パーソナルデータはなぜ「集めるほど価値が増す」のか

一般財団法人日本情報経済社会推進協会　電子情報利活用研究部
次長　　松下　尚史

1. はじめに

デジタル経済において、パーソナルデータの価値をめぐる議論には繰り返し現れる構図がある。特に注目される構図の一つが、個人レベルでのデータ価値の小ささと、それを集約した際に生じる巨額の価値との間のギャップである。この個人的価値と集合的価値のギャップという論点については、姉妹稿「パーソナルデータの価値とは何か～経済学から見た性質・測定・分配～」[1]において整理している。

本稿が問うのは、そのギャップがなぜ生じるのか、である。言い換えれば、「パーソナルデータは単体では価値が低く、集合・連携によって価値が生まれる」という命題を、経済学的・統計学的に解き明かすことを目的とする。

この命題は、データビジネスの実務に携わる者にとって直感的には自明に映るかもしれない。しかし、なぜそうなのかを説明し、具体的な事例によって検証することには独立した意義がある。なぜなら、この命題が確立されてはじめて、「いかにして集合・連携を実現するか」という制度設計の問いが正当化されるからである。

このため本稿では、まず単体データが価値的限界を持つ理由を整理したうえで、集合・連携によって価値が生まれる経済学的メカニズムを論じ、国内外の事例によってその構造を検証する。なお本稿は、姉妹稿「プライバシー強化技術（PETs）と個人情報保護法～データ連携の制度的課題～」[2]と合わせて読むことで、より体系的な理解が得られる。

2. 単体データの価値的限界

2-1. 文脈なしには解釈できない

単一の観測値は文脈なしにはほとんど意味を持たない。これはデータ全般に共通する性質であるが、パーソナルデータにおいてはとりわけ顕著である。ある個人が特定の商品を購入したという事実は、それ単体では、その人物の嗜好・所得・ライフステージ・購買意図のいずれについても何も語らない。購買の背景を解釈するためには、同一人物の過去の行動履歴、あるいは類似した属性を持つ他者の行動との比較が不可欠である。

1 「パーソナルデータの価値とは何か～経済学から見た性質・測定・分配～」
2 「プライバシー強化技術（PETs）と個人情報保護法～データ連携の制度的課題～」

単体データはその保有者（個人）と観察者（事業者）の双方にとって、解釈の根拠を欠いた状態にある。価値を持つのはデータそのものではなく、「データが他のデータと接続されたときに生じるパターンの認識」である。

2-2. 閾値性——分析が成立するための最低条件

統計的な分析を行うためには、一定以上のサンプルサイズが必要である。この「一定以上」という条件は、分析の種類と対象の複雑さによって変わるが、いずれの場合も無視できない下限値、すなわち閾値が存在する。

この閾値問題を理解するうえで有用な概念が「次元の呪い（curse of dimensionality）[3]」である。これは、分析に用いる変数（次元）の数が増えるにつれて、データの密度が急激に低下し、意味ある分析に必要なサンプルサイズが指数的に増大するという現象を指す。たとえば、年齢・性別・居住地域という三つの変数でセグメントを切るだけでも、その組み合わせは相当数に上り、各セグメントに統計的に有意な人数を確保しようとすれば、全体のサンプルは一定規模以上でなければならない。実際のパーソナルデータ分析では変数の数がさらに多く、小規模なデータセットでは分析の粒度が粗くなるか、そもそも分析が成立しない。具体的には、「30代・女性・東京都在住・過去1か月以内に購入あり」という4条件でターゲットを絞るだけで、中小企業が保有する数千件規模の顧客データでは該当者が一桁になることも珍しくなく、そこから傾向を読み取る分析はそもそも成立しない。条件（変数）を増やせば増やすほど、意味ある分析に必要なデータ量は急速に膨らむ。これが「次元の呪い」の実務的な意味である。

中小企業が単独で保有する顧客データが、精度の高いパーソナライゼーションや需要予測に活用しにくい理由の一つはここにある。保有するデータ量が閾値を下回っている場合、そのデータは潜在的な情報を内包していても、それを引き出す分析が成立しない。データは存在するが価値が顕在化しない状態、いわば「眠った価値」にとどまる。

2-3. 複製しても価値は増えない

データは原価ゼロで複製できる「非競合的な財」[4]であるが、同一データをいくら複製しても分析の精度は向上しない。これはデータ全般に共通する性質である。1万人分のデータを二つのシステムにコピーしても、利用可能な情報量は依然として1万人分にとどまる。価値が増えるのは、複製によって量を増やすことではなく、異なる個人・異なる場面のデータを新たに加えることによって、母集団の代表性と多様性が増すときである。

パーソナルデータにおいてはこの点がとりわけ重要である。行動・属性・選好といった多面的な情報を対象とするパーソナルデータは、「量的拡大」と「多様性の拡大」をともに必要とするが、これは単一事業者の内部に閉じた形では達成が難しい。ここに、複数事業者間でのデータ連携を必要とする構造的な理由がある。

---

3 「次元の呪い（curse of dimensionality）」はR. E. Bellmanが1957年の著作『Dynamic Programming』で提唱し、1961年の『Adaptive Control Processes』でさらに展開した概念（Bellman, R.E., Adaptive Control Processes: A Guided Tour, Princeton University Press, 1961）。

4 「非競合的な財」とは、ある者が使用しても他者の使用可能性が減少しない財を指す。デジタルデータは原価ゼロで複製可能であるため非競合的であり、この点で石油や土地のような競合財（一者の使用が他者の使用を排除する）とは根本的に異なる。なお、この性質はデータ全般に共通するものであり、パーソナルデータに限られるものではない。

3. 集合・連携によって価値が生まれるメカニズム

3-1. 規模の経済――量の拡大による精度向上

規模の経済とは、規模の拡大に伴って単位あたりの効用・価値が高まる現象を指す。データに関する規模の経済には、同一データの大量処理による単位コストの低減と、学習データの拡大による分析精度の向上という二つの側面があり、本節は後者に焦点を当てる。前者については姉妹稿「パーソナルデータの価値とは何か～経済学から見た性質・測定・分配～」を参照されたい。

機械学習モデルの精度は、学習に用いるデータ量に依存する。データ量が少ない段階では精度の向上は緩やかだが、一定の閾値を超えると精度が急速に改善し、やがて漸近的な上限へと収束するという特性がある。データ量の増加に伴い精度が向上するこの関係は、「学習曲線（learning curve）」として知られており、特に自然言語処理や画像認識の領域で実証的に確認されている。

パーソナルデータを扱うレコメンデーションや需要予測においても、同様の性質が観察される。単一事業者が保有するパーソナルデータだけでは閾値に達しない場合でも、複数事業者のパーソナルデータを集合させることで学習曲線の急上昇域に入り、実用的な精度を初めて達成できる。価値は量的な集合によって生まれる。

3-2. 範囲の経済――異種データの組み合わせによる新たなインサイト

規模の経済が量の問題であるとすれば、範囲の経済[5]は種類の問題である。異なる文脈や場面で収集されたパーソナルデータを組み合わせることで、単一ソースからは見えなかったパターンが顕在化する。

購買データと位置情報データを組み合わせれば、どの店舗でどのような商品が誰に購入されているかが見える。そこに医療・健康データが加われば、生活習慣と購買行動の相関が浮かび上がる。個別データは個別の事実を記録しているに過ぎないが、その組み合わせは単純な足し算を超えた洞察を生む。これが範囲の経済であり、異種のパーソナルデータが連携されることで初めて実現される価値である。

3-3. ネットワーク効果（外部性）――参加者増加による連携価値の逓増

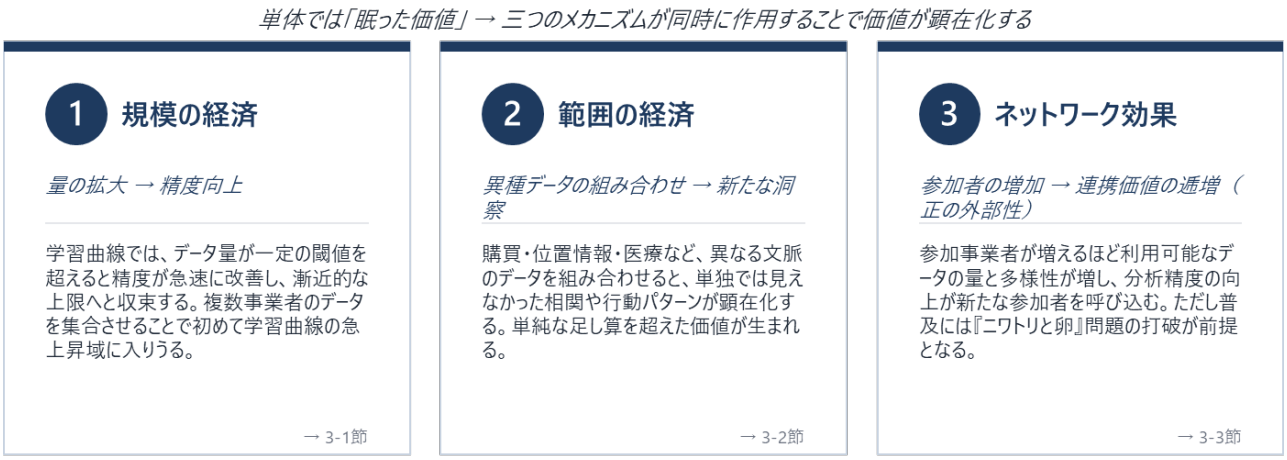
ネットワーク効果（外部性）とは、サービスの利用者が増えるほどそのサービスの価値が高まるという現象を指す。

データ連携基盤に参加する事業者が増えれば、利用可能なデータの量と多様性が増し、分析精度が向上する。精度の向上は連携の魅力を高め、新たな参加者を呼び込む。この正のフィードバック構造

5 範囲の経済とは、複数の財・サービスを組み合わせることで個別に扱う場合には生まれない付加価値が創出される現象。データの文脈では異種データの組み合わせによる新たなインサイト創出という形で現れる（基本的な定義は姉妹稿「パーソナルデータの価値とは何か～経済学から見た性質・測定・分配～」脚注2参照）。

のもとでは、連携基盤の価値は参加者数に比例するのではなく、より急勾配で増大する傾向がある[6]。

逆に言えば、参加者が少ない初期段階では価値が低く、投資回収の見通しが立ちにくい。これは「ニワトリと卵」問題として知られる参入障壁であり、データ連携基盤の普及を阻む構造的要因でもある。この問題を克服するためには、初期段階での参加インセンティブの設計や、信頼できる中立的な運営主体の存在が不可欠である。



図表 1 データ連携で価値が顕在化する三つのメカニズム

4. 事例による検証

4-1. 事例①：医療データ連携（国内）

日本の医療データ活用において、単一病院の電子カルテが保有する情報だけでは見えない知見が、複数機関のデータを統合することで初めて顕在化するという構造は、政策上も認識されてきた。

その代表的な実例が、厚生労働省が整備するNDB（レセプト情報・特定健診等情報データベース）である。NDBは全国の保険請求データを集積しており、2025年3月末時点で533件（オンサイト利用を含む）の研究に対してデータ提供が承諾され[7]、医薬品の副作用発生率や地域別の医療資源配分など、個別医療機関では把握不可能な統計的事実を明らかにしてきた。また、NDBと介護DB等の公的データベースを連結解析できる法制度も整備されており[8]、医療・介護の横断的な分析が可能になりつつある。

この事例が示すのは、単体では「個人の受診記録」に過ぎないデータが、数千万件規模で集積・統

---

6 ネットワーク効果の規模については、参加者数nのネットワークの価値がn^2に比例するとするメトカーフの法則（Metcalfe's Law）が広く参照されている（Metcalfe, B., "Metcalfe's Law after 40 Years of Ethernet" （Computer, Vol.46, No.12, pp.26–31, 2013））。ただし実証研究ではn^2よりも緩やかな成長を示す事例も報告されており、データ連携基盤においては参加者の質・データの多様性によっても増大の速度が左右される。

7 厚生労働省「匿名医療保険等関連情報の第三者提供の現状について（報告）」（第29回匿名医療情報等の提供に関する専門委員会、令和7年6月11日 資料2）。2025年3月末時点でNDBデータの第三者提供が承諾された研究は533件（うち7件は迅速提供。オンサイトリサーチセンター及びHICにおける提供を含む）。

8 NDBと介護DBの連結解析は、高齢者医療確保法・介護保険法の改正（令和2年10月1日施行）により法制化された。なお、内閣府健康・医療戦略推進事務局「改正次世代医療基盤法について（利活用編）」（2024年4月）は、2024年4月施行の改正次世代医療基盤法により新たに可能となった匿名加工医療情報と公的データベースとの連結解析について説明するものであり、あわせて参照されたい。

合されることで、医療政策の根拠となる集団レベルの知見へと変換されるという過程である。個々のレセプトデータに内在していた潜在的な価値は、集合によって初めて顕在化した。

なお、2022年10月に内閣に設置された「医療DX推進本部」は、複数医療機関で患者情報を共有する全国医療情報プラットフォームの構築を進めており[9]、データ連携による価値創出はさらに拡大する方向にある。

4-2. 事例②：購買・行動データの大規模連携活用（海外）

パーソナルデータの集合が価値を生む構造をより鮮明に示すのが、Amazonのレコメンデーションシステムである。Amazonは購買履歴のみならず、閲覧データ[10]を含む大規模な行動データを集積しており、これらが協調フィルタリングをはじめとする推薦アルゴリズムの精度を支えている。

Amazonは2003年に「アイテムベース協調フィルタリング（item-to-item collaborative filtering）」を発表した。この論文は2017年、IEEE Internet Computingの創刊20周年記念において「時代の試練に最もよく耐えた論文」として表彰されており[11]、レコメンデーション技術における画期的な成果として評価されている。

この技術の本質は、個々のユーザーの購買履歴や閲覧データを直接比較するのではなく、商品間の関係性（Aを購入した顧客がBも購入する傾向）を大規模なデータから抽出し、それをリアルタイムの推薦に活用するという点にある。このアルゴリズムは顧客数や商品数の規模に依存せず処理が可能になるよう設計されており[12]、数千万人規模の顧客基盤を持つプラットフォームとして機能する前提の下で開発された。

重要なのは、このシステムが生み出す価値の源泉が「データの集合規模」にあるという点である。既存のアルゴリズムではAmazonの数千万人規模の顧客と商品に対応できなかったため、同社は独自のアルゴリズムを開発するに至った[13]。これは裏を返せば、小規模なデータセットではこの水準のレコメンデーション精度が原理的に達成できないことを意味する。単一の中小規模事業者が個別に保有する顧客データでは、協調フィルタリングの「協調」が成立しないのである。

---

9 「医療DX推進本部」は2022年10月に内閣（本部長：内閣総理大臣）に設置された。本文で引用した出典は、厚生労働省「医療DXの更なる推進について」（2024年8月、第181回社会保障審議会医療保険部会資料）であり、全国医療情報プラットフォームの整備状況を示している。

10 Smith, B. and Linden, G., "Two Decades of Recommender Systems at Amazon.com"（IEEE Internet Computing, Vol.21, No.3, pp.12–18, May/June 2017） 同論文では、Amazon.comのレコメンデーションが2003年時点で既に「過去の購買と店内で閲覧したアイテム」に基づいていたことが述べられ（p.13）、また「行動ベースのアルゴリズム（購買・閲覧・評価を利用するもの）」が言及されている（p.17）。

11 Linden, G., Smith, B. and York, J., "Amazon.com recommendations: item-to-Item collaborative filtering"（IEEE Internet Computing, Vol.7, No.1, pp.76–80, January–February 2003） 同論文は2017年のIEEE Internet Computing創刊20周年記念において、同誌史上「時代の試練に最もよく耐えた論文（Test of Time Award）」として表彰された。

12 Linden, G., Smith, B. and York, J.（2003）前掲. 同論文において、アイテムベース協調フィルタリングのオンライン計算量が顧客数・商品数の規模から独立してスケールすることが示されており、数千万人規模の顧客基盤に対応可能な設計上の特徴として論じられている。

13 Amazon Science, "The history of Amazon's recommendation algorithm"（2021） 既存のユーザーベース協調フィルタリングがAmazonの顧客・商品規模に対応できなかったため、同社が独自のアイテムベースアルゴリズムを開発するに至った経緯が記述されている（Amazon社公式ブログによる社史的記述）。

比較軸	NDB（レセプト情報・特定健診等情報データベース）	Amazon（ECプラットフォーム）
領域	国内・公的領域	海外・民間領域
データの種類	保険請求データ（レセプト）、特定健診等情報	購買履歴・閲覧データ
連携・集約の形	厚生労働省が全国の保険請求データを集積。介護DB等との連結解析も法制度上可能。	プラットフォーム上の全ユーザー行動を集積。アイテムベース協調フィルタリングで商品間の関係性を抽出。
顕在化した価値	個別医療機関では把握不可能な統計的事実（医薬品副作用率、地域別医療資源配分等）。2025年3月末時点で533件の研究にデータ提供。	数千万人規模を前提とした高精度なレコメンデーション。2003年発表のアルゴリズムは2017年にIEEE Internet Computing 20周年記念で表彰。

図表 2 NDBとAmazonにみるパーソナルデータ連携の価値創出

5. 帰結と展望

本稿の議論を整理すれば、次のとおりである。

第一に、パーソナルデータは単体では価値が低い。文脈なしには解釈できず、統計的分析の閾値を下回り、複製によっても価値は増えない。この限界は、データの質の問題ではなく、データが単体である以上避けがたい構造的な制約である。

第二に、集合・連携によって価値が生まれる。規模の経済（量の拡大による精度向上）、範囲の経済（異種データの組み合わせによる新たなインサイト）、ネットワーク効果（外部性）（参加者増加による価値の逓増）という三つのメカニズムが、この過程を説明する。国内の医療データ連携と海外の購買・行動データ活用という対照的な事例は、いずれもこのメカニズムの実証を提供している。

第三に、データビジネスにおいて競争力のある価値を生み出すためには、単一事業者の内部に閉じたデータ活用ではなく、複数事業者間でのデータ連携を可能にする環境の整備が不可欠である。本稿で取り上げたNDBは公的部門が中央集権的に、Amazonは巨大プラットフォーム事業者が独占的に集約することで、いずれも集合による価値創出のメカニズムを実証している。しかし、こうした集約モデルを取りえない中小事業者や、業界横断的な価値創出を志向する主体にとっては、複数事業者間でのデータ連携基盤こそが、同等のメカニズムを享受する制度的経路となる。

単体では眠ったままの価値が、集合によって初めて顕在化する。そのメカニズムを確立することが、集合・連携をいかに実現するかという制度設計の問いを正当化する。

JIPDEC　電子情報利活用研究部　　次長　　松下　尚史

青山学院大学法学部卒業後、不動産業界を経て、2018年より現職。経済産業省、内閣府、個人情報保護委員会の受託事業に従事するほか、G空間関係のウェビナーなどにもパネリストとして登壇。その他、アーバンデータチャレンジ実行委員。

実施業務：

・自治体DXや自治体のオープンデータ利活用の推進
・プライバシー保護・個人情報保護に関する調査
・ID管理に関する海外動向調査
・準天頂衛星システムの普及啓発活動　など