

# ネットワーク環境上のデータ管理と社会センシングに関する研究

～進化するICT～ビッグデータ・AI技術が社会を映す～

大阪大学大学院情報科学研究科 教授(名誉教授) 原 隆浩

## 1. はじめに

この5～10年間で、私たちを取り巻く環境が大きく変化したことを多くの人々が実感しているでしょう。ブロードバンドネットワークや無線ネットワーク、スマートフォン、クラウドサービスの急激な普及、それを足掛かりとしたTwitterやFacebookなどのソーシャルネットワークサービス（SNS）の大流行、さらには様々なセンサー機器（防犯カメラ・マイク、気象センサー、車載センサーなど）や小型デバイスがネットワークで相互接続されるIoT（Internet of Thing：モノのインターネット）環境の整備などといった大きな変革が、ビジネスだけではなく私たちの日常生活を大きく変化させています。

このような変化にはもちろん良い面と悪い面がありますが、良い面の代表的例として、ビッグデータ解析による社会システムの改善や新たなサービスの創成が挙げられるでしょう。SNSやIoT環境から大量のデータ（ビッグデータ）が生成され、それを人工知能（AI）技術などで解析することにより、これまでは絶対に不可能であった、実社会で起きていることの詳細な分析やモデル化、それに基づく将来予測が可能になります。例えば、IoT環境から得られる人々の位置情報（GPSデータなど）や移動状況と、SNSから得られるそれらの人々のコメントなどを解析することで、特定の場所や時間帯での人々の行動やその動機・感想などが分かり、実社会における人の行動モデルを詳細に構築することができます。このようなモデルは、都市計画や防災、マーケティングなど様々な目的に応用することができます。

私を含む情報科学に携わる研究者は、このような時代が来ることを予見・期待し、実際に世の中の変革が起こるかなり前（10～20年前）から、多様なネットワーク環境におけるデータベース技術、ビッグデータ基盤、AI技術に関する研究開発に取り組んでいます。この度、私の約20年に亘る取組みである「ネットワーク環境上のデータ管理と社会センシングに関する研究」に対して、大阪科学賞を授与頂くことになりましたので、この研究の背景や内容、今後の可能性について簡単に紹介したいと思います。

## 2. 多様なネットワーク環境、ビッグデータ・AI時代において取り組むべき課題

上記のようにネットワークが多様化（ブロードバンド、無線・モバイル、IoT）し、ビッグデータやAIが注目されている状況において、新たに取り組むべき課題が生じています。私たちの研究では、その中でも特に次の課題に取り組んでいます。

### 課題1：個々のネットワーク環境を考慮したデータベース技術の確立

ブロードバンド、無線ネットワークなど個々のネットワーク環境には、それぞれに特徴があります。例えば、最近ではクラウドサービスなどの普及により、広範囲（国や世界規模）に分散したサービスを、ブロードバンド（高速）ネットワークを経由して利用することが一般的です。ここで、高速ネットワークは回線の大容量化が急速に進み、大量のデータを短時間で送ることは可能ですが、通信パケット（信号）がネットワーク内を進む速度（伝搬遅延）は物理的な制約（電波・光の速度など）があり急激に速くなっているわけではありません。そのため、広域な高速ネットワークでは、小さなデータを多数回送信するような処理は不向きということになります。

一方、モバイルネットワークの一種であるモバイルアドホックネットワーク（MANET）は、モバイル端末のみで無線通信により構成されるもので、災害時などインフラが破たんした状況での重要な通信基盤として注目されています。MANETでは、ネットワークを構成する端末自体が自由に移動するため、

ネットワークの形（トポロジー）が絶えず変化し、分断等が頻繁に発生します。さらに、端末のリソース（メモリ、通信速度、バッテリーなど）に大きな制約があります。

このように個々のネットワークには独自の特徴があるため、同じアプリケーション・サービスでも、それを実行するネットワークに応じて適切に処理方法を選択しなければ、不必要に処理の遅延が大きくなったり、リソースを浪費するようなことが生じます。特に、データ・情報の処理はアプリケーションやサービスの性能に直結するため、ネットワーク特性を考慮したデータベース技術が必要となるのです。

## 課題 2：SNS が生成するビッグデータを AI を用いて解析する基盤の確立

SNS から一般ユーザが投稿するメッセージは、一般的なセンサーなどでは決して得ることができない重要な情報（口コミ、評判、感想など）を多く含んでいます。そのため、大量の SNS データを解析した結果は、社会を映す特殊なセンサー（社会センサー）として捉えることができます。そのため、私たちは、SNS のビッグデータを解析することを「社会センシング」と呼んでいます。

社会センシングの研究は、学術分野におけるホットなテーマの一つであり、研究開発が盛んに行われています。その一方で、これらの研究開発は、個々のプロジェクトや機関・企業で単独に実施されており、それらの成果を統合して再利用する試みは、ほとんど行われていません。個々の取組みには限界があるため、それらを統合・共有することで、社会センシングの効果・効率は大幅に向上し、さらなる発展が期待できます。このような統合・共有のための基盤技術やプラットフォームの確立が急務と考えられます。

## 3. 研究内容の紹介

私たちは、上記の課題に取り組み、多様なネットワーク環境を考慮したデータベース技術と社会センシングに関する研究開発を推進しています。以下では、代表的な研究について、内容を簡単に紹介します。

### 3-1. 高速ネットワーク上のデータベース移動

私たちは、高速ネットワークの特徴を考慮し、最大限に活用する概念として「データベース移動」を提唱しました。データベース移動とは、データベースそのものを高速ネットワーク上で移動させることで、アクセス特性に応じてデータベースを最適な場所に再配置する際に利用されます。例えば、遠方からのアクセスが短期間に集中している場合には、その場所にデータベースを移動します（図 1）。この提唱の当時（1990 年代終盤）は、データベースは特定のサーバに固定されているのが常識であったため、データベース移動は非常に独創的なアイデアとして高く評価され、データベース分野で世界最高峰の論文誌や国際会議に論文が採択されました。

これらの研究の中で、データベースへのアクセスパターンや各サーバ・クライアントの位置、ネットワークの性能（容量、伝搬遅延）を考慮したデータベースの再配置技術や、移動中のデータベースに対しても処理を継続する技術を確立しました。そして、データベース移動の機能をもつ実システムを実装し、実測評価を行いました。その結果、データベース移動によって適切にデータベースを再配置するこ

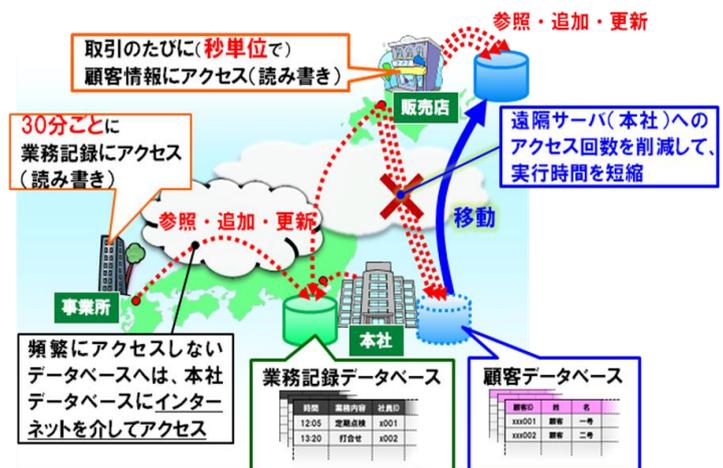


図 1：データベース移動の概念図

とで、処理時間を大幅に削減できることを確認しました。

データベース移動は、当時は大胆な発想という点が高く評価されましたが、高速ネットワークを経由して様々なクラウドが接続・連携している現在の環境は、まさにデータベース移動が想定しているものです。そういう意味では、私たちは先見的に現在の状況を予想していたと言えるかもしれません。データベース移動は、クラウド環境のデータアクセス性能を大幅に向上する可能性があるため、その価値が再度見直されることを期待します。

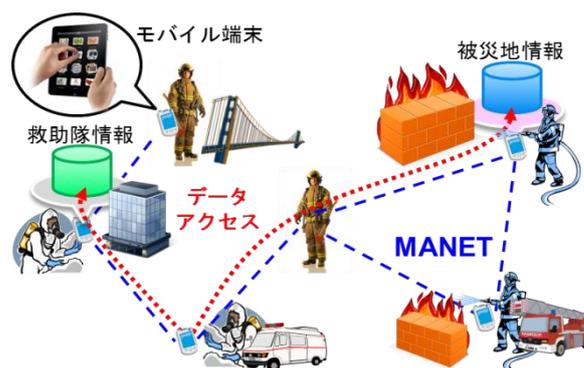


図 2: MANET 上のデータアクセスの概念

### 3-2. モバイルアドホックネットワーク (MANET) 上のデータ管理

MANET は、インフラが破たんした災害時において、被災状況や負傷者の情報、救助活動の状況など各種の情報を共有して協調作業を支援・円滑化する用途などに期待されています。そのため、情報（データ）をうまく共有、つまり有限のリソース（端末のメモリ、バッテリー、通信速度など）内で必要なデータを効率よく取得できる仕組みが必要となります。特に、MANET では、各端末が無線通信により通信パケットを中継し、バケツリレー的に遠距離の端末間の通信を実現するため（図 2）、ネットワークの分断により、必要なデータを取得できない状況が頻繁に発生します。

そこで私たちは、ネットワーク分断時にもデータへのアクセスを可能にする技術を確認しました。まず、ネットワークの分断に備えて、データの複製（コピー）を周期的に再配置する手法をいくつか考案しました。これらの手法では、再配置時のネットワークのトポロジーや各端末の各データへのアクセスの頻度を考慮します。提案の当時（2000 年頃）は、MANET 研究の最隆盛期でしたが、それらのほとんどがネットワーク技術（どのように効率的に通信パケットを通信相手に転送するか）に関するものであったため、私たちの研究は、MANET 上のデータの複製配置に取り組む世界最初の研究として注目されました。その後、「MANET 上のデータ管理」は新しい研究分野として認識され、盛んに研究されています。

私たちは、この研究の後も、MANET 内の複数の端末が所持する複製間のバージョン管理や、k 最近傍検索（指定地点に最も近い場所に関連する k 個のデータを取得）、Top-k 検索（指定条件に合致する最上位 k 個のデータを取得）などの高度なデータ検索、検索時のセキュリティ保護など、多角的に研究を行いました。

私たちの一連の研究成果は、MANET 上のデータ管理という新分野を確立した先駆的研究として高く評価され、他の研究者から頻繁に参照（2017 年 8 月の時点で Google Scholar で 2000 回以上）されています。さらに MANET は、災害時の救助・復旧活動や犯罪予防など社会的価値の高い応用が期待されており、本研究成果はそのような応用において、情報共有の効果・効率を大幅に向上する効果があります。そのため、社会的な貢献は大きいと考えます。

### 3-3. 社会センシング

私たちは 2006 年頃から、SNS データから実社会を反映する様々な情報を、AI 技術を用いて抽出する社会センシングの研究を推進してきました。その一例として、位置情報が付与されているツイート（Twitter データ）を収集して、時間的・空間的なツイート数の偏りから、ローカルなイベントを抽出する手法（図 3）を開発しました。その他にも、世の中のトレンドやユーザの興味などを抽出する手法の開

発を行いました。

先述のように、私たちの研究を含めて多くの既存研究・システムでは、解析結果の再利用を想定していなかったため、これを再利用できるようにする基礎技術やプラットフォームの構築にも取り組みました。まず基礎技術としては、ツイートなどノイズの多い短文中の語句の類似性や重要性を正確に見積もる手法を考案しました。

さらに、位置情報付ツイートが現状ではごくわずかの比率（1%以下）であることを考慮して、より多くのツイートに位置情報を付与するために、ツイートの投稿位置を推定する手法を考案しました。位置情報を付与することで、先述のように SNS データを時空間的に解析することが可能になるため、社会センサーデータの生成を促進する効果があります。

私たちは数年前から、SNS データ解析の結果を社会センサーデータとして共有することを目的として、新しいプラットフォームの構築を進めています。このプラットフォームは解析結果だけではなく、解析手法も社会センサーとして共有できるように、プログラムソースの共有を可能としています。

これらの研究は、最高峰の論文誌や国際会議に採択されるなど、学術的に高く評価されています。また、SNS データの解析結果を社会センサーデータとして共有することを促進する効果があるため、社会的にも大きな貢献があると考えます。

#### 4. 将来展望

私たちの研究は、当該研究分野の常識を覆すような先駆的な方法論を数多く考案しているため、それらのいくつかが今後の研究分野の新しい方向性を生み出すことを期待しています。なお、MANET 上のデータ管理については、既にこれを実現し、新たな研究分野となりました。

最後に紹介した社会センシングの研究について、社会センシングの解析結果と解析手法を共有するための基盤プラットフォームが実現すると、各機関・企業で個別に開発された解析システムや手法を相互に連携することが可能となります。その結果、新たな技術的なブレークスルーの起点になることが期待されます。さらに、災害対策・対応など社会的に重要性の高い応用やサービスの開発が活性化されることが期待されます。

#### 5. 情報技術の研究への誘い

皆さんもご存知のように、最近では情報技術に直接携わる分野以外でも、ほとんどの分野において、ビッグデータを AI 技術で開発することが業務改善や新サービス創出の観点で必須となっています。このような状況に反して、ビッグデータや AI の研究者・技術者の数は十分ではなく、その育成が急務となっています。ビッグデータや AI を始めとする情報科学の分野は、ベースとなる基盤技術（ネットワーク、デバイス・ハードウェア、プログラミング言語・開発環境、データベース、機械学習など）の進歩も急速であり、日々の継続的な勉強・研さんを必要とする学問分野ですが、その分、若い人たちでも頑張れば短期間で成長し、活躍することが可能です。また、情報科学は、真理を発見するその他の自然科学分野とは一線を画し、自分が想像した世界を自分で実現できる分野でもあります。私自身、このように技術と創造性の両方を必要とする情報科学の魅力の虜となり、この分野の研究者になりました。

この講演で、情報科学に興味をもち、情報技術者・研究者を目指す若い人が少しでも増えれば、至極の喜びです。ぜひ、一緒に情報科学について研究し、未来を切り拓きましょう。

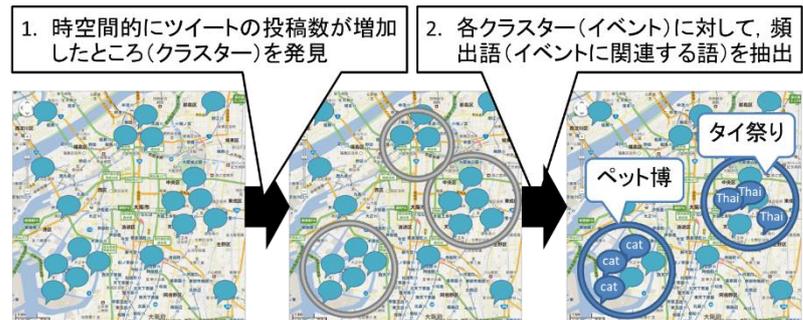


図 3：ツイートからのイベント抽出

## 用語集

### **データベース**

コンピュータシステム上でデータを再利用可能（抽出、加工、検索など）な形式で管理したものをデータベースという。データベースを構築、運用、管理するためのシステム、およびそのソフトウェアをデータベースシステム（厳密にはデータベース管理システム）という。

### **データ管理**

データベースシステムやファイルシステム上で、データをどのように蓄積し、加工し、アプリケーションやユーザからのデータアクセス要求に応答するかなどを含めた、データの扱い方の全般をデータ管理という。

### **クラウドコンピューティング**

サーバ、ストレージ、アプリケーション、サービス、ネットワークなどの計算機に関わるリソース（資源）を、インターネットを経由してオンデマンドに提供するコンピューティングモデルのことをさす。このようなサービスを提供するサーバ群をクラウドと呼ぶ。

### **ソーシャルネットワークサービス（SNS）**

Webなどをベースとしてインターネット上で、社会的ネットワーク（ソーシャルネットワーク）を構築するサービスのことをいう。ソーシャルネットワーキングサービスともいう。SNSでは、ユーザ同士が友人や支持者としてつながることで、情報共有やコミュニティ形成などのサービスを享受する。代表的なものに、FacebookやTwitterなどがある。SNSは、大規模数の一般人（ユーザ）が様々な情報を発信するため、社会を映す貴重な情報源であるとともに、ビッグデータの大きな要因の一つとなっている。

### **IoT（Internet of Things：モノのインターネット）**

パソコンやサーバなどのみがインターネットにつながっていた従来の環境とは異なり、最近では防犯カメラや各種センサー、監視・制御機器、家電など様々なデバイス（モノ）がインターネットで相互接続されており、これをIoT（Internet of Things：モノのインターネット）と呼ぶ。IoTにより、大量のデバイスが実社会の情報を詳細に収集するため、様々な新しいサービスを展開できる。その一方で、それらのデバイスが生成する大量のデータが、ビッグデータの大きな要因の一つとなっている。

### **ビッグデータ**

IoTやSNSなどの急速な普及により、インターネットなどのコンピュータネットワーク上で生成されるデジタルデータが質・量ともに急速に増加しており、これをビッグデー

タと呼ぶ。また、ビッグデータという用語は、話の文脈に応じて、大規模なデータそのものや、データが爆発している状況、大規模データを解析する技術、解析して知見を得る一連の行為など、さまざまな意味で用いられる。

## **人工知能**

コンピュータシステム上などで人間と同様の高度な知能を実現させようという試みであり、1960年代頃の第一次ブームを始めとして、1980年代の第二次ブームを経て、数年前から第三次ブームが到来している。第三次の人工知能ブームは、ビッグデータの流行が後押ししており、これまで長年に亘り成熟してきた機械学習などの人工知能技術が、利用可能なデータの増加により、様々なアプリケーションで重要な成果を達成している。