

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 17 日現在

機関番号：17401

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24500138

研究課題名(和文)大規模サーバにおける温度センサ監視と熱処理の効率化に関する研究

研究課題名(英文)Monitoring and Forecasting of Temperature Sensor Data for Large-scale Computing Servers

研究代表者

櫻井 保志(Sakurai, Yasushi)

熊本大学・自然科学研究科・教授

研究者番号：30466411

交付決定額(研究期間全体)：(直接経費) 4,100,000円

研究成果の概要(和文)：本研究では、計算機サーバのCPU温度をリアルタイムに予測するための時系列データマイニング技術を開発した。計算機科学の中でも時系列解析は、計算理論、データベース、データマイニングなど幅広い分野において数十年に渡って取り組みが行なわれている、いわゆる「よく知られた」研究課題である。しかし、センサデータなど、近年のデータは大規模かつ多様化しており、そのような複雑な時系列ビッグデータを解析し、将来予測のような高度な処理を行なうため、本研究では(1)自動特徴抽出、(2)大規模テンソル解析、(3)非線形解析、という研究の新しい3つの方向性を示し、先進的な解析技術を開発した。

研究成果の概要(英文)：We developed time-series data mining techniques to predict the future CPU temperature for data-center monitoring. Time-series data analysis is a well-known topic that has attracted huge interest in various research communities (e.g., theory, databases, data mining, networking) for a few decades. The increasing volume of online, time-stamped activity represents a vital new opportunity for data analysis, and the most fundamental requirements are the efficient and effective mining of big time-series data with high variety information. We presented three new directions for research on time-series analysis, which include: (1) automatic mining, (2) large-scale tensor analysis, (3) non-linear modeling, and we developed efficient and effective data mining techniques.

研究分野：時系列データマイニング

キーワード：データマイニング 時系列データ センサデータ

1. 研究開始当初の背景

信頼性および拡張性が高い計算機サーバを大量に抱える近年のデータセンタはネットワーク社会において重要な社会基盤となっている。その重要性とともにデータセンタの電力消費量は急増している。本研究では、計算機サーバのCPU温度をリアルタイムに予測し、その予測結果に基づき計算機の負荷を制御することにより、計算機の電力消費量を低減化させることを目的として取り組みを開始した。ただ、研究期間の途中で研究代表者の所属が企業から大学に変わり、研究環境が大幅に変わったため、負荷制御やシステム化に関する実践的な取り組みよりも、センサデータを始めとする時系列データの解析やマイニングアルゴリズムなど、理論的な研究に重きを置いて取り組んだ。

2. 研究の目的

データセンタにおいては、各計算機サーバの温度管理が極めて重要である。そしてセンサデータから異常を検知する、もしくは異常状態をその兆候から事前に検知するためには、センサデータのモデル化が必要となる。これはセンサデータの予測を行なう場合も同様である。本研究ではセンサデータのような大規模な時系列データ、すなわち時系列ビッグデータからその特徴やダイナミクスをとらえるための高速かつ先進的な時系列マイニング技術を開発することを目的とする。

3. 研究の方法

計算機科学の中でも時系列解析は、計算理論、データベース、データマイニングなど幅広い分野において数十年に渡って取り組みが行なわれている、いわゆる「よく知られた」研究課題である。しかし、センサデータストリームなど、近年のデータは大規模かつ多様化しており、そのような複雑な時系列ビッグデータを解析し、将来予測のような高度な処理を行なうため、本研究では従来の研究とは異なる、以下のような新たな方向性を示し、研究開発を行なった [T1]。

(1) 自動特徴抽出

実用的な技術やシステムの開発を考えた際、最も重要である課題が特徴抽出の完全自動化である。例えば、自己回帰やフーリエ変換、特異値分解をはじめとする従来の時系列解析では、使用する係数の個数や閾値等のパラメータの設定が必要である。しかしセンサデータのように、解析するデータが膨大になるほど、専門のエンジニアによる細やかなパラメータチューニングには時間的、金銭的コストが多くなり、重大なボトルネックとなる。このようにビッグデータの解析において、解析技術の自動化は極めて重要な課題である。

(2) 大規模テンソル解析

時間発展の情報を含むデータの多くは、テ

ンソルとして表現することができる。例えば、データセンタにおいて各計算機からいくつものセンサデータをモニタリングしている場合、(timestamp, sensor ID, machine ID, ...) という m 個の属性からなるレコード群は、 m 階のテンソルとして表現することができる。時間情報をともなう大規模なテンソル、すなわち時系列ビッグデータテンソルの解析技術は今後重要な要素技術の一つとなる。

(3) 非線形時系列解析

非線形モデルは、疫学、生物学、物理、経済など、様々な分野で用いられている。一方で、データマイニング分野においては、ソーシャルメディアやオンラインユーザアクティビティの分析が盛んであり、一部の研究ではソーシャルメディアのダイナミクスをとらえるために非線形モデルが用いられている。このような研究が行なわれている中で、著者らは非線形モデルとテンソルを統合した非線形テンソル解析技術を考案した。非線形テンソル解析は計算機科学において世界で初の取り組みであり、各時系列データに対し個々の独立したモデルを当てはめるのではなく、時系列ビッグデータを多角的に、かつ非線形性を考慮しながら分析することにより、各シーケンス間の潜在的な関係性をとらえることが可能となる。

4. 研究成果

本研究ではセンサデータの解析、そしてWeb情報や医療情報の解析も含めながら、大規模時系列データを対象として、情報抽出、パターン検出、モデル学習、将来予測などの研究を行なった。

4.1 大規模時系列シーケンスからの特徴自動抽出

大規模な時系列シーケンスの中から、典型的なパターンや異常値を発見することは非常に重要な課題である。考案手法である AutoPlait [C7] は、大規模時系列データを対象とし、重要な時系列パターンの抽出を自動的に行なうことができる。特に、多次元時系列シーケンスを扱い、これらのデータ全体を表現する要約情報を抽出する。

4.1.1 提案手法とその特徴

図1はモーションキャプチャセンサデータに対する AutoPlait の出力結果例である。この時系列データは、「チキンダンス (chicken dance)」と呼ばれるダンスのステップを表現している。このモーションは、4次元のシーケンスで構成され、それぞれの次元が左右の腕と足の加速度を表現している。チキンダンスは、beaks, wings, tail-feathers, claps の4つの代表的なステップから構成されている。ここで、ステップのような部分シーケンスの特徴パターンをレジームと呼び、この時系列データは4つのレジームから構成され

る。図1の下段は、AutoPlaitが自動抽出した4つのレジームを示している。提案手法は、ダンスに含まれる4つのレジームを抽出し、そして各レジームの切れ目も正しく発見することができる。AutoPlaitは、これらレジームに関する事前知識を必要とせず、適切な数のレジームとその位置を自動的に把握することができる。

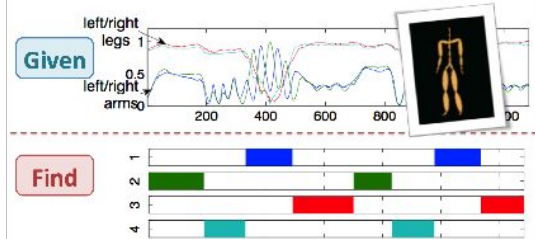


図1 モーションキャプチャデータからの特徴抽出

AutoPlaitの目的は、与えられた時系列シーケンス群 X の特徴を抽出し、すべての時系列パターンを表現するパラメータ集合 $C = \{m, r, S, \theta, F\}$ を発見することである。ここで、 m はセグメント数、 r はレジーム数、 S はセグメント、 θ はレジーム、 F セグメントメンバシップであり、これらを AutoPlait は自動的に抽出する。以下では、AutoPlait の2つの重要なアイデアについて説明する。

(1) 多階層連鎖モデル (MLCM: multi-level chain model)

複数のレジーム間の時系列パターンとその遷移を表現するために、多層的な連鎖モデル (MLCM) を提案する。提案モデルである MLCM は隠れマルコフモデル (HMM: Hidden Markov Model) を拡張しており、従来の HMM の遷移確率に加え、上位層の状態 (super-state) の概念を導入することによって、パターンのグループ化を行なう。

(2) モデル表現コストとデータ圧縮

適切なセグメントとレジームの検出のため、最小記述長 (MDL: minimum description length) の概念を用いる。MDL は情報理論に基づくモデル選択基準のひとつで、可逆圧縮を行なうことができる。本研究では、与えられた時系列シーケンス X を適切に表現するモデルを見つけるために、新しい符号体系を定義した。具体的には、(a) 最適なパラメータ集合 C を推定するためのコスト関数を定義し、(b) 最適解を発見するための効果的なアルゴリズムを提案する。

4.1.2 アルゴリズム

図2は、AutoPlaitの最適化アルゴリズムの概要を示す。提案アルゴリズムは、次に挙げる3つの部分問題に分割される。

Algorithm 1, CutPointSearch: レジームの個数 ($r = 2$) とモデルパラメータが与えられたとき、 X を2つのレジームに分割し、それぞれのセグメントの分割位置を検出する。

Algorithm 2, RegimeSplit: レジームの個数 $r = 2$ が与えられたときに、2つのレジームを表現するモデルパラメータ ($\theta_1, \theta_2, \Delta$) を推定する。

Algorithm 3, AutoPlait: 最適なレジームの個数 ($r = 2, 3, 4, \dots$) を求める。

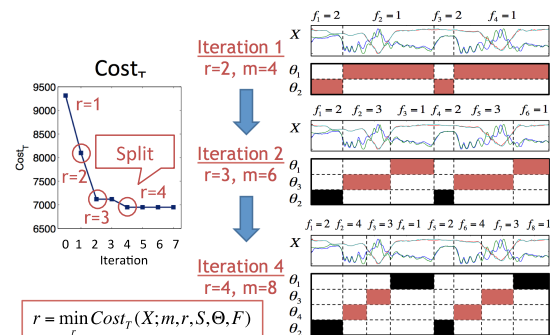


図2 提案アルゴリズムの概要

4.2 テンソルおよび非線形解析

4.2.1 TriMine: 複合イベントデータのためのテンソル解析と将来イベント予測

TriMine [C4] は複数の属性で構成される複合イベントデータが与えられたとき、イベントの中から重要な要素 (トピック) を抽出し、潜在的なトピックの時間的な推移を把握することにより、将来のイベントを予測するアルゴリズムである (図3)。

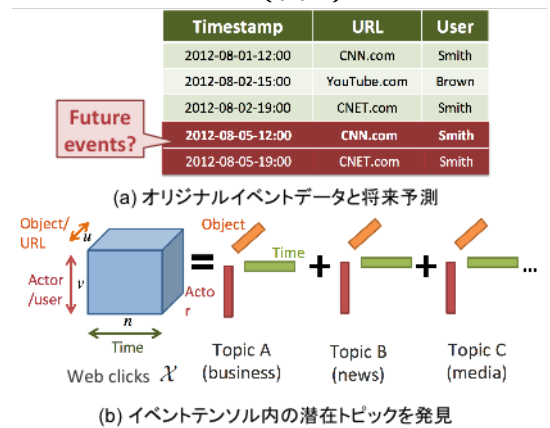


図3 複合イベントデータ解析

4.2.2 SpikeM: ソーシャルネットワーク上の情報拡散過程と非線形解析

SpikeM [C3] は、Blog や Twitter を始めとするソーシャルネットワーク上において、噂やニュース等の情報が、時間が経過するごとに、どのように拡散し減衰していくかを表現する (図4参照)。提案手法は、ユーザ間のグラフ構造に基づく情報拡散過程をパワー則に基づく非線形モデルで表現し、将来の情報拡散過程を予測することができる。

4.2.3 EcoWeb: 生態系モデルに基づくオンライン活動上の競合関係自動抽出

EcoWeb [C10] は生態系における種間競争モデルに基づき、Web 上のユーザ活動の中か

ら潜在的な競合関係や季節性等の重要パターンを自動抽出する非線形モデルである。

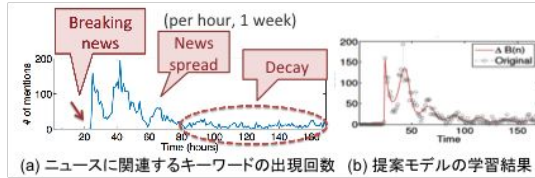


図 4 Blog 上の情報拡散過程とモデル学習

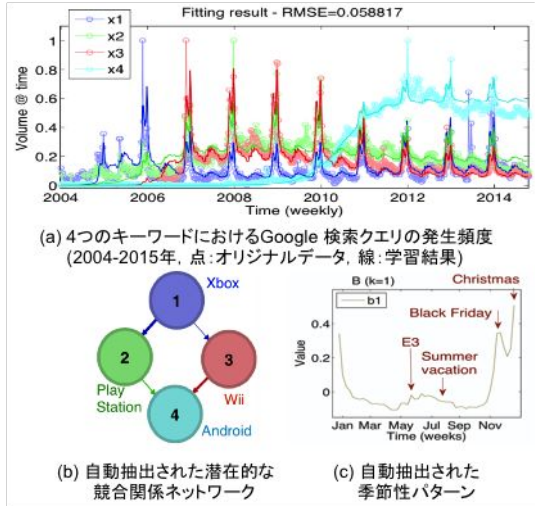
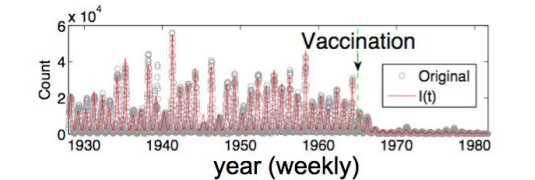


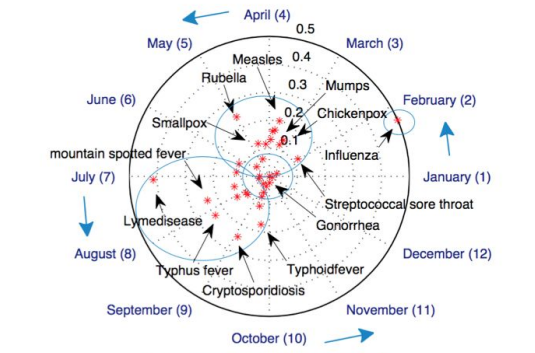
図 5 Google 上の活動パターンと特徴抽出

4.2.4 FUNNEL: 大規模疫病テンソルデータのための非線形解析モデル

FUNNEL [C8] は大規模な疫病感染データ (timestamp, disease, location) に対し、非線形モデル学習とテンソル解析技術を融合し、疫病の感染力、季節性、地域性、ワクチン効果等の重要なパターンを自動抽出する統合モデルである。



(a) はしかの患者数の推移とFUNNELの学習結果



(b) 主要な疫病の季節性

図 6 大規模疫病テンソルの非線形解析

以上のように、本取り組みでは、センサーデータのみならず、Web 情報や医療情報も含めながら、大規模な時系列データを対象として、

情報抽出、パターン検出、モデル学習、将来予測などの研究を行なった。特に、(1) 自動特徴抽出、(2) 大規模テンソル解析、(3) 非線形解析、という時系列データマイニングにおける新しい3つの方向性を示し、研究に取り組んだ。また、上記の成果を中心に、研究代表者らの過去の時系列解析に関する全ての成果をまとめ、ACM SIGMOD 2015 の Tutorial 講演を行なうこととなった [T1]。

本研究では学術的な貢献のみならず、実用性を重視し、研究成果は全てフリーソフトウェアとしてリリースし、広く一般に公開している。これまでの開発技術については様々な企業からの問合せがあり、例えば AutoPlait [C7] については、技術を発展させてセンサーデータ処理に関する共同研究を開始するに至り、現在も継続して企業との共同研究を行っている。今後、本研究を発展させるとともに、特に国産企業との連携を意識し、国内の経済と産業を支える基盤技術の確立を目指し、技術開発を続けていく予定である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

研究会などの国内発表は省略する。Google scholar において 50 件以上引用されている国際会議論文については、その被引用数情報を記載する。また、主な国際会議論文には採択率情報についても記載する。

[雑誌論文](計 6 件)

[J1] 前川 卓也, 柳沢 豊, 岸野 泰恵, 石黒 勝彦, 櫻井 保志, 亀井 剛次, 岡留 剛: “手首に装着したカメラ付きセンサーデバイスを用いた行動認識手法”, 電子情報通信学会論文誌 B, 査読有, Vol. J95-B, No. 11, pp. 1480-1490, 2012 年 11 月.

[J2] M. Toyoda, Y. Sakurai, Y. Ishikawa: “Pattern Discovery in Data Streams under the Time Warping Distance”, VLDB Journal, 査読有, Vol. 22, No. 3, pp. 295-318, DOI 10.1007/s00778-012-0289-3, June 2013.

[J3] T. Maekawa, Y. Kishino, Y. Sakurai, T. Suyama: “Activity Recognition with Hand-worn Magnetic Sensors”, ACM/Springer Personal and Ubiquitous Computing (ACM/Springer PUC), 査読有, Vol. 17, Issue 6, pp. 1085-1094, DOI 10.1007/s00779-012-0556-8, August 2013.

[J4] 松原 靖子, 櫻井 保志, B. Aditya Prakash, Lei Li, Christos Faloutsos: “情報拡散過程のダイナミクス: 非線形モデルの提案と情報予測”, 情報処理学会論文誌: データベース, 査読有, Vol. 6, No. 5, pp. 11-22, 2013 年 12 月.

[J5] 松原 靖子, 櫻井 保志, Christos Faloutsos, 岩田 具治, 吉川 正俊: “大規

模 Web クリックデータのためのイベント予測”, 電子情報通信学会論文誌 D, 査読有, Vol. J97-D, No. 4, pp. 822-834, 2014 年 4 月.

[J6] 松原 靖子, 櫻井 保志, Christos Faloutsos: “大規模時系列データの特徴自動抽出”, 情報処理学会論文誌: データベース, 査読有, Vol. 7, No. 2, pp. 37-50, 2014 年 6 月.

〔学会発表〕(計 18 件)

国際会議

[C1] T. Maekawa, Y. Kishino, Y. Yanagisawa, Y. Sakurai: “Mimic Sensors: Battery-shaped Sensor Node for Detecting Electrical Events of Handheld Devices”, International Conference on Pervasive Computing (PERVASIVE), pp. 20-38, Newcastle, UK, June 2012 (採択率: 28/138, 20.3%). 発表日: 2012 年 6 月 20 日.

[C2] T. Maekawa, Y. Kishino, Y. Yanagisawa, Y. Sakurai: “Recognizing Handheld Electrical Device Usage with Hand-worn Coil of Wire”, International Conference on Pervasive Computing (PERVASIVE), pp. 234-252, Newcastle, UK, June 2012 (採択率: 28/138, 20.3%). 発表日: 2012 年 6 月 21 日.

[C3] Y. Matsubara, Y. Sakurai, B. A. Prakash, L. Li, C. Faloutsos: “Rise and Fall Patterns of Information Diffusion: Model and Implications”, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp. 271-279, Beijing, China, August 2012 (採択率 (full presentation): 133/755, 17.6%, Citations: 70). 発表日: 2012 年 8 月 13 日.

[C4] Y. Matsubara, Y. Sakurai, C. Faloutsos, Tomoharu Iwata, M. Yoshikawa: “Fast Mining and Forecasting of Complex Time-Stamped Events”, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp. 6-14, Beijing, China, August 2012 (採択率 (full presentation): 133/755, 17.6%). 発表日: 2012 年 8 月 13 日.

[C5] Y. Matsubara, L. Li, E. E. Papalexakis, D. Lo, Y. Sakurai, C. Faloutsos: “F-Trail: Finding Patterns in Taxi Trajectories”, Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), pp. 86-98, Gold Coast, Australia, April 2013 (採択率: 59/344, 17.2%). 発表日: 2013 年 4 月 15 日.

[C6] Y. Kishino, Y. Sakurai, Y. Yanagisawa, T. Suyama, F. Naya: “SVD-based hierarchical data gathering for environmental monitoring”, ACM

International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp), pp. 9-12, September 2013 (Adjunct Publication). 発表日: 2013 年 9 月 10 日.

[C7] Y. Matsubara, Y. Sakurai, C. Faloutsos: “AutoPlait: Automatic Mining of Co-evolving Time Sequences”, ACM SIGMOD Conference, pp. 193-204, Snowbird, Utah, June 2014 (採択率: 107/421, 25.4%). 発表日: 2014 年 6 月 24 日.

[C8] Y. Matsubara, Y. Sakurai, W. G. van Panhuis, C. Faloutsos: “FUNNEL: Automatic Mining of Spatially Coevolving Epidemics”, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp. 105-114, New York City, August 2014 (採択率: 151/1036, 14.6%). 発表日: 2014 年 8 月 25 日.

[C9] Y. Matsubara, Y. Sakurai, N. Ueda, M. Yoshikawa: “Fast and Exact Monitoring of Co-evolving Data Streams”, IEEE International Conference on Data Mining (ICDM), Shenzhen, China, December 2014 (採択率 (full paper): 71/727, 9.8%). 発表日: 2014 年 12 月 15 日.

[C10] Y. Matsubara, Y. Sakurai, C. Faloutsos: “The Web as a Jungle: Non-Linear Dynamical Systems for Co-evolving Online Activities”, International World Wide Web Conference (WWW), Florence, Italy, May 2015 (採択率: 131/929, 14.1%). 採録決定, 発表予定.

チュートリアル講演

[T1] Y. Sakurai, Y. Matsubara, C. Faloutsos: “Mining and Forecasting of Big Time-series data”, ACM SIGMOD Conference, Tutorial 2, Melbourne, VIC, Australia, May/June 2015. 採録決定, 発表予定.

招待講演

研究代表者の招待講演のみ示す.

[I1] 情報処理学会連続セミナー2012 第1回 ビッグデータの解析と価値発見 招待講演 “データストリームのためのマイニング技術とその応用”

主催: 情報処理学会
場所: 化学会館 (情報処理学会)
日時: 2012 年 6 月 25 日

[I2] 情報処理学会第 155 回データベースシステム研究会 招待講演 “大規模 Web クリックデータのためのイベント予測”

主催: 情報処理学会データベースシステム研究会
場所: 東京都 秋葉原コンベンションホール

日時：2012年11月19日
[I3] 情報処理学会第155回データベースシステム研究会 招待講演
「情報拡散過程のダイナミクス：非線形モデルの提案と情報予測」
主催：情報処理学会データベースシステム研究会
場所：東京都 秋葉原コンベンションホール
日時：2012年11月19日
[I4] Invited Talk: "Automatic Mining of Big Time-series Data", the 8th Korea-Japan International Database Workshop (KJDB), Kumamoto, Japan, February 22, 2014.
[I5] 人工知能学会 第95回人工知能基本問題研究会(SIG-FPAI) 招待講演
「時系列ビッグデータからの特徴自動抽出」
主催：人工知能学会人工知能基本問題研究会
場所：大阪大学産業科学研究所
日時：2014年10月10日
[I6] Invited Talk: "Automatic Mining of Spatially Coevolving Epidemics", the 9th Korea-Japan Database Workshop (KJDB), Gangwon, South Korea, December 1, 2014.
[I7] 情報処理学会第77回全国大会「トップランナーが語るITの未来」招待講演、パネル討論
「時系列データマイニングの立場から見た近未来」
主催：情報処理学会、独立行政法人科学技術振興機構
場所：京都大学
日時：2015年3月19日

〔産業財産権〕
出願状況(計 2 件)

[P1]
名称：類似部分シーケンス検出装置，方法，及びプログラム
発明者：豊田 真智子，櫻井 保志
権利者：日本電信電話株式会社
種類：特許
番号：特願 2012-133152 (登録番号)
出願年月日：2012年6月12日
国内外の別：国内

[P2]
名称：時系列階層モデルのためのデータ更新処理装置，方法，及びプログラム
発明者：櫻井 保志，松原 靖子，吉川 正俊
権利者：日本電信電話株式会社，京都大学
種類：特許
番号：特願 2012-152857 (登録番号)
出願年月日：2012年7月6日
国内外の別：国内

取得状況(計 0 件)

〔その他〕

受賞

研究代表者の受賞のみ示す。

[A1] 第6回データ工学と情報マネジメントに関するフォーラム (DEIM2014) 最優秀論文賞 (2014年6月21日)
[A2] 平成26年度 情報処理学会 山下記念研究賞 (2015年3月17日)

6. 研究組織

(1) 研究代表者

櫻井保志
(Sakurai Yasushi)
熊本大学・自然科学研究科・教授
研究者番号：30466411

(2) 研究分担者

吉川正俊
(Yoshikawa Masatoshi)
京都大学・情報学研究科・教授
研究者番号：30182736

(3) 連携研究者

該当なし