

修士論文

クラウドセンシングにおける
ノード選択手法の研究

A Study on Node Selection Scheme for Crowd Sensing

2015年2月5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 48-136416

坂本 敬太

内容梗概

一般ユーザの持つスマートフォンに代表されるモバイルデバイスを用いてセンシングを行うクラウドセンシング（参加型センシング）において収集されるデータの質やカバレッジはそのデータの利用価値に直結するため非常に重要だが、それらは参加ユーザに依存してしまうという問題がある。そのため、データ収集の際に質やカバレッジを考慮したノード選択を行う必要がある。本研究では、データの質・位置・収集コストのそれぞれがバラバラなデータ群に対して、総コスト一定の条件下でデータの質・カバレッジを総合的に評価した最適なノード選択を行う手法を提案し、シミュレーション、および実データによる評価を行う。この手法により、クラウドセンシングだけでなく、クオリティコントロールが求められる様々な分野においてより利用価値の高いデータ収集の実現が期待できる。

目次

第1章	序論	1
1.1	はじめに	1
1.2	本論文の構成	1
第2章	研究背景	3
2.1	クラウドセンシングの概要	3
2.2	クラウドセンシングの課題	4
2.2.1	インセンティブの確保	4
2.2.2	データの質	4
2.2.3	データのカバレッジ	4
2.3	本章のまとめ	5
第3章	関連研究	6
3.1	概要	6
3.2	ユーザのインセンティブに関する研究	6
3.2.1	RADP-VPC	6
3.2.2	インセンティブに関するその他の関連研究	8
3.3	データのカバレッジに関する研究	9
3.3.1	GBMC アルゴリズム	10
3.3.2	Centralised Management	11
3.4	データの質に関する研究	12
3.4.1	Micro Payments	12
3.4.2	Steered Crowdsensing	13
3.5	本章のまとめ	14
第4章	データの質の定量化	18
4.1	概要	18
4.2	センサの性能を利用したデータの質の定量化	18
4.3	データの位置情報を利用したデータの質の定量化	18
4.4	ユーザの評価を利用したデータの質の数値化	19
4.5	本章のまとめ	20
第5章	ノード選択アルゴリズム	21
5.1	概要	21
5.2	データの質とカバレッジの両面を考慮したノード選択アルゴリズム	21
5.3	ユーザの移動を考慮したノード選択アルゴリズム	23

第6章	評価	24
6.1	概要	24
6.2	シミュレーションデータを用いた評価	24
6.2.1	シミュレーション環境	24
6.2.2	シミュレーション結果	24
6.3	実データを用いた評価	31
6.3.1	実験環境	31
6.3.2	動的なノード選択を行った場合	33
6.3.3	静的なノード選択を行った場合	34
第7章	結論	39
7.1	まとめ	39
7.2	今後の展望	40
	謝辞	41

目次

2.1	クラウドセンシング（参加型センシング）の概要	3
2.2	報酬金によるインセンティブの確保	4
2.3	クラウドセンシングのデータに関する課題	5
3.1	RADP-VPC における逆オークション方式	7
3.2	通常の逆オークションと RADP-VPC の比較の一例	7
3.3	RADP-VPC と RSFP のシミュレーション結果	8
3.4	Weather News	9
3.5	OpenStreetMap	10
3.6	GBMC アルゴリズムにおけるノードの決定法	10
3.7	GBMC アルゴリズム未適用時のシミュレーション結果	12
3.8	GBMC アルゴリズム適用時のシミュレーション結果	13
3.9	Centralised Management におけるカバレッジの評価	15
3.10	Micro Payments における写真データの収集	16
3.11	Micro Payments における質の評価	16
3.12	Steered Crowdsensing の概要	16
3.13	Steered Crowdsensing でユーザに配布されるアプリ	17
3.14	Steered Crowdsensing の実験結果	17
4.1	分解能とデータの質の対応の例	19
4.2	データの質の定量化	20
5.1	提案手法における評価値	21
5.2	提案手法におけるノード選択	22
5.3	ユーザの移動を考慮した評価値	23
6.1	ノード選択の結果：提案手法	25
6.2	ノード選択の結果：GBMC アルゴリズム	26
6.3	ノード選択の結果：質のみ重視	27
6.4	ノード選択の結果：ランダム	28
6.5	カバレッジと質の平均値のシミュレーション結果	29
6.6	重ね合わせの体積のシミュレーション結果	30
6.7	人の流れデータから作成した人の分布状況	32
6.8	今回対象とした範囲	32
6.9	動的なノード選択を行った場合のカバレッジの時間推移の平均の比較結果	34
6.10	動的なノード選択を行った場合の質の時間推移の平均の比較結果	35
6.11	動的なノード選択を行った場合のカバレッジと質の平均の比較結果	36
6.12	静的なノード選択を行った場合のカバレッジの時間推移の平均の比較結果	37

6.13 静的なノード選択を行った場合のカバレッジと質の平均の比較結果	38
---	----

表 目 次

6.1	シミュレーションのパラメータ設定	25
6.2	シミュレーション結果のまとめ	31
6.3	実験のパラメータ設定	33
6.4	動的なノード選択を行った場合の実験結果のまとめ	37
6.5	静的なノード選択を行った場合の実験結果のまとめ	37

第1章 序論

1.1 はじめに

近年，スマートフォンの急速な普及が進んでいる．技術の進歩によって，その小さな端末には多くの機能が詰め込まれており，人々の生活を豊かにしている．特に，センサー類は加速度センサ・デジタルコンパス・ジャイロセンサ・気圧センサ・GPS 受信機など多種多様であり，様々なアプリケーションに利用されている [1]．中でも，一般のユーザがスマートフォン内蔵センサで収集したデータを統合・解析し，環境モニタリングや交通状況の監視，ヘルスケアや世論調査など幅広い分野に利用するクラウドセンシングが期待されている [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]．クラウドセンシングは参加型センシングとも呼ばれ，従来のセンサネットワークシステムと比較してセンサ設置等のコストを必要とせずに一般のユーザを社会インフラとして利用できるという大きな利点がある．また，人間がデータを送信することから機械のセンサだけでは判断が難しいような情報を収集することが出来るというのも利点である．

しかし，クラウドセンシングには従来のセンサネットワーク技術に比べていくつかの課題が存在する [17]．一定以上の参加ユーザ数を保つためのインセンティブを確保しなければならないという課題 [4, 5] や，データ収集が参加ユーザに委ねられているがゆえにデータの質やカバレッジが安定しないという課題 [3, 5] が代表的である．特にデータの質やカバレッジはそのデータの利用価値に直結するものであり，場合によってはせっかく収集したデータのアプリケーション等への利用が不可能になってしまうことから非常に重要な課題として先行研究でも議論されている [3, 5, 15]．データの質やカバレッジを安定させるには，データ収集の対象とするノード（ユーザ）を取捨選択する方法が考えられる．また，クラウドセンシングでは参加ユーザに報酬を与える事でインセンティブの確保を行うことが一般的であるため [4, 7, 16]，データの質とカバレッジだけではなくコスト面でも効率的なノード選択手法が望ましいといえる．

本研究ではまずクラウドセンシングにおけるユーザのインセンティブやデータのクバレッジ，データの質に関する研究として RADP-VPC[4]，GBMC アルゴリズム [5]，Centralised Management[15]，Micro Payments[7]，Steered Crowd Sensing[16] などの紹介をして，クラウドセンシングにおけるデータのクバレッジや質の重要性について議論した後，データの質・クバレッジの両面を考慮した新たなノード選択手法の提案を行う．そして提案手法のアルゴリズムについて解説した後，シミュレーションと実データによる評価を行い，結論を述べる．

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、本研究の対象となるクラウドセンシングについてその概要と課題について解説を行う。

次に、第3章では第2章でまとめたクラウドセンシングの三つの課題について、それぞれに関連した研究を紹介し、問題点を明らかにする。

第4章、第5章では提案手法となるデータのカバレッジ・質の両面を高めるようなノード選択手法について、その下準備となるデータの質の定量化とノード選択アルゴリズムのそれぞれについて検討する..

第6章ではシミュレーションデータ、実データのそれぞれに手法を適用し、比較実験を行い、それぞれについて結果を示す。

最後に、第7章にて全体をまとめ、今後の展望を述べる。

第2章 研究背景

本章では、研究背景として参加型センシングの概要と、その課題について述べる。

2.1 クラウドセンシングの概要

半導体技術の進歩によって、超小型かつ低消費電力のCPUモジュール、無線通信モジュール、センサモジュールの開発が可能になった。これらのモジュールの出現によって生み出されたのがセンサネットワークである [18]。センサネットワーク技術は、センサを有する小型コンピュータが無線によって相互に通信を行いながら実空間上の情報をサイバー空間に取り込むための技術である [1, 18, 19] が、その中でも近年のスマートフォンの急速な普及を受けて期待されているのがクラウドセンシングである [1, 20]。

クラウドセンシングとは、スマートフォンユーザにセンシングへの参加を促し参加ユーザから集めた大量かつ広範囲のデータを統合・解析して様々なアプリケーションに利用するセンサーネットワーク技術である (図 2.1)。センサー類はスマートフォンに標準装備されているため、従来のセンサーネットワーク技術に比べてセンサー設置のコストが必要ないという利点がある [4]。また、スマートフォン本体によって電源が確保されているため、センサネットワークの大きな課題である消費電力の心配が不要であるという点も利点である [18]。しかし、いくつかの課題を抱えている。

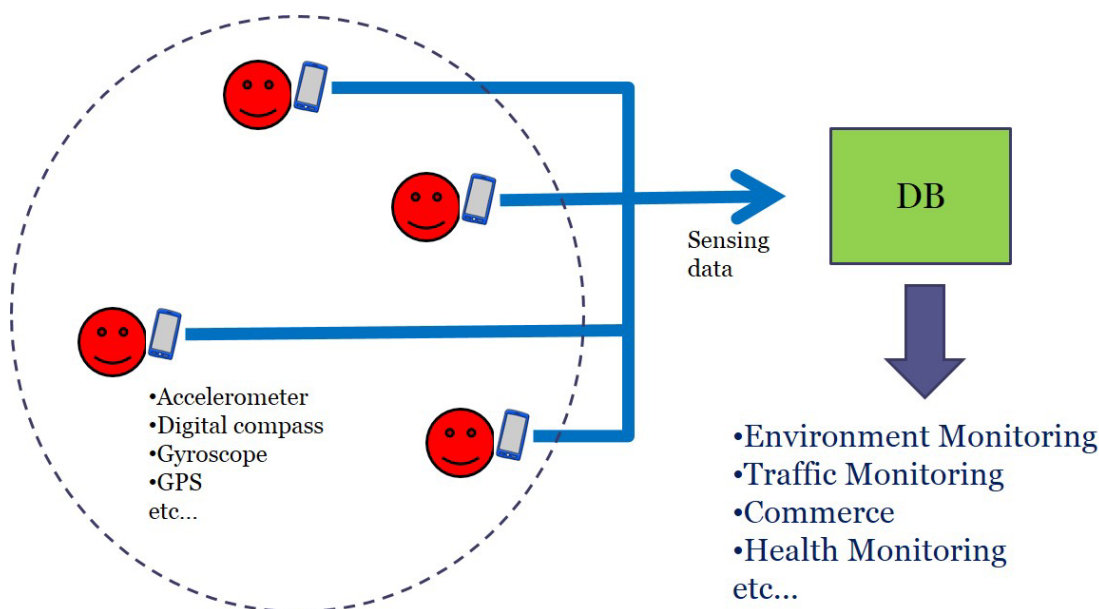


図 2.1: クラウドセンシング (参加型センシング) の概要

2.2 クラウドセンシングの課題

クラウドセンシングの課題について述べる。

クラウドセンシングにおいて課題となるのは大きく分けてインセンティブの確保、データの質、データのカバレッジ、の三点である。

2.2.1 インセンティブの確保

クラウドセンシングへの参加は、ユーザにとって様々な負担を伴う。まず挙げられるのがバッテリーやリソースの消費といったモバイルデバイスへの物理的な負担である。また、送信するデータには位置情報に代表される個人情報に該当するデータが含まれる場合がほとんどであるため、参加ユーザはプライバシー面でのリスクも負うことになる。以上のことから、参加へのモチベーションを上げる仕組みを取り入れ、インセンティブを確保することが必要であると考えられる。多くのクラウドセンシングは報酬金によってインセンティブの確保を行っている（図 2.2）が、歩合制 [7]、逆オークション制 [4, 5]、ランキング制 [16]、クーポン制 [9] とその支払い形態は多岐にわたっており、まだまだ議論の余地があるといえる。



図 2.2: 報酬金によるインセンティブの確保

2.2.2 データの質

クラウドセンシングにおいてデータを収集するのは参加ユーザであり、プロバイダ側では送られてくるデータの品質の管理が難しいという問題がある。具体的にはセンサーの精度に問題があったり、人為的なミスや悪意のあるユーザによって誤ったデータが送られてきたりといったようなことが考えられる。勿論、収集するデータの種類や適用先のアプリケーションの形式によって注意すべき点は様々であるが、ほとんどのケースにおいてデータの質を安定させるための何らかの措置は必要であると考えられる。

2.2.3 データのカバレッジ

データの質と同じく、収集されるデータのカバレッジが参加ユーザの現在位置に依存してしまうという問題がある。クラウドセンシングにおいて無作為にデータ収集を行った場合、人が集まりやすい場所（大きな町や駅など）にデータが偏ってしまうことが予想され

るが、多くの場合データの地理的なカバレッジの大きさはアプリケーションのカバレッジの大きさに直結するため非常に重要な要素となるため、対策が必要になる。関連研究ではノードの取捨選択や報酬の重み付けによる誘導などによってカバレッジの改善を図っている。

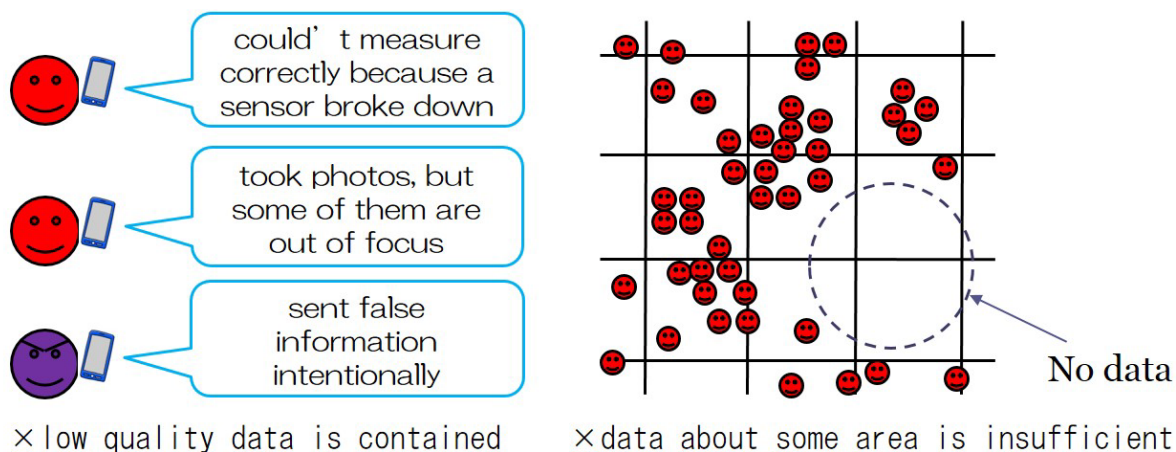


図 2.3: クラウドセンシングのデータに関する課題

2.3 本章のまとめ

本章では、クラウドセンシングの概要とそれが抱える課題として、インセンティブの確保、データの質、データのカバレッジについて解説を行った。

インセンティブの確保はクラウドセンシングの存続に欠かせないものであり、データの質、データのカバレッジは収集したデータの利用価値を左右する重要なものである。これらの課題をうまく緩和していくことでクラウドセンシングはより実用的なものになり、従来のセンシング技術に比べて主にコスト面で非常に効率的なセンシングの実現が期待できる。次章ではこれらの課題に対する理解を深めるため、関連する研究を複数紹介していく。

第3章 関連研究

3.1 概要

本章では、前章で述べたクラウドセンシングにおける三つの課題である、ユーザのインセンティブ、データのカバレッジ、データの質、の三点についてそれぞれに関する検討を行っている研究について各項目毎に紹介をする。最後に、紹介した関連研究をまとめ、今後の課題を明らかにする。

3.2 ユーザのインセンティブに関する研究

本節では、クラウドセンシングの存続に必要なユーザのインセンティブに関する研究を RADP-VPC[4] を中心に紹介する。

3.2.1 RADP-VPC

ほとんどのクラウドセンシングにおいて、インセンティブの確保は参加ユーザへの報酬金によって行われている [4, 5] が、その中でも逆オークションの方式を用いた関連研究として、RADP-VPC(Reverse Auction based Dynamic Pricing with Virtual Participation Credit)[4] を紹介する。

逆オークションとは参加ユーザが自分のデータに値段を付け、プロバイダはその中から必要なデータを選んで提示された競り値で競り落とすという方式である (図 3.1)。ユーザが自分のデータに付ける価値というのは本来プロバイダが知りえない情報であるが、それをオークションの過程で知ることができるのが逆オークション型の利点であるといえる。また、ユーザ同士で競争が起こることによってデータの競り値の低下が見込めるのも利点である。

RADP-VPC は通常の逆オークション手法に VPC(Virtual Participation Credit) と呼ばれるものを適用した手法である。通常の逆オークション手法では単純に低い価格を付けたユーザが勝つため、低い価格を付けにくいユーザ (参加への負担が大きいと考えるユーザ) が勝つことは難しく、結果勝てないユーザが離れて行ってしまうことになり、ユーザ数の確保が難しくなってしまう。ユーザ数の低下によって、収集できるデータ量の減少や競り競争の鈍化による競り値上昇などの問題が起こると考えられる。

VPC はこの問題を解決するために導入されたものである。

VPC はユーザ毎に管理される仮想の貨幣であり、オークションの参加に対する報酬のようなもので、 v を VPC、 r をオークションのラウンド数、 i を任意のユーザ、 α を定数とした時、以下のように変動する。

ユーザ i がラウンド $r - 1$ でオークションに負けていた場合

$$v_i^r = v_i^{r-1} + \alpha \quad (3.1)$$

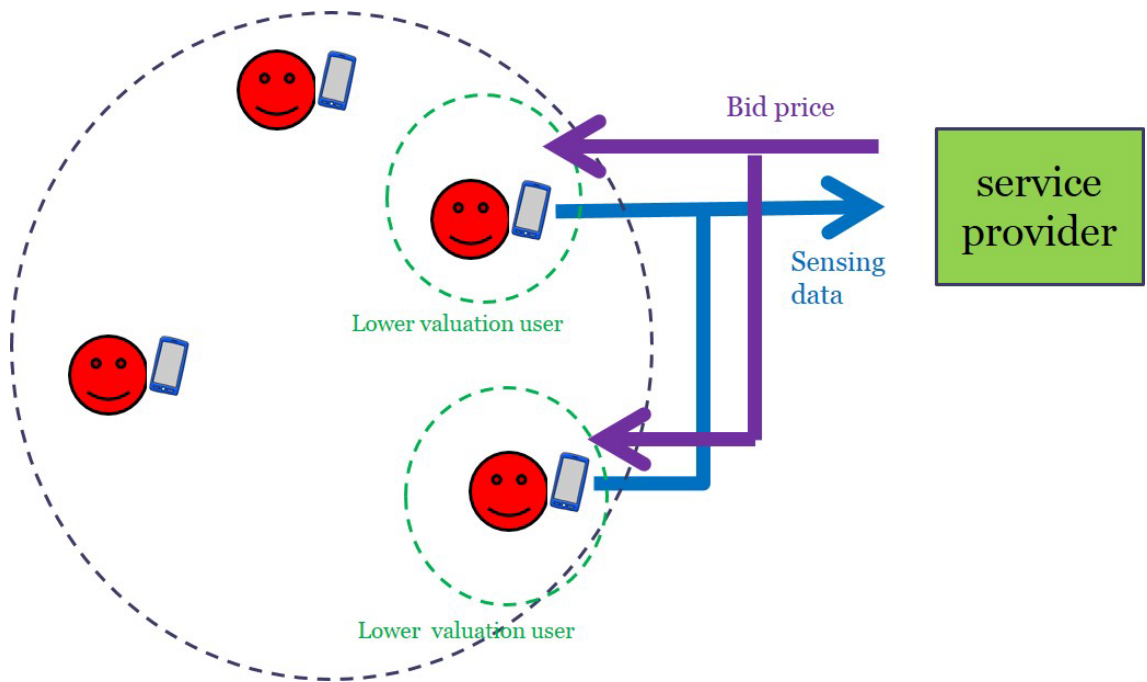


図 3.1: RADP-VPC における逆オークション方式

それ以外の場合

$$v_i^r = 0 \tag{3.2}$$

そして、ユーザはオークションの際自分の提示した競り値より VPC の分だけ安い価格で勝負することができる (図??)。つまり、VPC が大きければ大きいほどオークションで勝利できる可能性は高くなるので、高い競り値をつけるユーザ≒高い T を持つユーザでも VPC が大きければ勝利するチャンスがある。VPC は上式の通りユーザがオークションに負ければ負けるほど貯まっていき、勝った時にリセットされる。そのため、負ければ勝つ確率が上がるという設計になっており、ユーザ離れを起こしにくいシステムになっている。

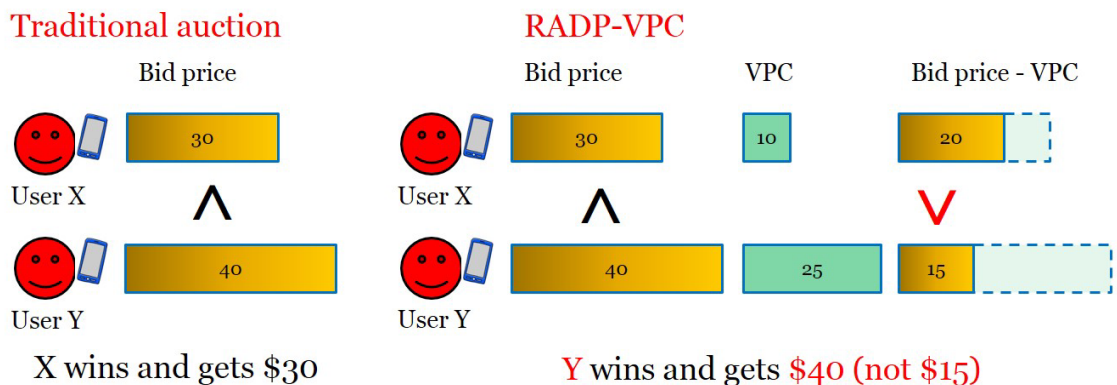


図 3.2: 通常の逆オークションと RADP-VPC の比較の一例

また RADP-VPC では、一度離脱したユーザにオークションの度に前回のオークションでの最高競り値を通知し、再参加を促す工夫も行っている。前回のオークションでの最高競り値という情報は参加ユーザは知り得ない情報であり、これによって離脱ユーザは参加ユーザに比べ次回のオークションの戦略が立てやすくなるため、再参加するきっかけとなる。RADP-VPC ではユーザの行動を ROI（投資収益率）を用いてモデル化し、シミュレーションを行っている。そして、ランダムに勝利者を選び一律の報酬を与える手法である RSFP(Random Selection based Fixed Pricing) と比較したシミュレーション結果が（図 3.3）である。

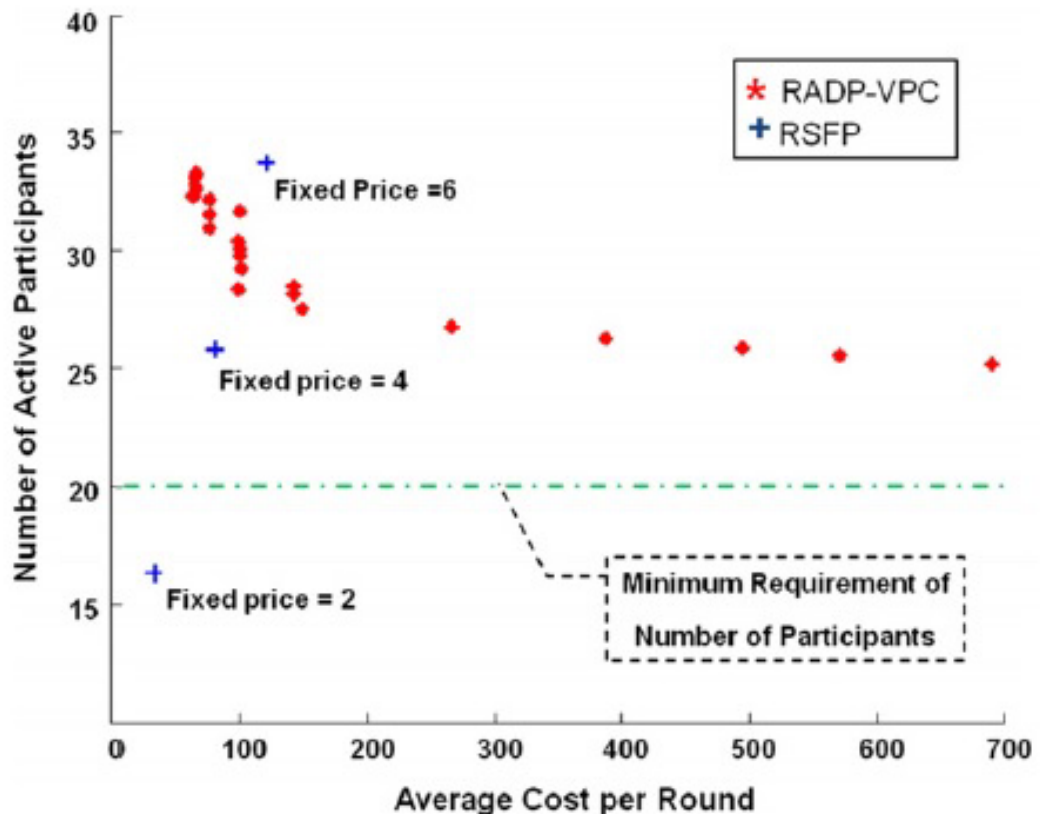


図 3.3: RADP-VPC と RSFP のシミュレーション結果

この研究では確保できるユーザ数のみが評価されており、データのカバレッジや質に関する評価は行われていない。

3.2.2 インセンティブに関するその他の関連研究

RADP-VPC の他にも金銭によるインセンティブの確保を行う関連研究として、RADP-VPC と同じく逆オークションの方式を利用した GBMC アルゴリズム [5] や、様々な支払い形態（定額制、歩合制等）を利用した Micro Payments[7]、データ収集のランキング毎に決められた賞金が支払われる Steered Crowdsensing[16] などがあるが、これらの研究はデータのカバレッジや質に関する部分が大きいいため、それぞれ後節にて紹介を行う。

その他に、ユーザへの還元を行うことなくユーザのボランタリーな参加によって成功を取めているクラウドセンシングが存在する [21, 22, 23]. その一つがウェザーニュース [21] である.

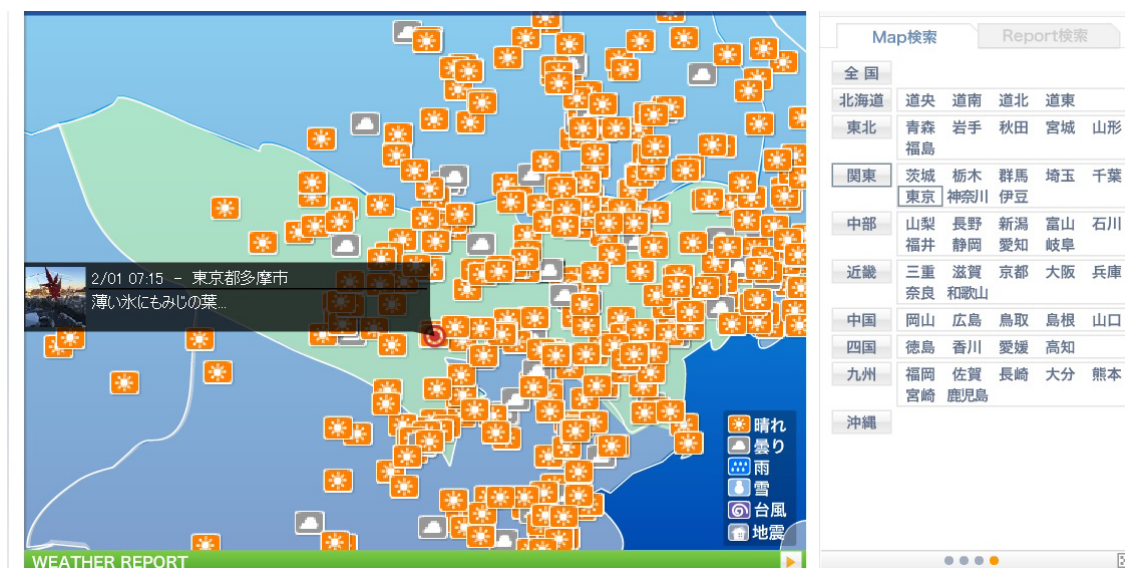


図 3.4: Weather News

ウェザーニュースはウェザーニューズ社の提供するサービスで、有償で会員登録したユーザのみにウェザリポーターとなる権利が与えられる。ウェザリポーターとなったユーザは自分の居場所の天気の様子について携帯電話を利用してテキスト形式で報告することができる。それを受信した同社側が一定のチェックを行った後、同社の気象予報サービスの一部としてその記事内容が公開される仕組みとなっている。同社の気象予報サービス自体はウェザリポーターではないユーザでも利用することが出来るため、ウェザリポーターになることはユーザにとって特にメリットはない。にもかかわらず、ウェザーニュースは100万人以上のウェザリポーターと一日1万件以上の投稿数の確保に成功している。

また、道路地図などの地理情報データを誰でも利用できるようなフリーの地理情報データを作成することを目的としたプロジェクトである OSM(OpenStreetMap)[22] も、ウェザーニュースと同様に多くのユーザのボランタリーな参加によって成立しているクラウドセンシングである。(図 3.5).

これらの事実から、ユーザの中には具体的なメリットがなくても参加する事自体に意義を見出し積極的にセンシングを行う層が一定数存在する事が分かる。これはクラウドセンシングのインセンティブ確保において非常に重要なファクターとなり得るが、この分野の研究はまだあまり進んでいない。参加自体に意義を見出すユーザの意思決定モデルの提案などによる定量的な評価が今後の課題といえる。

3.3 データのカバレッジに関する研究

本節では、クラウドセンシングにおける収集データの利用価値に影響するカバレッジに関する研究として GBMC アルゴリズム [5] と Centralised Management[15] を紹介する。

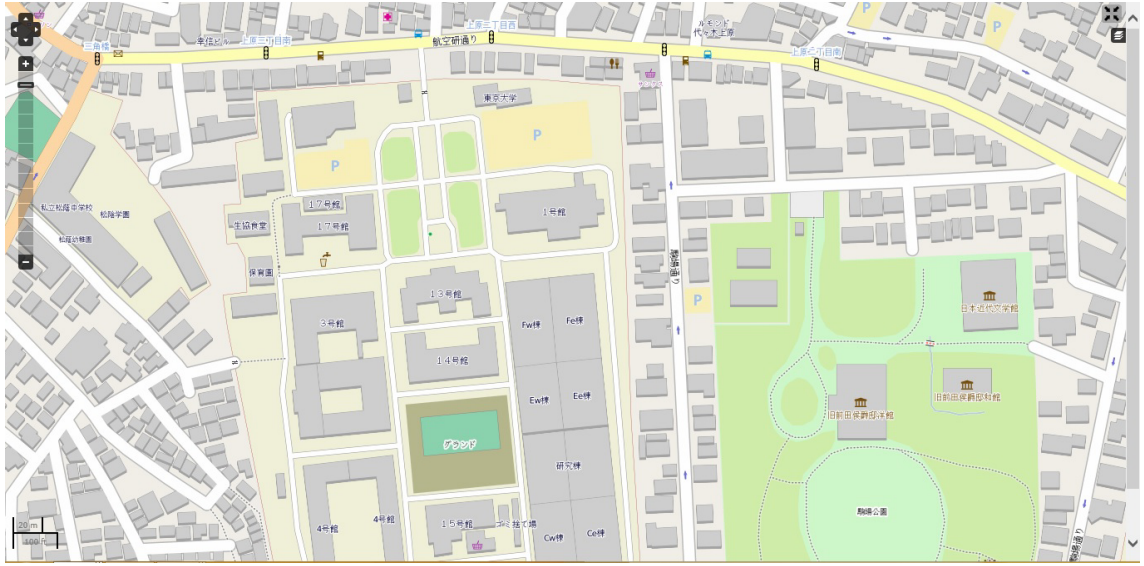


図 3.5: OpenStreetMap

3.3.1 GBMC アルゴリズム

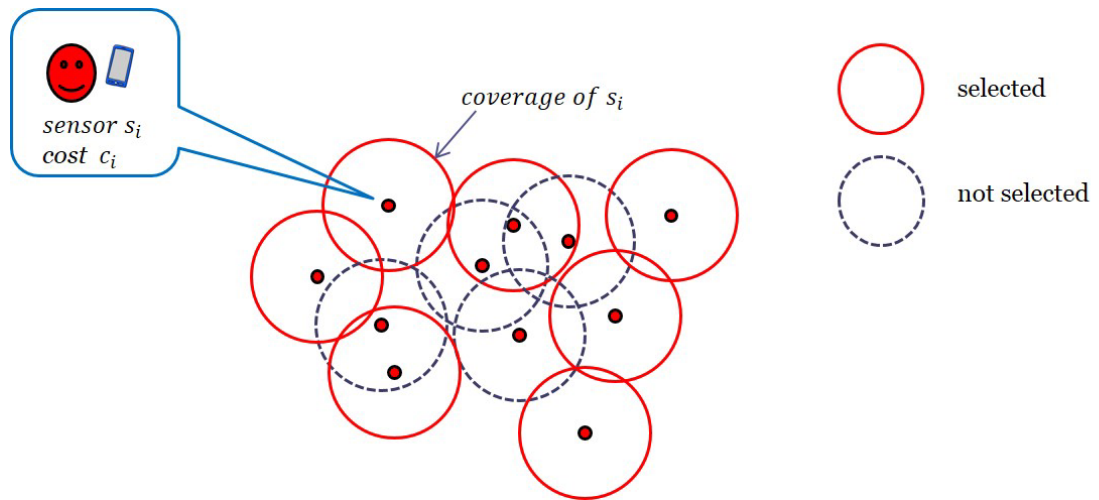


図 3.6: GBMC アルゴリズムにおけるノードの決定法

クラウドセンシングにおいてセンサノードの位置は参加ユーザの移動に合わせて動的に変化をする。そのため、収集したデータのカバレッジを安定させるためには、動的なノード選択を行う必要がある。そのような選択を行う関連研究として Jaimes らによる GBMC (Greedy Budgeted Maximum Coverage) アルゴリズム [5] を紹介する。

この研究は、3.2.1 で紹介した RADP-VPC をベースにして問題点の一つであるデータのカバレッジの改善を図ったものであり、ノード選択の際に GBMC アルゴリズムというアルゴリズムを使用している。GBMC アルゴリズムは、コスト一定の条件化で収集するデータのカバレッジを出来る限り大きくするようなノード選択を行う近似アルゴリズムで

ある。GBMC アルゴリズムでは、各センサノードを中心とした円を描き、それをそのノードのカバレッジとして定義する。そしてコスト一定の条件下で全体のカバレッジの重ね合わせに含まれるノード数を最大にするようなノード選択を目指す (図 3.6)。

しかしこれはよく知られた NP 困難な問題である [24]。この問題を解決するため、GBMC アルゴリズムは以下の手順でノード選択を行い、近似解を導き出す。

1. まだ選んでいないノードの中から、(そのノードを選択することによって増加するカバレッジ÷そのノードの収集コスト) を最大にするノードを選択する、という手順をゼロから総コストが規定値に達するまで繰り返す。
2. まだ選んでいないノードの中から、(そのノードを選択することによって増加するカバレッジ) を最大にするノードを選択する、という手順をゼロから総コストが規定値に達するまで繰り返す。
3. 1, 2 の結果を比べ、カバレッジの大きいものを最終的な結果として出力する。

このように、GBMC アルゴリズムは二種類の評価値 (単位コストあたりの増加カバレッジと単純な増加カバレッジ) を使用した貪欲法を比較して結果が良い方を選択することにより、貪欲法の欠点を緩和し、より近似した解を求めるよう工夫がされている。実際、既存手法に比べカバレッジが改善された事がシミュレーションによって示されている (図 3.7, 図 3.8)。図の横軸は X 座標、縦軸は Y 座標であり、ドットは各ノード、点線で描かれた円は選択されたノードのカバレッジを示している。GBMC アルゴリズムを適用しない場合のシミュレーション結果 (図 3.7) に比べて GBMC アルゴリズムを適用した場合のシミュレーション結果 (図 3.8) の方が選択されるノードが広範囲に散らばっており、全体のカバレッジが改善されているのが分かる。

しかし、この手法ではデータの質が考慮されていないという問題がある。データの質はクラウドセンシングを利用したアプリケーションの性能を左右する非常に重要な要素であり、いくつかの関連研究でも評価の対象となっている [3, 7]。そのため、GBMC アルゴリズムはクラウドセンシングにおけるノード選択の手法としては不十分であるといえる。

3.3.2 Centralised Management

データのカバレッジの評価に関するもう一つの研究として、Centralised Management[15]を紹介する。

Centralised Management は 400m × 400m の庭園内での蜂の目撃情報を収集する参加型センシングが行い、参加ユーザに対する質問の方法やその時間間隔の違いによって収集されるデータのカバレッジやその収集効率の違いを分析した研究である。Centralised Management では目撃情報のカバレッジを評価する際、空間を格子状に区切り、同じ大きさの正方形のブロックに分割する。そして、目撃情報が寄せられたブロックの数をデータのカバレッジとして評価している (図 3.9)。赤色の部分がデータが足りていないエリア、

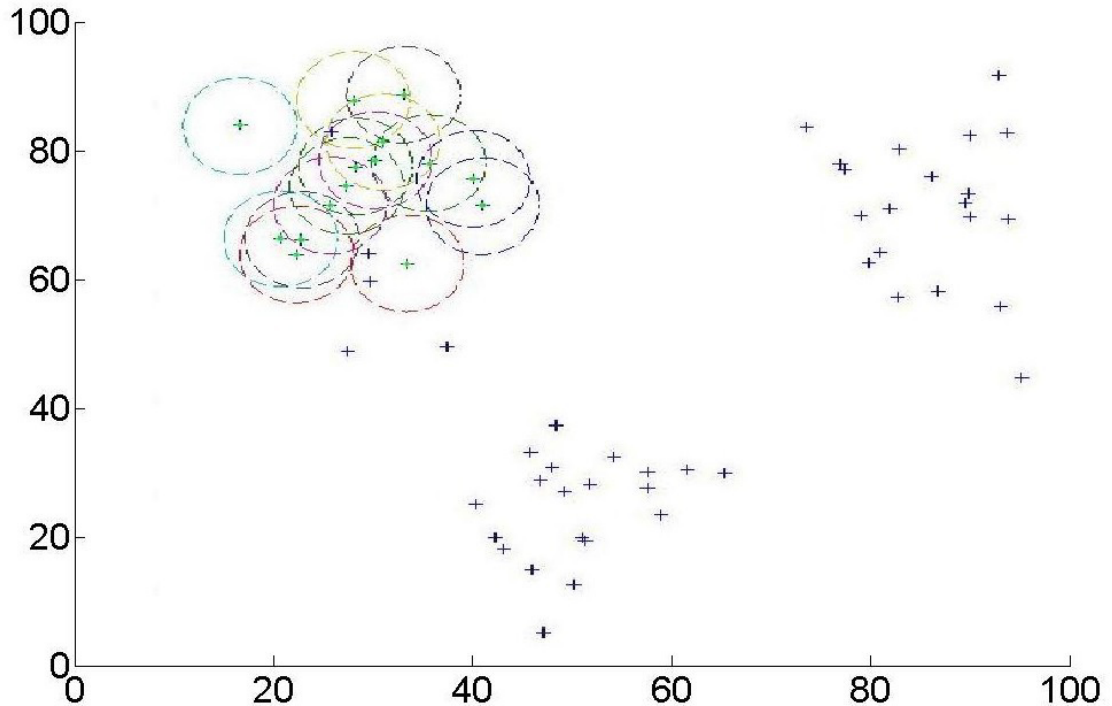


図 3.7: GBMC アルゴリズム未適用時のシミュレーション結果

緑色の部分がデータが過不足なく収集できたエリア，黄色の部分にデータが過剰に収集された（無駄が多い）エリアである．上の二つに比べて下の図において効率的なデータ収集が行われていることが分かる．このように，空間を小さなエリアに分割し，データが存在するエリアの数を数える事でカバレッジの評価は可能となる．

以上のように様々な関連研究においてデータのカバレッジが評価されており，クラウドセンシングにおけるデータのカバレッジが非常に重要であることが分かる．

3.4 データの質に関する研究

本節では，クラウドセンシングにおける収集データの利用価値に影響する質に関する関連研究として Micro Payments[7] と，Steered Crowdsensing[16] を紹介する．

3.4.1 Micro Payments

クラウドセンシングで収集するデータの質を評価する指標は収集するデータの種類によって変化すると考えられる．そのようなデータの質の評価が行われている関連研究として Micro Payments[7] を紹介する．

Micro Payments は写真をデータとして収集することを目的としたクラウドセンシングを実際に運用し，報酬としての賃金の支払い形態の違い（参加ユーザに与えるインセンティブの違い）によって収集されるデータの質・カバレッジに生じる変化を分析した研究である．この研究ではある大学構内に点在するゴミ箱の写真が収集対象となっているが，

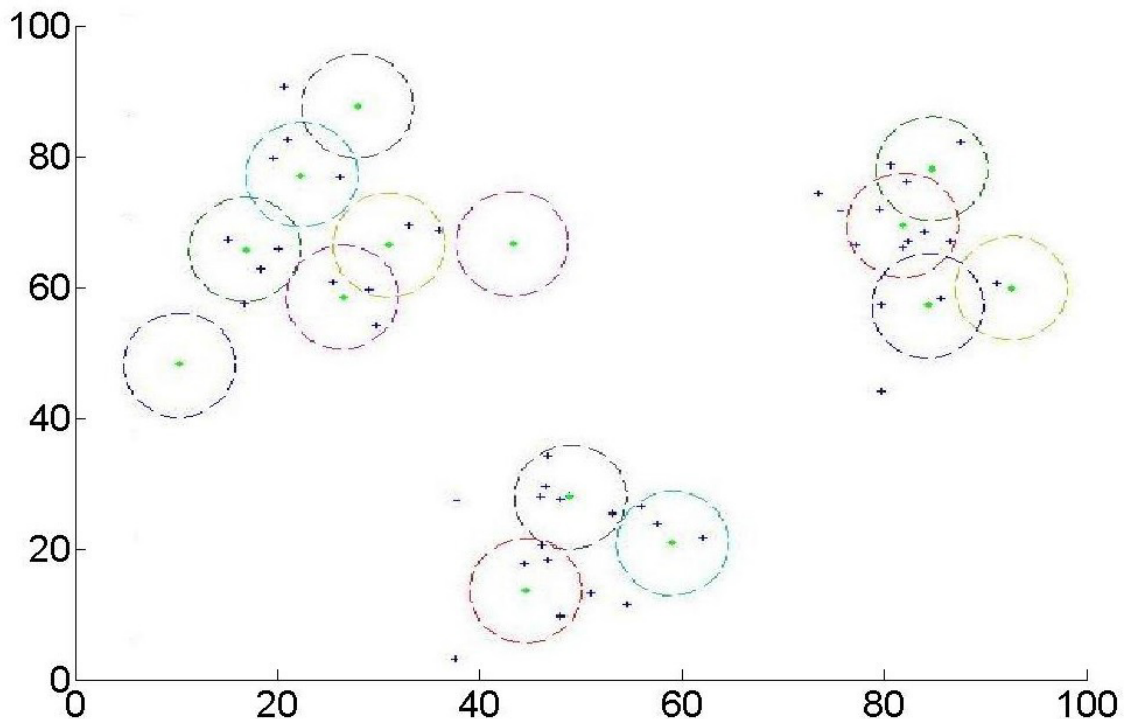


図 3.8: GBMC アルゴリズム適用時のシミュレーション結果

参加ユーザは写真に付加情報として任意でタグを付けることができる (図 3.10). そして, このタグの数や写真の写り具合がデータの質の指標として扱われ, 賃金の支払い形態による違いが分析されている (図 3.11).

3.4.2 Steered Crowdsensing

データの質に関するもう一つの研究として, Steered Crowdsensing[16]を紹介する. Steered Crowdsensing は, Wi-Fi を利用した屋内測位のための学習データを収集することを目的としたクラウドセンシングで, ゲーミフィケーションを利用しユーザにインセンティブを与えている. 参加ユーザはデータ収集によって得られる pt (ポイント) を集め, その総量によるランキングに応じた賞金を獲得することができる (図 3.12). 現在のランキングと各計測地点で得られる pt を示した地図が専用のアプリ上に表示されるようになっており (図 3.13), 参加ユーザはアプリを頼りにより高い pt を集めることでランキング上位を目指すことになる.

Steered Crowdsensing において, データの質は学習データとしての有用性によって評価される. Wi-Fi を利用した屋内測位においては, Wi-Fi フィンガープリントの分散が大きい地点における新たな計測データがより大きな学習効果をもたらすため, Steered Crowdsensing では分散の値がそのままデータの質として利用されている. 筆者らは各計測地点における pt の値を調節することでより高い質のデータが得られる地点へと参加ユーザを誘導することが可能であると述べており, 実際に大学内で学生を使った実験によって評価を行っている. その結果が (図 3.14) である. 各手法毎に収集したサンプルを利用した屋内測位にお

ける推定位置と実際の位置との距離が比較されており、Steered Crowdsensing は他の手法に比べ推定誤差が少なく、より質の高いデータサンプルを収集できていることが分かる。

以上のように、収集されるデータの種類によって適切な指標を用意することでデータの質に対する評価を行うことが可能となる。先述した通り、クラウドセンシングにおけるデータの質はその利用価値に直結するため、それぞれのプロジェクトにおける収集データを適切な尺度を持って評価することができるかどうかは非常に重要な課題となってくる。

3.5 本章のまとめ

本章では、クラウドセンシングにおけるユーザのインセンティブやデータのカバレッジ、データの質に関する検討を行っている関連研究についてそれぞれの項目毎に紹介を行った。

ユーザのインセンティブに関する研究としては RADP-VPC を主に紹介した。RADP-VPC ではユーザの行動に対し ROI を利用した簡単なモデル化を行い、クラウドセンシングにおけるインセンティブとして逆オークションが有用であることを示した。しかし、ウェザーニュースや OpenStreetMap がユーザへの還元なしでインセンティブの確保を行っている事実を鑑みると、現実の人間はそれほど単純な行動モデルでは表すことができない行動を取るといえるため、ユーザのインセンティブに関する定量的な評価はまだ難しいと結論付けられる。

データのカバレッジに関する研究としては GBMC アルゴリズムと Centralised Management を紹介した。GBMC アルゴリズムはユーザ毎に円状のカバレッジを想定しその重ね合わせの広さを測定することで、Centralised Management では空間をメッシュ状に区切り十分なデータ数が集まったメッシュの数を数えることでそれぞれカバレッジの測定を行い、その改善を示した。また、データの質に関する研究としては Micro Payments と Steered Crowdsensing を紹介した。Micro Payments は写真の写り具合と追加タグの数で、Steered Crowdsensing は収集したデータによる WiFi を利用した屋内測位の測定精度によってそれぞれデータの質を評価した。

データの質・カバレッジに関してはそれぞれについては様々な方法で定量的な評価が行われている一方で、この両面について有用な評価が行われた研究はいまだ数が乏しい。データの質とカバレッジはいずれもデータの利用価値に直結するという関わりが深い関係にあるため、この両面について一般的に評価を行うことは非常に重要な課題であるといえる。



図 3.9: Centralised Management におけるカバレッジの評価

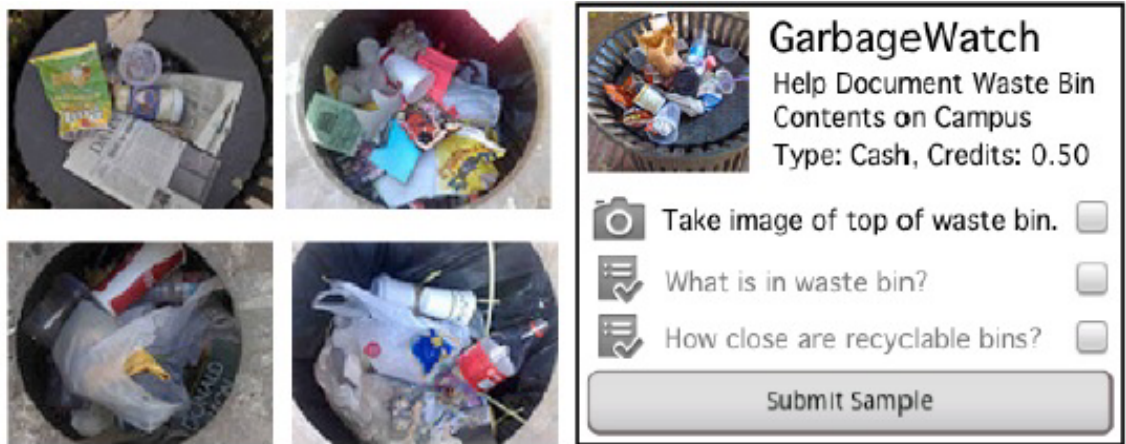


図 3.10: Micro Payments における写真データの収集

Incentive Type	Avg % of Invalid Pics per User	Avg % of Optional Tags per User
MACRO	7±5	70±21
MEDIUM _{μt}	6±4	47±25
HIGH _{μt}	6±3	60±25
LOW _{μt}	5±3	52±24
COMPETE _{μt}	6±3	6±4

図 3.11: Micro Payments における質の評価

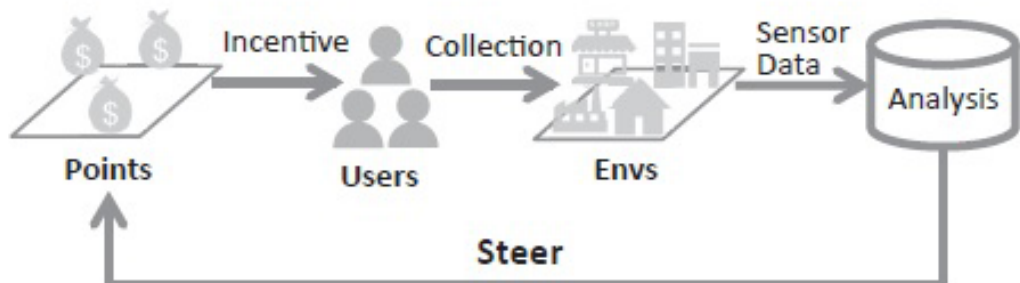


図 3.12: Steered Crowdsensing の概要

Stamp Rally	
RANKING	SCORES
1	Ryoma Kawajiri 35586pt
2	子安秀昇 21336pt
3	川尻亮真 14581pt
4	Saisho Osamu 3139pt
5	せいら荒木 1061pt
6	Matsudaira Kenei 1022pt
7	Sunakawa Takuya 435pt
8	Kentaro Nishi 95pt
9	Shimosaka Masamichi 75pt

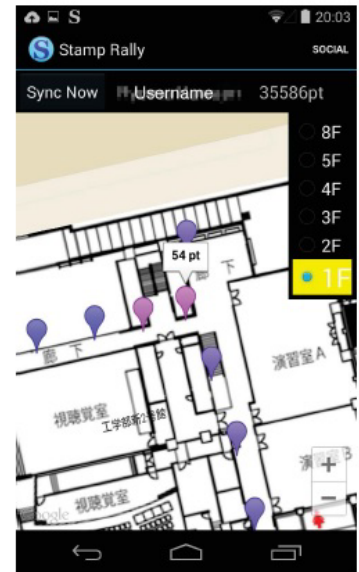


図 3.13: Steered Crowdsensing でユーザに配布されるアプリ

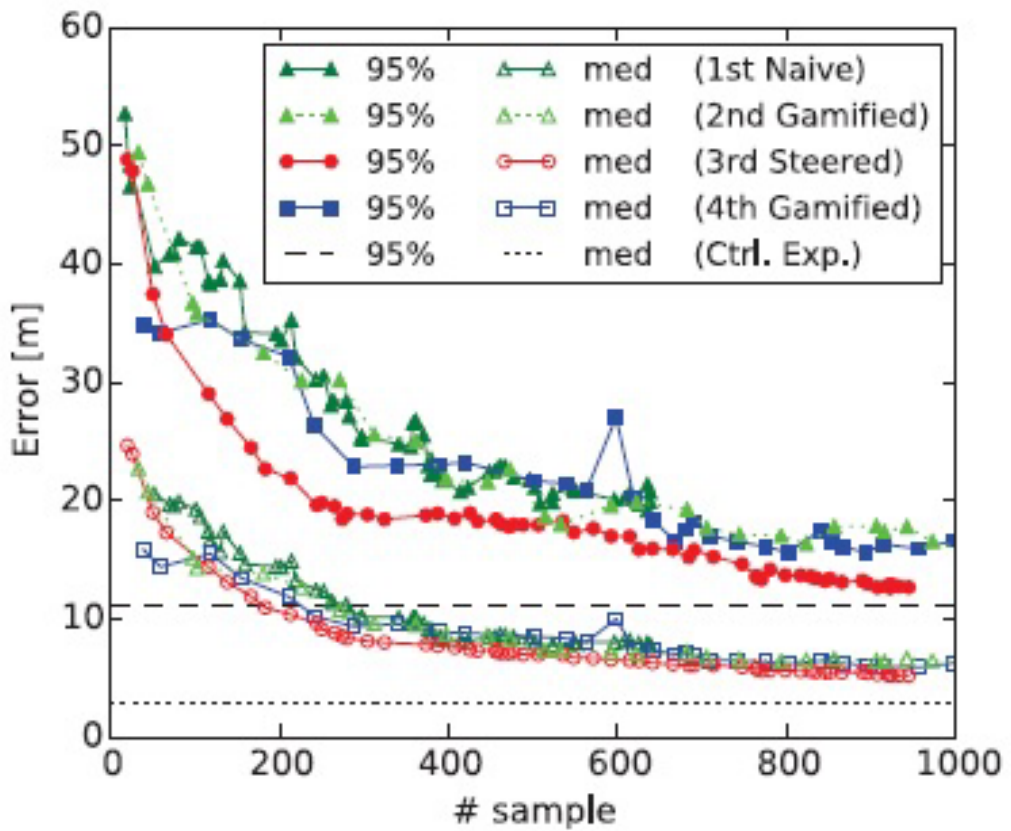


図 3.14: Steered Crowdsensing の実験結果

第4章 データの質の定量化

4.1 概要

前章の後半ではクラウドセンシングにおけるデータのカバレッジと質について評価を行った関連研究をそれぞれ二つずつ紹介したが、その両面について定量的な評価を行った有用な研究は未だ数が乏しい。そのため、本研究ではクラウドセンシングにおける重要な要素であるデータのカバレッジとデータの質の両面を考慮した新しいノード選択手法の実現を目指す。

データのカバレッジと質の両面を考慮したノード選択アルゴリズムを検討するにあたって、データの質を評価して定量化し、パラメータとして扱えるようにする必要がある。また、そのパラメータをもとにノードの取捨選択を行うため、定量化はデータ収集前の段階で行われる必要がある。関連研究の micro-payments[7] では写真の写り具合や付加情報であるタグの数が、steered crowdsensing[16] ではデータ収集地点の Wi-Fi フィンガープリントの分散の大きさが指標となっていたが、本章ではデータの質を定量化する方法として、センサの性能を利用する方法、データの位置情報を利用する方法、ユーザの評価を利用する方法、の三種類に分けてそれぞれ議論を行う。

4.2 センサの性能を利用したデータの質の定量化

センサの機械的性能を利用するという方法が考えられる。

例えば、数値を測定するタイプのセンサである場合、最終的に目標とする制御量に対するセンサの分解能の値は非常に重要である。どの程度の分解能が必要になるかは目指すデータの正確性によって変わってくるが、少なくとも目標となる制御量と同等の分解能が必要であり、そこから分解能が高くなるにつれてデータの価値は緩やかに高まっていくと考えられる。s のため分解能とデータの価値の対応は (図 4.1) のようになる。

このようにセンサの性能が予め定量化されている場合、それを利用することでデータ収集前の段階でデータの質の定量化を行うことが可能となる。

4.3 データの位置情報を利用したデータの質の定量化

データの位置情報を利用するという方法が考えられる。

収集するデータの利用目的によっては、データの収集位置毎にデータの価値が変化する場合がある。例えば、3.4.2 で紹介した Steered Crowdsensing では Wi-Fi を利用した屋内測位における学習データとしての有用性の指標として Wi-Fi フィンガープリントの分散の大きさがデータの質として利用されている [16] が、これはデータの計測位置に依存する値であり、データの位置情報を利用したデータの定量化の一例であるといえる。

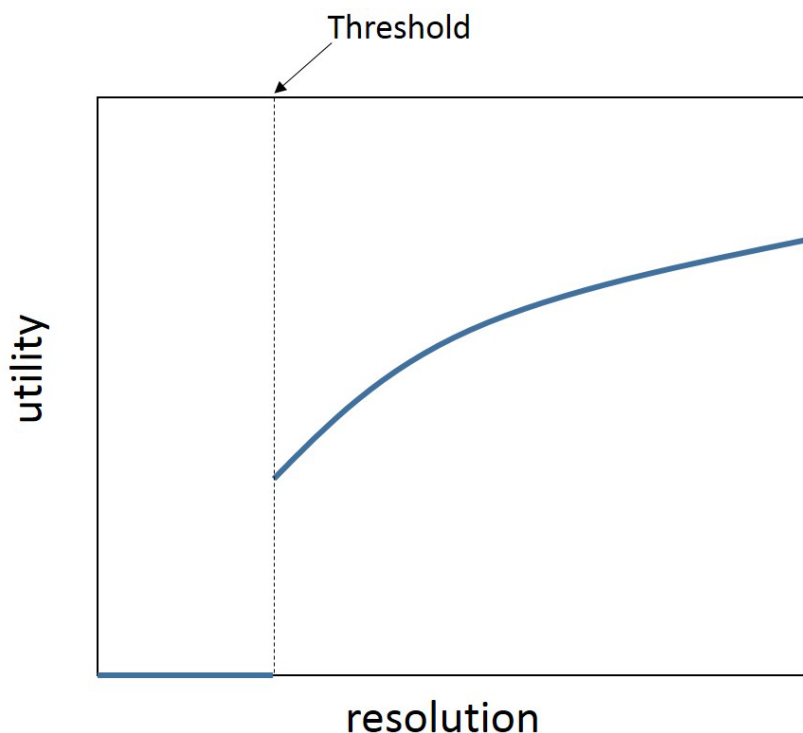


図 4.1: 分解能とデータの質の対応の例

他にもデータの位置情報を利用したデータの質の定量化が有効なケースはいくつか考えられる。例えば、単純な環境調査に利用する場合、データは位置毎のばらつきなく均一に集まることが望ましいため、収集の過程でデータが少なくなってきた場所のデータの質を高く評価することで全体の均一化を図ることが出来る。他の例として、人々の生活をサポートするアプリケーション（道案内やお店のレビューなど）に利用する場合を考えると、人々がよく訪れる場所の方が情報の需要がより高くなることが考えられる。そのような位置のデータの質を高く評価することで、より最適なデータの収集を行うことが可能になる。

以上のように、データの位置情報とデータの質に関係性があった場合、それを利用することでデータ収集前の段階でデータの質の定量化を行うことが可能となる。

4.4 ユーザの評価を利用したデータの質の数値化

ユーザに対する累積的な評価を利用するという方法が考えられる。

データ取得の方法が写真撮影やテキスト形式によるレポートなどのユーザによる介入が大きいものの場合、ユーザ毎の個体差によってデータの質に差が生じてしまう。例えば写真撮影の場合はユーザの撮影技術の差によって、レポートの場合はユーザの文章力の差によってデータの質に差が生じると考えられる。各ユーザとそのデータの質には相関があると考えられるため、過去の累積的なセンシングデータからユーザに対する評価を行い、その評価とデータの質を結び付けることで、データの質を評価することが可能となる。これ

はネットオークションなどで見られる出品者評価のシステムに近い。取引前に出品者の評価を見ることでどの程度の質の取引が期待できるか概ね予想できるように、ユーザの評価を見ることでデータ収集前の段階で概ねのデータの質が予想できるのである。

そして、ユーザに対する定量的な評価を行う単純な方法として、そのユーザが良質なデータを収集する確率を利用する方法が考えられる。3.4.1で紹介した Micro Payments のように写真撮影の場合は写り具合によってデータの質を評価することが可能であるが、そのように収集データがある閾値で良質なデータと悪質なデータの二種類に大別し、良質なデータの割合をそのままそのユーザの評価値として扱うことで定量化が可能である。

この方法は、悪意のあるユーザによる影響を緩和できるという利点があるが、累積的な評価を行うためにある程度長期的なプロジェクトの継続が必要である点が難点である。

4.5 本章のまとめ

本章では、データ収集前の段階でのデータの定量化の手法として、センサの性能を利用する方法、データの位置情報を利用する方法、ユーザの評価を利用する方法についてそれぞれ提案を行った。以下にそれぞれの特徴をまとめる。

- センサの性能を利用する方法は、分解能のようにセンサの性能が予め定量化されている場合に有効な手法となる。
- データの位置情報を利用する方法は、データの位置情報とデータの質に関係性がある場合に有効な手法となる。
- ユーザの評価を利用する方法は、データ取得の方法がユーザの能力によるところが大きい場合に有効な手法となる。

以上のように、データの取得方法や利用目的によってどの手法が有効に働くかは様々であることが分かる。データの質を適切に定量化するためには、クラウドセンシングのプロジェクト毎に適切な手法を一つあるいは複数選択し、最適な運用を行うことが必要となる。本研究では、データの質の定量化が行われた前提で、その値を「Utility」としてデータのパラメータとして扱う事を考える（図 4.2）。

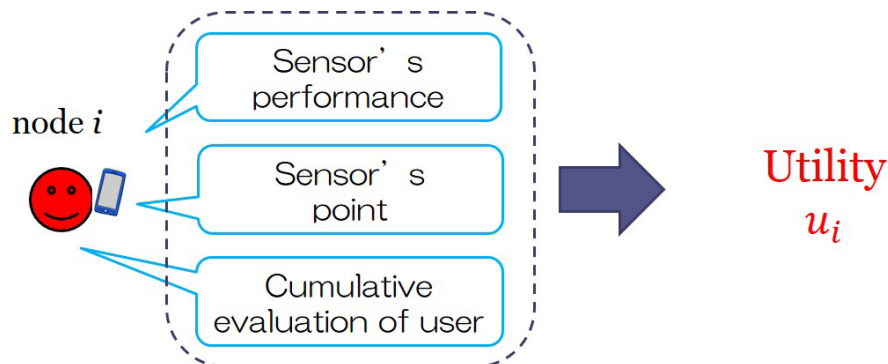


図 4.2: データの質の定量化

第5章 ノード選択アルゴリズム

5.1 概要

前章ではデータのカバレッジと質の両面を考慮したノード選択アルゴリズムを検討する下準備として、データ収集前の段階でデータの質を定量化する手法について示した。本章では前章で述べたような方法を用いてデータの質が定量化されている前提で、データの時とカバレッジの両方を考慮したノード選択が可能な手法を提案する。さらにその手法を拡張し、ユーザの移動を考慮し時間経過によるカバレッジの低下を緩和するようなノード選択が可能な手法を提案する。

5.2 データの時とカバレッジの両面を考慮したノード選択アルゴリズム

本節では、3.3.1で紹介した GBMC アルゴリズム [5] をベースに改良を加え、データの質とカバレッジの両方を考慮したノード選択が可能となる手法を提案する。

GBMC アルゴリズムでは各センサノードを中心とした円状のカバレッジを想定し、その重ね合わせの面積を評価値としてこれを最大化するような貪欲法によるノード選択を行っていた。本研究では GBMC アルゴリズムの考え方を拡張し、各センサノードを中心とした円状のカバレッジを底面、データの質を高さとする立体的な円柱を考え、その体積を評価値として扱うことを考える (図 5.1)。この円柱の重ね合わせの体積はデータのカバレッジとデータの質の両方を含んだ評価値となるため、これをより大きくするようなノード選択を行う (図 5.2) ことで、データのカバレッジと質を総合的に高めるような選択が期待できる。

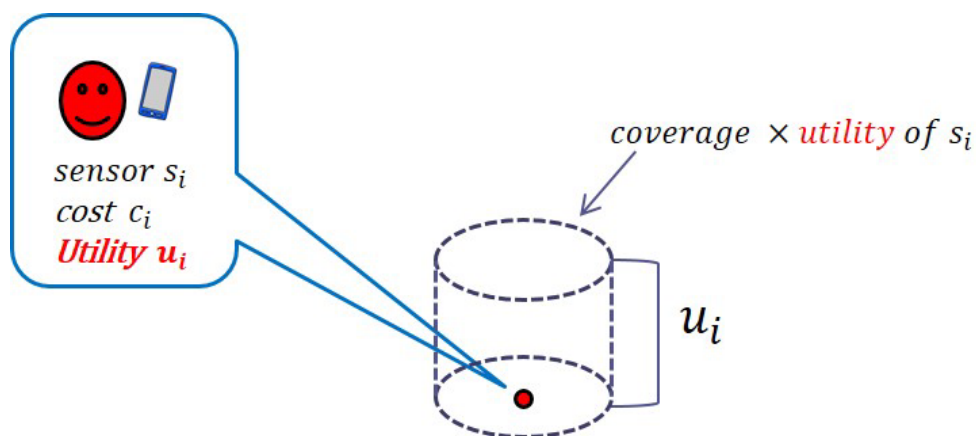


図 5.1: 提案手法における評価値

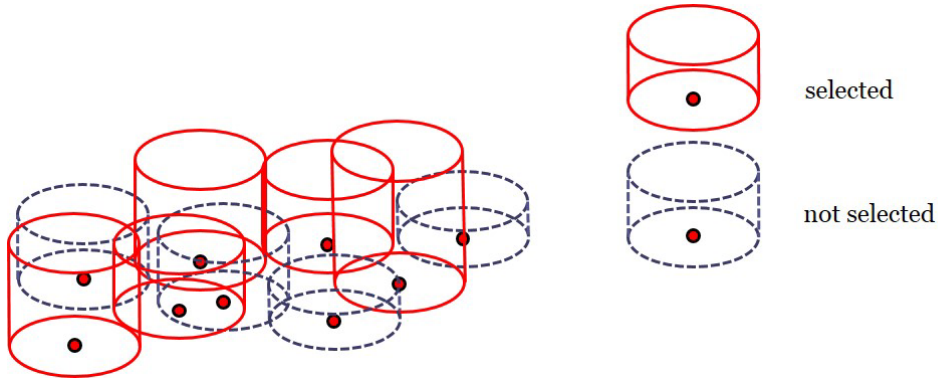


図 5.2: 提案手法におけるノード選択

この評価値をコスト一定の条件で最大化するノード選択を行うことは，GBMC アルゴリズム同様 NP 困難な問題となるため，貪欲法によって近似解を導き出す．筆者らが提案する近似アルゴリズムによるノード選択手法の概要を示した疑似コードを（Algorithm 1）に示す．

Algorithm 1 Maximize coverage and utility

S : the collection of nodes
 i : the ID of node
 V : volume
 C : cost
 $G \leftarrow \emptyset$
 $V \leftarrow \emptyset$
 $C \leftarrow 0$
 $U \leftarrow S$
while $U \neq \emptyset$ **do**
 select $S_i \in U$ that maximizes $\frac{\overline{V} \cap V_i}{C_i}$
 if $C + C_i \leq \text{MAXCOST}$ **then**
 $G \leftarrow G \cup S_i$
 $V \leftarrow V \cup V_i$
 $C \leftarrow C + C_i$
 end if
 $U \leftarrow U \cap \overline{S_i}$
end while
return G

5.3 ユーザの移動を考慮したノード選択アルゴリズム

現実にクラウドセンシングを行った場合、当然ユーザの位置は時間経過によって変化する。GBMC アルゴリズムや 5.2 で提案したノード選択手法はユーザの移動を考慮していないため、時間経過によってユーザの位置が変化した時、データのカバレッジの低下が生じる可能性がある。無論、ユーザの移動を完全に予測することは不可能であるため、カバレッジの低下を完全に防ぐことはできない。

そこで、時間経過によるカバレッジの低下を緩和するために、評価値に変化を加えることを考える。人間が移動を行う場合、一定時間後にその人間がある位置に存在する確率は現在位置からの距離が大きくなればなるほど減衰していくと考えられる。この減衰を評価値に反映させることで、ユーザ毎の評価値に移動後の存在位置の確率分布を含ませることが出来るため、時間経過によるカバレッジの低下の緩和が期待できる。

この減衰はノード中心からの距離 x 、カバレッジ半径 a 、データの質 u を用いて一般的に以下のように表すことができる。

$$\begin{aligned} f(0) &= u \\ f(a) &= 0 \\ f'(x) &< 0 \quad (0 \leq x \leq a) \end{aligned}$$

この減衰関数はユーザ移動モデル毎に様々なものが考えられるが、本研究ではデータを広く扱うことを想定し、最も単純な減衰として一次関数による単調減少を使用し、5.2 で提案した円柱状の評価値から円錐状の評価値へと変化させた (図 5.3)。

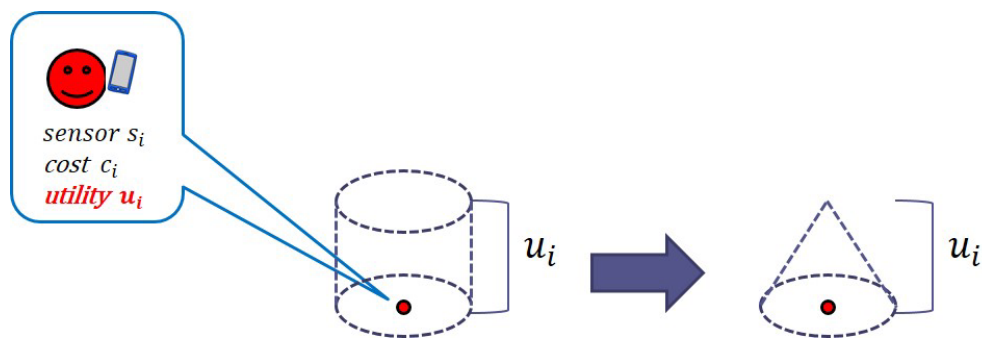


図 5.3: ユーザの移動を考慮した評価値

評価値に (図 5.3) の体積を用いて 5.2 で示したアルゴリズムによる貪欲法を行うことにより、データの質とカバレッジの両面とユーザの移動による影響を考慮しノード選択が可能となる。

第6章 評価

6.1 概要

本章では、前章で解説した提案手法と他の手法の比較評価について述べる。まず、事前実験としてシミュレーションデータを用いて比較評価を行い、提案手法がデータのカバレッジと質を総合的に高めるノード選択を行うことができることを示す。次に、実データである人の流れデータを用いた動的なノード選択を行った場合の比較評価を行い、提案手法が実データに対しても有効であることを確認する。最後に、静的なノード選択を行った場合の比較評価を行い、提案手法がノードの移動による影響を緩和できていることを示す。

6.2 シミュレーションデータを用いた評価

本節ではシミュレーションデータを用いた比較評価を行い、提案手法のアルゴリズムが狙い通り動作していることを示す。まず、シミュレーション環境について解説を行った

6.2.1 シミュレーション環境

シミュレーションには（表 6.1）のようなパラメータを使用した。各ノードの位置は、 500×500 の空間上に完全な一様分布によって生成した。総コスト上限 100 に対してノード毎のコストの平均を 3.5 に設定し、100 ノード中 30~40 ノード程選択することを想定している。また、各ノードのパラメータは試行毎に異なるシードの乱数によって生成されるが、一度生成した同じノード群に対してそれぞれの手法によるノード選択を行い、これを 10000 回繰り返して結果の平均値を比較した。今回比較対象としたノード選択手法は提案手法、GBMC アルゴリズム、データの質のみを考慮したノード選択手法、ランダムなノード選択手法の四種類である。提案手法、GBMC アルゴリズムは先述の通りのアルゴリズムである。データの質のみを考慮したノード選択手法とは、提案手法や GBMC アルゴリズムにおける貪欲法に用いる評価値を（データの質/コスト）に変えたアルゴリズムであり、データの質のみの最大化を目指す。ランダムなノード選択手法は総コストが上限に達するまで無作為に一つずつノードを選択していくだけの手法である

6.2.2 シミュレーション結果

10000 回のシミュレーションの比較結果の内の一例を（図 6.1）～（図 6.4）に、データのカバレッジと質の平均の比較結果を（図 6.5）に、提案手法で評価値とした重ね合わせの体積の平均の比較結果を（図 6.6）に、シミュレーション結果のまとめを（表 6.2）に示す。（図 6.1）～（図 6.4）に関しては選択された各ノードのカバレッジが赤い円で、選択されなかった各ノードが青い点で表されている。また、（図 6.6）に関してはそれぞれの手法の平均値

表 6.1: シミュレーションのパラメータ設定

各パラメータ	パラメータの値
試行回数	10000
総ノード数	100
総コスト上限	100
空間の x 幅	500
空間の y 幅	500
ノード毎のカバレッジ半径	40
ノード毎の X 座標の分布	$U(0, 500)$
ノード毎の Y 座標の分布	$U(0, 500)$
ノード毎のデータの質の分布	$N(5, 2)$
ノード毎のコストの分布	$N(3.5, 1.4)$

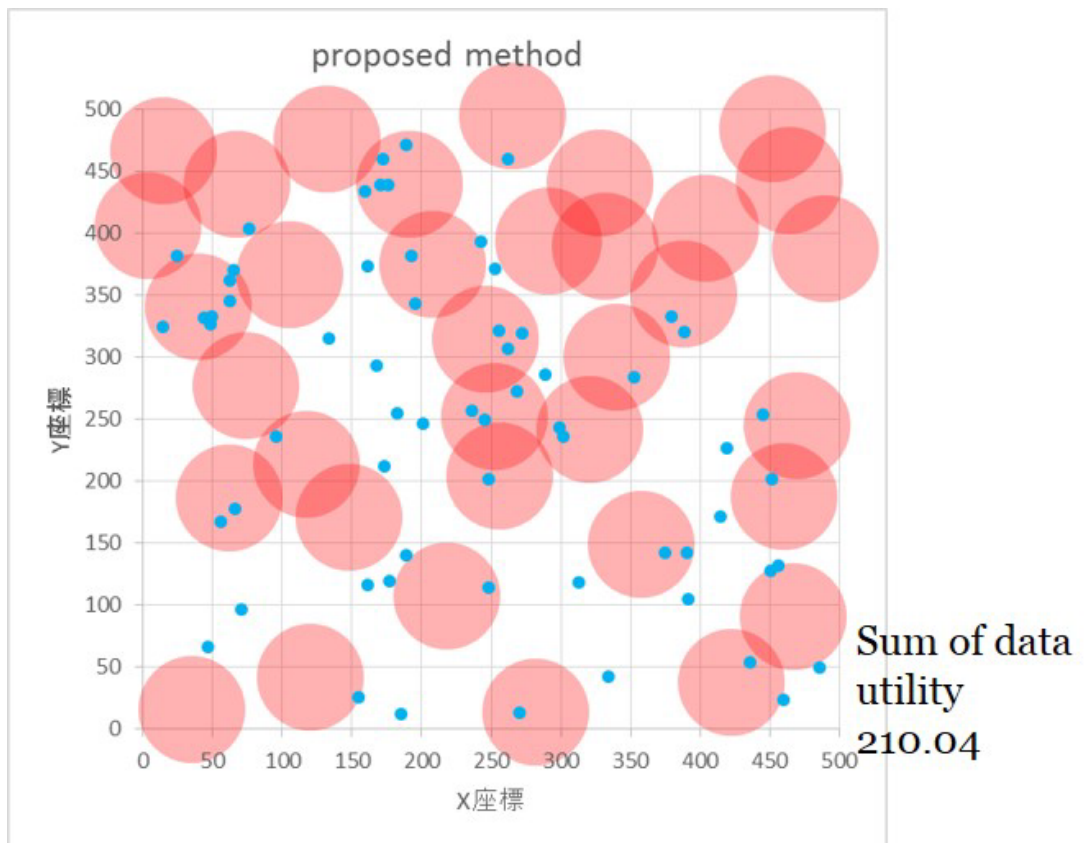
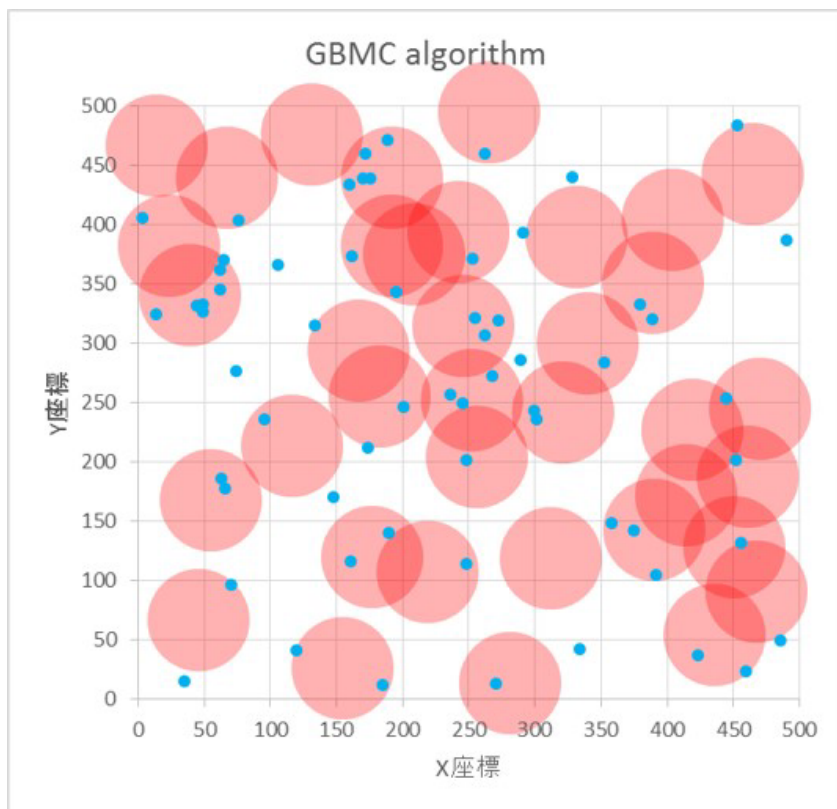


図 6.1: ノード選択の結果：提案手法



Sum of data utility
188.33

図 6.2: ノード選択の結果 : GBMC アルゴリズム

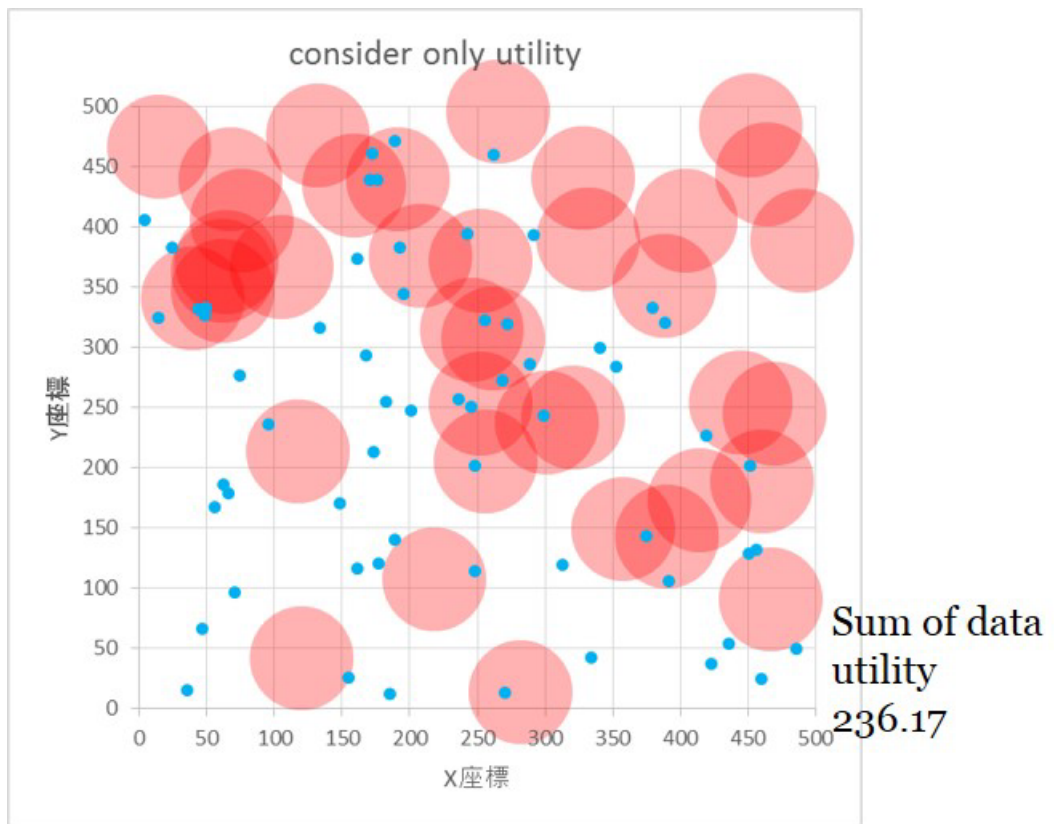
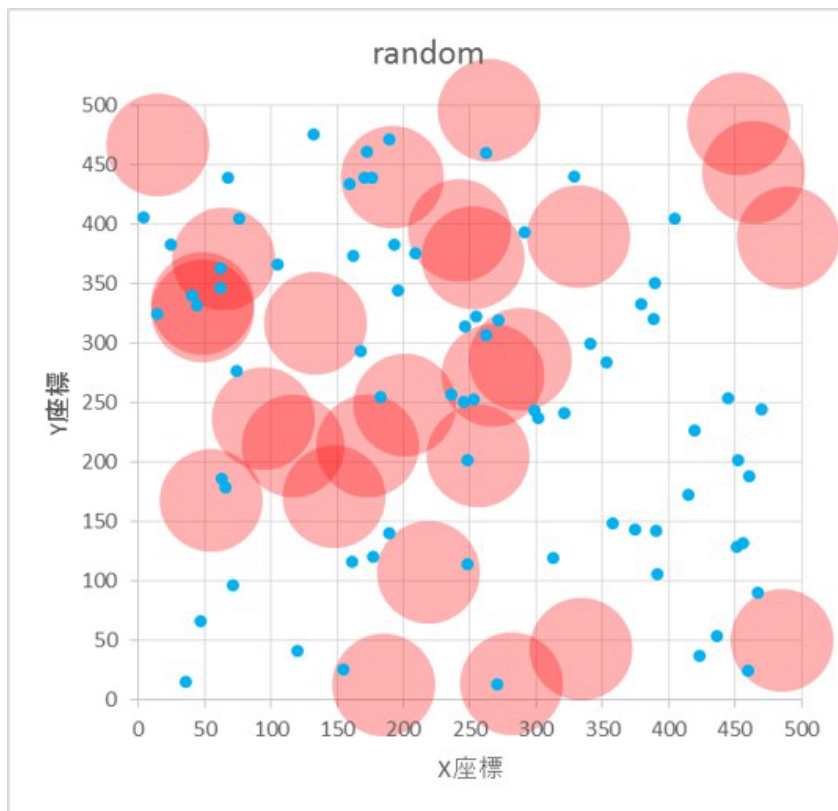


図 6.3: ノード選択の結果：質のみ重視



Sum of data
utility
155.00

図 6.4: ノード選択の結果：ランダム

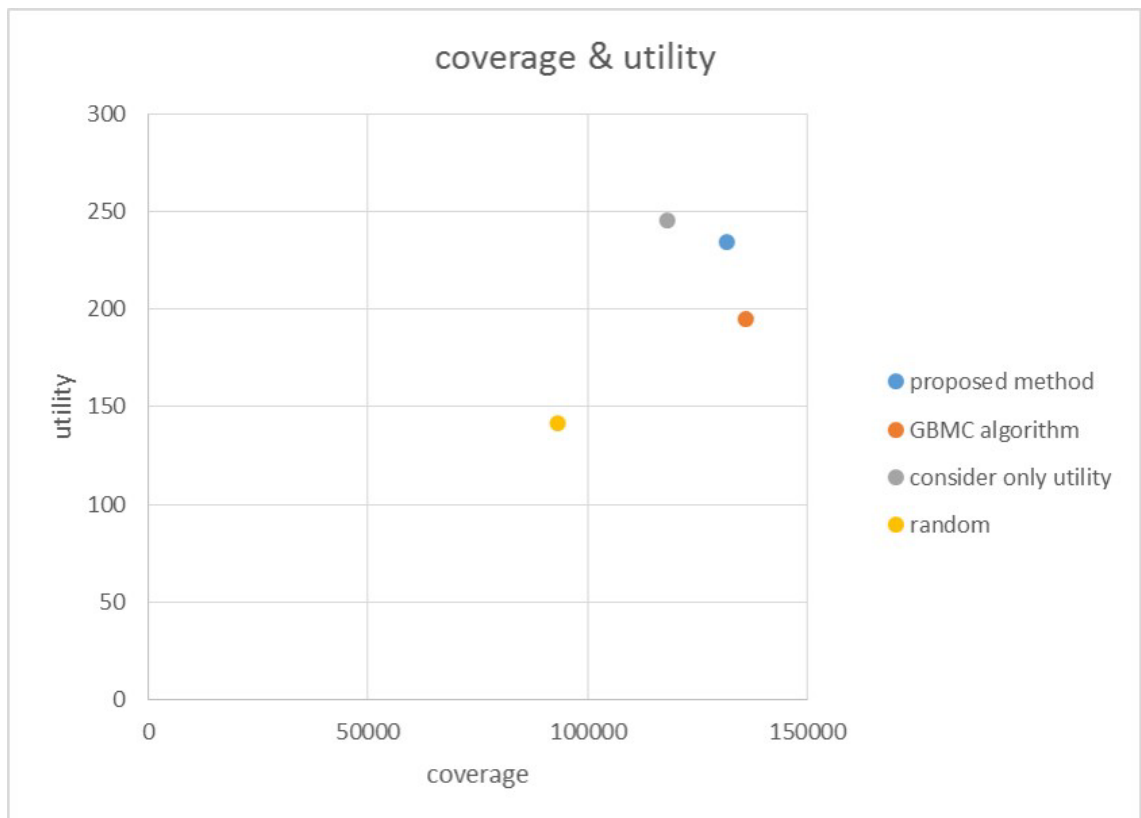


図 6.5: カバレッジと質の平均値のシミュレーション結果

の 99%信頼区間がエラーバーで示されている。エラーバーの短さから、10000 回もの試行によってほぼ正確な平均値が算出できているといえる。

まず、(図 6.1) ~ (図 6.4) から提案手法が質のみを重視したノード選択手法やランダムなノード選択手法と比較して GBMC アルゴリズムに近いレベルの広いカバレッジを確保できていることが視覚的に確認できる。実際の全体カバレッジと質の平均値の比較が(図 6.5)のグラフに示されている。横軸がデータのカバレッジ、縦軸がデータの質の平均値をそれぞれ示しており、提案手法が右上に位置していることから提案手法がカバレッジと質の両面を総合的に高めていることが分かる。(表 6.2) よりカバレッジの正確な平均値は提案手法、GBMC アルゴリズム、質のみを重視したノード選択手法、ランダムなノード選択手法の順に 131713, 136134, 118269, 93157 であり、提案手法はカバレッジのみを重視したノード選択手法である GBMC アルゴリズムと比較しても-3.2%程度のカバレッジの確保に成功している。また、質のみを重視したノード選択手法と比較すると+11.3%程度、ランダムなノード選択手法と比較すると+41.4%程度のカバレッジとなっている。また、質の正確な平均値は先述の順番に 234.4, 195.2, 245.2, 141.8 であり、提案手法は質のみを重視したノード選択手法と比較して-4.4%程度のデータの質の確保に成功している。また、GBMC アルゴリズムと比較すると+20.1%程度、ランダムなノード選択手法と比較すると+65.3%程度のデータの質となっている。

(図 6.6) のグラフには提案手法で評価値とした重ね合わせの体積の平均値の比較が示されている。提案手法が最も大きい値となっており、アルゴリズムが狙い通り働いて評価値

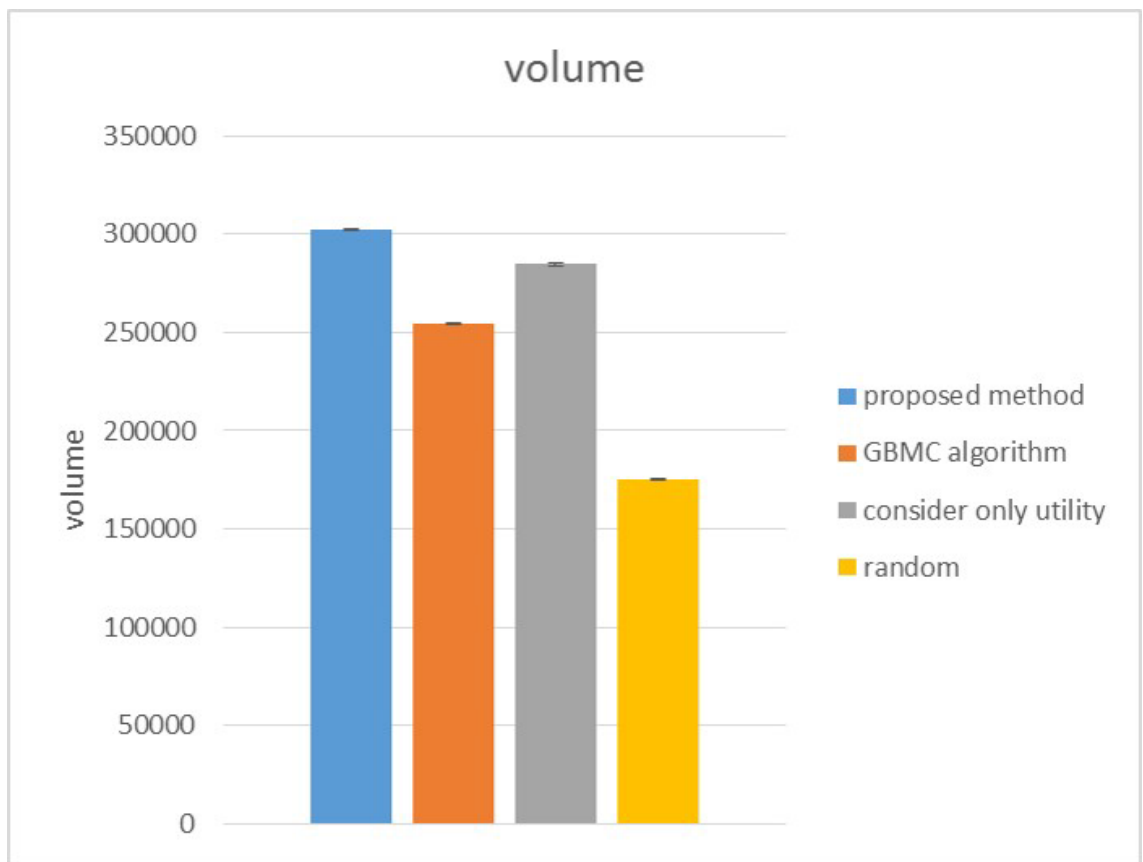


図 6.6: 重ね合わせの体積のシミュレーション結果

表 6.2: シミュレーション結果のまとめ

	総カバレッジ (Avg ± 95%CI)	総ユーティリティ (Avg ± 95%CI)	総体積 (Avg ± 95%CI)
提案手法	131713 ± 204	234.4 ± 0.4	302443 ± 576
GBMC	136134 ± 199	195.2 ± 0.4	254631 ± 608
質のみ重視	118269 ± 232	245.2 ± 0.4	284803 ± 622
ランダム	93157 ± 228	141.8 ± 0.4	175150 ± 562

を大きくすることに成功していることが分かる。(表 6.2) よりそれぞれの手法毎の正確な平均値は先述の順番に 302443, 254631, 284803, 175150 であり, 提案手法は GBMC アルゴリズムと比較して+18.8%程度, 質のみを重視したノード選択手法と比較して+6.2%程度, ランダムなノード選択手法と比較して+72.7%程度の値となっている。

これらの結果から, 提案手法はデータのカバレッジ・質の両面において安定したノード選択が実現できており, 提案したアルゴリズムはクラウドセンシングにおけるノード選択の手法として有用であるといえる。

6.3 実データを用いた評価

前節ではシミュレーションデータによる評価実験を行い, 提案手法の有効性を示した。本節では実データを用いた比較評価を行い, 提案手法が実データに対しても有効であることを示す。本節では, まず実験環境として利用する実データセットである人の流れデータとその他の環境設定について説明をする。次に, 動的なノード選択を行った場合の比較評価を行う。最後に, 静的なノード選択を行った場合の比較評価を行う。

6.3.1 実験環境

6.1 では比較評価のためのユーザの位置データは一様分布によって生成したが, 本節では人の流れプロジェクト [25] の人の流れデータを使用した。人の流れプロジェクトは東京大学空間情報科学研究センターによるデータ提供サービスで, 国土交通大臣や各都市圏の交通計画協議会, 国際協力機構 (JICA) の許諾をもとに, 国や地方自治体等で調査を行っているパーソントリップ調査データ (PT データ) から作成した時空間位置での人々の流動データを研究者や実務者などに提供している。(図 6.7) に人の流れデータの利用の一例として人の分布状況を可視化したものを示す。

今回, 実験には平成 20 年東京都市圏の 20 万人分の移動データを用いた。また, 対象とする空間の範囲は東京都市圏の中で北緯 35.4 度から北緯 35.85 度, 東経 139.4 度から 139.95 度の約 50km × 約 50km の正方形となる空間に絞った。その範囲を (図 6.8) に示す。

実データである人の流れデータは 6.1 で使用したシミュレーションデータとは異なり, 時間帯によって人々の位置に偏りが生じており, より実際のクラウドセンシングに近い環境での実験が期待できる。



図 6.7: 人の流れデータから作成した人の分布状況



図 6.8: 今回対象とした範囲

表 6.3: 実験のパラメータ設定

各パラメータ	パラメータの値
試行回数	100
総ノード数	300
総コスト上限	200
空間の x 幅	50km
空間の y 幅	50km
ノード毎のカバレッジ半径	2km
ノード毎のデータの質の分布	$N(5, 2)$
ノード毎のコストの分布	$N(3.5, 1.4)$
測定期間	5:00 ~ 24:00
測定間隔	1h

実験に使用したパラメータを（表 6.3 に示す．各ノードの位置については，20 万人分の人の流れデータから 300 人分のデータを無作為に抽出し，これを利用した．総コスト上限 300 に対してノード毎のコストの平均を 3.5 に設定し，300 ノード中 60~80 ノード程選択することを想定している．測定の期間は人間の主な活動期間である午前 5 時から午後 12 時とし，一時間毎の結果の変化を観察した．データに時間的な連続性があると仮定するとデータのカバレッジや質は前後の時間の測定値からの影響を受けると考えられるが，本実験では簡単のためにその時間のみのデータを参照した．今回はその実験を 100 回繰り返して結果の平均値を比較した．比較対象とするノード選択手法にはと同じく提案手法，GBMC アルゴリズム，データの質のみを考慮したノード選択手法，ランダムなノード選択手法の四種類を用いた．

6.3.2 動的なノード選択を行った場合

まず，動的なノード選択を行った場合について実験を行った．ここでいう動的なノード選択とは測定間隔である一時間毎にノード選択を再度行うことであり，常にノードの最適化が行われている．そのため人の流れによるカバレッジへの影響はあまり出ず，前節のシミュレーションと同じような結果が期待できる．

データのカバレッジの時間推移の平均の比較結果を（図 6.9）に，データの質の時間推移の平均の比較結果を（図 6.10）に，データのカバレッジと質の平均の比較結果を（図 6.11）に，実験結果のまとめを（表 6.4）に示す．

（図 6.9）と（図 6.10）を見ると，毎時間ノード選択をやり直したことにより，時間変化によるカバレッジや質への影響はあまり出ていないことが分かる．

（図 6.11）では横軸がデータのカバレッジ，縦軸がデータの質の平均値をそれぞれ示しており，シミュレーション時と同様に提案手法が右上に位置していることから提案手法がカバレッジと質の両面を総合的に高めていることが分かる．（表 6.4）よりカバレッジの正確な平均値は提案手法，GBMC アルゴリズム，質のみを重視したノード選択手法，ラン

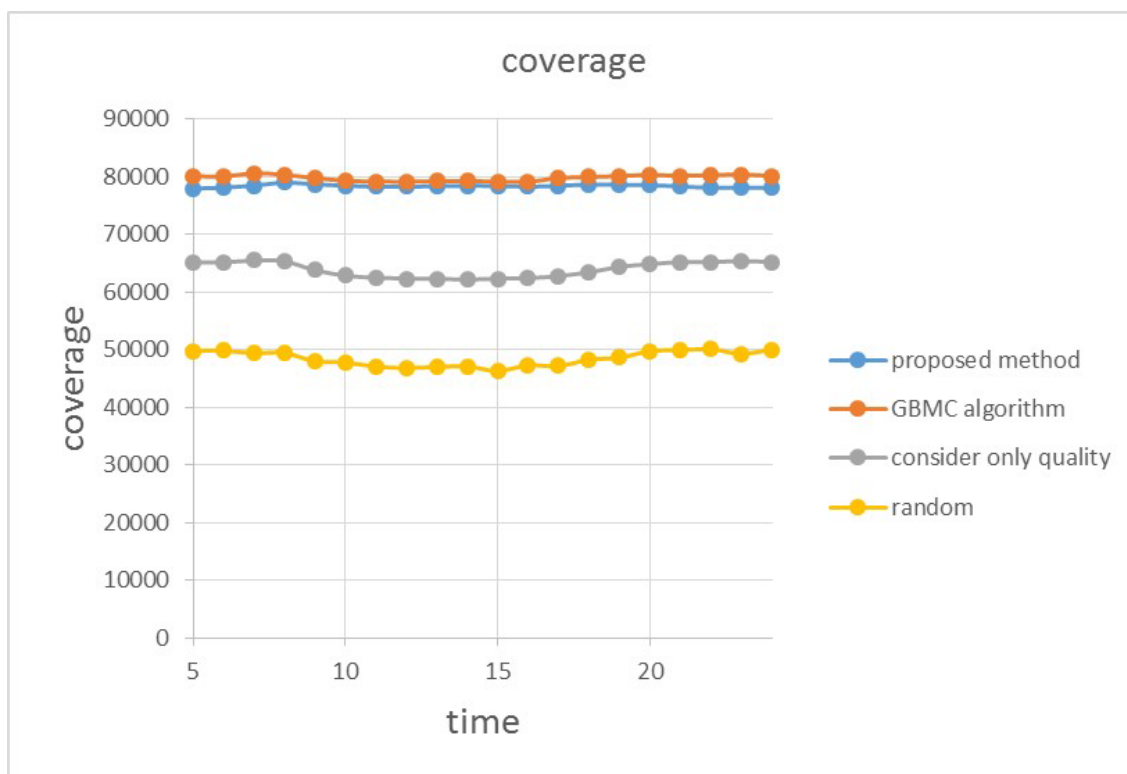


図 6.9: 動的なノード選択を行った場合のカバレッジの時間推移の平均の比較結果

ダムなノード選択手法の順に 78392, 79845, 63920, 48441 であり, 提案手法はカバレッジのみを重視したノード選択手法である GBMC アルゴリズムと比較しても-1.8%程度とほぼ同等のカバレッジの確保に成功している. また, 質のみを重視したノード選択手法と比較すると+22.6%程度, ランダムなノード選択手法と比較すると+61.8%程度のカバレッジとなっている. また, 質の正確な平均値は先述の順番に 522.2, 426.2, 567.2, 285.0 であり, 提案手法は質のみを重視したノード選択手法と比較して-7.9%程度のデータの質の確保に成功している. また, GBMC アルゴリズムと比較すると+22.5%程度, ランダムなノード選択手法と比較すると+83.2%程度のデータの質となっている.

これらの結果から, 提案手法は動的なノード選択を行う場合, シミュレーション時と同様にデータのカバレッジ・質の両面において安定したノード選択が実現できているといえる.

6.3.3 静的なノード選択を行った場合

実際にクラウドセンシングにおいては, 逐一位置情報を送信してノード選択を待つ必要のある動的なノード選択は消費電力やコストの観点からあまり現実的ではない. そこで, 最初の一度だけノード選択を行い, その後は選ばれたノードのみから定期的に取り得データを送ってくるというシナリオの静的なノード選択を行った場合についての実験を行った.

データのカバレッジの時間推移の平均の比較結果を (図 6.12) に, データのカバレッジと質の平均の比較結果を (図 6.13) に, 実験結果のまとめを (表 6.5) に示す.

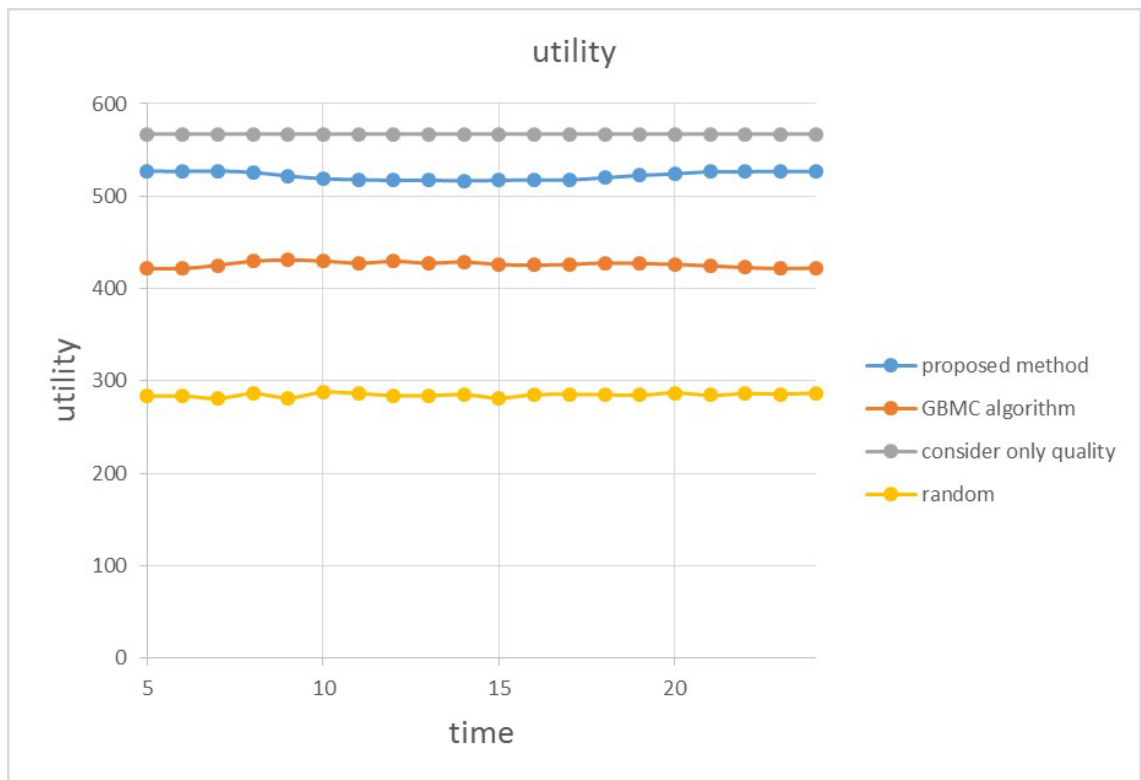


図 6.10: 動的なノード選択を行った場合の質の時間推移の平均の比較結果

(図 6.12) に動的なノード選択を行った場合では見られなかった時間経過による影響が見られる。7時から9時にかけておそらく通学や出勤による人々の移動が起き、全体的なカバレッジの減少へとつながっている。その中で提案手法と GBMC アルゴリズムに着目すると、5時時点では 2000 ほどあったカバレッジの差が 12 時前後ではほとんどなくなっていることが分かる。これは提案手法が 5.3 の円錐形の評価値を用いたことによって人の動きによる影響を緩和できたためだと考えられる。

(図 6.13) では横軸がデータのカバレッジ、縦軸がデータの質の平均値をそれぞれ示しており、これまでと同様に提案手法が右上に位置していることから提案手法がカバレッジと質の両面を総合的に高めていることが分かる。(表 6.5) よりカバレッジの正確な平均値は提案手法, GBMC アルゴリズム, 質のみを重視したノード選択手法, ランダムなノード選択手法の順に 70475, 71763, 63919, 48268 であり, 提案手法はカバレッジのみを重視したノード選択手法である GBMC アルゴリズムと比較しても-1.8%程度とほぼ同等のカバレッジの確保に成功している。また, 質のみを重視したノード選択手法と比較すると+10.2%程度, ランダムなノード選択手法と比較すると+45.9%程度のカバレッジとなっている。また, 質の正確な平均値は先述の順番に 525.3, 422.0, 565.1, 284.5 であり, 提案手法は質のみを重視したノード選択手法と比較して-7.0%程度のデータの質の確保に成功している。また, GBMC アルゴリズムと比較すると+24.5%程度, ランダムなノード選択手法と比較すると+84.6%程度のデータの質となっている。

これらの結果から, 提案手法は静的なノード選択を行う場合であっても, ノードの移動

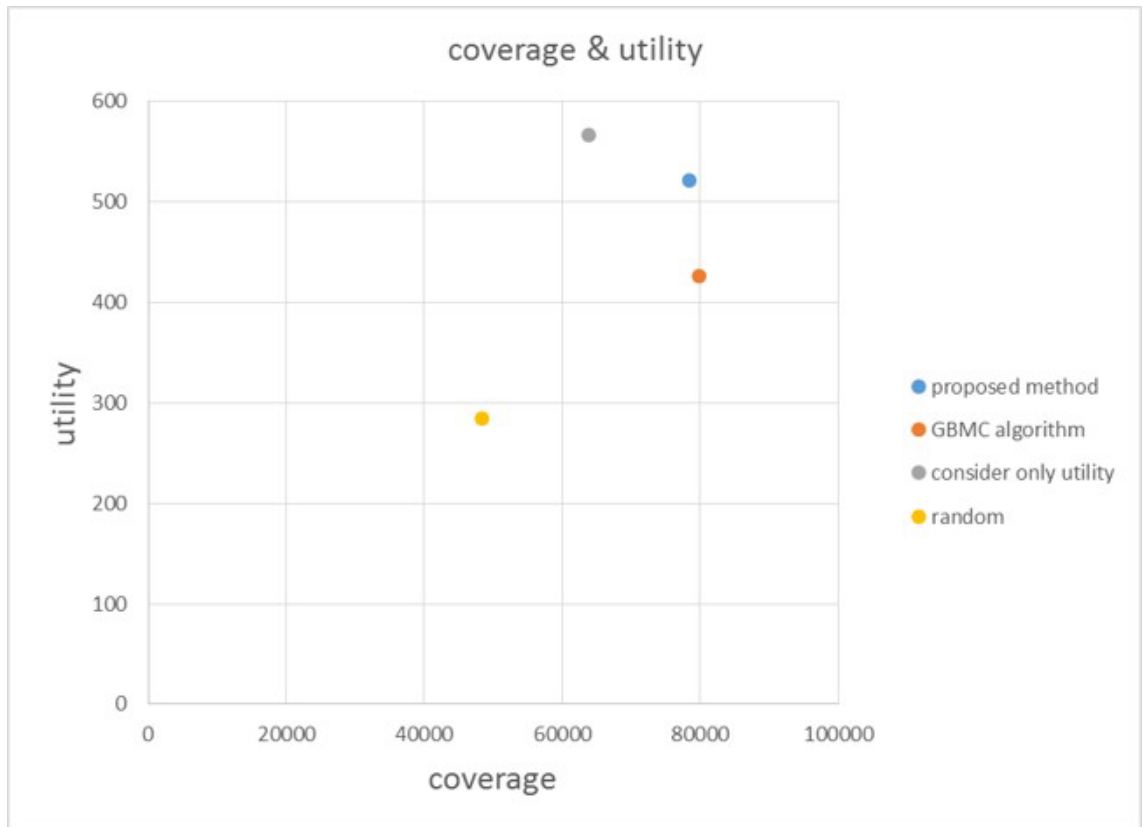


図 6.11: 動的なノード選択を行った場合のカバレッジと質の平均の比較結果

によるデータカバレッジへの影響を緩和しつつ，データのクバレッジ・質の両面において安定したノード選択を実現できているといえる。

表 6.4: 動的なノード選択を行った場合の実験結果のまとめ

	総カバレッジ (Avg ± 95%CI)	総ユーティリティ (Avg ± 95%CI)	総体積 (Avg ± 95%CI)
提案手法	78392 ± 515	522.2 ± 3.9	193040 ± 1365
GBMC	79845 ± 487	426.2 ± 4.6	155249 ± 1666
質のみ重視	63920 ± 621	567.2 ± 4.4	167680 ± 1578
ランダム	48441 ± 615	285.0 ± 4.2	95041 ± 1480

表 6.5: 静的なノード選択を行った場合の実験結果のまとめ

	総カバレッジ (Avg ± 95%CI)	総ユーティリティ (Avg ± 95%CI)	総体積 (Avg ± 95%CI)
提案手法	70475 ± 565	525.3 ± 3.7	194050 ± 1111
GBMC	71763 ± 552	422.0 ± 4.0	156701 ± 1381
質のみ重視	63919 ± 665	565.1 ± 4.0	170873 ± 1420
ランダム	48268 ± 594	284.5 ± 3.8	97647 ± 1457

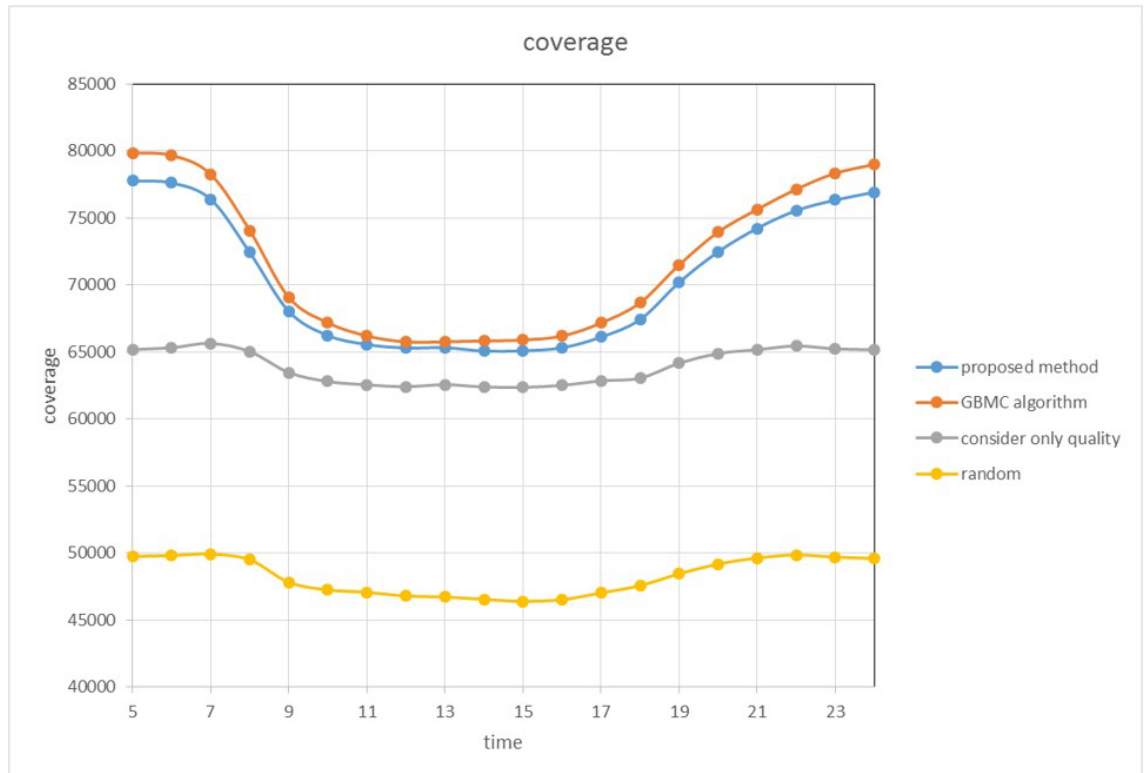


図 6.12: 静的なノード選択を行った場合のカバレッジの時間推移の平均の比較結果

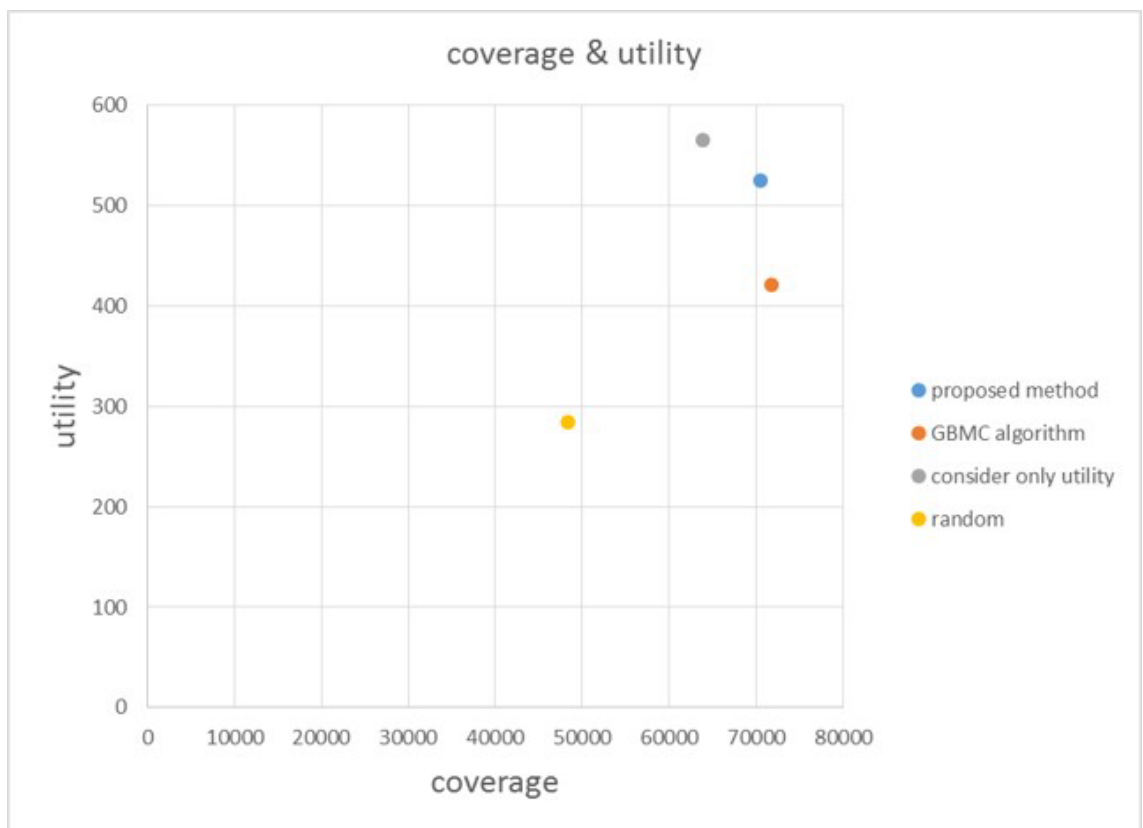


図 6.13: 静的なノード選択を行った場合のカバレッジと質の平均の比較結果

第7章 結論

7.1 まとめ

本研究ではクラウドセンシングの概要とその課題であるデータの質とカバレッジの重要性について関連研究を絡めて議論し，その両方を考慮したノード選択が可能な新しいアルゴリズムの提案と，シミュレーションによる既存手法との比較評価および実データによる比較評価を行った．本節ではそれらを整理する．

はじめに，通信モジュールを備えたモバイルデバイスの普及を背景に期待が高まっているクラウドセンシングについて解説を行った．クラウドセンシングには従来のセンサネットワークと比較してセンサ設置のコストが不要であるという大きな利点があり，環境調査や交通量調査，ヘルスマニタリングなど様々な方面へと活かすことが可能であるが，第2章にて以下の三つの課題があることを明らかにした．

- インセンティブの確保
- データのカバレッジ
- データの質

そして，第3章でこれらに関する研究をまとめた結果，インセンティブの確保に関しては不確定な要素が多く一般化が難しいこととデータのカバレッジ・質の両面について言及されている研究の数が乏しいことが分かったため，第4章，第5章でデータのカバレッジ・質の両面を高めるようなノード選択手法を提案した．

第4章ではデータのカバレッジ・質の両面を高めるようなノード選択を行うに当たってデータ収集の事前にデータの質の定量化を行う必要があること述べ，その方法論について議論した．

第5章ではデータのカバレッジのみを考慮した既存のノード選択手法である GBMC アルゴリズムに第4章で提案した定量化されたデータの質を取り入れ拡張することによって，データのカバレッジと質の両方を考慮したノード選択が可能なアルゴリズムを提案した．また，評価値の形状を変化させることでユーザの移動による影響を緩和する工夫も加えた．

第6章ではシミュレーションデータ，実データのそれぞれに手法を適用し，比較実験を行った．実データに関しては動的なノード選択，静的なノード選択のそれぞれについて実験を行い，結果を示した．実験の結果，いずれの場合も提案手法は他の手法に比べてデータのカバレッジ・質の両面を総合的に高めることに成功していることが示された．また，静的なノード選択を行った場合，提案手法はユーザの移動による影響を緩和できているということも示すことが出来た．これらの結果から，提案したノード選択アルゴリズムはクラウドセンシングにおいて安定したカバレッジと質のデータを収集するに当たって優れていることが分かった．

7.2 今後の展望

本研究で提案したノード選択アルゴリズムについて、今後の展望を述べる。まず、クラウドセンシングの実際の運用にあたっては収集するデータの質をどのように重視するかによってデータの質の定量化の方法は変わってくるため、より具体的な方法論が求められるということが今後の課題として挙げられる。また、データが時間的な連続性を持っていた場合、空間的なカバレッジだけでなく時間的なカバレッジについても同様に議論の余地があるため、これについても検討が必要であるといえる。

本研究で提案した手法とこれらの課題の解決によって、より効率的なクラウドセンシングが実用化され则认为している。環境調査や交通量調査といった単純な情報収集からヘルスマonitoringのような医療分野まで応用がきくクラウドセンシングの効率化は、人々の暮らしをより豊かでスマートなものに変えていくと確信している。

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る。

本研究の遂行にあたり、指導教官として研究に対する助言だけでなく、学生生活や人生訓など様々な面でご指導くださった 瀬崎薫 教授に、心より感謝の意を表す。また、研究の立ち上げ時に様々な方向性からの助言をいただき、研究の遂行にあたって終始ご指導いただいた東京大学 空間情報科学研究センター 小林博樹 助教と東京大学 生産技術研究所 伊藤昌毅 助教の両氏に深く感謝の意を表す。

最後に、日頃より有益な議論や助言をして下さった、東京大学 瀬崎研究室 博士課程の 青木俊介氏, Niu Hao 氏, Dunstan Matekenya 氏, 江甜甜氏, 同期としてお互いに高め合いながら楽しい時間を共有してくれた中村直人 氏, 稲葉瞳 氏, 加藤宗肖 氏, 後輩でありながら高い能力と意欲によって刺激を与えてくれつつも大変親しく接してくれた森英記 氏, 牧山紘 氏, 松野有弥 氏, 山本直人 氏, 鵜飼祐太 氏, 常日頃から研究生生活を支えてくださった秘書の 松本夏穂 氏, 内藤潤 氏に、感謝の意を表し、本論文を閉じる。

発表文献 / Publications

- [1] 坂本敬太, 青木俊介, 伊藤昌毅, 瀬崎薫, ”参加型センシングにおけるカバレッジと質を考慮したノード選択手法の検討,” 電子情報通信学会 ヒューマンプロープ研究会 (HPB), 2014年02月
- [2] 坂本 敬太, 青木 俊介, 伊藤 昌毅, 瀬崎 薫,”参加型センシングにおけるカバレッジと質を考慮したノード選択手法の提案と評価,” マルチメディア, 分散, 協調とモバイル (DICOMO2014) シンポジウム, 4G-2, 2014年7月
- [3] 坂本 敬太, 青木 俊介, 伊藤 昌毅, 瀬崎 薫,”クラウドセンシングにおけるノード選択手法の実データによる評価,” 電子情報通信学会 2015年総合大会, B-18-74, 2015年3月

参考文献 / References

- [1] N.D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A.T. Campbell, “A survey of mobile phone sensing,” *Communications Magazine, IEEE*, vol.48, no.9, pp.140–150, 2010.
- [2] 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.1, pp.41–50, 2014.
- [3] S. Reddy, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, “A framework for data quality and feedback in participatory sensing,” *Proceedings of the 5th international conference on Embedded networked sensor systemsACM*, pp.417–418 2007.
- [4] J.-S. Lee and B. Hoh, “Dynamic pricing incentive for participatory sensing,” *Pervasive and Mobile Computing*, vol.6, no.6, pp.693–708, 2010.
- [5] L.G. Jaimes, I. Vergara-Laurens, and M.A. Labrador, “A location-based incentive mechanism for participatory sensing systems with budget constraints,” *Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference onIEEE*, pp.103–108 2012.
- [6] D. Yang, G. Xue, X. Fang, and J. Tang, “Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing,” *Proceedings of the 18th annual international conference on Mobile computing and networkingACM*, pp.173–184 2012.
- [7] S. Reddy, D. Estrin, M. Hansen, and M. Srivastava, “Examining micro-payments for participatory sensing data collections,” *Proceedings of the 12th ACM international conference on Ubiquitous computingACM*, pp.33–36 2010.
- [8] M. Musthag, A. Raij, D. Ganesan, S. Kumar, and S. Shiffman, “Exploring micro-incentive strategies for participant compensation in high-burden studies,” *Proceedings of the 13th international conference on Ubiquitous computingACM*, pp.435–444 2011.
- [9] A. Albers, I. Krontiris, N. Sonehara, and I. Echizen, “Coupons as monetary incentives in participatory sensing,” *Collaborative, Trusted and Privacy-Aware e/m-Services*, pp.226–237, Springer, 2013.
- [10] W. Mason and D.J. Watts, “Financial incentives and the performance of crowds,” *ACM SigKDD Explorations Newsletter*, vol.11, no.2, pp.100–108, 2010.

- [11] M. Von Kaenel, P. Sommer, and R. Wattenhofer, “Ikarus: large-scale participatory sensing at high altitudes,” Proceedings of the 12th Workshop on Mobile Computing Systems and ApplicationsACM, pp.63–68 2011.
- [12] T. Luo and C.-K. Tham, “Fairness and social welfare in incentivizing participatory sensing,” Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2012 9th Annual IEEE Communications Society Conference onIEEE, pp.425–433 2012.
- [13] P. Mohan, V.N. Padmanabhan, and R. Ramjee, “Nericell: rich monitoring of road and traffic conditions using mobile smartphones,” Proceedings of the 6th ACM conference on Embedded network sensor systemsACM, pp.323–336 2008.
- [14] E. Kanjo, J. Bacon, D. Roberts, and P. Landshoff, “Mobsens: Making smart phones smarter,” Pervasive Computing, IEEE, vol.8, no.4, pp.50–57, 2009.
- [15] M. Linnap and A. Rice, “The effectiveness of centralised management for reducing wasted effort in participatory sensing,” Pervasive Computing and Communications (PerCom), 2014.
- [16] R. Kawajiri, M. Shimosaka, and H. Kahima, “Steered crowdsensing: incentive design towards quality-oriented place-centric crowdsensing,” Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous ComputingACM, pp.691–701 2014.
- [17] 瀬崎薫, “ユーザ参加型都市環境センシングとその課題,” マルチメディア, 分散, 強調とモバイル (DICOMO2013) シンポジウム, 2013.
- [18] 猿渡俊介, 森川博之, “ユビキタスセンサネットワーク,” 日本ロボット学会誌, vol.28, no.3, pp.1–4, 2010.
- [19] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, “The pothole patrol: using a mobile sensor network for road surface monitoring,” Proceedings of the 6th international conference on Mobile systems, applications, and servicesACM, pp.29–39 2008.
- [20] J.A. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M.B. Srivastava, “Participatory sensing,” 2006.
- [21] 株式会社ウェザーニューズ, “ウェザーニューズ,” <http://weathernews.jp/>.
- [22] OpenStreetMapFoundation, “Openstreetmap,” <http://www.openstreetmap.org>.
- [23] 武山政直, 津久井かほる, “地理空間的集合知による価値共創サービスの可能性,” 慶応義塾大学日吉紀要 社会科学, pp.73–89, 2009.
- [24] S. Khuller, A. Moss, and J.S. Naor, “The budgeted maximum coverage problem,” Information Processing Letters, vol.70, no.1, pp.39–45, 1999.

- [25] 東京大学空間情報科学研究センター, “人の流れプロジェクト,” <http://pflow.csis.u-tokyo.ac.jp/>.