

著者: Bora Beran, PhD

R のパワーとビジュアル 分析

高度な統計とデータビジュアライゼ
ーションとの連携

データ科学者や統計学者であれば、「R」というシンプルな名前の効果的な分析ツールを知っています。Rとはオープンソースの無料プログラミング言語であり、統計解析用の環境のことです。解析用プログラミング言語の中でも最も急速に成長している言語の1つであり、ほとんどのデータ科学者が選択するツールになりつつあります。実際にRは、SPSS StatisticsやSASを抜いて、ほとんどのデータ科学者が使用するツールとなっており、現在では全データ科学者のほぼ半数が使用しています。^{1 2}

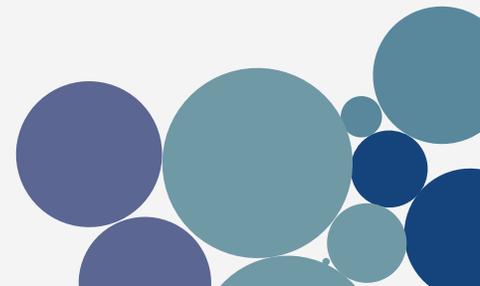
統計科学者、データ科学者、アナリストはビジネスに欠かせない存在になりつつあり、今日の膨大な量のデータをより分けて、新たな有益な分析情報を見つけ出す任務を担っています。複雑なパターンや関係性を特定する場合でも、データを使用して将来的なアクティビティを予測する場合であっても、これらの分析情報を得るには、統計モデルの構築と、結果の分析や共有といった能力が欠かせません。これらはまさに TableauとRの統合が得意とする分野です。

統計解析の「負荷のかかる仕事」はRが処理し、Tableauでは、統計結果が美しいダッシュボードに変換されるため、シンプルなドラッグアンドドロップ環境で結果を調査し、視覚化することができます。Tableau Server を利用することにより、これらのビジュアライゼーションをビジネス全体で安全に共有でき、データ科学者が行うような大変な作業を、ビジネスユーザーが手軽に行えるようになります。高度な分析結果を簡単に閲覧し、やり取りできるようになれば、ビジネス全体の意思決定を推進する上で、こうした分析がより重要な役割を担うことができます。

TableauとRを連携させることで、両者の得意分野を最大限活用することができ、今日のますます複雑になるビジネスの課題を解決するために、データをよりスピーディーに、そして、より簡単に利用することができます。

目次

雑音をより分け、お客様の声を見つけ出す	3
意思決定プロセスを自動化する	4
大規模で複雑なデータセットを理解する	5
偶然ではないことを証明する	5
著者について	6



雑音をより分け、お客様の声を見つけ出す

センチメント（感情）の分析

Twitter、Facebook、YouTube、Pinterest、Amazon などの人気の高いサイトで交わされる、コメント、レビュー、ディスカッションといったネット上の意見は、お客様との関係をなんとか確立したいと願っている企業にとって、まさに宝の山です。しかし、隠された宝を見つけるには、さまざまな雑音を排除していく必要があります。製品開発から評判管理まで、組織に必要とされるのは、内容を把握し、行動を起こすために、ネット上のやり取りを体系的により分けることです。そのために、多くの企業がセンチメント（感情）の分析を行うようになっています。

R には簡単にテキストを解析し、センチメントを判断できるツールが用意されていますが、ユーザーがその結果を使用して意思決定を行うには、結果を閲覧するための手段が必要になります。RをTableau と統合することにより、結果をビジュアライゼーションに変換し、プラスやマイナスのコメントに色付けし、分かりやすく表示することができます。

The screenshot shows a Tableau dashboard with a table of sentiment analysis results. The interface includes a 'Columns' shelf with 'CommentText' and a 'Rows' shelf. The 'Marks' card is set to 'Automatic'. The 'Sentiment' field is selected in the 'Marks' card, and a legend for 'AGG(Sentiment)' is visible at the bottom left, showing 'negative' (red), 'neutral' (grey), and 'positive' (green).

CommentText	Sentiment
Crisp and bright screen, well worth the money.	positive
Customer service was unresponsive. They wouldn't allow me to ret..	negative
Do not buy this product. It only lasts 2 weeks. It is garbage. And th..	neutral
Don't waste your time on this product. One word. Bad.	negative
I had a pleasant experience. Responsive customer service, great pr..	positive
I want a refund. It does not work! Ridiculous.	neutral
I would recommend this product to anyone. I will buy another one ..	positive
It is OK for what it costs. Not great but works as expected.	positive
Much better than the competing products for a much more reasona..	positive
My experience was terrible. The product was dead on arrival.	negative
Product description is misleading. It is much smaller than what's s..	neutral
This is an awesome product. I enjoy using it every day.	positive
Took 4 weeks to receive it even though I paid for 2 day delivery. W..	positive
What a joy to use.	positive

キャプション: Rのセンチメント分析の結果を視覚化することで、お客様の問題を誰でも簡単に見つけることができます。

センチメント分析では一瞬にして雑音を排除できますので、ビジネスユーザーは、お客様の声をお客様自身が発言したときに把握することができます。これにより、企業は特定の問題や変化するトレンドを迅速に特定し、対象となる顧客層の関係やマーケティング戦略に対応することができます。

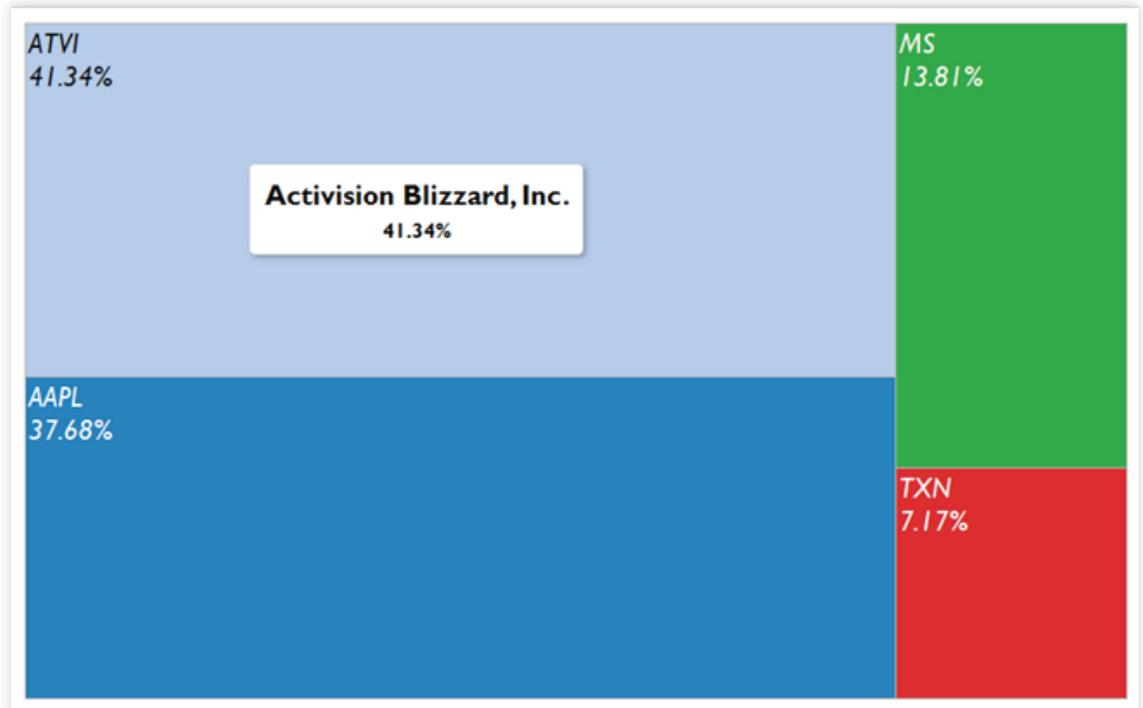
意思決定プロセスを自動化する

処方的な分析

めまぐるしく変化するチケットの価格を航空会社がどのようにして決定しているのか、不思議に思ったことはないですか？航空会社は、複雑な一連の旅行変数、顧客ニーズ、時期を分類することで、利益を損なわずに販売を促進するチケット価格をリアルタイムに提示しています。一連の制約を前提とし、最適なチケット価格を模索します。これが、処方的な分析です。

説明的な分析が既に発生したことを説明するのに対して、処方的な分析はこれから発生すると思われることを説明します。つまり、処方的な分析は、今後の予測を判断するのに役立ちます。判断結果が複数考えられる場合は、最適な結果を得るための分析を行います。処方的な分析は、ビッグデータで活用できます。

Rは最適化のための関数を使用し、処方的な分析を簡単に処理します。また、Tableauでは、それらの結果を視覚化し、やり取りできますので、ビジネスにおいてRの統計能力に基づいた意思決定を行うことができるようになります。たとえば、以下にあるのは、Rのoptim()の結果をツリーマップとして視覚化した例です。ここでは、4つの株式について、ポートフォリオの最適な割り当てを視覚化しています。



キャプション: Tableauを利用することで、このポートフォリオの割り当ての例のように、Rの処方的な分析機能の結果を簡単に視覚化できます。

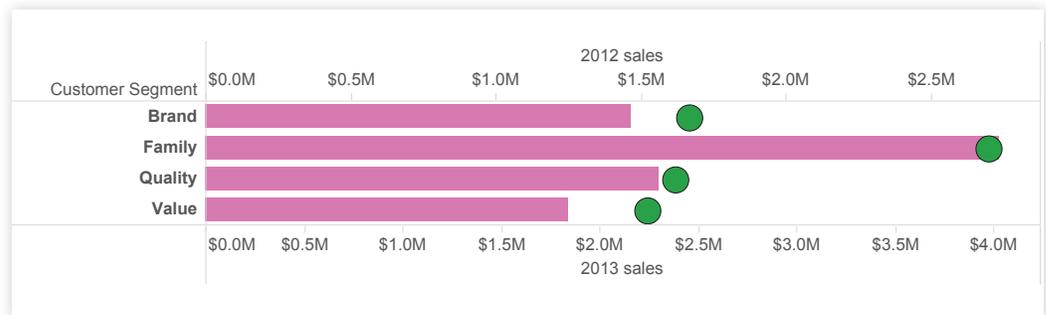
大規模で複雑なデータセットを理解する

デシジョンツリー

データが余りにも複雑なため、全体の回帰モデルや、それを説明できる単一の予測式を見つけるのが困難になることがあります。解決策の1つに、単純なモデルでそれぞれのサブセットを説明できるようになるまで、データをより小さなサブセットに細分化していくという方法があります。Rでは、複雑なデータを分類するためのデシジョンツリーを簡単に作成できます。

チームでデシジョンツリーを使用し、データを分類するためのRスクリプトを作成したら、その分類をあらゆるデータビジュアライゼーションに使用し、ドラッグアンドドロップするだけで色付け、形状の作成、グループ化を行うことができます。実際に、データを分類するためのRスクリプトを作成しておくことにより、あらゆるビジネスユーザーがそのスクリプトを使用した結果をデータビジュアライゼーションで活用できるようになります。

以下のビジュアライゼーションでは、顧客の購買傾向に基づいて、ある小売業者の顧客を分類するためのデシジョンツリーを作成しています。その後、分類をシンプルな棒グラフに使用し、グループごとに2012年の売上と2013年の売上を比較しています。TableauとRを連携させることで、データ科学者がRで作成したデシジョンツリー機能に基づいて、あらゆるビジネスユーザーがこのような方法でデータを調査することができるのです。こうすることで、データ科学者チームが行った成果を、ビジネスユーザーが日々有益に活用することができます。



キャプション: このビジュアライゼーションは、Rのデシジョンツリーが作成した顧客部門区分を表しています。ビジネスユーザーは、このデータを売上データ(この場合は、2012年、2013年のデータ)に組み合わせ、年ごとの各部門の支出の変化を把握することができます。

偶然ではないことを証明する

統計的な検定

法廷のように、有罪が証明されるまでは人は無罪です。統計的な検定は、誤りでないものを排除することで、ある対象が真実であることを証明するということを目的としています。ほとんどの場合、ある対象が偶然に発生したのではないことを証明することを意味します。通常、これは「統計的に有意である」と表現されます。たとえば、製薬調査では、特定の医薬品から変化がもたらされることを証明することを目的とします。成功であることが証明されれば、統計的に有意であり、偶然の結果ではないということになります。

サンプルには、性別や人種といったカテゴリーデータを含めることもできますし、体重、年齢、期間といったメジャーを含めることもできます。そういったデータを使用すれば、特定の薬品が、制御集合のメンバーに対してもたらす効果と比較し、特定の年齢層のメンバーにプラスの効果をもたらす可能性をテストすることができます。

Rにはさまざまな統計的な検定を行うための関数が用意されており、Tableauではそういったテストの結果を簡単に表示できます。Tableau のパラメーターを利用することにより、あらゆるユーザーがウェルチの2標本t検定などのテストのオプションを調整し、結果を簡単に視覚化できます。



キャプション: TableauではRの統計的な検定を行う関数の結果を簡単に視覚化でき、テストオプションを調整するためのフィルターも利用できます。

著者について

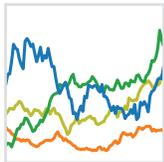
Bora Beran は、Tableau Softwareのプログラムマネージャーです。彼のチームは、統計および計算、技術的なパートナーシップ、クエリ生成パイプラインに重点的に取り組んでいます。Tableau に入社する前は、Bora は Microsoft Research、SQL Server、Windows HPC Server で、大規模な計算機のモデル化とシミュレーション、データマイニング、ビジュアライゼーション、知識表現のためのツールおよびインフラストラクチャに関する業務に携わっていました。彼はドレキセル大学の博士号を取得しており、テクニカルコンピューティング分野のいくつかの学術誌の査読者も務めています。

¹Karl Rexer, Heather Allen, Paul Gearan (2011)『[2011 Data Miner Survey Summary](#)』、Predictive Analytics World (2011年10月) で発表。

²Karl Rexer, Heather Allen, Paul Gearan (2011)『[2010 Data Miner Survey Summary](#)』、Predictive Analytics World (2011年10月) で発表。

Tableau について

Tableauは、お客様がデータを視覚的に理解できるよう手助けします。Tableauを利用することにより、情報をすばやく分析、可視化し、共有することができます。Tableauを導入することで、オフィス、又は外出先でも、すばやくデータ分析を行えるようになったお客様が既に19,000 人を超え、ブログやWebサイトでのデータ共有にTableau Public を利用しているユーザーも数万人にのぼります。Tableauがいかに役立つかを知るには、<http://www.tableausoftware.com/trial> から無料トライアル版をダウンロードしてください。



その他のリソース

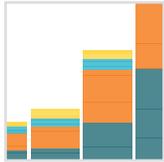
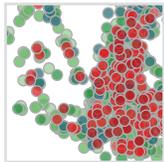
[無料トライアル版をダウンロード](#)

関連リソース

[統計分析のためのR プロジェクト](#)

[TableauとRの統合に関する詳細](#)

[Bora Beranのブログ](#)



その他のリソースを見る

- ・製品デモ
- ・トレーニングとチュートリアル
- ・コミュニティとサポート
- ・顧客事例
- ・ソリューション

