

データサイエンスのモデル教材開発の取組み

滋賀大学データサイエンス学系・教授
大学院データサイエンス研究科・副研究科長 清水 昌平



1. はじめに

2016年12月に、文部科学省より数理およびデータサイエンスに係る教育強化の拠点校として、北海道大学、東京大学、滋賀大学、京都大学、大阪大学、九州大学の6大学が選定されました。これら拠点校6校がそれぞれ、数理・データサイエンスを中心とした全学的・組織的な教育を行うセンターを整備し、各大学内での数理・データサイエンス教育の充実に努める構想です。さらに、その後、拠点校6校が個別に取組むのではなく6大学コンソーシアムを形成して、全国の大学に取組み成果の波及を図ることとなりました。それぞれの大学の周辺地域の数理・データサイエンス分野における拠点として他大学の数理・データサイエンス教育の強化に貢献することも期待されています。

6大学コンソーシアムは特に、カリキュラム、教材、教育用データベースに関して分科会をつくり活動しています。それぞれの分科会に6大学の教員が1名ずつ参加しています。例えば、カリキュラム分科会は全国的なモデルとなる標準カリキュラム・教材の作成およびその標準カリキュラム・教材の他大学への普及方策の検討を行っています。また、教材分科会は、数理・データサイエンスに関する教科書シリーズの企画および編集、そして各拠点大学が作成したeラーニング教材のポータルサイトの作成を行っています。そして、データベース分科会は、教育に適した実データに関する情報を集めデータベースを作成し公開するための活動をしています。

また、2019年には、拠点校6大学および6大学コンソーシアムによる数理・データサイエンス教育を全国の大学への普及・展開するスピードを加速させるため、新たに協力校20校が選定されました。今後、拠点校6大学と協力校20校が連携していきます（協力校20校：北見工業大学、東北大学、山形大学、筑波大学、宇都宮大学、群馬大学、千葉大学、お茶の水女子大学、新潟大学、長岡技術科学大学、静岡大学、名古屋大学、豊橋技術科学大学、神戸大学、島根大学、岡山大学、

広島大学、愛媛大学、宮崎大学、琉球大学)

本稿では、滋賀大学が主査を務める教材分科会の活動について紹介します。

2. 教材分科会

教材分科会の目的は、全国的なモデルとなる教材を協働して作成・普及に取組むことです。具体的な審議事項は、

- * 教科書シリーズの企画編纂
- * 各大学のeラーニング教材、講義動画等の統合的配信方法の検討
- * 普及方法やその他の可能性の検討

の3つです。分科会の委員は拠点6大学よりそれぞれ、

遠藤 俊徳	北海道大学	
寒野 善博	東京大学	
清水 昌平	滋賀大学	(主査)
中野 直人	京都大学	
朝倉 暢彦	大阪大学	
内田 誠一	九州大学	(副主査)

が務めています。

目標とするアウトプットは、

- 1) データサイエンスに関する教科書シリーズの編集および出版
- 2) 各拠点の教材情報を共有し、相互利用を検討するためにeラーニング教材・講義動画等のポータルサイトの立ち上げ

です。この2つについて次節以降で、紹介します。

3. 教科書シリーズ

データサイエンス入門シリーズは講談社より刊行されます。本シリーズの編集委員会は拠点6大学の教員が務めます：

竹村彰通（編集委員長、滋賀大学）、狩野裕（大阪大学）、駒本文保（東京大学）、清水昌平（滋賀大学）、下平英寿（京都大学）、西井龍映（九州大学）、水田正弘（北海道大学）。

特徴は、「数理・データサイエンス教育強化拠

点コンソーシアム」のスキルセットに依拠しており、具体的、体験的に学べる応用例および練習問題を収録している点です。またフルカラーで印刷され見やすい構成になっています。

竹村彰通（編集委員長、滋賀大学）による刊行に向けての言葉では次のように刊行の趣旨が述べられています：

「情報通信技術や計測技術の急激な発展により、データが溢れるように遍在するビッグデータ時代が到来しました。そして、データを処理・分析し、データから有益な情報をとりだす方法論であるデータサイエンスの重要性が広く認識されるようになりました。しかしながら、日本ではこの分野の人材育成が非常に遅れています。データサイエンスの知識は、ほとんどの分野で必要とされるものであり、大学の教養課程においても文系・理系を問わず多くの学生が学ぶことが望まれます。文部科学省も「数理及びデータサイエンスに係る教育強化拠点」を選定し、拠点校はコンソーシアムを作り、全国の大学におけるデータサイエンス教育の充実をはかっているところです。

データサイエンス教育は、従来からの統計学教育と、データサイエンスに必要とされる情報処理教育の二つを基礎としますが、ビッグデータの時代の中では、データという共通点からこれらの二つの分野を融合的に扱うことが必要です。また例題として、インターネット上のオープンデータや機械学習手法評価のためのデータなどを扱うことにより、今後求められるデータサイエンスの応用法を示すことも重要です。これらの点で、今回のデータサイエンス入門シリーズは、従来の統計学やコンピュータ科学の教科書とは性格を異にしており、今後のデータサイエンス教育にふさわしい教科書シリーズとなることを目指すものです。」

このように、本シリーズの対象は主に学部低年次の学生であり、文系理系を問わず教科書として利用してもらえるものと思います。

第1期は、次の3冊から成り、2019年8月29日に刊行予定です：

第1期

- 『データサイエンスのための数学』
椎名洋・姫野哲人・保科架風（著）清水昌平（編）
A5・288頁・予価2,800円・ISBN 978-4-06-516998-8
- 『データサイエンスの基礎』
濱田悦生（著）狩野裕（編）A5・192頁・予価2,200円・ISBN 978-4-06-517000-7
- 『最適化手法入門』
寒野善博（著）駒木文保（編）A5・256頁・予価2,600円・ISBN 978-4-06-517008-3



写真1 データサイエンス入門シリーズ

第2期は、次の3冊を2019年11月下旬刊行を予定しています：

第2期

- 『統計モデルと推測』
松井秀俊・小泉和之（著）竹村彰通（編）
 - 『Pythonで学ぶアルゴリズムとデータ構造』
辻真吾（著）下平英寿（編）
 - 『Rで学ぶ統計的データ解析の基本』
林賢一（著）下平英寿（編）
- そして、次の4冊が続刊予定です：
- 『データサイエンスのためのデータベース』
村井哲也・吉岡真治（著）水田正弘（編）
 - 『スパース回帰分析とパターン認識』
梅津佑太・西井龍映・上田勇祐（著）
 - 『モンテカルロ統計計算』
鎌谷研吾（著）駒木文保（編）
 - 『テキスト・画像・音声データ分析』
西川仁・佐藤智和・市川治（著）清水昌平（編）

4. 教材ポータルサイト

北海道大学、東京大学、滋賀大学、京都大学、大阪大学、九州大学の6大学から成る数理・データサイエンス教育強化拠点コンソーシアムでは、参画大学が協働して、全国的なモデルとなる教材の作成・普及に取り組んでいます。そこで、各大学のeラーニング教材や講義動画等の情報を集約す



写真2 数理・データサイエンス教育強化拠点コンソーシアムのWebサイト

るためにポータルサイトを作成しました。数理・データサイエンス教育強化拠点コンソーシアムのWebサイト (<http://www.mi.u-tokyo.ac.jp/consortium/index.html>) にバナーがあり、そこから辿ることができます。それぞれの拠点大学の数理・データサイエンスに関する教育研究センターのeラーニング教材情報のページへリンクが張られています。各大学が作成した個々のコンテンツの利用条件については各拠点大学にお問い合わせください。

5. 教材紹介

拠点6大学である北海道大学、東京大学、滋賀大学、京都大学、大阪大学、九州大学が作成した教材のうち、ポータルサイトにある教材を紹介します（令和元年7月現在）。

北海道大学

数理・データサイエンス教育研究センターでは、ブラウザ上で受講できる演習システムを作成しています。Pythonや数学に関して自習することが可能です。

東京大学

数理・情報教育研究センターでは、最適化手法や時系列解析など10以上の講義動画が公開されています。

滋賀大学

データサイエンス教育研究センターでは、講義動画を拠点校および協力校に提供可能です。「Python入門」、「データ構造とアルゴリズム」、「多変量解析」、「Rによるデータ解析入門」があります。また全学共通教育科目の「確率への招待」の講義資料を公開しています。

また、やや専門的なトピックとして、スパース推定法による統計モデリングと因果推論による講義動画を作成しています。また、MOOC教材の作成も行なっています。MOOC教材については、次の節で詳しく紹介します。

京都大学

高等教育院附属データ科学イノベーション教育研究センターでは、「臨床研究者のための生物統計学」と題する生物統計に関する講義動画を公開しています。

大阪大学

数理・データ科学教育研究センターは、医療データの収集・統合・分析・解釈の知識・スキルについてのe-Learning教材を作成しています。

九州大学

数理・データサイエンス教育研究センターでは、文系理系の双方の学生を対象にしたデータサイエンス入門科目「データサイエンス概論第一」および「データサイエンス概論第二」の講義資料を公開しています。

6. 滋賀大学のMOOC教材

最後に、筆者が所属する滋賀大学のMOOC教材を紹介します（Massive Open Online Course：大規模公開オンライン講義）。このMOOC教材は、データサイエンスを学ぶ大学生向けe-learning教材です。データサイエンス分野における国内屈指の教育研究拠点のひとつである滋賀大学の専任教員に加え、知見や技術を日々活用されている実務家等の外部講師がビジネスの視点から講義します。

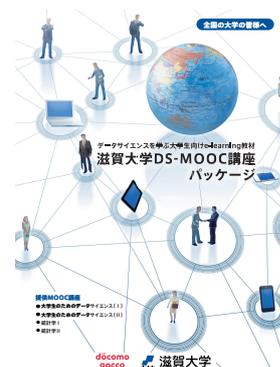


写真3 滋賀大DS-MOOC講座パッケージ

データサイエンスの基礎は統計学と計算機科学です。データから価値を生み出すためには、まずコンピュータを使ってデータの取得・整形などの処理を行い、統計学の手法を用いて分析します。その結果得られた知見を現実の問題解決に役立てることにより価値創造を実現します。本パッケージのうち「大学生のためのデータサイエンス(I)、(II)」では、コンピュータを用いたデータ処理・分析とその応用事例を重視し、統計パッケージのRや機械学習の機能の豊富なPythonを学びます。「統計学I、II」では、統計グラフの使い方から、統計的検定や回帰分析の方法まで統計学の標準的な手法と考え方を学びます。4つの講座を学ぶことにより、データサイエンスの全貌をつかむことができます。

このパッケージでは、授業管理および成績管理が行えます。インターネット上で、一般向けに無料で提供しているMOOC講座もありますが、個人毎の学習進捗状況等を管理・閲覧することはできません。学習の進捗には個人差がありますので、フォローアップをするには、教員が管理する仕組みが必要です。本パッケージでは、授業管理、成績管理が行えます。費用は、4講座セットで受講者1人当たり1,500円（税別）、利用期間6ヵ月です（令和元年7月現在）。

MOOCに関するお問い合わせ先

一般社団法人近江データサイエンスイニシアティブ
〒522-8522 彦根市馬場一丁目1番1号（滋賀大学内）
ds-mooc@biwako.shiga-u.ac.jp