



RIETI Policy Discussion Paper Series 17-P-012

AI技術の社会実装への取り組みと課題～ 産総研AIプロジェクトから学ぶ

小西 葉子
経済産業研究所

本村 陽一
産業技術総合研究所



Research Institute of Economy, Trade & Industry, IAA

独立行政法人経済産業研究所
<http://www.rieti.go.jp/jp/>

AI技術の社会実装への取り組みと課題～産総研 AI プロジェクトから学ぶ*

小西葉子（経済産業研究所）

本村陽一（産業技術総合研究所）

要 旨

現在、多くの企業や大学が AI 技術、IoT、センシング技術などの研究開発に取り組んでいる。サービスや製造の現場では、中小企業から大企業まで広くビッグデータを収集し、AI 技術を活用する意欲が高まっている。経済産業省でも、これらの技術を生産性向上に活かすべく、産業育成、中小企業支援、地方創生の視点で多岐に渡るプロジェクトが進行中である。しかし、どのような環境でどのような工夫をすればプロジェクトが成功し、継続していくかについて知見を得ることは難しい。本稿では、産業技術総合研究所が取り組んできた 28 件の AI プロジェクトについて、個々のプロジェクトの目的から成果までを独自に設定した分類を用いてレビューを行う。これにより、実際に企業や地方自治体に技術を実装する際に留意すべき点を明確化していく。

キーワード：AI 技術、ビッグデータ、人との相互理解、ベイジアンネットワーク、PLSA

JEL classification: C11, C14, C45, L81, L82, L84

RIETI ポリシー・ディスカッション・ペーパーは、RIETI の研究に関連して作成され、政策をめぐる議論にタイムリーに貢献することを目的としています。論文に述べられている見解は執筆者個人の責任で発表するものであり、所属する組織及び（独）経済産業研究所としての見解を示すものではありません。

*本稿は、独立行政法人経済産業研究所におけるプロジェクト「経済変動の需要要因と供給要因への分解：サービス産業を中心に」の成果の一部である。本稿を作成するに当たっては、深尾京司教授（一橋大学、RIETI-PD）、矢野誠所長（RIETI）、森川正之副所長（RIETI）並びに経済産業研究所リサーチ・セミナー参加者の方々から多くの有益なコメントを頂いた。また、執筆の過程において内藤真理子氏（RIETI 研究 G）、猪俣智子氏（RIETI 研究 G 副 D）、五十里寛氏（RIETI 研究 C）に多くのサポートとアドバイスを頂いた。記して感謝したい。

1. はじめに

近年、人工知能（Artificial Intelligence、以下 AI）という言葉が我々の日常生活に浸透している。また現在、多くの企業や大学が AI 技術、IoT、センシング技術などの研究開発に取り組んでいる。AI についての議論を行う際には、AI がまず何を指すのかを明確化しておくことが望ましい。本稿では AI 技術を議論する際には、ヒト型ロボットのような「汎用型の AI」ではなく、ある特定の技術についてヒトが行っている作業を自動化し、ヒトと同等もしくはヒト以上の作業ができる「特定型の AI」を対象とする。この AI 技術にはいくつかの分野があるが、議論される主なものが近年の AI 技術の主流となっている、ディープラーニングをはじめとする機械学習である。これらの技術の主な目的はコンピュータに何をさせるかをヒトが決め、目的変数と説明変数を含むデータを与えてその中から入力データの特徴やパターン、分類するクラスターを学習させ、新たにデータが現れたときに、どのクラスターに分類されるかを判断したり、出力を予測したりする。つまり近年の AI 技術は探索、判別、分類、予測を得意とする機械学習にもとづいている。この機械学習を使って多種、大量のデータ、いわゆるビッグデータから有用な知識を自動的に、あるいは半自動的に獲得し、AI 技術を応用することも期待されている。そこで獲得された知識は可視化されるだけでなく、計算モデルとして予測や推論、制御に活用できるものになることで、人が行うよりも高速、高精度で意思決定ができる自律的な知能システムが実装することを目指す。そのため、サービスや製造の現場では、中小企業から大企業まで広くビッグデータを収集し、AI 技術を活用する意欲が高まっている。経済産業省でも、これらの技術を生産性向上に活かすべく、産業育成、中小企業支援、地方創生の視点で多岐に渡るプロジェクトが進行中である。

しかし、どのような環境でどのような工夫をすればプロジェクトが成功し、継続していくかについて知見を得ることは難しい。そこで本稿では、2004 年から 2017 年 2 月の現在に至るまで、産業技術総合研究所が取り組んできた 28 件の AI プロジェクトについて、個々のプロジェクトの目的から成果までを独自に設定した分類を用いてレビューを行う。これらは、例えば医療・介護現場の業務改善、娯楽業の顧客へのサービスの推薦（レコメンド）、飲料業の商品物流の最適化など、サービス提供の現場を対象としている実サービス志向のプロジェクトである。本稿では、このレビューを通じて、実際に企業や地方自治体に技術を実装する際に留意すべき点、AI プロジェクトが通常のプロジェクトに加えてどのような特徴を持つかを明確化することを目的とする。

本稿での分析を通じて、どのような AI プロジェクトが持続的に発展するのだろうか？、というシンプルな問いに対して以下の答えが見えた。1. プロジェクトの目的や技術のユースケースが明確で、目標となる指標を観測可能なデータに落とし込めること、2. 現場のニーズとモチベーションが強いこと、3. データがセンサ、ネットワークやインターネット上を通じて低コストで集積できること、4. 業務プロセスの一部に組み込まれることで

データ収集が持続できること、5. データを分析して得られた知見や計算結果がさらにデータとして追加できること、などである。このように、AI プロジェクトにおいてはデータの担う役割が非常に高いことがわかる。一方、AI 技術と呼ばれる手法やアルゴリズムは研究現場において開発が進められているが、公表されればすぐに実装され普及することからコモディティ化の進行が速く、既存のものを各現場に合わせて改良し利用することも現在のプログラミング技術やコンピュータ環境においては容易である。

一方で、機械学習に基づく AI では、学習させるデータによって性能が左右される。そして、他にはない最新の質の高い学習データを大量に利用可能であることが競争力の源泉となる。したがって AI プロジェクトの成功の可否は、ビッグデータの質と収集コストを低くすること、そして何より継続できるかどうかにかかっている。質の良さは消費者現場でも企業側の現場でも、どれだけデータを残す、収集する技術が定着しているかに依存する。これらの得られた知見が、プロジェクトのレビューの中で何に依拠するのかを次章以降で観察していく。

以降では、2章で AI 技術と AI の周辺について議論する。3章では本稿が対象とするプロジェクトで採用されている、確率的潜在意味解析 (PLSA) とベイジアンネットワーク (BN) について他の手法との違いを考慮して説明する。4章では 28 件のプロジェクトについて体制、ビッグデータの役割、AI 技術の応用、プロジェクト評価に焦点を当てて議論する。5章は結論として、どのようなプロジェクトが成功しやすいかについて総括し、今後の課題を踏まえて政策含意を述べる。

2. 近年の AI ブームとその周辺

まず本稿では人工知能技術は、ヒト型ロボットのような「汎用型の AI」ではなく、ある特定の技術についてヒトが行っている作業を自動化し、ヒトと同等もしくはヒト以上の作業ができる「特定型の AI」を対象とする。この AI 技術にはいくつかの分野があるが、ここで議論される主なものが機械学習である。機械学習にもディープラーニングや決定木学習、帰納論理プログラミング、クラスタリング、サポートベクターマシン、強化学習、ベイジアンネットワークなどの様々な手法がある。主な目的はコンピュータに何をさせるかをヒトが決める、目的変数と説明変数を含んだ正解データを与えてその中から特徴やパターンを探索させ、入力データがどのグループに入るか等の適切な分類の仕方などを学習させる。その後、新たにデータが現れたときに、どのクラスターに含まれるか判断したり、正解の値や確率を予測する。つまり探索、判別、分類、予測を得意とする手法である。ディープラーニングは階層型ニューラルネットワークの発展形である。ニューラルネットワークは、判断や認識のアルゴリズムを脳内の神経回路に模し、多量の非線形な入力と出力のニューロンを（神経回路）を繋いで表現する。その際、入出力変数の間に中間層と呼ばれる潜在

変数で表現される階層を組み込んでいる¹。ディープラーニングは、中間層が複数に渡るもので、潜在変数の層が増える分、より複雑なモデルとなる。ビッグデータの増加とコンピュータのスペック向上により研究が盛んになり、現在は画像認識やチェスゲームにおいて、AI 技術とヒトとの勝負などで知名度が高まっている。最終層である出力層の役割は古くからの階層型ニューラルネットと同じく、判別、分類、正解の予測である。それにさらに強化学習を組み合わせた探索が可能となることによって、ゲームに応用したり、前段の階層を深くできるアルゴリズムが進歩したことによって、データの何を見るか、何が大事かという問題設定（特徴量の抽出）も学習できる能力が向上した。これにより複雑な問題に対処できるようになった。一方で、モデルが複雑になるとともに学習結果を理解することが困難となり、その自律性とブラックボックス化した決定アルゴリズムが、AI の暴走を恐れる議論と結びつけられている。これに対して最近、AI の計算過程や、内部表現を人が理解しやすいものとし、人と相互理解できる人工知能という提案がある。

以降では、人工知能やニューラルネットの歴史と背景を踏まえ、人工知能のブーム、その冬の時代、そしてわが国の大型プロジェクトの関係を概観する²。図1はおおまかな年表となっている。これにより、4章で紹介するプロジェクトがどの時期に行われているのかが明らかになる。

<図1 挿入>

1980年から1987年の第二次人工知能ブームの当時、1982年に通商産業省が第五世代コンピュータ開発プロジェクトを立ち上げ11年間実施された。当時の人工知能技術の特徴は論理に基づくアルゴリズムで、実際のビジネスに広く応用されたものはエキスパートシステムである。この当時はデータから学習する機械学習とAI技術は結びついておらず、エキスパートシステムを構築するためには、大量の知識を人間がプログラムに埋め込む作業がボトルネックとなっていた。

第二次AIブーム後期には、Rumelhart *et al.* (1986)から入力層と出力層の間に中間層を持つニューラルネットに対して、計算結果と正解の間の誤差を逆方向に各ニューロンに対して修正していく誤差逆伝播法（バックプロパゲーション）が提案された。その後、各種の工学的応用が多数発表されたことで三層のニューラルネットワーク構造を持つニューラルネットの応用研究が盛んになり、第二次ニューロブームが起こった。

しかし当時はまだコンピュータの性能不足、ビッグデータと呼べるほどの大量データがな

¹ 照井(2017)より、「経済学での効用という潜在変数を介して人間の行動を理解しようとする離散選択モデルも広い意味では同様」、つまり潜在変数を介した非線形モデルと解釈できる。

² 年代については総務省(2016)図表4-2-1-5「人工知能(AI)の歴史」を参考としている。

かったことから、過学習 (over fitting)³が起きるなど十分な精度が得られず 1990 年代前半には冬の時代を迎えることとなった。

第二次ニューロブームの時期、通産省は 1992 年から 11 年間、RWC (リアルワールドコンピューティング：実世界情報処理) プロジェクトを行い、パターンとシンボルを融合した柔らかなコンピューティング (情報処理) の開発を目指した⁴。このプロジェクトでは、並列計算機の開発だけでなく、自律移動対話ロボットのソフトウェアや、マルチモーダル対話アプリケーション、機械学習アルゴリズムの理論研究、確率モデリングプログラムの開発など、現在の人工知能技術の先駆けとなるソフトウェア研究も行われていた。

しかし、RWC プロジェクト終了後は、わが国では官主導の大きな人工知能研究プロジェクトが行われることはなく、市場経済においてインターネットの発展や、サービス現場における ID-POS 端末などの IT 技術の普及にともなうビッグデータの集積が進んでいった。

図 1 に見られるように、AI 技術、ニューラルネットワークは、1990 年代から 2000 年代前半、国主導の AI 技術大型プロジェクトは 2000 年代前半から 2015 年まで大きなブームの無い時期になっている⁵。その中で、2006 年に Hinton *et al.* (2006)により、ニューラルネットワークの中間層が多層になることを許した学習アルゴリズムが報告され、大量のデータと発達した計算機による学習が可能となり、多層ニューラルネットワーク (ディープラーニング) 研究が活発になっていった。2012 年にはディープラーニング技術が画像認識コンテストで優勝したことを契機に、世界中で注目度が高まった。このディープラーニングのブームと相まって、わが国でも 2013 年以降、第三次人工知能ブームが起こっている。その際に、これらのブームが 2012 年のビッグデータブームと重なっていることも、このブームの勢いを加速させている。さらに、2015 年 5 月に日本最大の AI 研究開発拠点として、国立研究開発法人産業総合研究所に「人工知能研究センター」が設立された。同年 9 月には国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) において次世代ロボット中核技術開発 (次世代人工知能技術分野) が開始され、人工知能研究センター辻井センター長がプロジェクトリーダーに就任し、総勢 100 名を超える研究者が集結している。この大型プロジェクトにおいて、2020 年 3 月まで次世代人工知能研究の研究開発が進められ、再び産官学一体の研究プロジェクトが始動している。そして現在は、機械学習に基づく AI 技術、ビッグデータ、大型プロジェクト、産業応用ニーズと全てが揃っているかつてないタイミングとなっており、社会全体に影響を与える大規模のブームとなっている。

上述の様に、この数年脚光を浴びている現在の人工知能ブームを支えているのは機械学習

³ 過剰適合問題は、線形回帰モデルにおいてもモデルの説明力を上げるために、データサイズに対してパラメータ数を多くすることによって起こりうる。ここではデータが少なすぎるために、学習したデータ以外では適合度が低いことや、モデルを複雑にしたために起こる過剰適合を指している。

⁴ 大型プロジェクトの目的や期間などは大津(2016)を参考としている。

⁵ その他にも、経済産業省主導の「情報大航海プロジェクト (2007-2009)」など短期間のプロジェクトは存在した。

の中の各種手法とビッグデータによる自動的な計算モデル構築と、その計算モデルの活用技術である。本稿が対象としたプロジェクトの中では、確率的潜在意味解析とベイジアンネットワークという計算モデルを使ったものが多い。次章では、この2つの手法を紹介していく。

3. 確率的潜在意味解析 (PLSA) とベイジアンネットワーク (BN) について

4章で取り上げたプロジェクトの多くでは、サービス現場やサービスを分析対象とし、現場で生成されるビッグデータと、そのデータから不確実性やばらつきを含む計算モデルを自動的に構築する確率モデリング技術を活用することで、リスク軽減、新たなサービスの創出、利益や生産性向上を目指す事例となっている。分析には、ベイジアンネットワーク（以降、BN）と確率潜在意味解析（Probabilistic Latent Semantic Analysis:以降 PLSA）がAI技術として導入されている。なぜこれらの手法が主に使用されているかを理解するために、前章での説明に加えてディープラーニングが得意とする分析とその特徴について簡単に説明する。繰り返しになるが、ディープラーニングは大量の非線形入力と出力層の間に位置する中間層が2層以上の多層化した非線形関数モデルとその学習アルゴリズムである。計算方法としては、誤差逆伝播法を用いて、正解と計算結果のずれを修正していく。以前はデータが十分になかったため良い精度での分析ができなかったが、多くのユーザの発話の収集、SNSを通じた画像のアップロードやテキスト情報の増加により正解ラベル付きの各種データも増え、音声認識、画像認識の精度が飛躍的に高まり、主にこれらの分野での応用が進んでいる。画像以外にも文章中の単語ベクトルのような高次元ベクトルに対して、例えば成功や失敗という離散の正解ラベルがあるようなビッグデータの場合には学習可能である⁶。また、正解ラベルがない場合でも、囲碁やチェスの勝敗のように報酬を明示的に与えることができる場合には、強化学習と組み合わせることで、プロのプレイヤーに勝利できたことで世界的に注目を集めている。

ディープラーニングはこの様に、非線形性が高い問題に対して有効で、非常に柔軟で応用可能性が高い手法である。一方、世の中には正解や報酬が明示的には与えられなかったり、決定論的な定式化が適切でない問題、つまり非線形性だけでなく非決定性が高い問題もある。例えば、消費者の異質性が存在する場合には、最適な商品の組み合わせについて正解が一つとは限らない。また消費者の選好は時間方向に変化し、確率構造を持つことも多い。その場合、選好（行動）は確率分布として表現され、正解も決定的な関数としては与えられないため、条件付確率分布モデルを想定することが自然である。また、予測精度が高い

⁶ 山本・松尾(2016)は、内閣府から毎月公表されている『景気ウォッチャー』の景気に敏感な人たちの景気判断（5段階）とその理由について書かれたテキストデータを使って学習させ、他の調査や資料の文書からも景気判断ができるマシンを開発した。

というだけでは、不十分なことも多い。例えば、来期売上を高精度で予測できるような自己回帰モデルを作成し高精度の予測ができたとしても、そのモデルが構造を含まなければ、何をすれば売上が上がるのかを知ることはできない⁷。また、リスクとなる現象を避けたい場合には、そのリスクの生起を予測するだけでは不十分で、その現象が起こった理由を知り、リスク回避するにはどうすればよいか、予防対策ができることが重要である。そのためには、現象が生起するばらつき、不確実性を条件付確率として構造的にモデル化することや、ばらつく理由となる分散の大きさに対する影響や相互作用も構造的に理解できるモデル化が重要になる。

本稿のプロジェクトではサービス提供の現場を対象としているサービス志向のプロジェクトがほとんどであり、業務の実記録やサービスの利用履歴、購買履歴のデータや IoT デバイスにより記録されたデータを用いて分析を行う。そのため、消費者や行動が異質性を持ち、人の行動を通じた不確実性を考慮し、起こっている現象を様々な状況や特性によって条件付けて説明する確率モデルを採用することが望ましい。

BN は、この様な不確実性や状況依存性を近似する非決定的な確率モデルであり、条件付確率モデルとして表現でき、確率分布の分布形も仮定する必要がないノンパラメトリックモデルである。今、条件付き確率を $P(Y|X)$ と表し、 X, Y を離散化してテーブル上に条件確率値を表現する。 $P(Y|X)$ は、目的変数 Y を説明変数 X によって説明する確率モデルであるが、この説明変数を複数与えることで、人による違い、時間や空間による依存性などを交互作用も含めて表現できる。例えば、化粧品の購買確率を考えると、まず条件ごとに異なる購買確率を $P(\text{購買}|\text{条件})$ で与えるとする。性別によって購入頻度が異なり、女性の方が化粧品を買う確率が高いならば、 $P(\text{化粧品購買}|\text{女性}) > P(\text{化粧品購買}|\text{男性})$ と表せる。また、平日と週末で異なる場合、 $P(\text{化粧品購買}|\text{女性、週末}) > P(\text{化粧品購買}|\text{女性、平日})$ と表せる。BN では目的変数と説明変数群の確率的関係を表現する計算モデルをグラフ構造によって表すので、「グラフィカルモデル」とも呼ばれる。このとき、目的変数である購買確率が最も高くなる条件の組み合わせをデータから探索する方法が BN で、条件の設定を工夫したり、推定結果の解釈することで説明変数を変化させたときの目的変数の予測の精度が高めることができる。上記の例の場合は便宜上、目的変数を設定して説明した。この様に、明らかに関係のないモデルを人が外して計算負荷を減らすこともできるし、全変数の観測頻度から各変数間の因果関係を探索することもできるのが BN である。さらに、上記の目的変数に対する予測だけでなく、目的（結果）が与えられたときの説明変数（原因）方向への逆推論も可能で、例えば子どもが怪我の種類から、怪我の種類に関わる室内での行動（原因）を探ることもできる。これにより、結果がある値に変化したときに、その原因となる変数群を探索し、目的変数が望ましい状態になるよう原因変数を制御する課題に対して適用さ

⁷ 計量経済学では、このような現象の構造や因果関係の識別に関心を持って研究を行っているため、これまででは AI 技術を含めた探索的に分析する統計手法を十分に応用してこなかった。

れ、「制御モデル」と呼ばれる。

確率的グラフィカルモデルの構造は、複数の変数の間の依存（因果）関係の有無をビッグデータの観測頻度から定量的に判定し、依存関係が強い変数の間にリンクを張る手続きを繰り返すことで自動的に学習することができる。BNの学習の場合、パラメータの学習はクロス集計表の正規化で求められるので、この構造学習を実行することが容易である。一方、ディープラーニングなどの階層型ニューラルネットは計算量の多くがパラメータ学習に費やされるため、構造自体は固定されて学習しないことが多い。以上より、BNでの学習の方が、より構造的な分析が可能であると言える。

さらに、BNにより、目的変数をインターネットユーザーやサービス利用者の行動として、その行動の生起確率をグラフィカルモデルとして構築できると、目的変数の予測だけでなく、前述の様に逆推論によってユーザや利用者の行動変化を制御できることになる。こうした目的に使われる計算モデルは「ユーザモデル」と呼ばれる。

個人にひもづけたIDと、購買履歴やWebの閲覧履歴などのデータを集めたビッグデータには、ユーザの特性が大量に含まれている。こうしたビッグデータからユーザモデルを構築できると、実際の消費者や生活者の行動を予測できるユーザモデルが活用できるため、マーケティングや健康問題の解決などに有用である。BNは構造学習や確率推論アルゴリズムが整備されており、ユーザモデルの構築に適した手法であると言える⁸。

しかし、数百万人といったビッグデータから直接BNを構築することは計算量、メモリ容量の点から難しい。こうした場合に、PLSAという潜在クラス分析手法を用いて、データに対してある種の次元圧縮を行い、数十程度のセグメントに全ユーザを分類してから、そのセグメントごとのユーザモデルをベイジアンネットとしてデータから構築する方法が多方面で使われ始めている。例えば表1のNo.9の「コープK」プロジェクトでは、会員数は100万人、1年間でデータ数は、669,511,467（約7億）トランザクション数となる。この中で分析に使用したのは、3981名、売上げ上位1000アイテムの420万トランザクションで、これに対してPLSAを行い、「顧客-商品」クラスタを作成した。図2はPLSAの分類イメージである。左の行列は、行は顧客で各列の商品を何個購入したかが記入してある。PLSAを実施すると行①、②、③で似た物を買う消費者が分類され、列①、②、③でよく一緒に買われる商品の分類を表す。例えば、クラスター①は消費者ID-1、3、6が分類されパン、お菓子、お弁当といった既製品の食品をよく買うグループとなっている。このクラスターでは消費者ID-6のお菓子の購入頻度が「0」、つまりまだ購入した経験がない。この消費者セグメントでは、本来お菓子の購入確率が高いため、同じ購入経験「0」でも他の消費者セグメントが買う確率の高い大根より、この消費者が属するセグメントが購入する可能性が高いお菓子を薦めた方が購入する可能性が高いと判断される。同様に、商品からみた場合

⁸ BNについての詳しい説明は繁梲・本村・植野（2006）、本村・岩崎（2006）、生活行動予測への応用については本村（2009）を参照のこと。

にもこの確率的な推論によって、商品ごとに誰が次に購入する確率が高いかを知ることができる。

こうして PLSA によって潜在クラスが抽出できるが、その潜在クラスはどういったクラスタなのか説明を求められることが多い。また、先の確率推論だけでは、購買頻度が高い消費者や商品でしか確率推論を行うことができない。これらの問題を解決するために、各潜在クラスタを目的変数として、それを予測するベイジアンネットを新たに構築する方法、「確率的潜在構造モデリング」が提案された。プロジェクトで PLSA と BN による「確率的潜在構造モデリング」を行っているものは、PLSA でビッグデータの次元を落として、ある程度カテゴリを作った後に、BN による確率モデルを構築し、それをユーザモデリングや消費者の行動予測モデルとして活用する。ユーザモデリングは、アマゾンなどで購入や閲覧した製品同士の類似度の近さで測る協調フィルタリングと異なり、ユーザの履歴や好みを探査し、確率モデルを構築するため、より個々人の嗜好を反映したレコメンドやそれに基づく経営予測を行うことができる。

また、この手法は顧客一人一人でなく、購買行動から顧客の好みやライフスタイルをしつかり反映しながら、個人の秘匿性が守られる小さなセグメントを作り、セグメント ID を紐づけることが可能なため、個人情報を利用したくないまたはできない企業や顧客にも安心して使用できる。

図 2 PLSA の分類イメージ

ID/商品	パン	お菓子	歯磨き粉	洗剤	りんご	大根	お弁当
ID_1	1	3	0	0	0	0	2
ID_2	0	0	0	0	3	2	0
ID_3	2	1	0	0	1	0	1
ID_4	0	0	2	1	1	0	0
ID_5	0	1	1	2	0	1	0
ID_6	1	0	1	0	1	0	1
ID_7	0	0	0	0	2	0	1

→ PLSA →

ID/商品	クラスター①			②		③	
	パン	お菓子	お弁当	りんご	大根	歯磨き粉	洗剤
ID_1	1	3	2	0	0	0	0
ID_3	2	1	1	1	0	0	0
ID_6	1	0	1	1	0	1	0
ID_2	0	0	0	3	2	0	0
ID_7	0	0	1	2	0	0	0
ID_4	0	0	0	1	0	2	1
ID_5	0	1	0	0	1	1	2

4. AI 技術の実装プロジェクト

表 1 は 2004 年から現在 (2017 年 2 月) までの産業技術総合研究所が取り組んだ AI 技術等の社会実装プロジェクトの一部である。28 件の AI プロジェクトは開始時期順に並べられ、二巡目、三巡目と派生したプロジェクトは関連がわかるように記述している。著者の一人の本村がデジタルヒューマン研究センター (2003-2008)、サービス工学センター (2008-2015)、AI 技術研究センター (2015 年 4 月～) で参加したものが中心である。本稿で対象とする 2004 年から 2017 年のわが国の AI 技術分野は図 1 で示したように、1980 年代から始まった通産省主導の第五世代コンピュータ (1982-1992)、リアルワールドコンピューティング (RWC) プロジェクト (1992-2001) が終了し、2015 年 9 月の NEDO の次世代

ロボット中核技術開発プロジェクト立ち上げまでは大きな人工知能関連プロジェクトが無い研究開発支援という観点では不遇な時代と重なっている。この期間、民間主導でインターネットの発展や、サービス現場における ID-POS 端末などの IT/ICT 技術の普及にともなうビッグデータの集積が進んだ。2011 年後半から本格化したビッグデータブーム、2014 年頃からの最近のディープラーニングを伴う第三次 AI ブームから現在に至るまでの期間を含んでいる。2015 年は前述の国主導の大型プロジェクトも再開され、AI 技術の開発、応用ともに活発になってきている。

そこで、表 1 では、技術進歩やその時々の流れがわかるようにと、実施開始期間の早い順に並べている。

<表 1 挿入>

プロジェクトで分析に使用された AI 技術の列を見ると、2009 年からベイジアンネットワーク（以降、BN）に加えて、確率潜在意味解析（Probabilistic Latent Semantic Analysis:以降 PLSA）が AI 技術として導入されている。プロジェクトの大半は、サービス現場やサービスを分析対象とし、現場で生成されるビッグデータと確率モデリング技術を活用することで、リスク軽減、新たなサービスの創出、利益や生産性向上を目指す事例となっている（サービス工学関連プロジェクトについては産業技術総合研究所(2014)を参照）。

実際には個々のプロジェクトを一つの研究成果として論文報告したとしても、工学系論文では適用した技術やその効果の評価⁹が主たる関心であるため、現場で得られた種々の知見をまとめて記述することは困難である。その点を補う目的で、本稿では 28 プロジェクトを総括し、各プロジェクトの開始から分析、評価の過程のなるべく多くの情報や特徴を汲み上げられる様、各プロジェクトの共通項を探索することに努めた。プロジェクト総括の表の列部分に当たる項目はかなり注力して設定した。その過程で、いくつかの知見も得ることができた。まず、プロジェクトの目的はデータの種類を決めるが、すぐに利用可能なデータの存在がプロジェクトを決定することもあること、次に業務系のデータは簡単には他組織に移管することが難しいため、プロジェクトの現場はデータへのアクセスが容易なデータベースが在る組織になること、データの秘匿の有無がプロジェクト形態や技術介入期間に影響を与えること、新規データ収集の困難さがプロジェクトの継続に影響を与えることなどがある。さらに当然だが、こうしたデータに関わる背景が、適用すべき AI 技術の選択にも大きく影響する。そこで、データ収集現場と分析現場を分け、データの種類、収集技術と形態、収集コストなどデータに関連するプロセスを多く項目として立てた。

表 2 は表 1 の 20 項目を 6 つの評価軸で分類したものである。各プロジェクトを(1)プロジェクトの主体となる相手先について、(2)プロジェクトの概要、(3)技術協力者である産業技

⁹研究開発プロジェクトの技術についての評価は参考文献[2]、[3]、[10]、[11]を参照。

術総合研究所について、(4)プロジェクトで使用するデータについて、(5)分析について、(6)プロジェクトの総括の視点で記述した。

表2 プロジェクトのレビューの分類視点

分類	分類項目
(1) 相手先について	企業（組織）名、業種、予算規模
(2) プロジェクト概要	時期、実装期間、目的、プロジェクトの主たる現場、現場での実施内容
(3) 産総研について	現場への技術介入期間、産総研の役割
(4) データについて	データの種類、持出しの可否、収集法、収集コストの負荷の程度、コストの内容
(5) 分析について	分析場所、AI技術・統計手法、AI技術を使って行ったこと
(6) プロジェクト総括	評価、困難だった点

4.1 プロジェクトの体制について

ここでは、プロジェクトの体制として、主体となる組織、技術協力者となる産総研、プロジェクトの概要について、表2の分類 (1)、(2)、(3) について説明する。

(1) 相手先について

順を追って、各項目の説明をすると、まずプロジェクトを行った**企業や大学名、業種、予算規模**の項目である。対象となる業種は小売、飲食、医療、介護、保険、娯楽、情報通信、教育などほぼ全てサービス業またはサービスの提供であった（例外はプロジェクト No.25 と 28）。予算規模は最大のもので数千万円、小さいものだと期間中で 100 万円のものがある。また、産総研の人材、データ、分析ツール、道具など既存リソースを使ったため無料のものもあった。これは産総研が公的研究機関であり、①利益を追求する必要がないこと、②研究者、機材、技術、前例となる資料などを有しており、それらを共同研究の名目で差し出せること、③産総研側にプロジェクト完了までの期限がないことなどに起因する。もしも、企業だけでプロジェクトを行う場合はより早く、より確実に利益が見いだせなければ簡単にプロジェクトは打ち切られるであろうし、国の事業であれば複数の委託・再委託先に事業を依頼することになるであろうから、予算規模はかなり大きくなることが予想される。

(2) プロジェクト概要

ここでは、**プロジェクト開始時期、期間、プロジェクトの目的、主たる現場と実施したこと**について記述している。傾向としては 2009 年以前に始まったプロジェクトは終了しているものが多く、それ以降は現在に渡るまで継続しており、二巡目、三巡目と続いているものや水平展開しているものもある。既に終了しているプロジェクトの実施期間は 2 年以下のものが多い。実際にプロジェクトの現場はどこで、そこで何をしたのかという視点で概観

すると、**主たる現場**は行動や嗜好を調べる対象である「ひと（消費者、患者など）」がいる場所で、**実施したことは**、28プロジェクトのうち20プロジェクトが現場でデータ収集を行っていた。そこで、表1では、「主たる現場」と「分析の現場」を分けることにした。実サービスにおけるビッグデータを対象としたサービス志向のプロジェクトにおいては、主たる現場は産総研になることはなく、相手先の企業や店舗、病院や施設、イベント会場、ネットワーク上など業務データが収集できる場所となっている。それ以外では分析、サービスやアルゴリズムの開発を行った場所、収集・購入したデータを統合する場所が主な現場となっている。約7割のプロジェクトが主な現場でデータ収集を行っていることより、データ収集がAIプロジェクトのプロセスの中で非常に重要であることが明らかになった。**プロジェクトの目的**は多岐に渡るが、表3でA~Fの6つに分類した。実際にはプロジェクトの目的は複数あることが多いが、代表的なものを抽出した。プロジェクトの目的で一番多かったのは、D. 相手先企業の支援で、次いでC. 消費者の行動や意思決定のための支援であった。一部紹介すると、A. 開発では、例えばプロジェクトNo. 1、2はカーナビ制作企業が、自身のカーナビシステム上でレストランを推薦するためのiPhoneアプリ開発を目的としている（詳細は本村・岩崎(2006)を参照のこと）。B. 分析のNo. 21はあるショッピングモールでの顧客の買い回り行動の分析を行った。C. 消費者支援の中の商品推薦(小野他(2007))では、映画会社が顧客の好みに合う映画を推薦し（No. 3）、No. 6のプロジェクトではこどもの事故を予防するためのリスク予測を行っている。また、健康・生活支援のNo. 26の医療機器メーカーは、血圧のビッグデータを用いた健康支援のための分析を行っている。D. 相手先企業支援の、利益・生産性向上に関するプロジェクトでは、小売業の企業を対象に（No. 9, 10, 21, 22）、ID-POS とその他のマーケティング情報や顧客情報を統合して、消費者の購買行動からセグメント分けして需要予測や施策の最適化を行い利益向上を目指している。接客・商品提案は、小売業の店舗においてタブレットを用いた顧客への商品紹介と、キャンペーンやプロモーションの企画支援を行った（No. 7, 8）。プロジェクトのNo.5は顧客の属性と映画の嗜好からDVDパッケージの映画の宣伝フレーズとして最適なものを選択するプロモーションの最適化を行った(落合他(2009))。レストラン(No. 11)では、購買履歴と従業員の行動の分析結果から、従業員のシフトや店内での立ち位置の最適化を行った。救急現場では、大学病院の病棟や救命センターの業務データの分析結果から業務の最適化支援を行い(No. 14)、No. 23は自動販売機の設置、商品配置、補充の最適化支援のための業務データの分析支援を行っている。業務改善は例えば、No. 16は介護施設での申し送り業務にタブレットを導入し、電子化することで情報共有と入所者のリスク抽出を行い、介護の質の向上と業務効率化に貢献している。No. 27では、保険関連団体が保険請求の履歴や、加入時の問診データ、属性を用いて、顧客の健康リスクや健康寿命の予測手法の研究を行っている。E. 地域貢献は、医療、介護の分野で健康診断や健康イベントを通じて、住民の健康データを集め、サービス系のビッグデータやオープンデータとも組み合わせることで地域の医療費削減や予防医療に貢献する。F. データ構築のNo. 20は、広告会社や異業

種の企業が連携して、新たな産業プラットフォームとして、ID-POS データ、メディア視聴データ、ネットショッピングデータ、クレジットカード履歴、共通ポイントカード情報などのビッグデータを統合した、共創的データプラットフォームを構築することを目的としたプロジェクトである(産総研人工知能技術コンソーシアムデータプラットフォームWG活動報告 2017.3.17 において報告された)。

表3 プロジェクトの目的

プロジェクトの目的		プロジェクトNo.	業種
A. 開発		1, 2, 3, 25, 27	機械、情報通信、自動車、保険
B. 分析		2, 13, 19, 21	機械、医療、小売、広告
C. 消費者支援	商品推薦	3, 6, 8, 22	情報通信、教育、小売
	リスク予測	6	教育
	健康・生活支援	15, 17, 18, 26	地域保健、介護、 電力・エネルギー、医療機器
D. 相手先企業支援	利益・生産性向上	4, 9, 10, 21, 22	映画、小売、コンビニなど
	接客・商品提案	7, 8	小売
	最適化	5, 11, 14, 23	DVDパッケージ、レストラン、医療・救急、自動販売機
	業務改善	12, 16, 24, 25, 28	医療、介護、ディーラー、自動車、重工
	顧客の健康判断	27	保険
E. 地域貢献		14, 15, 17	医療、介護
F. データ構築		20, 22, 27	広告、情報通信、保険

(3) 産総研の役割について

ここでは、産総研の現場への技術介入の期間とプロジェクトにおける産総研の役割をみている。(2)で述べたように、主な現場での作業の多くはデータ収集のため、この技術介入はデータ収集のためにどのように現場にアクセスしているのかを表しており、現場への負荷を観察することができる。現場への技術介入の期間には、常時(計14プロジェクト)、期間限定で常時(計6プロジェクト)、一時的(計8プロジェクト)の3パターンが見られた。まず、インターネット上や産総研がアクセス可能なネットワーク上にデータが存在する場

合は「常時」データ収集を行っている。これは、データ収集のシステムを作るのはコストを要するが、運用コストが低く、現場への介入もないため負担が少ない。「期間限定で常時」は、一定期間現場に介入して実験を行ったり、業務データ履歴とるが、それらのデータがネットワーク上にある場合である。「一時的」の場合は、収集データのアップデートをしない、ネットワーク上にデータがなく郵送のアンケートを行う、データを購入する、現場で実験を行ってデータを収集する様なケースである。店舗など実際のサービス現場での実験は現場への負担が大きいため、長期間のデータ収集を行うが困難である。

産総研が当該プロジェクトの中で果たした役割には、①技術移転、②共同研究、③コンサルティング、④計画立案、⑤研究・分析などがある。最も多い形が技術移転と共同研究を同時にするスタイルである。技術移転となるか共同研究となるかは、表2の(4)データについての項目の「データの持ち出しの可否」と(5)分析についての項目の「分析場所」と密接な関わりがある。データを持ち出せる場合には、共同研究の形態をとり、分析場所は産総研内とすることができる。一方、持ち出し不可の場合には産総研内で分析を行うことができないため、相手先の組織内か相手先の委託先が分析場所となる。その場合に取られる手段として、産総研の既存または新規の技術を相手先に移転して利活用してもらう方法である。また、相手先に明確なアイデアや目的がない場合には、産総研がコンサルティングやプロジェクトマネジメントをしたり、計画全体を管理することもあり、技術協力に留まらず臨機応変な対応が求められている。

4.2 AI プロジェクトとビッグデータ

ここでは表2の(4)データについての**分析に用いたデータの種類、データ持ち出しの可否、データ収集法、データの収集コストの程度、収集コストの内容**について説明する。前述のように、プロジェクトの流れを整理する際に、データに関する項目が5つあり一番多くなった。データの種類は、以降の他の項目と密接に関わり、その中で詳細に議論するので、ここでは大まかな特徴について説明する。表1から、最も頻繁に使用されているデータは、収集法は郵送、インターネット、タブレット等を使った対話方式、規模も数千人以上の大規模なものから数十人程度のもので種々あるが28プロジェクト中の15プロジェクトがアンケートデータを用いていた。これは、4.3で後述するがAI技術による分析では、アンケートから得られる属性に加え、「ひと」の感情や嗜好に関する情報をデータに加えることで、データから学習して得られた予測モデルが、感情や嗜好を説明できる確率的な因果構造を反映したものになり、より実際の消費者心理や行動を反映した分析や理解が可能となる。これが「人と相互理解できるAI」と呼ばれる所以である。また、タブレットや携帯端末などのIoT行動系ビッグデータを収集しているプロジェクトは9件であった。また、Webの閲覧履歴、端末の操作履歴のデータなどがよく使用されている。一方、センサを使って収集したデータはNo.11の従業員の位置情報とNo.23の自動販売機の位置情報のデータ、

No.28 の工場内のエンジンなどを制御するためのデータの3プロジェクトであった。

4.1 の(3)の産総研の役割で述べたが、データの持ち出しの可否はプロジェクトの中での産総研の役割と(5)分析の場所と大きく関わる。28プロジェクト中、データの外部に持ち出すことができるものは5件、持ち出せないものが22件、外部持ち出しに向けて倫理委員会の審議を検討しているものが1件で、その大半が秘匿性や他社から購入した際の契約上持ち出し不可なものが多い。持ち出し可能な場合は、産総研の役割は共同研究や全権を委託されて企画、分析、成果普及まで行っている。

AIプロジェクトでは、ビッグデータの存在が欠かせない。それは、ID-POS データのようにそれ自身のサイズの大きさだけでなく、ID-POS データとアンケートと気象情報など複数のデータが統合されたことで実現された大きさも含む。プロジェクトの継続性や成果のクオリティにはデータの質が大きく関わり、なるべくコストをかけずに長期間持続的に収集できることが望ましい。ここでは、**データ収集法**、**コストの程度**、**コストの内容**について考察する。データ収集のコストとは、金銭的、投入人員、システム構築の有無、物理的、時間コストを総合したものとする。そのため、まずコストの程度を調べその内容についても記述した。データ収集のコストの程度を分類すると、11プロジェクトが低い、12プロジェクトが高い、高いものと低いものが存在するプロジェクトが5件となった。コストが高いものと低いものが混在しているものは、プロジェクト初期時点でデータを収集するためのシステム構築にはコストがかかるが、構築後の運用コストは低いとなっている。

コストが低いプロジェクトの特徴は、使用するデータがPOS データ、ID-POS データなど業務中に自動収集されているデータ、元々業務中に使用しているデータ、Web の閲覧履歴やIoT 技術を使用してタブレットや機械から自動収集されるデータなどを使用する場合である。また、大学病院などで医学研究や介護研究の別目的でデータを集める想定があった場合である。また、アンケートや実験をする場合に、相手先の中に被験者がいて彼らの業務中に参加する流れができていてもコストが低くなる。自動であること、既存であること、通常業務の中に組み込む工夫がなされていることが重要である。

一方で、No. 7、8、23、25、26 のメーカー企業では、データの収集コストが高くなっている。これらは、自社データのみを使用しているが、インターネット上や企業ネットワーク上にデータがなく、支社や部署にそれぞれデータの保管場所が拡散し、またその形式が複数あり、集めることも繋げることも自動化が難しいという共通点がある。電子ファイルになっていれば、データ管理ができていないということではなく、ネットワーク上で一元的に管理され、構造化したデータベースになっていることが重要である。No.1 と No.2 は開発のための情報収集に協力してもらおうモニターへの謝金や招集、開発後も精度を高めるためには開発されたアプリの認知度を高めなければならないコストがかかる。No.14 の救急医療のプロジェクトでは、救急車出動履歴、救命救急士が救急車内でタブレットにより収集した情報などそれぞれはネットワーク上で管理されているが、一元管理がなされていないために、分析用に加工するためのコストが高くなっている。地域での健康イベントを行うこと

により住民参加者の健康データを収集し、地域貢献や予防医療を行うプロジェクト(No. 15)では、地域イベント開催のためのスタッフ配備、運営コスト、集客コストが高くなっている。その他では、No.20 と 21 の広告・マーケティング企業では、それぞれのクライアント企業が求める分析サービスを提供するために、消費者関連の様々なビッグデータを収集し、大規模データセットを構築している。この場合も、各データはインターネットやネットワーク上で収集可能だが、統合してデータのプラットフォームを構築するコストが高くなる。

4.3 AI 技術による分析

ここでは、表2の(5)分析についての項目の、分析場所、分析に使用したAI技術とAI技術で行ったことについて説明する。本稿で取り上げたプロジェクトでは、2007年以前に開始された6プロジェクトはBNを用いているが、2009年以降の22プロジェクトでは、BNのみ、PLSAのみ、PLSAとBNを組み合わせたプロジェクトが18プロジェクトであった。期間を通じて24プロジェクトがAI技術を使用している。分析場所に産総研が含まれるプロジェクトは、28プロジェクト中、10プロジェクトであった。分析場所が産総研で無い場合は、技術移転によって相手先での分析が可能となっている。4.1節の表3とほぼ同じことになるが、AI技術を使って行ったことのうち相手先企業の支援に役立つ分析を行ったのが21プロジェクト、ユーザや消費者支援のための技術は7件であった。具体的に行ったことは、ユーザモデリング、グラフィカルモデリングの作成、レコメンド、シミュレーション、セグメンテーション、サーチ、可視化、入力支援、リスク抽出、最適化分析など多岐に渡る。

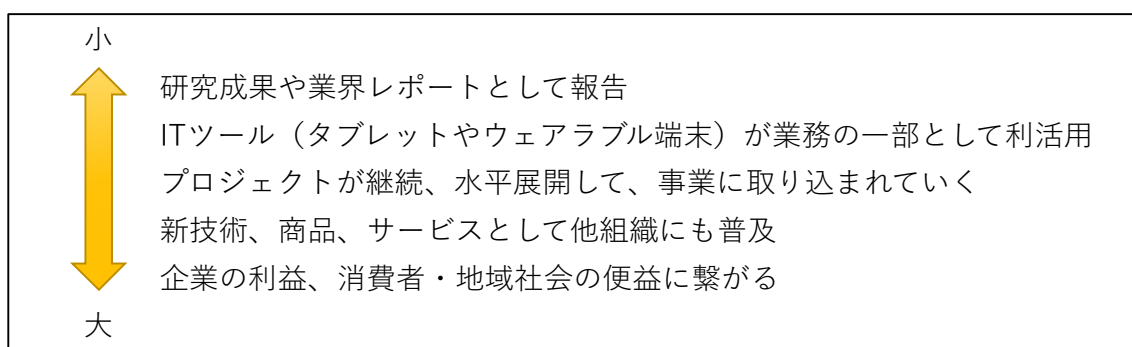
PLSAとBNを組み合わせた分析の流れを説明すると、まずID-POSデータなど顧客情報と行動がわかるビッグデータを収集し、PLSAによって意味のあるいくつかのセグメントを探索し、顧客と行動を分類する。その上で、分析の目的に応じてグラフィカルモデリング、ユーザモデリング、制御モデリングを行う。ここで、注視したいのが対象としているプロジェクトでは、15件のプロジェクトがアンケートデータの収集を行っている。アンケートはAI分析の中でどのように役立っているのだろうか。まず、PLSAでセグメントが決まると、各顧客は所属確率が高いセグメントに割りあてられる。その後、そのセグメント(目的変数)がどんな説明変数で特徴づけられるか、という関係性をBNのグラフィカルモデルとして構築する。これにより、その後、新しい顧客がアンケートに回答するだけで、どのセグメントに所属するかが推論でき、さらにそのセグメントへの所属確率が高い商品を購入する確率もわかるため、どんな商品を優先的に薦めればよいか、を推論できるようになる。また逆に、購買履歴からも顧客がどの消費者セグメントに属するか判断できる。ID-POSデータの様な大規模データの構成員全員にアンケートを取ることは現実的ではないが、もし一部の顧客からでも取れると、その顧客が代表する消費者セグメントの意見として利用

できることから、そのデータは非常に有用である。またアンケートへの回答データを入力条件にして、所属する消費者セグメントの推定もできるし、セグメント分けを探索的にした後、そのセグメントの特徴の説明変数としてアンケートへの回答を使うこともできる。アンケートの複数の設問と回答の間の関係も確率的なグラフィカルモデルとして構築し、ある設問にこう答えるならば、別の設問にはこう答えるだろう、といった確率推論も可能になる。従来型のアンケート収集法だけでなく、タブレットやスマホのアプリの中で少数の設問を埋め込む対話的なアンケート収集方法を活用することにより、今後、アンケートデータの効率的な収集と効果的な活用も可能になるだろう。こうした人が理解しやすい表現であるアンケートの設問と回答を変数とした計算モデルをAIが使うことで、推論プロセスの中に人の介入する余地が生まれ、ブラックボックス化が避けられる。また、データのみからの機械学習に加えて、人の知識を埋め込みやすくなることで、計算モデルを学習するための計算負荷を下げ学習速度や予測の精度をあげることができる。これが「人と相互理解できるAI」が目指すところである（本村(2016)を参照）。

4.4 プロジェクト評価：成功の種類と秘訣と好例紹介

表2の(6)のプロジェクトの評価と当該プロジェクトにおける困難な点について説明する。図3に、本稿での成功の種類とその程度を示した。全28プロジェクト中、17件が図3に見られる何らかの形での成功を収めたと言える。

図3 プロジェクト成功の種類と程度



通常の研究開発プロジェクトにおいては、研究開発の最終段階で技術の性能が十分でない場合が失敗であるが、実サービス志向のプロジェクトの場合はそれとは異なる基準で判断される。本稿における失敗の基準は、途中で事業が中止や立ち消えになったり、後継プロジェクトがない場合とする。なぜなら、これらのAIプロジェクトはサービスの現場から出てきたニーズに依拠するもので、その現場がある限り目的を変えながら継続することが求められるからである。従来の研究開発プロジェクトにおいて事業化が進まない困難について、「魔の川」、「死の谷」、「ダーウィンの海」と呼ばれるものがある。これはそれぞれ、基

礎研究→開発→応用→実用化が実際の現場ではリニアな過程となりにくいことを意味する。AI プロジェクトが途中で中止または継続しないときの主な理由には以下のものがある。

1. 担当者の異動や意識共有の不足

中心となる担当者が異動するなどした場合、後任の担当者がプロジェクトの目的や意義、内容を理解できなかつたり、説明が不十分であるためにプロジェクトが終了してしまう。また担当者が目的や意義を理解していても、チームや現場で意識が共有されていない場合には、特定の期間や予算内に費用対効果や結果が見えないことが原因でプロジェクトが終了することがある。

2. 社内の体制構築が問題になる場合

1.とも通じるが、研究開発（R&D）部門では重要性が理解されプロジェクトが進んでも、実際の業務の中心となる事業部に共有や理解が無い場合も、プロジェクトの社内実装や継続は難しい。また、事業部がプロジェクトに参加した場合でも、横断的な社内体制を作るための人材育成が上手くいかないと継続は難しい。このケースは各部署の独立性や権限が大きい大規模メーカーに多い。

3. データ収集の金銭的、人的、時間的コストが高い

コストを理由にプロジェクトが中断するケース。初期のデータ収集のためのシステム構築で予算不足になったり、収集、加工工程を自動化できず人員を割かなければいけないような場合がこれにあたる。また、ユーザの獲得にコストがかかり、データ収集が困難な場合もある。

次のステップとして、成功したプロジェクトは上記1～3についてどのような違いや仕組みがあるのかを整理していく。1と2については、AI プロジェクトに関わらず、いかなるプロジェクトが成功するためにも必要な事柄である。成功するプロジェクトは、まず強力な明確な目的意識があり、それが複数人で共有されている。さらにチームとして組成され、その中に現場や研究だけでなく事業部やマネジメントなどのメンバーが入っていたり、連携がとれている。そのため、実験的に行った事例を全社的に広げたり、デモンストレーションで被験者を使ったデータ収集を全顧客に拡大するといった、次のプロジェクトへの展開が可能となる。

AI プロジェクト特有の成否に影響を与える事柄は、前述の様にデータの存在の大きさで、4.2節ではデータ収集コストとプロジェクトの継続性について議論した。機械学習に基づくAI プロジェクトの継続性や成果のクオリティにはデータの質が大きく関わり、なるべくコストをかけずに長期間収集できることが望ましいからである。失敗するプロジェクトでは、データ収集の部分（上記3.）がコスト面でもシステム面でも上手くいっていないことが多

い。

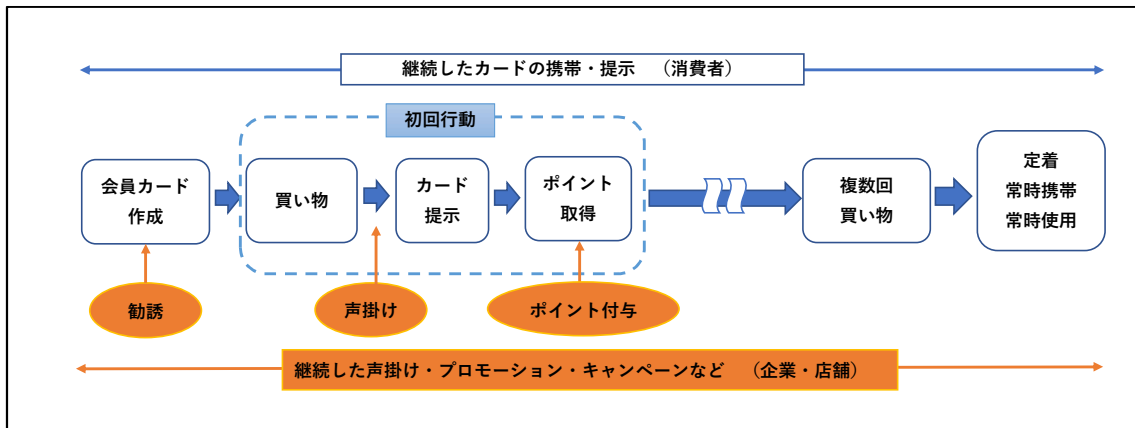
成功したプロジェクトは、POS データの様に既に業務上自動収集しているものや、他の目的で整備しているデータなどを使用してコストが抑えられている。また、データの新規・既存に関わらず、分析のためのデータベース構築システムを作る場合にも初期費用はかかっても、運用コストが低くなるような設計をしているところが多い。

より AI プロジェクトらしい点は、例えばタブレットなどの携帯端末を使って業務データを収集する際に、データが持続的に収集できる仕組みが現場に定着することにどれだけ注力できるかである。図 4 は共通ポイントカードの定着プロセスである。これを消費者、現場両側から観察したい。まず消費者は、会員カードを作成しプログラムのメンバーになる。登録などは済ませたとして、次の大事なステップはカードを携帯することである。その上で店舗に来店し、買い物をしてカードを提示してポイントを取得する。この一連の行動でポイント取得や値引きで便益を享受できたと思えば、携帯と提示行動は習慣化するだろう。定着した顧客が増えれば、情報は爆発的に増加していく。通常の研究開発プロジェクトと異なり、データ収集の継続のためのシステム、環境作りの成否がプロジェクト評価に直結する。

現場側から見ると、まずデータ収集対象となる消費者を増やすため勧誘を行わなければならない。次に来客時にカードの有無を聞き、所持している場合には提示を促す必要がある。ここが重要で、この時点ですべての従業員がすべての客に声がけしなければ、定着は上手くいかない。従業員にとっては声がけし、カードを認証し、ポイントを付与するという複数タスクが増える。しかし、これを徹底しなければ、質の良いビッグデータの収集にならない。この部分が、業務工程に新たなタスクを入れてデータを収集するすべての AI プロジェクトに共通する難しいところである。ここをいかに負担を軽く、いかに速やかに通常業務にできるかがプロジェクトの成否と大きく関わる。負担の軽減は IT、ICT、IoT 技術で工夫できるところで、通常業務への組み込みはプロジェクトの計画やデザイン、つまり最初の目的設定に依る。消費者の定着にはそれに加えて、常時携帯、常時提示をするインセンティブを与えなければならない。企業はそのためのポイント付与の告知、ポイント使用の利便性向上、キャンペーンやプロモーションを継続して行う必要がある。逆言えば、これらのプロセスが、プロジェクトが失敗する際の原因となり得る。共通ポイントカードにしても、近年急に、会計時の声掛けが徹底してきたように感じる。それは、例えば会員数を増やすノルマがあるのか、処理が簡単になったので聞くのが苦にならなかったのか、自分も消費者でポイント付与の利益を得ているので聞かないと心苦しいのか、理由は各企業の戦略で様々であろう。しかし、重要なことは、「すべての従業員が、全ての客に声掛けをすること」とポイントカードの所持と提示を定着させる努力を徹底することである。それができたプロジェクトは、間違いなく成功する。さらに、一部のサービスではネットにログインしてアンケートに答えるとポイントもらえるものもあるので、購買データと心理データを紐づけることもできる。そうすると、そのデータは当該企業の枠を超えて他企業、

業界団体、データ収集業者、さらには各省庁の統計調査部局に対しても付加価値を持つようになる。

図4 ポイントカード定着までのプロセス



最後に 28 件あるプロジェクトの中から、3 つの成功事例を紹介する。まず、2 つのプロジェクトは、IoT デバイスを活用することを通常業務に落とし込み、そのタブレットから行動系、感情系ビッグデータを収集している。一つは No. 16 の介護施設のプロジェクトで、介護業務の改善という共通の強い目的意識があり、かつその改善内容が「申し送り」と「業務記録データ」の入力支援と電子化しネットワーク上に置くことで介護士、入所者、家族と共有できることと非常に具体的であった。プロジェクト自体は、まだ続いており更なる業務改善やデータの活用などの課題はあるが、すべての介護士がすべての入所者との対話や介護にタブレットを利用することで、通常業務の中に完全に組み込まれ、追加のコストはタブレットの運用コストという小さなコストのみで、安定的に継続してビッグデータを収集できている。

二つ目の例は、No. 24 の自動車ディーラーのプロジェクトで、営業マンが来店した顧客の情報管理や、自動車に対する嗜好などを把握する取り組みを徹底し、日常業務で無理なくデータ収集するシステムの利用方法を工夫した。そのため通常業務に落とし込むことができた。今後は、ディーラー用支援アプリや端末の台数を増やし、No.16 のプロジェクト同様に、低コストで常時データ収集が可能となることを目指す。

三つ目のプロジェクトは、自社プロジェクトでの開発技術がサービスとして確立し、ビジネスになった現状での最終形と呼べるものである。表 1 では、あるプロジェクトから派生したり継続したプロジェクトには「二巡目」と記入している。巡るという字を当てたのは、サービス現場で重要とされている PDCA (Plan, Do, Check, Action) サイクルと同様に、プロジェクト自身もサイクルを持つことが重要であるという意味を込めている。

AI プロジェクトの場合は特に、データの収集を実用化の過程に近い現場で行うため、先に実用化可能な領域を設定し、そこでテスト運用を回しながら徐々にスケールを大きくする

スパイラル型の過程をとることが多くなる。この過程を表4で示している。No. 21の東急エージェンシーとの共同研究は2013年にスタートし、産総研からの技術移転を受けID-POSデータに対してAI技術(PLSA)を利用したツールと分析サービスを共同開発した。商用ツール「Target Finder」を商品化し、現在では自身のクライアントにシステムと分析サービスを提供する新ビジネスを立ち上げ、20社以上との取引実績を上げ、軌道に乗せた¹⁰。このプロジェクトでは、表4で示す実装とイノベーションのサイクルが、プロジェクト内でも、また他組織への波及という形で継続していく中でも実現できたと言えよう（ただし、5巡目まで）。

表4 プロジェクトの循環とその内容

サイクル	内容
1巡目	既存データ→分析→解釈（レポート作成）
2巡目	システムを導入することで持続的なデータ収集→分析→活用（研究部門から担当業務部門への共有）
3巡目	2巡目のフローを自動で実行できる新規データ統合システムを構築する。 → AIの学習（モデル構築）速度、精度が飛躍的に向上。モデルの活用場面が組織内で広がる。
4巡目	構築したモデルの再利用性が飛躍的に向上し、費用対効果増大。モデルを活用した業務支援のスピードと精度も飛躍的に向上することで生産性が向上。
5巡目	業務システムのAI化、デジタル化が急激に進むことで、各モジュールの接続性と、組み合わせ効果が増大。少ないリソース、限られた時間で高い効果が創出できる。他組織との連携、モジュール接続も高まる。
6巡目	組織や地域限定性が意味をもたなくなり、社会システムの融合が加速する。

5. 総括と今後の課題と政策含意

本稿での分析を通じて、どのようなAIプロジェクトが持続的に発展するのだろうか？、というシンプルな問いに対して以下の答えが見えた。1. プロジェクトの目的や技術のユースケースが明確で、目標となる指標（目的変数）や目標を達成するために制御可能な変数を観測データに落とし込めること、2. 現場のニーズとモチベーションが強いこと、3. データがセンサ、ネットワークやインターネット上を通じて低コストで集積できること、4. 業務プロセスの一部に組み込まれることでデータ収集が持続できること、5. データ活用や分析して得られた知見をもとに介入を行った結果がさらにデータとして追加できること、などである。このように、AIプロジェクトにおいてはデータの担う役割が非常に高いことがわかる。一方、AI技術やアルゴリズムは研究現場において開発が進められているが、ソフトウェアとして公表されれば、既存のハードウェアやクラウド上ですぐに利用可

¹⁰ 経済産業省広報誌（2017）に紹介された。

能になるため広く普及する可能性も高く、各現場で既に使われているシステムと組合わせて改良し利用することも現在のコンピュータ環境においては容易である。

AI プロジェクトの成功の可否は、目的に適合した質の高いビッグデータの収集と収集コストを低くすること、そして何よりその活動と改善を継続できるかどうかにかかっている。データの質の良さは消費者現場でも企業側の現場でも、目的に合った変数が適切に選択されているか、データを収集する技術が現場に定着しているかに依存する。できるだけ多くの人の多くの行動をデータとして収集することが分析の精度に影響を与えるし、データの価値を高めることとなる。AI プロジェクトはビッグデータの状態の善し悪しにかかっており、プロジェクトの規模を表すのにも、従来型プロジェクトでも用いられるプロジェクト費、人数、施設の大きさに加え、データ収集の頻度も示されることが望ましい。本稿では、表1において一部のプロジェクトにおいては、会員数やアンケート参加者数など記述しているが、より正確に規模と負荷がわかる指標として、今後はデータ収集の頻度を表すトランザクション数を示すことが有用であろう。

例えばプロジェクト No. 9 の「コープ K」プロジェクトでは、会員数は約 100 万人、1 年間で 669,511,467 (約 7 億) トランザクション数となる。トランザクション数は計算機やサーバーに与える負荷と非常に相関が高い。この中で分析に使用したのは、アンケートを行った 3981 名、売上げ上位 1000 アイテムの 420 万トランザクションで、これが実際の分析時のデータ負荷になる。データの移送や収集に掛かる時間は各組織の情報技術環境に依存するし、データサイズもイメージが沸きにくい。分析対象の行動の頻度の高低を知るためにもトランザクション数を調べることは意味がある。

今後の AI 技術を社会の中で活用するためには、持続的なデータ収集を行う仕組みが不可欠なため、自らがインターネットサービスを提供し、多数のユーザを獲得している Google, Amazon, Facebook, Apple(GAFA)のような巨大 IT 企業は強みを持っている。しかし、これからは彼らだけが勝者となるわけではない。なぜなら、今後 AI 技術が社会に浸透していく際には、表4に示したような循環型の研究が推進されていくからである。AI プロジェクトはデータを通じて AI が学習するだけでなく、プロジェクト自体も学習、進化しながら持続的に発展する。

今後は、より社会的なニーズの高い問題を設定し、生活やサービスの現場でビッグデータの収集と技術活用が並行して行われていく。その過程では、単独企業ではなく多くのステークホルダーが連携していくことが不可欠である。産総研人工知能技術コンソーシアムは 2015 年 5 月、産総研人工知能研究センターの設立と同時に発足し、2017 年 1 月現在で 85 社、100 名以上のメンバーで活動している¹¹。2016 年 11 月には関西支部、2017 年 2 月には九州支部も開設され各地域の特色を活かした業種間の連携を進めることで、日本全国で大手から中小、ベンチャーも含む企業や地域への人工知能技術の普及と促進、さらに活用の

¹¹ 経済産業省広報誌 (2017)に紹介された。

ための連携を進め、新プロジェクト発足を促進している。そして、その付加価値として、ビッグデータのプラットフォーム化が進んでいくことが期待される。近年の AI プロジェクトの状況では、アルゴリズム改良よりも、機械学習に適用するデータによる改善効果が高いため、データの価値が圧倒的に高いのである。その価値にいち早く気づき、どのようなデータを収集、活用するかを議論する場が国際的な競争力を高めるためには必要である。国、民間、学術分野それぞれで、または連携し、一つでも多くの広く、大きく、価値のあるデータプラットフォームの構築が望まれる。

6. 参考文献

邦文

- [1] 大津展之 (2016) 「RWC プロジェクトを回顧する」
https://sankoukai.org/secure/wp-content/uploads/untold_stories/nobuyuki-otsu_final.pdf
- [2] 小野智弘, 黒川茂莉, 本村陽一, 麻生英樹, ユーザ嗜好の個人差と状況差を考慮した映画推薦システムの実現と評価, 情報処理学会論文誌, 49,1, (2007)130.
- [3] 落合香, 下角哲也, 小野智弘, 麻生英樹, 本村陽一: ベイジアンネットを用いた映画コンテンツのマーケティング支援, 人工知能学会全国大会, 3D2-NFC1-1, (2009).
- [4] 経済産業省広報誌 (2017) 「もしもがかなう社会へ! 明日の AI」、METI Journal 2-3 月号、pp. 14-18.
- [5] 繁梈 算男、本村陽一、植野 真臣 (2006) 『ベイジアンネットワーク概説』、培風館.
- [6] 総務省(2016) 「人工知能 (AI) の現状と未来」『情報通信白書』第 4 章第 2 節、pp. 232-241.
- [7] 照井伸彦(2017) 「ビッグデータ統計解析入門 6」『経済セミナー』、2-3 月号、日本経済評論社、pp.78-88.
- [8] 本村陽一、岩崎弘利 (2006) 『ベイジアンネットワーク技術: ユーザ・顧客のモデル化不確実性推論』東京電機大学出版局.
- [9] 本村陽一(2009)大規模データからの日常生活行動予測モデリング -実サービスを通じたベイジアンネットワークの学習と推論-、シンセシオロジー, vol2, no.1, pp.1-11.
- [10] 本村陽一、竹中毅、石垣司 (2012)『サービス工学の技術: ビッグデータの活用と実践』東京電機大学出版局.
- [11] 産業技術総合研究所(2014)『社会の中で社会のためのサービス工学』、カナリア書房
- [12] 本村陽一(2016), 次世代人工知能技術, 情報処理学会誌, Vol.57, no. 5, pp. 466-469.
- [13] 山本裕樹、松尾豊(2016) 景気ウォッチャー調査の深層学習を用いた金融レポートの指数化、The 30th Annual Conference of the Japanese Society for Artificial Intelligence, pp. 1-4.

英文

- [1] Hinton, G. E. and Salakhutdinov, R. R. (2006) “Reducing the dimensionality of data with neural networks,” *Science*, Vol. 313. no. 5786, pp. 504 – 507.
- [2] Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986) “Learning representations by back-propagating errors”. *Nature* 323, pp.533–536.

図1 AI技術とその周辺のできごと

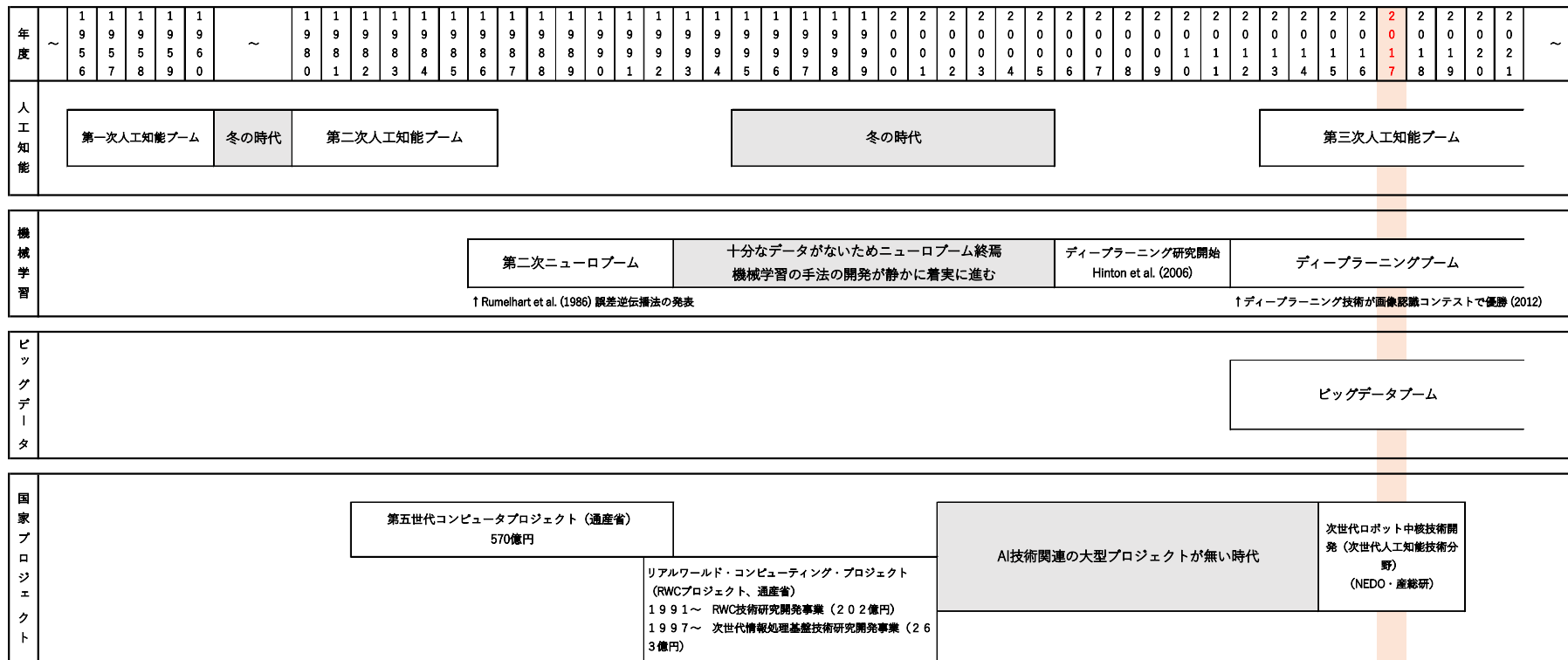


表1 AI技術を用いた産総研プロジェクトのレビュー①

No.	企業名	業種	実装期間	継続期間 2017.2現在	予算規模 (概算)	目的	プロジェクト の主たる現場	現場での実施内容	現場への技術 介入期間	産総研の役割	分析に使用した データの種類の	データの外部への持ち 出し	データ収集法	データ収集 コスト	コストの内容	分析場所	分析に使用した AI技術・統計手法	AI技術等 で行うこと	プロジェクト 評価	困難だった点
1	デンソーITL	機械製造	2004-2006	2年間	300万円/年	カーナビ上での レストラン推薦 アプリの研究開発	デンソー ITL開発部	レストラン推薦 のiPhoneアプリ 開発	実験期間 常時	①技術移転 ②共同研究	①ネットアン ケートデータ <u>(数千人)</u> ②対面インタ ビューデータ ③カーナビ操作 履歴	×	①ネットアン ケート ②疑似環境で のカーナビ操 作	高	疑似環境でのモニ ターへの謝金有 り。持続的なデー タ収集にはユー ザー増が必須だ が、アプリが評判 になることが条件 となる。	デンソー ITL 産総研	ベイジアンネッ トワーク	①グラフィカルモ デリング(店舗、 状況、顧客) ②相関 ③可視化	書籍化 iPhoneアプリ開 発に必要な 技術特許取得	R&D部門での開発には 成功したが、事業継続の ための事業部への橋渡し のためのビジネスモデル 構築が困難だった。
2	デンソーITL	機械製造	2006-不明	不明	300万円/年	①カーナビ上での レストラン推 薦アプリの完成 ②アプリからの 収集データによる 分析	インター ネットサー ビス上	①レストラン推 薦のiPhoneアプ リ開発 ②iPhoneのアプリ 利用者からの データ収集	常時	アプリを共 同開発	①ネットアン ケートデータ <u>(数千人)</u> ②クリック履歴	×	①ネットアン ケート ②アプリ利用 履歴	高	ユーザーが増えれ ば持続的なデータ 収集が可能だが、 アプリが評判になる 必要あり。	デンソー ITL	ベイジアンネッ トワーク	レコメンド	特許技術を利用 し、 iPhoneアプ リ開発成功 No.1の二巡目だ が継続なし	ユーザー獲得、ユーザー 数増加が困難。膨大な iPhoneアプリに埋もれる。
3	KDDI	情報通信	2005-2007	2年間	数千円/年	消費者への映画 作品の推薦	インター ネットサー ビス上	①データ収集 ②携帯・イン ターネットによる レコメンド実 験	実験期間 常時	①技術移転 ②共同研究	①クリック履歴 ②ネットアン ケートデータ 1600名の被験者	×	①インタ ビュー ②ネットアン ケート ③携帯端末・ PCネット操作 履歴	システム構築 高、運用は低	新規開発した情報 サービスのシステム 構築コストは高い が、運用コスト は低い。	KDDI研究 所	ベイジアンネッ トワーク	映画推薦のための ユーザモデリング	研究発表に成功 プレス発表、 NHK放映	企業内にビジネスモデル や業務内での技術活用の 場を企画・計画する人が 不在。研究者に現状を変 えられる権限がない。
4	松竹	娯楽業	2007-2008	1年間	600万円	興行収入増のため の映画館の上 映内容と上映ス ケジュール決定 の支援	松竹マーケ ティング部	データ収集	一時的	コンサルティング	①テキストデー タ ②ネットアン ケートデータ <u>(数千人)</u>	×	①インタ ビュー ②ネットアン ケート	低	業務で収集・使用 するデータを流用 するのでコストは 低い。	松竹の共同 研究者 (KDDI研究 所)	ベイジアンネッ トワーク	興行収入予想のため のユーザモデリ ング	研究発表に成功 No.3の二巡目	企業内にビジネスモデル や業務内での技術活用の 場を企画・計画する人が 不在。研究者に現状を変 えられる権限がない。
5	松竹	娯楽業	2008-2008	1年未満	数百万円台 +既存リ ソース	DVDパッケージ の設計支援 (商品キャッチ コピーの最適 化)	松竹DVD販 売部	業務支援ツール の導入・試用	一時的	①技術移転 ②コンサル ティング	①ネットアン ケートデータ (大規模) ②業務支援ツール の履歴データ	×	①ネットアン ケート ②日常業務支 援ツールの操 作履歴	低	ツール、データと も共同研究者が開 発したものを利 用。	松竹の共同 研究者 (KDDI研究 所)	ベイジアンネッ トワーク	「見たさ」のシ ミュレーション	研究発表に成功 商品利用に成功 No.3の三巡目だ がその後継続な し	社内システム構築予算の 不足から、継続して成果 を利活用するための社内 部署や人材が不足してい る。

表1 続き②

No.	企業名	業種	実装期間	継続期間 2017.2現在	予算規模 (概算)	目的	プロジェクト の主たる現場	現場での実施内容	現場への技術 介入期間	産総研の役割	分析に使用した データの種類の	データの外部 への持ち出し	データ収集法	データ収集 コスト	コストの内容	分析場所	分析に使用した AI技術・統計手法	AI技術等 で行うこと	プロジェクト 評価	困難だった点
6	A社	教育	2007-現在	10年間	数百万円+ 既存リソース	①子どもの行動 予測 ②リスク予測	Webサイト	①子どもの事故 予防アニメ (CG作成)の レコメンド ②データ収集	常時 期間限定	①技術移転 ②共同研究	①会員データ ②Web閲覧履歴	×	Web閲覧履歴 を保存	低	Web閲覧履歴のため コストは極小	B社	ページアン ネットワーク	①子どもの事故認 知のためのユーザ モデリング ②CGレコメンド	担当者異動のため 中止	読者向け無料サービスの ため、先方に活動継続の 意志が弱い。
7	B社	製造・小売	2009-2011	2年間	200万円/年	①消費者行動理 解 ②需要予測 ③消費者セグメ ント分けによる 接客支援	店舗	データ収集	常時	①技術移転 ②共同研究	①購買履歴 ②ポイントデー タ ③商品マスター ④キャンペーン 日時 ⑤アンケート データ	×	①ID-POSシ ステム ②携帯電話に よるアンケート	高	全国の店舗から複 数の形式で保存さ れたデータを集め て、統合する部分 に作業コストがか かる。	委託先	①PLSA(確率的 潜在意味解析) ②ページアン ネットワーク	①ロイヤル顧客育 成のためのユーザ モデリング ②感度分析	研究発表に成功 IT部門の新機能 となり、マーケ 事業部内にも実 行可能性のコン センサス(POC)が とれた。	アパレル系は技術につい てのリテラシーが低いた め、担当部署への日常業 務への埋め込みが難し かった。
8	B社	製造・小売	2009-2011	2年間	数百万円台/ 年+既存リ ソース	キャンペーンや 製品プロモー ションの提案支 援	事業部	①データ収集 ②顧客への商品 のレコメンド	常時	①技術移転 ②共同研究	①タブレット操 作履歴 ②アンケート データ	×	①ID-POSシ ステム ②店内タブ レットによる アンケート	高	全国の店舗から複 数の形式で保存さ れたデータを集め て、統合する部分 に作業コストがか かる。	B社と委託 先	①PLSA(確率的 潜在意味解析) ②ページアン ネットワーク	①店舗、施策マネ ジメントのための ユーザモデリング ②レコメンド	No.6の二巡目だ が、その後継続 なし	POCから事業部内で継続 運用への移行で予算不足 となり、体制変更のため 活動終了。
9	コープK	小売業	2009-2010	1年間	1000万円台	①消費者行動理 解 ②需要予測 ③消費者セグメ ント分けによる 利益向上	店舗	データ収集	一時的	①計画 ②データ分 析 ③研究	①ID-POS ②アンケート データ 100万人の会員情 報	○	①郵送アン ケート ②ID-POSシ ステム	低	DMアンケートは研 究成果から得られ た。ID-POSデー タは業務データのため コスト(人件費、 システム構築費用) は低い。	産総研	①PLSA(確率的 潜在意味解析) ②ページアン ネットワーク ③線形回帰	①グラフィカルモ デリング(購買行 動・ライフスタイル) ②各種データ間の 関係分析 ③セグメンテー ション	研究発表に成功	分析結果を活用できる部 署や業務、担当者を見つ けることが困難。新しい 知見は生まれるが、実行 に移すための方法や体系 を提案する必要がある。
10	コープK	小売業	2017-現在	---	既存リソ ースのみ利用 のため追加 コストなし	①消費者行動理 解 ②需要予測 ③消費者セグメ ント分けによる 利益向上	店舗	データ収集	常時	①一時的コ ンサルティ ング ②データ分 析 ③研究	①ID-POS ②アンケート データ 100万人の会員情 報	×	①ID-POSシ ステム ②業務で収 集・使用する データを流用	低	コープKを含むコー プ系全国組織から のデータ提供があ れば既存システム を利用できるため 運用コストは低 い。	コープ系全 国組織から の全国展開	①PLSA(確率的 潜在意味解析) ②ページアン ネットワーク ③線形回帰	①グラフィカルモ デリング(来店・ 購買行動) ②各種データ間の 関係分析 ③セグメンテー ション	未定 No.9の二巡目事 業	費用対効果を勘案し、担 当部署を選定中。執行役 員の了解は得られてい る。持続的活動を定着す ることが鍵。

表1 続き③

No.	企業名	業種	実装期間	継続期間 2017.2現在	予算規模 (概算)	目的	プロジェクト の主たる現場	現場での実施内容	現場への技術 介入期間	産総研の役割	分析に使用した データの種類の	データの外部 への持ち出し	データ収集法	データ収集 コスト	コストの内容	分析場所	分析に使用した AI技術・統計手法	AI技術等 で行うこと	プロジェクト 評価	困難だった点
11	がんこグループ	飲食業	2009ー現在	8年間	数千万円台	①従業員の最適シフトの発見 ②サービス時の従業員の店内の最適配置	店舗	①データ収集 ②実験	一時的	①計画 ②データ分析 ③研究	①従業員位置情報 ②売上げデータ	○	①センシング ②POSシステム	センシング高、POS低	センシングには店舗に合わせた専用機器と専門家のサポートが必要でコスト高い。POSデータは業務データのためコストは低い。	産総研	相関分析	①セグメンテーション ②行動推定	研究発表に成功	日常業務の邪魔にならない形で、持続的にデータを収集することが困難。
12	長崎大学	医療	2010ー現在	7年間	100万円/年	①医療現場の安全 ②看護業務改善	病院	①データ収集 ②看護業務改善	年に一回 一ヶ月を 複数年	共同研究 科研費取得	①看護プロセスの時間計測 ②看護行動の体系化 ③業務記録 ④病棟管理データ	×	①タブレット利用履歴 ②対話システム利用履歴	条件付き 低	看護研究の中で調査が実施できると教育コストの中に含まれる。	長崎大学 産総研	変動分析	①サーチ ②可視化 ③入力支援	看護研究としては成功	業務効率の向上を実現することが課題
13	佐賀大学	医療	2010ー現在	7年間	既存リソースのみ利用のため追加コストなし	包括医療費支払制度 (Diagnosis Procedure Combination: DPC) の統計解析	病院	医療データの収集	一時的	共同研究	医療データ	○	既存の医学用データと研究用データを利用	低	医学部の研究活動の中で、構築済みデータを流用するためコストは低い。	佐賀大学 産総研	①PLSA(確率的潜在意味解析) ②ベイジアンネットワーク ③線形回帰	医学研究のためのデータ分析、新たな分析手法の検討	研究発表に成功、学会でのデータ活用の活動などに発展	医療人材不足で新たな取組みの実現が困難。医療人材以外が院内で業務を行えないので、研究協力を結び経営改善プロジェクトとしたが予算が限定的。事業化の起案、稟議を上げる難易度が高い。
14	佐賀大学	医療・救急	2010ー現在	7年間	既存リソースのみ利用のため追加コストなし	①救急車の配車支援 ②救命救急センターの受け入れ最適化支援	救急現場 (家庭から救急病院まで)	救急医療の受け入れまでのデータ収集	一時的	共同研究	①救急車出動履歴 ②救命救急士がタブレットにより集めた患者のアセスメントデータ	×	①救急車出動記録データ ②患者データ ③タブレット	高	各業務データはネットワーク上にあるが一元管理されていないため、分析用に加工する人件費コストが大きい	佐賀大学 産総研	線形回帰	地域医療への貢献としてのデータマイニング、行政へのレポート	研究発表に成功 No.13の二巡目	救急は第三次医療機関、行政、消防が連携しており、新たな取組みを行う体制構築は非常に難しい。横断的な地域の取組みとして推進する必要があり、その成果報告が継続的であることが必要。
15	千葉大学	地域保健 看護	2014ー現在	3年間	学生数人+ 既存リソース	①医療費削減 ②予防医療 ③地域看護への貢献	地域社会	健康データの収集	常時	コンサルティング	①来場者データ ②タブレット操作履歴 ③参加履歴 ④アンケートデータ	×	①来場者の健康計測データ ②対話アンケート	高	地域健康イベントのスタッフ配備、運営コスト、集客コストが大きい。	千葉大学 産総研	①PLSA(確率的潜在意味解析) ②ベイジアンネットワーク	①対話アンケート ②来場者分析 ③レコメンド ④リスク抽出	研究発表に成功	人材不足で新たな取組みの実現が困難。研究協力を結び地域貢献プロジェクトとしたが予算が限定的。事業化の起案、稟議を上げる難易度は非常に高い。

表1 続き④

No.	企業名	業種	実装期間	継続期間 2017.2現在	予算規模 (概算)	目的	プロジェクト の主たる現場	現場での実施内容	現場への技術 介入期間	産総研の役割	分析に使用した データの種類の	データの外部への持ち 出し	データ収集法	データ収集 コスト	コストの内容	分析場所	分析に使用した AI技術・統計手法	AI技術等 で行うこと	プロジェクト 評価	困難だった点
16	石川県介護施設	介護	2013-現在	4年間	数千万円台	介護業務改善	介護施設	データ収集	常時	①常時遠隔 作業 ②データ分析	①申し送りデータ ②業務記録データ	×	タブレット利用 履歴	低	通常業務の中に組み込むことで追加コストは機器の運用コスト程度にできた。	産総研	テキストマイニング	①テキスト分析 ②可視化 ③リスク抽出	研究発表に成功 業務に利用することに成功	収集データのさらなる活用、効果評価、生産性向上、業務改善が次の課題
17	コープK介護	介護	2017-現在	---	500万円	①高齢入居者と デイケア利用者の健康及び生活 支援 ②地域交流・貢献	介護施設	①データ収集 ②介護サービス 支援	常時	①常時遠隔 作業 ②データ分析	①レセプトデータ ②タブレット操作履歴 ③アンケートデータ	検討中 要倫理委 員会審査	①ID-POSシステム ②対話システム ③タブレット 上でのアンケート	システム構築 高、運用は低	持続的なデータ収集は開発した支援ツールの評価次第。自動化端末のため初期導入コストはかかるが運用費は低い	未定	①PLSA(確率的 潜在意味解析) ②ベイジアン ネットワーク ③線形回帰	①アンケート推定 ②対話アプリ	未定 No.9の三巡目事業	担当部署の部長の了解は得られたが、業務のどの部分に技術導入するかとユーザーの確保が課題。リスクとコストが低いことが条件となる。
18	C社	電力・ エネルギー	2009-現在	8年間	既存リソース+新規センサ+学生 数名	利用者協力型節電行動の促進	産総研のセンサハウス	①行動データ収集 ②分析 ③モデル化	一時的	共同研究	①室内に設置した センサデータ ②アンケートデータ ③システム操作履歴	○	独自システム 構築	システム構築 高、運用は低、 人件費高	データ収集のための初期コストは高いが、運用コストは低い。ただし、実験は都度行うため、被験者を含む人件コストは大きい。	産総研	①標準的な統計 解析 ②ベイジアン ネットワーク	①データ収集 ②可視化 ③構造分析	共同研究期間終了	再現性は高いので、相手先の希望があればコストダウンをはかり『リビングラボ』化が可能。センサ、運用システムの低コスト化が鍵。
19	D社	小売業	2012-現在	5年間	既存リソース+学生1名 (超優秀)	①消費者行動理解 ②需要予測 ③消費者セグメント分けなどによる 買い回り分析	協力先の ショッピング モール管理会社	①モール共通ポイントカードの 分析 1年間280万 レコード、15万人、 200店舗 ②レポート作成	常時 期間限定	共同研究	①共通ポイント データ ②アンケート データ ③会員データ	○	業務データの ダウンロード	低	業務データのためコストは低い	産総研	①PLSA(確率的 潜在意味解析) ②ベイジアン ネットワーク ③セグメンテーション	①店舗マネジメントのための顧客 分析・ユーザモデリング ②顧客購買行動分析 ③セグメンテーション	担当者異動のため中止	ショッピングモール管理会社がデータを活用するタイミングが限られ、活用するモチベーションの維持が難しい。
20	博報堂	広告・ マーケティング	2013-現在	4年間	数百万円台 /年+既存リソース	新たなデータ構築	①博報堂 ②クライアント企業	複数データの 接合	常時	①技術移転 ②共同研究	①ID-POSデータ ②メディア視聴 ③ネットショッピング ④クレジットカード履歴 ⑤共通ポイントカード	×	①POSシステム ②インターネット ③各社のネットワーク	高	複数のネットワーク上のデータを集めて、統合する部分に作業コストがかかる。	博報堂	PLSA(確率的 潜在意味解析)	①データ統合 ②セグメンテーション	ビジネスモデル 特許取得 研究発表に成功	ユーザーへのコンタクト、持続的活動定着が課題。 コンソーシアムWGを活用開始。

表1 続き⑤

No.	企業名	業種	実装期間	継続期間 2017.2現在	予算規模 (概算)	目的	プロジェクト の主たる現場	現場での実施内容	現場への技術 介入期間	産総研の役割	分析に使用した データの種類の	データの外部 への持ち出し	データ収集法	データ収集 コスト	コストの内容	分析場所	分析に使用した AI技術・統計手法	AI技術等 で行うこと	プロジェクト 評価	困難だった点
21	東急エージェンシー	広告・マーケティング	2013-現在	4年間	100万円/年 + 既存リソース	①ID-POSデータを活用した分析サービスの提供 ②システム構築 ③消費者行動理解 ④需要予測 ⑤消費者セグメント分け	東急エージェンシーのクライアント企業	①予備データ分析 ②コンサルティング	常時	①技術移転 ②共同研究	①ID-POSデータ ② Web閲覧履歴	×	クライアントの業務データ	高	東急エージェンシーの営業コストが大きい。クライアント先コストは相手先による。	東急エージェンシー→クライアント先へ展開（ビジネス化）	PLSA(確率的潜在意味解析)	①施策最適化のためのユーザモデリング ②DM最適化	成功	東急エージェンシー社内での業務として立ち上げ、社内での体制構築が課題であった。
22	共通ポイントE社	情報通信	2014-現在	3年間	200万円/年	①消費者行動理解 ②需要予測 ③消費者セグメント分け	コンビニエンスストアなど	①データ収集 ②データ分析	常時	共同研究	共通ポイントデータ 5000万人の会員情報	×	各社のネットワーク	なし	通常業務データのため追加コストはない。	相手先企業	①PLSA(確率的潜在意味解析) ②ベイジアンネットワーク	①レコメンドのためのユーザモデリング ②顧客のアンケート回答シミュレーション（未回答の顧客の回答を予測→マーケティング提案に活用）	従来法より新しい知見の発見に成功	通常業務への貢献、本格システムの開発・実装、費用対効果向上が課題。
23	F社	飲料業	2014-現在	3年間	200万円/年 + 既存リソース	①自動販売機の設置最適化 ②商品配置最適化 ③補充最適化	①自動販売機 ②物流現場	データ収集	常時	共同研究	①購買データ ②商品マスター ③ポイントカード利用者データ	×	各データの保管場所からのダウンロード	高	インターネット、企業ネットワーク上ではなく、複数データが各々に保管されており、収集・統合に作業コストがかかる。	R&D部門	PLSA(確率的潜在意味解析)	自動販売機オペレーション支援のためのシミュレーションを検討中	未定	R&D部門の分析を事業部に活用してもらうために、費用対効果増大が課題。
24	G社	教育、販売	2015-現在	2年間	数百万円/年 + 既存リソース	ディーラー業務支援	ディーラー店舗	データ収集	常時	プロジェクトマネジメント	①販売記録 ②顧客管理データ ③アンケート ④インタビューデータ	×	①インタビュー ②アンケート	低	営業マンが顧客を理解するための通常業務に落とし込んだことで、業務コストの中に含めることができた。	広島大学（JST研究）	ベイジアンネットワーク	①レポート向上のためのユーザモデリング ②ワクワク感のシミュレーション	研究発表に成功業務に利用することに成功	開発システムの水平展開、運用コスト削減、導入マニュアルの整備などが課題。
25	G社	輸送機械製造業	2015-現在	2年間	数百万円/年 + 既存リソース	マーケティング業務支援	事業部	分析ツールの開発	一時的	コンサルティング	①Web閲覧履歴 ②マーケット調査データ ③マーケティングリサーチデータ ④試乗予約データ	×	多部門が連携して紙や電子ファイルのデータをかき集め、データ統合する	高	インターネット、企業ネットワーク上ではなく、複数データが各々に保管されており、収集・統合に作業コストがかかる。	R&D部門+事業部	①PLSA(確率的潜在意味解析) ②ベイジアンネットワーク	①消費者理解のためのユーザモデリング ②ユーザの感性シミュレーション ③マーケティング施策レコメンド	未定 No.24の二巡目	上層部の理解が得られたので、R&D部門の分析を事業部に活用してもらうために担当者の育成を支援中。

表1 続き⑥

No.	企業名	業種	実装期間	継続期間 2017.2現在	予算規模 (概算)	目的	プロジェクト の主たる現場	現場での実施内容	現場への技術 介入期間	産総研の役割	分析に使用した データの種類	データの外部 への持ち出し	データ収集法	データ収集 コスト	コストの内容	分析場所	分析に使用した AI技術・統計手法	AI技術等 で行うこと	困難だった点	
26	H社	医療機器製造業	2015-現在	2年間	数百万円台/ 年+既存リソース	血圧のビッグ データを使用した健康支援分析	家庭や健康 診断会場	血圧計測とデータ 収集	常時	①技術移転 ②共同研究	血圧、気象、健康、 アンケートデータ	×	R&D部門が紙 や電子ファイルのデータ をかき集め、データ統合 する	高	インターネット、 企業ネットワーク上 ではなく、複数データ が各々に保管されてい おり、収集・統合に作 業コストがかかる。	R&D部門	PLSA(確率的潜在 意味解析)	①セグメンテーション ②血圧変動分析	未定	R&D部門での研究成果 の認知は十分なので、 事業部門への橋渡しと 予算化が課題。
27	某保険関連会社	保険業	2016-現在	1年間	数百万円台/ 年+既存リソース	①顧客の健康判断 ②保険商品の開発 ③健康管理サービスの 開発	保険会社	データ収集、分析	常時	①技術移転 ②共同研究	①保険請求 ②保険加入時間 診データ ③顧客属性	×	①データ事業者 からの購入 ②グループ会社 のデータ活用	高	保管場所、形式の 異なる複数データを 収集・統合する部分 に作業コストがかか る。	保険関連会社	①PLSA(確率的 潜在意味解析) ②ベイジアン ネットワーク	①セグメンテーション ②健康リスク、 ライフコースシミュ レーション ③可視化	未定	生命保険のデータは 管理体制が複雑なた め、分析データの 入手、システム 構築が非常に難 しい。データを 活用したビジネ スモデルの構築 と実施体制の 構築が課題。
28	I社	重工業	2017-現在	----	千万円台+ 既存リソース	プラント制御	工場や工場 内のエンジン	工場やエンジ ンのデータ分析・ 制御	常時 期間限定	①技術移転 ②共同研究	①センサデータ ②気象データ	×	センサからの データを収集 するシステム 経由	システム構築 高、運用低	業務用のセンサ データを収集する 初期システム構築 にコストがかかる が、運用コストは 極小	相手先企業	①PLSA(確率的 潜在意味解析) ②ベイジアン ネットワーク ③線形・非線形 回帰	①変動予測 ②制御	未定	工場プラントの 複雑システムの 予測・制御は 産業競争力上 も重要な課題 のため、比較 的高いコスト 負担にも耐え られる。