

分散サンプリングストレージを用いた電波利用状況可視化

長縄 潤一[†] 井上 武^{††} 佐藤 浩史^{††} 山崎 敬広^{††} 猿渡 俊介[†]
森川 博之[†]

[†] 東京大学先端科学技術研究センター, 〒 153-8904 東京都目黒区駒場 4-6-1

E-mail:{naga,saru,mori}@mlab.t.u-tokyo.ac.jp

^{††} 日本電信電話株式会社 NTT 未来ねっと研究所, 〒 239-0847 神奈川県横須賀市光の丘 1-1

E-mail:sato.hiroshi@lab.ntt.co.jp

あらまし 現在, 周波数帯域の有効利用に向けて, コグニティブ無線等の技術開発や電波政策の議論が活発に行われている. 筆者らは, これらの取り組みを支援するために, 電波利用状況を表すデータであるスペクトラムを多地点で測定し, ウェブ上で可視化・公開する分散スペクトラムセンシングの研究開発を進めている. スペクトラムの可視化にあたっては, ユーザがウェブ上でインタラクティブな閲覧を行うことから, 多様な可視化要求を迅速に処理することが求められる. 本稿では, センサデータマイニング向けのストレージ技術である分散サンプリングストレージを用いることで, 可視化の処理時間を削減できることを示す.

キーワード 電波利用状況, 可視化, コグニティブ無線, 分散スペクトラムセンシング, 分散サンプリングストレージ

Spectrum Visualization using Distributed Sampling Storage

Junichi NAGANAWA[†], Takeshi INOUE^{††}, Hiroshi SATO^{††}, Takahiro YAMAZAKI^{††},
Shunsuke SARUWATARI[†], and Hiroyuki MORIKAWA[†]

[†] Research Center for Advanced Science and Technology, The University of Tokyo

Komaba 4-6-1, Meguro-ku, Tokyo, 153-8904, Japan

E-mail:{naga,saru,mori}@mlab.t.u-tokyo.ac.jp

^{††} NTT Network Innovation Laboratories, NTT Corporation

Hikarinooka 1-1, Yokosuka-shi, Kanagawa, Japan 239-0847

E-mail:sato.hiroshi@lab.ntt.co.jp

Abstract Towards more efficient spectrum utilization, R&D on cognitive radios and discussion on spectrum policy have been active. To support these activities, we have been working on the distributed spectrum sensing system, which realizes spectrum measurement at multiple locations and visualization of the measurement data on the web. Because users interactively browse the data, the system should quickly respond the user's visualization request with varying parameters. This paper shows fast spectrum visualization by using the distributed sampling storage for sensor-data mining.

Key words Spectrum Utilization, Visualization, Cognitive Radio, Distributed Spectrum Sensing, Distributed Sampling Storage

1. はじめに

筆者らは, 電波資源有効利用に向けては, 電波の利用状況をきめ細かく把握することが鍵であると考え, 多地点でスペクトラムを測定する分散スペクトラムセンシングシステムの研究開発を進めている [1]. 従来のスペクトラム測定 [2]~[4] では, 測

定者が可視化まで行っていた. そのため, 測定者の意図に基づく可視化結果のみが利用可能であり, 他のユーザの可視化要求を満たすことができていなかった. 本システムは, 様々なユーザの要求に合致した可視化結果を提供するために, 可視化サービスを通してウェブ上でスペクトラムデータを可視化・公開することを目指している.

図1に可視化サービスのシステム構成を示す。可視化サービスはユーザインタフェースと可視化サーバで構成される。ユーザインタフェースはブラウザ上で動作し、可視化手法や周波数・時間に関する範囲・分解能といった可視化パラメータの選択をユーザに提供する。可視化サーバは、ユーザインタフェースの発行する可視化クエリに応じて、収集・蓄積したスペクトラムデータの読み出しと可視化処理を行う。ユーザの利便性確保に向けては、可視化クエリへの迅速な応答が求められるため、膨大なスペクトラムデータを高速に可視化できなければならない。

本稿では、サンプリング処理を用いた高速なスペクトラムの可視化について述べる。可視化の精度に基づき、サンプルサイズを適切に決めることで、可視化処理を行うサンプルサイズを削減することができる。サンプリング処理では、ストレージへのアクセスがオーバーヘッドとなるため、センサデータマイニング向けの分散ストレージ技術である分散サンプリングストレージ[5]を利用する。分散サンプリングストレージの利用に向けて、サンプルサイズ調節のパラメータである抽出率の決定方法を、2値化と占有率推定の場合について示す。

本稿は以下のように構成される。2.では、既存研究を概観し、可視化サービス提供の必要性を示す。また、可視化サービスの実現に向けては、サンプルサイズ増大による可視化の処理時間増大という問題があることを示す。3.では、分散サンプリングストレージがサンプルサイズ増大の問題を解決することを示す。4.では、可視化の精度を保証可能なサンプルサイズおよび抽出率の決定方法を示す。5.では、模擬環境における性能評価を示す。最後に6.でまとめとする。

2. スペクトラムの可視化

周波数帯域の有効利用に向けた取り組みに向けては、基礎的なデータである電波利用状況の測定・可視化・公開が不可欠である。電波利用状況を表す指標としては、受信した電波の強度を示すスペクトラムが基本となる。スペクトラムデータは、コグニティブ無線[6]等の技術開発において、スペクトル検出といった要素技術の評価に使用することができる[7],[8]。また、電波政策の議論において、周波数帯域が二次利用可能かどうかといった政策決定の参考資料となる[9]。

従来のスペクトラム測定では、測定者が可視化まで行っている[2]~[4]。例えば、[2]では、屋外でスペクトラムを数ヶ月にわたって測定し、帯域占有率の時間的な変動のプロットを作成している。しかしながら、測定者が可視化まで行う場合、測定者

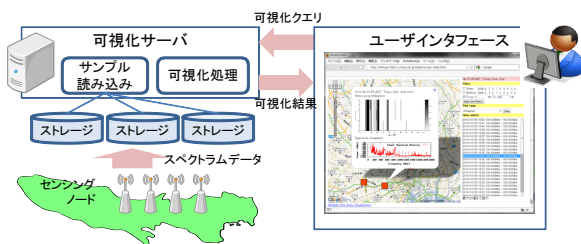


図1 可視化サービスのシステム構成

の意図した用途に向けた可視化結果のみが提供される。スペクトラムデータは様々な用途に利用できるため、他のユーザは異なる可視化手法や可視化パラメータを求める可能性がある。

測定者から可視化のプロセスを分離・公開することで、より多くのユーザの要求に合致した可視化結果を提供することができる[1][10][11]。可視化のプロセスを分離・公開する方法としては、ウェブ上での可視化サービスの提供[1],[11]と未加工データの公開[10]が挙げられる。特に、ウェブ上の可視化サービスは、ブラウザがあれば利用可能で、かつGUIも提供できるため、特別な機材や専門知識を持たない一般ユーザの利用を促進できる[12],[13]。

可視化サービスの提供に向けては、膨大な測定データの読み出しや可視化処理による処理時間の増大を解決しなければならない。例えば、スペクトラムアナライザ R&S FSL 6 を用いると、時間分解能 1 sec, 周波数分解能 300 kHz で 100 MHz ~ 1 GHz を測定できる。24 時間分のデータを可視化すると、サンプルサイズは $(900\text{MHz}/300\text{kHz}) \cdot (24\text{hour}/1\text{sec}) = 2.6 \times 10^8$ となる。このような膨大な数のサンプルの読み出しは、ストレージにおけるディスクアクセス、ストレージ-可視化サーバ間のデータ転送を増大させる。また、膨大な数のサンプルに対する可視化処理は CPU 時間を増大させる。

処理時間の増大に対する解決策としては、あらかじめユーザの希望する可視化要求を予想し、可視化済みのデータや間引き等を施した中間データを準備しておく方法が考えられる。しかしながら、可視化の手法や可視化のパラメータは様々に変化するため、全てのパターンに対応することができない。例えば、電波の利用状況の長期的な変化傾向を見る場合と現在の利用状況を見る場合では、時間的な縮尺や範囲が異なる。そのため、生データからの可視化を、より高速に実行できるようにしなければならない。

3. 分散サンプリングストレージの利用

より高速にスペクトラムの可視化を実行するために、可視化の際に何らかの統計処理が測定データに施されることに着目する。測定時には高分解能で測定が行われるが、可視化時には表示端末の解像度等の問題から、全ての測定データを可視化することができない。そのため、可視化する情報としては、平均値、電波の有無、占有率といった統計量が中心となる。

一般に、統計処理の精度はサンプルサイズの減少に従って低下する。統計処理の精度の劣化が許容範囲内であれば、サンプルを削減することで、可視化の信頼性を保ったまま高速化を実現できる。ただし、統計処理の精度を含めた可視化のパラメータは、可視化の目的に応じて様々に変化するため、可視化クエリ毎の動的なサンプルサイズの調節が必要となる。

本研究では、動的なサンプルサイズの調節に、センサデータマイニング向けの分散ストレージ技術である分散サンプリングストレージ[5]を用いる。分散サンプリングストレージは、サンプルの読み出し時に、抽出率を可視化クエリに応じて設定することで、サンプルサイズを動的に調節する枠組みを備える。抽出率はサンプルサイズ決定のためのパラメータで、保存した

サンプルサイズに対する読み出しサンプルサイズの比を表す。抽出率の決定方法は可視化手法毎に用意する。

抽出率を変化させるサンプリング機能は、ストレージからの読み出し機能に内包されているため、サンプリング対象外のサンプルには一切のアクセスを生じない。例えば、分散サンプリングストレージと比較して、ファイルシステムではディスクのシークが発生する。コンシステント・ハッシュ法 [14] を利用したデータベースでは全サンプルに対する検索が必要となる。データベースからのサンプリングを効率的に実行する手法も提案されているが [15]~[18], ソフトウェア上でサンプリングを実行するもので、分散サンプリングストレージの方がより高速なサンプリングが可能と考えられる。

分散サンプリングストレージの動作原理を図 2 に示す。書き込み時には、図 2(a) に示すように、スペクトラムを取得するごとに、全ストレージ台数 S_{all} からランダムに 1 台のストレージを選択して書き込みを行う。すなわち、あらかじめ全ストレージへランダムにスペクトラムを振り分けて保存することで、ランダムサンプリングした状態を作っておく。読み出し時には、図 2(b) に示すように、全ストレージ台数 S_{all} の中から S 台のストレージを選び、スペクトラムを読み出してマージする。このとき、抽出率は S/S_{all} となる。すなわち、アクセスするストレージ台数 S により抽出率を変化させることができる。所望の抽出率を α とした場合に、アクセスするストレージ台数は次式となる。

$$S = \lceil \alpha \cdot S_{all} \rceil$$

$\lceil x \rceil$ は実数 x 以上の整数のうち、最小の整数を表わす天井関数である。

4. 可視化サービスへの適用

分散サンプリングストレージを可視化サービスに適用するにあたっては、抽出率 α の決定方法が鍵となる。抽出率 α を低く設定すれば、サンプルサイズが減少し、可視化の速度は向上する。しかしながら、抽出率 α を下げすぎると、可視化結果に誤差を生じる。そのため、可視化の精度に基づく抽出率決定方法が求められる。抽出率 α の決定方法は、可視化の方法に依存するため、以下では 2 値化および占有率推定における抽出率決定方法を示す。

4.1 2 値表示への適用

図 3 に 2 値表示の例を示す。2 値表示は、特定の時間・周波数帯域における電波の存在を有り (1, 黒)・無し (0, 白) の 2 値

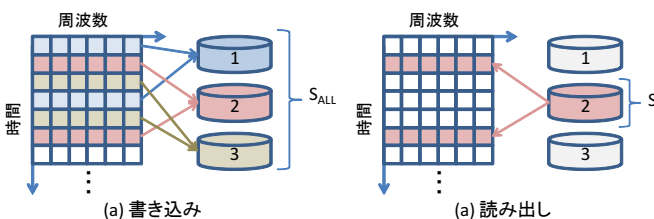


図 2 分散サンプリングストレージの動作原理

で表現したものである。2 値表示は、ユーザの直感的な理解を助けることができ、コグニティブ無線向けにホワイトスペースの可能性のある帯域を表示するために用いられる [19], [20]。また、帯域占有率といった統計量を計算する際の事前処理にも用いられる。

スペクトラムを 2 値化する方法の概要を図 4 に示す。2 値化は、スペクトラムの時間的・周波数的な平均値をとり、閾値判定することで実現される。平均化は熱雑音によるスペクトラムの変動の影響を低減し、誤判定を防ぐために行われる。しかしながら、完全に誤判定を防ぐことはできず、偽陰性と偽陽性の 2 つの誤りが生じる。

抽出率 α は偽陰性率と偽陽性率を保証するように設定する。ホワイトスペース発見といった用途では、保証すべき偽陰性率と偽陽性率が規定されているためである [23]。まず、1 回の判定に必要なサンプルサイズ N を求める。閾値判定における、偽陰性率 P_{FN} と偽陽性率 P_{FP} は次式で表わされる [21]。

$$P_{FN} = 1 - Q\left(\frac{\gamma - F}{\sqrt{2/N}}\right) \quad (1)$$

$$P_{FP} = Q\left(\frac{\gamma - (F+P)}{\sqrt{2/N}}\right) \quad (2)$$

ここで、閾値 γ 、測定機器の持つノイズフロア F 、標準正規分布の裾確率 $Q(x)$ である。 P は検出したい電波の強度 P であり、 P 以上の電力を持つ電波の検出が偽陽性率と偽陰性率の保証の対象であることを意味する。式 (1) および式 (2) を個別に N について解くと、 P_{FN} を保証するサンプルサイズ N_1 、 P_{FP} を保証するサンプルサイズ N_2 が次式のように求まる。

$$N_1 = 2 \left\{ Q^{-1}(P_{FP}) \frac{F}{\gamma - F} \right\}^2$$

$$N_2 = 2 \left\{ Q^{-1}(1 - P_{FN}) \frac{F}{\gamma - (F + P)} \right\}^2$$

$Q(x)^{-1}$ は標準正規分布における裾確率の逆関数である。 P_{FN} および P_{FP} の両方を満足するには、 N_1 と N_2 のうち、より大きい値をサンプルサイズ N として採用する。

$$N = \max(N_1, N_2)$$

なお、ユーザが閾値 γ を指定しない場合には、式 (1) を閾値 γ について解くことで、閾値 γ を決定する。このとき、 N_1 と N_2 が等しくなり、 N_1 と N_2 の比較が不要となる。 N および γ は次式で得られる [21]。

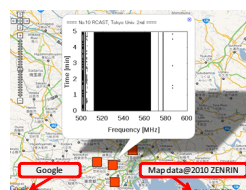


図 3 2 値プロットの例

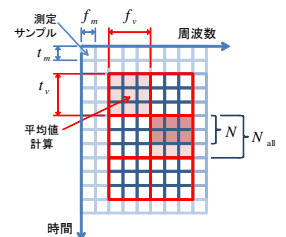


図 4 平均化の領域

$$N = 2 \left[\{Q^{-1}(P_{FP}) - Q^{-1}(1 - P_{FN})\} \frac{F}{P} - Q^{-1}(1 - P_{FN}) \right]^2$$

$$\gamma = F \left\{ 1 + \sqrt{\frac{2}{N} Q^{-1}(P_{FP})} \right\} \quad (3)$$

次に、スペクトラムを平均化する領域に含まれる全サンプル N_{all} を考える。図4にスペクトラムを構成する測定サンプルと、平均値をとる領域の関係を示す。測定の時間分解能を t_m 、周波数分解能を f_m 、平均電力をとる時間幅を t_v 、周波数幅を f_v とすると、 N_{all} は次式となる。

$$N_{\text{all}} = [t_v/t_m] \cdot [f_v/f_m] \quad (4)$$

抽出率 α は N と N_{all} の比で、上限が1となるため、次式で計算される。

$$\alpha = \min(N/N_{\text{all}}, 1) \quad (5)$$

4.2 周波数帯域の占有率推定への応用

図5に周波数帯域の占有率を可視化した例を示す。周波数帯域の占有率は、全測定時間に対する電波が存在する時間の割合である。占有率は、無線システムの電波送出の頻度を表し、どの程度周波数帯域が有効に利用されているかを検討する上で、重要な指標の1つである。

図6に占有率推定の概要を示す。占有率を推定するには、一定時間にわたって、2値化したサンプルの平均値をとれば良い。すなわち、 X_i を2値系列の i 番目のサンプル、 M を2値系列のサンプルサイズとすると、占有率の推定値 U' は次式となる。

$$U' = \frac{\sum_{i=0}^M X_i}{M}, \quad X_i = \begin{cases} 1 & (\text{電波有り}) \\ 0 & (\text{電波無し}) \end{cases}$$

4.1で述べたように、2値系列の各サンプル X_i は、 N 個のスペクトラムのサンプルの平均値を閾値と比較して得たものである。 M 個の2値系列のサンプルを用いた場合、推定全体に必要なサンプルサイズは NM である。一方で、占有率推定を行う時間幅を T 、周波数帯域幅を W 、測定の時間分解能 t_m 、測定の周波数分解能 f_m とすると、占有率推定を行う領域に含まれる全サンプルサイズは $(T/t_m \cdot W/f_m)$ となる。抽出率は、全サンプルサイズと必要なサンプルサイズの比で表わされ、上限1を持つことから、次式で得られる。

$$\alpha = \max\left(1, \frac{NM}{(T/t_m \cdot W/f_m)}\right)$$

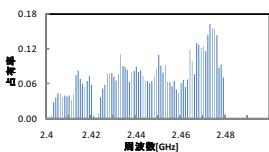


図5 占有率プロットの例

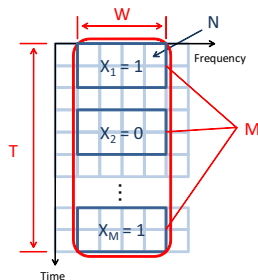


図6 2値系列による利用率計算

具体的な N と M の値は占有率推定の精度を保証するように決定する。まず、 N を2値化の偽陰性率と偽陽性率が占有率推定に与える誤差を保証できるように決定する。付録で示すように、偽陰性率と偽陽性率が占有率推定に与える誤差の期待値は、最大で偽陰性率と偽陽性率と等しくなる。誤差の期待値が ΔU 以下であることを保証するには、偽陰性率 P_{FN} および偽陽性率 P_{FP} が次式を満たす必要がある。

$$\max(P_{FN}, P_{FP}) \leq \Delta U$$

前式を満たすサンプルサイズ N は4.1で述べた手順で求めることができる。

次に、2値系列のサンプルサイズ M を占有率推定の信頼区間と信頼度を保証するように決定する。占有率の推定値 U' は、 M が十分大きい (例えば20以上 [22]) と仮定すると、中心極限定理によって正規分布 $\mathcal{N}(E[X], \text{Var}[X]/M)$ に従う [22]。ここで、 $E[X]$ は X の平均値、 $\text{Var}[X]$ は X の分散で、付録で示すように次式となる。

$$E[X] = U(1 - P_{FN}) + (1 - U)P_{FP}$$

$$\text{Var}[X] = E[X](1 - E[X])$$

正規分布 $\mathcal{N}(E[X], \text{Var}[X]/M)$ において、信頼区間の幅を 2β 、信頼度 Γ とすると、以下の関係が成り立つ [22]。

$$P(E[X] - \beta < U' \leq E[X] + \beta) = \Gamma$$

$$2\Phi\left(\frac{\beta}{\sqrt{\text{Var}[X]/M}}\right) = \Gamma$$

$\Phi(x)$ は標準正規分布の累積分布関数である。前式を M について解けば、信頼区間の幅 2β 、信頼度 Γ を満たすようなサンプルサイズ M は次式となる。

$$M = \left\{ \frac{1}{\beta} \Phi^{-1}\left(\frac{\Gamma}{2}\right) \right\}^2 E[X](1 - E[X])$$

$\Phi^{-1}(x)$ は標準正規分布の累積分布の逆関数である。 M は $E[X]$ によって変動するが、最大値をとる $E[X] = 1/2$ での値を使用すれば、 M は次式で得られる。

$$M = \frac{1}{4} \left\{ \frac{1}{\beta} \Phi^{-1}\left(\frac{\Gamma}{2}\right) \right\}^2$$

5. 評価

5.1 処理時間

評価として、次の3つの模擬環境で、可視化クエリの処理時間を計測した。1つ目は分散サンプリングストレージを模擬した環境である。3台のストレージにあらかじめランダムサンプリングした実データを配置し、アクセスするストレージ台数と可視化計算に利用するサンプルサイズを可視化クエリに応じて調節した。2つ目は分散サンプリングストレージを利用せずにサンプルサイズを調節する環境である。3台のストレージに時系列で3分割した実データを配置し、可視化計算に利用するサンプルサイズをクエリに応じて調節した。3つ目は分散サンプリングを利用せず、サンプルサイズの調節も行わない環境であ

る。3台のストレージに時系列で3分割した実データを配置し、可視化計算には全てのサンプルを使用した。

2値化において、偽陽性率と偽陰性率を変化させたときの処理時間の変化を測定した。実データとしては、スペクトラムアナライザ R&S FSL 6 を用いて、測定時間 24 hour、時間分解能 5 sec、周波数分解能 300 kHz で周波数帯域 100 MHz~1 GHz を測定したものを利用した。可視化クエリのパラメータは、ホワイトスペースの発見 [23] を参考に、検出する電力-107 dBm、周波数範囲 [470 MHz, 770 MHz]、時間範囲 [0:00, 23:59]、2値化の周波数幅 6MHz、2値化の時間幅 120sec とした。また、2値化に用いる閾値は上記の可視化パラメータを基に式 (3) から計算した。

図 7 に偽陽性率・偽陰性率を変化させたときの処理時間の変化を示す。3種類の環境のうち、分散サンプリングストレージを使用した場合が最も低い処理時間を示した。例えば、偽陰性率および偽陽性率が 0.1 のとき、サンプルサイズを調整しない場合と比較すると、処理時間は 65.6%削減された。サンプルサイズの調節を行った場合と比較すると、処理時間は 63.6%削減された。また、分散サンプリングストレージを利用した場合には、処理時間は階段状に変化した。階段状の変化は、サンプルを読み出すストレージ台数 S の変化に対応している。全ストレージ台数 S_{ALL} を増やすことで、よりきめ細かい読み出しストレージ台数の調整が可能となり、さらなる処理時間の削減効果を期待できる。

5.2 精度の保証

まず、偽陰性率と偽陽性率が保証可能であることを検証するために、抽出率に対する偽陰性率と偽陽性率の依存性を実データを用いて評価した。偽陰性率の検証には、電波の存在を想定できる TV 周波数帯域のスペクトラムデータを用いた。21 チャンネル (518 MHz - 524 MHz) 内の 520 MHz - 521 MHz を周波数分解能 100 kHz、時間分解能 50 msec で測定し、平均受信電力は -74.47 dBm であった。2値化に用いる閾値は、-77.5 dBm、-76.5 dBm、-76.0 dBm の 3種類を用いた。偽陽性率の検証には、電波が存在しないと想定できる電波天文帯域のスペクトラムデータを用いた。1400 MHz - 1427 MHz を周波数分解能 100kHz、時間分解能 50msec で測定し、平均受信電力は -106.93dBm であった。2値化に用いる閾値は、-106.5 dBm、-106.0 dBm、-105.0 dBm の 3種類を用いた。

図 8・図 9 に、それぞれ偽陰性率・偽陽性率の抽出率依存性を示す。全体の傾向として、抽出率に従って偽陰性率・偽陽性

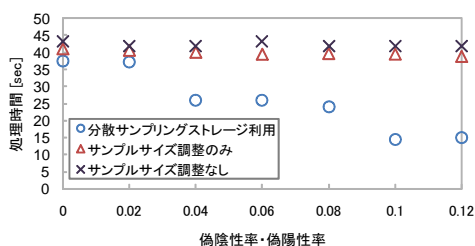


図 7 偽陰性・偽陽性に対する処理時間

率は減少する傾向を示した。式 (1)・式 (2) から求まる理論値と比較すると、実験値は理論値と同じか、理論値を下回った。理論値を下回る理由は、測定機器の特性により、ばらつきの少ないスペクトラムデータが取得できているためと考えられる。ただし、抽出率が小さい場合に、偽陰性率の実験値が理論値を上回るケースも確認された。抽出率が低く、サンプルサイズが少ないため、中心極限定理が成り立たないためと考えられる。しかしながら、実験値が理論値を上回るのは偽陰性率が 0.4 以上の領域に限られる。このような高い偽陰性率を用いることは稀であるため、通常の利用では問題無いと考えられる。

次に占有率推定における、信頼区間と信頼度が保証できることを検証した。実データとしては 2.4GHz 帯における無線 LAN チャンネルを測定したスペクトラムデータを利用した。全サンプルを用いて求まる占有率を真の占有率とみなしたとき、全サンプルの一部をサンプリングして得られる占有率の分布を図 10 に示す。占有率の分布は正規分布 $\mathcal{N}(E[X], \text{Var}[X]/M)$ に従っていることが確認できる。信頼度 95%、99%のときの信頼区間を求めると、それぞれ [0.1175, 0.1418], [0.1133, 0.1458] となった。理論値は、[0.1153, 0.1449], [0.1108, 0.1495] であり、実際の信頼区間は理論値よりも狭くなった。

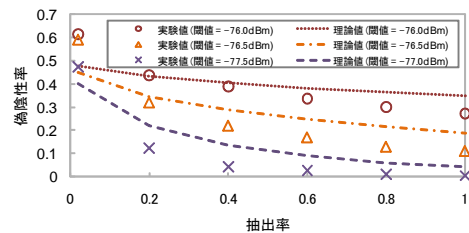


図 8 偽陰性率の抽出率依存性

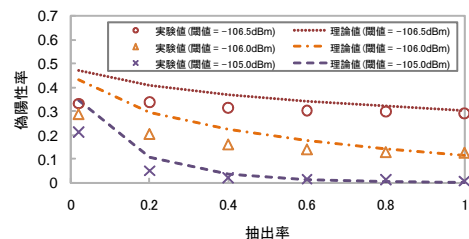


図 9 偽陽性率の抽出率依存性

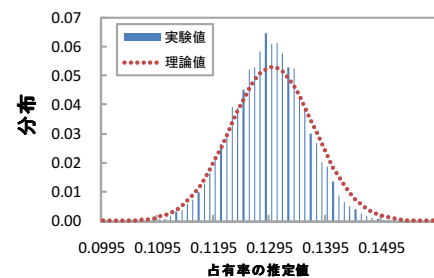


図 10 占有率推定値の分布

6. おわりに

本稿では、電波資源の有効利用に向けて、分散サンプリングストレージを用いたスペクトラムデータの可視化について述べた。分散サンプリングストレージを用いることで、可視化の精度を保証しつつ、処理時間を削減できる。現在、システムの実装・評価を進めている。 文 献

- [1] 長縄潤一, 西村亨輔, 金昊俊, 猿渡俊介, 翁長久,, 森川博之, “広範囲・高密度に展開可能な分散スペクトラムセンシングの設計と実装,” 信学技報, RCS2009-342, Mar. 2010.
- [2] R. Bacchus, T. Taher, K. Zdunek, and D. Roberson, “Spectrum utilization study in support of dynamic spectrum access for public safety,” Proc. IEEE DySPAN 2010, Singapore, Singapore, Apr. 2010.
- [3] D. Chen, S. Yin, Q. Zhang, M. Liu, and S. Li, “Mining spectrum usage data: a large-scale spectrum measurement study,” Proc. ACM MobiCom 2009, Beijing, China, Sept. 2009.
- [4] M. Wellens, J. Riihijarvi, and P. Mahonen, “Empirical time and frequency domain models of spectrum use,” Physical Communication, vol.2, issues 1-2, pp.10-32, Elsevier, Mar.-Jun. 2009.
- [5] 佐藤浩史, 井上武, 山崎敬広, 片山陽平, 高橋紀之, “センサデータマイニングのための分散サンプリングストレージの提案,” 信学技法, IN2009-183, 2009.
- [6] S. Haykin, “Cognitive radio: brain-empowered wireless communications,” IEEE Journal on Selected Areas in Communications, vol.23, issue 2, pp.201-220, Feb. 2005.
- [7] M. Lopez-Benitez, F. Casadevall, and C. Martella, “Performance of spectrum sensing for cognitive radio based on field measurements of various radio technologies,” Proc. 2010 EW, Lucca, Italy, Apr. 2010.
- [8] M. Wellens, A. de Baynast, and P. Mahonen, “Performance of dynamic spectrum access based on spectrum occupancy statistics”, IET Communications, vol.2, issue 6, pp.772-782, Jul. 2008.
- [9] 総務省, “電波利用状況の実地調査結果について,” http://www.soumu.go.jp/main_content/000068577.pdf, May 2010.
- [10] RWTH Aachen University Static Spectrum Occupancy Measurement Campaign, <http://download.mobnets.rwth-aachen.de/index.php>
- [11] Y. Li, T.T. Quang, Y. Kawahara, T. Asami, and M. Kusunoki, “Building a spectrum map for future cognitive radio technology,” Proc. ACM CoRoNet '09, Beijing, China, 2009.
- [12] Microsoft Research WhiteSpaceFinder, <http://whitespaces.msresearch.us/WSWebGUI/whitespaces.aspx>
- [13] U.S. Interactive TV White Space Map, <http://www.spectrumbridge.com/products-services/whitespaces/interactive-map.aspx>
- [14] D. Karger, E. Lehman, T. Leighton, M. Levine, D. Lewin, and R. Panigrahy, “Consistent hashing and random trees: distributed caching protocols for relieving hot spots on the World Wide Web,” Proc. ACM STOC '97, El Paso, TX, USA, May 1997.
- [15] F. Olken, D. Rotem, and P. Xu, “Random sampling from hash files,” Proc. ACM SIGMOD'90, pp.375-386, Atlantic City, NJ, USA, May 1989.
- [16] F. Olken and D. Rotem, “Random sampling from B+ trees,” Proc. VLDB '89, pp.269-277, Amsterdam, The Netherlands, Jul. 1989.
- [17] C. Jermaine, A. Pol, and S. Arumugam, “Online maintenance of very large random samples,” Proc. ACM SIGMOD

'04, pp.299-310, Paris, France, Jun. 2004.

- [18] B. Babcock, S. Chaudhuri, and G. Das, “Dynamic sample selection for approximate query processing,” Proc. ACM SIGMOD '03, pp.539-550, San Diego, California, Jun. 2003.
- [19] S. Pagadarail, A.M. Wyglinski, and R. Vuyyurul, “Characterization of vacant UHF TV channels for vehicular dynamic spectrum access,” Proc. IEEE VNC 2009, Tokyo, Japan, Oct. 2009.
- [20] M.H. Islam, C.L. Koh, S.W. Oh, Q. Xianming, Y.Y. Lai, W. Cavin, L. Ying-Chang, B.E. Toh, F. Chin, G.L. Tan, and W. Toh, “Spectrum survey in singapore: occupancy measurements and analyses,” Proc. CrownCom 2008, Singapore, Singapore, May 2008.
- [21] D.B. Cabric, “Cognitive Radios: System Design Perspective,” UCB/EECS-2007-156, 2007.
- [22] E. クライツィグ, 確率と統計, 技術者のための高等数学, 培風館, 2004.
- [23] FCC, Tech. Rep. 08-260, Nov. 2008.
- [24] J.C. ミラー, 統計学の基礎, 培風館, 1998.

付 録

2 値化の誤差が占有率推定に与える影響

2 値化の誤差が占有率推定に与える影響を考察するために、2 値化の誤差を含まない 2 値系列 Y_i を導入する。 Y_i は誤差を含まないため、真の占有率 U とすると、次式が成り立つ。

$$P(Y_i = 0) = 1 - U$$

$$P(Y_i = 1) = U$$

偽陽性率 P_{FP} および偽陰性率 P_{FN} とすれば、 X_i と Y_i の関係について、次式が成り立つ。

$$P(X_i = 1|Y_i = 0) = P_{FP}$$

$$P(X_i = 1|Y_i = 1) = 1 - P_{FN}$$

前式を用いると、

$$\begin{aligned} P(X_i = 1) &= P(X_i = 1|Y_i = 1)P(Y_i = 1) + \\ &P(X_i = 1|Y_i = 0)P(Y_i = 0) \\ &= (1 - P_{FN})U + P_{FP}(1 - U) \end{aligned}$$

$P(X_i = 1)$ は $E[U']$ と等しくなることが知られている [24].

$$\begin{aligned} E[U'] &= P(X_i = 1) \\ &= (1 - P_{FN})U + P_{FP}(1 - U) \end{aligned}$$

前式を用いて、誤差の期待値 $E[U' - U]$ を求めると次式となる。

$$\begin{aligned} E[U' - U] &= E[U'] - U \\ &= P_{FP} - U(P_{FN} + P_{FP}) \end{aligned}$$

誤差の期待値 $E[U' - U]$ は真の占有率 U に依存するが、 U に対して極値を持たないため、 U の持つ区間 $[0, 1]$ の両端で最大値をとる。

$$E[U' - U] = \begin{cases} -P_{FP} & (U = 0) \\ P_{FN} & (U = 1) \end{cases}$$

平均 $E[X]$ と分散 $V[X]$

$$\begin{aligned} E[X] &= P(X_i = 0) \cdot 0 + P(X_i = 1) \cdot 1 \\ &= (1 - P_{FN})U + P_{FP}(1 - U) \end{aligned}$$

$$\begin{aligned} \text{Var}[X] &= E[X^2] - E[X]^2 \\ &= P(X_i = 0) \cdot 0^2 + P(X_i = 1) \cdot 1^2 - E[X]^2 \\ &= E[X](1 - E[X]) \end{aligned}$$