

リアリティ・チェック：ビッグデータと予測データモデルの使用

人間は昔から、過去に関する十分な量のデータを機械（それがどんな機械であれ）に入力すると、未来に起こることを機械が予測できるという想像の虜になってきました。1984年に放映されたアメリカのテレビアニメーション作品『ザ・トランスフォーマー』では、オートボットたちが母星のサイバトロンのために見つけにくいスペースブリッジを探す場面があります。人間の友人で、典型的な神童のチップ・チェースは、オートボットたちに次のように言います。

スペースブリッジが最後に出現した時の全データをテレトラン1 [オートボットたちのスーパーコンピューター]に入力すれば、ブリッジが次に出現する地点を予測できるかもしれない¹。

現実がアニメーションと同じくらい単純であれば、その通りになるでしょう。

大量のデータを使用して、これから起こることを予想するという考えは目新しくありません。しかし、予測モデルの機能、精度、手頃感、データ全般のボリュームと可用性およびクラウドの計算能力が増大していることでますます高まっています。このような予測モデルの現実の使用に関して、これまで仮説のレベルで行われていたような厳しい疑問や問題に、より多くの組織が直面することになるでしょう。

多くの事業体にとって...予測データモデリングへの投資は、顧客満足度と効率性の向上はもちろん、人命救助にも役立つことさえある追求する価値のあることです。

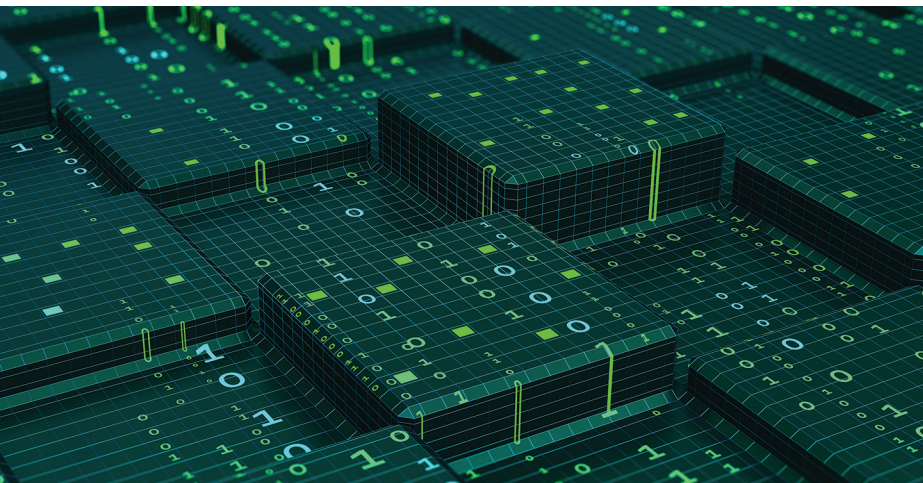
すべての組織が、未使用のデータという形で真正銘の水晶玉を抱えているという考えは、幻想にすぎません。しかしながら、幅広い業種にまたがる多くの事業体にとって、予測データモデリングへの投資は、顧客満足度と効率性の向上はもちろん、人命救助にも役立つことさえある追求する価値のあることです。但し、こうした価値を実現するためには、組織は、地に足の付いた、時に面倒な現実に対処しなければなりません。

データセキュリティおよび完全性

サイバーセキュリティは、すべての組織のリスク登録簿において、常に一二を争う位置に記載されています。データセキュリティおよび完全性に対する脅威は、組織の内外で高まっています。例えば、組織内では、データガバナンスの複雑さ、優先順位の競合、軽視などが見られ、組織外からでは、悪意のある行動、サードパーティに対するコントロールの欠如などが見られます。同時に、データ漏洩に伴う潜在的な金銭的損害、風評被害、法的/規制上の責任などのリスクも増大しています。

KEVIN M. ALVERO | CISA、CDPSE、CFE

Nielsen Companyの内部監査、コンプライアンス、ガバナンス担当シニアバイスプレジデント。事業体のグローバルメディア製品およびサービスを広くカバーする、内部品質監査プログラムと業界コンプライアンスイニシアティブの先頭に立って指揮しています。



2022年のProtivitiのレポートでは、次のような状況が報告されています。

IT監査チームや、その他の部門（例えば、法務、コンプライアンス、IT）は、新しいデータプライバシー規則およびデータセキュリティ規則のほか、組織のデータ管理や技術関連活動に大きな影響を与える法律や規制のコンプライアンス要件の変化に対応するために必死である²。

特に機密性の高いデータに関しては、組織がこれらのデータから得る価値が所有のリスクに見合うかどうかという根本的な疑問があります。この疑問に答えるためには、予測モデリングが組織のミッション、戦略、コアバリューをどのように支援することが期待されるかについて、経営幹部が明確な理解を共有しなければなりません。同時に、予測モデリングの価値を組織に分け与える機能は、モデルに入力されるデータの品質と完全性に直接、関連します。したがって、組織を過剰なリスクにさらすようなやり方で、組織が予測データモデリングの将来的な利益に手を出したり、飛びついたりしないように、強い目的意識とデータセキュリティおよび完全性に対する取り組みをトップダウンで整えなければなりません。

偏見、プライバシー、その他の倫理的懸念事項

一般市民の間では、（同意の有無を問わず）組織が収集する個人データが、プライバシー権や公平・公正な処遇を受ける権利を侵害する方法で 사용되는ことへの懸念がますます高まっています。人の思考や行動を予測しすぎると、相手はとても不快な気分になり、ブランドの印象や信用度が下がるかもしれません。そうした懸念を伝えたことで、政府でも市場でも変化が起きました。2022年の『Harvard Business Review』の記事には、次のような記述があります。

これまで、企業はできる限り大量のデータを収集してきたが...多くの場合、顧客側は何が行われているのか理解していなかった。ところが、コントロールが顧客側へシフトするのに伴い、同意を得て収集されたデータは、やがて最も価値のあるデータとなる。なぜなら、企業はそのデータに基づいてしか行動することが許されないからである³。

プライバシーへの注目度の高まりに加え、高度なデータ分析を活用する組織が、人々を公正かつ公平に取り扱っていることを確認するために、透明性を求める声

経営陣は、利用可能なデータが自分たちが行っている予測に関連するかどうか、また、そのデータをモデリングで取得、アクセス、保存、取り入れることのリスクとコストが釣り合うかどうかを判断できなければなりません。

も大きくなっています。特に、透明性に対する懸念が関連するのは、クラスタリングと呼ばれる、データ（およびデータが表す人間）を共通の属性に基づいて様々なグループに分類する予測データモデリングの一種のデータに対してです。ターゲティング広告以外にも、次のような使用事例があります。

...この予測データモデリング技法を使用するその他の事例には、融資応募者を融資の属性に基づいて「スマートバケット」にグループ分けしたり、都市の犯罪率の高い地区を特定したり、SaaS [Software-as-a-Service] の顧客データをグループ化してグローバルな使用パターンを特定するベンチマーキングなどが含まれます⁴。

一見すると、これは直感的かつ合理的に見えます。広告の目的が、人々が欲しいと思う製品やサービスを知らせることであるならば、その人々の情報を使って、関連する製品を提案する確率を高めることは、理にかなっていると言えるでしょう。破壊的な事象が発生する既知の歴史的可能性を、その事象に備えた保険料を顧客がどれだけ支払わなければならないかの決定に適用することについても、同じことが言えます。

しかしながら、特定の人々の集団が特定の事柄を行うことを希望することができる、希望すべき、または希望する可能性があるかどうかについて、（たとえ十分な裏付けがあっても）組織が判断するとき、組織は、慎重なリスク管理や顧客体験の有益な調整と差別との境界線を越えてしまう危険性があります。データ経済は、「業界の慣行を法律家や一般市民から隠すために作られた"デジタルカーテン"を中心に構成されていた.. [しかし]そのカーテンは取り除かれた」⁵。したがって、組織は、モデルを活用した意思決定プロセスがカーテンから差し込む直射日光（衆目）にどれだけ耐えられるかを理解しなければなりません。さらに、組織の予測データモデリングがどのようにそのコアバリューと整合するかを前向きに検討することも重要です。



LOOKING FOR MORE?

- Read Data Science Fundamentals Study Guide. www.isaca.org/credentialing/itca/data-science-fundamentals-certificate
- Learn more about, discuss and collaborate on information and cybersecurity in ISACA's Online Forums. <https://engage.isaca.org/onlineforums>

データの関連性

予測モデリングは、もう一つの一般的な使用例です。組織はこれまで、需要を予測するために過去のデータを使用していましたが、計算に含まれる情報の包括性と即時性により、データの関連性について新たなレベルの精査が必要になっています。これまで、事業体は取引量、収益データ、顧客ロイヤリティプログラムデータ、来店者の動向統計、一部の人口統計データなどを持っていましたが、今では、リアルタイムトレンド分析を用いてより幅広い種類の変数を計算に取り入れることができるようになりました。したがって、根本的な問題は、より多くのデータを必要とするということから、どこで線引きするかを判定することにシフトしました。経営陣は、利用可能なデータが自分たちが行っている予測に関連するかどうか、また、そのデータをモデリングで取得、アクセス、保存、取り入れることが予測の正確さを高める可能性があることを前提として、そのリスクとコストが釣り合うかどうかを判断できなければなりません。

組織のリーダーは、予測モデリング戦略を実行するために必要なスキルを有する人材がいるかどうか、あるいは獲得できるかどうかを真剣に検討しなければなりません。

混乱

その他にも経営陣にとって重要なのは、予測モデルを責任を持って効果的に活用するために必要な、継続的なモニタリングへの取り組みを完全に理解することです。予測モデルは、単に導入して放置しておくだけでは、バイアスがかかり、意思決定が損なわれ、ビジネスがリスクにさらされる可能性があります。

市場や、それよりも広い世界での変化がデータモデルを破壊しないことを確実にするには、警戒心が必要なことはもちろんとして、予測モデルの開発および使用を計画する早い段階で、レジリエンス（回復力）を前もって考慮しておく必要もあります。要するに、最終的に予測モデルが実行される環境は、

開発時には予測できなかった方向へ配置後に変化することが多いことを仮定して、予測モデルを構築する必要があります。これらは、予測モデリングに着手する前に経営陣が理解しなければならない投資利益率(ROI)の方程式での重要なファクターです。

ROI

特定の産業の組織にとって、高度なデータモデリングアルゴリズムへの投資に価値があることは確実でしょう。しかし、組織は、そのROIを定量化することもできなければなりません。ROIの予想精度を上げることができれば、以下のように定量化できる利益が数多くあります。

- 廃棄物の減少によるコスト削減
- 非効率性および無駄な手続の削減
- コストや危険な遅延の削減
- エラーの減少による品質の向上

例えば、緊急対応や災害への備えに関しては、高精度な予測モデルはそのまま救われる命の多さにつながります。しかしながら、他の状況では、コストと価値の比較図はあまり明瞭ではありません。例えば、タクシーやライドシェアリングサービスは、膨大な量の消費者および市場のデータを分析し、見込顧客を呼び込む事象の発生タイミングとその正確な数を（できる限り正確に）予測する機能があれば、大きな利益を得られます。他方、在庫水順を管理している実店舗は、過去の売上データを追跡するだけで十分に需要を予測できるかもしれないため、高度な予測モデルに投資しても、コストに釣り合わない可能性があります。

さらに、より適切な意思決定や行動を促す前向きな体験を求める顧客の需要は、「必ずしも予測可能なソリューションを提案するとは限りません」⁶。

繰り返しになりますが、経営陣は、予測モデルから高い正確さと洞察を得ることで、組織がどれほどの利益を得られるかを理解し、予測技術に投資する前に組織の利益を定量化できる必要があります。

組織的スキルと専門知識

2022年から2025年までに、AIに関連する雇用が9,700万件も創出されると予想されています⁷。「AIにはすべての産業を変革する力があります...しかし、企業は今なお、インテリジェントマシンを生み出

し、教育し、それと協力して働くために必要なスキルを持つ従業員を見つけることに取り組み続けています」⁸。

報道によると、分野における有能な人材を求める需要は、急激に供給を追い越していますが、予測モデリングに関する組織的スキルと専門知識の問題は、厳密には雇用の問題ではありません。マサチューセッツ工科大学(MIT) (米国マサチューセッツ州ケンブリッジ) 情報システム研究センターは、次のような研究報告をしています。

効果的なAIプログラムを作成することは、適切なAIシステムを構築して終わりではない。そうしたプログラムは、組織に統合する必要もある。また、利害関係者（特に従業員と顧客）は、AIプログラムが正確かつ信頼できると信用する必要がある⁹。

これが、事業体規模のAI説明力構築のケースである、と研究者は結論付けています¹⁰。

組織のリーダーは、予測モデリング戦略を実行するために必要なスキルを有する人材がいるかどうか、あるいは獲得できるかどうかを真剣に検討しなければなりません。また、予測モデリングの潜在的な効果をフルに発揮するために、組織内のリテラシーレベルを上げることができるかどうか判断しなければなりません。

結論

潜在的な投資機会を探ったことがある人なら、過去の実績は将来の結果を示唆するものではないと痛感しているはずでしょう。それにもかかわらず、多くの組織において、各顧客のニーズをこれまで以上に正確に予測し、市場全体の変化を予見する力を得ることは、高度な予測データモデルに伴う未知の事柄と潜在的な落とし穴に取り組む価値あるものです。

調査により、以下の事実が判明しています。

大企業の多くは、すでに顧客データを巡る内部の緊張にさらされている...現在のIT予算の約90%は、社内の複雑な状況を管理するためだけに費やされ、生産性や顧客体験を向上させるデータイノベーションに実際に使われるお金はごくわずかである¹¹。

こうしたタイプの機能不全は、どれほどアルゴリズムが高度になろうとも克服できません。むしろ、予測モデルが強固な基盤に基づいて構築されることを確実にするには、明確な戦略と整合性の取れた目標への取り組みが求められます。

後注

- 1 *The Transformers*, Season 1, Episode 6, "Divide and Conquer," directed by John Walker, written by Donald F. Cohort, 20 October 1984, syndicated
- 2 ISACA® and Protiviti, *IT Audit Perspectives on Today's Top Technology Risks*, USA, 2022, www.isaca.org/it-audit-2022
- 3 Rahnama, H.; A. Pentland; "The New Rules of Data Privacy," *Harvard Business Review*, 25 February 2022, <https://hbr.org/2022/02/the-new-rules-of-data-privacy>
- 4 Insightsoftware, "Top Five Predictive Analytics Models and Algorithms," 1 January 2022, <https://insightsoftware.com/blog/top-5-predictive-analytics-models-and-algorithms/>
- 5 *Op cit* Rahnama and Pentland
- 6 Blanchard, B., et al.; "Predictive Modeling and Influencing Customer Behavior," Microsoft, <https://docs.microsoft.com/en-us/azure/cloud-adoption-framework/innovate/considerations/predict>
- 7 Marr, B.; "What Are the Most In-Demand AI Skills?" *Forbes*, 13 June 2022, <https://www.forbes.com/sites/bernardmarr/2022/06/13/what-are-the-most-in-demand-ai-skills/?sh=4682e7b3249c>
- 8 *Ibid.*
- 9 Brown, S.; "Why Companies Need Artificial Intelligence Explainability," Massachusetts Institute of Technology (MIT), Swan School of Management, Cambridge, Massachusetts, USA, 21 September 2022, <https://mitsloan.mit.edu/ideas-made-to-matter/why-companies-need-artificial-intelligence-explainability>
- 10 *Ibid.*
- 11 *Op cit* Rahnama and Pentland