

身体データの解析による健康情報マイニング

北林 宏樹[†] 吉川 正俊^{††}

[†] 京都大学情報学研究科 〒606-8501 京都府京都市左京区吉田本町

^{††} 京都大学情報学研究科 〒606-8501 京都府京都市左京区吉田本町

E-mail: [†]kitabayashi@db.soc.i.kyoto-u.ac.jp, ^{††}yoshikawa@i.kyoto-u.ac.jp

あらまし 近年、あらゆるデータのビッグデータ化が進み、個人の身体データの記録も容易に行えるようになってきている。身体データとは、体温・血圧・心拍数などであり、センサを身につけることにより記録することが可能である。このような個人的な身体データを時系列で連続して記録し、ビッグデータとして有効に利活用するためのマイニング手法を提案する。ビッグデータを分析することにより、隠れたトレンドの発見や将来のイベント予測が可能である。ビッグデータを分析する手法として非負値行列因子分解（NMF）を用いる。血圧や心拍数といった身体データを分析することにより、個人の健康状態の将来予測などを行うこと目標とする。

キーワード ビッグデータ、マイニング、ライフログ

1. はじめに

近年、あらゆるデータのビッグデータ化が進み、個人の身体データの記録も容易に行えるようになってきている。身体データを取得する方法としてセンサを身につけるといことがある。近年、ウェアラブルなセンサが開発され、様々な身体データを記録できるようになった。ウェアラブルなセンサの例として Google のグーグルグラス [1] や Nike の Nike+ [2] などが挙げられる。

今後は、これらのウェアラブルなセンサを用いて身体データをライフログとして記録していくユーザが増えていくことが予想される。ライフログを記録していくことによって、後から自分のデータを振り返ることができ、日常生活や健康状態を改善することができる。また、単にデータを記録し、そのデータを自ら振り返るだけでなく、自動で健康状態に関する情報をマイニングし提供する機能があれば有益であると考え、マイニング機能があれば、自分自身では気づけなかったことに気づき、健康状態の改善のための糸口を発見する手助けになる。また、多くのユーザの身体データを集約することで、初めて見えてくるトレンドもあると考えられる。また、多くのユーザの身体データから得られるマイニングによって、各個人にアラートするといったことも応用可能である。本研究では、大規模な身体データからある具体的なトレンドを発見し、健康状態をマイニングしていくことを目標とする。

マイニングを行うためには、まずデータから特徴を発見しなければならない。特徴を発見するための手法として非負値行列因子分解（NMF）[3] を用いる。NMF によって、大規模な身体データから隠れたトレンドや特徴を発見することを実現する。NMF によって発見した特徴を学習アルゴリズムによって学習することでマイニングを実現していく。

また、身体データをライフログとして記録していくにあたって、そのデータを管理し、利活用できる仕組みが必要となる。我々はこの点に関して、情報銀行 [4] に着目している。情報銀行とは、個人の活動履歴情報を「銀行口座」に蓄積し、本人の管理の下でプライバシーを守りながら、自分や社会へ向けて高次利活用することを目的とした社会システムである。この情報銀行を活用し大規模な身体データを管理・利活用するといった展望を考えている。

2. 関連研究

ビッグデータの高速な解析手法として TriMine [5] がある。TriMine は得られたデータをテンソル分解することによって隠れたトレンドを発見する手法である。

テンソルとは、大きさと複数の向きによって特徴づけられる物理量である。テンソル分解について示した図を図 1 に示す。この例では、テンソル分解によってオブジェクト・アクター・時間の 3 つのトピックベクトルに分解する。分解した後、複数のスケールでイベントテンソルを作成する。

多次元時系列をもとに将来を予測する手法として TriMine-F [5] がある。イベントテンソルに分解した後、多次元時系列に変換する。それぞれの時系列に対し独立に予測を行う。テンソルから時間行列を予測し、3 つの行列からイベントを生成することによってマイニングを行う。

また、血圧や体温、体重といったデータを記録し、管理するサービスとして goo からだログ [6] がある。データをログとして記録することにより、自分の過去のデータを振り返ることが可能である。ユーザ数が多く、データの規模も非常に大きい。主にデータの記録・管理が中心で、解析・マイニング機能はあまり充実していない。

3つのベクトル集合 = 3つのトピック行列

- [O] オブジェクト行列 ($u \times k$)
- [A] アクター行列 ($k \times v$)
- [C] 時間行列 ($k \times n$)

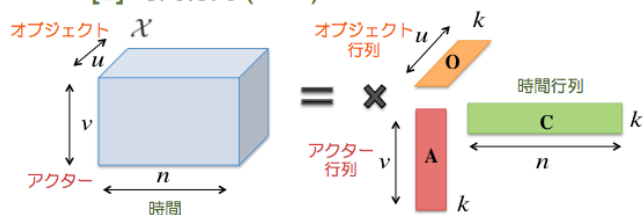


図1 テンソル分解 [5]

3. 身体データ取得から健康情報マイニングまでの全体プロセス

身体データ取得からマイニングまでの全体プロセスを図2に示す。

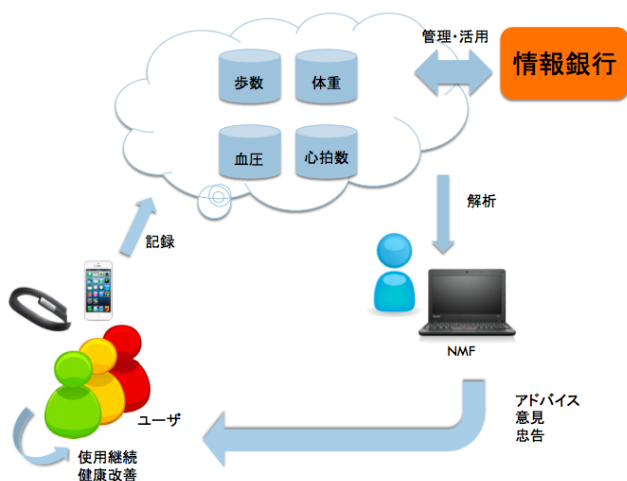


図2 全体プロセス図

3.1 身体データの取得

マイニングに用いる身体データを取得するため、ウェアラブルなセンサを利用する。取得するデータは、血圧や体温、心拍数といった身体データである。ウェアラブルなセンサを常に身につけることにより、身体データを時系列に記録する。

3.2 NMF

身体データをマイニングするにあって、個人から得られた身体データから特徴を発見することが必要となる。身体データから特徴を発見する手法として非負値行列因子分解 (NMF) を用いる。因子分解とは、積を求めることで再びこの行列を構築できるような二つの小さな行列を探し出すことである。ビッグデータに含まれる重要な情報は一部であり、大半は冗長な情報である。NMF は、ビッグデータから重要な特徴のある情報を発見することに優れていると考え、本研究で使用する。

NMF の手法について具体的に述べる。例えば、ある複数の記事中での単語の数の辞書と、それぞれの記事中での単語の数の辞書を手にしていることとする。これらの情報を単語の数を持つ記事の行列に変換する。この行列を因子分解することによって、二つの小さな行列を生成する。

二つの小さな行列は、特徴の行列と重みの行列である。

特徴の行列を図3に示す。

	hurricane	democrats	Florida	elections
特徴1	2	0	3	0
特徴2	0	2	0	1
特徴3	0	0	1	1

図3 特徴の行列の一部

特徴の行列では、行はそれぞれの特徴であり、列がそれぞれの単語を表している。値は単語が特徴に対してどれくらい重要であるかということを示している。それぞれの特徴は記事のセットから引き出したテーマを表している必要がある。

重みの行列を図4に示す。

	特徴1	特徴2	特徴3
hurricane in Florida	10	0	0
Democrats sweep elections	0	8	1
Democrats dispute Florida ballots	0	5	6

図4 重みの行列の一部

重みの行列では、複数の特徴を記事の行列にマップする。それぞれの行は記事であり、列は特徴である。値はそれぞれの特徴がそれぞれの記事にどの程度適合するかということを表している。

特徴の行列と重みの行列の積を求めることによって、元の行列が再構築されることになる。

以上のように NMF で特徴を取り出した後、マイニングのプロセスに進む。マイニングは、NMF によって発見した特徴を活用することにより、実現させるものとする。

3.3 マイニング手法

マイニング手法として、SVM と重回帰分析を用いる。SVM は、線形入力素子を利用して2クラスのパターン識別器を構成する手法である。SVM を用いることにより、特定の病気であるか否かをマイニングできると期待できる。

重回帰分析は多変量解析の一つで、回帰分析の独立変数が複数あるものである。重回帰分析を用いて、病気の原因となる要素のマイニングが可能である。

3.4 情報銀行

近年、データのパーソナライゼーションの進展により、サービス事業者による個人情報の利用・管理負担が増大しているといった背景がある。このような課題を解決するために提案されたものが情報銀行 [4] である。情報銀行の概念図を図 5 に示す。

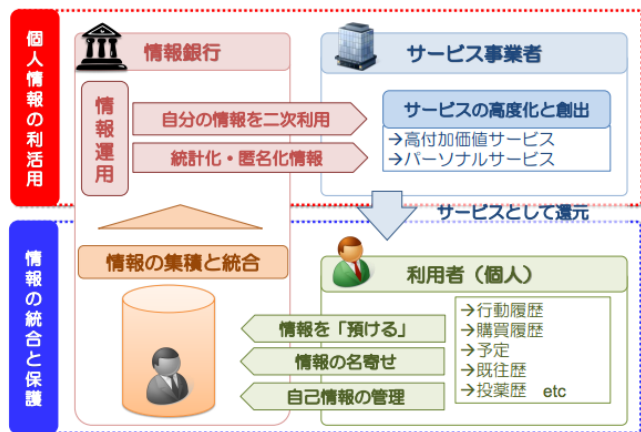


図 5 情報銀行 [4]

情報銀行とは、個人が位置情報をはじめとした自分自身の活動履歴情報を「銀行口座」に蓄積し、本人の管理の下でプライバシーを守りながら、自分や社会・産業へ向けて高次利用することを旨とする社会システムである。

個々人が自分の全ての情報にアクセス可能な唯一の主体であることに着目し、本人の管理の下で様々な活動情報（スマートフォンなどで取得可能な情報や事業者に散在する履歴情報）を統合することで、自分自身について広範に網羅された活動情報を作り出すことが可能になる。さらには、その網羅的な活動情報を本人の承認を通じて利活用することで、新たなサービスの創出や、既存サービスの高度化を期待することができる。

情報銀行は、個人から見た場合、パーソナル情報を信託すると、本人のため、社会のためにより役に立ちかつ安全な利用を促進する組織となる。また、企業から見た場合、統合されたパーソナル情報を安全に利活用し新しいサービス・ビジネス実現を可能にする組織となる。

情報銀行実現のポイントとして 2 点挙げられる。一つ目は、新しい価値を生み出す技術・ノウハウの蓄積と社会から信頼されるブランドとして「情報銀行」を確立することである。二つ目は、パーソナル情報のエコシステムを大きく育てることを通じて個人・産業・社会・世界に貢献していくことである。

この情報銀行を用いて、ウェアラブルなセンサによって取得した身体データを管理・利活用していくといった展望を考えている。ビッグデータ化によってデータの規模が大規模になっても情報銀行であれば扱うことが可能であり、個人ごとに様々なデータを集めることができるため、マイニングの可能性や精度

を高めることができると考えている。

4. 使用センサ

センサによって、体温・心拍数・血圧・歩数・カロリーなどを測定することが可能である。センサには様々な種類があるが、近年は体に身につけることができるウェアラブルなセンサが急増している。ウェアラブルなセンサの例を図 4～6 に示す。



図 6 グーグルグラス [1]



図 7 Nike+ [2]

ウェアラブルなセンサは大きく分けて眼鏡型と腕時計型がある。眼鏡型のセンサとして、図 6 のグーグルグラスなどがある。腕時計型のセンサとして図 7 の Nike+、図 8 のオムロン社の脈拍計などがある。

これらのウェアラブルなセンサは、昔から人間が常に身につけていて違和感を感じないのは、眼鏡と腕時計だけであるといった事実に基づいている。本研究では、センサから常にデータを取得できるという条件が重要であるのでウェアラブルなセンサを用いるということが理想である。

本研究では、ウェアラブルなセンサを用いて個人の身体データを記録するものとする。取得するデータは、主に体温・血圧・心拍数などである。どのウェアラブルセンサを使用するかといったことは今後決定していくものとする。



図 8 脈拍形 [7]

ウェアラブルなセンサーを選ぶ際に重要となる指標は、身につけやすさ・身につける箇所・大きさ・バッテリー・無線通信機能・デザインなどが挙げられる [8]。

5. 実 験

今後行っていく実験について述べる。患者から取得した身体データを解析することでマイニングを行い、マイニングの結果や精度を検証する。

身体データは、健康な人と不健康な人（特定の病気を持っている人）にウェアラブルなセンサを装着し、24 時間連続した時系列なデータとして取得する。取得するデータは、体温・血圧・心拍数・カロリー・睡眠時間・歩数などを想定している。また、取得の対象となる人数は、実際の状況を反映できる程度に設定する。

得られたデータから意味のある特徴を抽出するために NMF を用いる。得られた特徴を学習アルゴリズムなどで解析し、マイニングを行う。マイニングによって得られた結果と精度を検証し、考察につなげていきたいと考えている。

6. 今後の課題

本研究で身体データを用いたマイニングをするために、今後取得するデータの種類を具体的に決定していかなければならない。データの候補などは体温・血圧・心拍数などである。取得するデータに合わせて適切なセンサを選ばなければならない。分析対象となるデータの規模であるが、近年のビッグデータ化に伴い、実際の状況を反映できる規模の身体データを扱う予定

である。国内の個人達から時系列で連続したデータを取得することが理想である。また、これによって多ユーザの身体データによるマイニングも可能になる。個人の身体データを集約し、実際の状況を反映できる規模でのマイニングを目標とする。身体データを取得した後は、5 章で述べた実験を行っていく予定である。マイニングによって得られた結果や精度を検証し、考察につなげていきたい。

また、取得するデータは個人の身体データなどを含むためセキュリティといった観点でも課題がある。

情報銀行のセキュリティ強度についてはまだ明確ではないが、どこまでの情報ならば公開するといったアクセスコントロールや公開設定といった点を確固としたものにしていかなければならない。セキュリティの徹底によりプライバシー保護 [9] をできるようにしていきたい。

7. ま と め

本研究では、ウェアラブルなセンサを個人の身体に身につけることによって身体データを取得し、取得したデータから NMF を用いることによって特徴を発見し、個人の健康状態をマイニングすることを目指している。

データを管理する手法として、情報銀行を用いるという展望を考えている。情報銀行を用いることによって個人の身体データを集約し、よりマイニングの質と可能性を高めていきたいと考えている。

マイニングを行う手法は、NMF によってデータから特徴を発見した後に学習アルゴリズムによって学習することにより実現すると考えている。

文 献

- [1] <http://www.google.com/glass/start/>
- [2] <http://www.nike.com/jp/ja-jp/c/nikeplus-fuelband/>
- [3] Toby Segaran (2013), 集合知プログラミング, 245-270.
- [4] http://shiba.iis.u-tokyo.ac.jp/?page_id=410/
- [5] Yasuko Matsubara, Yasushi Sakurai, Christos Faloutsos, Tomoharu Iwata, Yasuko Matsubara, Masatoshi Yoshikawa, "Fast Mining and Forecasting of Complex Time-Stamped", ACM SIGMOD International Conference on Knowledge Discovery and Data Mining (KDD), pp271-279, Beijing, China, 2013.
- [6] <http://karada.goo.ne.jp/>
- [7] <http://www.healthcare.omron.co.jp/product/etc/hr/hr-500u.html/>
- [8] Alexandros Pantelopoulos and Nikolaos G. Bourbakis, "A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 40, NO. 1, JANUARY 2010.
- [9] 菊池浩明, "ライフログに関するプライバシーの課題とプライバシー保護データマイニングの展望", 信学技報, vol. 113, no.326, ISEC2013-64, pp. 35-38, 2013.