

日米データサイエンティスト教育の違い

佐藤 貴海^{†1}

概要：データサイエンティスト教育において日米とも Computer Science (CS) 学部の数理系の授業を集めた科目で大きい違いは無いが、卒論必須の日本方式とプロジェクトベース授業が苛烈だが卒論のない米国方式という違いがある。両方式でも教育の質については大きい違いは感じられないが、日米両国において高等教育を受けた人材に対する社会の受け入れ方が異なっており、待遇の差に繋がっていると考えられる。一方、日本の新卒採用は世界で稀に見る未経験者に優しい採用となっており、データサイエンティストを志す学生は日米の良いところを見極めてキャリアの構築が必要になっている。

キーワード：留学、教育、データサイエンティスト、機械学習エンジニア、データエンジニア

1. はじめに

この度「データサイエンスを担う人材を育成する」特集において、寄稿する機会を頂きありがとうございます。私は現在 University of California, Irvine にて Computer Science (CS) の修士プログラムに在籍をしており、それ以前は日本で修士まで経営工学を学んだ後5年ほどデータサイエンティストまたは機械学習エンジニアとして働いていました。

タイトルは「日米データサイエンティスト教育の違い」となっていますが、一般的な Computer Science 学部在籍なので米国のデータサイエンス教育についてはそれほど多くの事は語れません。しかし、米国の大学教育について実地で感じたことを、日本でのデータサイエンティスト経験ある私が語るのも一興と思ひ寄稿させて頂きました。

2. データサイエンティストに必要な教育

データサイエンティストという言葉は2010年頃から登場し、今や広く受け入れられるようになってきました。しかし、一重にデータサイエンティストと言っても Excel や SQL から仮説を導いて施策を次々回しているコンサルタント型のデータサイエンティストから、大規模データウェアハウスや HDFS 上のデータから機械学習モデルを作りシステムに組み込んでいるエンジニア型のデータサイエンティストまで幅広く存在し、厳密な定義はさながら宗教論争に近い状況なので不可能でしょう。

本稿では、大学教育に主眼を当てているので、丁度講演を聴く機会があった Stanford 大学の Ullman 教授のデータサイエンスの解釈の図(図1)からスタートしようと思います。この図には、大きく5つの円があり、それぞれ Computer Science (CS), Machine Learning (ML), Math, Stat, Domain Science となっています。Ullman 教授の解釈ではこの5つの円の交わるところに位置しています。大学教育においては、Domain Science 以外は CS 学部において数理系

の授業を履修していけば自ずと網羅できると思います。講演では「統計学は物事を正しく分析するのに役立つが、しばしば現実問題への解決策に乏しい事がある。ML もデータサイエンスの大きい部分だが全てでは無い。」とも述べており、統計学については教養課程レベル、ML については応用レベルまでが大学教育で求められるラインと考えられます。

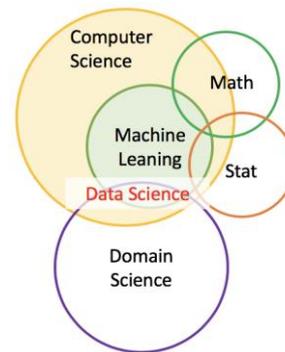


図1 Ullman 教授によるデータサイエンスの解釈

Domain Science については皆様ご存知かと思いますが、大学教育内で学ぶことがとても難しい項目です。しかし、これは大学に限った話ではなく、データサイエンティストもプロジェクトを始める際はドメイン知識の収集には苦労します。いわゆる“プロ勘”と呼ばれる現場の勘に機械学習モデルが惨敗するのはよく見られる光景です。さらに、これらの経験は研究を通して得られる経験によく似ています。日本の大学教育の優れた点は、基本的に学生全てが研究経験と卒業論文の執筆経験を持って卒業することだと感じています。

3. 日本と米国の大学教育の違い

では、日本の大学教育のほうが優れているかと言うとそ

^{†1} University of California, Irvine (連絡先: sharp2261@gmail.com)

ういうわけではなく、米国の大学教育と比較して日本の大学教育には2点問題を感じています。

- 卒業論文が必須のため、学生一人当たりの教員の指導が希薄化
- 米国の授業は課題が多く、プロジェクトベースの授業で応用レベルの教育が充実

米国の大学では学部と修士課程において研究室に所属することが必須ではなく、研究職を志す学生は自主的に教授にお願いする必要があります。この結果として教員はやる気のある学生のみを相手にでき、卒業自体は授業だけで出来るので無理やり論文審査を通す必要もありません。

ただし、その授業は軽いものではなく、学生は1学期に1つから3つ程度の授業を履修し、各授業が週2から3回の合計週3時間の講義と山の様な課題を出されます。新しく履修する分野で応用レベルの授業プロジェクトを成功させるのは容易ではないので、それを叩き込むために授業課題が重いというのがあります。さらに応用レベルを超えて授業プロジェクトをそのまま論文にして投稿することもよくあり、苛烈なプロジェクトベースの授業は米国大学教育の良い点の一つです。この様な形で、研究機関としての大学と教育機関としての大学が上手く棲み分けられているというのが米国の大学教育において私が感じたことです。

一方、米国の大学教育が日本よりも優れているかというと、必ずしもそういうわけではありません。米国では良くも悪くも様々な人がたくさんおり、留学生も世界各国からくるため、全員一律に論文指導をする様な事は米国の大学では機能しません。授業が重いため、研究に使える時間が少なくなり、特に修士では日本の様に研究をすることは難しいです。他にも、一律試験の大学入試により、教育機会の均等さや入学生の質の担保、それに伴う授業の開始レベルの高さは日本の大学の良い点です。結論的には性質は異なるものの、受けられる教育の質において日米で大きい違いはないと思っています。ただし、米国方式の方がよりシステムチックで、日本方式は教員の献身によって成り立っているように感じます。

4. 高等教育の捉え方の違い

教育の質において違いはあるものの日米で大きな違いは無いと考えています。しかし、社会の高等教育および高等教育を受けた人材に対する社会の捉え方が日米で大きく異なっていると感じています。

米国では先に触れた様に多種多様な人々が存在しており、人材レベルの下限を設定することが不可能です。そのため大学の授業を重くして、それを乗り越えた者には能力証明を与えるという証明機関としての大学の役割が重視されています。これは学位やGPAが就職において重視される事か

らも伺えます。結果として、専門職が同質の教育(例えばCS修士)を受けた者にのみ開かれ、Ph.D.ホルダーや研究者に対する敬意も強くなり、その専門職に就職するために働いてから大学院に行く人も多く、産学の交流も盛んになるという好循環が生まれています。

一方、日本において専門知識は業務を通じて獲得するという感覚が強くなります。充実した研修や終身雇用を通して一人前になっていく事が、かつてのロールモデルでした。そのため大学教育は人材の土壌を富ますものといった捉え方をしている人が多いと感じています。たとえ特定の勉強をしてなくてもモラトリアムもその人の土壌を富ます可能性があると捉えられ、一方、勉強をしていても、学生のままだと土いじりばかりで作物を作っていないと思われ一人前と扱われません。

この思想は、技術進歩の緩やかな時代には適合して日本の経済成長を促したものの、技術進歩が急激になり、新卒のほうが勤続年数の高い社員よりも高度な専門知識を持ち、高度人材が常人の100倍以上の生産性を持ちえるような情報技術の分野ではむしろ足かせになっています。現状日本においては高等教育とは何かの再定義が必要な場面となっています。

5. 日米どちらの大学で学ぶのがよいか

このように、データサイエンティストのような専門職は米国のほうが良い待遇なのが現状です。ただしその職につけるかはまた別の問題となります。データサイエンティストは統計や機械学習の知識を持つ専門家であるものの、Domain Scienceの比重が大きいため、実務経験のないデータサイエンティストはやはり一人前ではないでしょう。

この点において、日本の新卒採用は世界でも稀に見る未経験者に優しい採用システムを持っており、これを活用して実務経験を先ず得るのが得策です。近年のデータサイエンスブームにより会社を選ばず、真面目に勉強をしていれば職を見つけることは難しくありません。実務経験があればそれを糧に、より良い待遇を求めることも容易になります。

ただし、米国の大学出身でも日本の新卒採用には応募できるので、結論からいうと日米どちらの大学でも良いと思われれます。個人的には教育の質に大きい差は無いと思っています。各自の信条で選ぶのが良いと思います。確実に米国の大学が良い点としては、米国の大学を卒業するとOPTと呼ばれる制度でH1-Bなどの就労ビザなくとも数年働ける点です。米国で働く最も大きい壁はビザと言われていたので、留学してしまうのも良いです。ただし授業料は高額という欠点はあります。修士でもOPTは貰えるので、私のように一度企業勤めをしたあと修士で留学というのも選択肢の一つです。

日本と米国ともに良い点と悪い点があり、簡単に制度は変わっていかないのが、なるべく双方の良いところを渡り歩ける様なキャリアを積んでく事が理想だと思われま

6. データサイエンティストとしてのキャリア構築

データサイエンティストという職種が生まれた当初は、様々な職がデータサイエンティストと一括りにされていましたが、近年は整理が進んでおり、データコンサルタント、機械学習エンジニアやデータエンジニアなど細分して募集されるようになってきました。

これらの職種は大きく分けて、コンサルタント型とエンジニア型に分類されます。コンサルタント型は SQL, Excel, 各種 BI ツールを武器にデータから仮説立案して施策立案をしていく職種です。仕事のスタイルとしては数字やデータに強いマーケターが一例で、例えば既存のメーカーに仮説に基づく人にメールを送って効果検証をするといった仕事があります。日本で最初にデータサイエンティストが受け入れられたのが WEB マーケティングとソーシャルゲーム業界なのは、データが取りやすく施策実施が容易だったというのが大きい要因でした。リーマンショックに伴い金融業界から数理に強い人材がこれらの新興業界に流入したという要因もありました。

一方、機械学習エンジニアやデータエンジニアは、データ分析を機械学習や統計の知識を使用して、大規模な自動処理を行う事を指すものです。機械学習エンジニアはより予測モデル構築などの数理モデルに責任を持ち、データエンジニアはログなどの大量データを処理する基盤をいかに高速・堅牢に構築するかというところに責任を持ちます。おそらく今最も需要が高いのはデータエンジニア職です。大規模データを流すパイプラインは大企業でも構築する人数は少なく、経験を積む場が多くないにも関わらず投資金額が大きいので経験がある人間の需要はとても高いです。

個人的にも大学卒業してからエンジニア型のデータサイエンティストを目指すのであればデータエンジニアがお勧めです。新卒エンジニアの最初の大きい壁は開発経験をいかに積むかというところにあります。コンサルタント型のデータサイエンティストや一部の機械学習エンジニアは直接の開発部隊から部署が離れていることがあり、開発経験が積みにくいことがあります。またデータサイエンティストのようなキラキラした肩書ですとエンジニア側が敬遠して開発指導をしてもらえない事もあります。開発力のないデータサイエンティストは自分の技術の出口を持たないため価値を出すのが難しいです。開発経験をしっかり積むためにも最初の数年はエンジニアとして働くのが個人的にはお勧めです。

しかし、近年のクラウドサービスの充実はかなり早く、

データエンジニア職も安泰とはいえません。機械学習エンジニアはそもそも倍率が高く、機械学習ライブラリを表面上使えるだけの人から、実務で叩き上げの人および CS や物理の Ph. D. を持つ理論からしっかりした人などとの競争になるので違いを見せるのはとても難しいです。コンサルタント型のデータサイエンティストはそもそも戦略コンサルタントと業務内容が絡むので更に熾烈な競争が待っています。ただし、コンサルタント型のプロジェクトマネジメントや受託開発でのお金の稼ぎ方といったスキルは、エンジニアであっても身に付けていかないと営業頼みになり、技術的に面白いと思える案件に巡り会えるのが難しくなります。自分が面白いと思う仕事は結局の所、自分で見つけて作り上げていくしかありません。

そのため、現状一社に勤め上げて、データサイエンティストのスキルを網羅するのはとても難しいです。ただし幸いにして選択肢は一昔前よりも充実しており、データ分析ベンチャー企業で機械学習エンジニアやコンサルタントとしての経験を積んだり、コンサルティングファームも最近ではデータサイエンティストを募集しているのでコンサルコンサルティング経験を積んだり、大規模データを扱える企業に就職して大規模データを捌けるデータエンジニアとしての経験を積んだり、正解がない中で自分の目指すデータサイエンティストのポートフォリオを組んでいく必要があります。

7. さいごに

日米データサイエンティスト教育という物々しいタイトルで書き始めましたが、日米ともにデータサイエンティストとしての教育は主に CS 学部において数理系の授業を強めたカリキュラムになっており大きい違いはありません。ただし、日米の大学教育の大きい違いは、日本は卒論修論での研究・執筆経験の充実、米国はプロジェクトベースの苛烈な授業メインで卒論は必須ではない事でした。米国は希望する学生のみならず研究指導をすることで、研究機関としての大学と教育機関としての大学が上手く棲み分けられているなと感じています。

日本方式は、現状では教員の献身によって成り立っているように感じており、システムチックな米国方式を一部取り入れる事で、昨今問題になっている大学教員の雑務の多さを軽減できると思われます。ただし米国大学と異なり、日本の学生はほぼそのまま日本社会の構成要員になるため、全員が卒論を書かなくなった時にどういった社会構造になっていくのかも問題です。個人的には産業界に如何に大学が質の高い学生を供給してきたかを知らしめるのに一度やってみてはと思いますが難しいでしょう。

学生の皆様は、これからデータサイエンティストとしてキャリアを積んでいくには、自身で主体的にスキルのポー

トフォリオを管理して大学の授業や研究から転職を含めて考えて行く必要があります。データサイエンティストは一過性のブームという心配があるかもしれませんが、既に周りの多くのサービスに機械学習が組み込まれており、今後

なくなることは無いでしょう。データに基づく意思決定、果ては人工知能というのは人生でやるには面白いテーマの一つなので、研究と実務どちらでも将来一緒にお仕事出来ることを願っています。