

「時間情報」に着目し、ビッグデータから価値を生み出す

新谷 研究室



新谷 隆彦
Takahiko SHINTANI

日々蓄積され、刻々と変化する

多様かつ膨大なデータは「ビッグデータ」と呼ばれ、その活用近年、注目が集まっています。しかし、増え続けるビッグデータからいかに価値ある情報を抽出し、知識に変えていくかという研究はまだ緒に就いたばかりです。

種々雑多なビッグデータを解析する手法の一つに「データマイニング」があります。同手法は、データベースに蓄積された大量のデータを解析する技術として登場しま

した。マイニングは「探鉱」という意味ですから、データマイニングとは、いわば広大な鉱山(ビッグデータ)から金鉱脈を探し出す(マイニング)イメージです。そこでは巧みな探鉱技術(アルゴリズム)が鍵を握ります。

データマイニングの説明で、よく例に出されるのが『おむつとビール』の法則です。何がいつ誰にどれだけ買われたかという商品の購買データを解析した結果、「おむつとビールが一緒に買われることが多い」という意外な併買パターンが明らかになったのです。こうしたパターンを見つけ出す「パターンマイニング」がデータマイニングの原点にあります。

新谷隆彦准教授は正統派のデー

タマイニングの分野で、特に実用性を重視したアルゴリズムの研究を行っています。以前、企業の研究所で実用価値の高い研究を指向していた経験から、「極端に言えば、美しいアルゴリズムを作るよりも、さまざまな制約条件のある実データから価値ある情報を取り出すことを目指したい」と考えています。

そこで新谷准教授は、これまであまり行われてこなかった「時間情報」を効果的に生かす研究に着手しました(図1)。具体的には、人の生活を記録した「ライフログ」をデータとして用い、時間情報に注目して人の行動パターンを抽出することを目指した研究です。この研究によって、「ライフ

ログマイニング」という新しい分野が誕生しました。

例えば、24時間365日ずっと手首に装着し、ライフログが取得できるリストバンド型のセンサが製品化されています。実験では、このようなセンサを使って約600日間にわたって生活を記録しました(図2)。

得られたデータはまず、1日ごとの「行動」に分類します。行動は複数の運動状態の集まりだと考えられるので、センサに内蔵されている加速度センサなどから、「静止時」「安息時」「歩行時」「活動時」「スポーツ時」といった運動状態のレベルに分けます。その上で、各状態が何時から何時まで行われたかという時間情報に変換しま

す。

各状態の「継続時間」と、各状態間の「時間間隔」とを両方考慮するきめ細かな時系列パターン処理を行うのが特徴です。例えば、「8時間睡眠後、朝起きて30分間ランニングする」というパターンと、「30分間昼寝をしてから3時間ジムで汗を流す」というパターンとを分けて抽出することができま

す。継続時間を考慮していなかった従来の手法では、これらは「寝て起きて運動する」という同一のパターンとして分類されていました。さらに、これらの行動が行われた順序を考慮することで、普段よく行うパターンと、そうでないパターンとに分けることができま

キーワード

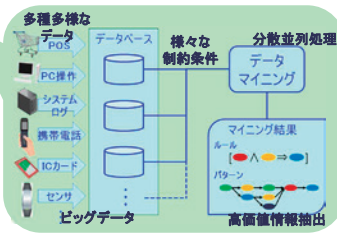
データ工学、データマイニング、大規模データ処理、並列分散処理、ライフログ

所属	大学院情報理工学研究科 情報・ネットワーク工学専攻
メンバー	新谷 隆彦 准教授
所属学会	情報処理学会、電子情報通信学会、日本データベース学会
E-mail	shintani@is.uec.ac.jp

大規模データの活用を実現するデータ処理基盤と高価値情報創出

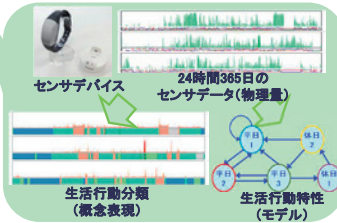
データマイニングの高可用性・高性能化

様々な制約条件の下で多種多様かつ膨大なデータを効率良く処理し価値ある情報を抽出するアルゴリズム、データ/結果の管理と表現、分散並列処理、秘匿性保持、更新データ処理の研究



ライフログマイニング/センサデータマイニング

腕に装着したセンサデバイスで長期間継続して測定したセンサデータからの生活行動分類、生活行動特性抽出、生活状態可視化の研究



実データからの高価値情報抽出

大規模な実データにデータマイニング技術を活用して価値ある情報を実際に抽出する実データ解析

図 1

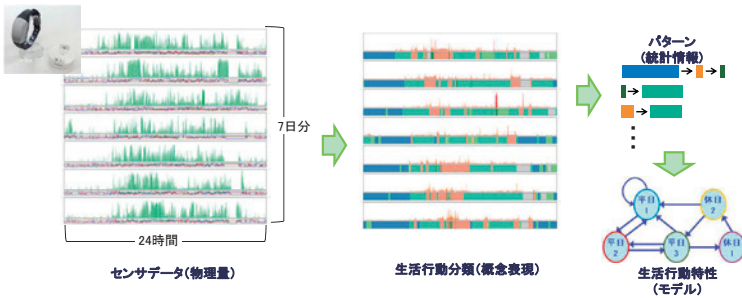


図 2

異なる期間の生活比較

● 生活行動<安静, 歩行, デスクワーク>の変化

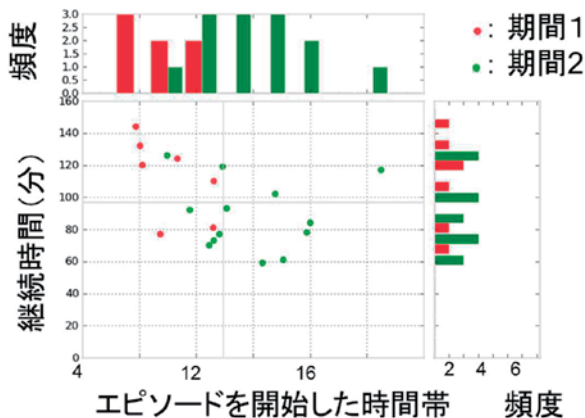


図 3

す。これによって、ある一日と別の一日の「生活パターンの類似度」を数値で表すことに成功しました。平日と休日の生活の違いなどを明確に導き出すことができず(図3)。

ともとライフログは、過去の記憶を呼び覚ます目的で作られました。記憶を無くした人の記憶補助ツールとして、よりきめ細かな治療にも役立つでしょう。また、生活の記録として考えれば、自分がどのくらい規則的に行動しているかが一目瞭然です。病気の治療中やトレーニング中、ダイエットなどに自ら生活スタイルをチェックできるでしょう。「時

間情報を取り入れたライフログマインニングを使って、大規模なデータから人に役立つ情報を提供したい」と新谷准教授は考えています。ビッグデータはさまざまな領域に入り込んでおり、その適用可能範囲は想像するよりずっと広いかもしれません。実際にゲーム会社と協力し、ソーシャルゲームの過去ログを使って、どんなパターンで行動す

るプレイヤーが高額課金者になりやすいかなどを予測する研究も行っています。また、ビッグデータを手軽に複数のマシンに分散して処理できるオープンソースのプラットフォーム「Apache Hadoop」を使った研究も手がけています。連携企業に対しては、「大規模なデータと顧客ニーズさえあれば、さまざまな形の共同研究が見込める。何ができるかを一緒に考

え、知恵をほしい」と新谷准教授は言います。研究室で取得したデータではなく、大規模な実データを用いることで実用が加速できるとみているからです。データを提供してもらえただけでありがたく、決して多くの研究費は望んでいないとのこと。【取材・文：藤木信穂】