

ビジネスデータサイエンス：ビジネスデータのモデリング

東京理科大学 経営学部 ビジネスエコノミクス学科 教授 て る い 照井 のぶひこ 伸彦

1. IoT 社会とビッグデータ

現代社会では、製品や機械に組み込まれたコンピュータがネットワークと繋がり、機器やモノが情報を発信します。この発信された情報は、ビジネス、健康、農業、高齢化社会など、さまざまな分野で活用され、特に「Society 5.0」における超スマート社会の実現に寄与することが期待されています。これは、サイバー空間とフィジカル空間の融合により、経済発展と社会的課題の両立を目指す社会です。先進諸国では、製造業からサービス業へのシフトが進んでいます。日本でも、サービス業が経済の中心となり、労働者の多くがこの分野で働いています。この変化により、企業はビッグデータを活用して新たな知識を発見し、新しいサービスを創出しています。特に、社会の多様なニーズに応えるためには、膨大なデータを適切に収集・解析し、個別化されたサービスを提供することが重要です。

現代の企業は、自社が保有するデータだけでなく、社会全体に存在するさまざまなデータを活用しています。例えば、気温や天候データを活用して商品の売上予測を行い、廃棄ロスや機会損失を減らすことができます。また、顧客の購買履歴を分析し、個別にサービスを提供することで顧客満足度を高め、企業のロイヤリティを強化することができます。

ビジネスにおける意思決定では、企業の経験や直感が重要ですが、近年では収集したデータを活用してより精度の高い意思決定を行うことが求められています。それに対して、近年急速に発展し応用が進んでいるベイズ統計やマルコフ連鎖モンテカルロ（MCMC）法などの計算技術を用いることで、企業は膨大なデータを基により正確な予測や意思決定が可能となります。

これらを背景に、本稿では著者らのビジネスデータサイエンスの研究成果を紹介します。

2. ビジネスにおける個別化対応の統計モデリング——価格の個別化

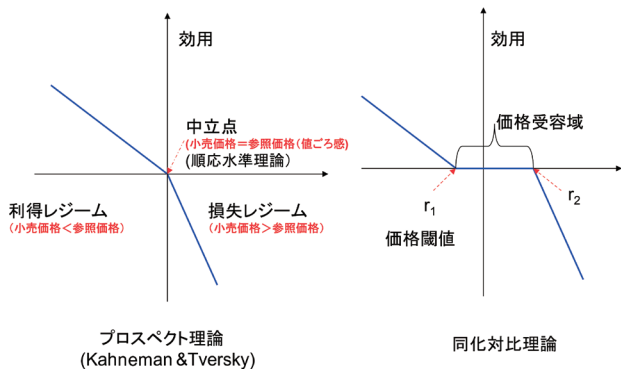
ビジネス、とりわけマーケティング分野では、消費

者の嗜好や購買行動を個別に理解し、適切な対応をすることが求められています。この「個別化」へのアプローチとして、階層ベイズモデルが有効です。階層ベイズモデルは、消費者ごとの異質性を考慮し、少ないデータでも安定した推測を可能にします。これにより、各消費者に合わせたマーケティング戦略が立てられます。階層ベイズモデルは、消費者の行動や市場反応を予測するために、消費者の異質性（個々の違い）を考慮した推定が可能であることで、マーケティング分野で広く応用されています。

具体的な応用例として、消費者の参照価格に基づく価格戦略の策定に関する研究を紹介します。

まず参照価格（RP: Reference Price）は、消費者が購買に先立って商品に対して持っている“値ごろ感”であり、この概念を取り入れた消費者行動の分析が数多く行われてきました。

【図1左】は、横軸が価格を表し、縦軸が消費者が商品のある価格で買うときに生じる効用の変化を表しています。原点は参照価格であり、原点より右側は店頭価格が参照価格より高い領域で、ここで消費者は割高感を感じることから“損失”領域、そして逆に原点の左側は割安感を生み出す“利得”領域を表しています。ここで原点は、利得に関して損とも得とも思わない中立な参照点です。これは認知心理学の枠組みにおいて、人がある判断をする場合に基準とする参照点であり、価格反応の場合は参照価格と呼ばれます。図中の二つの直線は、損と感じるか得と感じるかによって効用がどのように変化するかを表し、これらの直線の傾きが同じでないことは、損と感じたときの効用の変化分と得と感じたときの効用の変化分が同じではない非対称市場反応を表しています。さらには損失領域での直線の傾きがより急で大きいことから、消費者の損失回避行動を示しており、これは提唱者の Kahneman and Tversky（1979）によるプロスペクト理論を背景にしています。彼らは、消費者の利得・損失と効用の関係を提示する際に、経済学の基本的仮定である合理的消費者像とは異なり、消費者は必ずしも合理的に行動するとは限らないことを実験によって示し、実験経



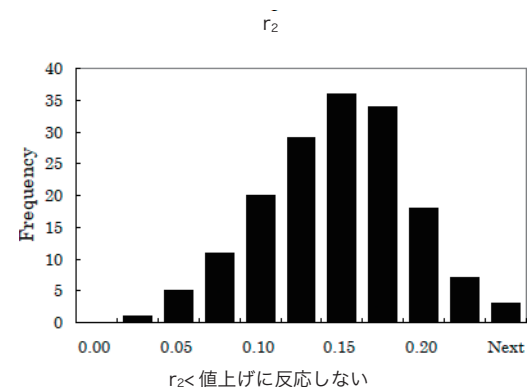
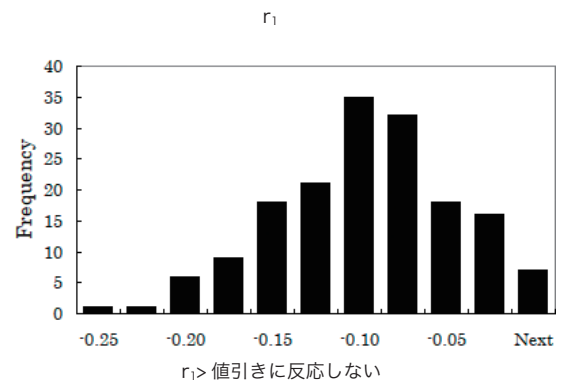
【図 1】 参照価格とプロスペクト理論

経済学という分野を切り開いたことでノーベル経済学賞を授賞しています。

さらに【図 1 右】は、参照価格は 1 点ではなく幅を持つという理論により、消費者には価格変化に反応しない「価格受容域」が存在することが示唆され、それが下限 (r_1)、および上限 (r_2) の「価格閾値」によって構成されることを示しています。下限の r_1 は、企業が商品の価格をディスカウントした場合でも r_1 以上に下げないと消費者は値下げと感じず、無駄なプロモーションとなる。逆に、値上げの場合でも、上限の価格閾値 r_2 を超えない限り、消費者は値上げと感じないため、純粋な利益増となる。これらの価格閾値の値は人により異なると考えられ、これを消費者の購買履歴データから消費者ごとに個別に推定することが問題となります。

Terui and Dahana (2006a) は、これを表現する区分的線形確率効用関数によりモデル化を行い、インスタントコーヒーの購買履歴データを用いて r_1 、 r_2 を顧客ごとに推定し、その結果が【図 2】で示されています。

図は顧客ごとに得られた閾値パラメータ推定値のヒストグラムです。いずれも非対称で歪んだ分布をしており、下限閾値 r_{1h} の平均は -0.113 であり、値下げ戦略の場合には 11.3% 以上のディスカウントを行う必要があること、また上限閾値 r_2 の平均は 0.138 であり、値上げの場合は、13.8% 以内の値上げであれば許容される可能性を示しています。もちろん、異質性の分布から数パーセントでも反応する顧客や 25% でも反応しない顧客がいること、値上げの場合も同様であり、ここに階層サイズモデルを用いた個別化アプローチの実装可能性があります。価格閾値の定義から、下限の価格閾値を越えないディスカウントは消費者に値下げを感じさせないため、売上げに影響は与えません。値下げ分が単純に損失として計上されることにな



【図 2】 価格閾値の分布

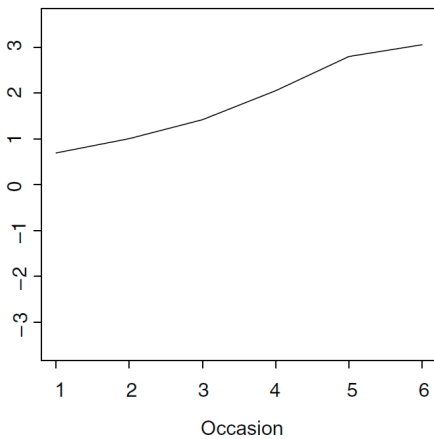
る。他方、値上げの場合でも、上限の価格閾値を越えない限り消費者は値上げを感じないことから、同様に売上げに影響を与えることはなく純粋な利益増となります。このように価格閾値の情報は、下限の閾値を越えないディスカウントから生じる損失を最小化し、上限の価格閾値を越えない値上げから生じる利得を最大化することに貢献し、Terui and Dahana (2006b) では、消費者ごとに異なる価格を提供する個別価格戦略は、効率的な価格付けを与えることが期待できることを示しました。

3. 消費経験による飽きと消費者バラエティシーキング行動のモデル化

次には、消費者の購買経験が商品への「飽き」を誘発し、その結果、多様な商品を購入する行動（バラエティシーキング行動）をモデル化した研究を紹介します。

顧客のブランド選択行動の分析で使われるモデルは、顧客が 1 つのブランドを選ぶ場合に限りて利用可能でしたが、実際の顧客は同じカテゴリ内で複数のブランドを同時に選んだり、1 つのブランドで複数の数量を購入することがよくあります。これを捉えるために、複数の選択肢を同時に選ぶ行動を説明する「直接効用モデル」が用いられます。

A: Panelist 99 (Increasing Trend)



【図 3】 飽きのスコアと要因

直接効用モデルでは、消費者の選択肢に対する効用（満足度）を数式で表現し、その効用が単一選択と複数選択の両方に対応できるようにします。また、消費者が複数のブランドを購入する行動を「飽きたから他のブランドを試す」という動機に関連づけて、飽きによる行動の変化もモデルに組み込んだモデルを Hasegawa, Terui and Allenby (2012) で提案しました。

さらにこのモデルを利用して、商品やブランドが店頭から撤去された場合の影響を金額で評価する「補償価値」の考え方も導入し、マーケティングの現場で商品ラインアップの価値を評価する方法を提案しました。

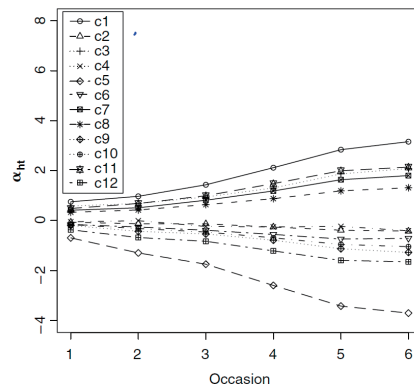
実際のデータ分析では、米国の大学の学生を被験者とした購買実験データを使用し、被験者の購買データをもとに飽きの度合いや消費者の行動変化を分析しました。【図 3】は実証分析の結果として、ある被験者 No. 99 について、左図は飽きスコアの時間変化を表わしており、購買経験が増えるにつれて飽きが増加していることを意味しています。また右図は、その飽きスコアを製品成分の空間に回帰した結果を表し、成分 c1, c3, c7, c8, c11（チーズ、ハーブ、およびコショウ）が飽きを増加させる要因であることを意味しています。この他、経済学における補償価値分析を適用することにより、当該カテゴリでは“品揃えの多さ”が重要であることを実証しています。

4. ネットワークとコミュニティの統計モデリング

最近、消費者は e-コマースだけでなく、実店舗での購買時にも SNS や口コミを参考にしているため、消費者同士や消費者と企業の間でのコミュニケーション

Figure 3
PARTWORTH DYNAMICS

A: Panelist 99 (Increasing Trend)



ンが重要になっています。したがって、これらの情報を活用して効果的なマーケティングを考える必要があります。社会ネットワークを分析する際には、ネットワーク情報だけでなく、人々がソーシャルメディアで発信するテキスト情報を加えることが重要です。これにより、ネットワーク内で密に関係が結びついている人々が持つ興味や関心に基づいた潜在的なコミュニティを特定できます。これらのコミュニティ

を識別するための統計モデルも提案されており、近年では、ソーシャルメディアやユーザー生成コンテンツ（UGC）を取り入れた新しいモデルが開発されています。

そこでは、ネットワーク情報だけでは捉えきれないコミュニティ構造を、テキスト情報も加えることでより正確に識別できることがメリットです。例えば、同じ学校の生徒が形成するネットワークでは、フォロー/フォロワーの関係だけでなく、音楽や読書、スポーツなどの共通の趣味を持つグループも見つけることができ、これがビジネスにおいて有用なセグメントになります。

Igarashi and Terui (2020) の研究では、これらネットワークとテキスト情報を同時に利用して、コミュニティを発見し、構造を可視化する方法を提案しています。実証分析では任天堂アカウントを中心とするネットワーク（2018年5月1日におけるフォロー関係）タイムライン上に投稿された Tweets（2017年9月1日から2018年2月28日）を用いています。モデルの詳細は割愛しますが、分析結果の一部は、【図 4】に示されており、これによって我々のモデルで何が分かるのかを説明します。図では6人のネットワーク構造が示されています。ネットワークデータから各人のフォロー/フォロワーの関係は記述できますが、それに各人がコミュニケーションしているテキストデータを上乗せすることによって、潜在的なサブコミュニティを抽出できることを図は示しています。例えば、#95, #804, #2241, #3476 の4人は、音楽（トピック4）を共有したコミュニティであり、#95, #336 はスポーツ（トピック9）、#2241, #3476 はビジネス（トピック6, 7）、#336, #1683, #2241 はアニメーション

(トピック1), ゲーム(トピック2)をそれぞれ共有するコミュニティであると推測できます。

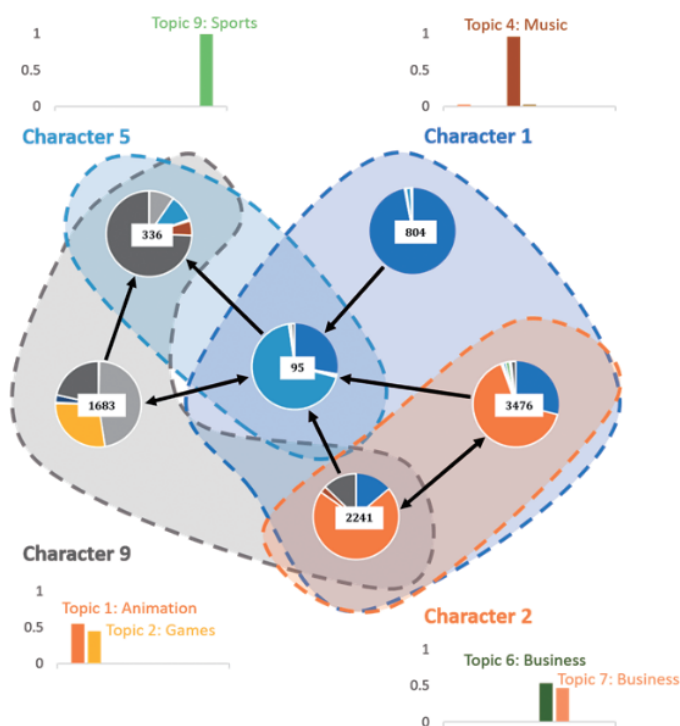
このようにネットワークとテキストデータを同時に使うことで直接は見えないコミュニティの関係が見えるようになります。

5. おわりに

データ分析の伝統的手法は統計学であり、さまざまな分野に応用されてきました。特に文系分野では、経済学において経済理論の検証や経済予測のための計量経済学がビジネス分野に応用され、経営学部や商学部でもその重要性が認識されてきています。現在では、社会学、言語学、政治学、歴史学、教育学など、統計学を活用する研究や教育が増えています。

一方、統計学は保守的な側面を持ち、データは貴重であるという認識のもと、緻密な計画に基づいて実験や観測で得られた無作為標本を丹念に分析する方法が取られてきました。しかし、IoT社会では、さまざまなものがネットを通じてつながり、無計画かつ無意識的にデータが生成されます。このような状況では、無作為標本を前提にした統計学の理論が適用できないことが多いです。また、ソーシャルネットワークでの人々のコミュニケーションに関連するテキストや画像などの非構造データは、機械学習分野で新たな手法が開発されてきました。ビジネス分野は、このような社会的変化に迅速に対応する必要があり、新しいデータ分析手法のいち早い導入が求められます。

AIの深層学習技術は大きな進展を遂げていますが、現時点では予測性能に重点を置いたモデル開発が中心で、モデルの結果を戦略策定に結びつける議論には限界があります。ビジネス計画を立てるためには、因果関係に基づく計量モデル $Y=f(X)$ が必要です。ビッグデータは非常に大量の情報を含んでいるものの、その情報の質は浅いといえます。また自然現象とは異なり、社会経済システムにおいて物理定数が安定的に存在するわけではありません。したがって、データに対して複雑な事象に潜む構造をモデリングする必要があります。これにより、先験的な知識(経験, 理論, 合意)と実際のデータを統合し、複雑な事象を理解し、予測や制御に生かすことが可能となります。このような背景から、ビジネスデータにはベイズ統計に基づく統計モデリングが有効であり、ベイズ統計が重要な基盤技術となりま



【図4】潜在コミュニティの可視化

す。

また経済学分野でも、実証分析手法が精緻化・高度化しており、特に政策評価を因果推論で精密に検証する方法や、実証産業組織論のミクロ動学モデルによる構造推定などが注目されています。今後は、大規模データに対応する手法が進展し、ビジネス分野への応用が進むことが期待されます。さらに機械学習の深層学習技術や強化学習モデルは、大規模なビジネスデータへの適用が進むと思いますが、経済学、マーケティング、社会学など、関連する分野の知見を取り入れ、新しい分析手法を組み合わせ、課題解決や戦略策定を行うことがますます重要になります。

【参考文献】

- Nobuhiko Terui and Wirawan Dony Dahana (2006a), "Estimating Heterogeneous Price Thresholds," *Marketing Science*, 25 (4), 384-391.
- Nobuhiko Terui and Wirawan Dony Dahana (2006b), "Price Customization Using Price Thresholds Estimated from Scanner Panel Data," *Journal of Interactive Marketing*, 20 (3), 58-70.
- Shohei Hasegawa, Nobuhiko Terui and Greg Allembly (2012), "Dynamic Brand Satiation," *Journal of Marketing Research*, vol. XLIX, 842-853.
- Mirai Igarashi and Nobuhiko Terui (2020), Characterization of topic-based online communities by combining network data and user generated content, *Statistics and Computing*, 30, 1309-1324.
- Kahneman, D. and A. Tversky (1979), Prospect theory: An analysis of decision under risk, *Econometrica*, 47, 263-291.