-82，pp. 73-78，2013 年 2 月．(**学生研究奨励賞 受賞**)

[13] 青木俊介，岩井将行，瀬崎薫，“既知データ分布に対する個人情報保護を考慮したユーザ参加型センシングに関する検討，” 電子情報通信学会 技術研究報告，vol. 112，no.466，LOIS2012-83，pp. 83-88，2013 年 3 月.

[14] 青木俊介，岩井将行，瀬崎薫，“ユーザ参加型センシングにおいてプロファイル情報を対象としたプライバシー保護手法の検討，” 電子情報通信学会 総合大会，B-19-50, 2013 年 3 月.

[15] 青木俊介，劉広文，岩井将行，瀬崎薫，“ユーザ固有の雑音を考慮する参加型環境センシングのデータ校正手法，” 情報処理学会 研究報告ユビキタスコンピューティングシステム，2013-UBI-38，no. 18，pp. 1-7，2013 年 5 月.

[16] 青木俊介，岩井将行，瀬崎薫，“ユーザ参加型環境センシングにおける状態依存型プライバシー保護手法，” 電子情報通信学会 技術研究報告，vol. 113，no. 38，ASN2013-14，pp. 73-78，2013 年 5 月．(**若手研究奨励賞 受賞**)

[17] 重田航平，青木俊介，劉広文，岩井将行，瀬崎薫，“モバイル端末を用いたユーザ参加型環境センシングにおけるデータ欠損の検知および処理に関する考察，” 情報処理学会 マルチメディア，分散，協調とモバイル (DICOMO2013) シンポジウム，2A-4，2013 年 7 月． (ヤングリサーチャ賞 受賞)

[18] 青木俊介，劉広文，清水和人，岩井将行，瀬崎薫，“ユーザ参加型環境センシングにおける効率的なシステム運用モデルの構築とユーザ分析，” 情報処理学会 マルチメディア，分散，協調とモバイル (DICOMO2013) シンポジウム，8G-4，2013 年 7 月． (**優秀プレゼンテーション賞 受賞**)

[19] 重田航平，青木俊介，劉広文，瀬崎薫，岩井将行，“ユーザ参加型センシングアプリにおいて計測者の人為的ミスを軽減する改良手法，” 第 12 回情報科学技術フォーラム FIT 2013，M-018，2013 年 9 月.

[20] 青木俊介，小林博樹，近藤紀子，塚原直樹，瀬崎薫，“スマートフォン搭載の GPS とマイクを用いた都市圏における野鳥観察手法の提案，” 日本鳥学会 2013 年度大会，2013 年 9 月．(Poster)

参考文献 / References

[1] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy and M. B. Srivastava, "Participatory Sensing," World Sensor Web Workshop, ACM SenSys 2006, Colorado, Oct 2006.

[2] "Our Mobile Planet," http://www.thinkwithgoogle.com/mobileplanet/en/

[3] 青木俊介，岩井将行，瀬崎薫，"参加型環境センシングを用いた統計情報構築のための プライバシー保護手法，" 電子情報通信学会 論文誌 B，vol. J97-B，no. 01，pp. 41-50，2014.

[4] J. Ko, C. Liu, M. B. Srivastava, J. A. Stankovic, A. Terzis, and M. Welsh, "Wireless Sensor Networks for Healthcare," Proceedings of the IEEE, vol. 98, no. 11, pp. 1947-1960, 2010.

[5] P. Germanakos, C. Mourlas, and G. Samaras, "A Mobile Agent Approach for Ubiquitous and Personalized eHealth Information Systems," Proceedings of Workshop on Personalization for e-Health of the 10th International Conference on User Modeling, Edinburgh, pp. 67-70, Jul 2005.

[6] A. K. Dey, K. Wac, D.Ferreira, K. Tassini, J. Hong, and J. Ramos "Getting Closer: An Empirical Investigation of the Proximity of User to their Smart Phones," Proceedings of ACM International Conference on Ubiquitous Computing (UbiComp 2011), pp. 163-172.

[7] L. Bao and S. S. Intille, "Activity Recognition from User-annotated Acceleration Data," Pervasive Computing, pp. 1-17, 2004.

[8] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity Recognition using Cell Phone Accelerometers," ACM SIGKDD Explorations Newsletter, Vol. 12, No. 2, pp.74-82, Dec. 2010.

[9] A. Sellen, A. Fogg, M. Aitken, S. Hodges, C. Rother, and K. Wood, "Do Life-logging Technologies Support Memory for the Past?: An Experimental Study Using Sensecam," Proceedings of the ACM SIGCHI Conference on Human factors in computing systems (CHI 2007), pp. 81-90, Apr 2007.

[10] K. L. Huang, S. S. Kanhere, W. Hu, "Preserving Privacy in Participatory Sensing Systems," Computer Communications, vol. 33, no. 11, pp.1266-1280, Jul 2010.

[11] S. Aoki, M. Iwai, and K. Sezaki, "Privacy-aware Community Sensing using Randomized Response, " Proceedings of IEEE 8th International Workshop on Security, Trust and Privacy (IEEE STPSA), pp. 127-132, Kyoto, Jul 2013.

[12] 青木俊介，岩井将行，瀬崎薫，"ユーザ参加型環境センシングにおける状態依存型プライバシー保護手法, " 電子情報通信学会 技術研究報告，vol. 113，no. 38，ASN2013-14，pp. 73-78，2013.

[13] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, "A Survey of Mobile Phone Sensing," Communications Magazine, Proceedings of the IEEE, vol. 48, no. 9, pp. 140-150, 2010.

[14] A. T. Campbell, S. B. Eisenman, N. D.Lane, E. Miluzzo, R. A. Peterson, H. Lu, X. Zheng, M. Musolesi, K. Fodor, and G. S. Ahn, "The Rise of People-Centric Sensing," IEEE Internet Computing, vol.12, no. 4, pp. 12-21, 2008.

[15] N. Maisonneuve, M. Stevens, M. Niessen and L. Steels,"NoiseTube: Measuring and Mapping Noise Pollution with Mobile Phones," Proceedings of 4th International Symposium on Information Technologies in Environment Engineering (ITEE), pp. 215-238, May 2009.

[16] "NoiseTube," http://www.noisetube.net/.

[17] "Haze project," http://www.pollution.ee.unsw.edu.au, 2010.

[18] P. Mohan, V. N. Padmanabhan, and R. Ramjee, "Nericell: Rich Monitoring of Road and Trac Conditions using Mobile Smartphones," Proceedings of 6th ACM Conference on Embedded Networked Sensor Systems (SenSys 2008), NC, USA, pp. 165-178, Nov 2008.

[19] S. Eisenman, E. Miluzzo, N. Lane, R. Peterson, G. Ahn and A. Campbell, "BikeNet: A Mobile Sensing System for Cyclist Experience Mapping ," ACM Transactions on Sensor Networks, vol. 6, pp. 1-39, 2009.

[20] M. Annavaram, N. Medvidovic, U.Mitra, S. Narayanan, G. Sukhatme, Z. Meng, S. Qiu, R. Kumar, G. Thatte, and D. Spruijt-Metz, "Multimodal Sensing for Pediatric Obesity Applications," Proceedings of International Workshop on Urban, Community, and Social Applications of Networked Sensing Systems (UrbanSense08), pp. 21-25, Nov 2008.

[21] E. Kanjo, J.Bacon, D. Roberts, and P. Landshoff, "MobSens: Making Smart Phones Smarter," IEEE Pervasive Health Computing, vol. 8, no. 4, pp. 50-57, 2009.

[22] 斉藤裕樹，菅生啓示，間博人，テープウィロージャナポン ニワット，戸辺義人，"sBike：参加型センシングを志向したモバイルセンシングによる自転車走行状態収集・共有機構，" 情報処理学会論文誌，Vol. 53，No. 2，pp. 770-782，2012.

[23] R. Istepanian, S. Laxminarayan, and C. S. Pattichis, "M-Health: Emerging Mobile Health Systems," Springer Science+Business Media, 2006.

[24] R. Istepanian and J. C. Lacal, "Emerging Mobile Communication Technologies for Health: Some Imperative Notes on m-Health," Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 2, pp. 1414-1416, 2003.

[25] M. Azizyan, I. Constandache, and R. Roy Choudhury, "SurroundSense: Mobile Phone Localization via Ambience Fingerprinting," Proceedings of the 15th Annual International Conference on Mobile Computing and Networking (MobiCom 2009), pp. 261-272, 2009.

[26] S. P. Tarzia, P. A. Dinda, R. P. Dick, and G. Memik, "Indoor Localization without Infrastructure using the Acoustic Background Spectrum," Proceegings of the 9th ACM International Conference on Mobile Systems, Applications, and Services (MobiSys 2011), pp.155-168, 2011.

[27] S. Aoki, H. Kobayashi, and K. Sezaki, "Cicada Fingerprinting System: From Artificial to Sustainable," Proceedings of the International Conference on Mobile and Ubiquitous Multimedia (MUM 2013), 2013.

[28] Y. Chon, N. D. Lane, F. Li, H. Cha, and F. Zhao, Automatically Characterizing Places with Opportunistic Crowdsensing using Smartphones," Proceedings of ACM International Conference on Ubiquitous Computing (UbiComp 2012), pp. 481-490, 2012.

[29] H. Lu, D. Frauendorfer, M. Rabbi,M. S. Mast, G. T. Chittaranjan, A. T. Campbell, and T. Choudhury, "StressSense: Detecting Stress in Unconstrained Acoustic Environments using Smartphones," Proceedings of ACM International Conference on Ubiquitous Computing (UbiComp 2012), pp. 351-360, 2012.

[30] R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong, "MoodScope: Building a Mood Sensor From Smartphone Usage Patterns," Proceedings of the 11thACM International Conference on Mobile Systems, Applications, and Services (MobiSys 2013), pp. 389-402, 2013.

[31] K. K. Rachuri, M. Musolesi, C. Mascolo, P. J. Rentfrow, C. Longworth, and A. Aucinas, "EmotionSense: A Mobile Phones based Adaptive Platform for Experimental Social Psychology Research," Proceedings of ACM International Conference on Ubiquitous Computing (Ubicomp 2010), pp. 281-290, 2010.

[32] J. Vaidya, C. W. Clifton, and Y. M. Zhu, "Privacy Preserving Data Mining," Springer Science+Business Media, 2006.

[33] G. Drosatos, P. S. Efraimidis, I. N. Athanasiadis, E. D'Hondt and M. Stevens, "A Privacy-Preserving Cloud Computing System for Creating Participatory Noise Maps," Proceesings of 36th IEEE International Conference on Computer Software and Applications (COMPSAC 2012), pp.581-586, 2012.

[34] E. De Cristofaro and C. Soriente, "PEPSI- Privacy-Enhanced Participatory Sensing Infrastructure," Proceedings of the 4th ACM Conference on Wireless Network Security, pp. 23-28, 2011.

[35] E. De Cristofaro and R. Di Pietro, "Preserving Query Privacy in Urban Sensing Systems," In Distributed Computing and Networking, Springer Berlin Heidelberg, pp. 218-233, 2012.

[36] V. Bozovic, D. Socek, R. Steinwandt, and V.I. Villnyi, "Multi-authority Attribute-based Encryption with Honest-but-curious Central Authority," International Journal of Computer Mathematics, vol. 89, no. 3, pp. 268-283, 2012.

[37] L. Sweeney, "k-anonymity: A Model for Protecting Privacy," International Jounal of Uncertainty, Fuzziness, and Knowledge-based Systems, vol. 10, no. 5, pp. 557-570, 2002.

[38] P. Samarati, “Protecting Respondents Identities in Microdata Release,” IEEE Transactions on Knowledge and Data Engineering, vol. 13, no. 6, pp. 1010-1027, 2001.

[39] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, “l-diversity: Privacy beyond k-anonymity,” ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 1-1 no. 3, 2007.

[40] B. Gedik and L. Liu, “Protecting Location Privacy with Personalized k-anonymity: Architecture and Algorithms,” IEEE Transactions on Mobile Computing, vol. 7, no. 1, pp. 1-18, 2008

[41] M. Shao, Y. Yang, S. Zhu, and G. Cao, “Towards Statistically Strong Source Anonymity for Sensor Networks,” Proceedings of IEEE 27th Conference on Computer Communications (INFOCOM 2008), pp. 51-55, 2008.

[42] “Differential Privacy for Everyone,” http://www.microsoft.com/en-us/download/details.aspx?id=35409.

[43] 五十嵐大，高橋克巳，“注目のプライバシー Differential Privacy,” コンピュータソフトウェア, vol. 29，no. 4，pp. 40-49, 2012.

[44] C. Dwork, F. McSherry, K. Nissim, and A. Smith, “Calibrating Noise to Sensitivity in Private Data Analysis,” In Theory of Cryptography, Springer Berlin Heidelberg, pp. 265-284, 2006.

[45] C. Dwork, “Differential Privacy,” In Automata, languages and programming, Springer Berlin Heidelberg, pp. 1-12, 2006.

[46] F. D. McSherry, “Privacy Integrated Queries: An Extensible Platform for Privacy-preserving Data Analysis,” Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data, pp. 19-30, 2009.

[47] F. McSherry and R. Mahajan, “Differentially-Private Network Trace Analysis,” Proceedings of ACM SIGCOMM Computer Communication Review, vol. 40, no. 4, pp. 123-134, 2010.

[48] H. Haddadi, R. Mortier, S. Hand, I. Brown, E. Yoneki, J. Crowcroft, and D. McAuley, “Privacy Analytics,” ACM SIGCOMM Computer Communication Review, vol. 42, no. 2, pp. 94-98, 2012.

[49] W. Du and Z. Zhan, “Using Randomized Response Techniques for Privacy-preserving Data Mining,” Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data mining, pp. 505-510, 2003.

[50] H. Kargutpa, S. Datta, Q. Wang, and K. Sivakumar, “On the Privacy Preserving Properties of Random Data Perturbation Techniques,” Proceedings of the IEEE International Conference on Data Mining (ICDM 2003), pp.99-106, 2003.

[51] Z. Huang, W. Du, and B. Chen.“Deriving Private Information from Randomized Data,” Proceedings of ACM SIGMOD International Conference on Management of Data, pp.37-48, Jun 2005.

[52] R. K. Ganti, N. Pham, Y. Tsai and T. F. Abdelzaher,"PoolView: Stream Privacy for Grassroots Participatory Sensing," Proceedings of 6th ACM Conference on Embedded Networked Sensor Systems (SenSys 2008), NC, USA, pp.281-294, 2008.

[53] "PoolView," http://smart-attire.cs.uiuc.edu/poolview/.

[54] N. Pham, R. K. Ganti, Y. S. Uddin, S. Nath, and T. Abdelzaher, "Privacy-preserving Reconstruction of Multidimensional Data Maps in Vehicular Participatory Sensing," In Wireless Sensor Networks, Springer Berlin Heidelberg, pp. 114-130, 2010.

[55] F. Zhang, L. He, W. He, and X. Liu, "Data Perturbation with State-Dependent Noise for Participatory Sensing," Proceedings of IEEE 29th Conference on Computer Communications (INFOCOM 2012), pp. 2246-2254, 2012.

[56] S. L. Warner, "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," Journal of the American Statistical Association, vol. 60, no. 309, pp. 63-69, 1965.

[57] R. Agrawal, R. Srikant and D.Thomas, "Privacy Preserving OLAP," Proceedings of ACM SIGMOD International Conference on Management of Data, pp. 251-262, 2005.

[58] F. Esponda, "Negative Surveys," ArXiv Mathematics e-prints, 2006.

[59] M. M. Groat, B. Edwards, J. Horey, W. He and S. Forrest, "Enhancing Privacy in Participatory Sensing with Multidimensional Data," Proceedings of IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 144-152, 2012.

[60] D. Wenliang and Z. Zhan, "Using Randomized Response Techniques for Privacy-Preserving Data Mining," Proceedings of 9th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2003.

[61] Z. Huang, H. J. Wang and N. Borisov, "OptRR: Optimizing Randomized Response Schemes for Privacy-Preserving Data Mining," Proceedings of 24th International Conference on Data Engineering, pp.705-714, 2008.

[62] D. Quercia, I. Leontiadis, L. McNamara, C. Mascolo, and J. Crowcroft, "SpotME If You Can: Randomized Responses for Location Obfuscation on Mobile Phones," Proceedings of 31st International Conference on Distributed Computing Systems (ICDCS), pp. 363-372, 2011.

[63] F. Esponda and V. M. Guerrero, "Surveys with Negative Questions for Sensitive Items," Statics & Probability Letters, vol. 79, no. 24, pp. 2456-2461, 2009.

[64] S. Aoki, H. Kobayashi, M. Iwai, and K. Sezaki, "Perturbation with General Tendency for Mobile Community Sensing," Proceedings of IEEE International Conference on Mobile Services (MS 2013), pp. 23-30, 2013.

[65] H. Xie, L. Kulik, and E. Tanin,"Privacy-aware Collection of Aggregate Spatial Data," Data & Knowledge Engineering, vol. 70, pp. 576-595, 2011.

[66] J. Horey, M. M. Groat, S. Forrestand F. Esponda, “Anonymous Data Collection in Sensor Networks,” Proceedings of 4th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous),pp. 1-8, 2007.

[67] M. M. Groat, B. Edwards, J. Horey, W. He and S. Forrest, “Application and Analysis of Multidimensional Negative Surveys in Participatory Sensing Applications,” Pervasive and Mobile Computing, 2013.

[68] F. Esponda, E. S. Ackley, S. Forrest, and P. Helman, “On-line Negative Databases,” In Artificial Immune Systems, Springer Berlin Heidelberg, pp. 175-188, 2004.

[69] 岩井将行，齋藤修，中嶋紀夫，戸辺義人，瀬崎薫，“モバイル CO2 センサを用いた通勤者自身による生活圏レベル空気計測システム,” 情報処理学会　DICOMO，pp. 1345-1351，2010.

[70] J. S. Lee, B. Hoh, “Dynamic Pricing Incentive for Participatory Sensing,” Pervasive and Mobile Computing, vol. 6, no. 6, pp. 693-708, 2010.

[71] S. Aoki, M. Iwai, and K. Sezaki, “Limited Negative Surveys: Privacy-Preserving Participatory Sensing, ” Proceedings of the 1st IEEE International Conference on Cloud Networking (IEEE CLOUDNET), pp. 158-160, 2012.

[72] 財団法人日本規格協会，“JIS Z 8731 環境騒音の表示・測定方法,” 1999.

[73] “国立環境研究所 環境展望台,” http://tenbou.nies.go.jp/gis/monitor/.

[74] T. Zimmerman and C. Robson, “Monitoring Residential Noise for Prospective Home Owners and Renters,” In Pervasive Computing, Springer Berlin Heidelberg, pp. 34-49, 2011.

[75] E. D’Hondt, M. Stevens, and A. Jacobs, “Participatory noise mapping works! An evaluation of participatory sensing as an alternative to standard techniques for environmental monitoring,” Pervasive and Mobile Computing(2012), doi: 10.10.16/j.pmcj.2012.09.002.

[76] S. Santini, B. Ostermaier, and R. Adelmann, “On the Use of Sensor Nodes and Mobile Phones for the Assessment of Noise Pollution Levels in Urban Environments,” Proceedings of 6th International Conference on Networked Sensing Systems (INSS 2009), pp.1-8, 2009.

[77] N. Maisonneuve, M. Stevens, and B. Ochab, “Participatory Noise Pollution Monitoring using Mobile Phones,” Information Polity, vol. 15, no. 1, pp. 51-71, 2010.

[78] ISO, “ISO 266:1997 acoustics - preferred frequencies,” 1997.

[79] J. A. Hartigan and M. A. Wong, “Algorithm AS 136: A K-Means Clustering Algorithm,” Journal of the Royal Statistical Society, Series C (Applied Statistics), vol. 28, no. 1, pp. 100-108, 1979.

[80] 澤上佳希，岩井将行，瀬崎薫，“異種スマートフォン間の音圧校正手法の提案, ” 電子情報通信学会 技術研究報告，vol. 111，no. 386，USN2011-76，pp. 111-116，2011.

[81] “Audacity,” http://audacity.sourceforge.net/.

[82] 青木俊介，劉広文，岩井将行，瀬崎薫，“ユーザ固有の雑音を考慮する参加型環境センシングのデータ校正手法，” 情報処理学会 研究報告ユビキタスコンピューティングシステム，2013-UBI-38，no. 18，pp. 1-7，2013.

[83] L. G. Jaimes, V. Idalides, and A. L. Miguel, “A Location-Based Incentive Mechanism for Participatory Sensing Systems with Budget Constraints,” Proceedings of IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 103-108, 2012.

[84] “Open Street Map,” http://www.openstreetmap.org/.

[85] A. L. McIlraith and H. C. Card, “Birdsong Recognition using Backpropagation and Multivariate Statistics,” IEEE Transactions on Signal Processing, vol. 45, no. 11, pp. 2740-2748, 1997.

[86] M. Gogala, “Songs of Four Cicada Species from Thailand,” Bioacoustics, vol. 6, no. 2, pp. 101-116, 1995.

[87] L. H. Yang, “Periodical Cicadas as Resource Pulses in North American Forests,” Science, vol. 306, no. 5701, pp. 1565-1567, 2004.

[88] K. S. Williams and C. Simon, “The Ecology, Behavior, and Evolution of Periodical Cicadas,” Annual Reviews of Entomology, vol. 40, no. 1, pp. 269-295, 1995.

[89] S. Amari, A. Cichocki, and H. H. Yang, “A New Learning Algorithm for Blind Signal Separation,” Advances in neural information processing systems, pp. 757-763, 1996.

[90] A. Bertaud and S. Malpezzi, “The Spatial Distribution of Population in 48 World Cities: Implications for Economies in Transition,” 2003, unpublished manuscript.

[91] I. Dagan, L. Lee, and F. C. N. Pereira, “Rapid Detection of Rare Geospatial Events: Earthquake Warning Applications,” Proceedings the 5th ACM International Conference on Distributed Event-Based System, pp.89-100, 2011.

[92] C. Thompson, J. White, B. Dougherty, A. Albright, and D. C. Schmidt, “Using Smartphones to Detect Car Accidents and Provide Situational Awareness to Emergency Responders,” Proceedings of the 3rd International ICST Conference on Mobile Wireless Middleware, Operating systems, and Applications, pp. 29-42, 2010.

[93] I. Dagan, L. Lee, and F. C. N. Pereira, “Similarity-Based Models of Word Cooccurrence Probabilities,” Machine Learning, vol. 34, no. 1-3, pp. 43-69, 1999.

[94] R. S. Hayano and R. Adachi, “Estimation of the total population moving into and out of the 20 km evacuation zone during the Fukushima NPP accident as calculated using ”Auto-GPS” mobile phone data,” Proceedings of the Japan Academy Series B, vol. 89, pp.196-199, 2013.

[95] S. Aoki and K. Sezaki, “Negative Surveys with Randomized Response Techniques for Privacy-aware Participatory Sensing,” IEICE Transactions on Communications, vol. E97-B, No. 04, 2014.

[96] R. E. Mickens, D. N. Brewley and M. Russell,"A Model of Dieting," SIAM Review, vol. 40, no. 3, pp.667-672, 1998.

[97] K. P. Tang, J. Lin, J. I. Hong, D. P. Siewiorek, and N. Sadeh, "Rethinking Location Sharing: Exploring the Implications of Social-Driven vs. Purpose-Driven Location Sharing," Proceedings of the ACM International Conference on Ubiquitous Computing (UbiComp 2010), pp. 85-94, 2010.

[98] S. Aoki and K. Sezaki, "Privacy-Preserving Community Sensing for Medical Research with Duplicated Perturbation," Proceedings of IEEE International Conference on Communications (ICC), 2014.

[99] G. Loukides, A. Gkoulalas-Divanis, and B. Malin, "Anonymization of Electronic Medical Records for Validating Genome-wide Association Studies," Proceedings of the National Academy of Sciences, vol. 107, no. 17, pp. 7898-7903, 2010.

[100] J. Cranshaw, E. Toch, J. Hong, A. Kittur, and N. Sadeh, "Bridging the Gap Between Physical Location and Online Social Networks," Proceedings of the ACM International Conference on Ubiquitous Computing (UbiComp 2010), pp. 119-128, 2010.

[101] E. Toch, J. Cranshaw, P. H. Drielsma, J. Y. Tsai, P. G. Kelley, J. Springfield, L. Cranor, J. Hong, and N. Sadeh, "Empirical Models of Privacy in Location Sharing," Proceedings of the ACM International Conference on Ubiquitous Computing (UbiComp 2010), pp. 129-138, 2010.

[102] E. M. Rogers, "Diffusion of Innovation," Free Press, Social Science, 1962.