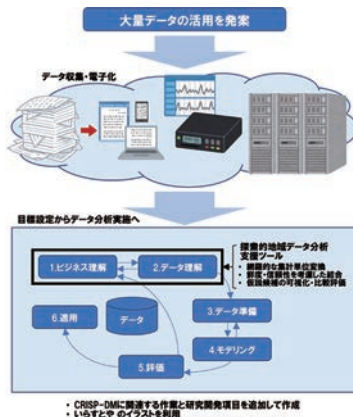


大量データをスムーズに活用するためのデータサイエンス

コンピュータの高性能化と記憶装置の大容量化、そしてインターネット上で流通する情報の爆発的な増加により、私たちを取り巻く生活環境は大きく変化しました。研究開発の現場では、個人や研究室で蓄積した知識や経験に加え、実験装置で計測した大量のデータや日々増え続ける文献やデータベースを活用することが重要になりつつあります。

しかし、手持ちのデータを活用しようとしても、データが整理されていなかったり、印刷されたデータしかなかったりと、データ分析を簡単に始められないケースもあります。また、データが準備できたあとも、結論に向けて見通しよく作業を進めるために、データから何を導くのか、どの分析方法を使うのか、どのように評価するかなど、一定のスキルと経験に基づいた意思決定が必要となります。こうした課題を解決するための技術の総称がデータサイエンスであり、現在、データの理解をスムーズに進めるための探索的データ分析法に関する研究と、専門知識とデータサイエンティストの素養を備えた技術者・研究者の育成に取り組んでいます。

大量データを処理する流れの概要



産業界へのアピールポイント

- 企業における実務経験と大学・大学院における教育経験に基づき、データサイエンスに関する研究開発、技術者・研究者の育成に取り組んでいます。
- 既存データの活用に関する各種相談を受け付けています。

実用化例・応用事例・活用例

- オープンソースソフトウェアを利用したデータ分析環境の構築
- OCR を利用した紙資料の電子化と既存データと紐づけ・名寄せの実施
- 地理的データの集計単位（地域メッシュ、行政区等）の高速変換



平松 薫（ヒラマツ カオル） 教授
大学院理工学研究科

【最近の研究テーマ】

- 地理的要素を含むオープンデータの探索的データ分析の効率化
- 探索的データ分析におけるインタラクティブなデータ可視化の高速化
- OCR を利用した紙資料の電子化とそのデータクレンジング

