

参加型センシングにおけるセンシング誤差を考慮した サンプリングレート制御手法

倉沢 央¹ 佐藤 浩史¹ 山本 淳¹ 川崎 仁史¹ 中村 元紀¹ 松村 一^{1,†1} 山下 靖貴² 鈴木 誠²
森川 博之²

概要: ユーザ参加型センシングは、ユーザが集う時空間領域において必要以上に測定値が集まってトラフィックが大きくなってしまふことや、測定の誤差や多様なセンサの混在によって個々の測定値の信頼性が低いことが懸念される。測定値の信頼性を考慮せずにトラフィック抑制を行うと、誤差が大きく精度の低い移動型センサノードの影響によって、測定結果が不正確になりうる問題があった。これに対して、我々は、センサデータの収集トラフィックを抑制しつつ、時空間領域における測定値の母平均の信頼区間幅を狭めて測定結果の信頼性を確保する、サンプリングレート制御手法を提案する。提案手法は、時空間領域における測定値の信頼区間幅を基準に分布の推定に十分なセンサデータの量を算出し、すべての移動型センサノードに均一な低いサンプリングレートを設定して、集めるセンサデータの量を削減する。実験では、収集対象の移動型センサノードを間引く手法よりも、誤差の影響を抑えられることを確認した。また、測定値の信頼区間幅を指定した大きさ以下にする条件下で、常に一定のサンプリングレートで収集するよりも9分の1にまでトラフィックを抑制できることを確認した。

Adaptive Sampling for Estimating True Values in Participatory Sensing Environment

HISASHI KURASAWA¹ HIROSHI SATO¹ ATSUSHI YAMAMOTO¹ HITOSHI KAWASAKI¹
MOTONORI NAKAMURA¹ HAJIME MATSUMURA^{1,†1} YASUTAKA YAMASHITA² MAKOTO SUZUKI²
HIROYUKI MORIKAWA²

1. はじめに

各種センサを搭載したスマートフォンの急速な普及に後押しされ、参加型センシング [3], [4] が注目を集めている。参加型センシングは、固定型センサノードを実世界上に配備して実世界の様々なデータを測定するのではなく、ユーザが持ち運ぶスマートフォンなどの移動型センサノードを利用して測定するのが特徴である。我々は、これまで固定型センサノードで検討されてきた多様なユビキタスアプリ

ケーションを、参加型センシングによって実現することを考えてきた。

参加型センシングには、固定型センサでは膨大にかかった設置・運用コストを大幅に低減できるというメリットがある [1]。しかしながら、参加型センシングでは、ユーザの保持するセンサを測定に利用するため、ユーザが集う時空間領域において必要以上に測定値が集まって収集トラフィックが大きくなってしまふことや、実際的なアプリケーションで必要となる測定値の信頼性が低いことが懸念される。これらの問題のうち、トラフィック抑制については従来研究で取り組まれてきたが、測定値の信頼性については十分に検討されていない。

我々は測定値の信頼性が低い要因として3つ認識している。1つめは、測定の誤差である。参加型センシングでは、固定型センサで使われるプロユースの専用センサに比べて

¹ 日本電信電話株式会社 NTT 未来ねっと研究所
NTT Network Innovation Laboratories, NTT Corporation,
Musashino, Tokyo 180-8585, Japan

² 東京大学
The University of Tokyo, Meguro, Tokyo 153-8904, Japan

^{†1} 現在, NTT アドバンステクノロジー株式会社
Presently with NTT Advanced Technology Corporation,
Kawasaki, Kanagawa 210-0007, Japan

一般的に誤差の大きいセンサが使われたり、測定環境が劣悪だったりするため、測定する度に生じる誤差が固定型センシングよりも大きくなりうる。2つめは、多様なセンサの混在である。すべてのセンサノードが同じ仕様である保証がなく、偶然誤差や系統誤差の大きさが異なる多様なセンサが混在し得る。各センサノードの精度は不明なことが多い。3つめは、時刻や位置によるセンサデータの量の偏りである。測定の時刻と位置は参加型センシングに寄与するセンサノードの移動軌跡に依存するため、時刻や位置によって集まるセンサデータの量の偏りが生じ得る。

測定値の信頼性を考慮せずにトラフィックを抑制すると、測定値を集約して得られる測定結果が不正確になりうる。例えば、移動型センサノードから集める測定値の量を減らしすぎると、測定値の分布を確率的に推定することが難しくなる。また、測定値の収集対象とする移動型センサノードを間引くと、誤差が大きく精度が低い移動型センサノードを収集対象に選んでしまった場合に、測定値の分布を確率的に推定することがより一層難しくなる。

そこで我々は、参加型センシングにおいて、移動型センサノードからセンサデータを収集するのに必要なトラフィックを抑制しつつ、時空間領域における測定値の母平均の信頼区間幅を狭めて測定結果の信頼性を確保する、サンプリングレート制御手法を提案する。測定値のばらつき具合によって、測定値の母平均の範囲を確率的に推定するのに必要なセンサデータの量は異なる。提案手法はこの点に着目し、時空間領域における測定値の母平均の信頼区間幅を基準に分布の推定に十分なセンサデータの量を算出して、サンプリングレートを設定する。また、この際、すべての移動型センサノードに均一な低いサンプリングレートを設定して、集めるセンサデータの量を削減する。これにより、測定値の分布を確率的に推定するのに十分な量にまでトラフィックを抑制しつつ、誤差が大きく精度の低い移動型センサノードから偏ってセンサデータを収集してしまうことを避けることを実現する。

参加型センシングを模したシミュレーションで評価した結果、提案手法は測定に誤差が生じやすい環境や、誤差の大きさが異なる多様なセンサが混在している環境においても、信頼区間幅以下に抑えつつ、サンプリングレートを適応させてトラフィックを抑制できることを確認した。実験では、収集対象の移動型センサノードを絞らずに低いサンプリングレートに設定したほうが、誤差の影響を抑えられることを確認した。また、測定値から計算される信頼区間幅を指定した大きさ以下に抑える条件下で、すべての移動型センサノードで常に一定のサンプリングレートで収集する場合と比べて9分の1にまで収集するセンサデータの量を削減できた。

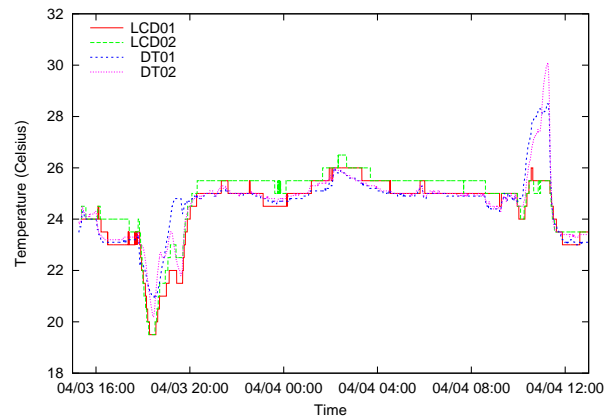


図 1 温度センサ値

Fig. 1 Values of Temperature Sensors

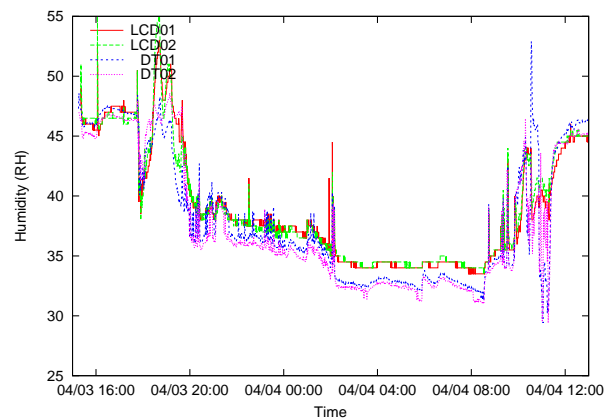


図 2 湿度センサ値

Fig. 2 Values of Humidity Sensors

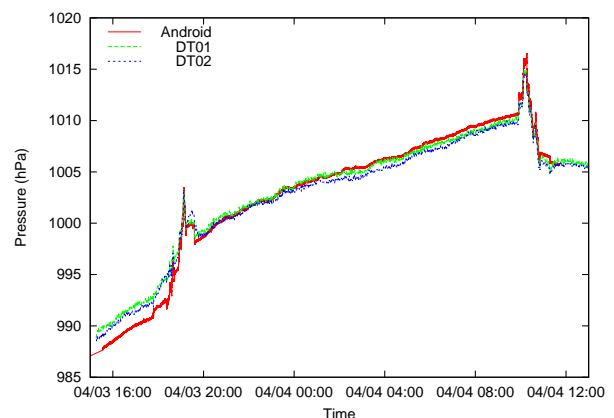


図 3 気圧センサ値

Fig. 3 Values of Pressure Sensors

2. 参加型センシングにおける誤差

測定誤差は、偶然誤差と系統誤差に分類される。偶然誤差は、測定の度に偶然的に起こる誤差である。一方、系統誤差は、同じ条件のもとで測定する限り繰り返し起こる誤差である。

我々は、参加型センシングは固定型センシングよりも、集まったセンサデータに含まれる誤差が大きいと見込んでいる。参加型センシングでは、一般的に固定型センサと比べて安価で汎用的なセンサノードが使われるため、センサノードに由来した偶然誤差や系統誤差が大きいと想定される。

我々は、参加型センシングを模した環境で測定したセンサデータにどの程度の誤差が生じるか検証した。実験には、温度と湿度センサを搭載したデータロガーの EL-USB-2-LCD を 2 台、温度と湿度、気圧センサを搭載したデータロガーの DT-174 を 2 台、気圧センサを搭載したスマートフォンの Galaxy Note II を 1 台使用した。筆者が、合計 5 台を一括して携帯して、各種センサ値を記録した。

図 1 から図 3 はそれぞれ、4 つのセンサノードの温度、4 つのセンサノードの湿度、3 つのセンサノードの気圧値の時間変化を表している。これらの図から、すべてのセンサノードを同じ時刻でかつ同じ位置に携帯していたにも関わらず、異なる値を測定していることは明らかである。この誤差を偶然誤差と系統誤差に分類して検証していく。

厳密に誤差の大きさを求めることは難しいため、誤差の生成と真の値の変化量を単純化して考えた。本実験では、測定値は真の値と偶然誤差、系統誤差の 3 つの和でのみ構成され、偶然誤差の大きさは測定値毎に異なり、一方、系統誤差は各センサノード固有の一定値とした。そして、各センサノードの連続した前後 5 つの測定値の移動平均が、真の値にセンサノード固有の大きさの系統誤差を加えた値に等しいと仮定した。

まず、温度の値を使って、偶然誤差の分布について検証した。連続した前後 5 つの測定値の移動平均を基準として測定値のずれの大きさを計算した。図 4 は EL-USB-2-LCD、図 5 は DT-174 の偶然誤差の分布を表している。センサノードの種類によらずに偶然誤差が生じていて、ほぼ正規分布に一致していることがわかる。標準偏差は EL-USB-2-LCD が 0.160、DT-174 が 0.596 であった。

次に、同様に温度の値を使って、系統誤差の分布について検証した。同時刻の各センサノードの移動平均間のずれの大きさを計算した。図 6 がその結果を表している。偶然誤差と比べて、若干歪んでいる分布となった。これは、センサノードの数が 4 つしかなく、サンプル数が非常に少ないためと考えられる。標準偏差は 1.66 であった。

以上のように、参加型センシングを模した環境で複数のセンサノードを携帯して測定した結果から、偶然誤差や系統誤差を確認できた。参加型センシングにおいて測定結果の信頼性を確保するためには、これらの誤差を起因とした測定値のばらつきを考慮しなければならない。本実験では偶然誤差よりも系統誤差のほうが大きい結果となったが、系統誤差の影響を小さくするには、センサノード 1 つあたりの測定回数を増やすよりも、センサノードの数を増やし

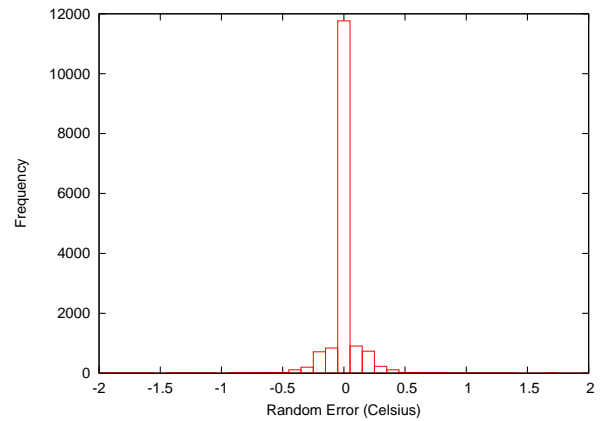


図 4 温度センサの偶然誤差 (EL-USB-2-LCD)
Fig. 4 Random Error of Temperature Sensor (EL-USB-2-LCD)

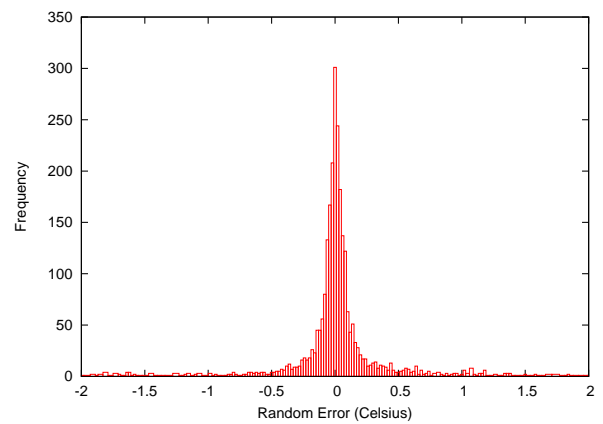


図 5 温度センサの偶然誤差 (DT-174)
Fig. 5 Random Error of Temperature Sensor (EL-USB-2-LCD)

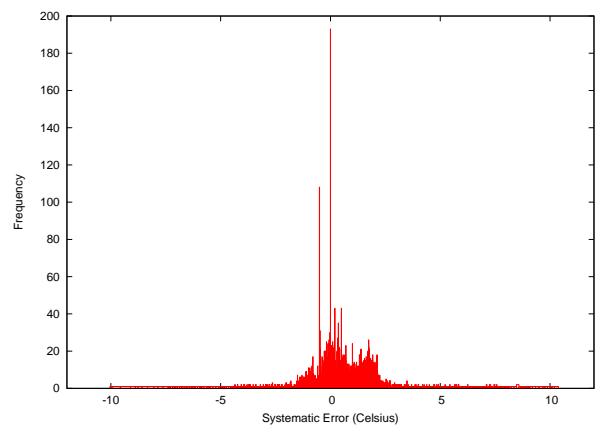


図 6 温度センサの系統誤差
Fig. 6 Systematic Error of Temperature Sensors

たほうが良い。この考えにもとづいて、提案するサンプリングレート制御手法はできるだけ多くの移動型センサノードから測定値を集める設計とした。

3. 関連研究

参加型センシングは、街中へ大量の固定型センサを配置

せずとも各種センサの情報収集が可能なのが最大のメリットである。これまで、タクシーからのプローブデータをもとに交通量の推定をしたり [1], 多様な車から燃費情報を集めて燃費効率の良いルートを検索したり [6], 街なかの騒音マップを構築する [11] といったアプリケーションが検討されている。今後は、従来の固定型センサノードで検討されてきた山火事検知 [8] や橋梁モニタリング [5] といった多様なアプリケーションを、参加型センシングで実現されることが期待されている。

参加型センシングは、よく整備されたセンシング環境を構築するよりも広い範囲を低コストで観測できる利点を持つ反面、センサデータが不十分、不正確、そして、不完全になりうる欠点をもつ [15]。ボランティアのみでのデータ収集では、データが不十分にしか集まらない恐れがある。また、参加型センシングでは固定型センサで使われるプロユースの専用センサに比べて一般的に誤差の大きいセンサが使われるため、不正確なデータになりうる。さらに、いどこで測定されるかは各センサを携帯するユーザの移動や行動に依存するため、欠損のある不完全なセンサデータが集まる。これに対して我々は現在までに、ユーザのセンサデータ提供意欲向上を目的とした多次元階層型集合でのランキング提示手法 [9], [17] や、蓄積された不正確なセンサデータに対して統計的な信頼区間の幅を要求された一定の幅以下になるように時空間領域を統合する手法 [16], 欠損が含まれた不完全なセンサデータに対して欠損部分を精度高く推定する手法 [18] を提案してきた。本研究は、参加型センシングでは不正確な測定値が集まるという問題に、移動型センサノードのサンプリングレートに着目して取り組むものである。

移動型センサノードのトラフィック抑制技術は、移動型センサノードが密に存在するエリアや、測定値が高い相関性を示すエリアにおいて、測定する移動型センサノードや測定値を間引く手法が提案されてきた。文献 [10] では、変化の乏しい時空間領域では少ないセンサデータでも傾向が把握できることに着目し、測定値のばらつき度合いが小さい時空間領域ほど収集対象とする移動型センサノードの数を減らしてトラフィックを抑制している。同様に、文献 [12] では、移動型センサノードが前回の測定値と現在の測定値の差が小さいときは測定値の送信を控えることで、時間変化の乏しい空間で収集される測定値を間引いている。

アドホックネットワークで近傍の移動型センサノードとデータを集約して、基地局への送信トラフィックを抑制する手法も提案されている。文献 [2], [7] では同一の空間領域に属する移動型センサノード間で測定値を共有し、相関性を利用して回帰近似で圧縮している。

しかしながら、これらのトラフィック抑制手法は、測定の誤差や誤差の大きさが異なる多様なセンサの混在を考慮していないため、想定通りの動作は期待できない。文

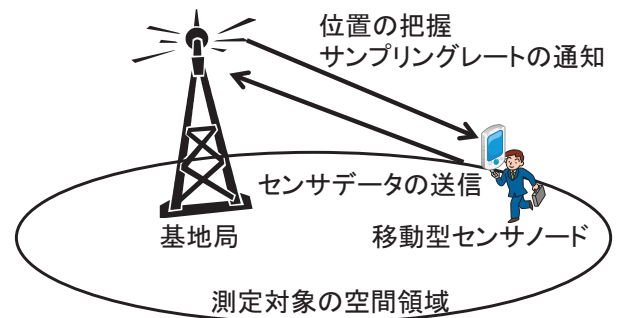


図 7 システムの概要図

Fig. 7 System Architecture

献 [2], [7], [10] の手法は、誤差が含まれた測定値では測定値のばらつき度合いや相関性を正しく算出することが難しい。また、文献 [12] の手法は、同一の移動型センサノードでも偶然誤差によって値がばらつくときは測定値を間引くことが難しい。そこで我々は、参加型センシング環境における誤差を考慮したトラフィックを抑制手法を検討することにした。

4. 誤差を考慮したサンプリングレート制御手法

我々は、参加型センシングにおいて、測定値の収集トラフィックを抑制しつつ、信頼性の高い測定結果を得ることを目的とした、サンプリングレート制御手法を提案する。本手法では、測定結果の信頼性を、測定値の母平均の信頼区間幅の大きさを表すことにした。提案手法は、測定値の母平均の信頼区間幅が常に条件とする大きさ以下になるサンプリングレートに抑制する。

4.1 特徴

提案手法は2つの特徴を有する。

1つめは、測定対象とする時空間領域の測定値のばらつき度合いに応じてサンプリングレートを設定する点である。測定値のばらつき度合いは、時空間領域における測定値の母平均の信頼区間幅を基準とする。例えば、移動型センサノードから得られる測定値のばらつきが少ない時空間領域ではサンプリングレートを低く設定する。この特徴は、測定値の分布を確率的に推定するのに十分な量にまでトラフィックを抑制する効果がある。

もう1つは、すべての移動型センサノードに均一な低いサンプリングレートを設定する点である。つまり、誤差が大きく精度が低い移動型センサノードから偏ってセンサデータを収集してしまうことを避ける。この特徴は、誤差の大きさが異なる多様なセンサの混在の影響を抑えつつ、トラフィックを抑制する効果がある。

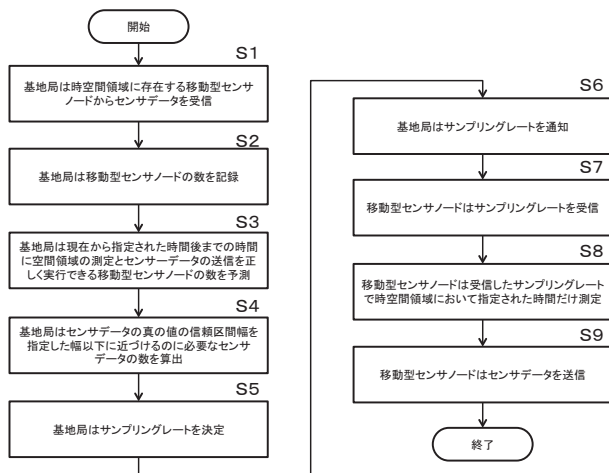


図 8 本方式のフロー
Fig. 8 Procedure

4.2 設計

提案手法は、基地局が移動型センサノードのサンプリングレートを制御する。提案手法では、図 7 に示すように、基地局と移動型センサノードで構成されるシステムを前提としている。基地局は、測定対象となっている時空間領域に存在する移動型センサノードを常に把握を行い、サンプリングレートを通知したり、センサデータを集める役割を担う。移動型センサノードは、基地局から通知されたサンプリングレートで測定して、「測定時刻」と「測定値」を基地局に送信する。例えば、時刻「2012-01-31 09:15:00」に温度「25.4 度」といったような測定結果を送信する。

移動型センサノードは、サンプリングレートと測定時間の初期値が事前に設定する。例えば、0.01Hz のサンプリングレート、および測定時間 10 分というように設定する。

また、基地局は、サンプリングレートを制御するパラメータとして用いる信頼係数と信頼区間幅、時間幅が事前に設定される。例えば、信頼係数 0.95、信頼区間幅 0.4、時間幅 10 分というように設定される。

本方式は 9 つのステップから構成される。図 8 に本方式のフローの詳細を示す。下記に、ステップ S1 から S9 について順番に説明する。

ステップ S1 において、基地局は時空間領域に存在する移動型センサノードからセンサデータを受信する。本方式では、移動型センサノードのサンプリングレートを制御するが、測定時刻は同期することを前提としていない。このため、計測時刻は等間隔になるとは限らない。

ステップ S2 において、基地局は移動型センサノードの数を記録する。図 9 の例は、時間「2012-01-31 09:00」から「2012-01-31 10:20」の 10 分おきの期間で空間領域に存在した移動型センサノードの数を表している。

ステップ S3 において、基地局は、現在からサンプリングレートを制御するパラメータとして設定した時間までに空間領域の測定とセンサデータの送信を正しく実行でき

時間	移動型センサノードの数	サンプリングレート
2012-01-31 09:00	10	0.01
2012-01-31 09:10	13	0.01
2012-01-31 09:20	10	0.01
2012-01-31 09:30	8	0.02
2012-01-31 09:40	6	0.02
2012-01-31 09:50	4	0.05
2012-01-31 10:00	4	0.02
2012-01-31 10:10	6	0.02
2012-01-31 10:20	8	0.01

図 9 移動型センサノードの数とサンプリングレートの例

Fig. 9 Example History of No. of Mobile Sensor Nodes and Sampling Rates

る移動型センサノードの数を予測する。上述した通り、基地局には時間幅「10 分」を設定したので、10 分間に空間領域を測定できる移動型センサノードの数を予測する。提案手法では、現在から指定された時間後までの時間に空間領域の測定とセンサデータの送信を正しく実行できる移動型センサノードの数の予測値は、基地局で記録した過去の移動型センサノードの数をもとに予測する。この予測は、フーリエ変換で算出した時間的周期性を用いた方法や、類似した時間変移があった別の時空間領域の変化をもとにした方法、自己回帰や移動平均によるモデル化をもとにした方法などが有効であると考えられる。さらに、センサノードの数や時空間情報以外にも、気象情報や移動型センサノードの所有者の属性情報のような説明変数を使った方法も有効と考えられる。予測方法の最も単純な例として 1 次微分を使ったものは、時刻 t_i と t_{i-1} の移動型センサノードの数を $x(t_i)$, $x(t_{i-1})$ とすると、時刻 t_{i+1} の移動型センサノードの数の予測値 $x(t_{i+1})$ を、

$$x(t_{i+1}) = x(t_i) + \frac{x(t_i) - x(t_{i-1})}{t_i - t_{i-1}} \times (t_{i+1} - t_i) \quad (1)$$

で求める。この予測方法で、図 9 の時刻「2012-01-31 10:30」の移動型センサノードの数の予測値 $x(2012-01-31 10:30)$ を時刻「2012-01-31 10:20」の移動型センサノードの数 $x(2012-01-31 10:20)$ と時刻「2012-01-31 10:10」の移動型センサノードの数 $x(2012-01-31 10:10)$ から求めると、 $x(2012-01-31 10:30) = 8 + 2 = 10$ が求まる。時刻「2012-01-31 10:20」から時刻「2012-01-31 10:30」にかけて、移動型センサノードの数が 8 から 10 に増えるという予測値が得られた。

ステップ S4 において、基地局はセンサデータの真の値の信頼区間幅を指定した幅以下に近づけるのに必要なセンサデータの数を算出する。基地局はまず、移動型センサノードごとに測定値の平均値を計算する。例えば、時刻「2012-01-31 10:20:00」から「2012-01-31 10:30:00」の間に、

移動型センサノードから 25.449, 25.364, 25.663, 25.542, 25.535, 25.692 のセンサデータが収集されていたとする*1。上述の例において、基地局には信頼係数 0.95, 信頼区間幅 0.4, 時間幅 10 分を設定していた。この 6 つの値の標準偏差は 0.12451 である。標本数 6 なので、信頼度 95% のときの信頼区間は、統計学における定義を用いると、25.441 から 25.640 の範囲と計算される。この信頼区間幅は 0.199 である。信頼区間幅を 0.4 以下にする条件であったのに対してこの信頼区間幅はほぼ 0.2 であるため、必要なセンサデータは現在の 4 分の 1 の数で十分であると算出される。

ステップ S5 において、基地局のサンプリングレート決定部はサンプリングレートを決定する。ステップ S3 では移動型センサノードの数は $x(2012-01-31\ 10:20)$ から $x(2012-01-31\ 10:30)$ で 8 から 10 に増えると予測され、ステップ S4 では信頼区間幅を「0.4」以下にする条件ならばセンサデータは現在の 4 分の 1 の数で十分だとわかった。また、時刻「2012-01-31 10:20」のサンプリングレート $R(2012-01-31\ 10:20)$ は、図 9 によると「0.01」と設定されていた。以上の条件から、時刻「2012-01-31 10:30」のサンプリングレート $R(2012-01-31\ 10:30)$ は、

$$\begin{aligned}
 & R(2012-01-31\ 10:30) \\
 &= R(2012-01-31\ 10:20) \\
 &\times (\text{移動型センサノードの増加率の逆数}) \\
 &\times (\text{信頼区間幅の条件を満たすための標本数の増加率}) \\
 &= 0.01 \times \frac{8}{10} \times \frac{1}{4} = 0.002 \quad (2)
 \end{aligned}$$

と求める。

ステップ S6 において、基地局の送信部はサンプリングレートを送信する。ステップ S5 で算出された 0.002Hz を新しいサンプリングレートとして、移動型センサノードに送信する。

ステップ S7 において、空間領域に位置する移動型センサノードはサンプリングレートを受信し、ステップ S8 において、移動型センサノードの観測部は受信したサンプリングレートで指定された時間だけ測定する。つまり、時刻「2012-01-31 09:15」までは 0.01Hz で 10 分間だけ測定していたのを、0.002Hz で 10 分間だけ測定するように変更する。なお、空間領域に途中から入ってくる移動型センサノードに対しては、入ってきたタイミングでサンプリングレートを通知する。

さらに、ステップ S9 について説明する。移動型センサノードの送信部はセンサデータを送信し、ステップ S1 からの手順を繰り返す。

以上の手順で、基地局は、現在から指定された時間後ま

*1 本説明では簡単のため、測定されたセンサデータの数をもとにしたが、時刻「2012-01-31 10:20:00」の移動型センサノードの数が 8 でサンプリングレートは 0.01Hz のとき 10 分間に測定されるサンプル数は 6 を上回るだろう。

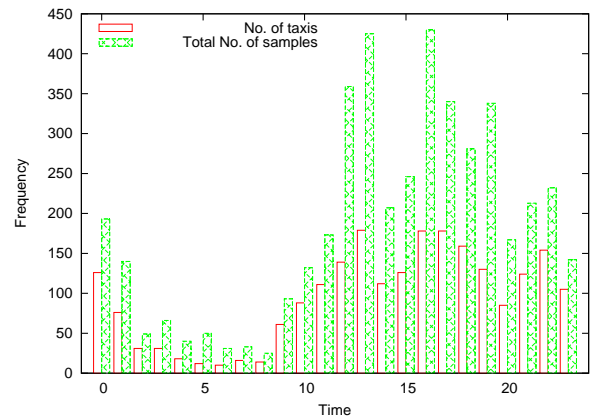


図 10 タクシーの台数と総サンプル数

Fig. 10 No. of Taxis and Total No. of Samples

での期間に時空間領域に居合わせると予測される移動型センサノードの数や、集まる測定値のばらつき度合いに応じた、測定のサンプリングレートを、収容するすべての移動型センサノードに対して通知することが可能となる。

5. 評価

提案手法が、測定結果の信頼性を確保しながら移動型センサノードからセンサデータを収集するのに必要なトラフィックを抑制できるかをシミュレーション実験で評価した。実験では、基地局が指定した区画に居合わせた移動型センサノードにサンプリングレートを通知し、移動型センサノードは基地局に温度の測定値を送信する状況を想定した。

5.1 シミュレーション設定

参加型センシングの環境を模したシミュレーションとするために、指定した区画に居合わせた移動型センサノードの数の時間変移、測定値の真の値の時間変移、測定値の偶然誤差と系統誤差、比較手法は以下のように設定した。

移動型センサノードの数の時間変移は、北京市内のタクシーの移動経路データセットの T-Drive Data[13], [14] を使って設定した。T-Drive Data は、北京市内を走行する 2008 年の 2 月 2 日から 8 日の期間に、10,357 台のタクシーに装着した GPS の測定値で構成されたデータセットである。GPS の測定頻度は一律ではないが、平均で 177 秒おきで 623 メートルの間隔で位置情報が取得されている。すべての測定値には位置情報の他、タクシーの ID も付与されている。本実験では、2008 年 2 月 3 日の 0 時から 23 時に北緯 39.90 度から 39.91 度、東経 116.39 度から 116.40 度の区画に居合わせたタクシーの ID を利用した。参考までに、この区画は天安門広場に面した大通りのおよそ 1 キロメートル平方である。図 10 は、この区画に居合わせたタクシーの台数と収集された GPS の測定値の総数を表したものである。横軸は時刻を表している。この図から、早朝

の時間帯のタクシーが少ないことがわかる。

測定値の真の値の時間変移は、気象庁観測の気温を使って設定した。上述の北京の区画の気象データを得ることができなかったため、同日の2008年2月3日の0時から23時に東京（北緯35度41.4分，東経139度45.6分，標高6.1メートル）で観測された1時間おきの気温を使った。提案手法は測定時刻の同期をせずにサンプリングレートだけを共有する方式のため、シミュレーション実験では任意の時刻の気温の真の値を必要とする。我々は任意の時刻の真の値を、近傍2時刻の気温の線形補間で計算した値とした。

各移動型センサノードの測定値は、上述の東京で観測された気温に偶然誤差と系統誤差を加えた値とした。偶然誤差と系統誤差の大きさは、第2章の実験で得た誤差の分布を参考に設定した。この実験では、偶然誤差の標準偏差は0.160と0.596，系統誤差の標準偏差は1.66であった。系統誤差のほうが偶然誤差よりも大きい値であった。そこで、系統誤差のほうが偶然誤差よりも平均的に大きくなるように、タクシーのID（移動型センサノード）ごとに誤差のパラメータを定めた。具体的には、偶然誤差のパラメータは0から0.8の範囲の一様分布で、系統誤差は0から2の範囲の一様分布で設定した。偶然誤差はこのパラメータを標準偏差とする正規分布で測定の際に生成した。系統誤差はこのパラメータを標準偏差とする正規分布でタクシーIDに対して1つ生成した。

参加型センシングにおいて精度を考慮したデータ収集手法は、筆者らの知る限り、過去に存在しない。そこで、一定のサンプリングレートで測定した場合の信頼区間幅と比較した。サンプリングレートは、測定値の信頼区間幅が常に提案手法が条件とする大きさ以下になるように設定した。これをUniform samplingと呼ぶ。さらに、参考までに、1時間おきに区画に居合わせた移動型センサノードの中からランダムに1つを選んで測定した場合の結果とも比較した。これをOne sampleと呼ぶ。

提案手法のパラメータとして必要な信頼係数は0.95に、時間幅は1時間に、センシング時間と基地局がサンプリングレートを通知する頻度はともに1時間に設定した。移動型センサノードの数の予測に関しては、簡単のため本シミュレーションでは省略して、正しい移動型センサノードの数をシミュレータに与えることとした。

本シミュレーション実験では、サンプリングレートの通知に必要なトラフィックは評価から除いている。通知手段に関しては今後の課題として認識しており、「エリアメール」のように一斉同報のプロトコルなどが有効であると考えている。

5.2 信頼区間幅

提案手法が、母平均の95%信頼区間幅が0.5度以内を満たすようにサンプリングレートを制御した場合の、気温の

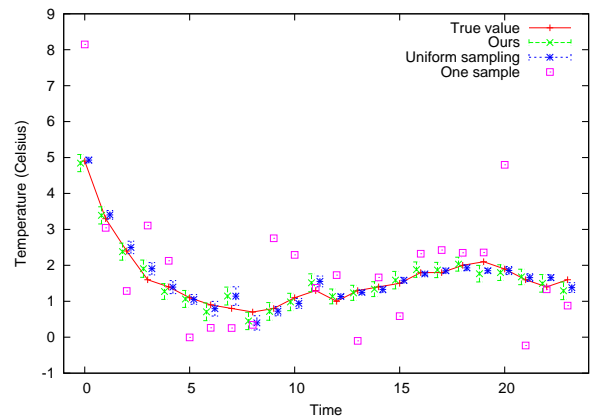


図 11 気温の測定値の信頼区間

Fig. 11 Confidence Interval of Temperature

測定値の信頼区間を表したものが図 11 である。縦軸が温度を、横軸が時刻を表している。真の値、提案手法、Uniform sampling、そして、One sample の結果を表している。

One sample の結果は、誤差の影響によって真の値から大きく外れた測定値が混在していた。一方、提案手法とUniform sampling は比較的正しい測定結果であった。提案手法とUniform sampling は、測定する移動型センサノードの数を絞らず、すべての移動型センサノードに均一な低いサンプリングレートを設定することで、系統誤差の影響を抑えることができたと考えられる。つまり、誤差が大きく精度の低い移動型センサノードから偏ってセンサデータを収集してしまうことを避けることができたといえる。

提案手法とUniform sampling を比較すると、例えば22時の結果のように、Uniform sampling よりも提案手法の総測定回数大幅に下回るような時間帯において、Uniform sampling の信頼区間幅の結果が小さくなっているが、真の値に近づいていなかった。これは以下のように解釈できる。サンプリングレートが低いと、移動型センサノードが測定するタイミングに該当区画に居合わせる確率が小さくなり、結果として測定する移動型センサノードの数が小さくなる。一方、ある程度サンプリングレートを高くしていき、測定する移動型センサノードの数が上限に達すると、1つの移動型センサノードで測定する回数が増加する。前者の状況では、測定する移動型センサノードの数が変化するため、偶然誤差と系統誤差の両方の影響を抑える効果が期待できる。しかし、後者の状況では、測定する移動型センサノードの数は変化しないため、偶然誤差の影響のみを抑える効果となる。Uniform sampling では、後者の状況となっていて、提案手法よりもサンプリングレートが高くなっているにも関わらず精度がさほど向上していなかったと考えられる。

5.3 サンプリングレートと総測定回数

提案手法が、母平均の95%信頼区間幅が0.5度以内を

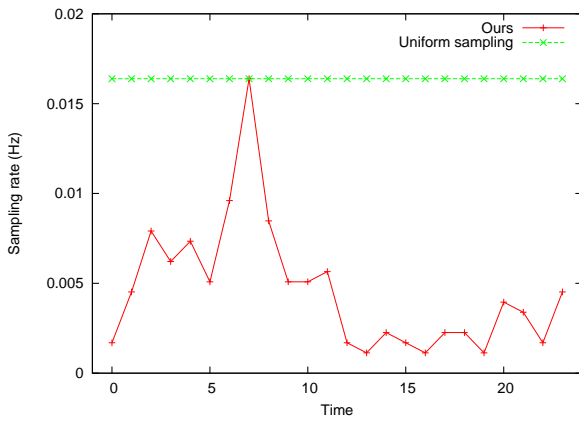


図 12 サンプリングレート
Fig. 12 Sampling Rate

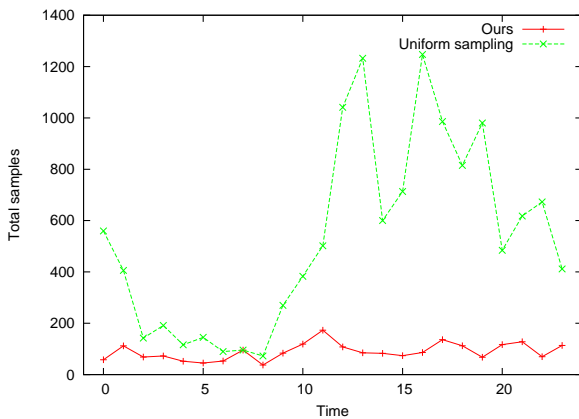


図 13 総測定回数
Fig. 13 Total Samples

満たすようにサンプリングレートを制御した場合の、サンプリングレートと総測定回数を表したものが図 12 と図 13 である。横軸は時刻を表している。提案手法と Uniform sampling の結果を表している。

図 12 から、提案手法は Uniform sampling と比べて大幅にサンプリングレートを抑制できていることがわかる。最も効果的な時刻のサンプリングレートでは、提案手法は Uniform sampling のおよそ $\frac{1}{10}$ にまで抑制できている。平均的にも $\frac{2}{7}$ にまで抑制できている。また、図 13 から、提案手法の総測定回数は、区画に居合わせたタクシーの台数に関係なく、ほぼ一定の値となっていることがわかる。総測定回数は、平均で提案手法は Uniform sampling のおよそ $\frac{1}{9}$ にまで抑制できていた。以上より、本研究の目的としていた、センサデータの精度を保証しながら、移動型センサノードからセンサデータを収集するのに必要なトラフィックを抑制するようなサンプリングレートの制御ができているとわかる。

図 10 と図 12 を比較すると、必ずしもサンプリングレートは移動型センサノードの数（タクシーの台数）の逆数になっているわけではないが、これは、提案手法は、単純に

区間に居合わせた移動型センサノードの数だけでなく、測定値の分散度合いにも適応してサンプリングレートを定めているためであると考えられる。

6. おわりに

本稿では、参加型センシングにおいて、測定値の母平均の信頼区間幅が常に条件とする大きさ以下になるサンプリングレートに制御して、移動型センサノードからセンサデータを収集するのに必要なトラフィックを抑制する、サンプリングレート制御手法を提案した。提案手法は、参加型センシングで集まるセンサデータの信頼性の確保とトラフィック抑制を実現し、実世界情報を活用するアプリケーションで参加型センシングを利用しやすくするものである。

現在の手法は測定値の母平均がどのような範囲の値をとるか推定するものであるが、将来的には真の値を推定するように改善したい。厳密に真の値を推定するには、真の値の時空間領域における分布を考えた上で、さらに移動型センサノードごとの偶然誤差と系統誤差を推定する必要がある。他にも、将来の検討課題として、基地局と移動型センサノードで構成されるネットワークの設計や移動型センサノードへのサンプリングレートの通知方法といった実装面での課題が残っている。現在、実環境での評価を進めている。

参考文献

- [1] Aslam, J., Lim, S., Pan, X. and Rus, D.: City-scale traffic estimation from a roving sensor network, *Proceedings of the 10th international conference on Mobile systems, applications, and services (MobiSys '12)*, pp. 141–154 (2012).
- [2] Banerjee, T., Chowdhury, K. and Agrawal, D. P.: Tree based data aggregation in sensor networks using polynomial regression, *Proceedings of the 8th international conference on Information Fusion*, pp. 25–29 (2005).
- [3] Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S. and Srivastava, M. B.: Participatory sensing, *Proceedings of the Mobile Device Centric Sensor Networks and Applications (WSW '06)*, pp. 117–134 (2006).
- [4] Campbell, A. T., Eisenman, S. B., Lane, N. D., Miluzzo, E., Peterson, R. A., Lu, H., Zheng, X., Musolesi, M., Fodor, K. and Ahn, G.: The Rise of People-Centric Sensing, *IEEE Internet Computing*, Vol. 12, No. 4, pp. 12–21 (2008).
- [5] Chebrolov, K., Raman, B., Mishra, N., Valiveti, P. K. and Kumar, R.: Brimon: a sensor network system for railway bridge monitoring, *Proceedings of the 6th international conference on Mobile systems, applications, and services (MobiSys '08)*, pp. 2–14 (2008).
- [6] Ganti, R. K., Pham, N., Ahmadi, H., Nangia, S. and Abdelzaher, T. F.: GreenGPS: a participatory sensing fuel-efficient maps application, *Proceedings of the 8th international conference on Mobile systems, applications, and services (MobiSys 2010)*, pp. 151–164 (2010).
- [7] Guestrin, C., Bodi, P., Thibau, R., Paski, M. and Madden, S.: Distributed regression: an efficient framework

- for modeling sensor network data, *Proceedings of the 3rd international symposium on Information processing in sensor networks (IPSN '04)*, pp. 1–10 (2004).
- [8] Hartung, C., Han, R., Seielstad, C. and Holbrook, S.: FireWxNet: a multi-tiered portable wireless system for monitoring weather conditions in wildland fire environments, *Proceedings of the 4th international conference on Mobile systems, applications and services (MobiSys '06)*, pp. 28–41 (2006).
- [9] Kawasaki, H., Yamamoto, A., Kurasawa, H., Sato, H., Nakamura, M. and Matsumura, H.: Top of Worlds: Method for Improving Motivation to Participate in Sensing Services, *Proc. Ubicomp 2012*, pp. 594–595 (2012).
- [10] Mendez, D.: Density Maps: Determining Where to Sample in Participatory Sensing Systems, *Proceedings of 2012 Third FTRA International Conference on Mobile, Ubiquitous, and Intelligent Computing (MUSIC)*, pp. 35–40 (2012).
- [11] Rana, R. K., C. T. Chou, S. S. K. and N. Bulusu, W. H.: Ear-phone: an end-to-end participatory urban noise mapping system, *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN 2010)*, pp. 105–116 (2010).
- [12] Sharaf, M. A., Beaver, J., Labrinidis, A. and Chrysanthis, P. K.: TiNA: a scheme for temporal coherency-aware in-network aggregation, *Proceedings of the 3rd ACM international workshop on Data engineering for wireless and mobile access (MobiDe '03)*, pp. 69–76 (2003).
- [13] Yuan, J., Zheng, Y., Zhang, C., Xie, W., Sun, G. and Huang, Y.: T-drive: driving directions based on taxi trajectories, *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '10)*, pp. 99–108 (2010).
- [14] Yuan, J., Zheng, Y., Xie, X. and Sun, G.: Driving with knowledge from the physical world, *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '11)*, pp. 316–324 (2011).
- [15] 佐藤浩史, 山本淳, 倉沢央, 川崎仁史, 中村元紀, 松村一: 仮想センサ環境の実現へ向けて — 参加型センシングの可能性と課題, 信学総大 (2013).
- [16] 山本淳, 川崎仁史, 倉沢央, 佐藤浩史, 中村元紀, 松村一: 参加型センシングのためのデータ精度を考慮した仮想センサの提案, 第11回情報科学技術フォーラム (FIT 2012) 論文集, Vol. 4, pp. 17–20 (2012).
- [17] 川崎仁史, 山本淳, 倉沢央, 佐藤浩史, 中村元紀, 松村一: 多次元階層型集合でのランキング提示によるサービスへの参加意欲向上手法, 信学総大, pp. 654–654 (2012).
- [18] 倉沢央, 佐藤浩史, 山本淳, 川崎仁史, 中村元紀, 松村一: 不完全センサデータの重回帰における変数選択手法, 信学技報, USN2012-54 (2012).