

ビッグデータ解析とクラウドソーシング

Big data analysis and crowdsourcing

鹿島 久嗣

Hisashi KASHIMA

京都大学

Kyoto University

kashima@i.kyoto-u.ac.jp

Abstract

機械学習をはじめとするデータ解析技術の進歩が実世界において様々なブレークスルーを起こしている一方で、ビッグデータの解析や処理のプロセスはいまだ極めて労働集約的であり、これらを行う人手をいかに調達するかが重要な課題である。この人的ボトルネックの問題を解消するための有望なアプローチの一つとしてクラウドソーシングの考え方が注目されている。クラウドソーシングを利用して人間による判断や処理をプロセスに組み込むことによって、機械だけでは解決できない、いわゆる「データの外側」を人間の知識や判断によって補うことが可能となる。本講演ではビッグデータ解析・処理をクラウドソーシングで実現するための要素技術となる品質保証技術、クラウドソーシングで収集したデータからの機械学習、クラウドソーシングを利用したデータモデリング事例などを紹介するとともに、セキュリティやプライバシー、人間と機械の協働問題解決といった今後の課題についても述べる。

1 ビッグデータ解析のボトルネック：人材不足

近年、機械学習をはじめとするデータ解析技術は様々な分野における差別化のカギとして認識されつつある。しかしながらデータ解析研究においてしばしば中心的に捉えられるこれらの自動解析技術はデータ解析のプロセス全体からみるとごく一部にすぎない。データの収集や洗浄・結果の解釈などを含むデータ解析プロセスの多くの部分がデータを解析する人間に依存する極めて属人的で労働集約的なものであり、急速に高まるデータ解析需要に反して、データ解析において主導的な役割を果たすいわゆる「データサイエンティスト」の不足が各所で指摘されている。

2 クラウドソーシングの台頭

米国政府が2012年に打ち出した「ビッグデータ研究開発イニシアティブ」の中で、注力すべき情報技術分野として「機械学習」「クラウドコンピューティング」とともに挙げている技術が、インターネットを介して不特定多数の人に仕事や作業を依頼する「クラウドソーシング」である。2005年に登場した米Amazon社の提供するクラウドソーシング市

場 Mechanical Turk はクラウドソーシングの利用を広く浸透させる契機となったが、国内においても同様の商用サービスが多数登場しており、発注側にとってはオンデマンドで労働力を調達する手段として、働き手にとっては場所や時間にとらわれない新しい働き方として注目されている。クラウドソーシングの対象範囲は、マイクロタスク（特別なスキルを要しない比較的単純な労働）から、より高度で専門的な業務を行うものへと拡大しつつあり、データ解析業務はその最たるものであるといえる。

計算機科学分野においても HCI、メディア処理など様々な分野でその利用が拡大しており、従来の計算機を中心としたパラダイムに変革を起こしつつある。

3. クラウドソーシングによるビッグデータ解析

データ解析のプロセスには比較的誰にでも実行可能なデータ収集や電子化のステップ、多少の専門知識やドメイン知識を要するデータクレンジングやキュレーションのステップ、そしてデータ解析手法の高い専門技能を要するモデル化・視覚化のステップへと続く。最終的に得られた結果の評価や解釈には対象ドメインの深い知識が必要であり、また、そもそもの課題立案にはビジネス的な洞察も必要となる。このようにデータ解析のプロセスの各々のステップが要する様々な種類・レベルの専門性や適性を少人数でカバーすることは極めて困難であり、クラウドソーシングによってこれらの人材をオンデマンドで調達し、並列・協調的にプロセスを実行することが、この人的資源のボトルネックの解消へ向けた極めて有望なアプローチとなるだろう。その実現のためにはプラットフォーム技術・品質保証技術・インセンティブ設計・セキュリティ/プライバシー保護技術など様々な観点からの技術開発が必要である[鹿島 14, 鹿島 16]。

参考文献

- [鹿島 14] 鹿島久嗣, 馬場雪乃: ヒューマンコンピューテーション概説. 人工知能学会誌, 29(1) (2014).
- [鹿島 16] 鹿島久嗣, 小山聡, 馬場雪乃: ヒューマンコンピューテーションとクラウドソーシング. 講談社サイエンティフィック (2016). [刊行予定]