

大規模オープンオンラインコース (MOOC)での海外統計教育・データサイエンス教育の展開

櫻井尚子 (東京情報大学) 和泉志津恵 (大分大学) 深澤弘美 (東京医療保健大学)

1. はじめに

MOOC (Massive Open Online Courses) は 2012 年前半から情報通信技術の進展に伴い本格的にサービスが開始された高等教育のプラットフォームである。MOOC の代表格である Coursera では、2015 年 6 月中旬現在、1049 のコースを提供し、1350 万強の人々が受講している。コンテンツ提供側の高等教育機関は 121 にのぼる。本発表では海外 MOOC 内の特徴的な統計およびデータサイエンス教育プログラムを概説し、国内におけるデータサイエンス教育の今後の展開について検討する。

2. MOOC での統計教育の内容

現在、海外では情報分野に特化した MOOC である UDACITY や Coursera, FutureLearn 等、多くのプラットフォームで統計に関連した講座が開講されている。UDACITY が提供する Introduction to Data Science コースは、データサイエンスの基本である「データオペレーション」「統計学・機械学習に基づくデータ解析」「データ通信およびデータの視覚化」「データの尺度化-ビッグデータ活用」の 4 つのテーマを含む約 2 か月のコースである。有料 (199 ドル/月) であるが、一部無料でも提供されている。

また、Introduction to Data Science コースは上級コースである Nanodegree : Data Analyst の一環である。1 つのテーマを深く掘り下げることで、データサイエンス全体のイメージをつかむことに主眼を置いている。授業内容は 1~5 章に分かれ、各章とその内容の説明は以下に示すとおりである。図 1 に 3 章のビデオの 1 例を示す。

- 1 章 データサイエンス入門：現実のエンジニアの講義等
- 2 章 データラングリング：RDB, API, 欠測値, 補完等
- 3 章 データ解析：統計学, 検定, 機械学習, 決定係数等
- 4 章 データ視覚化：視覚刺激, Python による図表作成等
- 5 章 MapReduce : Mapper/Reducer, 簡易ツール Pig 等



図 1 UDACITY データサイエンス (3 章)

3. まとめ

ビジネスの各領域では既に大量に蓄積されたデータの活用がオンデマンドである。大量で扱いにくいデータ形式を持つさまざまな生のデータをコード化して分析対象に整形する必要がある。たくさんの工程をこなさねばならないデータサイエンスはチーム連動ワークの最たるものであり、一つが欠けても成果は出ない。統計学の知識を持つ人がデータサイエンスの各プロセスを学び、現実の仕事の中で生かすことでデータサイエンスはさらに発展する。発表時には、教育コースの中身と評価の方法について詳述の予定である。

【参考文献・参考 URL】

- [1] 櫻井尚子・和泉志津恵・深澤弘美(2015): MOOC での certification 問題について. 統計数理研究所共同研究レポート 335, 73-78.
- [2] 櫻井尚子(2014): MOOC を用いた統計教育について. ESTRELLA, 8-14, (財)統計情報研究開発センター.
- [3] <https://www.coursera.org/>, <https://www.edx.org/>, <https://www.udacity.com/>