

データライフサイクルマネジメントに基づいた データリテラシー教育の提案

Proposal of data literacy education based on data lifecycle management

中鉢直宏[†]

Naohiro Chubachi[†]

[†] 高崎商科大学 商学部

[†] Faculty of Commerce, Takasaki University of Commerce.

要旨

政府の方針などに見られるように AI や機械学習の重要性が認知されるようになり、大学教育において、文系理系問わず情報系以外の一般教育としてデータリテラシーの科目が広く導入され始めている。しかし、社会が必要とするデータリテラシーに関する教育はどのような学習内容が相応しいのかまだ確立されていない。そこで、本稿ではデータリテラシー教育を考える際、狭義と広義のデータリテラシーを定義した。そして、広義のデータリテラシーを教育し、理解させるためにデータ発生から、活用、蓄積、保管、廃棄などを管理するデータライフサイクルマネジメントに注目し、それに基づいた学習教材の開発の必要性について提案する。

1. はじめに

現代社会において、AI や機械学習などのデータに基づいた情報技術が社会的に大きな役割を持つようになった。それを受けて文系理系を問わない一般教育において大学におけるデータに関する教育の必要性が叫ばれるようになった。データに関する教育環境は大きく変わる時期に差し掛かっている。文部科学省は、2016 年度に「大学の数理・データサイエンス教育強化方策について」を「数理及びデータサイエンス教育の強化に関する懇談会」において取りまとめて発表した。それを機に国としてはデータに関する教育を重視するように舵を切ったが、「一般教育では、数理・データサイエンスが社会における課題発見・解決や新たな価値の創出や様々な学問分野の発展に役立つものであることを認識し、それがどのような原理で行われているのかを理解できることを目指す。」[1]とあるが、データリテラシーではなく、データサイエンスの専門教育を主としたものであった。その後「AI 戦略 2019」[2]においては、デジタルトランスフォーメーションを背景に「文理を問わず、全ての大学・高専生（約 50 万人卒/年）が、課程にて初級レベルの数理・データサイエンス・AI を習得」「大学生、社会人に対するリベラルアーツ教育の充実（一面的なデータ解析の結果やAI を鵜呑みにしないための批判的思考力の養成も含む）」とデータ教育に関する目標が提示された。これは AI 戦略 2022[3]においても引き続き目標とされている。また、「AI 戦略 2022」では、第三部にある「2. 社会実装の推進に臨む姿勢」において次のことが記述されている。「“データが全て” ⇒ “ループの形成が重要”」「AI は、データをアルゴリズムによって学習又は処理するものである。このため、ややもすると「データが全て」であると言わんばかりに、AI の利活用においては膨大なデータを持っていなければ勝てないとの思い込みがある。データは確かに重要であるが、デジタル化された状況においてはサービスの提供を通じてデータを取得し、AI の強化、ひいてはサービスの向上につなげる手段も有効である。このため、データは重要ではあるけれども、それ以上に重要であるのは、サービス等の構築や提供の際に、AI を強化するデータ収集等を行うような持続的なサイクル（ループ）を形成するように配慮することである。」[3]。このことはこれまでの数学や統計などのデータサイエンスに特化したデータ教育の方向性とは異なる。このようなデータ教育を実現するにはデータサイエンス的アプローチよりも情報システム的なアプローチの方が向いていると考えた。特に一般教育としての情報教育においてこのような考え方は重要であり、データリテラシーにおいて学習すべき内容として取り上げるべきと考えている。

2. データリテラシーについて

筆者は2016年にデータ従事者に対してのデータリテラシーの必要性について言及している[4]。ここではデータ従事者とは、データに関連する作業を担うすべての人に対して、データリテラシー教育の必要性について述べており、現在の一般教育におけるデータ教育のニーズに通じていると考えている。日本では、データリテラシーという言葉はあまり広く使われていなかったが、日本以外では、以前からデータリテラシーについて取り組んでおり、情報リテラシー、統計リテラシー、データリテラシーは分けて考えている場合もあり、データリテラシーは、コンピュータサイエンスや経営情報システム(MIS)よりも技術的ではないが、データにアクセスし、変換し、操作するためのさまざまなツールを理解(構造化照会言語(SQL)、リレーショナルデータベース、データ操作技術、統計ソフトウェア、およびデータ表示ソフトウェアを理解することである[5])とある。日本では、データリテラシーの科目に関して標準的な科目の内容が確立されておらず、これから一般教育として設置するデータリテラシーの科目などをAIや機械学習などの影響から統計やデータサイエンスの基礎的な位置づけでとらえている科目が多い。しかし、海外の例では統計・データサイエンス以外の内容が多く含まれていることが分かる。そこで、日本にデータリテラシーとして統計・データサイエンス以外の重要性を考えるために、既存の統計・データサイエンスの基礎としてのデータリテラシーを狭義のデータリテラシーとし、一般教育にふさわしいデータリテラシーを広義のデータリテラシーとして定義して捉えることにする。筆者は、広義のリテラシーのもとになる考えは、データライフサイクルとその管理ではないかと考えている。

3. データライフサイクルマネジメント教育の必要性

本稿では、データライフサイクルというモデルに注目する。「データライフサイクルとは、情報システムのデータの流をその作成から最初の保存まで、それが時代遅れになり削除されるまでの間、管理するためのポリシーベースのアプローチである」[6]とある。そして、先に挙げたようにデータに関するループの形成が重要性や、AIを強化するデータ収集等を行うような持続的なサイクル形成という目的を達成するためには、どのようにデータライフサイクルを意識させる教育が重要になると考える。図1には本研究が想定する狭義のデータリテラシーと広義のデータリテラシーをデータライフサイクルとともに示している。学習者はデータリテラシーとしてこのようなデータライフサイクルを理解、管理するデータライフサイクルマネジメントという考えを学ぶことが重要なのではないかと考える。改めて現在のデータリテラシーは、狭義のデータ教育であり、加工・分析・提示の部分のみにフォーカスされがちである。しかし、AIなどこれからのデータ主導社会に求められているのは、図1で示すようなデータの収集、統合、蓄積、共有、破棄するまでループ型のデータ理解であり、社会システムとしてAIなどの事例を理解するためには狭義のデータリテラシーのような一過性のデータ利用の理解にとどまらず、データが蓄積され、データを再利用するなど、データに関して長期的な視野を身に着ける広義のデータリテラシーを学習すべきであると提案する。

4. データライフサイクルマネジメント教育の実現のために

先に示したような広義のデータリテラシー教育の実現には、データライフサイクルマネジメントの必要性を広く理解させることが重要である。長くコンピュータなどを使用してデータを保存管理してきている人がこの考えを理解してもらうことはそう難しくないと考えられる。しかし、そういった経験が不足している学習者にとって、重要なデータが消えてしまったり、数年後に突然過去のデータが必要になったり、またデータを責任をもって削除する経験などは乏しいと考える。これは、例えば課題のレポートなどで提出されるファイル名一つとっても、ファイルがどのようなものか、採点する側がそれを管理すること、後から見返す可能性への配慮など考えられていると思えないものが散見された経験をなどからも推

測される。まず学習者がデータについてどのような経験や考えを持っているか把握する必要がある。各自がデータ取得、データ加工だけではなく加工後のデータの保管方法、セキュリティ対策、バックアップ、データの削除などについて、学習者とくに大学生にとってこのデータライフサイクルに即した感覚や考えを実生活で持っているか知り、それを活かした教材作りが重要になってくると考える。データライフサイクルマネジメント教育に必要なのは会社の情報システムにおけるそれと異なり、大学生にとって身近なデータ管理を想定した、データライフサイクルマネジメントの事例にあてはめ提供することが出来ないか、例えば、データライフサイクルをシナリオ的に体験させて、そのマネジメントの重要性を理解させるような教材を開発し広く、理解してもらえる可能性を探るなどが考えられる。

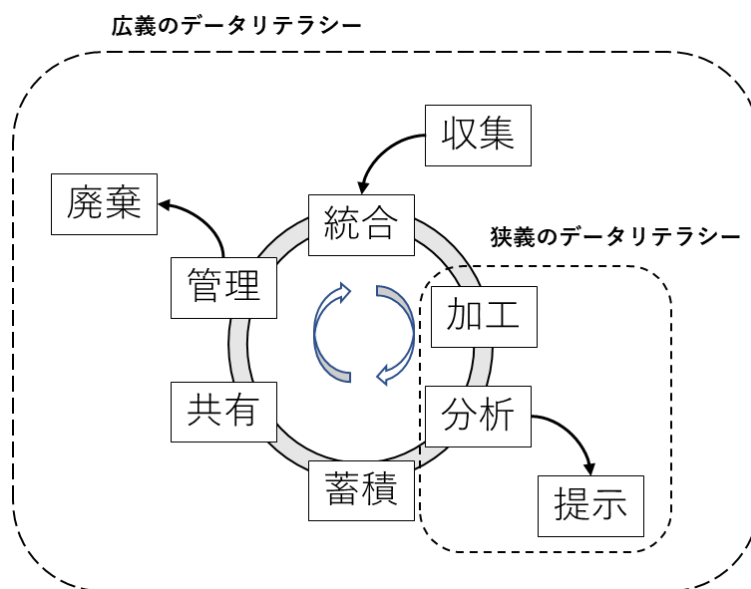


図1 データライフサイクルとデータリテラシーとの関係図

5. まとめ

政府の方針などから、AIや機械学習の重要性が認知されるようになってきた背景を受け、データ関連の教育、特に大学教育において、情報系以外の一般教育としてデータ関連の授業導入が必要とされている。そこで一般教育としてデータリテラシーと称した科目などが多く導入されるようになってきている。現状大学の一般教育においてデータリテラシーに関する教育はどのような学習内容が相応しいのかまだ確立されていない現状であり、データリテラシー教育を考える際、狭義と広義のデータリテラシーを定義し、広義のデータリテラシーの必要性を述べた。それをその教育を実現するためにデータ発生から、活用、蓄積、保管、廃棄などのデータライフサイクルマネジメントに注目し、それに基づいた教材開発の必要性を示した。

謝辞

本研究はJSPS 科研費21K02867の助成を受けている。

参考文献

- [1] 統合イノベーション戦略推進会議, “人間中心のAI 社会原則”, <https://www8.cao.go.jp/cstp/aigensoku.pdf>, 2019
- [2] 統合イノベーション戦略推進会議, “A I 戦略2019”,

- <https://www8.cao.go.jp/cstp/ai/aistrategy2019.pdf>, 2019
- [3] 統合イノベーション戦略推進会議, “A I 戦略2022”,
https://www8.cao.go.jp/cstp/ai/aistrategy2022_honbun.pdf, 2022
- [4] 中鉢直宏, “データ従事者の育成を目的としたデータ・リテラシー教育の試み”, 情報処理学会第135回コンピュータと教育研究会, 2016
- [5] Milo Schield, “Information Literacy, Statistical Literacy and Data Literacy”, IASSIST quarterly / International Association for Social Science Information Service and Technology 28(2):7-14,
https://www.researchgate.net/publication/253283220_Information_Literacy_Statistical_Literacy_and_Data_Literacy, 2004
- [6] Gray, Jonathan and Gerlitz, Carolin adn Bounegru, Liliana. “Data infrastructure literacy.”, Big Data & Society. (2018).