

産業技術大学院大学におけるIT技術者のための データ解析の学びと実践

中野美由紀[†] 柴田淳司[†] 渡邊紀文[†]

産業技術大学院大学[†]

1. はじめに

21世紀に入り、「ビッグデータ」という言葉に代表されるように、IT技術は現代社会における情報の在り様、あるいはその情報を利用する社会の在り方を大きく変えている。それは、世界中に広がる個人が種々様々な情報を自ら発信し、自ら収集可能である社会であり、また、情報（コンテンツ）が社会において経済的な観点も含めて大きな価値をもつ社会である。IT技術の変化は激しくまた様々な分野と融合しており、現場で開発にいそむIT技術者も十分に追随しているとは言い難い。

短期間で変遷するIT技術の中でも、データ解析技術はこの数年で急速に着目され、統計解析の知識に始まり、最先端の機械学習技術の理解まで幅広い知識と技術力が求められている。

産業技術大学院大学では、主として職業人を対象として最先端のIT技術の学びの場を提供している。学生の80%以上が社会人と学生を両立しているなか、今後のIT業界を担う人材の育成、リカレント教育も含め、ビッグデータ時代に対応できるデータサイエンティスト育成が業界のみならず様々な分野から強く望まれている。特に社会に流通する様々なデジタルコンテンツに柔軟かつ即座にデータ解析を行うためには、データ解析の理論と共に現代社会に即応した実践が必要不可欠である。産業技術大学院大学では、講義およびPBL活動という異なる教育の場を通して、現代におけるIT技術者のデータ解析力をアップしている。本論文では、本学で行っているe-learning、講義、PBLを通してのデータサイエンティスト教育について紹介し、e-learningによる初学者に向けた実データを利用した学びの試みについて報告する。

2. データサイエンスカリキュラムの動向

データサイエンスの知識体系は、計算機科学、情報学等のカリキュラム標準としてはいまだ米国のACM等でも検討中である²⁾。

数学における統計解析の学びにおける知識体系の共有、あるいは、社会学、生物学、医学等他分野における副専攻として、共通で学ぶべき知識体系の整備も現在米国で調査が進められている段階であり、確たるのが定められているわけではない。2016年に内閣府から出された第5期科学技術基本計画においてSociety 5.0の実現にあたりデータサイエンスの重要性がうたわれ、また総務省からもデータサイエンティストが現在20万人近く不足しているとの推計が出ている¹⁾。データサイエンティストとして育成される学生数は年間数千人にとどまっておらず、データサイエンティスト学部が創設されているが到底間に合うものではない。IT業界ではデータ解析に由来から携わっている人材を中心に自主的な学びの場の提供が行われているが、体系的かつ継続的な学習の機会が保証されているわけではない。

データ解析技術を理解し、利用できる即戦力となる人材を継続的に育成するためには、現在の社会人に向けてなんらかの形で基礎学習からその時点での最新技術の情報提供も含めた学びの場が望ましい。総務省統計局ではMOOCを利用したデータサイエンティスト育成のための講座を提供し始めており、初心者への学びの場としてよい教材が利用可能となっている。一方で、このような質の高い学びの場は教材等の準備には時間もかかり、発展的な学習のための知識体系は定まっていなかったり、忙しい社会人に向けた情報発信なども十分ではない。

3. 産業技術大学院大学におけるデータサイエンス教育

本学では、大学院の講義として「データインテリジェンス特論」「ビッグデータ解析特論」を開講し、大学院学生に向けたデータ解析技術の理解および利活用の学習の場を提供している。本学は科目等履修生制度による単位バンク制度として本学の学生以外の方でも学べる機会を提供しているが、決まった時間帯の講義形態ということもあり、卒業生からも、より柔軟な講義形態への希望が寄せられている。また、本学では修士2年で一年間のPBLを通じて、最先端IT技術を利用したサービス、システムの研究

Study and Practice of Data Science for IT Engineers in AIIT
[†]Miyuki NAKANO, Atsushi SHIBATA, Norifumi WATANABE · Advanced Institute of Industrial Technology

開発を実施している。近年の傾向として、データ解析を利用しなければならない事例も多く、前節のデータサイエンスカリキュラムの動向にて、米国においても副専攻のための知識体系の調査、整備が行われていると伝えたが、本学においてもまさに同じ状況となっている。サイエンスでは従来からデータ解析は必要に応じて利用されていたが、個別の分野に限らない基礎的な知識体系の共有として、講義形式に限らない柔軟な学習支援の手段は今後 IT 技術を学ぶ場では必須になると考えられる。

3. 1 データサイエンス育成道場

本学では、社会人に開いた大学院大学として、今後のリカレント教育の実現も念頭にデータ解析の学びを支援するサイトを試験的に開設している。

- ・データサイエンスの学びおよび情報のポータルサイト
 - ・基礎的な統計解析技術の理解
 - ・ビジネスにおける利用を前提とした学習教材および利用可能な実データの提供
- に主眼をおいたウェブサイトを構築している³⁾。

データ解析技術を実践的に学ぶ教科書は数多くあるが、適切なものを選ぶ、あるいは、具体的に利用可能なデータの提供まで含めた教科書はさほど多くない。これは、変化の激しい IT 業界において、その変化に追従したデータの提供が難しいこと、および特定の機械学習ツールで利用するためには事前にデータクレンジング等が必要となる、等によると考えられる。

また、データ解析においては、計算機の性能、収集されているデータの特性、あるいは求められている予測モデル等により、適切なデータ解析ツールが異なる場合も多い。本サイトでは異なる応用分野を取り上げることにより、異なる手法を実データの利用により理解できるように構築している。

具体的には、総務省統計局が提供している e-stat の比較的整備されたデータを利用した相関の理解からウェブ上のテキストデータを対象としてデータクレンジング、簡単な自然言語処理による要素の抽出による文書理解まで幅広く実習可能な形式で教材を提供している。

3. 2 講義・PBLにおける実践的学び

開講されているデータ解析に関する講義は「データインテリジェンス特論」「ビッグデータ解析特論」がある⁴⁾。本大学院では学部で統計解析の基礎等は学んできたことを前提としているが、IT 技術者であっても社会人の経歴は様々であるため、他の科目（「情報アーキテクチャ特論 I」等）を履修することで、実践的なデータ解析技術の学びへの準備を可能としている。IPA が提供する情報技術者のためのスキル体系においても、各スキル項目のなかに含まれた形になっており、明示的な指示はないが、マーケティング

における動向理解、戦略的企画における市場理解など、データ解析技術を背景とする技術力が求められており、本学においても「e ビジネス特論」「情報ビジネス特論」等で積極的にデータ解析技術も含めて取り上げている。

また、本学の特徴である修士 2 年生で実施される一年間の PBL でも「ビッグデータ解析」をテーマとした PBL も開講されており、データ解析を発展的に学びたい学生への機会を提供している。具体的には、2016 年度は東京オリンピック・パラリンピック 2020 に向けて、スポーツイベントを考慮した大都市における混雑度予測⁵⁾などの企画が実装され、実践的かつ新規のテーマが取り上げられている。混雑度予測では、ウェブ上からのデータ収集、クレンジング、統計解析とその結果の混雑に相関のある特徴量の抽出、予測モデルの構築、検証と予測モデルの再構築まで、知識発見プロセスの理解とデータ解析が実践された。データ解析の初心者であった学生達が一年の経験を経て利用可能なサービスとして実装する力を持つに至っている。

PBL は実践的学びの場として大変効果的であるが、学べる人数には限りがあり、また、指導のための人的資源を多く必要とする。この経験を元にデータサイエンス育成道場では、学びの場として初心者にも経験者にも効率のよい教材の提供、リカレント教育として最先端情報の提供の場として、改善をしていく予定である。

4. まとめ

データサイエンティスト育成が強調されるなか、現在のデータサイエンス教育の動向および課題を挙げるとともに、産業技術大学院大学におけるデータ解析教育の現状および現在試験的に立ち上げているデータ解析の学びの場を紹介した。本学の試みはまだ始まったばかりであり、社会人がデータサイエンスを学ぶにあたって、どの程度有効であるかは未知数である。プロトタイプを利用した学生の活躍を期待している。

参考文献

- 1) <http://www.stat.go.jp/training/2kenkyu/pdf/gakkai/toukei/2015/sue.pdf>
- 2) http://www.computingportal.org/sites/default/files/Data%20Science%20Education%20Workshop%20Report%201.0_0.pdf
- 3) <https://sites.google.com/a/aiit.ac.jp/datascientist-startup/>
- 4) https://aiit.ac.jp/admission/pdf/scholarship/h27_syllabus.pdf
- 5) 都市における混雑度推定のためのスポーツ観戦者数の予測手法：日本におけるプロ野球観戦に関する考察，小林峻 新井教広 一ノ木繁 金丸正憲 鎌柄拓史 吉野悦朗 中野美由紀，第 79 回情報処理学会全国大会、5B-01，2017.3