

データサイエンスにおける データ抽象化によるデータ理解へのアプローチ

小特集編集にあたって

編集チームリーダー 堀山貴史

「データサイエンスにおけるデータ抽象化によるデータ理解へのアプローチ」小特集と、何ともデータの多い小特集タイトルで恐縮ながら、データは近年とみに注目を集めており、ビッグデータ、データ解析、データ分析、データマイニング、データサイエンティストなどなど、データにまつわる用語が世をにぎわしている。情報とデータは、JIS X 0001 情報処理用語—基本用語の 01.01.01 と 01.01.02 に掲載される基本中の基本であり、情報化社会の到来とともにデータが持つ情報の重要性は増し、データから情報を得る技術への熱望は高まっている。

本小特集では、JST CREST「ビッグデータ統合活用のための次世代基盤技術の創出・体系化」の研究課題の一つである「データ粒子化による高速高精度な次世代マイニング技術の創出」で培われた研究成果を核に、より良いデータ理解への取組みについて解説する。具体的には、データを抽象化することでデータの俯瞰や着目点の移動を容易にし、意味解釈や因果推定など人間の思考を効果的にサポートする技術について、数理理論やアルゴリズムからインタラクション、応用の事例まで統合的に論じる。

第1章「人の思考に寄り添ったデータ解析技術への道」は、上記 CREST 課題の研究代表者である宇野毅明による。意味解釈を大事にしたデータ解析の新しい潮流、データの抽象化を行うデータ研磨アルゴリズム、そしてその利用について紹介する。計算技術による思考支援を行う際にどのようなポイントが重要になるか、データを効果的に抽象化するデータ研磨など、本小特集を御

覧頂くにあたって、まず最初に目を通して頂きたい。

第2章「データ可視化におけるデータインタラクティブティ」は、中小路久美代、山本恭裕、松原伸人、北雄介による。データ解析を行う際には、データ・解の両方を表示する必要がある、ここに大きな問題がある。この問題を解決するための、人間の認知的側面を上手に使う視点や着目点を上手に誘導、探索するためのビジュアルインタラクション技術について紹介する。

第3章「データ研磨を用いたクラスタリングの安定化」は、岩崎幸子による。クラスタリングはデータ解析において広く利用される手法であるが、計算結果がランダムに生成した初期値に依存して大きく変動してしまう。そこで、データ研磨によるクラスタリングの安定化手法について紹介する。

第4章「社会科学におけるデータマイニングの役割」は、羽室行信、岡田克彦、寺田克彦による。データ解析技術が社会科学の分野で活発に活用されるための考察を行い、映画評価コメントのポジ／ネガ分析を通して、機械学習の成果を社会科学分野で活用するための方向性を述べる。

第5章「データ解析における特徴選択」は、申吉浩による。巨大なデータにおいては、特徴量や意味構造を高速に選び出すことが重要となる。そのためには、シンプルかつ効果的な評価方法のモデル化が重要である。そこで、情報量基準を効果的に用いる新しい手法について紹介する。

第6章「個人特性を考慮したグラフ研磨によるクラスタリングと特徴抽出」は、中原孝信による。データ研磨のマーケティングへの応用として、消費者の旅行意向に関するデータを例に、抽象度を高めた意味解釈のしやすい特徴量生成と分類モデルの説明変数への利用について紹介する。

小特集編集チーム	堀山 貴史	澤島 康仁	藤沢 匡哉	荒川 元孝	荒木 徹也	岩田 哲
	小川 祐樹	熊木 武志	高取 祐介	中野 允裕	東野 武史	平井 経太
	宮田 孝富	山口 真悟	山脇 大造	米澤 和也		