

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 13 日現在

機関番号：94301

研究種目：挑戦的萌芽研究

研究期間：2015～2016

課題番号：15K12112

研究課題名(和文)複数情報源からの異種データに対する統合的解析法

研究課題名(英文)Integrated analysis for heterogeneous data from multiple sources

研究代表者

川鍋 一晃(KAWANABE, Motoaki)

株式会社国際電気通信基礎技術研究所・脳情報通信総合研究所・主幹研究員

研究者番号：30272389

交付決定額(研究期間全体)：(直接経費) 2,700,000円

研究成果の概要(和文)：実環境で計測される異種情報源データから心身の健康に役立つ個性・心理指標を連続的に推測するための機械学習法の研究開発と実データによる評価を行った。まず、心身の健康に影響を与える因子として個々の注意レベルと行動パターンに着目し、前者については聴覚SART実験時の脳波と生体信号の同時計測データから心電パルス間隔などの注意レベルに関連する特徴量を抽出し、後者については移動履歴データから行動パターンを効率的に推測するためのテンソル分解法を開発した。また、複雑な欠測パターンを持つ異種情報源データを最大限活用するため、行列分解やARモデルを用いた欠測補完法や傾向スコア法の改良を提案し、その有用性を示した。

研究成果の概要(英文)：We developed machine learning techniques to infer continuously personality traits and psychological measures useful for mental/physical health based on heterogeneous data from multiple sources in real environments. As factors affecting on health status, we focused on attention levels and behavioural patterns. For the former, we measured simultaneously EEG and biomedical signals from several subjects during an auditory sustained attention response task (SART), and extracted features related to attention levels such as ECG R-R intervals. For the latter, we developed a tensor decomposition method to infer behavioural patterns efficiently from mobility traces. Furthermore, in order to make full use of heterogeneous multi-sensor data with complex missing patterns, we proposed completion methods of missing observations based on matrix factorisation or AR models, and improvements of propensity score estimation with deep neural networks.

研究分野：医用工学・数理統計学

キーワード：確率的情報処理 ライフログ 異種データの統合解析

1. 研究開始当初の背景

(1) 近年、情報通信技術の発達により、カメラや GPS などによる行動履歴に加え、ウェアラブル 生体センサや脳活動計測機器を用いることで、様々な種類の情報源から構成されるライフログの取得が容易になりつつある。これに対して、日常生活環境では過去と全く同じ状況は存在しないという問題の複雑さゆえに、このような異種情報源ライフログから「健康寿命の延伸」といった社会的課題の解決につながる有益な情報を精度良く推測できていない。ビッグデータを活用して実問題応用が進んでいる機械学習法の高度化により、この分野でのブレイクスルーが望まれている。

(2) われわれはこれまで BMI

(brain-machine interface) に役立つ機械学習法の研究を通じて、脳信号からの意図推定による生活支援に挑戦してきた。その過程で、カメラなどの各種センサを設備した実験住宅を ATR 敷地内に構築し、日常生活の複数情報源データを常時計測できるデータ取得環境を整備した。失われた機能を補完する BMI から一歩進んで、この実験住宅ライフログから心身の健康に影響を及ぼす因子を捉えられるようになれば、「健康寿命の延伸」という社会的課題への貢献が期待できる。

2. 研究の目的

心身の健康状態を調べる方法として、現在は質問紙法や脳構造 MRI (核磁気共鳴画像) 計測データ解析が用いられているが、コストや利用者にかかる負担の問題から頻繁には実施できない。そこで、本研究では心身ともに健康的な生き方を支援するため、従来の質問紙や構造MRI解析から得られている知見に基づき、日常環境で常時計測される複数情報源データを活用して、心身の健康に役立つ個性・心理指標を連続的に推定するデータ解析技術の開発をめざす。特に、複数情報源データの複雑な欠測パターンへの対応と、多人数データからの共通成分と個人特有の要素の分離法に重点を置く。

3. 研究の方法

(1) 生体・脳信号や画像といったデータと質問紙等により定められる個性指標との関連を正準相関分析などの探索的データ解析法により同定し、ライフログの各ドメインに共通し、個性との関連を最大化する特徴空間を求める。

(2) 個性の計測はアンケートなどに頼るため、常時計測するライフログに比べて稀にしか得られない。また、カメラ死角への移動などにより常時計測データにも部分的に観測値が得られないことがある(欠測)。欠測がランダムに発生していない場合、単純に欠測データを無視すると、推定結果にバイアスを生じることが知られている。そこで、高次元データ

を効率的に処理するための低ランク近似という機械学習法と、バイアスがないことを保証するセミパラメトリック欠測統計法の両者を統合して、複雑な欠測のある複数情報源ライフログからの情報復元法を開発する。

(3) 欠測が補完された多人数の完全データを対象とし、教師なし学習により個人ごとの特性と、類似した人々に共通する性質をそれぞれモデル化し、複数人のデータを効率よく共有するための転移学習法を構築する。これにより、大規模データの効率的な利用が可能となり、より高精度で安定した個性推定が可能となる。

4. 研究成果

(1) 経験サンプリングを用いた英国での大規模調査によれば、日常生活においてどれだけやらなければいけないことに注意を向けられているか、あるいは現実から離れたことをあれこれ考えている (Mind Wandering) が個人の幸福度に大きく関係している (Killingworth et al., Science, 2010)。そこで、複数情報源データから心身の健康に関連する因子を調べるために、安静時および持続的注意レベルを測るための聴覚 SART (Sustained Attention Response Task) という実験課題を行っているときに、脳波と生体信号(心電、皮膚反応、血中酸素飽和度、眼電)の同時計測を行った。この聴覚 SART では、1セッション中に約 2.5 秒おきに 0~9 の数字がランダムに 13~36 回読み上げられる。読み上げられた数字が 3 以外のときにはただちにスペースキーを押すが、3 のときには何もしないという単純な課題を各被験者に対して 72 セッション以上行った。また、各セッションの最後には、被験者の注意状態に関して、課題に集中していたか、他のことを考えていて集中していなかったかを質問し、さらに、そのときの気分が快か不快かどちらでもないかを聞いた。計 12 名被験者に対してデータを収集した。

まず、生体信号については、アーチファクトの影響が少なく、十分な試行数が揃っている 9 名分のデータを解析した。心電パルス間隔 (R-R 間隔) のばらつきの指標である LFHF と VLF、および血中酸素飽和度 (SpO₂) で、数字の 3 を読み上げられたときに課題に集中しておらず誤ってキーを押してしまった試行 (誤警報群) と課題に集中しておりキーを押さなかった試行 (正解群) で差が見られた。しかし、このイベントの数が少ないために標準偏差が大きく、統計的有意性は言えなかった。

また、脳波についてはノイズの多い被験者を除いた 7 名のデータを対象とし、パイロット解析として事象関連電位などの群平均の比較を行った。持続的注意レベルの高い正解群 (NMW) と注意散漫状態の誤警報群 (MW) の電位時系列を比較すると、前者の方がより

P3成分が大きくなっており、その空間パターンを調べたところ、正解群ではCzからPzあたりでP3成分が特に高くなっていることがわかる(図1)。

以上をまとめると、脳・生体信号から計算される特徴量の中には持続的注意レベルとの関連傾向が見られるものが存在した。しかし、個人間で誤警報率が10~50%と大きくばらついたため、統計的に有意な結果を得るには至らず、さらに実験課題を改善する必要がある。

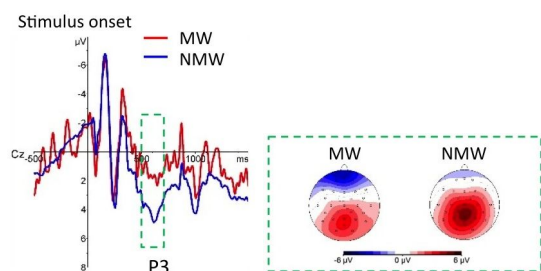


図1 脳波の事象関連電位解析。チャンネルCzの信号時系列(左)と事象関連電位P3の空間パターン(右)。

(2) 日常生活における注意レベルやMind Wanderingの頻度だけでなく、個々の行動パターンも心身の健康に大きな影響を与える因子である。近年モバイル端末のGPS機能などにより日常空間における人の位置情報系列を記録、活用することができるようになっており、このような情報から個々の行動パターンを抽出することは重要なテーマである。実環境ライフログ全般的に言えることであるが、移動履歴データも過去と全く同じ経路は存在しない複雑な問題であるため、行動に関連しない差異を取り除いて、各種の行動パターンに要約するための優れた機械学習法が必要になる。本研究ではテンソル分解法を用いた移動履歴データの要約法を提案した(Murakami et al. 2015)。特に、群疎性(group sparsity)を仮定することで、移動履歴データの効率的な要約を可能とし、複数の実データセットを用いてその性能を評価した。

GPSなどの移動軌跡データベースではプライバシー保護のために一部の移動履歴を削除することがあり、このような欠測に対応するために、上記の研究で用いた群疎性正則化付テンソル分解に基づく欠測データの処理法を提案した(Murakami et al., 2017)。この手法は少数訓練サンプルにおいても移動パターンの個人特性を捉えることができる特徴量表現を与える。一方で、公開データ内の匿名化された移動履歴のユーザ同定性能が従来法より高まるため、プライバシー攻撃への対応策の必要度も増すことがわかった。

(3) 日常生活中にモバイル端末で利用者に質問する経験サンプリングの手法がライフログ

研究でも活用されるようになってきたが、状況あるいは心境が整わずに回答が得られないということが起こる。このような複雑な欠測パターンを持つ質問応答データに対して、行列分解法に基づいて欠測推定を行う方法論を開発した(兼村ら2015, 2016)。本手法を適用することで、欠測値のある質問応答データを有効活用することができ、結果的に質問回数や煩雑さの軽減につながることを期待される。

また、実環境で生体信号を継続的にモニタリングする場合、計測エラーや予期せぬノイズが混入し、複雑な欠測が生じ得る。特に、脳波時系列の欠測補完を行うためにARモデルを応用した方法論を開発し、実データを用いて評価を行った(金子ら2015, 鄭ら2016)。

(4) 複雑な欠測パターンを持つ場合、欠測データを無視した推論では結果にバイアスを生じる危険性がある。欠測統計学で知られている傾向スコア(propensity score)は、統制の不十分な実験データ(調査観察データ)を補正して、統制群を計算的に作るのに有用な概念である。従来は線形モデルであるロジスティック回帰により傾向スコアを計算していたが、複雑な実環境ライフログに適用する場合には、非線形学習法を用いた方が良かった。具体的には、複数情報源データを統合解析するため、情報源ごとに傾向スコア法で欠測値による偏りを補正する際に、代表的な非線形学習法である畳み込みニューラルネットワークとGBDT (gradient boosting decision tree)により傾向スコアを求める方法を試み、ベースラインであるロジスティック回帰法やランダム法を性能で上回ることを確認した(内橋・兼村2016, Uchihashi & Kanemura, 2016)。

(5) 連続的に計測、保存されるライフログは長時間にわたるが、その中で意味のある日常生活行動、心身の健康に関連する箇所は一部であると考えられる。そのような重要な情報を持った部分を膨大なデータストリームから抽出する技術がますます重要になる。その解決のために、異種情報源からなるライフログの長時間記録の中から、有用な情報を持つ(key frame)を抽出するための機械学習法を開発した(Li et al., 2017)。一人称映像に加えて加速度データを活用することにより、日常行動に無関係なアーチファクト区間(frame)を除去することが可能となった。

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)

〔雑誌論文〕(計7件)

Li, Y., Kanemura, A., Asoh, H., Miyanishi, T., and Kawanabe, M., Key frame extraction from first-person video with multi-sensor integration,

Proc. of IEEE ICME 2017, 査読有, 2017, 6 pp.

Murakami, T., Kanemura, A., and Hino, H., Group sparsity tensor factorization for re-identification of open mobility trace, IEEE Trans. on Information Forensics and Security, 査読有, Vol.12, No.3, 2017, 689-704, DOI: 10.1109/TIFS.2016.2631952

Uchihashi, K., and Kanemura, A., Modeling the propensity score with statistical learning, Lecture Notes in Computer Science, 査読有, Vol.9950, 2016, 261-269, DOI: 10.1007/978-3-319-46681-1_32

Kanemura, A., Lipowski, G., Komine, H., and Akaho, S., Automatic categorization of health indices for risk quantification, Procedia Computer Science, 査読有, Vol.63, 2015, 325-331, DOI: 10.1016/j.procs.2015.08.350

Murakami, T., Kanemura, A., and Hino, H., Group sparsity tensor factorization for de-anonymization of mobility traces, Proc. of IEEE TrustCom-15, 査読有, Vol.1, 2015, 621-629, DOI: 10.1109/Trustcom2015.427

兼村厚範、大成弘子、鹿内学、橋本将崇、赤穂昭太郎、能力テスト得点の非負行列分解、信学会技報 IBISML、査読無、115巻、2015、203-208

兼村厚範、畑中哲生、鹿内学、能力テストにおける非負行列分解を用いた適応型出題の精度解析、信学技報 NC、査読無、115巻、2015、71-75

[学会発表](計10件)

Li, Y., Kanemura, A., Asoh, H., Miyanishi, T., and Kawanabe, M., Key frame extraction from first-person video with multi-sensor integration, IEEE ICME 2017, 2017.7, Hong Kong, China

鄭郁森、兼村厚範、福永修一、繰り返し推定による脳波時系列の欠測補完、情報論的学習理論ワークショップ (IBIS2016)、2016.11.17、京都

Uchihashi, K., and Kanemura, A., Modeling the propensity score with statistical learning, ICONIP2016, 2016.10.18, Kyoto, Japan

川鍋一晃、ブレイン・マシン・インタフェース、日本磁気学会 第4回岩崎コンファレンス、2016.5.17、東京(招待講演)

兼村厚範、畑中哲生、鹿内学、能力テストにおける非負行列分解を用いた適応型出題の精度解析、通信学会 NC 研究会、2016.3.23、東京

兼村厚範、大成弘子、鹿内学、橋本将崇、赤穂昭太郎、能力テスト得点の非負行列分解、情報論的学習理論ワークショップ (IBIS2015)、2015.11.27、つくば

金子拓光、兼村厚範、福永修一、多変量自己回帰モデルを利用した脳波時系列の欠測値推定、情報論的学習理論ワークショップ (IBIS2015)、2015.11.27、つくば

Kawanabe, M., On possibility and difficulties in adaptive brain-computer interface in real environments, Germany-Japan Workshop on Adaptive BCI, 2015.10.28, Kyoto, Japan (招待講演)

Kanemura, A., Lipowski, G., Komine, H., and Akaho, S., Automatic categorization of health indices for risk quantification, IEEE ICTH 2015, 2015.9.28, Berlin, Germany

Murakami, T., Kanemura, A., and Hino, H., Group sparsity tensor factorization for de-anonymization of mobility traces, IEEE TrustCom-15, 2015.8.20, Helsinki, Finland

6. 研究組織

(1) 研究代表者

川鍋 一晃 (KAWANABE, Motoaki)
国際電気通信基礎技術研究所・脳情報通信総合研究所・主幹研究員
研究者番号: 30272389

(2) 研究分担者

兼村 厚範 (KANEMURA, Atsunori)
国立研究開発法人産業技術総合研究所・人間情報研究部門・研究員
研究者番号: 50580297

(3) 連携研究者

金井 良太 (KANAI, Ryota)
株式会社アラヤ・ブレイン・イメージング・代表取締役兼 Chief Scientist
研究者番号: 80607177