

心身の健康増進に寄与するビッグデータ活用法

企画者	日本健康心理学会研究推進委員会*
司会者	嶋田洋徳(早稲田大学)*
話題提供者	吉本潤一郎(奈良先端科学技術大学院大学)
話題提供者	荒牧英治(奈良先端科学技術大学院大学)
話題提供者	山本哲也(徳島大学)*
指定討論者	山蔦圭輔(大妻女子大学)*

企画趣旨

近年の国際動向として、心理・行動・生体情報といった、さまざまな健康情報に関するデータベースが公開され、多くのオープンデータが利用可能な状況となってきた。わが国でも、厚生労働省が所有する電子レセプトのデータベース(National Database)が、使用目的の制限はあるものの、一般に利用可能な状況になるなど、ビッグデータの活用の際に門戸が開きつつある。そのため、従来は扱うことが困難であった健康に関わるこうしたビッグデータを活用することは、健康心理学の発展に寄与するのみならず、今後の健康心理学の新たな方向性を見出す上でも有用である可能性がある。

そこで本シンポジウムでは、ビッグデータを用いた研究を実践されている吉本潤一郎氏、荒牧英治氏、山本哲也氏から、ビッグデータを用いた具体的な研究例を紹介していただく。加えて、ビッグデータ解析を健康心理学的研究手法の新たなアプローチの1つとするための方向性について、山蔦圭輔氏に指定討論を行っていただく。

今後、ビッグデータの活用した研究は大きく広がりを見ることが予想される。本シンポジウムを通して、「健康心理学にビッグデータをどのように活用していくべきか」という点について、フロアを交えて活発な議論を行いたい。

話題提供者 吉本潤一郎(奈良先端科学技術大学院大学)

【機械学習によるビッグデータ解析と精神医学への活用法】

機械学習とは、与えられたデータからその背後に潜む規則性を発見し、予測や意思決定に役立てようとする情報技術の総称である。近年のIoTやウェアラブルデバイスを活用した計測技術の発展により、我々はライフログに代表されるような大規模かつ高次元のデータを小コストで容易に取得できるようになってきた。一方で、データの大規模化は解析の手間を増大させることにも繋がるため、機械学習を活用した知識発見の自動化が生命科学や医療の分野でも試みられている。しかしながら、一言で機械学習といっても多種のアルゴリズムが提案されており、各アルゴリズムの特性を知らないままに活用しても、多くの場合は望ましい結果を得ることはできない。そこで、本講演では、機械学習手法をユーザとして利用する際に最低限理解しておきたい概念や理論をまず紹介する。

続いて、機械学習を用いたデータ解析に関する精神医学への応用事例をいくつか紹介する。第1は、認知課題中に計測した脳活動データからうつ病患者群と健常対照群の判別を試みたものである。この応用では、サンプル数に比べて入力変数(説明変数)の数がはるかに大きい場合、実効的な入力変数の次元を削減する工夫を入れなければ、汎化性能の低い判別器が構築されてしまう。この問題を解決するために、我々は変数選択能力を持つL1正則化ロジスティック回帰法を用いた判別器を構築する手法を提案した。10分割交差検証により提案手法を評価した結果、精度91.95%±2.02%(感度:91.35±2.90%, 特異度:92.55±2.70%)という高い判別精度を実現することができた。

第2は、安静時fMRIデータから計算される脳領域間の機能結合性、問診データ、血中マーカーを属性に持つ多次元データから、明示的な教師ラベルを与えず、データ駆動的にうつ病サブタイプの発見を試みたものである。ここでは、類似性の高いサンプルと属性群を同時にグループ化し、かつ、最適なグループ数も推定できるベイズ多重クラスタリング手法を開発し、データ解析を行った。その結果、グループの違いがうつ病重症度を表す指標(HRS-DやBDI-IIなど)と相関を持つ属性群が同定された。この属性群を詳細に調べたところ、セロトニン再取り込み阻害剤に対する治療反応性が幼児期のトラウマ体験を表す指標(CATS)と右角回をハブとするネットワークの機能結合性と相関が高いことを見出した。

計測機器間あるいは施設間の汎化能力など克服すべき課題はまだ残されているものの、以上の結果は、機械学習に基づく多次元データ解析がうつ病の診断やサブタイプを同定する上で有用なツールであることを示唆している。

なお、発表に関連し、開示すべき利益相反関係にある企業などはない。

話題提供者 荒牧英治(奈良先端科学技術大学院大学)

【ソーシャルメディア情報や医療情報などに基づくビッグデータの活用方法】

現在、医療現場には、画像、音声、テキスト、検査数値など様々なデータがあり、その多くはそのまま機械学習で処理可能である。しかし、医療現場で生成される多様なデータの形式は、診断であったり、アセスメントであったり、相当な

部分は自然言語文で表現されている。そして、その自然言語文こそ、もっとも扱いが難しいデータである。つまり、いかに自然言語文を処理するかが、医療分野におけるビッグデータの大きな課題である。本稿では、この言語を扱う技術について概観する。医療分野に存在するテキストは、大別すると、

(1) 論文や教科書などのフォーマルなテキスト(公開テキスト)、(2) 医療者が現場で記述するテキスト(臨床テキスト)、そして、(3) 患者が作成するテキスト(患者テキスト)の3つがある。

公開テキスト

公開テキストの利活用としては、主に医学論文が材料とされている。医学論文の蓄積は膨大なものになりつつあり、例えば、医学論文データベースであるPUBMEDには現在すでに2400万件を超える論文のabstractが登録されている。この膨大な成果を1人の専門家、または、少数の研究グループが把握することは困難となりつつあり、学問領域を検索する技術、及び、俯瞰する技術が研究されている。

この解析の基盤となる大規模なコーパス(用語の関係などの教師ラベルが付与されたテキストデータ)も必須である。

臨床テキスト

現在、より注目を浴びているのは、臨床現場で業務を行なっている最中に自然と蓄積される臨床テキストの利活用である。臨床テキストと呼んだが、一口に臨床テキストといっても、膨大な種類が数と種類がある。これらは基本的には、病院運営の業務のために使用されるが、蓄積されたテキストを利用することでかつてない大規模な研究が実施可能となる。ただし、記述方法については医療従事者に一任されてきた面が多く、医療情報(例えば、傷病名や愁訴、検査、治療に関する記述)を抽出し、標準化することは難しく、新たな自然言語処理の課題として活発に研究されている。

患者テキスト

患者の医療の語りを集め、利活用するというまったく新しい試みが行われはじめています。患者の多くは、ある日、突然病気になる、十分に知識がないまま闘病生活を迎えている。そのような患者にとって、すでに長い闘病生活を経ている先輩患者のノウハウには重い価値がある。そこで、患者の記録の体系だった集積や整理を目指して、闘病記録や患者ソーシャルメディアへの投稿記事のライブラリ化や検索サービスなどを付加して、利活用する試みが始まっている。また、インフルエンザや花粉症、アレルギーなどのコモンディーズについては、これまでの材料の量を遥かに凌駕したビッグデータが利用可能である。とくに感染症については実用化間際である。患者テキストは、3つの効果が期待されている。第一に、患者の悩みを患者達が共有することでQOLの向上に寄与する可能性がある。また、患者が正しい知識を身に付けることで、医療従事者とのコミュニケーションがスムーズになり、結果的に、医療の質の向上することが考えられる。さらに、医療従事者側にとっても、医療従事者には語らない患者の不満や不安を知ることができ、教育用のコンテンツとなりうる。このように、患者と医療従事者の双方に利点があるものの、

患者同士のコミュニケーションの円滑化のためには、プライバシーを保護しつつ、必要な情報にアクセスできるという情報技術が必要となる。

利益相反開示；発表に関連し、開示すべき利益相反関係にある企業などはありません。

話題提供者 山本哲也(徳島大学)

【心理学研究におけるビッグデータの活用方法】

心理学におけるビッグデータの活用方法に注目が集まっている。たとえば、心身の健康、物質使用、行動変容、ソーシャルメディア、職場のウェルビーイング、学校場面における適応や学習効率をはじめ、多岐にわたる心理学研究領域での応用が期待されている。一方で、どのように多様なデータを測定・解析し、知見を見出すか、といった研究方法の実際については、いまだ発展の余地を大いに残しており、健康心理学領域における研究技術の創出がのぞまれる。そこで本講演では、容易で簡便なデータ測定を可能とするデバイスを用いて、個人の生活情報を1年以上にわたって調査した講演者の研究例を紹介する。

ウェアラブルデバイスやスマートフォンなどの様々な機器を活用し、人々の生活に関わる多様な情報(以下、ライフログ)を記録する情報技術が、近年急速に発展している。こうした技術は、人々の行動的・生理的情報から位置情報に至るまで、リアルタイムで収集可能であり、私たちの健康増進に寄与する可能性を大きく有していることが指摘されている(山本・吉本, 2017)。一方で、データ量が膨大になりやすく、データのモダリティも様々であることから、ライフログから有用な知見を見出すには困難を伴うことが少なくない。そのため、煩雑なライフログの情報を効果的に整理する解析手法が極めて重要となると考えられる。今回紹介する研究例では、(a)ポジティブ/ネガティブ気分などの気分状態の頻度、(b)思考内容や反すうなどの認知行動傾向の頻度、(c)腹痛や抑うつなどの心身の症状の有無、(d)睡眠時間や運動量などの生活習慣、(e)ポジティブ/ネガティブな出来事といったライフイベントの頻度など、計43種類の特徴量を用いる。これらの特徴量は、2種類のスマートフォンアプリに加え、活動量や心拍数、睡眠ステージなどの記録を可能とするリストバンドを用いて測定されている。

上記の研究例を題材としながら、本講演では近年の人工知能の発展に大きく寄与した「機械学習法」を活用し、ライフログに内在する規則性をデータ駆動的に発見することを試みる。そして、ライフスタイルを可視化することを通して、ライフログを用いた健康心理学研究の応用可能性について、議論を深めることを目的としたい。

利益相反開示；本シンポジウムにおける発表に関連し、開示すべき利益相反関係にある企業などはない。

(SHIMADA Hironori, YOSHIMOTO Junichiro,
ARAMAKI Eiji, YAMAMOTO Tetsuya,
YAMATSUTA Keisuke)