

多様な大規模センシングデータ間の時系列相関分析 技術の研究

メタデータ	言語: ja 出版者: 静岡大学 公開日: 2013-01-09 キーワード (Ja): キーワード (En): 作成者: 峰野, 博史 メールアドレス: 所属:
URL	http://hdl.handle.net/10297/7024

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年3月31日現在

機関番号：13801

研究種目：挑戦的萌芽研究

研究期間：2010～2011

課題番号：22650012

研究課題名（和文） 多様な大規模センシングデータ間の時系列相関分析技術の研究

研究課題名（英文） Study on time-line correlation analysis method for analyzing the spatial and temporal relation among heterogeneous sensing data

研究代表者

峰野 博史 (MINENO HIROSHI)

静岡大学・情報学部・准教授

研究者番号：40359740

研究成果の概要（和文）：ICT を活用して環境情報を収集するセンサネットワークにおいて、多様な大規模センシングデータ間の空間的、時間的な複雑な相関関係を様々な解像度で分析可能な時系列相関分析技術の開発を目指した。連続するいくつかの時系列データをセグメントとしてまとめ、一定の長さを持つ時系列データをセグメント単位に分割し、多段階のクラスタリングと分類の組合せによって頻出パターンを抽出する方式を開発した。

研究成果の概要（英文）：We aimed to develop a time-line correlation analysis method for analyzing the spatial and temporal relation among heterogeneous sensing data in various resolution. Our developed method is a multiple-stage clustering method that is composed of three stages. The first is a single stream clustering of continuous time-line data as a segment, second is a multi-stream clustering and the last is a multi-stream classification.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	1,500,000	0	1,500,000
2011年度	1,400,000	420,000	1,820,000
年度			
年度			
年度			
総計	2,900,000	420,000	3,320,000

研究分野：総合領域

科研費の分科・細目：情報学・計算機システム・ネットワーク

キーワード：センシングデータ・データ系列・多段階クラスタリング・K-NN・大規模可視化・時空間情報

1. 研究開始当初の背景

ユビキタスネットワーク社会における ICT 利活用の高度化や、多彩なサービスの実現には、対象物の状況・状態やそれらの周辺環境など、様々な状況や環境を自動的に認識して、自律的な情報流通に基づいて最適な制御を実現する技術が必要不可欠である。これまで、申請者も含め多くの企業、大学、研究機関等でシーズ先行型の研究開発が中心に進められてきたが、社会が求める具体的なアプリケ

ーションの創出を意識したセンサネットワーク研究に関する報告も相次いでいる (VolcanoMonitoring, FireWxNet, BikeNet, CowNet, BusNet)。しかし、数値や事実の集まりであるデータから、ある基準などを基にしてデータに意味を持たせた情報として扱えるようにするところまでが大半で、データと情報で構成される文脈の中で専門知識（ノウハウ）による判断の影響まで踏み込んで研究し十分な成果が出ているとは言い難い。ま

た、多数のセンサを等間隔に並べ、その位置も正確に管理されたような均一かつ較正されたセンサ群であればデータ処理としては扱いやすいが、現実問題としてこのようなセンサ群を実際の日常環境中に用意するのは容易ではない。つまり、非均一、非較正、非均質な多種多様なセンシングデータに対して統一的特徴抽出を実現でき、ノウハウによる状況の分類や暗黙知の形式化を促進することで、多種多様な時系列データ間の相関関係の解明をサポートする時系列相関分析技術の開発が必要不可欠であると考える。

センサネットワークを活用することで、これまで意識されなかった様々なデータや専門家の持つ暗黙知を形式化できる環境が整備されつつある。しかし、膨大な量のデータと情報で構成される文脈の中でノウハウによる判断がどのような影響を与えるかまで踏み込んで研究し十分な成果が出ているとは言い難い。空間的・時間的に非均一、非較正、非均質な多種多様なセンシングデータに対して、きめ細かい観測、作業や観測データの記録・解析といった要素を統一的に実現可能とする時系列相関分析技術の開発に取り組み、他プロジェクトで遂行中の実験環境で有効性を検証することで、きめ細かい観察、作業や観察データの記録・解析がどの程度の解像度で必要なのか、異種データ間の遅延相関や相関ルール解析、環境制御との因果関係などの詳細解明を促進できると考える。

例えば、精密農業に関して、必要な土壤栄養素にも注意を払い環境負荷をより低減させた高品質・安全・低コスト施設生産システムの確立や持続可能農業の実現、自給率の向上だけでなく、多品種少量生産や家庭菜園、車載菜園といった小規模農業分野にも適切な規模で柔軟に展開でき、ICTの活用に慣れた新たな担い手の参入にも大きな期待を持てる。以上のように、国内外の研究動向および申請者のこれまでの研究成果から、申請者の持つ情報通信技術に関する知識や技術と農林技術研究等のエキスパートからのノウハウを連携させることで、双方の分野において学術的かつ実効的な新たな波及効果が期待できると確信する。

2. 研究の目的

多種多様なセンシングデータに対して、きめ細かい観測、作業や観測データの記録・解析、時系列相関分析といった要素を統一的に実現可能とする時系列相関分析技術の開発に取り組み、他プロジェクトで遂行中の実験環境を用いてその有効性を検証する。具体的には、申請者らが既に実施中の知的クラスター創成事業(第II期)『自律分散協調ユビキタスセンサネットワーク』や科学研究費若手(A)『高精密農業を可能とするマルチベンダ

センサグリッドの実証的研究』にて構築中のセンサネットワークからのセンシングデータに対し、空間的、時間的なデータの相関関係を多様な解像度で分析可能なユーザインタフェースを実現させる。例えば、果樹の各育成ステージの環境条件(日照時間、CO₂濃度、気温、養液量、相対湿度、光強度など)に関して多様な解像度で統一的特徴抽出を実現可能とする。これを用いて、ノウハウを持つがICTの利活用に長けていない専門家らと共に状況の分類や暗黙知の形式化を促進させ、各育成ステージでどのような判断がどのような影響を与えるかを定量的に分析可能な時系列相関分析環境を構築する。

- 在席パターン分類精度(下グラフ参照)
 - 0.18(単一)⇒0.38(時系列)⇒0.9(複合)
- 離席パターン分類精度
 - 0.9(単一)⇒0.8(時系列)⇒0.98(複合)

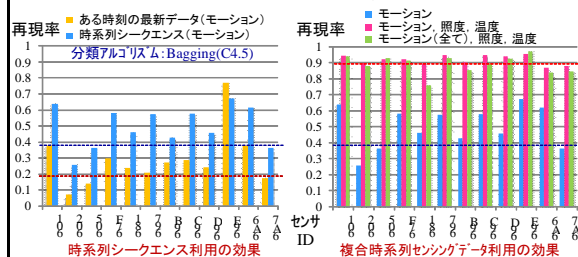


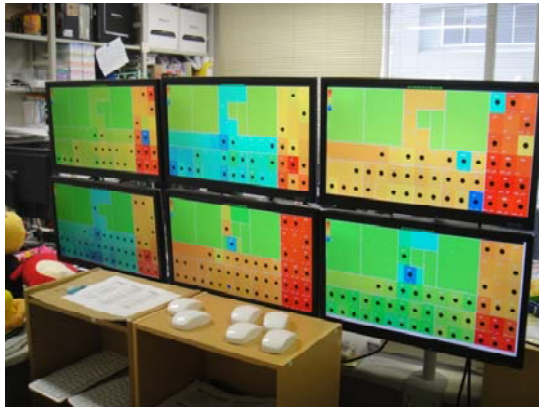
図 1. 在席推定のための属性組合せ比較

3. 研究の方法

まず課題となるのは、多種多様なセンシングデータからどのように有用な部分情報の抽出を行うかである。センサは情報を受動的に獲得可能であるが、環境中の背景とみなしたい構成要素の中には、定常的なノイズ発生源や、緩やかな変動要因が存在することで、周囲環境からのノイズが混入することが考えられる。そこで、多種多様なセンシングデータ間の波形パターン組合せの動的特性に着目し、ノイズが混入するような環境でも有用な部分情報を抽出することを考える。これは、申請者らが遂行している知的クラスター創成事業の研究成果からの着想で、複数の異種センシングデータを統合して得た時系列シーケンスを利用した在席状況推定では、90%の再現率を得ることができ、人の有無を検知することが目的でないセンサ群(モーション、照度、温度)であっても、複数の異種センシングデータを統合して得た時系列シーケンスを利用することで高精度な在席状況推定を実現できる可能性が示されたという事実に基づく(図1)。

次に課題となるのは、センサフュージョンの方法である。多種多様なセンシングデータ間を完全に同期させることは困難なため、多種多様なセンシングデータ間の波形パターン組合せが完全に一致するとは限らない。そ

ここで、各センシングデータの特徴量を統一的に扱うための粒度の均質化について検討し、観測された時系列シーケンスを類似したものと判別するための距離の定義を行う。また、実際に取得した時系列シーケンスにおいて、定義した距離を用いたクラスタリングを行い、その結果について考察を行う。以上のアイデアは、これまでに類似したものは提案されておらず、斬新なアイデアかつチャレンジングな手法であると考えられる。



4. 研究成果

図 2. ビル 6 階分に見立てた表示例

大規模可視化のミドルウェアをベースに開発することで、センシングデータの時間的な変化に関して空間的解像度を変更可能とするようにした。また経過時間バーを異動させることで自由にその時間変化の様子も分析できるようにした。さらにフローアや研究室など選択された多種多様なセンシングデータ（モーション、温度、照度など）間の関連性を視覚的に認識できるように、複数種類のセンサを組合せて表示できるセルという単位を考え、それらを空間的に並べて表示させることで多様な次元に対応させるだけでなく、時系列データの範囲指定することで時間的解像度（曜日、時刻など）も柔軟に変更可能にすることができるようになった（図 2）。Web ブラウザを利用して閲覧可能とすることで、一般的な Web コンテンツとして開発可能である。また、本システムでは各画面の同期に係るネットワークトラフィックを除けば、任意の画面数を扱えるように実装することが可能で、大規模なセンシングデータストリームの閲覧にも適用可能である。

また、有意部分の抽出とデータ粒度の均質化に関して、連続するいくつかの時系列データをセグメントとしてまとめ、一定の長さを持つ時系列データをセグメント単位としてその後の判別処理に用いることを考えた。まずセンシングデータ（モーションセンサ）を用いた実験により複数の同時刻の短いセグメント（30 秒）をクラスタリングし、さらに結果を用いて系列内の長いセグメント（30

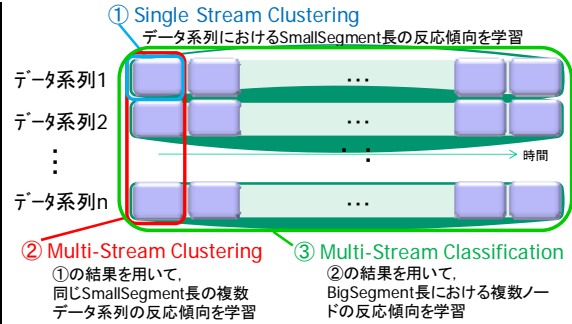


図 3. 多段推定の概要

分)を対象にまとめるという二段階のクラスタリングにより、有意なデータ系列（在席状態）を発見することができた（図 3）。そのようにして見つけたデータ系列から他の手段で得た情報を用いて正解データを作った。そのデータを正解として対象となる時系列データの状態（在席か離席か）を判定する方式を複数の分類方式（C4.5、ナイーブベイズ、k-NN など）を適用して実験し、k-NN で 95% 以上の正解率を達成することができた（図 4）。

本多段推定アルゴリズムを簡単に表現すると、時系列クラスタリングの結果を複数のセンサ群、長期データ推定にクラスタリングと分類学習を使い、拡張させた手法と言える。

最初の SingleStream Clustering では、各センサの短期間（部分シーケンス）のセンサ反応傾向の推定を行う。センサデータは、各ノード毎に別々のストリームデータであり、 $X=(x_1, x_2, \dots, x_n, \dots)$ の値からなる半無限長のシーケンスである。最初に、このデータストリーム X に対して時間軸方向に一定の長さで分割する。この部分シーケンスを SmallSegment と呼ぶこととする。DTW (Dynamic Time Warping) などセグメントを火変調とすることも可能である。その後、固定長に切り出した部分シーケンスから特徴量を抽出し、クラスタリング手法を用いた部

多段階推定の効果 (Recall)

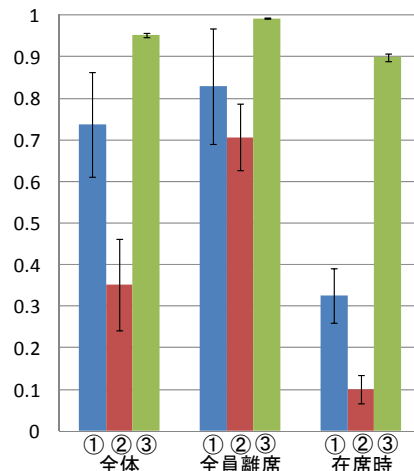


図 4. 多段階推定の効果 (Recall)

分シーケンスの特徴をグループ分けする。部分シーケンスの特徴量は、検知回数と検知のばらつきを用いる。クラスタリングには、k-means を用い、距離の定義は標準ユークリッド距離を用いた。また、滑走窓 (Sliding Window) 方式を用いて、反応の推移も漏れを少なくなるようにした。

MultiStream Clustering では、Single Stream Clustering による推定結果を利用し、短期間の複数台のセンサ群の反応傾向の推定を行う。同時刻帯の複数センサ群の反応傾向を推定するため、特徴量として各センサノードの SingleStream Clustering の推定結果、つまり、所属クラスターを用い、クラスタリングによりグループ分けする。距離の尺度としてハミング距離を用いた。

MultiStream Classification では、MultiStream Clustering の推定結果を利用し、複数台のセンサ群全体における SmallSegment 長より時間幅の大きい区間の在席推定を分類器学習を用いて行う。長期間のデータを使い、また、教師あり学習による推定を行うことで、即時性は低いが高精度な在席推定が可能となる。この時間幅の大きい区間を BigSegment と呼ぶこととする。MultiStream Classification では、BigSegment における SmallSegment の集約値と共に、在席状況の正解データを教師データとして用い分類器学習させ、その分類器を用いて推定フェーズで在席状況を推定する。

以上のような方法で、推定粒度の抽象化と、多段階の出力による即時性と制度のトレードオフを可能とし、設置センサの台数や設置場所、センサ個数に柔軟な推定を行える手法の有効性を検証した。その結果、多様なセンシングデータの粒度や位相のずれをある程度吸収し、有意なデータのみ注目できることを確認することができた。

今後の課題として、機械学習のモデル作成データに依存しすぎる点で、モデル作成に使う教師データに依存しすぎてしまうとモデル作成データでない別データにあてはめた際、精度よく推定ができなくなることである。この過学習という課題に対しては、再サンプリングを行うことや自動学習機能によって解決していけると考える。

5. 主な発表論文等

[学会発表] (計 11 件)

- ① Nobuhiko Matsuura, Seiji Suzuki, Maki Ohata, Ken Ohta, Hiroshi Inamura, Tadanori Mizuno, Hiroshi Mineno, “An Efficient and Highly Available Distributed Data Management System,” The 6th International Conference on Information Systems, Technology & Management (ICISTM2012), 2012. 3. 30 ,

Grenoble Ecole de Management (Grenoble, France)

- ② 江田政聡, 賀新剛, 横山昌平, 福田直樹, 峰野博史, 石川博, “赤外線センサを用いた在席推定に基づく証明制御手法の提案,” 第 4 回データ工学と情報マネジメントに関するフォーラム (DEIM フォーラム 2012), 2012. 3. 5, シーサイドホテル舞子ビラ神戸 (兵庫県)
- ③ 中根傑, 江田政聡, 横山昌平, 福田直樹, 峰野博史, 石川博, “WebSocket を用いた大規模センサデータストリームの可視化システムの開発,” 第 4 回データ工学と情報マネジメントに関するフォーラム (DEIM フォーラム 2012), 2012. 3. 5, シーサイドホテル舞子ビラ神戸 (兵庫県)
- ④ 江田政聡, 中根傑, 横山昌平, 福田直樹, 峰野博史, 石川博, “センサルームにおける赤外線センサを用いた人の移動・在籍状況の推定と利用,” 第 3 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2011), B8-6, 2011. 2. 28, ラフォーレ修善寺 (静岡県)
- ⑤ 鈴木誠二, 峰野博史, 太田賢, 水野忠則, “多種多様なデータストリームを用いた家電制御システムの構築,” 情報処理学会 BCC (放送コンピューティング研究グループ) 第 25 回研究会, 2011. 1. 22, 薩摩川内市里公民館 (鹿児島県)
- ⑥ 松野智明, 増井崇裕, 安部恵一, 峰野博史, 大須賀隆司, 水野忠則, “無線センサネットワークを利用した農業支援環境の見える化に関する検討,” 情報学ワークショップ 2010 (WiNF2010) 論文集, pp. 103-106, 2010. 12. 10, 名古屋工業大学 (愛知県)
- ⑦ 増井崇裕, 松野智明, 安部恵一, 峰野博史, 大須賀隆司, 水野忠則, “高密度無線センサネットワークを利用した農業技術の形式知化に関する検討,” 情報処理学会 DPS ワークショップ, Vol. 2010, No. 11, pp. 95-100, 2010. 10. 28, 青島サンクマール (宮崎県)
- ⑧ 中根傑, 横山昌平, 福田直樹, 峰野博史, 石川博, “Tiled Display を用いた大規模センシングデータの可視化システム,” 電子情報通信学会データ工学研究会 (DE2010), 2010. 6. 2, 名古屋大学 (愛知県)

6. 研究組織

(1) 研究代表者

峰野 博史 (MINENO HIROSHI)
静岡大学・情報学部・准教授
研究者番号: 40359740

(2) 研究分担者

石川 博 (ISHIKAWA HIROSHI)

静岡大学・情報学部・教授

研究者番号：60326014

(3)連携研究者

()

研究者番号：