

# 応用生物学課程

## “生命科学のデータサイエンス演習”の紹介

北 島 佐紀人\*  
sakito@kit.ac.jp

### 1. はじめに

バイオ分野の少なからぬ数の研究者は、実験で得たギガバイトときにはテラバイトに及ぶ巨大データから計算機の力を借りて新しい知識を発見しようとする。対象となるデータの形式は、塩基配列のような文字情報、遺伝子発現・代謝物量のような数値情報、診断機器と顕微鏡の画像・映像情報など多岐にわたる。必要なスキルは、統計学的あるいはそうでない計算機技術を駆使し、そのためのソフトウェア群を使いこなし、そのためのちょっとしたプログラムを書くことができる、というようなものだろうと私は理解している。

### 2. 応用生物学課程での関連の教育

京都工芸繊維大学の応用生物学課程における関連の講義・演習として、1年次には統計数理(2単位)、情報処理演習(2単位)、生物統計学(2単位)を実施している。これらのうち1つ目と3つ目は数理統計学的内容で、いずれもPCを使わない講義が主体である。2つ目はMicrosoft Officeの使い方の演習が主体である。PCを使う統計学的内容としては、2年次の学生実習の一部として、実験計画法のような内容をExcelとRを使って実施している(表1。担当者は私、北島)。1日だけの演習だが、3～5時限の実施なので3回分に相当する。

3年次向けには、2020年度前期から本稿の主題である“生命科学のデータサイエンス演習”を開講した(2単位)。正確には、当初“バイオインフォマティクス演習”と呼んでいたが2022年度に前述の名称に変更した。

### 3. “生命科学のデータサイエンス演習”の内容

本演習で学ぶ内容を表2に示した。15回中12回を応用生物学系所属の私、北島が担当し、画像・映像編集の3回を同じく応用生物学系の井上喜博先生が担当している。以下、私の担当の13回について書く。

本演習では、何を調べる手法なのか、どのようなデータを対象にする手法か、計算結果をどのように解釈するべきか、を理解すること、そして実践してみることに重点を置いている。数学的な背景の説明が蔑ろにされているが、そこを頑張ると実施時間内に実践に到達できないのでやむを得ない(担当教員の私には理論を教えるのは荷が重いというのも理由の一つである)。統計と機械学習に加えて、塩基・アミノ酸配列解析というバイオ分野固有のデータ解析を含む点は応用生物学課程ならではの特徴である。RとPythonを使う回がそれぞれ5回ずつあるが、文法の詳しい理解よりも、ともかく使ってみる、結果を出力させてみる、ことを重視している。履修者がプログラミングを面白いと思えれ

表 1. 応用生物学課程 2 年次後期学生実験の一部で学ぶこと

実施回	学習内容のキーワード	使用するソフトウェア
	実験計画法 正規分布 ヒストグラム 平均 分散 標準偏差 相関 回帰直線 F 検定 t 検定 分散分析 Tukey の HSD 法	Excel R

\* 応用生物学系 准教授

表 2. 応用生物学課程 3 年次前期 ”生命科学のデータサイエンス演習 ” で学ぶこと

実施回	学習内容のキーワード	使用するソフトウェア
1	多重検定 Benjamini-Hochberg 法	Excel R
2	ノンパラメトリック検定 Wilcoxon の順位和検定 Kluscal-wallis 検定 Bonferoni の調整法	R
3	主成分分析	R
4	Fisher の直接確率検定 Gene ontology のエンリッチメント解析	R
5	R の簡単なプログラミング繰り返し文 if 文 塩基配列・アミノ酸配列の取得と相同検索 ドメイン検索	R BLAST SMART
6	塩基・アミノ酸配列の多重アラインメント 系統樹	MEGA
7	塩基配列の編集と人工遺伝子配列の作成	ApE
8	Python の簡単なプログラミング 繰り返し文 if 文	Python
9	Python と Biopython による塩基配列解析	Python
10	Python と Biopython による塩基配列解析	Python
11	Python による機械学習 ランダムフォレスト	Python
12	Python による機械学習 ランダムフォレスト	Python
13	画像データの編集	ImageJ Photoshop
14	画像データの編集	Photoshop
15	動画データの編集	Realplayer

ば、自主的に勉強していけるだろうし、この程度でも学生のうちに体験しておけば、近い将来業務でプログラムを自分で書くことになったとしても心理的な障壁はだいぶ低くなるであろう。

#### 4. 実施形態

コロナ禍1年目の2020年度前期の開講となったため、当初から実施方法に工夫が求められたのであるが、それもきっかけとなり3年目となる2022年度までに、アクティブラーニング、PBL（問題解決型学習）、反転学習、リモートと対面の同時実施、そしてクラウドベースのグループワークというてんこ盛りの実施形態へと発展した（図1）。ポイントは、(1) 履修者はテキストと解説動画で予習する。(2) 当日は教員が用意した課題を毎回メンバーの異なるグループで取り組む。(3) グループ課題のレポートは、クラウドのMicrosoft one drive に置いた共有ファイルに記述するため、演習室での履修者とリモート参加の履修者がシームレスにリアルタイムに共同作業できる。(4) メンバー同士は対面であるいはwebex等を用いて議論できるので理解度の確認と向上が可能。(5) 教員は演習時間中、各グループの共有ファイルを閲覧して、直接あるいはwebexを使ってリアルタイムに助言を与える。

映像の準備など、教員が事前に準備しなくてはならないことは多いが、演習時間中は割と暇である。

以下は解決すべき問題点である。(6) 履修者が真面目に予習するためには個人所有のPCに必要なソフトウェアを自分でインストールしなくてはならない。これは履修者個人のPCスキルに強く依存している。(7) 事前に真面目にテキストを読み動画を視聴したかを確認する良い方法がない。小テストで理解度を調べることは可能だが、履修者に過度の負担となる恐れがある。

#### 5. 履修者からの評価

大学が実施する期末の授業評価アンケート調査に加え、出席確認を兼ねて毎回moodleのフィードバック機能を使って履修者に自由にコメントをさせている。“手作りのテキストがわかりやすかった”、“機械音痴の私でも取り組めた”、“毎回課題が出されていたので実際に自分の手を使って理解を深められた”、“受講してよかった”、“機械学習がおもしろい”、“Pythonを今後も勉強したい”というような前向きなコメントが寄せられている。クラウドでのグループワークとした点については、履修者からの意見は特に聞こえないが、演習室で見ていると履修者同士で熱心に相談しながら取り組んでいる様子を感じられる。一方、留意すべき点としては、2020、2021、2022年度の履修者数はそれぞれ18、28、14人で、応用生物学課程の在籍者が50人近くいることを考えるとあまり多くない。シラバスとテキストを見

て、興味をもった且つ自分にも出来そうだと  
思った学生だけが履修しているのであろう。

数年かけて準備してきたので、本演習は私に  
とって思い入れのある演習である。今後も内容

と実施方法を改善して、履修してよかったと履  
修者たちが卒業後にも思えるような演習に仕立  
てていきたい。

### Moodleからリンクされたテキストと動画で事前に学習

The Moodle page shows course details for '第8回 6/3 (金) 第5限 5IS (担当者: 北島)'. It includes instructions for students to download Python files and watch a video. The YouTube player shows a video with a presenter and a code editor.

### 当日はグループで演習課題

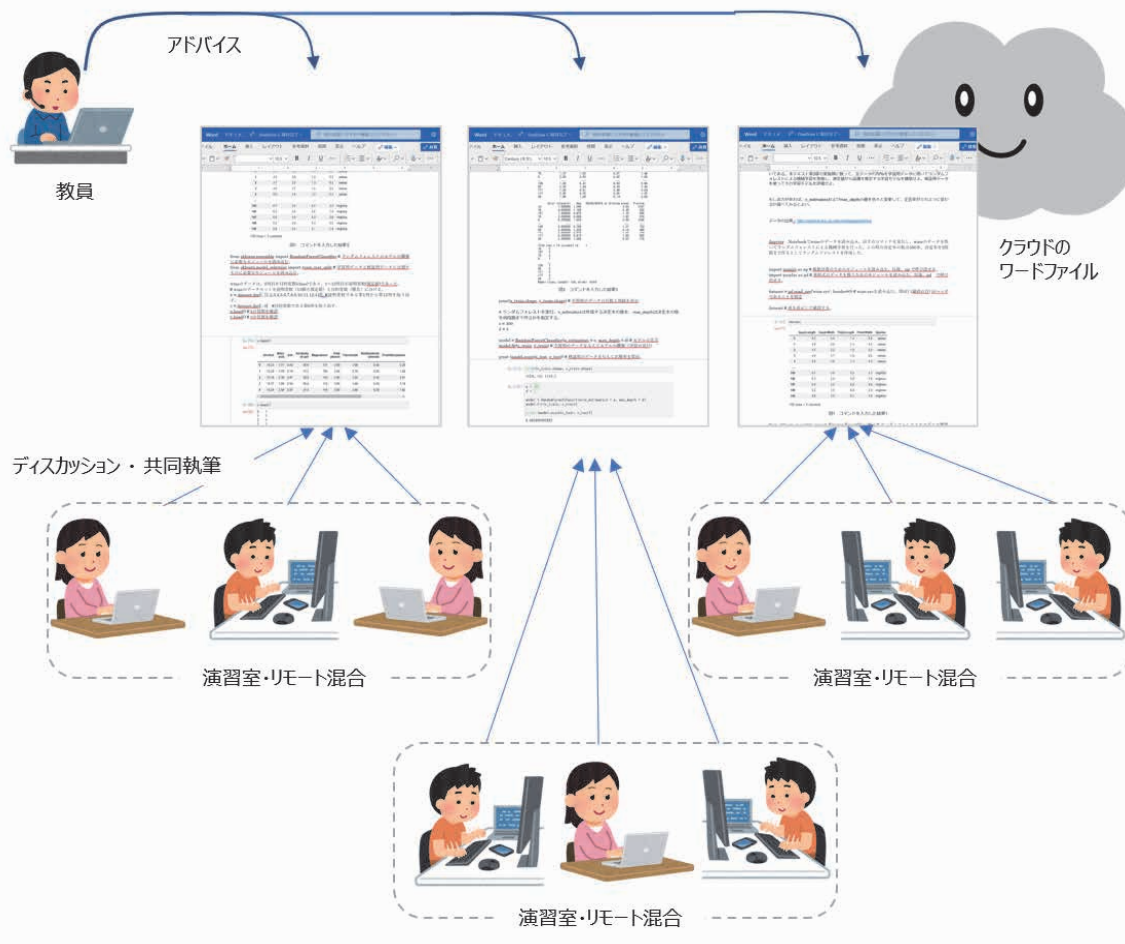


図1 ”生命科学のデータサイエンス演習の実施の様子