

3.6 ITメディアとデータマネジメント

本俯瞰区分ではウェブ、ソーシャルメディア、センサーの多種多様で超大量なデータを統合、管理、検索、分析するための技術に関する研究開発を扱う。

モバイル技術の発展・普及によりスマートフォンなどの利用者が移動しながら情報の発信や利用を行うようになり、個人適応技術に支えられた SNS などプラットフォーム上にテキストや画像映像などのユーザー生成コンテンツ（Consumer Generated Media, CGM）が大量に生成されている。こうした CGM は Wikipedia のような知識の蓄積を生み出すだけでなく、人々が社会の状況を主観的に捉え情報発信する Twitter などは速報性の高いソーシャルセンサーと考えることができる。最近では位置情報が付加された CGM も増えており活用に向けた期待が一層高まっている。

もう一方で、工業生産設備の稼働状況のセンシングによる生産設備の最適運用や生産された工業製品の稼働状況をモニタリングにより適切なメンテナンスなどのサービスを組み込んだ付加価値づけといった流れが、社会インフラなどにも対象を拡大していくなかで、様々なセンサーから得られるデータの時間的・空間的な動態を把握し、解析する技術の重要性が高まっている。

このような背景から本俯瞰区分では次の 6 つを重要な研究開発領域として採りあげる。

1. ビッグデータの統合・管理・分析技術

スマートフォンや自動車のテレマティクスなど様々な応用領域で移動を伴う情報の発信や利用がデフォルトとなりつつあり、異なる場所で生成したデータを連携するための分散データベース技術、データセンター仮想化技術などは重要な研究領域となっている。データセンターの消費電力低減化も重要である。また、異なるステークホルダーの保有するデータを連携させることで新たな知識を抽出できる可能性が高まっており、このための ID 連携やプライバシー保護技術も重要な研究領域となっている。

2. ユーザー生成コンテンツとソーシャルメディア

SNS などに発信される情報は人々が社会の状況を主観的に捉えたソーシャルセンサー情報と考えることができる。これらは、専門家の手を経ずとも人々が活用できる情報源となり得ること、スマートフォンなどを介した Crowd-sourced sensing により空間的に固定された専用センサーの捕捉エリアを補える可能性があることなど、特に突発的な災害など非常時におけるセンサーデータ収集の手段として注目されつつある。しかし、こうした情報や知識の信頼性・信憑性に対する問題も浮上しており、センチメント分析、関連する言説との意味的關係の可視化、画像・映像、放送メディアなどとのクロスメディア分析などにより、信頼性・信憑性を評価・保証・検証する技術は重要な研究領域となっている。

3. センサーデータ統合・検索・分析技術

IoT に接続された多種・多量なセンサーや SNS などソーシャルセンサーからアプリケーション要求を満たすセンサーを検索し、必要な要件を満たすデータを収集するセンサーネットワークを動的に構築する技術は重要な研究領域である。また、収集されるセンサーのストリームデータを構造化し、特定の条件を満たすデータをイベントと

して検索する技術、複数のイベントが組み合わさった複雑な状況をリアルタイムに検出する複合イベント処理などは、現在盛んに研究開発が進められているが、今後はこうした処理を数値データのストリームだけでなく動画や SNS など異なる種類のデータに拡張する技術は重要になってくる。さらに、異分野のセンサーデータを統合して、より複雑な分析を行う試みが進んでいる。

4. 時空間データマイニング技術

Web 上の情報伝播や SNS におけるユーザー間のリンク構造の変化、実社会の車・人・モノの位置情報、地域や時間の情報とともに蓄積される医療データなどは、時空間データといえる。近年、時系列データマイニング、ストリームマイニング、グラフマイニングが特に注目されている。時系列データマイニングの応用として、ソーシャルネットワーク上の情報の拡散力や興味を持つ潜在的ユーザー数の推定、医療データの時系列解析による医療サービスの質の向上や効率化などがある。時系列データをリアルタイムに分析するストリームマイニングの応用として、センサーデータの解析の他、ネットワークセキュリティ、人や自動車など移動物体の監視、ソーシャルネットワークへの応用などがある。グラフマイニングとしては時間発展グラフデータへの適用が近年特に注目を集めている。

5. 次世代情報検索・推薦技術

モバイル環境における効果的な情報提供を実現するために、ユーザーの位置、運動状況、生体情報などのユーザーの状況や、ランドマーク、気温、湿度など周囲の状況などのコンテキストを把握し、情報要求の明確化、移動時やコンテキストが突然変化した時などに明示的な情報要求がない場合でも適切なタイミングで有用な情報をシステム側から能動的に提供する技術などが重要な研究課題となっている。また、必要な情報をコンパクトに適切な方式で提示したり、インタラクションを通じて必要な情報を明確化と確実な情報提供を行う技術が求められている。

6. 個人ライフログデータの記録・利活用技術

例えばロケーションログ、活動ログ、ビジュアルログ、食事の記録、読書の記録、睡眠の記録、スマートフォンのアプリ利用履歴などのユーザー自身の生活のありさまを蓄積・可視化することが主として行われている。現状では、健康、ヘルスケアに関連した情報提供などが直接的な活用領域である。多様なセンサーやアプリケーションにより計測されたデータを共有管理・利用できるプライバシーにも配慮したプラットフォーム構築やデータの標準化、魅力的なアプリケーションの開発、ウェアラブルデバイスを含む生体情報等のセンシング技術の開発などが重要な研究領域である。健康やヘルスケア以外に個人の体験を伝えるといった活用も重要になってくると考えられる。

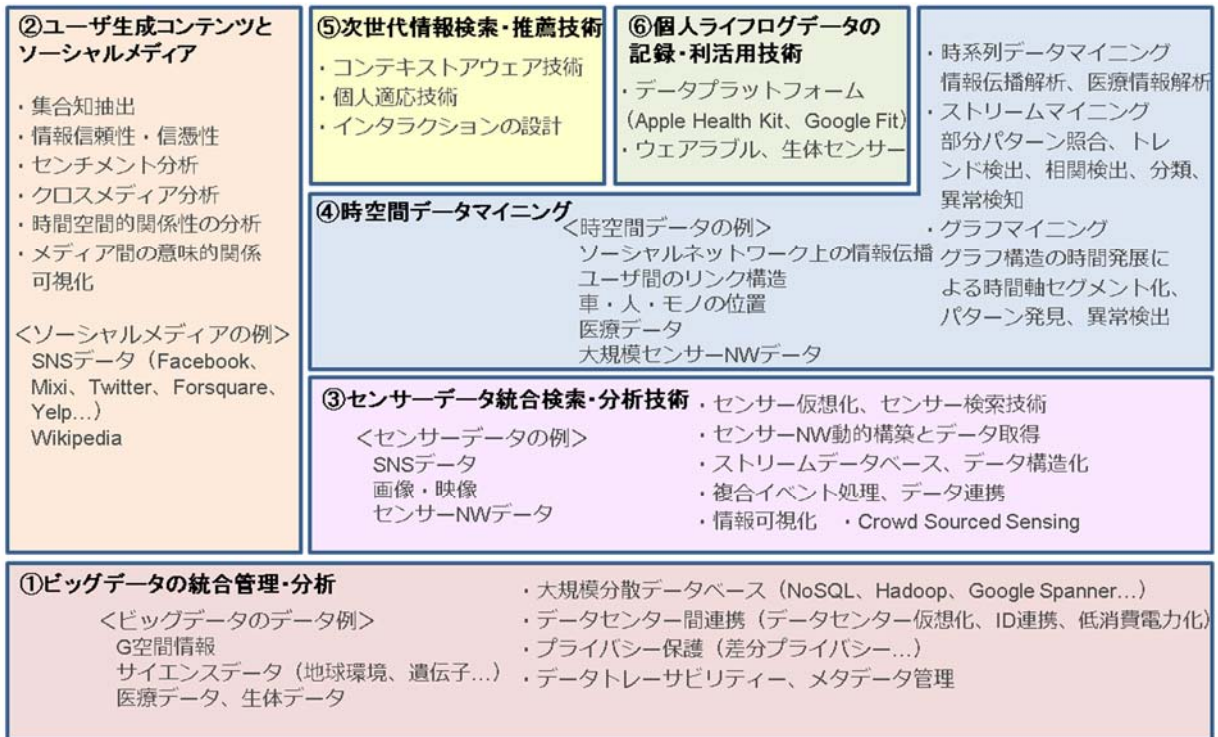


図 3. 6. 1 ITメディアとデータマネジメントの俯瞰図

3.6.1 ビッグデータの統合・管理・分析技術

(1) 研究開発領域名

ビッグデータの統合・管理・分析技術

(2) 研究開発領域の簡潔な説明

各種センサー技術、観測技術の発達に伴い自然、社会の広い領域において大規模データを取得できるようになってきている。ビッグデータ処理のための情報データ基盤として、複数データセンターの一括管理やその上での Hadoop の展開、地球規模のトランザクション処理を可能とする大規模分散データベースなどの研究開発がされている。各国ともビッグデータ応用のための政策を展開しているが、米国の民間企業および大学が圧倒的な研究開発力を有する。最近の動向としては、健康・医療ビッグデータの取り扱いが注目される。

(3) 研究開発領域の詳細な説明と国内外の動向

[対象ビッグデータの広がり]

対象となるビッグデータは、当初、購買、決済などのトランザクション履歴、Web アクセス履歴などが主要なものであった。しかし、現在ではこれらに加え、以下のような広い分野のビッグデータを対象とする応用開発が急速な発展を見せている。(1) 地理空間情報の整備が進められ、GPS データが広く利用可能になるとともに「G 空間情報」(人や物の位置情報とそれに関連づけられた情報)、(2) 地球環境データ、遺伝子データ、大規模実験データなどのサイエンスビッグデータ、(3) 電子カルテデータやレセプトデータなどの医療データ、およびウェアラブルセンサにより取得される生体データなどを含む健康データ。

[ビッグデータ処理のための情報システム基盤]

最近の重要な研究開発課題としては、複数データセンターに跨る情報システム基盤がある。データセンターの規模によるコストメリット、データ発生源の地理的分散化、対象データの大容量化などの理由により、複数のデータセンターを高速ネットワークなどで接続し一括運用する。SDN (Software Defined Network) などによるデータセンターの仮想化も活発に研究開発が進められている。データセンター運用コストにおいて大きい比重を占める電力消費低減化も重要な課題である¹⁾。

ビッグデータの情報システム基盤をデータベーストランザクション機能により大きく分類すると、伝統的なデータベース管理システムのような ACID (Atomicity, Consistency, Isolation, Durability) 特性を遵守しない、いわゆる NoSQL やバッチ解析処理のために広く使われている Hadoop のようなシステムと、ACID 特性に基づくトランザクション機能を提供するものに分けることができる。前者に関する複数データセンター化に関する研究としては、例えば Hadoop を複数データセンターに跨って実行するための性能改善 (予測に基づくジョブ局所化、再構成可能な Hadoop 分散ファイルシステム HDFS データ配置、データプリフェッチなどシステムソフトウェア層の改善²⁾) の研究開発を挙げることができる。一方、ACID 特性に基づくトランザクション機能を地球規模で提供する大規模分散データベースとしては、Google の Spanner³⁾ が注目される。従来、地球規模の分散システムでは時刻の同期を取ることは困難であるとされてきたが、Spanner では、各データセンターに設置した原

子時計と GPS データによりサーバー間の時刻同期を行い、地球規模のトランザクション実行を可能としている。これにより、一貫性を保ったままデータをデータセンター間で自動的に移動し、負荷分散や障害回復を実現している。

〔諸外国の政策と研究開発の動向〕

米国では、2012年にオバマ政権が Big Data Research and Development Initiative を発表し、NSF, NIH, DOD, DOE, DARPA, USGS の六つの連邦政府機関に対して 2 億ドルの研究開発資金を投資している。2013年にはこの Initiative から生まれた産学官のビッグデータパートナーシップによる共同研究プログラムが始まっている。その例としては、NASA と Amazon Web Services の協力による地球観測データの公開⁴⁾やノースカロライナ大学を始めとする州内の大学や IT 企業が参画するコンソーシアムなどがある。EU では、各国政府が研究開発を主導すべき分野を明確化した上でそれらの分野に重点投資を行う European Innovation Partnerships (EIPs) が、健康高齢社会、農業持続性と生産性、スマートシティーおよびコミュニティなどの分野を対象とし、これらの分野におけるビッグデータの応用研究が進められている。

（４）科学技術的・政策的課題

ビッグデータ管理においてプライバシー保護は重要な課題である。基礎研究としては、データベースに格納されているプライベートデータがどの程度特定され得るかを理論的に定量化しその定量化の基でプライベートデータが特定され得る確率をある閾値以下にする差分プライバシー⁵⁾が提案され活発に研究されている。しかし、現実の応用場面では閾値をどの程度に設定すべきかの指針がなく広く実用化されるには至っていない。一般にプライバシーデータの利活用とプライバシーの保護は相反するため適切な均衡点を見いだす必要があるが、技術の進展やプライバシー情報開示の受容性に対する個人差により議論の前提自体が固定的ではなく、継続的な見直しが必要となる。また、個人のデータを異なるプロバイダーが収集した場合にそれらを統合することにより利用価値は高まるが、一方では漏洩時の損失は大きくなる。プライバシー情報を保有する複数の組織が相互に必要なプライバシー情報のみを交換する ID 連携の技術開発と運用も必要になる。

また、我が国の場合は、新規技術開発当初から国際標準化を見据えることや、交通、医療・健康を始めとする分野において規制緩和により新規技術開発を促進することや特区の有効利用を行うことが重要である。

（５）注目動向（新たな知見や新技術の創出、大規模プロジェクトの動向など）

センサーの小型化、スマートフォンの普及などを背景とし、近年多くの IT 企業が健康・医療ビッグデータの収集、管理、分析に乗り出している。例えば、Apple は HealthKit と呼ぶモバイルヘルスプラットフォームを iOS8 に搭載し、心拍センサーを持つ Apple Watch のようなウェアラブル機器から収集する生体情報を一元管理できる環境を整えている。Google も同様に、各種センサーやサードパーティーのアプリで計測した健康データを共有管理できるプラットフォーム Google Fit をリリースした。Google は、ベンチャー企業

MC10 と共同で、直接肌に貼ることのできる伸縮する電子回路であるバイオセンサーBio-Stamp を開発し、脈拍数、体温、紫外線吸収量、脳の活動などをモニターすることを構想している。また、無線チップとセンサーを搭載する医療用スマートコンタクトレンズを開発し涙の成分から血糖値の変化を検出する構想を発表している。さらに、Baseline Study プロジェクトでは、多数の協力者から生体データを取得、解析することにより人間の健康状態の基準値を定義し、そこからの逸脱をいち早く検出することにより病気の兆候を早期発見することなどを構想している。また、個人向け遺伝子解析は、Google や Yahoo!JAPAN など が手掛けている。

また、IBM は、クイズ番組で人間のチャンピオンに勝ったことで有名になった Watson を医療分野向けにチューニングし、医療データから心臓病患者を予測分析する取り組みを行っている⁶⁾。

このような動向から、健康・医療ビッグデータの収集、管理、分析は今後規模の大きい産業へと発展する可能性がある。自動車保険の場合は、カーナビなどの車載機器から得られた情報を基に安全運転の度合いに応じて保険料が変動するテレマティクス保険が英米で普及し始めているが、将来的には健康データを利用し同様の考えに基づく生命保険や健康保険が普及する可能性も考えられる。しかし、遺伝子情報は究極の個人情報と言われ、健康・医療情報と共に機微情報であり、プライバシー情報の十分な保護が前提となることは言うまでもない。

（6）キーワード

ビッグデータ、Hadoop、データマイニング、データセンター、分散データベース、健康・医療データ

（7）国際比較

| 国・地域 | フェーズ | 現状 | トレンド | 各国の状況、評価の際に参考にした根拠など |
|------|---------|----|------|--|
| 日本 | 基礎研究 | ○ | → | グラフアルゴリズムやソーシャルネット系データ解析など一部の分野では研究が進展しているが、大規模データ管理の基盤技術などは米国に比べ研究力が見劣りする。 |
| | 応用研究・開発 | ○ | → | 人口規模、経済力の面から多様な応用分野の研究を同時並行的に進める研究力を持つが、米国に比べた場合、大規模データ管理基盤技術に基づく応用研究が脆弱である。 |
| | 産業化 | ○ | ↗ | 技術展開力はあるが、交通や健康・医療などの分野における規制が産業化の足枷になっている場合もある。 |
| 米国 | 基礎研究 | ◎ | → | 大学や民間企業研究所による研究は質、量ともに群を抜いており、この傾向は続く予想される。 |
| | 応用研究・開発 | ◎ | → | 産官学が多様な連携の元でプロジェクトを立ち上げている。特にIT系企業の豊富な資金力を背景にした中長期を見据えた研究に厚みがある。 |
| | 産業化 | ◎ | → | アイデアが生まれてから産業化までの時間が短い。政府による規制が少ないことも一因であると考えられる。 |
| 欧州 | 基礎研究 | ○ | → | セマンティックウェブやそれをWikipediaに応用したdbpedia、LOD (Linked Open Data)などの研究では世界をけん引している。 |
| | 応用研究・開発 | ○ | → | イノベーション・ユニオン(Innovation Union)および欧州イノベーション・パートナーシップ(European Innovation Partnership) ⁷⁾ の枠組みの下で研究機関や企業などの共同研究が計られている。 |
| | 産業化 | ○ | → | P2P型音楽配信サービスであるSpotifyなど特徴のある商用サービスが生まれている。大規模製造業では、IoT (Internet of Things)の概念を融合し、部品、製造過程、製品のデータをすべて統合管理する高度化が推進されている。 |
| 中国 | 基礎研究 | ◎ | ↗ | 大学、企業研究所による研究が国際的にも顕著であり、この傾向は続くことが予想される。 |
| | 応用研究・開発 | ○ | ↗ | 国家レベルで工業分野におけるビッグデータの利活用が推進されている。また、研究所や企業と日本など外国企業との共同研究もいくつかの事例が見られる。 |
| | 産業化 | ○ | ↗ | 国内ネット企業が人口のスケールメリットを背景に業績を拡大している。ビッグデータ解析が各種産業に浸透するにはしばらく時間を要すると考えられる。 |
| 韓国 | 基礎研究 | △ | ↗ | 一部の大学等では国際レベルの研究が行われている。米国大学での博士号取得者が帰国後大学教員となっている事例が増えており、研究の質が向上すると予想される。 |
| | 応用研究・開発 | ○ | ↗ | 政府の産業通商資源部が国民の健康管理サービスを強化し、スマートヘルスケア産業の輸出を拡大するための支援を行うなど、政府がいくつかの事業に対し重点的に政策的な取り組みを進めている。 |
| | 産業化 | △ | → | 中小企業およびソフトウェア産業育成の基盤が脆弱である。 |

（註1）フェーズ

基礎研究フェーズ：大学・国研などでの基礎研究のレベル

応用研究・開発フェーズ：研究・技術開発（プロトタイプの開発含む）のレベル

産業化フェーズ：量産技術・製品展開力のレベル

（註2）現状

※我が国の現状を基準にした相対評価ではなく、絶対評価である。

◎：他国に比べて顕著な活動・成果が見えている、○：ある程度の活動・成果が見えている、

△：他国に比べて顕著な活動・成果が見えていない、×：特筆すべき活動・成果が見えていない

（註3）トレンド

↗：上昇傾向、→：現状維持、↘：下降傾向

（8）引用資料

- 1) K. Rajamani, C. Lefurgy, S. Ghiasi, J. Rubio, H. Hanson, and T. Keller, "Power management solutions for computer systems and datacenters." In Low Power Electronics and Design (ISLPED'08), ACM/IEEE International Symposium on, Bangalore, India, 2008, pp. 135-136.
- 2) Qi Zhang, Ling Liu, Kisung Lee, Yang Zhou, Aameek Singh, Nagapramod Mandagere, Sandeep Gopisetty, Gabriel Alatorre, "Improving Hadoop Service Provisioning in A Geographically Distributed Cloud," 2014 IEEE Seventh International Conference on Cloud Computing (CLOUD 2014).
- 3) Corbett, James C; Dean, Jeffrey; Epstein, Michael; Fikes, Andrew; Frost, Christopher; Furman, JJ; Ghemawat, Sanjay; Gubarev, Andrey; Heiser, Christopher; Hochschild, Peter; Hsieh, Wilson; Kanthak, Sebastian; Kogan, Eugene; Li, Hongyi; Lloyd, Alexander; Melnik, Sergey; Mwaura, David; Nagle, David; Quinlan, Sean; Rao, Rajesh; Rolig, Lindsay; Saito, Yasushi; Szymaniak, Michal; Taylor, Christopher; Wang, Ruth; Woodford, Dale, "Spanner: Google's Globally-Distributed Database", Proceedings of OSDI 2012.
- 4) NASA NEX, <http://aws.amazon.com/jp/nasa/nex/>
- 5) Cynthia Dwork, "Differential Privacy,"
<http://research.microsoft.com/pubs/64346/dwork.pdf>
<http://www-03.ibm.com/press/us/en/pressrelease/43231.wss>
- 6) European Innovation Partnerships
http://ec.europa.eu/research/innovation-union/index_en.cfm?pg=eip
- 7) 平成 26 年度版 情報通信白書
<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h26/pdf/n3100000.pdf>
- 8) 特許庁 平成 25 年度特許出願技術動向調査報告書（概要）ビッグデータ分析技術，平成 26 年 2 月

この分野の重要な会議としては以下がある。

- IEEE International Conference on Big Data
- IEEE International Conference on Data Engineering (ICDE)
- IEEE 13th International Conference on Data Mining (ICDM)
- ACM International Conference on Research and Development in Information Retrieval (SIGIR)
- ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)
- ACM International Conference on Management of Data (SIGMOD)
- International Conference on Very Large Data Bases (VLDB)
- International World Wide Web Conference (WWW)

3.6.2 ユーザー生成コンテンツとソーシャルメディア

(1) 研究開発領域名

ユーザー生成コンテンツとソーシャルメディア

(2) 研究開発領域の簡潔な説明

ユーザー個人から発信されるデータとそれら個人ならびにデータに興味や関連あるユーザー同士のソーシャルネットワークを構成し、情報を共有・拡散配信することで新たな知識を創出する基礎・応用技術に関する研究開発

(3) 研究開発領域の詳細な説明と国内外の動向

ユーザー個人が配信するメディアは Consumer Generated Media (CGM) と呼ばれ、不特定多数あるいは限られたコミュニティ内のメンバが情報（テキスト、画像、音声、動画）へアクセスできる。CGM の特徴として、ユーザーの匿名性の低さがあげられ、そのため匿名性の高い掲示板サイトと比較すると、情報の質が高く、信頼性が高い。また、2000 年初頭に創業された Facebook（米国）や Mixi（日本）は、ユーザー同士の関係性からコミュニティを形成するソーシャル・ネットワーキング・サービス（SNS）を特徴としており、信頼性だけでなく、速報性も高い。これら信頼性、速報性の高い情報共有が可能である SNS は、2014 年には世界の 18.2 億人が利用しており、今後 2017 年までには 23.2 億人になると予想されており、特に SNS の代表的なものとして近年台頭している Twitter と Facebook の利用者は世界で約 14 億人、2009 年からの年平均成長率は Twitter が 70.5%、Facebook が 35.9% となっており¹⁾、情報通信のグローバル市場でも国際競争力の強化が進んでいる。また、2009 年以降には、Forsquare（米国）や Yelp（米国）にみられる位置情報に基づく SNS が普及し、ユーザーが情報を posting（送信）すると場所情報（緯度経度や店舗名等）がメタデータとして付与され（Forsquare ではチェックインと呼ばれる）、推薦や検索技術の研究開発の重要なプラットフォームとなっている。2014 年には Forsquare のチェックイン機能と特定の場所に関する情報検索機能とを分離し、チェックイン機能専用のアプリケーションである Swam が提供された。

Wikipedia に代表される一般ユーザーが編集可能なソーシャルメディアは、機械的に抽出・集約されたデータとは異なり、人手により編集されて知識ベース化された集合知である。この知識ベースには極めて大量で幅広い分野のデータが蓄積されており、これらのデータを基にして、言語処理、情報検索、あるいはデータマイニングなどの手法により、より高度な知識の抽出や集約を行うアプローチが世界的に数多く提案されている。また、これらの知識ベースやその利活用が活発化する一方、それらに含まれる情報や知識に対する信頼性・信憑性に関する問題も浮上している。情報の信頼性や信憑性、あるいはデータ品質に関する研究はネットワークなどの分野をはじめこれまでに多くのアプローチが提案されてきた。しかしながら、与えられた知識ベースからコンテンツの内容や知識そのものに対する判断を機械的に導くための手法は、わが国では総務省の情報信憑性検証技術プロジェクト²⁾ など極めて数が少ない。

テレビや新聞といったマスメディアと異なり、ソーシャルメディアは参加者同士の双方向通信が可能である。近年では、ソーシャルメディアとマスメディアが融合したメディア横断

的知識共有が行われているが、マスメディアは公共性、信頼性の高い情報発信を行わなければならないため、ソーシャルメディアの個人から発信される情報の信頼性判定やノイズ判定が課題となっている。これらについては、NHK や民放テレビ局でいくつかの取り組みがある。

（４）科学技術的・政策的課題

- ・知識ベースとしての集合知の収集・抽出技術については近年数多くの研究プロジェクトが遂行されているが、情報の信頼性・信憑性を保証する技術や検証のための手法などの研究は緒に就いたばかりであり、今後重点的に取り組むべき課題のひとつである。
- ・マスメディアにより Twitter 等のソーシャルメディアを利用したメディア横断的知識共有が行われているが、ソーシャルメディアの個人から発信される情報の信頼性判定やノイズ判定が課題となっている。
- ・位置依存情報や地理情報サービスに関する技術については、他の主要国と比較して先進的な研究やサービスが多いが、映像メディアや音声メディアなどの他メディアと連携・融合したクロスメディア環境における先行技術開発が我が国の優位性を保つために必要である。
- ・ソーシャルデータ解析、集合知分析を含む当該分野では、他の主要国と比較して研究者の絶対数が少なく、国際トップコミュニティで活躍している研究者は非常に少ない。トップレベルの人材育成が急務である。また、本分野の国際標準化に寄与できる技術者を育成することはわが国の国際競争力を高める上でも急務である。
- ・我が国ではソーシャルメディアを含むビッグデータに対する集合知分析、知識抽出や信頼性・信憑性向上に対する戦略的なファンディングが他の主要国と比較して十分ではなく、大規模なプロジェクトを推進できていない。
- ・データベース、ビジョン、センサーネットワーク、モバイルコンピューティング分野でデータ取得から分析に至るまでそれぞれが独立して研究を行っており、また、多種多様なデータや分析結果は個々の研究者により独立して管理されており、分野を横断して連携した研究開発を進めることが課題である。

（５）注目動向（新たな知見や新技術の創出、大規模プロジェクトの動向など）

- ・iPhone やアンドロイド端末等のモバイル端末には、GPS や加速度等の各種センサーが搭載されており、ユーザーがソーシャルデータを配信する際にそれらセンサーデータをメタデータとして取得・付与することができ、モバイル端末の各種アプリケーションを開発するためのプラットフォームとして急速に整備されている。
- ・ソーシャルメディアに関する研究開発としては、学会活動では SocialCom (IEEE International Conference on Social Computing) や SocInfo (International Conference on Social Informatics) などの国際会議が開催されている。
- ・コンテンツやサービスの信頼性・信憑性を保証する研究開発としては、総務省の「電気通信サービスにおける情報信憑性検証技術等に関する研究開発」(2007-2011) などの研究プロジェクトなどが挙げられる。学会活動としては、WICOW (Workshop on Information Credibility on the Web)³⁾ や World Wide Web Conference における WebQuality Workshop⁴⁾、IEEE TrustCom (International Conference on Trust, Security and Privacy in

Computing and Communications)などが開催されている。

(6) キーワード

ソーシャル・ネットワーキング・サービス、集合知、Facebook、Twitter、Consumer Generated Media

(7) 国際比較

| 国・地域 | フェーズ | 現状 | トレンド | 各国の状況、評価の際に参考にした根拠など |
|------|---------|----|------|--|
| 日本 | 基礎研究 | ○ | ↑ | <ul style="list-style-type: none"> データマイニング、機械学習に関する基礎研究は長年取り組まれており、ソーシャルメディアという新たなユーザー生成コンテンツに対する取り組みが、専門の学会（情報処理学会⁵⁾⁶⁾⁷⁾、電子情報通信学会⁸⁾⁹⁾¹⁰⁾、人工知能学会¹¹⁾などで行われている。 国内の研究会では、情報検索（情報基礎とアクセス技術⁵⁾、データベースシステム研究会⁶⁾）、自然言語解析（自然言語処理研究会⁷⁾）、人工知能（人工知能と知識処理研究会⁸⁾）、画像処理（パターン認識・メディア理解研究会⁹⁾）の分野で多くの発表¹²⁾¹³⁾が行われているが、主要ジャーナルや国際会議に貢献できる研究成果は、米国や欧州に比べるとまだまだ少ない。 |
| | 応用研究・開発 | △ | ↑ | <ul style="list-style-type: none"> FacebookやTwitter、Forsquare等のソーシャル・ネットワーキング・サービスのデータ解析から知識創出まで取り組まれているが、海外の主要ジャーナルやトップカンファレンスでのプレゼンスは、米国や欧州に比べると非常に低い。 データベース、ビジョン、センサーネットワーク、モバイルコンピューティング分野を横断した研究開発の連携がなされていない。 ソーシャルデータ解析、集合知分析を含む当該分野では、他の主要国と比較して研究者の絶対数が少なく、国際トップコミュニティで活躍している研究者は非常に少ない。トップレベルの人材育成が急務である。 |
| | 産業化 | △ | ↑ | <ul style="list-style-type: none"> Mixiなど世界展開を行っているサービスもあるが、主要国のサービスの利用者数を比較すると評価が低い。 ソーシャル・ネットワーキングのサービスの基盤にはユーザー生成コンテンツ分析技術が必要であり、つまりサービスの規模がデータの質に関わるが、主要国のサービスと比較して規模が小さい。 |
| 米国 | 基礎研究 | ◎ | ↑ | <ul style="list-style-type: none"> UC Berkeley、University of Illinoisなどの大学におけるソーシャルメディアのデータマイニング、機械学習、ソーシャルネットワーク分析の基礎研究は、WWW¹⁴⁾、ICDE¹⁵⁾、KDD¹⁶⁾、WSDM¹⁷⁾、CHI¹⁸⁾のトップカンファレンスでのプレゼンスが高く、研究成果はトップレベルである。 |
| | 応用研究・開発 | ◎ | ↑ | <ul style="list-style-type: none"> NISTなど公的機関、Stanford、UC Berkeleyなどの大学、Google、Yahoo!Labs、IBMなどの企業との産官学連携が非常に強く、基礎研究から応用研究、さらにサービスへの展開までのモデルが構築されている。 |
| | 産業化 | ◎ | ↑ | <ul style="list-style-type: none"> Google、Yahoo!Labs、Twitter、Facebook、Microsoftなどシリコンバレーを中心とした企業では、優秀なエンジニアを世界各国から集め、多様なサービスを展開しており、世界規模のサービスとなっている。 若手の優秀なエンジニアを世界各国から集め、開発だけでなく技術研究にも取り組まされており、人材育成の面でも高く評価できる。 |
| 欧州 | 基礎研究 | ◎ | ↑ | <ul style="list-style-type: none"> ソーシャルネットワーク分析、ソーシャルメディアのデータマイニングなど、Hannover大学、Amsterdam大学における基礎研究は、トップカンファレンスであるECML/PKDD2014¹⁹⁾やSIGIR2014²⁰⁾のsocial mediaセッションでプレゼンスが高く、研究成果はトップレベルである。 |
| | 応用研究・開発 | ◎ | ↑ | <ul style="list-style-type: none"> INRIA、CNRS、DFKI、DERIなど公的機関、EC各国の大学、IBM Research、Google、Yahoo!Labsなどの企業との産官学連携がEU各国内外で形成されており、基礎研究から応用研究への展開が成されている。 EU圏内での人材交流が活発になされており、人材の育成の面でも高く評価できる。 |
| | 産業化 | ○ | ↑ | <ul style="list-style-type: none"> IBM Research、Yahoo!Labs、Microsoftなどで主要ジャーナル、トップカンファレンスでのプレゼンスは高いが、米国と比較すると世界規模のサービスの構築が成されているとはいえない。 EU圏内で若手の優秀なエンジニアを集め、技術研究や開発に取り組んでおり、EU各国間での人材交流が活発になされており、人材の育成の面でも高く評価できる。 |

| | | | | |
|----|---------|---|---|--|
| 中国 | 基礎研究 | ○ | ↑ | ・清華大学、中国科学院、香港大学、香港技術大学等では、ソーシャルメディアのユーザー生成コンテンツに対してデータベース、機械学習、ネットワーク、ユーザーインターフェースに関する基礎研究に取り組んでおり、VLDB ²¹⁾ 、CIKM ²²⁾ 、PAKDD ²³⁾ 、ICDM ²⁴⁾ 等のトップカンファレンスにおいて高いプレゼンスを保っている。 |
| | 応用研究・開発 | ◎ | ↑ | ・Microsoft Research Asia (MSRA) は清華大学や中国科学院、香港大学等の国内の優秀な学生や国外の優秀な研究者が多く、主要ジャーナルやトップカンファレンスにおいてもプレゼンスが高い。 |
| | 産業化 | △ | ↑ | ・中国国内では、Facebook等の利用規制のため、騰訊が開発した微信（別名：WeChat）が国内の代表的サービスとして普及している。 ・MSRAでの応用研究のプレゼンスは高いが、国内外のサービスへの制限が顕著なため、応用研究がサービスとして十分な活用がされていない。 |
| 韓国 | 基礎研究 | ○ | ↑ | ・Korea Advanced Institute of Science and Technology (KAIST) では、ソーシャルメディアのユーザー生成コンテンツに関するデータベース、機械学習によるデータマイニングの基礎研究に取り組んでおり、ICDE ¹⁵⁾ 、CIKM ²²⁾ 等のトップカンファレンスにおいて高いプレゼンスを持っている。 |
| | 応用研究・開発 | △ | ↑ | ・Korea Institute of Science and Technology Information (KISTI) では、SNS等のビッグデータの応用研究が推進されている。 |
| | 産業化 | △ | ↑ | ・NAVERなど世界展開を行っているサービスもあるが、主要国のサービスの利用者数を比較すると評価が低い。 |

(註1) フェーズ

基礎研究フェーズ：大学・国研などでの基礎研究のレベル

応用研究・開発フェーズ：研究・技術開発（プロトタイプの開発含む）のレベル

産業化フェーズ：量産技術・製品展開力のレベル

(註2) 現状

※我が国の現状を基準にした相対評価ではなく、絶対評価である。

◎：他国に比べて顕著な活動・成果が見えている、○：ある程度の活動・成果が見えている、

△：他国に比べて顕著な活動・成果が見えていない、×：特筆すべき活動・成果が見えていない

(註3) トレンド

↑：上昇傾向、→：現状維持、↓：下降傾向

(8) 引用資料

- 1) 総務省 平成26年版 情報通信白書（第1部、第1節：全世界でのICTの急速な浸透）
- 2) 総務省 情報の巨大集積化とその活用に向けた基盤技術の開発（情報集積活用基盤技術）「電気通信サービスに関する情報信憑性検証技術等に関する研究開発」
- 3) Proceedings of the 6th International Conference on Social Informatics (SocInfo 2014)
- 4) Proceedings of the 4th Joint WICOW/AIRWeb Workshop on Web Quality (WebQuality 2014)
- 5) 情報処理学会 情報基礎とアクセス技術研究会：<http://ipsj-ifat.org/>
- 6) 情報処理学会 データベースシステム研究会：<http://www.ipsj.or.jp/katsudou/sig/sighp/dbs/>
- 7) 情報処理学会 自然言語処理研究会：<http://www.nl-ipsj.or.jp/>
- 8) 電子情報通信学会 人工知能と知識処理研究専門委員会：<http://www.ieice.org/iss/ai/jpn/>
- 9) 電子情報通信学会 パターン認識・メディア理解研究会：<http://www.ieice.org/~prmu/jpn/>
- 10) 電子情報通信学会 データ工学研究専門委員会：<http://www.ieice.org/iss/de/jpn/>
- 11) 人工知能学会：<http://www.ai-gakkai.or.jp/>
- 12) 第6回 集合知シンポジウム
- 13) 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
- 14) Proceedings of the 23rd international conference on World wide web (WWW2014)

- 15) Proceedings of the 30th IEEE International Conference on Data Engineering (ICDE2014)
- 16) Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (SIGKDD2014)
- 17) Proceedings of the 7th ACM international conference on Web search and data mining (WSDM2014)
- 18) Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI2014)
- 19) Proceedings of the 2014 European conference on Machine learning and knowledge discovery in databases (ECML/PKDD2014)
- 20) Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval (SIGIR2014)
- 21) Proceedings of the 40th International Conference on Very Large Data Bases (VLDB2014)
- 22) Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (CIKM2014)
- 23) Proceedings of the 18th Pacific-Asia conference on Advances in knowledge discovery and data mining (PAKDD2014)
- 24) Proceedings of the IEEE International Conference on Data Mining series (ICDM2014)

3.6.3 センサーデータ統合検索分析技術

(1) 研究開発領域名

センサーデータ統合検索分析技術

(2) 研究開発領域の簡潔な説明

実世界に埋め込まれたセンサーなどから生み出されるバイナリデータから、人間によって生み出される SNS のテキストデータ等の異種・異分野のデータの収集・管理・分析をシームレスに実施できる統合的な分析技術の研究開発

(3) 研究開発領域の詳細な説明と国内外の動向

20 世紀においては、地球科学などの科学分野における研究や企業における製品の開発や生産設備の自動制御のために、実環境の状況を計測する目的でセンサーが設置され、ある特定の現象を捉えるために計測・分析されてきた。21 世紀になり、ユビキタス技術が社会に浸透するに従い、実世界の隅々までネットワークとセンサーが配置される事によって、多様な情報がネットワークを通じて計測・収集・分析できる環境が整備されつつある。この技術動向は、ネットワーク基盤技術に近い分野では、データを計測する大量のセンサーと、そのデータを処理する大規模計算機を結び合わせる M2M(Machine to Machine)技術や、実世界のモノをネットワークに接続させるモノのインターネット技術(IoT: Internet of Things)として見られ、情報処理分野においてはビッグデータ処理技術の発展として見られる。

その一方で、実世界の状況を分析する新たなデータ収集対象も生まれている。それが、Twitter や Facebook 等のソーシャルデータや Web サービスログ等である。このような SNS から発信される情報は、社会の活動を各個人が観察して発信されたものであり、それらはソーシャルセンサーと呼ばれる新たなセンサーからのデータである。

これまでのセンサーデータは、実世界の物理現象を計測した数値データであったが、新たなソーシャルデータは人間の知識や要求、活動が現れている。前者はセンサーの計測誤差を除けば、正確に実世界の物理現象を捉えた数値データであるが専門家の手を経なければ一般の人々が活用できる情報源とはなり得ない。一方、SNS 等の情報は一般の人々でも理解しやすい情報源であるが、個人の主観が含まれているため、客観性がなく信頼性が低い場合がある。これらは性質が全く正反対のデータであるが、それ故、両者を統合的に検索・分析する事ができる様になれば、データから新たな価値が生まれる可能性が高い。センサー統合検索分析技術は、まさにこの新たな価値創造を作り出すための技術であり、大別すると以下の 3 つの研究領域が存在する。

超多様なセンサーデータの発見と収集技術

IoT 技術等によって接続された大量のセンサーや、Web や分散データベース等ネットワークに接続された大量の情報源を探索し、さらに必要なデータを検索して選択的かつ透過的に種別の異なるデータを収集するための技術である。気温や湿度などの単純な環境計測データから、マイクロブログ上の特定のハッシュタグやユーザーアカウントに紐付いたテキストデータまで透過的にアクセスし統一的に収集・管理するために、各種センサーの機能や能力、提供するデータの属性や特徴量などをセマンティックに記述しソフトウェアサービスを介して

統一的にアクセスするセンサー仮想化技術や、SDN(Software Defined Network)等に基づき既存のセンサーネットワーク資源を有効活用し様々なアプリケーション要求を満たすセンサーネットワークを動的に構築する技術、どこからでもセンサーデータを蓄積し共有できるクラウドサービスを提供したりする技術等が注目されている。さらに、モバイルセンサーやスマートフォンの普及により、人々の手によって収集したデータを使って物理センサーを効率的かつ効果的に補完する Crowd-sourced sensing⁸⁾ も、特に突発的な災害など非常時におけるセンサーデータ収集の手段として注目されつつある。

異種センサーデータの構造化と検索技術

実世界の物理情報を取得するセンサーは、人手を介さずにデータを定期的に生産する。また、その多くが数値バイナリデータであるため、サイズの小さな数バイトのデータが大量に生み出される。このようなデータは個別に取得される場合よりも一定期間のまとまったストリームデータとして活用される事が多い。こうしたストリームデータから時々刻々と変化する特徴量を抽出しオンライン処理でデータを構造化し、継続的に問い合わせを行うことにより、膨大なストリームデータから必要な部分のデータのみをリアルタイムに検索するストリームデータベースの研究開発が盛んに行われている⁹⁾。さらに、こうしたデータストリームから特定の条件を満たすデータをイベントとして検出するとともに、複数のイベントを組み合わせて複雑な状況をリアルタイム性高く検出する複合イベント処理(complex event processing, CEP)技術の研究開発も盛んに行われている¹⁰⁾。一方、これらの技術を、物理センサーから発信される数値データのストリームだけでなく、監視カメラの動画像や、SNS から発信されるテキストや写真など、異なる種類のデータストリームにも拡張する技術も次第に開発され始めており、今後は、こうした異種センサーデータストリームから検出された様々なイベントを横断的に処理し、より広い分野にまたがる複雑なイベントをリアルタイムに発見できるようにする技術の必要性が高まってくる。

異分野センサーデータの統合と分析技術

従来、気象や機械、自動車、交通、健康、株価など、各分野の応用ごとに決められたセンサーデータの分析を高いパフォーマンスとスケーラビリティで行うことに重点が置かれてきたが、様々な種類のセンサーデータが利活用できる環境が広まるとともに、分野ごとのセンサーデータを統合し分析を連携させ、より総合的で複雑な分析を行おうという試みが進められている。例えば、スマートシティでは、環境や交通、治安など街の中で起こる様々な状況を網羅的にセンシングし、公共サービスの効率化や非常時への迅速な対応を実現することを目指している。また、スマートヘルスケアでは、人々の健康状態だけでなく、日常の行動生活環境までセンシングし、継続的な健康管理を行おうとしている。さらに、近年では、スマートファクトリーに代表されるように、製造機械のセンサーデータ分析だけでなく、作業員の行動や物流の状態まで幅広いセンサーデータを横断的に統合分析し社会全体での生産性向上を目指している。こうした中で、各分野のセンサーデータに特化した分析だけではなく、それらを横断的に統合し、分野をまたがる相関性を網羅性高く発見する分析手法の研究開発が今後ますます重要になってくる。従来の蓄積データを対象としたデータウェアハウス構築技術や多次元データ分析技術をセンサーデータに応用発展させる研究や、異分野センサーデ

ータの相関関係を試行錯誤的で直感的に分析する情報可視化技術、さらにこれらをリアルタイムに処理する技術などが必要になってくる。さらに、近年では、異種センサーデータやスマートフォンを組み合わせ、人々の行動や群衆の動きをリアルタイム性高く分析し各種のスマートサービスをより高度化するための研究開発も注目されつつある。

（４）科学技術的・政策的課題

センサーデータ統合技術の技術的なボトルネックとしては、計測の精度も頻度も異なるデータをいかに正確かつリアルタイムに取得するかの問題や、数値からテキストまで多種多様なデータフォーマットをいかに利活用に適した形式に変換し相互に連携させるか、また、ネットワーク上に遍在する各種データやサービスをいかに発見し統合するか等が挙げられる。また、個別のデータには含まれないものの、複数のデータを統合する事によって見えてくる個人情報管理などが課題となる。監視カメラや携帯電話、ソーシャルネットワークサービス等で得られた利用者の行動情報などを活用することは、社会の動きを明らかにするソーシャルセンサーとしての期待が大きい反面、個人のプライバシー情報にも深く関係するため、我が国では利用許諾や利用者へのインセンティブ提供等のガイドライン等が確立していない。

（５）注目動向（新たな知見や新技術の創出、大規模プロジェクトの動向など）

<注目動向>

センサーデータ統合検索分析技術は、実世界に埋め込まれたセンサーデータや企業における販売データ、携帯端末などによる利用者の行動履歴などのデータを統合的に活用して社会の動きを明らかにすることが目的であり、昨今では、いわゆる CPS(Cyber Physical Systems)として、企業を始め大学・研究機関などで広く研究が行われている。また、当該技術は内閣府の科学技術イノベーション総合戦略 2014¹⁾において記述されている「世界に先駆け次世代インフラの構築」等における中核技術として注目を集めている。

具体的な我が国の大規模プロジェクトの動向としては、総務省が独立行政法人情報通信研究機構（NICT）に M2M 技術の研究開発を推進するためにモバイルワイヤレステストベッド²⁾を整備した上で、ソーシャルビッグデータの利活用・基盤技術の研究開発プロジェクトを推進している。このプロジェクトにおいては、NICT のみならず公募にて 22 件の研究課題が採択され、医療・保健分野、防災・減災分野、観光分野、交通・運輸分野、農業分野、教育分野、インフラ保全分野、スマートシティ分野等の課題について研究が進められている。

センサーネットワークにおいては、スマートシティ等に必要不可欠なスマートメーターの通信インフラとして WiSUN 規格³⁾が注目を集めており、様々なセンサーネットワークが構築されつつある。WiSUN 規格は IEEE802.15.4g/4e や IPv6 にも対応しており、広く普及していくことが予想されている。

また、空間と通信技術を融合させる目的で G 空間プロジェクト⁴⁾が推進されている。このプロジェクトにおいては、「G 空間オープンデータプラットフォームの構築」や「G 空間防災システムの構築」、「G 空間シティの構築」などが目標としてあげられており、位置情報を核として、各種センサーデータの統合利用を可能とする事を目指している。平成 26 年度には、内閣府のアクションプランにも「G 空間プラットフォームの高度化に関する研究開発」

⁵⁾ が挙げられている。さらに科学技術振興機構（JST）では、社会から生み出されるビッグデータを統合利活用するための「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化プロジェクト」⁶⁾や「科学的発見・社会的課題解決に向けた各分野のビッグデータ利活用推進のための次世代アプリケーション技術の創出・高度化プロジェクト」⁷⁾が推進されている。

（6）キーワード

M2M、Cyber-Physical Systems (CPS)、SNS、WiSUN、センサーネットワーク、Semantic Sensor Network、ソーシャルセンシングデータ、スマートヘルスケア、センサーデータ統合、異種・異分野データ統合、センサーデータウェアハウス、スマートシティ、センサークラウド、Internet of Things (IoT)、Participatory Sensing、Crowd-sourced Sensing

（7）国際比較

| 国・地域 | フェーズ | 現状 | トレンド | 各国の状況、評価の際に参考にした根拠など |
|------|---------|----|------|---|
| 日本 | 基礎研究 | ○ | ↑ | <ul style="list-style-type: none"> ・センサーネットワークやセンサーデータベース、モバイル情報システム等の研究開発が個別に行われている。特に、自動車や工場、ロボットなどを対象にしたセンサーデータのリアルタイム処理や機械学習などで、世界的に高い水準の基礎研究が行われている。また、異種・異分野のセンサーデータの統合や可視化などでも、世界に先駆けた基礎研究が行われている。 ・研究の競争力を強化するための共通基盤（テストベッドやデータセット）の整備や産官学連携が主要国と比べ遅れており、日本から国際的にインパクトのある研究成果として世界最高峰の国際会議、論文誌に公表される論文数が少ない。 |
| | 応用研究・開発 | ○ | ↑ | <ul style="list-style-type: none"> ・情報通信研究機構のモバイルワイヤレステストベッドや各省庁の委託研究・研究助成などを利用して、実証実験を目的とした応用研究・開発のプロジェクトが活発化してきている。 ・主要国と比較すると小規模なプロジェクトが個別に行われ、依然国際的な認知度を十分に得るに至っていない。国家的戦略の下で、これらを有機的に連携させることが必要不可欠である。 |
| | 産業化 | ○ | ↑ | <ul style="list-style-type: none"> ・フィールドセンサーや、組込みセンサー、モバイル端末、スマートフォンなど、センサー技術はあらゆる分野において世界最高水準である。 ・国内の隅々にまで普及したモバイル通信網、高機能なカーナビの普及、交通系ICカードの普及などを背景に、人や車の動きに関する膨大かつ高精細なセンサーデータを収集・分析し、高度交通システムや人口統計、モバイルサービスへの応用が活発化してきている。 ・産業界から得られる各種の膨大なセンサーデータは、依然各分野や企業内で閉じた利活用にとどまり、欧米のように社会インフラの共有財産（ソーシャルビッグデータ）として横断的に利活用しスマートシティーなど革新的な産業を生み出すまでには至っていない。オープンデータ化などへの一層の取り組みが必要である。 |
| 米国 | 基礎研究 | ◎ | ↑ | <ul style="list-style-type: none"> ・IT分野での世界的なリードを背景に、センサーネットワークやセンサーデータベース、モバイルデータ管理などの分野で世界最高水準の論文誌や国際会議に多数の論文を発表している。特に、メモリデータベース技術や時空間データベース、ストリームデータ処理などの基礎研究が盛んである。 ・近年、センサーネットワークとクラウドコンピューティングを融合した Sensor Cloud の研究開発が注目を集めており、仮想化によるセンサーネットワーク資源の有効活用や、分散協調によるセンサーデータ収集と共有、センサーデータの集約・統合などの基礎研究が盛んに行われている。また、スマートフォンやソーシャルメディアを活用し人々 (crowd) がセンサーデータ収集に協力する Crowd-sourced sensing が近年注目を浴びており、環境問題や災害対応などで物理センサーを補完する有力な手段として研究開発が進められている。 ・サイバーフィジカルシステムを国の重要課題の一つと位置付けており、NSFなどの支援を得ながら、科学や医療分野と連携したIT基盤の研究開発が行われている。代表的なものとして、生態学を対象とした全国規模のセンサーデータ収集統合基盤 (NEONプロジェクト) やヘルスケアを対象とした異種センサーデータ分析基盤 (DELPHIプロジェクト) などが挙げられる。 |
| | 応用研究・開発 | ◎ | → | <ul style="list-style-type: none"> ・電力網や高度道路交通システム、ヘルスケア分野を対象に、様々なセンサーや監視カメラなどのデータを利用したスマートサービスの応用研究が行われており、実用サービスも多数開発されている。 ・サイバーフィジカルシステムの基礎研究の成果を応用し産業化につなげることを目的とした数百万ドルに上る新たなファンドがNSFによって設立され、センサーデータなどのビッグデータ利活用によるサイバーフィジカルシステムのリアルタイム制御も重要課題の一つに位置づけられている。 |

| | | | | |
|----|---------|---|---|--|
| | 産業化 | ◎ | ↑ | <ul style="list-style-type: none"> • Google, Apple, Oracle, Microsoft, IBMなど、米国に拠点を置くIT分野のトップ企業の多くが、センサーデータの収集や分析に関する製品やサービスの研究開発を積極的に行っている。 • 特にヘルスケア分野への応用が著しく、GoogleやAppleは自社のスマートフォンアプリと連携し、個人の健康情報を一元化するためのサービスを提供している。 • 国立標準技術研究所(NIST)が、様々なセンサーを横断的に組み合わせてスマートサービスを開発する技術の標準化を産業界と連携して行っている。 |
| 欧州 | 基礎研究 | ◎ | ↑ | <ul style="list-style-type: none"> • 欧州連合では、FP7を始め、社会システム構築のための先端的基礎研究に莫大な予算を投じており、「もののインターネット」(Internet of Things, IoT)関連の研究開発が多数実施されている。その結果、世界最高水準の論文誌や国際会議において多数の論文が掲載されるようになってきており、当該分野における基礎研究のレベルは非常に高い。 • FuturICTプロジェクトでは、10年間で数百人のトップレベルの研究者を動員し、環境や社会、経済活動にまで至るグローバルなセンサーネットワークを構築し、様々な科学モデル、データを統合し大規模な社会シミュレーションを行うための基盤技術の研究開発を行っている。 • スマートシティを対象に様々なセンサーデータを統合し分析する技術の基礎研究が盛んに行われている。さらに近年では、ソーシャルメディアやマルチメディアも統合し、様々な種類のデータストリームを横断的に統合し人や社会の活動を分析する研究も活発になってきている（FP7 MultisensorプロジェクトやSocialSensorプロジェクトなど）。 |
| | 応用研究・開発 | ○ | ↑ | <ul style="list-style-type: none"> • ヘルスケア分野でデジタル・ソリューションの活用を図る「eヘルス行動計画」が2014年に欧州議会で承認され、ほぼ同時期に、米国との間で患者データの共有を図るeヘルス関連のプロジェクト「Trillium Bridge」の設立が公表された。 • 欧州委員会のスマートシティ戦略的実践計画に対して、研究・イノベーション資金配分プログラム(Horizon 2020)から2億ユーロを拠出する意向を示すとともに、スマートシティとスマートビジネスの構築を支援するアプリケーションに、賞金総額で40万ユーロを提供する欧州史上最大のアプリ・コンテストの開催を計画するなど、応用研究・開発に向けた取り組みを加速している。 |
| | 産業化 | ◎ | ↑ | <ul style="list-style-type: none"> • 産学官が一体となりスマートシティへの取り組みを活発に進めており、環境センサーや交通センサーと連動した各種公共サービス（交通規制や防犯対策、市民生活支援など）の実証実験を行っている。 • 製造機械だけでなく、作業人や物流サービスとの間でもデータをやり取りし製造業全体をスマート化するインダストリー4.0への取り組みがドイツを中心に進んでいる。 |
| 中国 | 基礎研究 | ○ | ↑ | <ul style="list-style-type: none"> • 欧米に留学し学位を取得した優秀な大学教員や留学生在が多数帰国し、これらの人材を中心に基礎研究の水準が急速に上昇している。研究分野の多くは米国と重なっており中国独自のものは少ないものの、世界最高水準の論文誌や国際会議に多数の論文を発表している。 |
| | 応用研究・開発 | △ | → | <ul style="list-style-type: none"> • クラウドコンピューティングやビッグデータの国家戦略的な推進により、北京、上海、深圳、杭州、無錫をモデル都市として、各種コンピューティング資源を統合し、国レベルのインフラ構築を加速させている。 • 膨大な人口をカバーする高速無線通信網の整備が急速に進んでおり、今後センサーやモバイルデータを利活用する応用研究・開発が進展してくると思われる。 |
| | 産業化 | ○ | → | <ul style="list-style-type: none"> • PM2.5など環境問題への関心の高まりを背景に、センサーデータの収集・分析や公開を行うサービスが登場し始めている。 • 論文としての研究成果を重視する習慣が根強く残っており、産業化は未だ発展段階と思われる。 |

| | | | | |
|----|---------|---|---|---|
| 韓国 | 基礎研究 | ○ | ↑ | ・欧米に留学し学位を取得した優秀な大学教員や留学生が多数帰国し、これらの人材を中心に基礎研究の水準が急速に上昇している。地理情報システムやモバイル分野の研究に力を入れ、世界最高水準の論文誌や国際会議でのプレゼンスが上昇してきている。 |
| | 応用研究・開発 | △ | ↑ | ・科学・ICTと既存産業の融合によるスマートな社会の実現をめざす「創造ビタミンプロジェクト」を韓国政府が打ち出し、農畜水産食品、文化観光、保健医療、主力製造業、教育学習、小商工業創業、災害安全を対象とした産業の高度化と問題解決を支援する。 |
| | 産業化 | ○ | → | ・サムスン電子の世界スマートフォン市場での躍進に加え、他国と比べて高いスマートフォン普及率とブロードバンド、LTEなどのICTインフラの高度化が進んでおり、モバイル分野で大変注目度が高い |

(註1) フェーズ

基礎研究フェーズ：大学・国研などでの基礎研究のレベル

応用研究・開発フェーズ：研究・技術開発（プロトタイプの開発含む）のレベル

産業化フェーズ：量産技術・製品展開力のレベル

(註2) 現状

※我が国の現状を基準にした相対評価ではなく、絶対評価である。

◎：他国に比べて顕著な活動・成果が見えている、○：ある程度の活動・成果が見えている、

△：他国に比べて顕著な活動・成果が見えていない、×：特筆すべき活動・成果が見えていない

(註3) トレンド

↑：上昇傾向、→：現状維持、↓：下降傾向

(8) 引用資料

- 1) 科学技術イノベーション総合戦略 2014

<http://www8.cao.go.jp/cstp/sogosenryaku/2014/honbun2014.pdf>

- 2) モバイルワイヤレステストベッド

<http://www.nict.go.jp/press/2014/01/29-1.html>

- 3) WiSUN アライアンス

<http://www.wi-sun.org>

- 4) G 空間×ICT 推進会議

http://www.soumu.go.jp/main_sosiki/kenkyu/gkukan-ict_suishin/

- 5) 平成 26 年度 科学技術関係アクションプラン

<http://www8.cao.go.jp/cstp/budget/h26gaisan2.pdf>

- 6) ビッグデータ統合利活用のための次世代基盤技術の創出・体系化プロジェクト

http://www.jst.go.jp/kisoken/crest/research_area/ongoing/bunyah25-6.html

- 7) 科学的発見・社会的課題解決に向けた各分野のビッグデータ利活用推進のための次世代アプリケーション技術の創出・高度化プロジェクト

http://www.jst.go.jp/kisoken/crest/research_area/ongoing/bunyah25-5.html

- 8) Lane, N.D. Miluzzo, E. ; Hong Lu ; Peebles, D. ; Choudhury, T. ; Campbell, A.T.: A Survey of Mobile Phone Sensing, Communications Magazine, 48(9), IEEE, pp. 140 – 150 (September, 2010).

- 9) Golab, L., Özsu, M. T.: Data Stream Management, Synthesis Lectures on Data Management, 2(1), Morgan and Claypool, pp. 1-73 (2010).

- 10) Gupta, A., Jain R.: Managing Event Infoamation: Modeling, Retrieval, and Applications, Synthesis Lectures on Data Management, 3(4) pp.1-141 (2011).

3.6.4 時空間データマイニング技術

(1) 研究開発領域名

時空間データマイニング技術

(2) 研究開発領域の簡潔な説明

センサーデータのような時系列データ、GIS のような地理情報、Web アクセス履歴のような時刻付きレコードデータ、ソーシャルネットワークにおける時間発展グラフデータなど、時空間情報を分析し、知識発見を行う技術に関する研究開発

(3) 研究開発領域の詳細な説明と国内外の動向

時空間データ

時系列データとは、ある現象の時間的な変化を計測して得られた値の系列である。従来の時系列解析は、物体の動き、株価や経済指標のような金融データ、温度や湿度などの気象データの解析が中心であった。しかし近年、取り扱う情報が飛躍的に増大しており、またデータの種類も多様化している。そのようなビッグデータ時代においては、多様な時系列データ、もしくは空間情報を含む大規模な時空間データの解析が極めて重要な課題となっている。

時空間データに関しては、特に Web 空間におけるデータが増大の一途をたどっている。例えば現在、Facebook や Twitter などの巨大なソーシャルネットワーク上では日々大量の情報が高速に流通している。Facebook のユーザー数は 2008 年の 1 億人から現在は 10 億人を超え、ユーザー間のリンク数は 1400 億を超える。2 億人以上のユーザーを持つ Twitter では毎日 2 億件以上のツイートが投稿されている。これらの情報はネットワーク上を刻々と流れており、時間変化の伴う時系列データと言える。また、ソーシャルネットワークにおけるユーザー間のリンク構造、あるいはユーザーと組織間の関係を表現したリンク構造も時間経過とともに変化しており、Web 空間の変化を表現した時空間データと言える。

Web 情報やソーシャルネットワーク以外にも様々な分野で時空間データマイニングが注目されている。GPS 付き携帯端末や高度交通システムの導入、普及により、実社会での車や人、物の位置情報がリアルタイムに把握できるようになっている。また、センサノードのような超小型デバイス技術の発達により、大規模なセンサーネットワークが様々なところで構築されている。そのような位置データやセンサーデータの時系列解析技術も必要不可欠となっている。さらに近年、電子レセプトの普及に伴い、電子カルテからの診察記録や症例登録など、様々な医療データが収集、蓄積されるようになっている。医療データは、地域や時間に関する情報とともに蓄積されるため、時空間データとしての解析が求められている。それら時空間データに対する解析技術として、最近特に注目されているのが、(a) 時系列データマイニング、(b) ストリームマイニング、(c) グラフマイニングである。

時系列データマイニング

時系列データから有用な情報を検索あるいはマイニングするために様々な処理が必要とされる。従来は、ARIMA (Autoregressive integrated moving average) のような回帰モデルを用いた時系列解析が行われていたが、データの大規模化によって、時系列データベースに対する問い合わせパターンの検索、頻出パターンの探索やモデル化、時系列シーケンスの分

類、時系列の部分パターンの照合、情報予測などに関する高速処理技術の研究開発が行われている。さらに、時刻付きイベントデータ、すなわち購買データのようなトランザクションデータも時系列データの一つと考えられ、例えば Web アクセス履歴のような時刻付きの多項目のレコード集合からなるデータもこれに含まれる。このように時系列データマイニングの内容は多岐にわたる。

それら時系列データマイニングの応用として、近年注目を集めているのがソーシャルネットワークにおける情報伝搬の解析である。ブログや Twitter をはじめとするインターネットメディアの普及により、オンライン上でのニュースや噂の伝播速度が増している。情報は緩やかに拡散してゆっくりと減衰していくのか、それとも急激に伝わると同時にすぐに消えていくのか、ニュースをはじめとする情報がオンラインメディア上でどのように伝わり、減衰していくかに着目した解析技術の開発が行われている。そして、ソーシャルメディア上における情報伝播の解析によって、特定の情報に関する拡散力、興味を持つ潜在的なユーザの数などを推定し、さらには外れ値検出や時系列予測などの実用的なタスクを処理することができる。

時系列データマイニングにおいて、もう一つの注目すべき応用として医療情報解析を挙げることができる。医療データは検査や診断、投薬などの医療サービスの内容が患者毎に、そして時間情報とともに蓄積される。従来、医療データは個々の医療機関内で管理されていたが、EHR（Electronic Health Record）に関する取り組みは、医療機関連携ネットワークを利用して個人の医療健康情報の共有を目指すものである。医療データの時空間解析は医療サービスの質の向上および効率化を図り、医療分野における様々な問題を解決できる重要なアプローチとして期待されている。

ストリームマイニング

データストリームとは、ネットワークから連続的かつ高速に流れてくる大量の時系列データのことである。センサーネットワークの普及に伴い、大量のセンサーから送られてくるデータストリームを高速に監視、分析するストリーム処理技術が重要となっている。

データストリームに対する問合せ処理を目的としたストリーム処理エンジンが、主にアメリカ主導でこれまで開発されてきた。STREAM（スタンフォード大学）や TelegraphCQ（カリフォルニア大学バークレー校）などの研究開発プロジェクトから、InfoSphere Streams（IBM）や Gigascope（AT&T）などの商用システムまで幅広くある。日本国内では研究開発プロトタイプとして StreamSpinner（筑波大学）が、商用システムとして uCosminexus Stream Data Platform（日立）や M2M ソリューション CONNEXIVE（NEC）がある。

そしてストリームマイニングとは、ストリームから有用なパターンを見つけ、特徴を捉えるための解析技術である。時系列データに関しては、データベースに蓄積された有限長の時系列データの解析のみならず、継続的に増え続ける時系列データをリアルタイムに分析し、監視するためのストリームマイニング技術が必要となっている。センサーデータの解析処理、ネットワークセキュリティー、人や自動車のような移動物体の監視、ソーシャルネットワークなどへの応用がある。増え続ける大規模なデータを分析するため、また利用者に情報をリアルタイムに提供するため、ストリームマイニングの技術は高速化と省メモリ化を図ってい

る。ストリームマイニングの技術として、部分パターン照合、トレンド検出、相関検出、分類、異常検知などが挙げられる。例えば、大規模なセンサーネットワークにおいて、類似した傾向のパターンを出力するセンサノード群をリアルタイムに検出、もしくは故障や経年劣化によって異常値を出力するセンサーを即座に発見することが可能になる。

グラフマイニング

グラフ構造とは、ノードおよびノード間のつながりを表現したエッジで構成されるものであり、分子生物学におけるタンパク質間の相互作用、タンパク質の立体構造、Web におけるリンク構造、さらにソーシャルネットワークにおけるユーザー間の関係性を表現したソーシャルグラフもその例として挙げられる。クラスタリングや異常ノードの発見、部分グラフの探索、PageRank や HITS (Hyperlink induced topic search) に代表される重要ノードの発見など多くの取り組みがある。

そして、近年特に注目を集めているのがグラフ構造の時間的変化を示す時間発展グラフデータである。例えばソーシャルネットワークにおいては、ユーザー間のリンク構造の変化を表現することができる。グラフの構造変化に基づいた時間軸のセグメンテーション、パターン発見、異常検出などの技術開発が行われている。また、時間発展グラフの効率的なデータの格納・管理技術、さらにはストリーム処理を指向したグラフストリームに関する研究も行われている。

(4) 科学技術的・政策的課題

- ・ソーシャルネットワークデータには多くの場合個人情報が含まれるため、これを広く利用するためにはプライバシー保護技術の適用が不可欠である。これは医療においても同様であり、個人の健康医療情報を解析するため、プライバシー保護は重要な課題である。
- ・ビッグデータ解析は医療において、リソース配分の適正化、現場業務の効率化、サービスの質の向上を図る重要なアプローチとして期待されている。そして、その効果を高めるため、医療情報連携基盤である EHR の普及が期待される。
- ・我が国において、自動車などの組込・制御機器分野で強みがある点を鑑みれば、センサーデータの時系列マイニングは重要な課題である。特に今後、我が国の製造業において付加価値を高めるためには、データマイニングをはじめとするソフトウェア技術が極めて重要になると考えられる。
- ・我が国では基礎研究と応用開発が個別に進められている感がある。競争的研究資金に基づくプロジェクトにおいて、産学が連携する試みもあるものの、多くの場合、製品化のような産業展開にはつながっていない。産学連携と産業展開をより一層重視した投資が適切であると考えられる。

(5) 注目動向（新たな知見や新技術の創出、大規模プロジェクトの動向など）

- ・データマイニングのトップ国際会議である KDD (ACM SIGKDD Conference)において、2014 年に最も発表数が多い研究テーマは、情報伝搬と医療情報解析である。その中には、時間発展グラフを含めた時空間データ解析の研究が数多く含まれる。

- ・近年注目されているグラフデータの応用として、RDF (Resource Description Framework) がある。これは機械処理可能な知識をグラフとして表現するためのフレームワークである。また、RDF 形式でデータを公開する動きとして、Linked Data / Linked Open Data がある。今後はこの種のデータが急増することが予想されるため、グラフマイニング技術の発展が期待される。
- ・オープンソースの大規模分散処理フレームワーク Hadoop は世界的に広く普及しており、大規模グラフ処理にも利用されている。例えば、Apache Giraph は Hadoop によるオープンソース大規模グラフ処理フレームワークである。
- ・医療情報システムの開発と医療データベース構築については、諸外国でも様々な試みがなされている。アメリカのダートマス大学では医療のレセプトデータを用いた Dartmouth Atlas Project、イギリスの National Health Service (NHS) では公的医療サービスの利用履歴データを活用した Right Care Program、カナダのオンタリオ州におけるトロント大学とウォータールー大学の協同による州内の医事管理データを活用した Health System Performance Research Network Project などが挙げられる。また、わが国でも厚生労働省のレセプトなどの National Database (NDB) が構築されている。これらは先駆的な取り組みであるが、医療データの管理と活用に主眼を置いており、医療データから価値を引き出すためのビッグデータ解析技術の開発が望まれる。
- ・製造業においては、センサーデータを対象としたビッグデータマイニング技術が製品の付加価値を高めるために重要視されている。先進的な例として以下が挙げられる。
 - ▶ テレマティクス（トヨタ自動車）：リモートメンテナンス、安全・セキュリティに関するエマージェンシーサービスなど、顧客へ新たな自動車運転サービスを提供するための ICT インフラを構築している。
 - ▶ KOMTRAX（コマツ）：建機を監視し、建機の位置や稼働情報など、データに基づいた顧客への保守サービスの強化と効率向上を図っている。
 - ▶ インダストリアル・インターネット（GE）：ジェットエンジンやガスタービンなど、様々な機械から多数のセンサーデータを収集し、データ解析に基づく新たな産業サービスを提供している。

（6）キーワード

時空間データ、時系列データマイニング、ストリームマイニング、グラフマイニング、ソーシャルネットワーク、センサーデータ、医療情報

(7) 国際比較

| 国・地域 | フェーズ | 現状 | トレンド | 各国の状況、評価の際に参考にした根拠など |
|------|---------|----|------|---|
| 日本 | 基礎研究 | △ | ↑ | <ul style="list-style-type: none"> 大学、研究所、企業などにおいて研究が行われているが、全般的に国際的なイニシアチブをとるには至っていない。 ソーシャルグラフマイニングの研究はかなり活発に行われている。また、時系列データマイニングに関しては世界トップレベルである。 2014年におけるデータベースおよびデータマイニング分野のトップ国際会議³⁶⁾の論文採択数は7件（全体の約3%）であり、依然として件数は少ないものの、若干ではあるが全般的に研究レベルは上昇傾向にあると思われる。 |
| | 応用研究・開発 | △ | → | <ul style="list-style-type: none"> データマイニングにおいてはGoogleなどの米国企業が強みを有するため、応用研究については日本がいかにかに伍していくか早急な検討が必要である。 分散オンライン機械学習フレームワークであるJubatusが現れるなど、一定の進展が見られる。 |
| | 産業化 | △ | → | <ul style="list-style-type: none"> データストリーム処理のための商用システムとして、uCosminexus Stream Data Platform（日立）やM2MソリューションCONNEXIVE（NEC）がある。 |
| 米国 | 基礎研究 | ◎ | ↑ | <ul style="list-style-type: none"> データマイニングについては、米国の大学、研究機関における基礎研究のレベルは高い。 2014年におけるデータベースおよびデータマイニング分野のトップ国際会議³⁶⁾の採択数は147件（全体の約57%）であり、件数は極めて多い。その中には世界をリードする研究も数多く見られる。 |
| | 応用研究・開発 | ◎ | ↑ | <ul style="list-style-type: none"> 医療情報解析に関する研究開発が本格化している。国際会議KDD2014では11件の医療情報解析に関する発表があり、その中で9件が米国の企業、大学からの発表である。また、IBMは基礎研究と応用研究の両面で医療情報解析の技術開発をリードしている。 |
| | 産業化 | ◎ | ↑ | <ul style="list-style-type: none"> Google、Facebook、IBM、AT&Tなどの企業によって、各社のために有用なデータマイニング技術が研究開発され、様々な商用製品がリリースされている。 著名なグラフデータベースNeo4jは米国発である。 |
| 欧州 | 基礎研究 | ○ | ↑ | <ul style="list-style-type: none"> データマイニングについては、特にドイツの大学、研究機関における基礎研究のレベルが高い。またデータベース技術に関してはドイツのみならず、スイスの研究レベルも高い。 2014年におけるデータベースおよびデータマイニング分野のトップ国際会議³⁶⁾の採択数はドイツ11件（全体の約4%）、スイス7件（約3%）、イギリス4件（約2%）、フランス4件（約2%）、イタリア3件（約1%）となっている。その中にはインパクトのある研究も見られる。 |
| | 応用研究・開発 | ○ | ↑ | <ul style="list-style-type: none"> StreamCloudプロジェクトなどデータストリームの研究開発、さらにグラフデータ処理の研究も盛んに行われている。 |
| | 産業化 | △ | → | <ul style="list-style-type: none"> 時空間マイニングについては、応用研究・開発の動きは見られるものの、全体的に産業化が進んでいることを示す明確な情報は見当たらない。 |
| 中国 | 基礎研究 | ○ | ↑ | <ul style="list-style-type: none"> 2014年におけるデータベースおよびデータマイニング分野のトップ国際会議³⁶⁾の論文採択数は35件（全体の約14%、香港の11件を含む）であり、件数は急増している。また、米国の大学との共同研究が多いことも特徴である。 |
| | 応用研究・開発 | ○ | ↑ | <ul style="list-style-type: none"> グラフデータ処理の研究が盛んに行われている。 地理情報を用いたソーシャルネットワークマイニング、移動軌跡データ解析など、位置情報を用いたデータマイニングに関する応用研究も盛んである。また、大気汚染に関する研究も行われている。 |
| | 産業化 | △ | ↑ | <ul style="list-style-type: none"> Baiduをはじめとして、IT企業の発展が目覚ましい。研究所の拡充も積極的に行われている。 |

| | | | | |
|----|---------|---|---|---|
| 韓国 | 基礎研究 | △ | ↘ | ・2014年におけるデータベースおよびデータマイニング分野のトップ国際会議 ³⁶⁾ の論文採択数は1件と大幅に件数は下がっている。しかし、韓国はこれまで質の高い研究成果を長年に亘って出してきており、また若手の研究者の台頭もあり、一時的なものと思われる。 |
| | 応用研究・開発 | △ | → | ・グラフデータの間合せ処理において質の高い研究が行われているものの、応用研究・開発が進んでいることを示す明確な情報は見当たらない。 |
| | 産業化 | △ | → | ・時空間マイニングについては、応用研究・開発の動きは見られるものの、全体的に産業化が進んでいることを示す明確な情報は見当たらない。 |

(註1) フェーズ

基礎研究フェーズ：大学・国研などでの基礎研究のレベル

応用研究・開発フェーズ：研究・技術開発（プロトタイプの開発含む）のレベル

産業化フェーズ：量産技術・製品展開力のレベル

(註2) 現状

※我が国の現状を基準にした相対評価ではなく、絶対評価である。

◎：他国に比べて顕著な活動・成果が見えている、○：ある程度の活動・成果が見えている、

△：他国に比べて顕著な活動・成果が見えていない、×：特筆すべき活動・成果が見えていない

(註3) トレンド

↑：上昇傾向、→：現状維持、↘：下降傾向

(8) 引用資料

- 1) Yasushi Sakurai, Yasuko Matsubara, Christos Faloutsos: "Mining and Forecasting of Big Time-series data", ACM SIGMOD Conference, Tutorial, 2015.
- 2) Christos Faloutsos, Gary L. Miller, Charalampos E. Tsourakakis: "Large Graph-Mining: Power Tools and a Practitioner's Guide", ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), Tutorial, 2009.
- 3) SIGMOD: <http://sigmod.org/>
- 4) VLDB: <http://vldb.org/>
- 5) ICDE: <http://tab.computer.org/tcde/>
- 6) KDD: <http://www.kdd.org/>
- 7) ICDM: <http://www.cs.uvm.edu/~icdm/>

3.6.5 次世代情報検索・推薦技術

(1) 研究開発領域名

次世代情報検索・推薦技術

(2) 研究開発領域の簡潔な説明

ウェブブラウザで検索エンジンを訪れ、クエリを入力してウェブページのリストを得るデスクトップ型ウェブ検索に代わる技術として、モバイル環境におけるスマートフォンなどを利用した検索・推薦技術の確立が求められている。(I) モバイル環境におけるコンテキストの利用 (II) クエリ入力を待たない能動的な情報の提示 (III) 小画面における必要最小限の情報提示 などの研究開発課題がある。

(3) 研究開発領域の詳細な説明と国内外の動向

[背景と意義]

1990年代に始まったウェブ検索のサービスは今日では日常生活と切っても切り離せないものとなっている。中国で百度 (Baidu)、韓国で Naver が国産検索エンジンとして国内トップシェアを堅持していることや、米国などで Microsoft の Bing が健闘していることを除けば、世界の検索エンジン市場は事実上 Google の支配下にある。我々は検索窓にクエリを入力し、URL (ウェブページ) のリストを出力として受け取り、この結果を順番に調べていくという検索スタイルに慣れ親しんでいる。大手検索エンジン会社は、ユーザーがどの URL をクリックしどれくらいの時間閲覧していたかといった詳細な行動記録を収集しており、クリックデータと呼ばれるこのようなデータを検索結果の質の向上に役立てている。

しかし今や、検索はデスクトップパソコンよりもスマートフォンなどのモバイル端末を通して行うのが主流である。モバイル検索では、デスクトップ型検索とは異なる以下のような重要な研究課題がある。

- (I) モバイル環境におけるコンテキストの利用：ここで、コンテキストとは例えばユーザーの絶対的位置 (緯度経度)、相対的位置 (ランドマークや他の通信機器などとの距離)、ジャイロセンサによる運動状況、生体センサーによる生体情報、気温、湿度など、モバイル端末で収集しうる情報全般をいう。モバイル検索ではこのような情報を利用してユーザーの情報要求を高精度で推定できる可能性がある。
- (II) クエリ入力を待たない能動的な情報の提示：モバイル環境においては、システムがユーザーからの詳細なクエリ入力を待つという前提が崩れる。クエリの入力手段が貧弱であるという問題だけであれば音声入力によりある程度解決されるが、移動時やコンテキストが突然変化した時など、情報要求をクエリとして表現する、もしくは情報要求自体を自覚する余裕がない場合がある。このような場合、システム側から能動的に役立つ情報を適切なタイミングで提供することが求められる。
- (III) 小画面における必要最小限の情報提示：デスクトップパソコンとは異なり、モバイル端末では一度に提示できる情報の量に制限がある。また、情報を閲覧する時間にも制限がある場合が多い。従って、従来のウェブ検索のようにウェブページをひとつひとつ精査して情報を探すという形態ではなく、システムが必要最小限の情報をコンパクトにまとめて提示することが求められる。ユーザーとシステムの間でインタラクショ

ンが必要な場合も生じるが、この場合のユーザーの入力および操作も最小限にとどめるよう工夫する必要がある。

[これまでの取り組み]

2012年2月、国際的にアクティブな情報検索研究者が豪 Lorne に集結し、情報検索技術の今後の課題について議論を行い報告書¹⁾をまとめた。この中で特に、モバイル検索を想定した文書単位ではなく情報単位の要約型検索や、クエリ入力を待たない検索の重要性が指摘された。以下の TREC (Text Retrieval Conference) および NTCIR (NII Testbeds and Community for Information Access Research)²⁾ における最近の取り組みはこの流れを汲むものである。

<<米 TREC の取り組み>>

TREC³⁾は米 NIST (National Institute of Standards and Technology) が主催する情報検索に関するコンペティション型国際会議である。ここでは様々な情報検索の問題（研究種目）が設けられており、これは今後の技術発展を占う判断材料となる。TREC では 2012 年にコンテキスト型推薦トラック (Contextual Suggestion Track) がスタートしており、モバイル環境においてユーザーのクエリ入力の代わりにユーザーの嗜好データを利用し、コンテキストをもとに情報推薦を行うシステムの試作が始まっている。TREC のその他の種目が常にクエリ入力を前提としてきたのに対し、クエリ入力なしの状況を扱っている本種目は注目に値する。

<<アジアにおける NTCIR の取り組み>>

NTCIR は NII (国立情報学研究所) が主催する「アジア言語を対象とした TREC」に相当する国際会議である。NTCIR では 2011 年より、スマートフォン上の検索ボタンを一度押すだけで所望の情報が画面に表示されることを狙った One Click Access という種目が行われており、これは 2014 年に MobileClick という名前が変わっている。対象言語は英語および日本語である。また 2015 年には、モバイル端末上でユーザーとシステムが必要最小限の対話を行うことを狙った Short Text Conversation という種目がスタートしている。こちらの対象言語は中国語である。

[今後必要となる取り組み]

前述の三大課題 (I) モバイル環境におけるコンテキストの利用 (II) クエリ入力を待たない能動的な情報の提示 (III) 小画面における必要最小限の情報提示 は、それぞれ多くの未解決問題を有している。まず (I) および (II) についてであるが、個々のモバイル端末にどのようなセンサーを盛り込んでコンテキスト情報を収集し、各種コンテキスト情報をどのような情報要求に対しどのように組み合わせ活用すべきかは明らかになっていない。また、ユーザー端末からのコンテキスト収集には常にプライバシーの問題が付きまとうので、この点に関してフェイルセーフな戦略が必要である。さらに (II) については、何をユーザーに提示すべきかだけでなく、どのようなタイミングで提示すべきかが重要課題である。(III) については、これまで行われている研究がもっぱらテキスト情報を対象としたものが中心で

あることから、テキスト以外のメディアや情報提示方法（例えばグラフによる簡潔かつ直観的な可視化）を含めた検討が必要である。さらに、最小限のインタラクションによりユーザーを所望の情報に導く手法の体系化も必要である。

（４）科学技術的・政策的課題

デスクトップ型検索においては米 Google や Microsoft がユーザーから大量のクリックデータを収集し、これを検索の質の向上に役立てている。中国には百度（Baidu）、韓国には Naver という国内トップシェアを維持しているウェブ検索エンジンがあり、同様の取り組みを行ってきたと考えられる。一方、巨大市場を抱える中国では Weibo, Weixin などの国産ソーシャルメディアが普及しており、韓国の Naver は LINE をアジアのみならずグローバルレベルで展開している。このようにソーシャルメディアにおいて大規模ユーザベースをもつことは、デスクトップ型検索がそうであったように、今後のモバイル検索の質の向上のキーとなると考えられる。

ユーザーからのコンテキスト収集には Apple の iPhone、Google の Android Phone といったスマートフォンや、将来的には Google Glass（これ自体の一般向け販売は 2015 年 1 月で終了）、Microsoft SenseCam のようなウェアラブルデバイスとの連携が不可欠である。この点でも上記米国企業や、韓国のスマートフォンメーカーなどは有利である。

上記を鑑みると、大規模ユーザベース、魅力的なデバイス、いずれの観点からも日本は次世代モバイル検索分野において不利である。日本の成熟したモバイルユーザを活用し、例えば「災害時に命を救うモバイル検索」のように頑健な通信インフラの構築から末端ユーザーの満足度向上までを包括的にとらえた大規模研究プロジェクトが急務であると考えられる。さらに、このとき開発する技術は日本国内のみならずグローバルにインパクトを与えうるものにするのが必須である。

（５）注目動向（新たな知見や新技術の創出、大規模プロジェクトの動向など）

モバイル検索においてユーザーを満足させるには、ユーザーからの暗黙的フィードバックの収集（ユーザーが満足しているかどうかの手がかりをユーザーに意識させずに収集すること）が重要である。ウェブ検索におけるユーザーからの暗黙的フィードバックは主にウェブページのクリックデータであったが、モバイル環境ではよりリッチな情報を利用できる可能性が高い。例えば、スマートフォンにおけるタッチインタラクションの履歴や、小さい画面に表示されるウェブページの領域情報、心拍数や脳波などの生体情報をモバイル検索に活かそうという研究が始まっている。

（６）キーワード

モバイル検索、情報検索、情報アクセス、コンテキスト、暗黙的フィードバック、レコメンデーション、パーソナライゼーション、スマートフォン、ウェアラブルデバイス

（7）国際比較

| 国・地域 | フェーズ | 現状 | トレンド | 各国の状況、評価の際に参考にした根拠など |
|------|---------|----|------|---|
| 日本 | 基礎研究 | ○ | ↑ | 国立情報学研究所が主催するアジア言語を対象としたコンペティション型の国際会議NTCIRでは、スマートフォンの小さい画面に必要最小限の情報を提示することを前提としたMobileClickという研究種目が設けられており、基礎研究レベルでの今後の発展が期待できる。 |
| | 応用研究・開発 | △ | → | 例えばモバイル端末向けに2つのレストランの特徴を比較するインターフェースなどの試作がある。しかし、個々のプロトタイプレベルである。 |
| | 産業化 | × | → | 上記のような基礎・応用研究が産業に活かされることが期待されるが、デスクトップ型ウェブ検索（Google, Bing）の知見やモバイルOSをもつ米国、携帯端末およびLINEのようなアプリケーションをもつ韓国、国産アプリケーションと大規模ユーザーベースをもつ中国のような際立った強みが日本には見当たらない。 |
| 米国 | 基礎研究 | ○ | ↑ | 米NISTが主催するコンペティション型国際会議TRECでは、ユーザーの嗜好を用いてモバイル環境における推薦を行うContextual Suggestion Trackという研究種目が設けられ、クエリ入力が必要としないモバイル推薦の研究が始まっている。 |
| | 応用研究・開発 | ○ | ↑ | デスクトップ型ウェブ検索に関する研究で多くの成果を発信しつづけているMicrosoftなどの企業がモバイル検索にも着手しており、今後も応用研究が加速すると考えられる。 |
| | 産業化 | △ | ↑ | 例えば米AppleはiPhone、GoogleはAndroid Phone、MicrosoftはWindows Phoneを有しており、これらの製品を活用したモバイル検索・推薦の産業化が展開しやすい状況にある。特にGoogleとMicrosoftはデスクトップ型ウェブ検索において深耕した技術を応用できるという強みもある。 |
| 欧州 | 基礎研究 | △ | → | アジアのNTCIR、米TRECに対応する欧州のコンペティション型会議CLEF (Conference and Labs of the Evaluation Forum) ²⁾ では本研究分野に直接インパクトを与える研究種目が設定されていない。 |
| | 応用研究・開発 | △ | ↑ | 例えばDublin City大学はライフログ収集の研究を長期間に亘って行っており、もしこのようなコンテンツに効率的にアクセスする技術が確立できれば応用研究の道が開ける可能性がある。ただし、ライフログ収集のために用いている技術はGoogle GlassやMicrosoft SenseCamなど、米国のものが目立つ。 |
| | 産業化 | × | → | 上記の状況により、明確な産業化への糸口が見えていないと言えない。 |
| 中国 | 基礎研究 | × | → | 本分野に関する基礎研究レベルでの存在感は大きくない。 |
| | 応用研究・開発 | △ | ↑ | 研究レベルでの存在感は大きくないものの、中国のIT企業は国産ソーシャルメディア（Weibo, Weixin）などを発展させており、これらの基盤を利用したモバイル検索・推薦の研究開発が今後発展する可能性はある。また、通信機器メーカーである華為技術は香港に研究所を新設し、モバイル分野の研究開発に取り組んでいる。 |
| | 産業化 | △ | ↑ | 上記の状況と、国内市場が極めて大きいことから今後の発展が見込まれる。また、中国語のデスクトップ型ウェブ検索で大成功を収め、最近では日本国内でもユーザログ情報無断送信の件で話題となった百度（Baidu）も、モバイル検索への展開を進めている。 |
| 韓国 | 基礎研究 | × | → | 本分野に関する基礎研究レベルでの存在感は大きくない。 |
| | 応用研究・開発 | △ | ↑ | 研究レベルでの存在感は大きくないものの、LINEを運営するNaverがアジアにとどまらずグローバルに展開している例もあり、このような企業が今後大規模ユーザーデータを活用した応用研究開発で存在感を増す可能性はある。 |
| | 産業化 | △ | ↑ | 上記の状況より、産業レベルでも今後の発展が見込まれる。 |

- (註1) フェーズ
基礎研究フェーズ：大学・国研などでの基礎研究のレベル
応用研究・開発フェーズ：研究・技術開発（プロトタイプの開発含む）のレベル
産業化フェーズ：量産技術・製品展開力のレベル
- (註2) 現状
※我が国の現状を基準にした相対評価ではなく、絶対評価である。
◎：他国に比べて顕著な活動・成果が見えている、○：ある程度の活動・成果が見えている、
△：他国に比べて顕著な活動・成果が見えていない、×：特筆すべき活動・成果が見えていない
- (註3) トレンド
↑：上昇傾向、→：現状維持、↓：下降傾向

(8) 引用資料

- 1) Allan, J. et al. Frontiers, Challenges, and Opportunities for Information Retrieval: Report from SWIRL 2012 The Second Strategic Workshop on Information Retrieval in Lorne. SIGIR Forum. 2012. Vol.46, No.1, pp.2-32 <http://www.cs.rmit.edu.au/swirl22/>
- 2) CLEF (Conference and Labs of the Evaluation Forum) オンライン予稿集. <http://www.clef-initiative.eu/publication/working-notes>
- 3) Dean-Hall, A. and Clarke, C.L.A. Assessing Contextual Suggestion, EVIA 2014. pp.29-32
- 4) Jin, H. et al. ReviewCollage: A Mobile Interface for Direct Comparison Using Online Reviews. Proceedings of ACM MobileHCI 2014. pp.349-358
- 5) NTCIR (NII Testbeds and Community for Information Access Research) オンライン論文集. <http://research.nii.ac.jp/ntcir/publication1-ja.html>
- 6) Sakai, T. et al. Click the Search Button and Be Happy: Evaluating Direct and Immediate Information Access. ACM CIKM 2011. pp.621-630
- 7) TREC (Text Retrieval Conference) オンライン論文集. <http://trec.nist.gov/pubs.html>
- 8) 情報検索関係の国際会議論文集 (SIGIR, CIKM, WWW, WSDM など)
- 9) 情報推薦関係の国際会議論文集 (RecSys など)
- 10) モバイルインタフェース関係の国際会議論文集 (CHI, MobileHCI, Ubicomp など)
- 11) 酒井哲也. 情報アクセス評価方法論～検索エンジンの進歩のために～. コロナ社. 2015.

3.6.6 個人ライフログデータの記録・利活用技術

（1）個人ライフログデータの記録・利活用技術

（2）研究開発領域の簡潔な説明

個人の行動をデジタル記録・アーカイブし、そのデータの解析や可視化を行う技術であり、自分自身の体験や活動の把握に利活用することで、日常の記憶や健康の増進など、様々な目的へ資することが期待されている。なお、必要に応じ行われる新しいウェアラブルデバイスの開発も含まれる。

（3）研究開発領域の詳細な説明と国内外の動向

ライフログの記録と利活用の取り組みはいまだに未成熟な状態にあり、メインストリームの一つになるかどうかの分岐点をこれから迎える。ライフログの意味するものの実態も明確には定まっていない。生活で生じたことすべての記録というのが究極ではあろう。しかしながら、現実の技術、サービスとしてライフログが論じられる場合は、滞在・移動を示す位置の記録（ロケーションログ）であったり、運動や活動、心拍の記録（活動ログ）、ウェアラブルカメラで取得する画像映像の記録（ビジュアルログ）、食事の記録、読書の記録、睡眠の記録、スマホのアプリ利用履歴と様々である。そのため生活の一側面を記録し、アーカイブすることが主として行われている。最終的なライフログは、それら個々のものの複合体となると考えられる。

また、ユーザー個人の記録からユーザー自身の生活の有様を定量化して把握することは、**quantified self** とか **personal informatics** とも称され、ライフログの利活用の一面をあらわしている。なお、最も身近なデバイスとして広く普及したスマートフォンを用いたアプリとして、ライフログを銘打つものが増え、その内蔵デバイスやウェアラブルデバイスと連動して情報収集するものを中心になっている。

学術活動

ライフログと称するアプリやサービスは増加するものの、まだ、学術領域として十分に確立されている状況に至っていないように感じられる。そのため、中国や韓国など確立した学術領域での充実を図る傾向の強い国からの学術的な活動はほとんど見られない。これまでの研究は、欧米、日本に依存していると思われる。現状の研究発表は、マルチメディア系会議、ユビキタスコンピューティング系会議等に散見される。なお、ライフログを主要課題とした学術集会を列記すると以下があげられる。

<国内>

- ・電子情報通信学会ライフインテリジェンスとオフィス

<海外>

- ・ACM Workshop on Personal Informatics¹⁾
- ・Quantified Self Conference & Expo²⁾ なお、この会は学術集会というより発表イベントに近いように思われる。

また、過去には、次のような学術集会が開かれてきた。

<海外>

- ・ ACM CARPE (2004-06): Workshop on Capture, Archival and Retrieval of Personal Experiences
- ・ SenseCam Conference (2009-2013)
- ・ IMAPA (2010-11) Information Access for Personal Media Archives
- ・ International Workshop on Lifelogging for Pervasive Health (2013)

ウェアラブルデバイス、アプリ

ウェアラブルデバイス、スマホアプリとして、現状、多くのものが現れている。人の活動量計が最もにぎやかであり、FitBit、JawBone UPなどが先行し、2014年には東芝、ソニーからもリストバンド型のウェアラブルデバイスがでてきている。なお、ソニーの場合には、ライフログと称するアプリを提供し、活動量、位置、スマホアプリの使用履歴の記録をとり、可視化するツールを提供している。

スマホアプリでの記録としては、位置（FourSquare 他）、食事記録（FoodLog 他）、睡眠（Sleep Cycle 他）、消費（Mint 他）と多彩なものがすでに一般利用に提供されている。

カメラに関しては、SenseCam というマイクロソフトの開発したウェアラブルカメラが、Vicon から売り出されていたが、現在、生産は終了している。クラウドファウンディングからスタートした Narrative が、Clip カメラ（もともと、Memoto Lifelogging Camera と称した）を発売し、GPS も内蔵した位置記録を有している。同様のものとして、SnapCam も2015年に新たに出た。

その他、様々な Personal Informatics、Quantified Self のアプリやツールが多数存在する。

（4）科学技術的・政策的課題

- ・ 「ライフログ」という個人の生活の中で生み出されるデータを扱うには、プライバシーへの十分な配慮がかかせない。取得したデータを、安易に共有することができないことは、この分野の進展を阻む大きな要因である。
- ・ プライバシーに関する議論では、カメラはその場に居合わせた他人を映してしまうために最も敏感なものになる。ウェアラブルカメラとして、Google Glass が存在するが、このプライバシーの問題があり、一般への販売を取りやめ、法人利用に特化するとの報道も2015年1月に行われた。
- ・ 「ライフログ」については、学術研究よりもアプリケーションが先行している。学術研究のためには、十分な量のデータを用い、目的に対しての評価が必要になるが、それを満たすためには、多人数のデータの取得を長期間行う必要があり、困難が多い。
- ・ アプリケーションが先行しているため、例えば、活動量に至っては、多くのウェアラブルのデバイスが利用可能である一方、データそのものは企業のクラウドに蓄積される。データ処理の観点からは、それらのデータを直接利用できない限り、何もすることができない。ライフログの研究では、データ取得のデバイスやアプリの構築を行うことから始める必要がある。
- ・ ライフログのデータを共有することが可能となるならば、そのためのデータの標準化を

行うことは大いに価値がある。ただし、現時点では、共有すること自体への問題がクリアされていない。

- ・ライフログの社会的な目標は、個人の体験を伝えることにあるとも言える。そのためには、プライバシーも考慮しながら共有のための議論が一層必要である。

（5）注目動向（新たな知見や新技術の創出、大規模プロジェクトの動向など）

- ・ウェアラブルデバイスが大きな進展をみせつつある。それらと連携したスマートフォンアプリが提供されている。2014年には、ソニーからも「ライフログ」アプリ、連動して利用するリストバンド型の活動量計がでている。現状では、ヘルスケア、健康分野が、ライフログのもっとも直接的な活用領域である。
- ・ウェアラブルカメラでは、スウェーデンの Narrative Clip (2014)、英国の Snap-Cam(2015)とライフログカメラと称するものが市場にあらわれた。その一方、GoogleGlass は一般販売を取りやめることが報じられた。なお、ビジュアルライフログの医療系への応用では、人の短時間記憶の補助やリハビリへの利用が論じられてきた。短時間記憶への効果についても論じられている³⁾。
- ・アイルランドのダブリンシティ大学のグループが 2006 年より継続して、ウェアラブルカメラで記録するビジュアルライフログを用いたライフログ研究のプロジェクトを継続している。2014年には個人ライフログについてかなり網羅的に記述した本⁴⁾を出版した。
- ・国内の比較的大きなプロジェクトとしては、次のものがあげられる。
JST CREST 「“食”に関わるライフログ共有技術基盤 (H21-26)」研究代表者 相澤（東大）FoodLog という食事記録のためのツールを構築し、アプリの提供、WebAPI の提供を行った
- ・ダブリンサイエンスギャラリーにて、2015年2月～4月に「Lifeloggging Lab」と称する展示が行われる。自分自身を知るためのツールを用いた分析と可視化の体験が提供される。

（6）キーワード

Lifelog、Lifeloggging、Quantified Self、Personal Informatics、Wearable

(7) 国際比較

| 国・地域 | フェーズ | 現状 | トレンド | 各国の状況、評価の際に参考にした根拠など |
|------|---------|----|------|--|
| 日本 | 基礎研究 | ○ | → | ライフログの黎明期から、研究がおこなわれていた。国内の電子情報通信学会に研究会 ⁵⁾ も設けられた。 |
| | 応用研究・開発 | ○ | → | 同上。 |
| | 産業化 | ○ | ↑ | ウェアラブルデバイスへの取り組みが活発化している。また、スマートフォンアプリへの取り組みも行われている。 |
| 米国 | 基礎研究 | ○ | → | 黎明期に、Microsoft ResearchにてMylifeBitsといった試みが見られたものの、アカデミアからの技術的な貢献はその後発展しているように見えない。 |
| | 応用研究・開発 | ○ | → | 同上 |
| | 産業化 | ◎ | ↑ | ウェアラブルデバイスで先行し、なおかつ活発化している。Quantified Selfの記録、分析のためのアプリも数多く出ている。 |
| 欧州 | 基礎研究 | ◎ | → | 継続的に研究を続けるグループがあり、医療系への応用も含めて、幅広い取り組みが行われてきた。 |
| | 応用研究・開発 | ◎ | → | 同上 |
| | 産業化 | ○ | → | 現時点での産業化のタネは、ウェアラブルデバイスであろうが、顕著なものが見えない。 |
| 中国 | 基礎研究 | × | → | 学術分野で、ライフログへの取り組みは、見たことがない。 |
| | 応用研究・開発 | × | → | 同上 |
| | 産業化 | △ | → | ウェアラブルデバイスなどへの取り組みはあると思われる。 |
| 韓国 | 基礎研究 | × | → | 学術分野で、ライフログへの取り組みは、見たことがない。 |
| | 応用研究・開発 | × | → | 同上 |
| | 産業化 | △ | → | ウェアラブルデバイスなどへの取り組みはあると思われる。 |

(註1) フェーズ

基礎研究フェーズ：大学・国研などでの基礎研究のレベル
 応用研究・開発フェーズ：研究・技術開発（プロトタイプの開発含む）のレベル
 産業化フェーズ：量産技術・製品展開力のレベル

(註2) 現状

※我が国の現状を基準にした相対評価ではなく、絶対評価である。
 ◎：他国に比べて顕著な活動・成果が見えている、○：ある程度の活動・成果が見えている、
 △：他国に比べて顕著な活動・成果が見えていない、×：特筆すべき活動・成果が見えていない

(註3) トレンド

↑：上昇傾向、→：現状維持、↓：下降傾向

(8) 引用資料

- 1) ACM Workshop on Personal Informatics
<http://www.personalinformatics.org/>
- 2) Quantified Self Conference & Expo
<http://quantifiedself.com/>
- 3) A. R. Doherty et al., Wearable Cameras in Health The State of the Art and Future Possibilities, Am J Prev Med. 2013 Mar;44(3):320-3.

- 4) C.Gurrin, A.F.Smeaton, A.R.Doherty, Lifelogging, Personal Big Data, Foundations and Trends, Volume 8, Issue 1, ISSN: 1554-0669, 2014.
- 5) 電子情報通信学会 ライフインテリジェンスとオフィス情報システム研究会
<http://www.ieice.org/iss/ois/jpn/>