

生命科学における データサイエンス教育のプランニング

SATテクノロジー・ショーケース2020

■ はじめに

今日では科学、工学の非常に幅広い分野にビッグデータ解析の潮流が押し寄せているが、生命科学分野では1980年代のヒトゲノム計画や遺伝子発現観測手法の登場などにより、その中でも最も早い時期からデータ肥大化の悩みを抱えてきた。悩みの中でも大きなものは二つである。それは(1)その時代の計算機資源での解析が困難な規模のデータが、解析の速度を上回って蓄積されていくこと、および(2)解析の知見を持つ人材が、データ観測をする分野内に乏しいこと、である。前者は計算機のハードウェア及びソフトウェアの両方の技術の発展により解消される一方で、観測技術の発展による大規模化が進むという、進歩のイタチごっこの状況がこれからも続くだろう。この進歩の一方で後者の問題は、教育によって解決を図らねばならない。産総研ではこれまでも科学技術振興調整費などによる社会人及びプロの研究者向けの座学及び実習コースを実施してきたが、平成28年度に早稲田大学と共同で同大学内に研究ラボを設置し、共同研究による学生指導および連携大学院での講義と平行して、学部生に向けたデータ解析の実習を共同で行っている。

■ 活動内容

1. 連携大学院

座学講義とOJT的学生指導を行っている。講義では産総研の4人の研究員が講師を務め、それぞれ最新の研究成果を背景知識とともに解説することで具体的なウェット&ドライ連携の実例を紹介している。

2. 学生実習

主に分子生物学分野を対象として、統計解析のもっとも基礎的な作業の進め方、ゲノム/核酸配列解析からたんぱく質立体構造解析までの情報科学的解析を2コースに分け、それぞれ二年次および三年次を対象として実施している。受講する学生もデータ解析手法が重要性を増していることは知られており、ここ数年受講者数は増加傾向にある。

● バイオインフォマティクス実習 I

2コマを一回として、統計解析の基礎的な考え方と作業の進め方に6コマ、配列解析の理論と実習に2コマ、文献検索に2コマ、分子動力学法によるシミュレーションに4コマをあて、生命科学分野以外でも有用な概念理解と技術

の習得、最新技術の理解に必要な基礎知識の習得を目指す。

● バイオインフォマティクス実習 II

より発展的な統計解析の作業について2コマ、遺伝子発現解析に2コマ、次世代シーケンサーのデータ解析(アセンブリとマッピング)に4コマをあて、研究現場での実際の作業に役立つスキルの習得を目指す。

3. 学部生講義

情報科学を専攻する学生に対し、応用分野の一つである生命科学分野でどのようにITが駆使されているか、具体的なアルゴリズムとその適用の方法、バイオデータ特有の問題点、現在の着目されている課題は何かなどを具体的に紹介する講義を行うことで興味の喚起を図っている。例年受講生は100名を超えており、情報科学を専門としている学生からも生命科学が興味深い対象であることをうかがわせる。



実習資料例

代表発表者 富永 大介(とみなが だいすけ)
所 属 国立研究開発法人産業技術総合研究所
生体システムビッグデータ解析オープンイノベーションラボラトリ (CBBDOIL)
問合せ先 〒169-8555 東京都新宿区大久保3丁目4-1
早稲田大学63号館5階 20 号室
TEL: 03-3599-8044

■ キーワード: (1) バイオインフォマティクス
(2) 統計解析
(3) データ科学

■ 共同研究者: 竹山春子(早稲田大学)
富永大介(産総研)
由良尊敬(早稲田大学)
森一樹(鹿児島高専)