
バイオ産業情報化コンソーシアム
「産業界横断的なバイオ分野データサイエンス
関連人材」を育成する講座の開発」

背景と事業の狙い

背景

電子カルテやレセプトデータ等の膨大な診療情報のビッグデータ解析、がんゲノム医療における遺伝子変異のシーケンス解析、自動病理画像診断を可能とするAIシステム構築、多様なタンパク質相互作用ネットワーク解析によるAI創薬などの研究開発が精力的に行われているが、これらの技術を駆使できる技術者、すなわちデータサイエンティストと呼ばれる人材が不足しており、人材育成が急務となっている。この人材不足は今後ますます加速すると考えられる。産業界におけるバイオインフォマティクス技術者の人材育成を目的とした「バイオインフォマティクス技術者認定制度」を平成16年度から弊社が、平成19年度からは日本バイオインフォマティクス学会が実施しているが、現在も毎年200名を超える受験者がいる。その半数強が社会人であることから、産業界におけるニーズが伺える。

事業の狙い

東京大学大学院農学生命科学研究科では、「アグリバイオインフォマティクス教育研究プログラム」、東京大学大学院新領域創成科学研究科では、製薬、バイオテクノロジー関連企業と連携した社会連携講座DSTEP (Data Scientist Training /Education Program)、東京医科歯科大学では、文部科学省の科学技術人材育成費補助事業「データ関連人材育成プログラム」として製薬やIT企業も参画した「医療・創薬データサイエンスコンソーシアム」を設立し、医療・創薬・ヘルスケア分野向けの教育プログラムを実施している。

これらの教育プログラムは、主に博士後期課程の学生を対象とした教育プログラムで専門的な内容も多く、通年の受講で単位取得を目指しているため、社会人が受講するにはハードルが高い。本事業では、これらの博士課程向けの教育プログラムを社会人向けに発展させ、ビッグデータやAIの医療応用といったより実践的な企業ニーズを踏まえた「**産業界横断的なバイオ分野データサイエンス関連スキル標準**」と**各産業界の実情に合わせた個別の「専門スキル標準」**を作成する。このスキル標準に基づき、製薬、ヘルスケア、農業などの産業界の社会人を対象とした教育プログラムを新たに開発し、受講者側の意見を聴取し、その有効性を実証することを目的とする。最終的には教育機関以外の事業者がビッグデータやAIの医療応用といった新しいタイプのデータサイエンスの講習を実施できる体制づくりを目指す。

実施内容(1.企業ニーズ調査を行った)

JBIC、JBAの会員企業（バイオ関連企業）に送ったアンケートの内容

- ◆ AIやビッグデータ解析などの先端的ICT技術の活用に関しまして、具体的にどのような業務で、どのような課題がございますか？
(具体的な事例を記載ください)
- ◆ 上記に関してどのようなスキルを持った人材が必要でしょうか？
(基礎的な知識、及び専門的知識について具体的に記載ください)
- ◆ 上記の人材育成のため、どのような講座を開催してほしいでしょうか？
- ◆ 専門知識とITスキルのバランスはどのようにお考えでしょうか？
- ◆ 現場（企業）での現状の人材育成方法についてお教えてください（特に課題）
- ◆ 大学等で実施されているデータサイエンス関連の教育プログラムを受講されている、或いは受講を検討されている場合は、そのカリキュラムに対するご意見があれば記載ください。
- ◆ 社会人向けの人材育成プログラムに関するご要望やご意見について記載ください。
(年間講義、集中講義、e-learning、テキストなど)

！ 具体的企業ニーズが引き出せるように！！

企業が困っていることは何なのだろう？

具体的にどのような人材が必要なのだろう？

具体的にはどのような講座が必要なのだろう？
専門知識とのバランスは？

どのような講座であれば企業は教育プログラムに参加して下さるのだろうか？

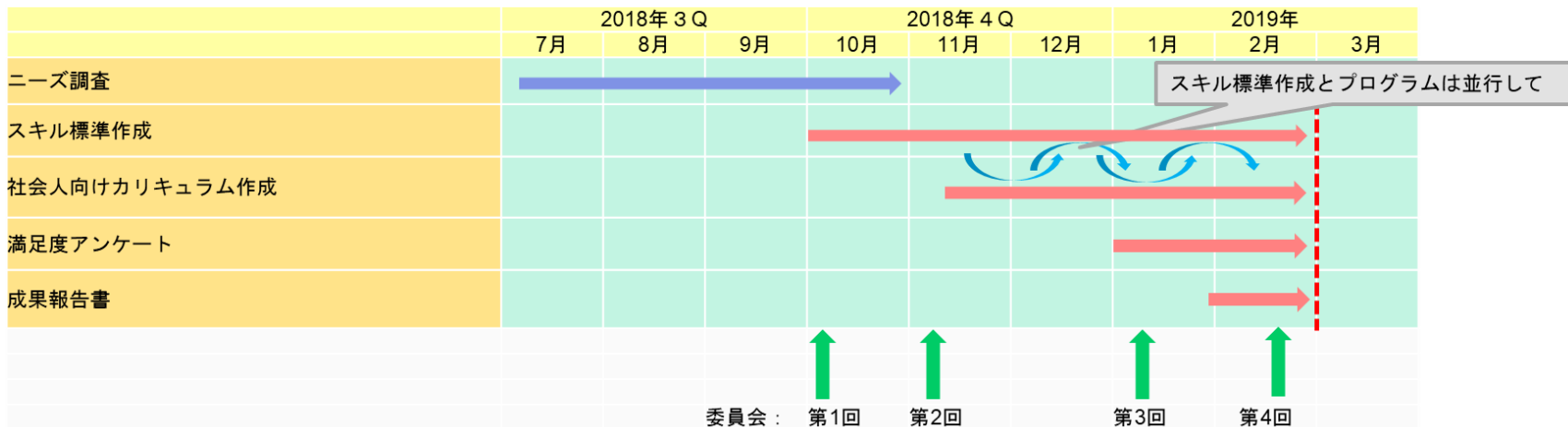
実施内容(2.3) バイオ分野データサイエンス関連スキル標準策定、並びに 社会人向け教育講座プログラムにむけて

- ✓ 企業ニーズ調査:興味深い回答を寄せてくれた企業を対象に個別訪問を行い、さらに具体的な企業の要望をヒアリングした。
- ✓ 個別訪問を行った企業を中心にスキル標準設定委員会への参画を要請した。
- ✓ 最終的に参加を承諾してくださった企業の皆様、大学の皆様、計10名を迎えて
バイオ関連スキル標準策定委員会を発足・開催し、事業の進め方の討議を行った。

【メンバー】

アカデミア	東京医科歯科大学	田中博先生、神沼英里先生
	東京大学大学院農学研究科	清水謙多郎先生、門田幸二先生
	国立遺伝学研究所	有田正規先生
	長浜バイオ大学	白井剛先生
企業	田辺三菱製薬	斎藤隆太氏
	第一三共	長谷川雅司氏
	出光興産	高嶋明人氏
	島津製作所	小林裕氏
	味の素	渡辺信弘氏
JBIC (事務局)		永田晃一、南多善、島田神生
JBA		中川智氏

実施内容(2.3) バイオ分野データサイエンス関連スキル標準策定、並びに 社会人向け教育講座プログラムにむけて



実施内容(3.4.社会人向け教育講座プログラムの開催とアンケート実施)

バイオインフォマティクス講習会/データサイエンティスト人材育成講座 (e-learning方式) (東京医科歯科大学)

主催：東京医科歯科大学データ人材育成プログラム、一般社団法人バイオ産業情報化コンソーシアム (JBIC)

配信期間：2019年2月12日 (火) ~2月19日 (火)

URL：<https://sites.google.com/site/mddsc1902ss/>

バイオインフォマティクス講習会 (集中講義方式) (東京大学)

主催：東京大学大学院農学生命科学研究科・アグリバイオインフォマティクス教育研究ユニット、一般社団法人バイオ産業情報化コンソーシアム (JBIC)

開催日時：2019年2月19日 (火) 13:00~17:30 (希望者のみ18:45まで延長して終了)

場所：東京大学 農学部2号館2階 化学第1講義室 (東京都文京区弥生1-1-1)

参加者：参加者*：79名(欠席3名)

タイムスケジュール

11:00-12:00 自習 (任意参加)

12:00-13:00 休憩

13:00-13:05 趣旨説明

13:05-15:30 前半「R基礎およびパッケージの利用法」

15:45-17:30 後半「RStudioの利用、Rによる機械学習 (分類や回帰)」

* 20営業日の申込期間を設定していたが、申し込み開始10営業日で120名ほどの受講者が集まり、申し込みサイトをクローズした。

成果：概要

達成したい状態

東京医科歯科大学及び東京大学農学系研究科 博士課程向けの教育プログラムを社会人向けに発展させ、ビックデータやAIの医療応用といったより実践的な企業ニーズを踏まえた「**産業界横断的なバイオ分野データサイエンス関連スキル標準**」と各産業界の実情に合わせた個別の「**専門スキル標準**」を作成する。このスキル標準に基づき、製薬、ヘルスケア、農業などの産業界の社会人を対象とした教育プログラムを新たに開発し、受講者側の意見を聴取し、その有効性を実証することを目的とする。最終的には教育機関以外の事業者がビックデータやAIの医療応用といった新しいタイプのデータサイエンスの講習を実施できる体制づくりを目指す。

実際の達成度

企業ニーズ調査報告書

- ✓ 完了（34社より受領）し解析を行った

バイオ分野データサイエンス関連スキル標準

- ✓ 現存のITスキル標準に、バイオインフォマティクスを追加する
- ✓ ドラフト版の作成を完了した

社会人向け教育講座プログラム

- ✓ 東京大学大学院農学生命科学研究科にて集中講義、東京医科歯科大でe-learningを実施した

受講者満足度アンケート調査結果

- ✓ 上記、講座受講者にアンケートを実施
- ✓ 好評であったと考える

理由・改善/発展の方向性

幅広いバイオ関連企業に企業ニーズ調査を実施し、現状に即した形でニーズを抽出できたものとする。そのうえで、産学官連携でバイオ関連バイオインフォマティクススキル標準（ドラフト）を作ったことは、産学官での共通言語、物差しができたという非常に大きなメリットであると考えられる。

企業ニーズ、バイオインフォマティクススキル標準に基づくプログラムを作成し、実際に開講したところ産業界から高い関心を持って受講者が集まったことから、実証作業は成果を出したものと考えられる。

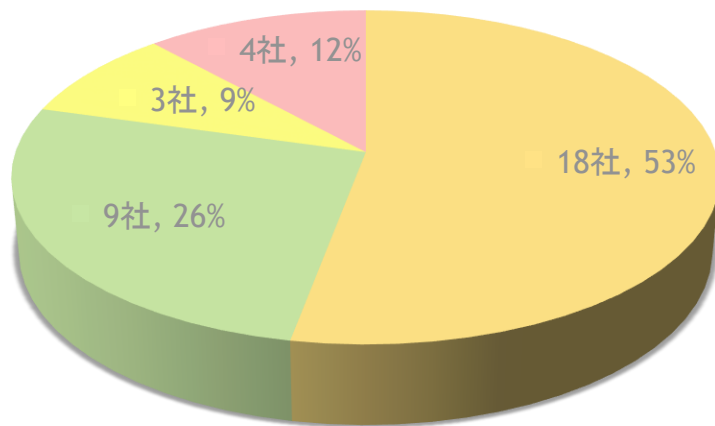
その一方で、実施できた講座（プログラム）は、当初想定したもの比べ、内容が若干易しく講座開講期間も短くなった。ただ、受講者アンケートを見ると、大多数が本講義に関して好意的な回答が大多数であり、実証作業が機能したものと考えられる。

しかし、プログラム実施者が、本プログラムのようなデータサイエンティスト人材育成を継続的に実施していくための問題点も見つかった。産学官、すべてがwin-winの形で人材育成を行うための仕組みづくりを早急に検討すべきである。

成果：詳細

(1) 企業ニーズ調査の実施（解析）

回答企業（10月末現在）34社



■ 1 ■ 2 ■ 3 ■ 4

データ活用

- 社内で蓄積、構築されつつあるビックデータ（必ずしもビックではないものも含め）、および社外のビックデータ（リアルワールドデータを含め）が、ほぼ全社的（調査・研究・開発・生産・営業）に活用

課題

- データを集め解析し最終的に何を成果とするか、を議論（あるいはコンサルテーション）するために、幅広い層（経営職・管理職・専門職）への教育（データを解析、データの構築、収集、統合、そして出口の戦略立案）が不足

ITスキルと専門スキルバランス

- 両者を完全に充足する必要はないが、両者が必要

人材育成

- 少なくとも、統計と簡単なプログラミングの知識があり、実際のデータを扱える人材が必要
- 各社のニーズに合わせて自社のビジネスに活かしていける即戦力を渴望
- データサイエンスは分野日進月歩のところが大きく、将来技術が普遍化した状態への対応が必要

求められる講座像、等

- 短期講座、遠隔講座、e-learning等々への対応は必須
- 昨今のAI、ビックデータブームにより様々な講座が乱立しており、参加講座の選択に戸惑う企業の声も多く聞かれた。

成果：詳細

(2) バイオ関連データサイエンススキル標準

- ✓ 「ニーズ→スキル標準→プログラム」というステップを踏む進め方から、「ニーズ→プログラム」と「ニーズ→スキル標準」を並行させる。
- ✓ 積み上げ式ではなく（スキル標準を作って講座を作るのは困難）、ニーズから追加講座を作り、スキル標準に対応させる。現存のITスキル標準に、バイオインフォマティクスを追加する。具体的な項目・内容ではなく、経験・実績のクライテリア、レベルを決める。スキル標準は大まかな記述で良い。
- ✓ 4カテゴリー「ドライ、ウェット、プログラマ、キュレータ」とレベル感の合意
- ✓ 学会（バイオインフォマティクス学会、日本オミックス医学会）承認の方向性確認

バイオ関連データサイエンススキル標準ドラフト版は別紙にて報告。

成果：詳細

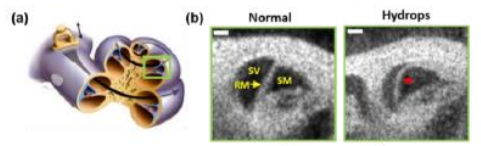
(3) 社会人向け教育講座プログラム

ELHNet: マウス蝸牛の内リンパ水腫検出
(文献番号 = PMID:2868782)

8100333, 2017 Sep 28;119(4):404-404. doi: 10.1369/000068017070401.
ELHnet: a convolutional neural network for classifying cochlear endolymphatic hydrops imaged with optical coherence tomography.
Liu GS¹, Zhu MF², Kim J³, Ikenoishi E⁴, Anzaike BE⁵, Ohtsuka JS⁶

マウス	1,759匹
データセット 画像枚数	2,159

光干渉断層撮影 (Optical Coherence Tomography : OCT)



▼ [2] データ拡張を理解する

データ拡張 (Data Augmentation) とは、機械学習分野で学習データ数を増やす操作を指します。画像分類課題の場合は、入力画像に対して移動、回転、拡大・縮小などの加工操作をして、加工画像を元の入力画像と合わせて学習データに使います。まず、深層学習ライブラリkerasでのデータ拡張方法を確認しましょう。画像の平行移動をkerasで、出力結果画像を視覚化して確認します。データ拡張の出力結果を確認し易くするために、ここでは、OCT画像ではなくマウスの画像を使用します。



参考LINK

東京医科歯科大学で実施された実証プログラムの様子 (上記)

「深層学習による画像分類と前処理効果」というタイトルで、オープンデータのマウスOCT画像を用いて、疾患と正常を分類する深層学習モデルを構築します。

東京大学大学院農学系研究科で実施された実証プログラムの様子 (下記)

企業ニーズが高い「統計と簡単なプログラミングの知識があり、実際のデータを扱える人材」を効率的に育成すべく、データ解析環境Rに特化した実践的な講座を開講しました。



成果：詳細

(4) 受講者満足度アンケート調査結果

アンケート報告

【まとめ】

受講した講座で良かった点、勉強・参考になった点をお書きください

- ✓ 今回の講義は、基礎から始めて具体的な事例を用いた機械学習まで、集中的に一通り（俯瞰的に）学べた点良かった。また、集中講義では同じ目的を持つ参加者が集まって講義を受けることで集中力やモチベーションが上がり、その場で疑問点が解決できる点良かった。
- ✓ e-Learningでは短時間に自分のペースで行うことができる点、集中講座では資料が作りこまれていて、説明が丁寧な点などが挙げられていた。

受講した講座で改善してほしい点をお書きください

- ✓ 今回は、無料で、盛沢山（内容が豊富、説明が丁寧）だったため、満足されている受講者が多かった（従って、不満な点は、時間が足りない点に集中している）。また、バイオ分野への使用例、深層学習の活用例などを盛り込んで欲しい等の要望があった。
- ✓ e-Learning：コメント・質問等への対応が欲しかった。
- ✓ 集中講座：時間が足りなかった。レベルをいくつか分ける等の工夫が欲しかった。