

社会システム・コネクティッド領域におけるデータ活用と人材育成・知財活用

福島真太郎 (トヨタ自動車株式会社社会システム PF 開発部 主査)

Data Utilization, Human Resource Development, and Intellectual Property Utilization for Social Systems and Connected Domains

Shintaro Fukushima

General Project Manager, Toyota Motor Corporation, Social System PF Development Div.

【要旨】 現在、自動車業界は「100年に一度の大変革の時代」を迎えており、産業構造が大きく変貌しようとしている。「CASE」(Connected (コネクティッド), Autonomous/Automated (自動化), Shared (シェアリング), Electric (電動化))に代表されるように、コネクティッド、自動運転、MaaS (Mobility as a Service)、スマートシティ、電気自動車、カーボンニュートラル等の領域において、新たなビジネスやサービスの創出が期待されている。それに伴い、車両や都市から大量のデータを収集・蓄積した上で解析し、商品やサービスに結び付ける人材の育成が急務となっている。それとともに、サービスや商品の実現に向けたデータ解析技術、データ基盤・システム実装技術などが必要とされ、発明や知財の観点からも考慮すべき事項を生み出していると感じる。本稿では、こうした取り組みにおけるデータ活用の人材育成や知財活用について、所感も交えながら説明する。

【キーワード】 データ解析 機械学習 人工知能 社会システム コネクティッド

【Abstract】 The automotive industry is currently undergoing a “once-in-a-century transformational period,” and its industrial structure is undergoing a major transformation. It is highly expected to create new businesses and services in areas such as connected domain, autonomous driving, MaaS (mobility as a service), smart city, electric vehicle, and carbon neutrality. Accordingly, there is an urgent need to develop human resources who can collect and accumulate a large amount of data from vehicles and cities, analyze it, and utilize it to products and services. At the same time, I feel that data analysis, data infrastructure and IT system implementation technologies are necessary to realize the services and products, which creates matters to be considered in terms of inventions and intellectual property. This paper describes the human resource development and utilization of intellectual property in the activity.

【Keywords】 Data analysis Machine learning Artificial intelligence Social system Connected

1. はじめに

現在、自動車業界は「100年に一度の大変革の時代」を迎えており、産業構造が大きく変貌しようとしている。「CASE」¹ (Connected (コネクティッド), Autonomous/Automated (自動化), Shared (シェアリング), Electric (電動化))に代表されるように、

コネクティッド、自動運転、MaaS (Mobility as a Service)、スマートシティ、電気自動車、カーボンニュートラル等の領域において、新たなビジネスやサービスの創出が期待されている。それに伴い、車両や都市から大量のデータを収集・蓄積した上で解析し、商品やサービスに結び付ける人材の育成が急務となっている。それとともに、サービスや商品の実現に向けたデータ解析技術、データ基盤・システ

ム実装技術などが必要とされ、発明や知財の観点からも考慮すべき事項を生み出していると感じる。

筆者が所属するトヨタ自動車株式会社社会システム PF 開発部 InfoTech-AS² では、これまでコネクティッドに軸足を置きながら、前述の領域を含む幅広い商品やサービスを対象として、機械学習 (machine learning) やデータマイニング (data mining) 等のデータ解析技術の研究開発や先行開発、またそのサービス応用に取り組んできた (NTT グループ・トヨタ自動車株式会社 2021; Yong, Wakabayashi, Okayasu, Miki, Sasai, Inoue, Fukushima 2022; Jiang, Wang, Yong, Jeph, Chen, Kobayashi, Song, Fukushima, and Suzumura 2023; Fukasawa, Yokoi, Yamada, Sasai, and Fukushima 2023; Xu, Suzumura, Yong, Hanai, Yang, Kanazashi, Jiang, and Fukushima 2023)。今後は、コネクティッド領域で培った知見や技術蓄積等をいかして、社会システム領域へと範囲を拡大していく予定である。本稿では、こうした取り組みにおけるデータ活用の人材育成や知財活用について、所感も交えながら説明する。なお、本稿で述べる見解は著者個人のもので、所属する企業や団体を代表するものではないこと、また筆者は発明や知財そのものを専門とするのではなく、機械学習・データマイニング・統計学等を駆使したデータ解析・活用分野における研究者、技術者であり、その立場や取り組みと関連付けながら私論を展開することに留意されたい。

2. データ活用プロセス

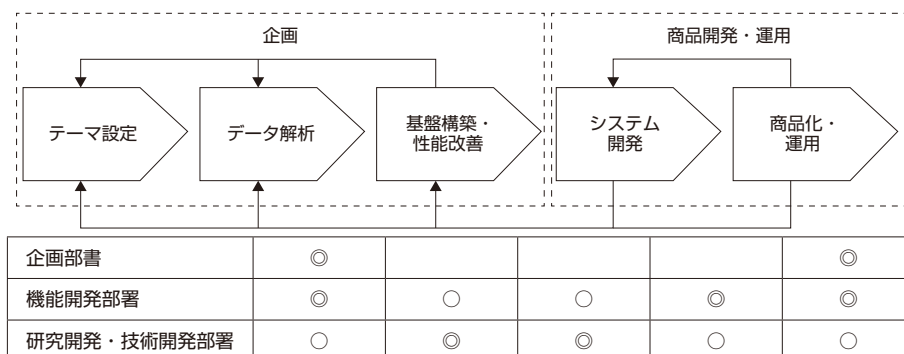
本章では、データ活用の人材育成や知財活用について説明する前に、まず筆者が従事する領域におけるデータ活用のプロセスの全体像を俯瞰する。データ活用のプロセスはこれまでもいくつか提案されているが³⁻⁵⁾、筆者の周囲では図 1 に示すように大きく以下 5 つの工程から構成される。カッコ内は、主として担当する部署を表す。

- テーマ設定 (企画部署／機能開発部署)
- データ解析 (研究開発・技術開発部署)
- 基盤構築・性能改善 (研究開発・技術開発部署)
- システム開発 (機能開発部署)
- 商品化・運用 (企画部署／機能開発部署)

「企画部署」は商品やサービスの企画を実施する部署、「機能開発部署」は企画の実現に向けて IT システムを含めた開発を実施する部署、「研究開発・技術開発部署」は企画の実現に向けてデータ解析やそれを実行する基盤のアーキテクチャや処理フロー等を明確化する部署である。筆者が所属する組織は、この分類では最後の研究開発・技術開発部署に相当する。

図 1 にも示されているように、これらの工程は一方に進むのではなく、適宜、前工程に戻って検討が必要になることに注意されたい。また、当初はテーマの設定もアイデアレベルから始まり、「データ解析」や「基盤構築・性能改善」の工程を経ることにより実現可能性が向上し、企画が具体化するケー

図 1. データ活用のプロセスと大まかな役割分担



凡例 ◎: 中心的な役割で推進 ○: サポート的な役割で推進

スも珍しくない。そのため、図1に示されるように「テーマ設定」、「データ解析」、「基盤構築・性能改善」の最初の3工程は一体となり、大きく見れば企画のフェーズであると言える。それに対して、最後の2工程である「システム開発」、「商品化・運用」は、商品化に向けた開発および商品リリース後の運用のフェーズである。以下では、上記5つの工程について概観する。

2.1. テーマ設定

商品やサービスの開発に向けて、テーマを設定する工程である。社会や市場の動向や自社サービスに対する顧客のフィードバック等を把握し、商品やサービスのあるべき姿と現状のギャップを認識した上で、提供する商品像や展開イメージ等について検討する。

従来、本工程は商品・サービスの企画部署や機能開発部署が主体で担っていたことが多い。一方で、社会システムやコネクティッドの領域のように都市や車両から大量のデータが収集され、商品やサービスの実現性がデータの解析結果やそのデータパイプライン（データの収集・蓄積、抽出、前処理、機械学習・データマイニング・統計学等のモデル、出力等を含む一連のデータ処理を実行するための基盤やITシステム）の性能やコスト等が技術的な実現性を大きく左右する領域においては、これら技術の実現性を示しながら企画を具体化し、その結果を例えばサービス企画書のような形でまとめることも必要となってきた。なお、実際の企画では、上記で説明した商品性だけでなく、利益やコスト、システムの開発体制等も考慮する必要がある。

2.2. データ解析

商品やサービスの実現に向けて、収集・蓄積したデータを抽出し、機械学習・データマイニング・統計学等の手法を駆使して解析を行いながら、実現性を評価する。昨今は、車両から様々なデータを取得することが可能になっている。一例として、CANデータ（Controller Area Network）、プローブデータ、車載カメラで取得される動画像データ等が挙げられる。

2.3. 基盤構築・性能改善

前工程である「データ解析」の結果、ある程度有望と判断された場合は基盤構築に移行し、データパイプラインを組み立て、性能やコストを検証する。この工程の目的は、構築した機械学習・データマイニング・統計学等のモデルを含むデータパイプラインが現実的な性能やコストでの実現性を有しているかについて検証し、必要があれば改善し実現性を向上させることにある。

「データ解析」の時点では、構築したモデルの検証対象とする車両の台数や期間、エリア等は実際のサービスでの想定と比較すると限定的であるケースも多い。そのため、技術の実現性は精度等で測られることが往々にしてある。一方で、実際のサービスで想定される規模の車両や利用シーンを対象としたときに、コストが非現実的になることがしばしば生じる。こうした場合、精度だけでなく、センター側（クラウド）と車載側の双方、およびこれらの連携において性能やコストの改善まで実施しなければ、商品やサービスの実現性は非常に低くなってしまう。そこで、本工程を設けて、企画としての実現性を向上させている。また、次工程である「システム開発」に着手するにあたり、研究開発・技術開発部署からはデータパイプラインの構築や検証をもとにして、機能開発部署に対して例えば検証報告書、また必要であれば要件定義書の形で成果を提示する。

2.4. システム開発

前節までの工程を経て実現性の目途が立ってきたら、企画部署や機能開発部署に対して成果を提示し技術的な実現性を示した上で、企画としての目処付けを行う。そして、商品性や利益、技術成立性、開発体制等がクリアできたと判断されたら、システム開発に移行する。システム開発は機能開発部署を中心に行われるが、その間も構築したモデルの精度改善、基盤の性能改善等で研究開発・技術開発部署も関与する。

2.5. 商品化・運用

システム開発が完了したら商品をリリースし、その後システムの運用を実施する。この工程は、企画

トの育成について説明する。機械学習・データエンジニアについては説明を割愛するが、図2の「基盤構築・性能改善」に示したように、クラウド技術や車載技術、およびそれらを結びつけた上でのデータフローの構築や改善を行うためのエンジニアリングスキルが必要とされる。

本稿では紙面も限られているため、AIやデータサイエンスに中心的に携わるデータサイエンティスト

データサイエンティストに求められるスキルや経験は多岐に亘る^{4,6}。図2にも示されているように、以下に代表的なものを挙げ、それぞれの育成上のポイントについて説明する。

- なお、筆者の環境においては、データサイエンティストに相当するメンバーは30名程度であり、全員修士号以上の学位を有する。メンバーのバックグラウンドは、数学、物理、化学、情報学、工学（材料等）等の多岐に亘っている。

データや解析結果を理解した上で、仮説を立案し検証するスキルである。解析の手法としては集計や可視化等の比較的単純な場合も含まれるが、重要なのはデータや解析結果から何が読み取れるのか、そして次に深掘りするポイントを見極めて仮説を構築し検証する、といった科学的な検証プロセスを組み

	テーマ設定	データ解析	基盤構築・性能改善	システム開発	商品化・運用
	企画部署・機能開発部署と連携	仮説立案・検証 データ抽出・加工 機械学習・データマイニング・統計学 有用性評価 等	データパイプライン設計・構築 分散処理技術 クラウド 車載器 要件定義 等		企画部署・機能開発部署と連携
データサイエンティスト					
機械学習・データエンジニア					

立て実行する能力である。機械学習やデータマイニングの手法と比べると、このスキルは一般化しづらくテーマによって必要とされる内容も変わるため、重視されにくい傾向にあると感じる。しかし、実際にデータから知見を得て商品やサービスの実現性の提示に向けて解析の方向性を見出すために、非常に重要なスキルである。そのため、メンバーにはデータや結果をよく見て、目的に照らし合わせて計画を立てるように指導している。個人的には、このスキルは企業でOJT（On the Job Training）により育成可能な部分もある一方で、大学院等での研究経験が大いに活かされる側面もあると感じている。

3.1.2. データ抽出・加工

SQLやHive等のスクリプトを記述してデータベースから解析に用いるデータを抽出するスキル、および前述のツールに加えてPython等も交えながら抽出したデータの解析に適した形式への整形や外れ値・欠損値等への対応等を行うスキルである。このスキルは、経験を積むことで一定のレベルに達することが多いと感じている。そのため、新規にジョインしたメンバーには、まずはデータの仕様や統計値等の理解を通してデータに慣れてもらうことも多い。一方で、データ解析に適した効率的なデータ構造や中間データの生成等については、ある程度、後工程となるデータ解析を意識した上で行う必要があるため、経験を必要とする。そこで、経験として3-5年程度は有するメンバーをアサインすることが多い。

3.1.3. 機械学習・データマイニング・統計学

機械学習、データマイニング、統計学等の解析手法を用いてモデルの開発や評価を行うスキルである。このスキルは、アカデミックな内容の理解に加えて、目的やデータの特性に応じて適切な手法を選択し、その結果を評価するという経験値も必要である。OJTで経験を積んだメンバーがデータの特性を把握した上で手法選択の指導を行うほか、手法に関する知識を向上するために専門的な内容の教育講座（統計数理研究所の公開講座³等）、書籍、論文等の情報ソース等を勧めるなどしている。また、メンバー間で先端的な技術トピックの論文を紹介し合う勉強会も開かれており、良い刺激となっている。

3.1.4. 解析を実行するためのコーディング

プログラミング言語を使用して解析を実行するスキルである。現在、機械学習やデータサイエンスで頻繁に使用されているプログラミング言語として、Python⁴が挙げられる。Pythonには深層学習（deep learning）のフレームワーク（PyTorch⁵・Tensorflow⁶・Keras⁷等）、機械学習（scikit-learn⁸）、データ解析（Pandas⁹）、可視化（matplotlib¹⁰・seaborn¹¹等）などのデータ解析で頻繁に用いる機能を実装したライブラリが揃っており、典型的な手法を試すのであればコーディングの負荷が過大になることなく実行することが可能である。一方で、例えば統計学であればR¹²、高速な科学技術計算や数値計算が必要であればJulia¹³といったように、適材適所でプログラミング言語を選択することも重要である。

昨今は、Jupyter Notebook/JupyterLab¹⁴のようにノートブック形式で探索的に解析を行う環境を簡単に使用できる。メンバーのコーディングに対しては、Jupyter Notebook/JupyterLab上でレビューを実施することも多い。その際、データ解析は試行錯誤を伴うプロセスであるため、一般的なソフトウェア開発におけるコードレビューとは目的や方法が異なっている。例えば、試行錯誤の途中においてはデータ解析の目的と処理やその結果の解釈の妥当性、データ構造やその保持方法の効率性を中心にコメントする。そして、ある程度解析方法に目処が立ってきたら、典型的な処理のクラスや関数の設計等にまで踏み込んでフィードバックしている。

3.1.5. 解析結果を組み込んだ商品やサービスの有用性評価

モデリングの結果、構築したモデルの有用性を評価するための計画策定、実施、その結果を考察するためのスキルである。有用性の評価では、予測精度だけでなく、商品やサービスとして見たときにどうかという視点が必要不可欠である。そこで、POC（proof of concept）の実施やMVP（minimum viable product）の構築を通してユーザ評価を実施し、サービスとしての嬉しさや機能としての過不足を評価するとともに、データ解析の出力結果に対するユーザの受容性や違和感等を把握した上で、改善すべきポイントを洗い出し、次の解析フェーズでの取り組み

を具体化する。通常は、このプロセスを何周か回しながら徐々にサービス像の具体化とそれに伴う処理を明確化していく。筆者の周辺を見ると、このスキルの獲得には少なくとも数案件で経験を積む必要があり、2-3年以上の経験を有するメンバーと経験値の浅いメンバーを織り交ぜながらプロジェクト組成することが多い。

3.2. 育成の方向性

全般を通して、経験の浅いメンバーにはデータの抽出や解析を、ある程度経験を積んだメンバーには解析計画立案や解析の結果構築したモデルを用いたユーザ評価の計画立案を中心にアサインすることが多い。また、それ以上の経験を積んだメンバーには、企画部署や機能開発部署と折衝しながらテーマの設定、企画の具体化、システム開発の要件定義支援等の業務まで担当してもらう。

最後に説明した熟練メンバーの担当範囲については、研究開発・技術開発部署としてデータの解析やモデルの構築だけに止めてしまうと成果が活用されにくい。前工程や後工程と密に連携しながら実現性を向上させている点とも関係している。例えば、「基盤構築・性能改善」の性能やコストの評価結果と合わせて、企画部署や機能開発部署に技術的な実現性を提示し、設定したテーマを実現する上での課題を抽出し改善するサイクルを回していく。さらに、前工程である「テーマ設定」に対しても適切な課題を設定できるように、データ解析の観点を中心に企画部署や機能開発部署を支援することもある。例えば電気自動車の動向について、メンバー自身による机上の調査に加えて実車の走行を実施し、提供する機能の実態の把握や比較等を行う。それを通して、開発すべき商品価値の明確化、目標とする精度の設定に活用するなどの工夫を行っている。こうした作業を行うことにより、データサイエンティスト自身が自ら携わる商品やサービスの価値を認識したうえでデータ解析に取り組むことができるようになってきている。

以上の取り組みを通して、実作業レベルからプロジェクトリーダーレベルまで段階的にキャリアアップする仕組みを提供している。

4. 知財への対応

本章では前章までとは多少視点を変えて、私見も交えながら筆者が従事する領域でのデータ解析・活用における発明や知財への考え方、対応について論じてみたい。これまで説明してきたデータ解析技術、データ基盤・ITシステム実装技術に関連して、ビジネス面、技術面の双方に関して新たな発明が生まれる可能性がある。一方で、個人的には、従来と同様に、発明を公知にすべきかについては、慎重な判断が必要な場面が少なくないと感じている。特に、機械学習を含む人工知能（artificial intelligence, AI）のアルゴリズムやモデルを搭載したITシステムに対する特許侵害の立証は一般的には容易ではない。そのため、自社内に留めるケースも多くあると思われる。また、機械学習を含むAIのアルゴリズムは、基本的には学習に使用するデータから帰納的に仕様が決まり、従来の演繹的なシステム開発とは異なるアプローチを採らざるを得ない。こうした背景から、近年では、機械学習工学¹⁵（丸山2019；柿沼・竹内・土橋・中川・原・堀内・鷲崎2022）、AIソフトウェア工学（佐藤・小川・来間・明神2021；澁井2021）、機械学習の品質管理や保証のガイドライン策定（QA4AIコンソーシアム2020；産業技術総合研究所2020）や研究（中島2020；中島2022）も活発に行われている。こうした動向を敏感にキャッチアップし、適合していくことも重要である。

本稿では、主に技術面にフォーカスして、技術要素を「解析プログラム」、「データ」、「ITシステム・実装」の3つに大別し、これらの技術要素の単体または組み合わせにより発明が生まれうるパターン例を挙げた上で、それぞれのパターンに対する私見を論じることとする。ここで、解析プログラムとは、入力データに対してデータ処理を行い、出力を返すプログラムを指すこととする。筆者が従事している社会システム、コネクティッド領域におけるデータ解析・活用分野では、3.1.3節で説明した機械学習、データマイニング、統計学等のアルゴリズムが使用されることが多いが、これらのアルゴリズムやデータ処理だけでなく、データの前処理や加工（外れ値

や欠損値への対応、データの正規化（標準化）等）なども含んだプログラムを指すこととする。次に、データとは、一次的に取得される「生データ」や、学習に使用される「学習データセット」（訓練データセットなどとも呼ばれる）などがあるが、本稿ではこれらを総称して「データ」と呼ぶことにする。筆者が従事する分野においては CAN データ、プローブデータ、動画画像データ、およびそれらを用いて生成された学習データセットなどが挙げられる。最後に、IT システム・実装については、前述のデータの収集・蓄積部、データのパイプラインを実装した解析基盤、ユーザが解析結果を確認するための表示部なども含む構成などを指すものとする。

本稿では、上記 3 つの技術要素の単体、または組み合わせにより、発明のパターンを以下 2 つに大別する。

- ドメインやデータの性質を積極的には利用しない発明（「解析プログラム」が中心）
- ドメインやデータの性質を積極的に利用した発明（「データ」×「解析プログラム」、または「データ」×「解析プログラム」×「IT システム・実装」）

1 点目については、AI を利用する場合は狭義には「AI コア 発明」と呼ばれることもある（特許庁 2023）。それに対して、2 点目は「AI 応用発明」と呼ばれることもある（特許庁 2023）。

以下では、各パターンにおける発明の考え方について説明した後、今後の展望について述べる。

4.1. ドメインやデータの性質を積極的には利用しない発明

まず初めに、ドメインやデータの性質を積極的には利用しない発明について説明する。これは、学習方法やそのパラメータ等を含むデータ解析の処理に重きを置いた発明である。機械学習を含む人工知能の知財や特許については、これまでもいくつかの考え方や事例等が提示されている（浅井 2019；福岡・濱野 2019；松下 2019）。また、機械学習の中でも、近年は特に深層学習に注目が集まっている。深層学習の中でも特に成功しているのは、深い層のニューラルネットワーク（neural network）を構成し、デー

タで学習するアプローチである。深層学習のアルゴリズムについては、近年、機械学習、表現学習（representation learning）、データマイニング等の分野で精力的に研究が行われており、例えば米国では特許も多く出願されている。一例として、深層学習の学習時に用いるドロップアウト（dropout）（Nitish, Hinton, Krizhevsky, Sutskever, and Salakhutdinov 2014；Google Inc. 2016）、バッチ正規化（batch normalization）（Ioffe and Szegedy 2015；Google LLC2019a）、Transformer（Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, and Polosukhin 2017；Google LLC2019b）などが挙げられる。

筆者は、解析プログラムが中心の発明は、ある程度汎用的なタスクが対象であり、その中でも当該タスクを実行するには学習方法やパラメータの設定等が不可欠であることが重要と考えている。例えば、ドロップアウトやバッチ正規化は、特定の分野または領域におけるデータの特性に著しく依存したタスクではなく、深層学習を用いて学習する際に広く用いられる技法である。また、画像認識については、実際の応用においてはデータの性質や各領域の特性などを考慮して、個々のチューニングが必要になるケースも多いと思われるが、データ処理・解析技術として捉えれば、ある程度汎用化・一般化が可能と捉えられる。一方で、このパターンの発明は、学習手法やパラメータ等を具体的に開示しており、第三者が発明を事実上自由に利用可能な状況が生じかねない。したがって、対外的な技術力の宣伝等を目的としなければ、一般的には特許出願のメリットは大きくないと考えられる。

筆者が従事している分野でこのパターンに属する発明はそれほど多いわけではないが、例えば学習済みのニューラルネットワークの入出力範囲を効率的に保証するプログラム（トヨタ自動車株式会社 2023, Toyota Motor Corporation 2023a）などが挙げられる。入力データの集合が与えられた際に学習済みのニューラルネットワークの出力データの範囲を厳密に同定する先行技術（Xiang, Tran, and Johnson 2017）の計算量が膨大であり現実的ではないため、計算量の低減という観点で改善余地がある点、また、学習済みモデルにデータを入力して得られる出力

データの信頼性を評価するものではない点に着目して、計算量を低減しつつ、出力データの信頼性評価を可能とすることを目的となされた発明である。

4.2. ドメインやデータの性質を積極的に利用した発明

次に、ドメインやデータの性質を積極的に利用した発明について説明する。このパターンでは、データ解析プログラム単体ではなく、データの性質を利用したり、うまく合わせたりする効率的なデータ処理・解析技術を提案する。このパターンは、4.3節で説明するパターンと並び、筆者の周囲では多く行われている。CAN データやプローブデータのように非定常な時系列や位置情報の特性、また車両挙動や運転操作、交通などの特性を考慮し、これらを活かす形でデータ解析プログラムを考案する。また、複数のデータを組み合わせる解析するマルチモデルデータ解析なども、このパターンに属する発明となる。その際、記載要件や進歩性をどのように捉えるかという点が検討課題になる。

特許庁が公開している特許・実用新案審査ハンドブックの付属文書（特許庁 2015, 特許庁 2018）では、AI を用いた特許・実用新案の記載要件と進歩性について、事例も交えて説明されている。記載要件については、以下 4 点が挙げられている：(i) 教師データの間の相関関係に関するもの、(ii) 相関関係が出願時に技術常識であるもの、(iii) 相関関係が明細書等に記載された統計情報に裏付けられているもの、(iv) 相関関係が実際に作成した人工知能モデルの性能評価により裏付けられているもの。(i)-(iii) については、データ間の相関関係が出願時の技術常識、統計情報、AI の性能評価等によってそれぞれ裏付けられるとされている（事例 47, 49, 50）（浅井 2019）。一方で、進歩性については、以下 3 点が挙げられている：(a) 単純な AI の適用に関するもの、(b) 教師データの変更に関するもの、(c) 教師データに対して前処理を行うもの。体感的には、記載要件については、ケースに応じて (i)-(iv) のいずれも取りうるが、(iv) のように性能評価により裏付けられる場合は比較的少ないように感じる。また、進歩性については、(a) のように単純な AI の適用に関

するものについては比較的少なく、(b) や (c) のように、教師データを変更したり前処理を行ったりするものが多い。

筆者が従事する分野において、このパターンに属する発明の例として、（トヨタ自動車株式会社 2022；Toyota Motor Corporation 2022）などが挙げられる。これは、検索文が表す運転シーンに対応する動画像と車両挙動データのペアを検索するための検索装置、学習装置、検索システム、検索プログラム、学習プログラムを提案している。すなわち、運転者の運転行動データをクエリとして、類似する運転行動データを抽出して出力する技術や、検索文をクエリとして類似した動画像を検索する技術を提供することを目的とした発明である。

なお、記載要件については、将来的には、機械学習工学、特に品質保証に関する知見の集積が、記載要件の充足性の客観的な評価に寄与すると考えるが、当面は、出願時の技術常識と審査実務とを不断に注視しつつ、明細書の記載を整えざるを得ないという見方もある（浅井 2019）。筆者の現場においても、同様の対応を行っている状況との認識である。

本パターンは、さらにシステム構成などにも踏み込む場合がある。筆者が従事する分野においては、例えば学習がセンターサーバ側においてバッチ処理で行われるのか、それとも車両側でストリーム（オンライン）で行われるかにより、データ収集・蓄積や学習のフローや形態が大きく変わる。本パターンでは、このようにデータが「どこから収集され」、「どこに蓄積され」、「そのデータを用いてどこでどのように学習されるか」までカバーしている。筆者が従事する分野において、このパターンに属する発明の例として、地点の危険予測装置及び危険予測システム（トヨタ自動車 2023, Toyota Motor Corporation 2023b）などが挙げられる。これは、十分にデータを確保できていない地点において危険予測する場合でも、類似する属性を有する地点のデータを集約することで予測精度を向上させることを可能とする危険予測装置及び危険予測システムを提供することを目的となされた発明である。

4.3. 今後の展望

以上では、2つのパターンに分類して、知財や特許の考え方について筆者の経験を交えて説明した。最後に今後の展望について、ここでは簡単に3点触れてみたい。

1点目は、対象が社会システムへと広がる中で、異種のデータの掛け合わせが行われ、その中で知財や特許が生まれる可能性があるという点である。例えば、電気自動車の充電に伴う行動を考えた時に、自車の充電量だけでなく、充電ステーションの有する情報や、道路事業者が把握している交通状況等との掛け合わせが生じてくる可能性がある。

2点目は、1点目とも関連し、データ収集・蓄積、解析、活用の一連の流れが自社内に留まるケースばかりではなく、官公庁の多くの事業者との連携が増加すると予測される。こうした連携においては、生データを共有できるケースばかりではないだろう。効果的、効率的に学習を行うためのフレームワークとして、例えば連合学習（federated learning）（Jakub, Brendan, and Daniel 2015；Yang, Liu, Chen, and Tong 2019）がある。データのプライバシー保護を行いながら、複数のデータソースを組み合わせる効果的な学習を行う方法や、それに関連したプライバシー保護、セキュリティの技術などは、これまでも精力的に研究や発明が行われている。今後、実用化を見据えた際に、社会システムのスケールでの知財や特許を考える必要性が生じる可能性があると考えている。

3点目は、データ処理・解析やそのデータパイプラインだけでなく、通信やセキュリティ、計算基盤等と組み合わせる方向性である。例えば標準化の動向と合わせた知財などが挙げられる。

5. おわりに

本稿では、筆者が従事する社会システムおよびコネクティッド領域におけるデータ活用のプロセスで必要なスキルや知識、人材育成の取り組み、および知財の考え方について、私見を交えて説明した。「CASE」に象徴されるように自動車業界はビジネス面からも技術面からも、業態が大きく変わろうとし

ている。その中で、データ利活用を含むソフトウェア開発や人材育成、知財への対応などがますます重要になることは論をまたない。都市や車両からのデータの収集・蓄積基盤が整いつつある昨今、これらのデータを利活用できる人材の重要性は今後ますます高まっていくであろう。その中で、データ解析や基盤の技術に加えて、商品・サービスを実現するために全体を俯瞰して技術的な実現性を考慮した上で、マネージしていく人材の必要性を強く感じている。また、こうした領域における発明や知財の考え方についても、社会的な動向と合わせて、戦略的に行う必要性を痛切に感じている。それに向けて、研究開発・技術開発部署としても、企画や機能開発を実施する部署の人材との相補的な関係だけでなく、互いにスキルがオーバーラップする人材の育成等も検討していく必要があるのではないかと考えている。また、近年では、AIや機械学習の技術の進展は目覚ましく、生成系AIなどの活用も期待されている。車両や都市から膨大なデータが収集され、その活用に期待が集まる中、こうした取り組みの推進は非常にチャレンジングであり、今後も果敢に挑戦していきたい。

注

- 1 トヨタ自動車株式会社企業公式サイト「モビリティ CASE」。
<https://global.toyota.jp/mobility/case/>
- 2 トヨタ自動車株式会社大手町オフィス サイト。<https://www.toyota-tokyo.tech/>
- 3 <https://www.ism.ac.jp/lectures/kouza.html>
- 4 <https://www.python.org/>
- 5 <https://pytorch.org/>
- 6 <https://www.tensorflow.org/>
- 7 <https://keras.io/>
- 8 <https://scikit-learn.org/stable/>
- 9 <https://pandas.pydata.org/>
- 10 <https://matplotlib.org/>
- 11 <https://seaborn.pydata.org/>
- 12 <https://www.r-project.org/>
- 13 <https://julialang.org/>
- 14 <https://jupyter.org/>
- 15 機械学習工学研究会 <https://sites.google.com/view/sig-mlse/>

参考文献

- Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, and Rüdiger Wirth (2000) "The CRISP-DM User Guide" <https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISP-WP-0800.pdf>
- Yuta Fukasawa, Yuho Yokoi, Kota Yamada, Takeyuki Sasai, and Shintaro Fukushima (2023) "Finding energy-efficient and fast detour routes in unusual traffic events," In Proceed-

- ings of IEEE International Conference on Intelligent Transportation Systems (ITSC), pp. 2618-2625.
- Ian Goodfellow, Yoshua Bengio, and Aaron Courville (2016) "Deep learning," The MIT Press. (邦訳) 岩澤有祐・鈴木雅大・中山浩太郎・松尾豊監修, 味曾野雅史・黒滝純生・保住純・野中尚輝・河野慎・富山翔司・角田貴大訳 (2018)『深層学習』KADOKAWA.
- Google Inc. (Inventors: Geoffrey E. Hinton, Alexander Krizhevsky, Ilya Sutskever, and Nitish Srivastva) (2016) "System and method for addressing overfitting in a neural network," US9406017B2.
- Google LLC (Inventors: Sergey Ioffe and Corinna Cortes) (2019a) "Batch normalization layers," US10417562B2.
- Google LLC (Inventors: Noam M. Shazeer, Aidan, Nicholas Gomez, Lukasz Mieczyslaw, Kaiser, Jakob D. Uszkoreit, Llion Owen Jones, Niki J. Parmar, Illia Polosukhin, and Ashish Teku Vaswani) (2019b) "Attention-based sequence transduction neural networks," US10452978B2, CA3144674C.
- Sergey Ioffe and Christian Szegedy (2015) "Batch normalization: accelerating deep network training by reducing internal covariate shift," In Proceedings of the 32nd International Conference on Machine Learning, pp. 428-456.
- Konečný Jakub, McMahan Brendan, and Ramage, Daniel (2015) "Federated optimization: distributed optimization beyond the datacenter," arXiv preprint arXiv: 1511.03575.
- Renhe Jiang, Zhaonan Wang, Jiawei Yong, Puneet Jeph, Qianjun Chen, Yasumasa Kobayashi, Xuan Song, Shintaro Fukushima, and Toyotaro Suzumura (2023) "Spatio-temporal meta-graph learning for traffic forecasting," In Proceedings of Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI), pp. 8078-8086.
- Microsoft Team Data Science Process (TDSP), 2016. <https://docs.microsoft.com/ja-jp/azure/architecture/data-science-process/overview>
- Srivastava Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov (2014) "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning*, Vol. 15, No. 56, pp. 1929-1958.
- Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich (2015) In Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-9.
- Toyota Motor Corporation (Inventors: Shintaro Fukushima and Takeyuki Sasai) (2022) "Retrieval of video and vehicle behavior for a driving scene described in search text," US11403343.
- Toyota Motor Corporation (Inventors: Shintaro Fukushima and Takeyuki Sasai) (2023a) "Machine learning model for image recognition used in autonomous vehicles," US11810338B2.
- Toyota Motor Corporation (Inventors: Shintaro Fukushima, Zenna Nagata, Akie Sakiyama, Kaoru Yamada, Sayaka Yoshizu, and Takeyuki Sasai) (2023b) "Danger prediction device, danger prediction system, method of danger prediction, and storage medium storing program," US2020009505A1.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin (2017) "Attention is all you need" In Proceedings of 31st Conference on Neural Information Processing System (NIPS).
- Weiming Xiang, Hoang-Dung Tran, and Taylor T. Johnson (2017) "Reachable set computation and safety verification for neural networks with ReLU activations," arXiv preprint arXiv: 1712.08163.
- Xiaohang Xu, Toyotaro Suzumura, Jiawei Yong, Masatoshi Hainai, Chuang Yang, Hiroki Kanezashi, Renhe Jiang, and Shintaro Fukushima (2023) "Revisiting mobility modeling with graph: a graph transformer model for next point-of-interest recommendation," In Proceedings of 31st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL), Vol. 94, pp. 1-10.
- Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong (2019) "Federated machine learning: concept and applications," *ACM Transactions on Intelligent Systems and Technology*, Vol. 10, No. 2, pp. 1-19.
- Jiawei Yong, Yuichi Wakabayashi, Akihiro Okayasu, Reiji Miki, Takeyuki Sasai, Masaaki Inoue, and Shintaro Fukushima (2022) "Estimating total traffic volume with statistical modeling approach," In Proceedings of IEEE International Conference on Intelligent Transportation Systems (ITSC), pp. 304-309.
- 浅井俊雄 (2019)『AI分野における NEC の知的財産マネジメント』「パテント」Vol. 72, No. 8, pp. 53-64.
- 石川冬樹・丸山宏編, 柿沼太一・竹内広宜・土橋昌・中川裕志・原聡・堀内新吾・齋崎弘宣著 (2022)『機械学習工学』講談社.
- 一般社団法人データサイエンティスト協会・独立行政法人情報処理推進機構 (2022)『データサイエンティストのためのスキルチェックリスト/タスクリスト概説 第2版』.
- NTTグループ・トヨタ自動車株式会社 (2021)『コネクティッドカー向け ICT 基盤に関する技術資料』https://www.toyota-tokyo.tech/news/pdf/technical_document_ictpf.jp.pdf
- QA4AI コンソーシアム (2020)『AI プロダクト品質保証ガイドライン』.
- 経済産業省 (2018)『AI・データの利用に関する契約ガイドライン』.
- 産業技術総合研究所 (2020)『機械学習品質マネジメントガイドライン』<https://www.digiarc.aist.go.jp/publication/aiqm/guideline-rev3.html>
- 佐藤直人・小川秀人・来間啓伸・明神智之 (2021)『AI ソフトウェアのテスト』リックテレコム.
- 澁井雄介 (2021)『AI エンジニアのための機械学習システムデザインパターン』翔泳社.
- 特許庁 (2015)『特許・実用新案審査ハンドブック』https://www.jpo.go.jp/system/laws/rule/guideline/patent/handbook_shinsa/index.html
- 特許庁 (2018)『AI 関連技術に関する事例の追加について』.
- 特許庁 (2023)『AI 関連発明の出願状況調査』https://www.jpo.go.jp/system/patent/gaiyo/sesaku/ai/ai_shutsugan_chosa.html
- トヨタ自動車株式会社 (発明者: 福島真太郎, 笹井健行) (2022)『検索装置, 学習装置, 検索システム, 検索プログラム, 及び学習プログラム』特許 7151654.
- トヨタ自動車株式会社 (発明者: 福島真太郎, 笹井健行) (2023a)『情報処理システム』特許 7279685.
- トヨタ自動車株式会社 (発明者: 福島真太郎, 永田善也, 崎山亮恵, 山田薫, 吉津沙耶香, 笹井健行) (2023b)『危険予測装置及び危険予測システム』特許 7294259
- 中島震 (2020)『ソフトウェア工学から学ぶ機械学習の品質問題』丸善出版.
- 中島震 (2022)『AI リスク・マネジメント』丸善出版.
- 福岡真之介・濱野敏彦 (2019)『AI の知財戦略と知財保護における課題』「パテント」Vol. 72, No. 8, pp. 26-40.
- 福島真太郎 (2015)『データ分析プロセス』共立出版.
- 松下外 (2019)『AI 技術関連発明の特許出願及び権利行使』「パテント」Vol. 72, No. 8, pp. 41-52.
- 丸山宏 (2019)『機械学習工学の狙いと展開』「情報処理」Vol. 60, No. 1, pp. 2-16.