

人文・社会科学系大学における データサイエンス授業の効果分析の試み

辻 智

1. はじめに

職業としてのデータサイエンティストの人気は過熱しており、米国では相変わらず人気のトップを走り続けている。毎年、米国の大手企業就職口コミサイト“glassdoor”は、米国の“50 Best Jobs”（人気職業ランキング）を発表しているが、データサイエンティストの人気は衰え知らずである（2019年1月現在）⁽¹⁾。データサイエンス分野の極端な人手不足により、米国では年俸基本給の中央値が\$108,000と高騰している。日本においても、2019年の夏に集中して、伝統的大企業が優秀な社員を別格で厚遇する施策を次々に発表した⁽²⁾。比較的穏やかな給与体系の日本では、これまでにはあり得ない厚遇の宣言となっている。

成城大学は、時代を先取りして人文・社会科学系大学の中では最も早期（2015年度）にデータサイエンス科目群の授業を開始し⁽³⁾⁽⁴⁾、この原稿の執筆時（2019年度）ですでに5年目になる。2019年度はデータサイエンスを概観する概論、統計学を中心とした入門、スキルを伸ばすための応用など6科目群全体で、前・後期合わせて延べ300名を越える履修登録があり、人気の科目群のひとつに成長した。2019年3月には、最初にデータサイエンスの授業を履修した学年が卒業し、新しい分野の就職先としてIT企業へ就職した学生も現れた。

成城大学において実践してきたデータサイエンス教育に関して、データサイエンス科目群の中でも最初に開講した「データサイエンス概論」の授業内容を2018年度に定性的に紹介した⁽⁵⁾⁽⁶⁾。本稿では、この「データサイエンス概論」の授業の効果分析を定量的に試みる。学生の認知度をスコアでフィードバックしてもらったアンケートのみならず、学生の感想や要望などのコメント文を言語分析し

て、多角的に学生の認知度や感情を捉えてみる。また、誰でも容易に分析方法をトレースできるように、本稿では分析ツールは無料範囲で使えるもので、操作が直観的かつ容易なものを利用している。

2. 学生の受講の動機

「人間とコンピューターの新たな関係を築くビッグデータの活用」と副題がつけられたこの「データサイエンス概論」の授業は、成城大生全学部全学年を受講対象とした全15回の構成である。在学中にいつでも受講できることが利点である反面、クラスの編成毎に学年や学部の偏りが出るため、クラスの雰囲気はクラス毎に異なる。そのためか、クラスが異様に静かなこともあり、授業の途中で学生に話しかけても反応してくれず、インタラクティブな授業が進められないことがある。元々、最近の傾向として授業中に沈黙する学生は多く、その理由と対策が研究されている⁷⁾。「データサイエンス概論」を履修する学生の率直な意見や本音を得るには、授業中のインタラクティブな進行より、授業の最後にコメントを書いてもらう方が、情報量が多くなることは、これまでの授業からも学んできた。

2019年度の「データサイエンス概論」は、前期2クラス、後期2クラスとなっている。この原稿の執筆時点では、後期のクラスが開講中なので、終了した前期の1クラス分をサンプルデータとして扱う。図1は、その前期の1クラス分の履修動機をまとめたものである。データサイエンスそのものに興味があり、詳しく学びたいと思っている学生が47%、数学やITの苦手意識を克服したい学生が20%、就活や資格取得など社会人になってからも役に立つスキルを身に付けたいと思っている学生が24%、プログラミングにチャレンジしたい学生が9%となっている。これらのように履修動機も様々となっている。

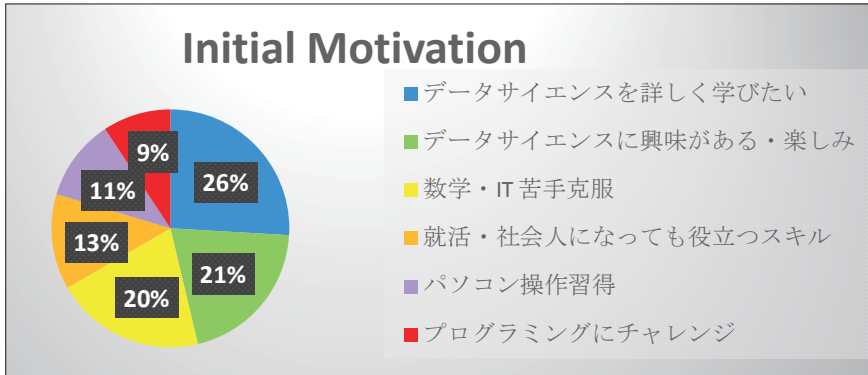


図 1. データサイエンス概论の授業を受講する動機の例

3. 調査方法

「データサイエンス概论」の授業は全 15 回の構成であるが、データサイエンスに関して初学者が多いのも特徴である。そのため、開始時の第 1 回と終了時の第 15 回には同じ質問内容の認知度アンケート調査を行っている。「次のデータサイエンスに関する用語の認知として、あなたに当てはまるレベルをクリックしてください。」という問いかけで、あらかじめ設定した 85 のデータサイエンスに関する用語に関して認知度を聞いている。学生にとってよく耳にする用語から、聞いたことがないような専門用語までを意図的に選定している。表 1 は、その認知度アンケート調査用の用語例であるが、抜粋して 20 例を示す。これらの用語に対して、次の 4 段階の順序尺度で認知度を聞いている。

- ・よく理解していて、他者に説明できる
- ・何となく知ってはいるが、他者には説明できない
- ・名前くらいは聞いたことがある
- ・ほとんど知らない

また、アンケートの最後部には、第 1 回には「この授業への期待、要望、質問、不安など」、第 15 回では「この授業の良かった点、残念だった点」などを

自由記述形式で書き込みできるように設定している。第2回から第14回までの途中の回でも、毎回授業の終わりに理解度や要望を50～300字程度の自由記述形式で書いてもらっている。第1回から第15回までのアンケート調査には、Microsoft Forms⁽⁸⁾を使用して、学生のコメント文をデジタル的に収集している。

4. 第1回授業前と第15回授業後の学生コメントの比較分析と結果

図2は、表1の用語(1～10)に対する第1回授業前(以下、Beforeで表す)と第15回授業後(以下、Afterで表す)の学生の認知度をMicrosoft Formsにより集計し、出力したものである。分析サンプルとしてのコメント数は、Beforeが42、Afterが37となっている。Afterの方が少ないのは、第15回の授業が定期試験期間直前の最終授業なので欠席する学生が増えたためである。用語(1～10)は、データサイエンスそのものと、データに関係した用語がリストアップされている。Beforeでは、よく耳にするデータサイエンスとビッグデータに関しては分かつ

表1. 認知度アンケート調査用の用語例(抜粋)

用語 (1～10)	用語 (11～20)
データサイエンス (Data Science)	人工知能 (AI: Artificial Intelligence)
ビッグデータ (Big Data)	機械学習 (Machine Learning)
データセンター (Data Center)	ニューラルネットワーク (神経網: Neural Network, 略称: NN)
データウェアハウス (Data Ware House)	ディープラーニングまたは深層学習 (Deep Learning)
テキストマイニング (Text Mining)	サポートベクターマシン (Support Vector Machine, SVM)
データマイニング (Data Mining)	チャットボット (Chatbot: 別名「人口無脳(じんこうむのう)」 または「人口無能(じんこうむのう)」)
データクレンジング (Data Cleansing)	IBM Watson (アイ・ビー・エム・ワトソン)
ダークデータ (Dark Data)	Microsoft Azure (マイクロソフト・アジュール)
データレイク (Data Lake)	Google Tensor Flow (グーグル・テンソルフロー)
メタデータ (Metadata)	Amazon Web Services (AWS)

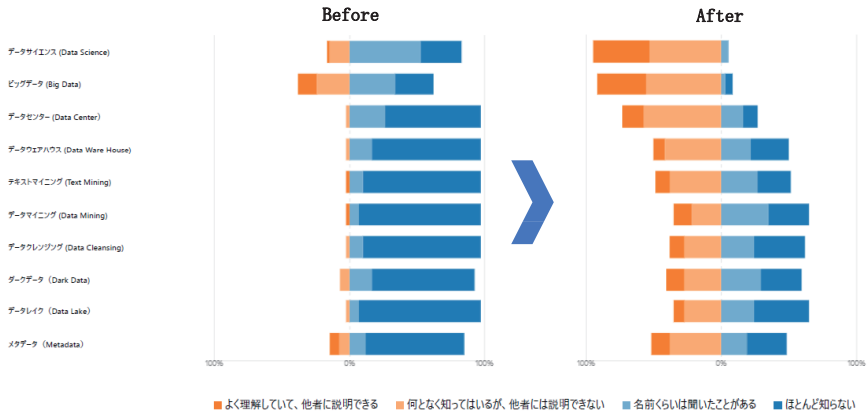


図 2. データに関する用語の学生の認知度比較
(Before: 第 1 回授業前、After: 第 15 回授業後)

ている学生が若干おり、メタデータに関しても認知しているが、その他の用語はほとんど知らない結果となっている。クラス全体としては、データ関係の用語に関して、認知がほとんどされていない状態となっている。

一方、After になるとデータサイエンスおよびビッグデータに関しては認知が劇的に進み、その他の用語に関しても認知が進んでいる様子がうかがえる。「よく理解していて、他者に説明できる」割合も増えている。After でも、まだ「名前くらいは聞いたことがある」、「ほとんど知らない」と回答している学生が半数くらいいるのは、学生のためらいがちな性格を反映しているのか、本当に認知していないのか、Before と After のアンケートの集計だけでは分かりにくいので、途中回のコメント文の追跡も大切である。

図 3 には、興味深い点がいくつかある。用語 (11 ~ 20) は、人工知能 (AI) に関係した用語である。AI に関しては、高校において学習してきたためか、Before の段階でほとんど認知されており、機械学習についてもかなり認知されている結果となっている。Before では、他の用語に関しても、図 2 のデータに関する用語と比較して、認知している割合が多くなっている。

After になると、AI および機械学習もよく理解していて、他者に説明できる割合が増え、授業でツールとして使用してきた IBM Watson、Microsoft Azure など

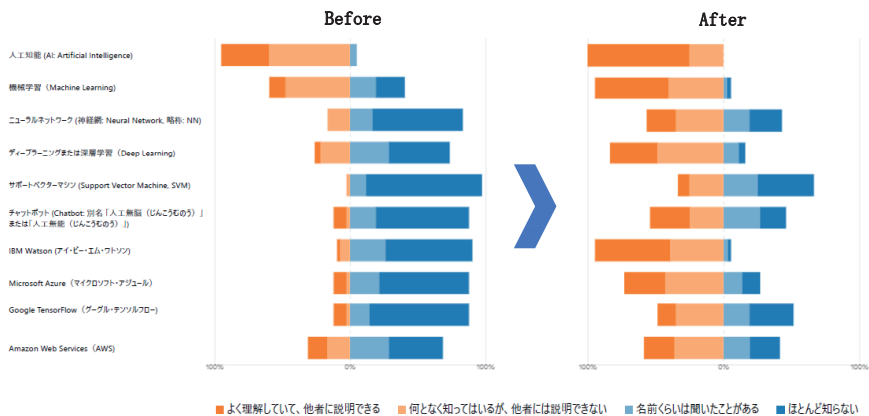


図 3. 人工知能 (AI) に関する用語の学生の認知度比較
(Before: 第1回授業前、After: 第15回授業後)

への認知度も大幅に増えている。不思議なことに、チャットボットなど途中回の授業中にトピックとして扱った内容であれば、Afterで認知度が進んでいるのは分かるが、授業ではサポートベクターマシンに関しては詳しく説明していないにも関わらず、Afterで認知度が大幅に上がっている。同様に、不思議な点として、AIに関する認知度のスコアはBeforeとAfterであまり変わらないが、その質が大きく変化することが挙げられる。これらの理由は、途中回のコメントを追跡することで明らかとなっていく。

図4は、BeforeとAfterの自由記述形式の学生コメントを、ワードクラウドで表したものである。ワードクラウドの出力には、ファンブライトラボのワードクラウド(テキスト)⁹⁾を使用し、出現頻度上位の40ほどの名詞を可視化した。Beforeでは、興味があり楽しみにしている様子うかがえるが、全体的に様々な単語が離散的に弱々しく並んでいる形である。一方、Afterの方は、AIやWatson、Pepperなど単語も具体的で、授業や知識、データサイエンス、アプリなどの単語も力強い並びとなっている。15回の授業を通して、学生が成長している様子が俯瞰的にわかる。

図5～7に、形態素解析関連の結果の中から、単語出現頻度および共起キーワードを例として示す。BeforeとAfterの学生コメントに対して、ユーザーロー

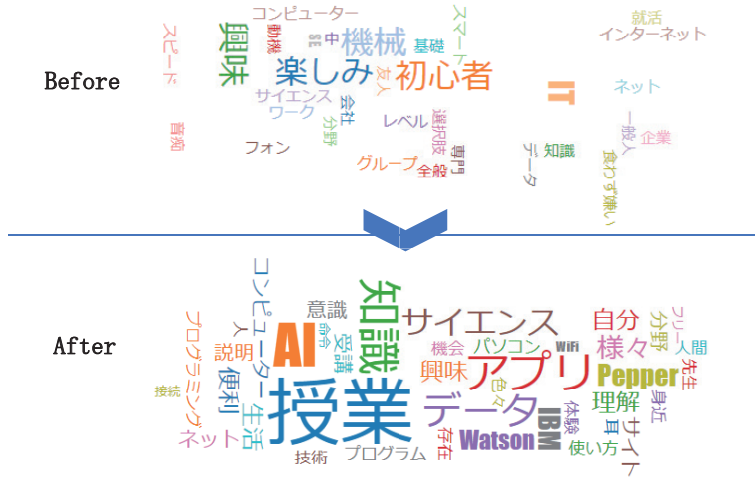


図 4. 学生コメントのワードクラウド例 (Before: 第 1 回授業前、After: 第 15 回授業後)

■ 名詞	スコア	出現頻度	■ 動詞	スコア	出現頻度
パソコン	0.33	5	学ぶ	0.19	3
授業	0.19	4	頑張る	0.03	3
心配	0.26	4	つく	0.05	3
不安	0.08	2	いける	0.04	3
楽しさ	0.03	2	使える	0.03	2
興味	0.05	2	ついていける	0.53	2
機械	0.20	2	追いつける	0.32	1
苦手	0.07	2	つくる	0.02	1
大丈夫	0.03	2	いく	0.00	1
初心者	0.04	2	働く	0.01	1
スピード	0.05	1	とる	0.00	1
友人	0.01	1	駸る	0.03	1
グループワーク	1.00	1	わかる	1.00	1
来年	0.02	1	教える	0.01	1
このまま	1.00	1	楽しむ	0.01	1

■ 形容詞	スコア	出現頻度	■ 感動詞	スコア	出現頻度
詳しい	0.01	1	---	---	---
弱い	0.02	1	---	---	---
楽しい	0.00	1	---	---	---
ほしい	0.01	1	---	---	---
嬉しい	0.01	1	---	---	---

図 5. 学生コメントの単語出現頻度 (Before: 第 1 回授業前)

■ 名詞	スコア	出現頻度	■ 動詞	スコア	出現頻度
授業	3.33	17	できる	0.84	30
AI	44.41	12	知る	0.72	16
知識	2.04	10	思う	0.06	10
アプリ	0.46	9	知れる	2.47	9
データサイエンス	16.84	6	学べる	3.48	7
様々	1.02	5	使う	0.09	7
IBM	5.00	5	動かす	1.86	6
pepper	12.92	5	わかる	16.84	6
興味	0.19	4	学ぶ	0.75	6
Watson	9.29	4	使える	0.17	5
生活	0.16	4	増える	0.10	4
理解	0.20	4	受ける	0.08	3
便利	0.25	4	思える	0.18	3
君	9.29	4	持つ	0.01	2
説明	0.12	3	分かる	0.02	2

■ 形容詞	スコア	出現頻度	■ 感動詞	スコア	出現頻度
よい	0.27	10	ありがとう	3.17	2
面白い	0.17	5	---	---	---
良い	0.03	4	---	---	---
やすい	0.06	3	---	---	---
楽しい	0.04	3	---	---	---
難しい	0.04	2	---	---	---
少ない	0.01	2	---	---	---
大きい	0.01	1	---	---	---
細かい	0.08	1	---	---	---
すごい	0.00	1	---	---	---
詳しい	0.01	1	---	---	---

図 6. 学生コメントの単語出現頻度 (After: 第 15 回授業後)

カルのテキストマイニングツール⁽¹⁰⁾を使って出力した例を示す。図 5、図 6 は、Before と After の学生コメントの単語出現頻度をそれぞれ示す。名詞、動詞、形容詞、感動詞に分けて出力している。出現回数が多い単語でも、意味が薄い、あまり重要ではないことがあるので、重み付けした特徴語を抽出するためのロジックとして、TF-IDF (Term Frequency- Inverse Document Frequency) 法が用いられている。Before では、名詞、動詞、形容詞ともに 5 回を越えて出現する単語はなく、偏りがあまりないような結果である。一方、After では名詞、動詞、形容詞ともに出現頻度の最上位の単語は、それぞれ 17、30、10 と突出しており、上位に位置する他の単語とともに傾向がはっきりとしている。たとえば、名詞では「授業」の出現頻度が 17、「AI」が 12、動詞では「わかる」の出現頻度が 30、「知る」が 16 となり、授業で AI への関心が進んだことがわかる。

図 7 は、共起キーワードの出力例である。一緒に出てくる隣接単語を線で結び、

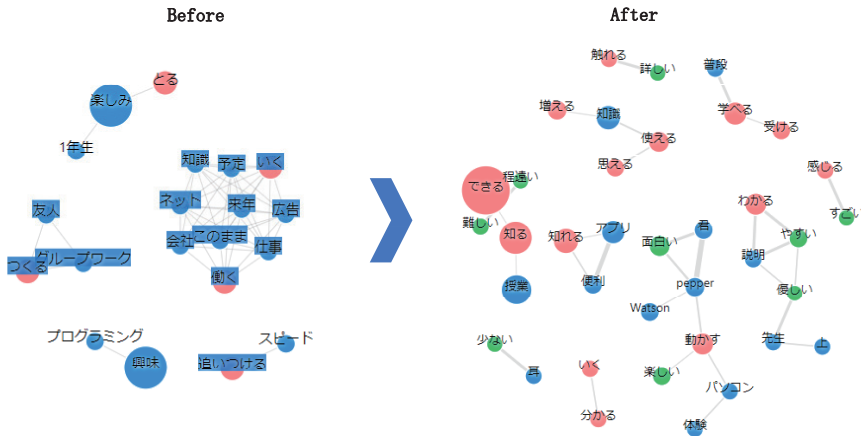


図7. 学生コメントの共起キーワード (Before: 第1回授業前、After: 第15回授業後)

「共起ネットワーク」で表す。共起とは、一文の中に単語のセットが同時に出現するという意味で、共起回数とは一緒に出現した回数を指す。Beforeでは、共起キーワードとして「楽しみ」、「興味」が目立っているが、固有名詞などの具体的な単語が見当たらず、力強さが感じられない。一方、Afterの方は、「できる」、「知る」、「授業」が共起キーワードとして隣接して大きくなっており、他のところでは、「Pepper」、「Watson」が固有名詞として表れている。また、「学べる」、「わかる」、「感じる」などの認知的な表現も多くなっており、学生の認識が強くなっていることがわかる。

従来の形態素解析を中心としたテキストマイニングに加えて、テキストに表れる語調(トーン)や感情を定量的に分析するためにIBM WatsonのTone Analyzer⁽¹¹⁾を利用している。このTone Analyzerは言語分析により、テキストから感情と文体のトーンを検出できる。感情としては、怒り(Anger)、不安(Fear)、喜び(Joy)、悲しみ(Sadness)を検出する。文体については、分析的(Alytical)、確信的(Confident)、あいまい(Tentative)なスタイルを検出する。また、このTone Analyzerは、ドキュメントとセンテンスの両方のレベルでトーンを分析できる。記述されたテキストのトーンは、Tone Analyzerのアルゴリズムを使って計算される。トーンの数値は、その強さによって0から1までの間の値を取る。IBM

【Tones 0~1.00】


	Anger	Fear	Joy	Sadness	Analytical	Confident	Tentative
Before	0	0.59	0.61	0	0.51	0	0.69
							
After	0	0	0.65	0	0.83	0	0.58

図 8. 学生コメントの語調（トーン）分析（Before: 第 1 回授業前、After: 第 15 回授業後）

Watson Tone Analyzer には今のところ日本語対応がないので、学生のコメントは IBM Language Translator⁽¹²⁾ により一手間かけて英訳したものを使って分析する。図 8 は、Before と After の学生コメントの Tone Analyzer による分析結果である。怒り、悲しみ、確信的なトーンに関しては、Before、After とともに検出されていない。不安なトーンは、Before では 0.59 であったが、After では 0 にまで払拭されている。喜びのトーンに関しては、Before は 0.61、After では 0.65 と僅かではあるが向上している。分析的なトーンは、Before では 0.51 であったが、After では 0.83 まで急上昇している。あいまいなトーンに関しては、Before は 0.69、After では 0.58 とかなり解消している。

5. 第 2 回から第 14 回までの授業途中回の学生コメントの追跡

先に示した IBM Watson Tone Analyzer による第 1 回授業前 (Before)、第 15 回授業後 (After) の分析結果を考察するために、第 2 回から第 14 回までの授業途中回の学生コメントを抜粋例として示す（順不同）。各トーンの変化を、これらの途中回の学生コメントから理解することができる。

【AI への先入観関係】 不安なトーンの払拭に関連

- ・高校では、将来 AI が仕事を奪ってしまうと教えられたが、映像を見て、奪うのではなく、まだこれから新たな可能性を AI は秘めているんだなと思った。
- ・「AI は危険だ」「AI に将来仕事を奪われてしまう」と教わってきてしまった

のでプラスのイメージがあまりなかったのですが、講義や動画を見て AI をこれからどう利用していくか考えるのが楽しみになりました。

- ・人間社会に情報技術を導入すると仕事を奪ってしまうといわれる時代ですが、効果的に活用することで、医療、金融など多岐にわたる分野で活用できることがわかりました。
- ・AI は人工知能という印象が強かったですが、世に出ているものは拡張知能がほとんどを占めていることに驚きました。知れば知るほど細かく分類されていることが分かったのでよかったです。

【興味・楽しみ】 喜びのトーンの向上に関連

- ・大学で学ぶことの中で一番楽しそうで興味があります。今日家に帰ったら色々な文章を、授業で教わったアプリで試したいと思います。
- ・自他共に認めるアニメファンである自分としては、今回の授業、特に IBM Watson のくだりは非常に興味深いものだった。つい先日フィリップ・K・ディックの『アンドロイドは電気羊の夢を見るか?』を読了したばかりなこともあり、人工 / 拡張知能が決して SF の中だけの夢物語ではなく、そう遠くない未来に起こる現実として存在するのだということが分かって胸が躍る思いがした。
- ・Pepper に触るのが楽しみです。
- ・Pepper をプログラムして動かすの、めっちゃおもしろかったです!!! 毎週やってほしいくらいです!!!

【実習関係】 喜びのトーンの向上、あいまいなトーンの解消に関連

- ・クラウドで AI をやってみるとというのが面白かった。実際にどのサイトでデモができるなどを知っておくと、後で家でも試せるので興味深かったです。
- ・ツイッターや文章の分析が出来てすごい時代だなと思いました。また、今後も家などで性格の分析などをやってみたいと思い、とても面白い実習時間でした。
- ・写真や顔の分析について知れてよかった。家でもやってみたいと思った。
- ・今日の実習部分ですごいと思ったのは、海外の海の向こうにサーバーがあるのに、画像を載せると数秒で解析が終わり、しかも解析内容がギターの画像

をあげればギター、ベースの画像を上げればベースと認識されるところだ。

- ・今日も新しいアプリを知れてよかった。家に帰ったら自分の写真を使って、感情などを見てみたいと思いました。
- ・顔認識は実際にも使われている場面を見たことがあったので、自分で操作できて楽しかったです。

【新たな理解】分析的なトーンの急上昇、あいまいなトーンの解消に関連

- ・特化型と汎用型の AI の違いが分かったし、自分たちの性格などがビッグデータから数秒で診断されるのはすごいと思った。実用的に使用されているのもすごいことだと思う。
- ・AI の原理や、その発祥、進化についての概要をよく理解することができました。
- ・AI は元々、完璧なものだと思っていたので、開発するのに失敗とかもあってびっくりしました。
- ・人間と AI との思考が違うことがよく分かった。AI は可能性を求める能力があるということで、私たちの手助けに大いに貢献し、また、それが人間の代替となるわけではないこともしっかり分かった。
- ・AI は今後世の中で重要な役割を果たすと思います。その社会で生きる身として、詳しくなる必要があると実感しました。
- ・AI が医療でも役立つと聞いて、最初はそんな大事なことに活用して信頼できるのかも思いましたが、根拠も同時に示してくれるという点で有能だと感じました。
- ・0 と 1 で判断しているコンピューターに、人格や特性が分かるのかつい疑ってしまいます。理解はされたいけれど、分析・把握はされたくないというか…。ただ、AI の研究や開発が進むほど、翻って私たちの体に備わっているシステムは大変複雑なプログラムで動いていることが分かって面白いと、動画を見て思いました。
- ・映画の予告を作らせると、人間が趣向を凝らして新たなものを足したり音声に躍動感を持たせたりして、面白いものを作ることができるのに対して、Watson はもともとあるものを切り出して並べる、などの面白みに欠ける作

業しかできず、そこから人間の感性に基づく思考を AI で再現する難しさがわかりました。人工知能は計算やデータ処理など、人間が行うことのできる行為の一部を人間より早く正確にできますが、人間にしかできないこともまだまだ多く、AI で補えるところは AI に手伝ってもらい、うまく使っていけたらより便利な社会になると思います。

- ・ SNS (Social Networking Service) の扱い方には十分気を付けないといけない世界なんだと改めて実感しました。プライバシーというのは簡単に侵害されてしまい、その結果これから先の人生にレッテルが張られたりすることは身近に起こり得るということを自覚して SNS を扱いたいです。

【課題発見】 分析的なトーンの上昇に関連

- ・ 約 10 年前の段階で Watson がこれほどまでの処理速度、正確度を有していたことはとてもすごいことだと感じました。しかし、現在の段階でもまだ少し低予算では AI の処理速度は遅いです。これからはいかに低コストで高性能を目指すのが課題になると考えます。
- ・ AI は人のストコーマ (心理的盲点) を指摘する素晴らしいものだと考えます。しかし、AI によって指摘されたことを理解・行動できるかどうかは人の能力次第だと感じました。また、テキスト入力での同時翻訳ですが、これも文字の打ち手の能力によって精度が左右されます。よって、高度な AI を完璧に使いこなすには人間にもそれ相応の能力が必要だと考えます。
- ・ 医学の AI 利用に関しては、最も合理的なやり方だと思った。しかし、動物の体に関して絶対はあり得ないので、そもそも論文が間違っていたり、例外が発生するリスクをできるだけ回避することも重要だと思った。
- ・ IT 業界の発達に対応する知識を身に着けることが、人類のこれからの課題になってくると感じました。

6. おわりに

本稿では、「データサイエンス概論」の授業の効果分析を、一クラス分 (42 名履修登録) のデータを基に定量的に試みた。授業中に沈黙する学生が多く、イン

タラクティブな授業が進めにくい場合でも、学生の理解度や感情を知るためには、スコアによるアンケートのみならず、コメントも集めて形態素解析を行い、さらに行間を読むような感覚で語調（トーン）分析を行うことが効果的であることを示した。毎回の授業中に体験したことに対して「家でもっと詳しくやってみたい」、分からなかったところは「家で復習したいと思いました」などのコメントも多く混じることから、自ら学習する意欲も増進していることが分かった。本稿で述べたように、15回分の授業の効果分析を初回と最終回のみで比較するだけでなく、途中回においても行うことで、その時のクラスの理解や感情の変化が分かり、その後の授業の進め方の調整に役立つことを示した。

(文責 辻)

参考文献

- (1) glassdoor: 50 Best Jobs in America for 2019,
https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm, 2019.
- (2) 日経 xTECH:
<https://tech.nikkeibp.co.jp/atcl/nxt/column/18/00001/02729/>, 2019.
- (3) 成城学園創立 100 周年サイト：日本アイ・ビー・エム東京基礎研究所と包括協定を結び、社会の発展に寄与します, 第 2 世紀への取り組み, 地域・社会連携,
<http://www.seijo100th.info/torikumi/chiiki/000082.html>, 2014.
- (4) 成城学園創立 100 周年サイト：成城大学で IBM 提供「データサイエンス概論」を開講, 第 2 世紀への取り組み, 教育改革,
<http://www.seijo100th.info/torikumi/kyoiku/000332.html>, 2015.
- (5) 辻 智：IBM Watson を文系大学の授業で使う, Rad-It21,
<https://rad-it21.com/ai/tsuji20180814/>, 2018.
- (6) 辻 智：コグニティブ・コンピューティングとデータサイエンス授業とのいい関係, 成城大学共通教育研究センター紀要, 成城大学共通教育論集, 11 (2018), 137-150.
- (7) 小橋 康章：学生はなぜ沈黙するのか, 成城大学共通教育研究センター紀要, 成城大学共通教育論集, 11 (2018), 151-160.
- (8) Microsoft Forms: <https://support.office.com/ja-jp/forms>
- (9) ファンブライト ラボ：ワードクラウド (テキスト),
<http://lab.fanbright.jp/wordcloud/text>

- (10) ユーザーローカル：テキストマイニングツール, <https://textmining.userlocal.jp/>
- (11) IBM Watson: Tone Analyzer, <https://tone-analyzer-demo.ng.bluemix.net/>
- (12) IBM Watson: Language Translator, <https://language-translator-demo.ng.bluemix.net/>