Vol.2019-MPS-123 No.40
Vol.2019-BIO-58 No.40
2019/6/19

変化点検出に基づく可変ビン幅ヒストグラムの構築

伏見 卓恭[1]　岩﨑 清斗[2]　大久保 誠也[3]　斉藤 和巳[4,5,3]

概要：本研究の目的は，与えられた数値データの頻度情報について，データ濃度が密な部分をより詳細に分析・可視化する手法の構築である．そこで，与えられた数値データより，K 個の可変ビン幅で描かれるヒストグラムを構築する手法を提案する．提案手法は，データの分布に粗密の偏りがあるとき，分布が疎な部分では広く，密でな部分では狭くするように，ビン幅を適切に自動調整する．具体的には，要素を昇順にソートしたデータ集合を時系列データと見なし，L2 誤差または L1 誤差に基づく変化点検出法を適用して変化点集合を検出し，得られた変化点集合より可変ビン幅のヒストグラムを構築する．検証実験では、静岡県内の 4 つのバラ農家から収集した飽差データに対し提案手法を適用することにより、標準的に採用される平方根選択法やスタージュの公式でビン数を決める等幅ビンのヒストグラムと比較して，データの要素値分布の粗密に応じて適切なビン幅となるヒストグラムの構築が可能なことを示す．また，エントロピーに基づく定量評価では，L2 誤差と比較して，L1 誤差で構築するヒストグラムが望ましい性質を持つことも示す。

キーワード：ヒストグラム，変化点検出，可変ビン幅，可視化，農業データ

Construction of Histogram with Variable Bin-width based on Change Point Detection

Takayasu Fushimi[1]　Kiyoto Iwasaki[2]　Seiya Okubo[3]　Kazumi Saito[4,5,3]

Abstract: In this paper, we address a problem of constructing a histogram drawn by K bins with variable widths from a set of numerical values, so as to have relatively large numbers of narrow bins for some ranges where values distribute densely and change substantially, while small numbers of wide bins for the other ranges. For this purpose, we propose a new method, i.e., after arranging a given set of values in ascending order, regarding them as a time-series dataset, and applying a change point detection method to this dataset based on an $L1$ or $L2$ error criterion, we produce a step function consisting of K steps, and then by using these change point information, we construct a histogram drawn by K bins with variable widths. In our experiments using four datasets of humidity deficit (HD) collected from vinyl greenhouses owned by four rose farmers by setting our original IoT devices, we show that our proposed method can construct more natural histograms with appropriate variable bin widths than those with an equal bin width constructed by the standard method based on square-root choice or Sturges' formula. In addition, by performing quantitative evaluation based on an entropy function, we also show the histograms constructed with the L1 error criterion has more desirable property than those with the L2 error criterion,

Keywords: periodic environment data, agricultural environment, visualization

[1] 東京工科大学　コンピュータサイエンス学部
School of Computer Science, Tokyo University of Technology
[2] 静岡県工業技術研究所　機械電子科
Industorial Research Institute of Shizuoka Prefecture
[3] 静岡県立大学　経営情報学部
School of Management and Information, University of Shizuoka
[4] 神奈川大学　理学部
Faculty of Science, Kanagawa University
[5] 理化学研究所　革新知能統合研究センター
Center for Advanced Intelligence Project, RIKEN

Vol.2019-MPS-123 No.40
Vol.2019-BIO-58 No.40
2019/6/19

1. はじめに

近年，IoT（Internet of Things）技術が急速に発展している．それにともない，農業・医療・教育などの専門知識と熟練が不可欠な分野においても，専門家や熟練者がどのような状況でどのような活動をしているかに関する情報を，容易に入手可能となってきている．一方，特に農業分野では担い手の高齢化による労働力不足が深刻化しており，作業の合理化や技術の継承が課題となっている．

農業において，環境は非常に重要な要素である．熟練農家は，様々な手段を用いて環境を制御することにより，より多くの収穫や高い品質の作物を得ている．しかし，熟練農家の持つ技術が，生産性や環境にどのような影響をどのような形で与えているのかは，明らかになっていない部分が多い．明らかにするためには，まず，環境データに偏りや分布があるかを調べ，その後，主要な偏りや集合について，生じる理由を明らかにする必要がある．つまり，データを適切な集合に分割し，その頻度を求める必要がある．

分布の傾向を分析するために，ヒストグラムによる可視化が広く使われてきた．一般的に，ヒストグラムのビン幅は固定長である．このビンを適切に設定することは重要であり，さまざまな手法や指標が提案されている [5]．本研究で扱う農業環境データは，望ましい値を取るように環境を制御する関係で，集中したものとなる場合が多い．加えて，集中した部分に農業的に重要な集合が複数含まれている可能性がある．このようなデータに対して固定長のヒストグラムによる分析を行った場合，一つのビン幅内に重要な情報が埋もれてしまい，分析を達成することは難しい．

本研究の目的は，データの偏りを考慮して頻度分布を可視化することで，重要な情報が埋もれてしまわないようにする手法の構築である．そこで，データの偏りを考慮してビン幅を調整する可変ビン幅ヒストグラムを提案する．提案手法は，クラスタリングとヒストグラムを組み合わせたものと考えることができ，偏りがあるデータに対して，密な部分は詳しく，素な部分は粗く可視化することができる．また，評価実験により，提案手法の有効性を検証する．具体的には，静岡県内にある 4 件のバラ農家から収集した環境データに提案手法を適用し，その妥当性を評価する．

本稿は以下に示す構成である．第 2 節で提案手法に関連する既存研究について整理する．第 3 節で提案手法の定義と計算アルゴリズムについて解説する．第 4 節で実データを用いた評価実験およびその結果について議論する．最後に本稿のまとめと今後の課題について述べる．

2. 関連研究

本研究では，ビン幅を求める際，変化点検出のアイデアを用いる．そこで，本節では，変化点検出の手法，ならびに農業データ解析の関連研究について述べる．

本研究では，農業環境データの変化を検出するため，データ生成の背後に内在する基本メカニズムの変化検出をレジーム切換（regime switching）問題 (e.g., [2], [3]) として定式化し，その変化点検出のために技術 [3] を利用する．この問題設定は，従来から使用されてきた，定常モデルと比較して統計的に有意な短期的外れ値を求める異常検出（anomaly detection）や，確率分布の混合モデルとして定常モデルを設定する統計的機械学習の枠組み [1] とは，一線を画すものである．従来の異常検出に使用される統計的手法は，与えられたデータに対して統計モデル（インスタンスの大多数は正常であるという仮定）を適合させ，統計的検定によって未知のインスタンスがこのモデルに属するか否かを決定する．このような手法では，適用された統計的検定に基づき，学習モデルから生成される確率が低いインスタンスは異常とされる．一方，本研究では，モデルやメカニズムの変化を検出するため，データ背後に内在する規則性や知識の抽出と親和性が高い．特に，時間で変化するモデルパラメータをレジームスイッチングとして扱うため，従来の典型的異常検出技術とは方向性が異なっている．

各種農業データを取得し，そこから環境制御技術について評価する研究も進められている [6], [8]．たとえば，代表日について，環境変化の波形や頻度のヒストグラムを描き，それぞれに理由を付与することなどが行われている．一方で，偏りのあるデータを自動で分割・可視化して，その原因についての検討することは行われていない．

3. 提案手法

提案手法は，数値データ集合 $\mathcal{X} = \{x_t \mid t = 1, \cdots, T\}$ が与えられたときに，可変ビン幅のヒストグラムを構築する．

通常のヒストグラムでは，横軸は階級の範囲，縦軸は各階級に属するデータの要素数が記述される．たとえば，“値が 0〜10 のデータは 20 個”，“値が 11〜20 のデータは 30 個” のようになる．一方，本研究では，ヒストグラムを横軸が s，縦軸が $h(s)$ の階段グラフとして記述する．先の例だと，$0 \le s \le 10$ ならば $h(s) = 20$，$11 \le s \le 20$ ならば $h(s) = 30$ となる．

この記法に基づくと，ビン数が K の標準的な等幅ビンのヒストグラムは次のように構築される．まず，データ集合 $\mathcal{X}$ の要素を昇順にソートし，データ集合の要素のインデクスを設定する．以後，簡単のため，各 $t(< T)$ に対して $x_t \le x_{t+1}$ を満たすとする．次に，ビン幅を $\delta = (x_T - x_1)/K$ に設定する．ビン幅 δ となる K 個のデータ集合に分割するため，各ビン間の区切りを $F(k) = x_1 + k\delta$（ただし $k \in \{1, \cdots, K-1\}$）として求める．また，$F(0)$ は x_1 未満のある定数 x_0，$F(K)$ は x_T とする．このとき，第 $k(\le K)$ ビンに対応するデータ集合は $\mathcal{X}_k = \{x_t \mid F(k-1) < x_t \le F(k)\}$ となるため，最終的にヒストグラム $h(s)$ は次式として得られる．

$$h(s) = |\mathcal{X}_k| \quad (k = \lceil (s - x_1)/\delta \rceil, \quad s \in [x_0, x_T]). \qquad (1)$$

以降，この等幅ビンのヒストグラム構築法を NM と呼ぶ．データ集合 $\mathcal{X}$ の要素値分布に粗密の偏りがあるとき，等幅ビンのヒストグラムでは分析に限界が起こり得るため，要素値分布が疎と密の部分で，ビン幅は前者では大きく，後者では小さくする手法が望まれる．

提案手法は，要素値分布の疎と密に応じてビン幅を自動的に調整する．提案手法のアイデアを図 1 に示す．まず，数値データ集合 $\mathcal{X}$ を小さい順にソートする．結果を図 1(a) に示す．ここで，横軸は順位、縦軸は値である．次に，変化点検出を行う．求められた変化点を用いて階段関数を描くと図 1(b) のようになる．図中赤線の縦に変化している部分が，変化点である．変化点の前後で要素数がどのように変化するかは分析で重要である．一方，変化点間のデータは，他と比較して値が一定であると解釈することができる．つまり，変化点で区切ることにより，主要な K 個の集合に分割することができる．最後に，変化点間にある要素数を数え，図 1(c) のようなヒストグラムを描く．

詳細なアルゴリズムの流れは，次のようになる．

入力 :データ集合 $\mathcal{X}$，ビン数 K．

(1) データ集合 $\mathcal{X}$ の要素を昇順にソートし，データ集合の要素のインデクスを設定する．今後，各 $t(< T)$ で $x_t \le x_{t+1}$ を満たしているとする．

(2) 変化点数を $K-1$ とし，L2 法では $\ell_K^2(\mathcal{G})$，L1 法では $\ell_K^1(\mathcal{G})$ の最小化により $\mathcal{G}$ を求める．

(3) $\mathcal{G}$ から $F(k)$ を求め，式 (1) に基づき．ヒストグラム $h(s)$ を構築する．

出力 :ヒストグラム $h(s)$ をプロットした可視化結果．

ここで，$\mathcal{G}$ は変化点の集合，$\ell_K^2(\mathcal{G})$ と $\ell_K^1(\mathcal{G})$ は変化点を求めるために使用する目的関数，$F(k)$ はビンの区切りの集合である．ステップ 2 が，ビン幅を適切に設定する処理である．以下に，ステップ 2 の詳細について述べる．

変化点検出問題では，与えられた変化点数 $K-1$ で，データ集合 $\mathcal{X}$ との誤差が最小となるような階段関数を求める．誤差を定義するノルムとして，L2 と L1 のそれぞれに基づく手法が考えられる．本稿では，これらを L2 変化点検出法（L2 法）と L1 変化点検出法（L1 法）と呼ぶ．

いま，時系列データ $\mathcal{X}$ には求められた変化点が存在しないとする．一つの値で近似するならば，L2 法では，次式で示すように，平均値 $\mu(1,T)$ により L2 誤差 ℓ_0^2 が最小化される．

$$\ell_0^2 = \sum_{t=1}^{T} (x_t - \mu(1,T))^2$$

ここで，平均値 $\mu(a,b)$ は次式である．

$$\mu(a,b) = \frac{1}{b-a+1} \sum_{t=a}^{b} x_t.$$

一方，L1 法では，次式で示すように，中央値 $\nu(1,T)$ により L1 誤差 ℓ_0^1 が最小化される．

$$\ell_0^1 = \sum_{t=1}^{T} |x_t - \nu(1,T)|$$

ここで，中央値 $\nu(a,b)$ は次式である．

$$\nu(a,b) = \begin{cases} x_{(a+b)/2} & \text{if } a+b \text{ is even,} \\ (x_{(a+b)/2} + x_{(a+b)/2+1})/2. & \text{otherwise.} \end{cases}$$

ここで，$\nu(a,b)$ の計算には，データ集合 $\mathcal{X}$ の要素がソートされていることを利用している．

次に，データ番号 τ が唯一の変化点であるとし，その前後をそれぞれ一つの値で近似するならば，次式で示すように，前後それぞれの区間での平均値と中央値により，L2 誤差 $\ell_1^2(\tau)$ と L1 誤差 $\ell_1^1(\tau)$ が最小化される．

$$\ell_1^2(\tau) = \sum_{t=1}^{\tau} (x_t - \mu(1,\tau))^2 + \sum_{t=\tau+1}^{T} (x_t - \mu(\tau+1,T))^2$$

$$\ell_1^1(\tau) = \sum_{t=1}^{\tau} |x_t - \nu(1,\tau)| + \sum_{t=\tau+1}^{T} |x_t - \nu(\tau+1,T)|.$$

ここで，望ましい変化点 τ は，L2 誤差 $\ell_1^2(\tau)$ や L1 誤差 $\ell_1^1(\tau)$ を最小化する τ になる．

等幅ビンのヒストグラム構築法でデータ分割に用いた配列 $F(k)$ と同様なアイデアで，このような誤差関数を一般化する．変化点の個数は $K-1$ 個であり，変化点に対応する x_i のインデクス i が小さい順に $G(1)$ から $G(K-1)$ に格納されているとする．また配列 $F(k)$ と同様に，便宜上 $G(0) = 0$ かつ $G(K) = T$ と設定し，$(K+1)$ 個の要素からなる集合 $\mathcal{G} = \{G(0), \cdots, G(K)\}$ を考える．すると，L2 誤差 $\ell_K^2(\mathcal{G})$ と L1 誤差 $\ell_K^1(\mathcal{G})$ は次式で求められる．

$$\ell_{K-1}^2(\mathcal{G}) = \sum_{k=1}^{K} \sum_{t=a(k)}^{b(k)} (x_t - \mu(a(k), b(k)))^2$$

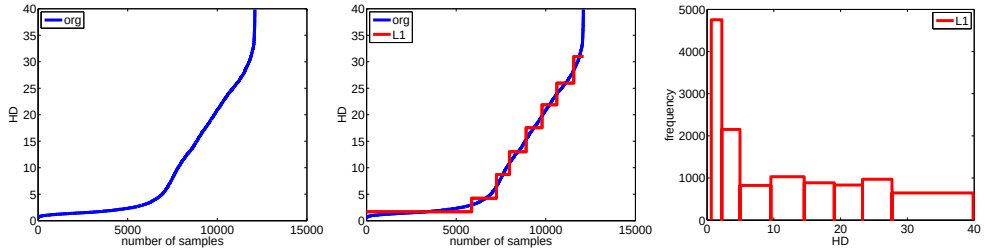
$$\ell_{K-1}^1(\mathcal{G}) = \sum_{k=1}^{K} \sum_{t=a(k)}^{b(k)} |x_t - \nu(a(k), b(k))|.$$

ここで，$a(k) = G(k-1)+1, b(k) = G(k)$ である．すなわち，提案手法における変化点検出問題は，$\ell_K(\mathcal{G})$ を最小化する $\mathcal{G}$ を求める問題として定式化できる．文献 [3] と同様の，逐次的に変化点を求める方法を採用する．

変化点の集合 $\mathcal{G}$ が求まれば，$k \in \{1, \cdots, K\}$ に対し $F(k) = x_{G(k)}$ とし，$F(0)$ を x_1 未満の定数とすれば， 次式によりヒストグラム $h(s)$ を構築することができる．

$$h(s) = |\mathcal{X}_k| \quad (F(k-1) < s \le F(k),\ s \in [x_0, x_T]).$$

提案手法を効果的に用いるには，ビン数 K を適切に設定する必要がある．K の値は，一般に，概要の分析では小さな値に、詳細な分析では比較的大きな値に設定すること



(a) (b) (c)

図 1 アルゴリズムのアイデア

になる．

適用例を図 1 を用いて説明する．L1 法を用いて求めた変化点の集合 $\mathcal{G}$ に対応する $x_{G(k)}$ は，{2.21049, 4.933572, 9.562568, 14.54854, 19.04688, 23.27846, 27.69275} となった（図 1(b) の赤線の縦棒部分）．これらの値に x_1 未満の値 0.631704 と x_T の値 39.83653 を加えたものが，ビンの区切り $F(k)$ となる．各ビンの要素数は {4754, 2152, 824, 1031, 887, 833, 970, 645} であったため，ヒストグラム $h(s)$ が求まる（図 1(c) 参照）．ヒストグラムにおいて，横軸のビン幅が各階級の範囲を，縦軸が各階級の要素数を表す．たとえば，図 1(c) の一番右の棒（枠）は “27.69275 より大きく 39.83653 以下の要素数は 645 である” ことを表している．

4. 評価実験

4.1 評価実験概要

評価実験により，提案手法の妥当性を評価する．具体的には， 静岡県内にある 4 件のバラ農家のビニールハウス（ハウス A〜D）から得られた環境データに対して，提案手法を用いた解析を行うことにより，有効性を評価する．用いたデータは，[4] で用いたものと同様であり，2018 年 3 月 27 日 0 時 0 分から 2018 年 5 月 7 日 24 時 00 分までのデータである．1 日あたり 288 次元ベクトルで，それが 42 日分であるため，各農家に対して 12096 個のデータとなる．

本稿では，飽差を対象とした解析を行う．飽差とは，ある温度と湿度の空気に，あとどれだけ水蒸気の入る余地があるかを示す指標であり，農産物の成育に，ある一定範囲での飽差制御が重要であるとされている [7]．小型デバイスは飽差 HD を直接測定するセンサーを備えていないため，以下の式により求めた．

$$\mathrm{HD} = (100 - H) * \frac{217 * \frac{6.1078 * 10^{\frac{7.5 * C}{C + 237.3}}}{C + 273.15}}{100}$$

ここで，H は湿度，C は気温である．一般には，飽差の値は 3 ∼ 6g/m^3 がよいとされている [7]．

4.2 従来法での可視化結果

比較のため，ビン数 $K = 16, 32, 64, 128$ のときの等幅ビンによるヒストグラム構築結果を図 2 に示す．データ数 $T = 12,096$ に対し，代表的なビン数決定方法のスタージェスの公式では $\lceil \log_2 T + 1 \rceil \approx 15$，平方根選択法では $\sqrt{T} \approx 100$ となる．図 2(a) と (d) がこれらの結果に概ね符合し，図 2(b) と (c) の結果はこれらの中間的なものとなる．これらの結果より，比較的ビン数の少ない図 2(a) と (b) のケースと，比較的多い図 2(c) と (d) のケースを比較すれば，要素値の分布が疎な $HD > 10$ 部分では，前者のケースと比較して，後者のケースの冗長度は高く，逆に，要素値の分布が密な $HD < 10$ 部分では，後者のケースと比較して，前者のケースの解像度は低くいことが分かる．

4.3 提案法での可視化結果

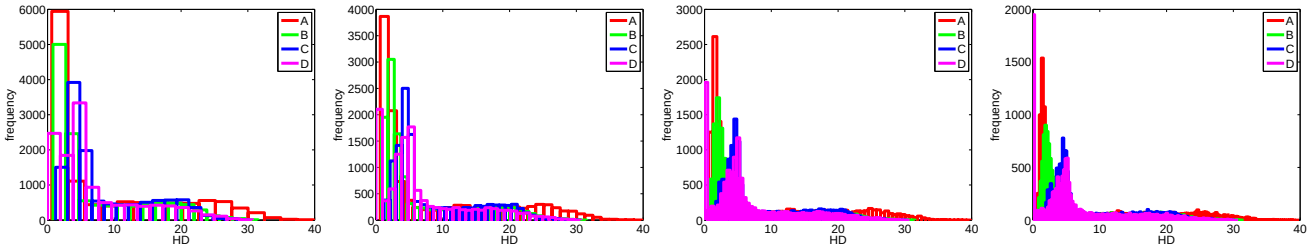
各ハウスのデータに対し，$K = 8$ として L2 法と L1 法を適用し，得られた変化点検出結果を図 3 に示す．結果より，L2 法でも L1 法でも，妥当な精度で階段関数近似できてることが確認できる．

各ハウスのデータに対し，ビン（変化点）数 $K = 4, 8, 16, 32$ として L2 法と L1 法で構築した可変ビン幅ヒストグラムを図 4 から 7 に示す．また，比較のため NM で構築した等ビン幅ヒストグラムも併せて示す．ここで，各図の (c) はスタージェスの公式に基づくビン数と概ね符合する．

データ集合 $\mathcal{X}$ の要素値分布に粗密の偏りがあるとき，要素値の分布が疎な部分では大きく，密な部分では小さくするように，ビン幅を適切に自動調整することが提案法の目的である．目的が実現できているかを定量評価するため，構築したヒストグラム $h(\cdot)$ に対し，次式で示すように，分割されたデータそれぞれの要素数に対する分布のエントロピーを求めた．

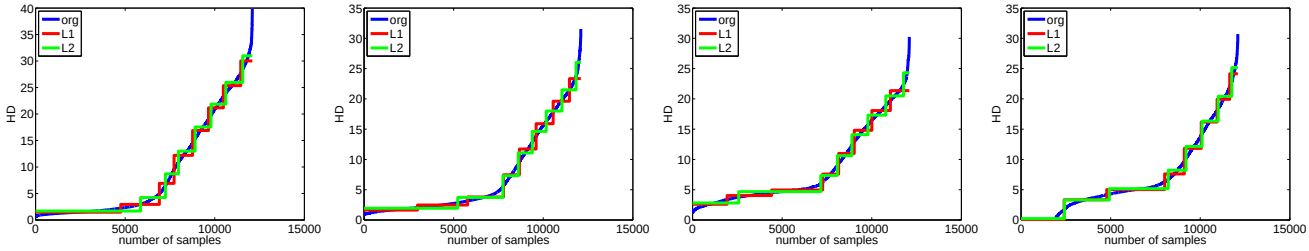
$$\begin{aligned} E(h(\cdot)) &= -\sum_{k=1}^{K} \frac{|\mathcal{X}_k|}{T} \log \frac{|\mathcal{X}_k|}{T} \\ &= -\sum_{k=1}^{K} \frac{h(F(k))}{T} \log \frac{h(F(k))}{T}. \end{aligned} \tag{2}$$

$E(h(\cdot))$ が大きいほど望ましいヒストグラムである．式 2 のエントロピーに基づく定量評価結果を図 8 に示す．それぞれの図において，横軸はビン数，縦軸はエントロピー $E(h(\cdot))$ を示している．L2 法よりも L1 法によって構築するヒストグラムの方が $E(h(\cdot))$ の値が大きいため，望ましい性質を持つことがわかる．



(a) $K = 16$　(b) $K = 32$　(c) $K = 64$　(d) $K = 128$

図 2　等幅ビンのヒストグラムの例



(a) ハウス A　(b) ハウス B　(c) ハウス C　(d) ハウス D

図 3　ソート後のデータに対する L2 法と L1 法での変化点検出結果の比較　($K = 8$)

5. 考察

農業環境データには粗密の偏りがあることと，分析後に個別の原因の解析を行うことを考えると，データの分布を少ないビン数 K で可視化できることが望ましい．

図 2 から，従来手法でも，データの値は 0 から 7 の間の狭い範囲に集中していることがわかる．$K = 16$ の場合，ハウス A とハウス B は同様の低い値にピークがあること，ハウス C とハウス D は，それよりも高い位置にピークがあることがわかる．一方，$K = 128$ の場合，農家 A は 1 カ所に高いピークが立つのに対して，ハウス B はハウス A より分布がなだらかであり，特徴が異なっていることがわかる．これらことから，$K = 16$ では特徴を捉えるのに不十分であり，より大きな K が必要であるといえる．一方，$K = 128$ では非常に細かく区切られており，どの区間に対して原因の解明をすればよいのかが明らかでない．

一方，図 4〜6 から，提案手法は，L1 法と L2 法のどちらでも，値が集中している 0 から 7 の間について，他の部分より細かく分割していることがわかる．$K = 4$ のときでも，ハウス A とハウス B の特徴の違いや，ハウス C とハウス D の類似性がわかる．そして，$K = 16$ では各特徴がより明確となっている．また，値が集中していない 10 以上については，NW 法よりもビン幅は大きくなっているが，NW と同様の分布を示すことができている． このことから，提案手法は K が小さい値でも，重要な特徴を捉えることができており，今後の解析等が行いやすいといえる．

6. おわりに

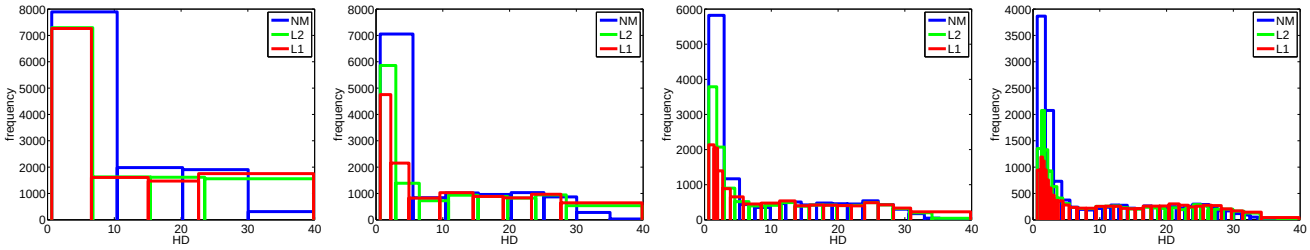
本研究では，偏りがあるデータの頻度分布を可視化するため，可変ビン幅ヒストグラムの構築手法を提案した．また，バラ農家の環境データに適用することにより，等幅ビン長よりも特徴が明らかにしやすいことを示した．今後の課題として，得られた結果から，熟練農家の技術が環境に与える影響を評価することなどがあげられる．

謝辞　本研究は，JSPS 科研費 (C)(No.18K11441) の助成を受けたものである．

参考文献

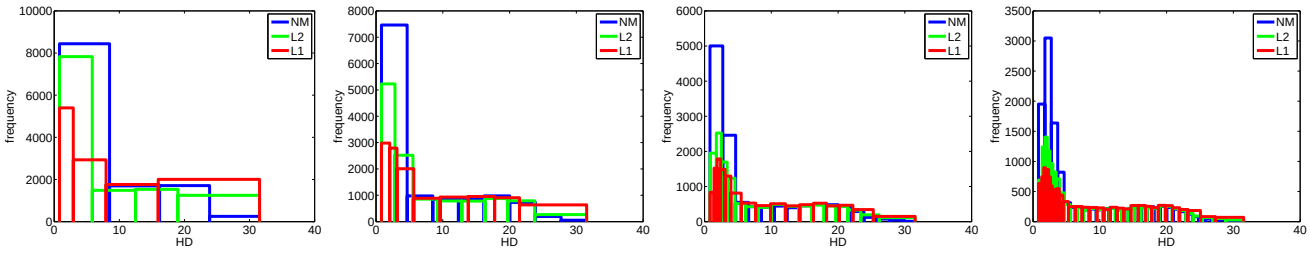
[1] Chandola, V., Banerjee, A. and Kumar, V.: Anomaly Detection: A Survey, *ACM Comput. Surv.*, Vol. 41, No. 3, pp. 15:1–15:58 (online), DOI: 10.1145/1541880.1541882 (2009).
[2] Kim, C.-J., Piger, J. and Startz, R.: Estimation of Markov regime-switching regression models with endogenous switching, *Journal of Econometrics*, Vol. 143, No. 2, pp. 263–273 (2008).
[3] Saito, K., Ohara, K., Kimura, M. and Motoda, H.: Change point detection for burst analysis from an observed information diffusion sequence of tweets, *J. Intell. Inf. Syst.*, Vol. 44, No. 2, pp. 243–269 (online), DOI: 10.1007/s10844-013-0283-2 (2015).
[4] 岩﨑清斗，伏見卓恭，大久保誠也，斉藤和巳：差分累積値に基づく農業環境データの可視化分析法，情報処理学会研究報告. MPS， Vol. 2018-MPS-121, No. 1, pp. 1–6 (2018).
[5] 坂元慶行，石黒真木夫，北川源四郎：情報量統計学 (情報科学講座 A・5・4)，共立出版 (1983).
[6] 山田寛乃，渡辺知恵美：観察記録とヒストグラムの変遷を用いた自然画像からの特徴抽出に向けて，技術報告 16 (2010).
[7] 農文協 (編)：野菜 vol.7: もっと知りたい環境制御技術-日中 CO2 濃度, 飽差, 葉面積を管理する，農山漁村文化協会 (2014).
[8] 農文協 (編)：野菜 vol.8 特集:ここまで見えた環境制御技術，農山漁村文化協会 (2015).

Vol.2019-MPS-123 No.40
Vol.2019-BIO-58 No.40
2019/6/19



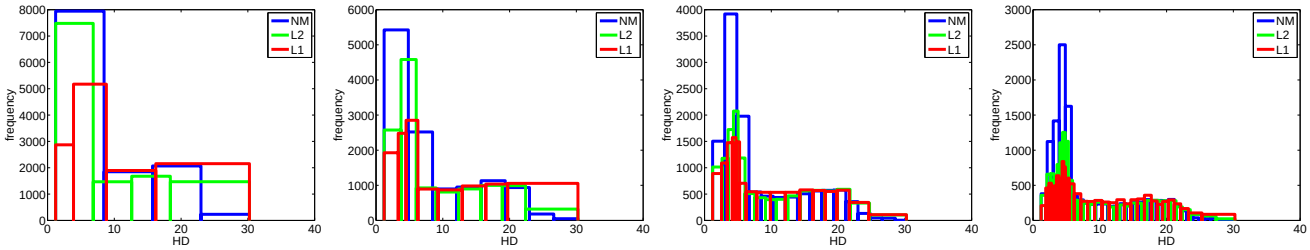
(b) $K = 4$ (c) $K = 8$ (d) $K = 16$ (d) $K = 32$

図 4 NM 法，L2 法，L1 法で構築したヒストグラムの比較評価 (ハウス A)



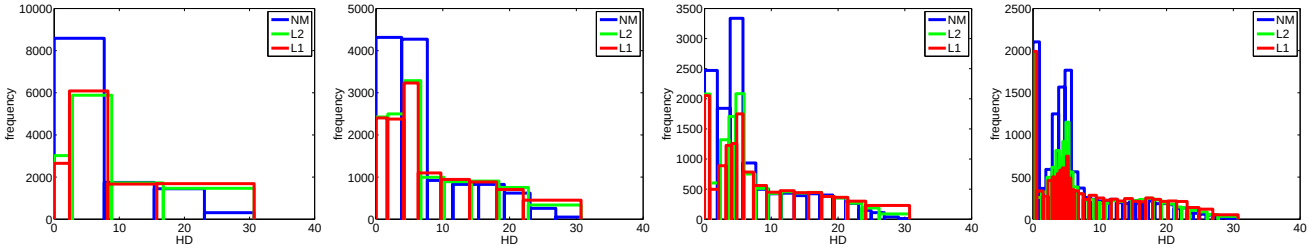
(b) $K = 4$ (c) $K = 8$ (d) $K = 16$ (d) $K = 32$

図 5 NM 法，L2 法，L1 法で構築したヒストグラムの比較評価 (ハウス B)



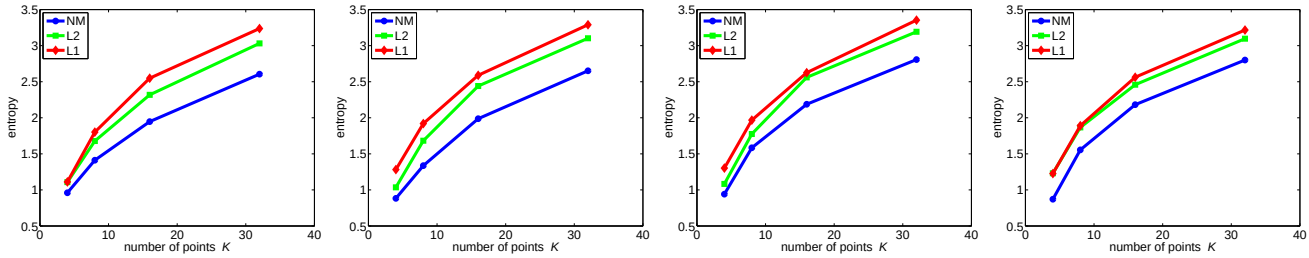
(b) $K = 4$ (c) $K = 8$ (d) $K = 16$ (d) $K = 32$

図 6 NM 法，L2 法，L1 法で構築したヒストグラムの比較評価 (ハウス C)



(b) $K = 4$ (c) $K = 8$ (d) $K = 16$ (d) $K = 32$

図 7 NM 法，L2 法，L1 法で構築したヒストグラムの比較評価 (ハウス D)



(a) ハウス A (b) ハウス B (c) ハウス C (d) ハウス D

図 8 エントロピーに基づく偏り解消度の評価