

ビジネスにおける商圈分析と機械学習

酒井嘉昭

Catchment analysis and Machine learning in Business

Yoshiaki SAKAI

Abstract: Applying Machine learning to Catchment analysis in business becomes popular method for understanding of regional market and consumer insights. Machine Learning can be defined by two perspectives; Supervised learning and Non-supervised learning. In business, there are some challenges to make consensus with non-specialists when we utilize Machine learning. This paper will provide some arguments on catchment analysis and Machine learning.

Keywords: 商圈分析 (catchment analysis), 機械学習 (Machine learning), 意思決定 (decision making), 可視化 (visualization), 最適化 (optimization)

1. はじめに

大量のデータから複雑なパターンを分類・回帰 (推計・予測) することができる機械学習は近年、商圈分析などマーケティング分野への応用が広がっている (酒井 2017)。

その一方で、ビジネス分野における意思決定者は、地域分析の専門家ではないため、分析結果が妥当かどうかを検証する方法は、依頼主と分析者の合意形成プロセスが重視される。

本発表では、実ビジネスにおける機械学習の商圈分析への適用における諸問題と事例について述べる。

2. 機械学習適用の諸問題

2.1 商圈分析における分類問題

機械学習は、大きく「教師なし」モデルと「教師あり」モデルとに分類される。

立地環境の理解と洞察を得るためには、立地環境に係わる諸データの比較と分類が主要な課題

となるため「教師なし」モデルによるクラスター分析が主要な手法となる。すなわち、複数の店舗立地を類型化する、地域区分問題を解くことと言い換えることができる。

クラスター分析においては、いくつに分類するか、また分類するための類似度をなにによって評価するかについては「正解」はないため、依頼主と分析者の間で、手法と結果の妥当性についての合意形成をしておくことが求められる。そのパターンは、大きく3つに類型化できる。

- 担当者の知見と照らし合わせて妥当と判断される場合 (例: 担当者の知っている既知の場所と比較して、その結果が納得できるのでこの分類は妥当だ)
- 担当者には十分な知見はないが、上司もしくは、信頼できる専門家が妥当性を示唆した場合 (例: あの方が妥当だというのならたぶん妥当なのだろう)
- 分類手法とロジックについて合理的かつ妥当と判断した場合 (例: この分析手法でかつこれらの変数を用いているのであればその結果は妥当なものである)

著者氏名 酒井嘉昭

所属 ジオマーケティング株式会社

E-mail, sakai.yoshiaki@geomarketing.co.jp

2.2 商圈分析における回帰問題

商圈分析における回帰問題では、売上予測が主要なテーマとなる。回帰分析を代表とする「教師あり」機械学習モデルは、あらかじめ正解となる学習データを目的変数として準備し、商圈情報などの説明変数を投入することによって構築される。

すでに店舗網があり売上データがあるものは、売上データを目的変数とし、商圈に係わる立地環境を説明変数とした「教師あり」モデルを構築することができる。

売上予測の場合、モデルの良し悪しはどのように評価したらよいかという問題は、依頼主との間で常に議論となる。店舗開発に応用する売上予測の場合、実績売上データへの当てはまりの良さ (goodness of fit) と予測性能の良さ (goodness of prediction fit) のバランスについて、担当者と認識を共有することができるかが課題となる。

3. 事例研究

大手チェーン店の商圈分析を行ったプロジェクト事例では、依頼者の希望により出店アセスメント用の売上予測モデルの精度を±10%から20%で既存店の売上が説明できることが求められた。

機械学習モデルの構築にあたっては、店舗開発担当部門に説明しやすいように店舗立地タイプを分類しそれぞれの立地タイプ毎にモデルを構築した。店舗立地タイプを分類するにあたっては、「教師なし」機械学習のクラスタリング手法である k-means 法を用い、売上予測モデルにおいては「教師あり」機械学習の回帰分析手法であるランダムフォレストをベースとした複数のモデルを組み合わせたアンサンブルモデルを用いた。

モデルの評価には、散布図を作成し10%から20%の誤差の範囲にどのくらい予測値が収まるかを表示し、担当者とのモデルの妥当性について議論し合意形成を行った (図1)。

20%以上の誤差のある店舗については現地調査

を行い、説明変数として想定していなかったものは何かについて議論をし、新しい説明変数を投入し検証する作業を複数回実施した。最終的には、10%以内の誤差で予測されたものは、全体の93%で期待される性能を確保することができた。

担当者からのフィードバックをベースにモデル構築を実施した結果、現場の知見がモデルに反映され、肌感覚になじむものを構築することができた。

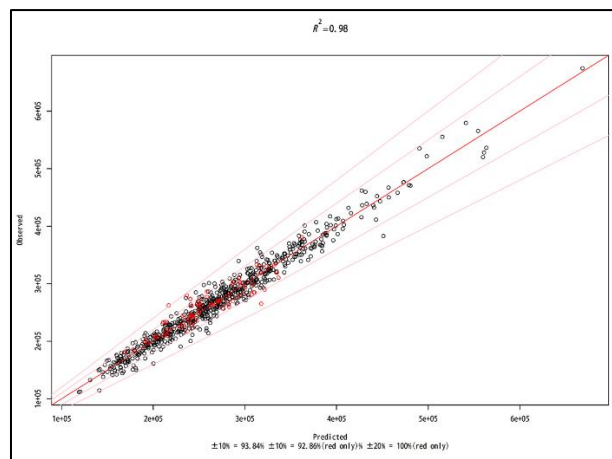


図1 予測結果の表示事例 (R2 乗値:0.98)

4. おわりに

商圈分析への機械学習モデルの応用は、分類・回帰問題に柔軟に対応することができることから、現場担当者の課題解決に対するモチベーションを高めることができる。

また、依頼主と共同で実施する仮説検証プロセスがモデル構築には不可欠で、信頼関係ならびに相互のリテラシーを高める上で重要な役割を果たしている。

参考文献

酒井嘉昭, 2017. 「ジオマーケティング戦略」 幻冬舎

Trevor Hastie, Robert Tibshirani, Jerome Friedman, 2009. *The Elements of Statistical Learning Data Mining, Inference, and Prediction. (Second Edition)* Springer