クラウド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

坂本 敬太[1] 青木 俊介[1] 伊藤 昌毅[1] 瀬崎 薫[1,2]

概要：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．本稿では，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．

A Node Selection Scheme Considering Coverage and Quality of Data for Crowd Sensing

Keita Sakamoto[1] Shunsuke Aoki[1] Masaki Ito[1] Kaoru Sezaki[1,2]

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]．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．

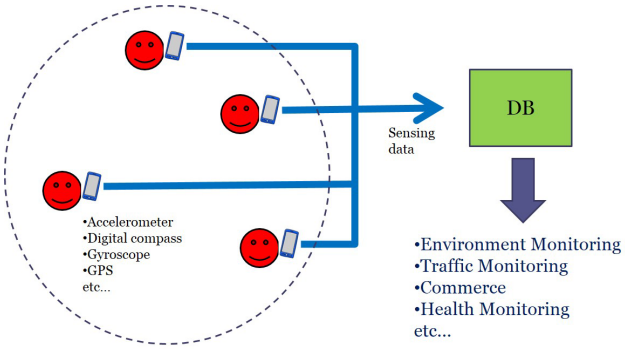

図 1 クラウドセンシング（参加型センシング）の概要

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．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], [4] や，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 [5] が代表的である．特にデータの質や

1 東京大学生産技術研究所 〒153-8505 東京都目黒区駒場 4-6-1
Institute of Industrial Science the University of Tokyo 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, JAPAN
2 東京大学空間情報科学研究センター
Center for Spatial Information Science the University of Tokyo

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], [4], [12]．データの質やカバレッジを安定させるためには，データ収集の対象とするノード（ユーザ）を取捨選択することが考えられる．また，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手法 [5], [7] が存在することも踏まえると，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を行うようなアルゴリズムが望ましい．

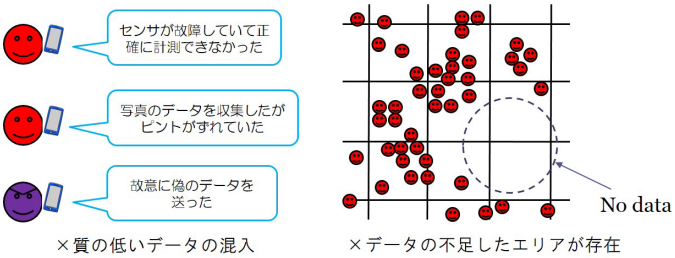

図 2　クラウドセンシングの課題

本稿ではまずクラウドセンシングにおけるデータのカバレッジや質に関する関連研究として GBMC アルゴリズム [4] や centralised management[12]，micro-payments[7] について紹介をして，クラウドセンシングにおけるデータのカバレッジや質の重要性について議論した後，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，シミュレーションによる評価を行い，結論を述べる．

2.　関連研究

本節では，クラウドセンシングにおけるデータのカバレッジや質に関する検討を行っている関連研究について紹介をする．

2.1　GBMC アルゴリズム

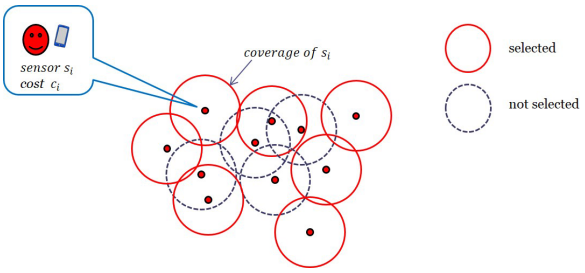

図 3　GBMC アルゴリズムにおけるノードの決定法

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 GBMC（Greedy Budgeted Maximum Coverage）アルゴリズム [4] を紹介する．

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を最大にするようなノード選択を目指す（図 3）．

しかしこれはよく知られた NP 困難な問題である [13]．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 4，図 5）．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 アルゴリズムを使用しない場合のシミュレーション結果（図 4）に比べて GBMC アルゴリズムを使用した場合のシミュレーション結果（図 5）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．

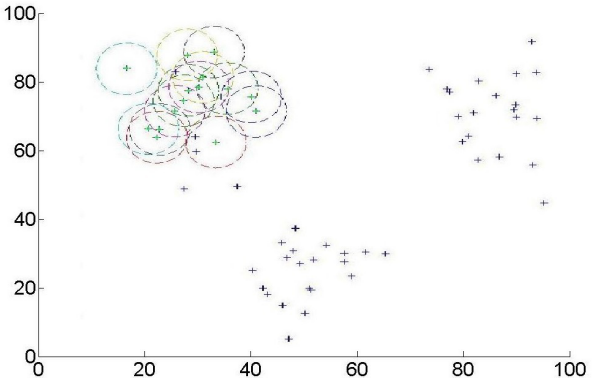

図 4　GBMC アルゴリズムのカバレッジのシミュレーション結果 1

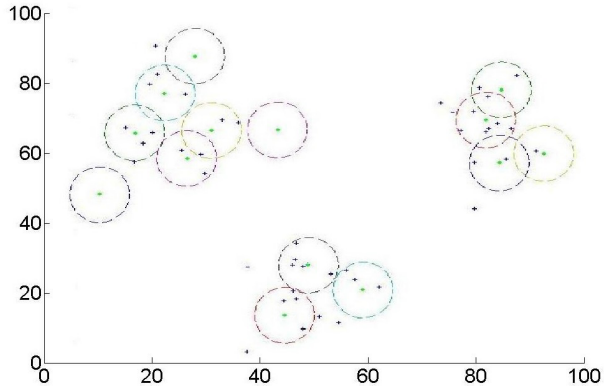

図 5　GBMC アルゴリズムのカバレッジのシミュレーション結果 2

2.2　centralised management

データのカバレッジの評価に関するもう一つの関連研究として，centralised management[12] 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 6）．赤色の部分がデータが足りていないエリア，緑色の部分がデータが過不足なく収集できたエリア，黄色の部分が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も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．

2.3　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．

図 6　centralised management におけるカバレッジの評価

micro-payments は写真をデータとして収集することを目的とした参加型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，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出来る (図 7)．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 8)．

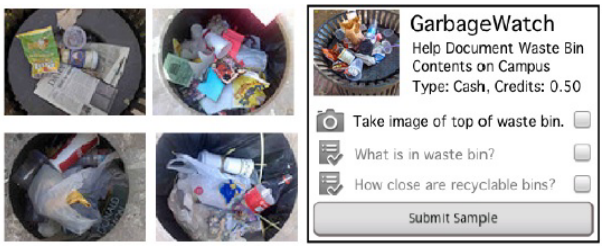

図 7　micro-payments における写真データの収集

このように，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．

Incentive Type	Avg % of Invalid Pics per User	Avg % of Optional Tags per User
MACRO	7±5	70±21
MEDIUMμ	6±4	47±25
HIGHμ	6±3	60±25
LOWμ	5±3	52±24
COMPETEμ	6±3	6±4

図 8　micro-payments における写真の質の評価

3.　提案手法

本節では，クラウドセンシングにおけるカバレッジと質の両面を考慮した新しいノード選択手法を提案する．

3.1　データの質の数値化

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数字化し，パラメータとして扱えるようにする必要がある．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指標となっていたが，他にはそのユーザが持つセンサデバイスの単純な機械的性能（カメラであれば画素数，加速度センサであれば分解能など）や，収集されたデータの数の多さ等もデータの質を計る指標として考えられる．その他では，長期的にセンシングが行われる場合，ネットオークションにおける出品者の評価のように過去の収集データの精度からユーザ毎に評価を付け，そのユーザの評価とデータの質を結びつけるような方法も考えられる．関連研究においてもこれらの指標を用いてデータの質が評価されており，データの質の重要性が示されているといえる [3], [7]．本稿では，以上のような方法を用いてデータの質を評価・数値化し，「data utility」としてデータのパラメータとして扱う事を考える（図 9）．

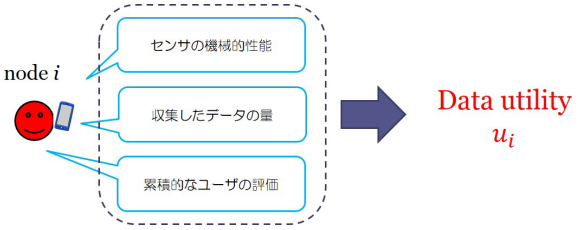

図 9　データの質の数値化

3.2　ノード選択のアルゴリズム

前節で述べたような方法を用いてデータの質が数値化されている前提で，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な手法を提案する．GBMC アルゴリズムの考え方を拡張して，各センサノードを中心とし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 10）．この円柱の重ね合わせの体積をより大きくするようなノード選択を行うことで，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 11）．この手法はクラウドセンシングだけに限らず，クオリティコントロールが求められるデータ収集を行う様々な分野に応用が可能であり，幅広い可能性を秘めているといえる．

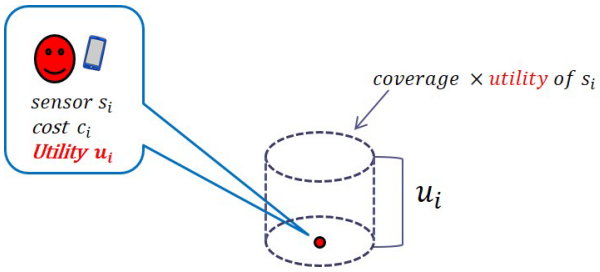

図 10　提案手法におけるカバレッジを底面，質を高さとした円柱を用いた評価値

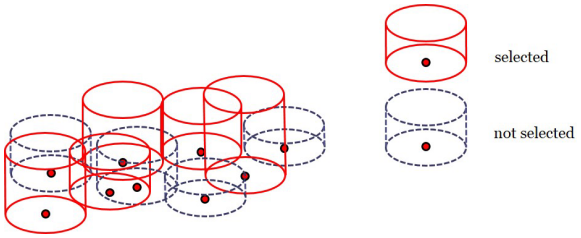

図 11　提案手法：円柱の重ね合わせの体積を最大化することでカバレッジと質を総合的に最適化するノード選択

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 $C + C_i \leq$ 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

表 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)

4.　シミュレーション評価

本節では，提案手法と他の手法のシミュレーションによる比較評価の結果について述べる．

シミュレーションには（表 1）のようなパラメータを使用した．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．

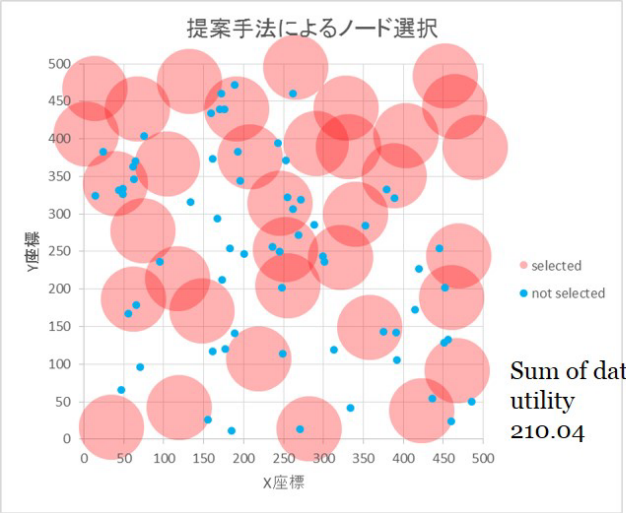

図 12　ノード選択の結果：提案手法

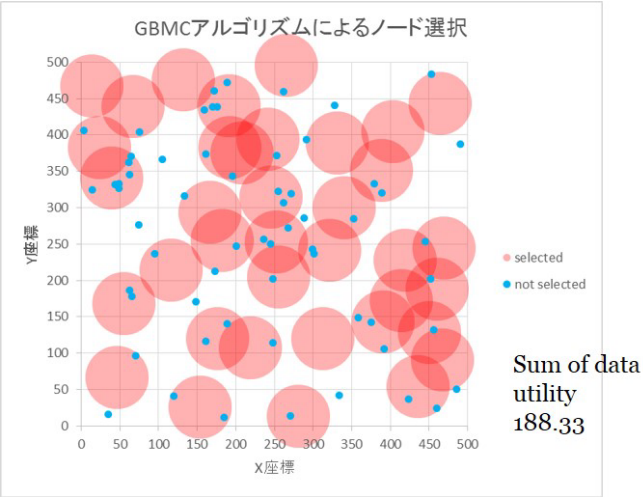

図 13　ノード選択の結果：GBMC アルゴリズム

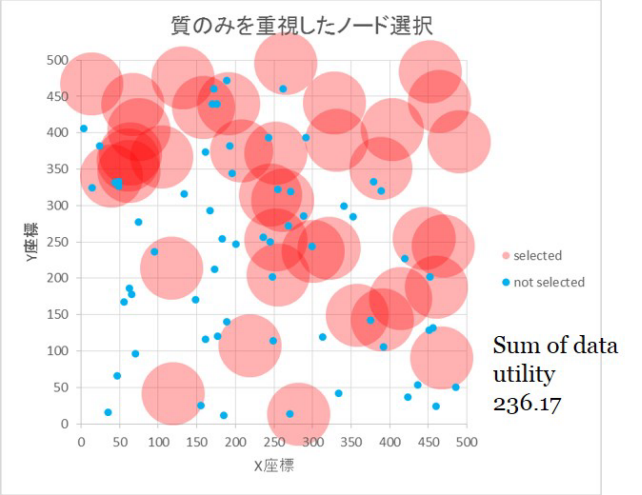

図 14　ノード選択の結果：質のみ重視

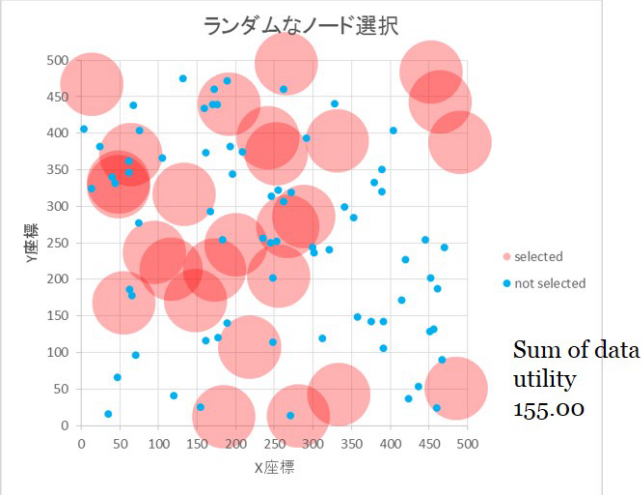

図 15　ノード選択の結果：ランダム

10000 回のシミュレーションの比較結果の内の一例を (図 12）〜（図 15) に，データのカバレッジの平均の比較結果を (図 16) に，データの質の平均の比較結果を (図 17) に，シミュレーション結果のまとめを（表 2) に示す．(図 12）〜

(図 15) 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 16) と (図 17) に関してはそれぞれの手法の平均値の 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 12）～（図 15) 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．そして，実際の全体面積の平均値の比較が (図 16) のグラフに示されている．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 129816，136065，118204，92909 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-5%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+10%程度，ランダムなノード選択手法と比較すると+40%程度のカバレッジとなっている．

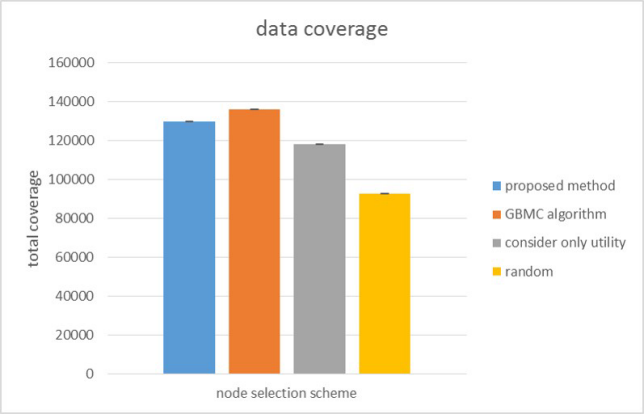

図 16　データの総カバレッジの比較

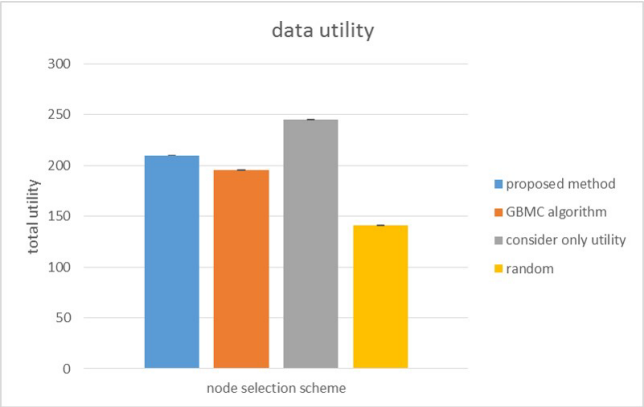

図 17　データの総ユーティリティの比較

（図 17）のグラフにはデータの質の合計の平均値の比較が示されている．それぞれの手法毎の値は先述の順番に 210.0，195.1，245.0，141.4 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-14%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+8%程度，ランダムなノード選択手法と比較すると+49%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．

表 2　シミュレーション結果のまとめ

	総カバレッジ (Avg ± SD)	総ユーティリティ (Avg ± SD)
提案手法	129816 ± 7746	210.0 ± 13.4
GBMC	136065 ± 7680	195.1 ± 16.0
質のみ重視	118204 ± 8938	245.0 ± 15.0
ランダム	92909 ± 8790	141.4 ± 15.0

5.　おわりに

本稿では，クラウドセンシングにおけ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を行った．本節では，それらを整理し，今後の課題を述べる．

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 micro-payments 等に見られるデータの質の評価の概念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，10000 回のシミュレーションによる比較評価によって，クラウドセンシングにおけるこのアルゴリズムの有用性を示した．このアルゴリズムの利用によって，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幅広い分野でのより利用価値の高いデータ収集の実現が期待できる．

今後の課題として，より効率的なノード選択アルゴリズムの実現や，データの質を数値化する際のより具体的な方法論，各ノードの位置とデータが時間変化する場合の動的なノード選択手法の確立などが挙げられる．現在，筆者らはそれらに対する検討を進めている．

参考文献

[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] 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.

[5] J.-S. Lee and B. Hoh, "Dynamic pricing incentive for participatory sensing," Pervasive and Mobile Computing, vol.6, no.6, pp.693–708, 2010.

[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] 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.

[10] 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.

[11] 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.

[12] M. Linnap and A. Rice, "The effectiveness of centralised management for reducing wasted effort in participatory sensing," Pervasive Computing and Communications (PerCom), 2014.

[13] S. Khuller, A. Moss, and J.S. Naor, "The budgeted maximum coverage problem," Information Processing Letters, vol.70, no.1, pp.39–45, 1999.