

分析前工程におけるデータ研磨スキルの重要性と その体系的な教育プログラムの開発

株式会社帝国データバンク

大里 隆也

株式会社帝国データバンク

菊川 康彬

滋賀大学データサイエンス教育研究センター

保科 架風

1. 分析前工程の教育カリキュラムの現状

技術の進歩や産業の高度化によって、全世界で取り扱われるデータは急増しつつあり、米国の調査会社 IDC によるとそのデータ量は 2020 年までに 40 ゼタバイトに到達すると言われている。このようにデータ量が増大化・複雑化するビッグデータ時代において、膨大なデータを分析可能な形に適切に処理することが正確な知見を獲得するために必要となる。データ分析においてデータの前処理工程が 8 割と言われる中で、「データ研磨」＝「データの構造的な整理」に関するスキルは必要不可欠である。

このようにデータ研磨スキルの重要性が増す一方で、そのスキルを学ぶための体系的な教育カリキュラムは存在しておらず、データ研磨スキルを保有する人材は限定され、また、それらの人材も個人学習によってスキルを習得しているため、習熟度にバラツキが存在する。

2. データ研磨の実践的教育カリキュラムの構築

このような状況を改善するために、株式会社帝国データバンク（以下、TDB）が持つ実践的なデータ研磨スキルを、国立大学法人滋賀大学（以下、滋賀大）が持つデータサイエンス教育のプラットフォームと連携することで、体系的な教育カリキュラムの開発に取り組んでいる。

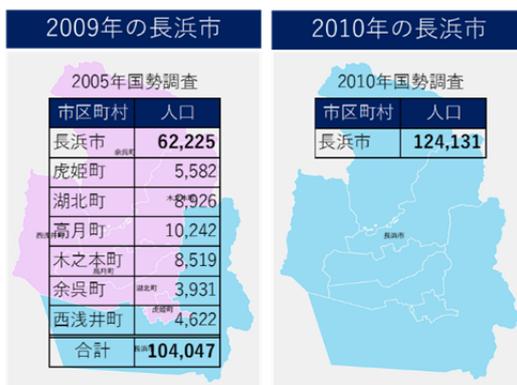


図 滋賀県長浜市を中心とした市区町村統廃合

TDB では、内閣府が 2015 年 4 月に公開した地域経済分析システム（RESAS）に搭載するための官民様々な 60 種類以上にも及ぶデータの研磨を、約 3 年半に渡り行った実績があり、様々なデータ形式に対応した研磨や、データ研磨プロセスのチェック体制を構築してきた。左図はデータ研磨の一例であるが、時系列データを扱う際には市区町村の統廃合を考慮しなければならない。具体的には、滋賀県長浜市は 2010 年に 6 つの町と合併しており、2005 年の国勢調査の人口を再集計する必要がある。

当日の報告では、滋賀大と TDB が共同で開発するデータ研磨スキルの教材と 2018 年 10 月～11 月に滋賀大学データサイエンス学部向けに実施する講義内容の紹介を軸に、データサイエンスにおけるデータ研磨スキルの重要性について報告する。

参考文献

- [1] 河本 薫 (2013), 会社を変える分析の力, 講談社現代新書
- [2] 柴田 里程 (2001), データリテラシー, データサイエンス・シリーズ
- [3] 竹村 彰通 (2018), データサイエンス入門, 岩波新書