

飲食店における蓄積データの利活用

Analyzing accumulation data of a restaurant

徐晃生¹ 張迎¹ 菅愛子¹ 高橋大志¹

Hwangsaeng Seo¹, Ying Zhang¹, Aiko Suge¹, and Hiroshi Takahashi¹

¹ 慶應義塾大学大学院経営管理研究科

¹ Graduate School of Business Administration, Keio University

Abstract: 本研究は、飲食店の POS データと顧客データを用いて企業の経営力向上に繋げるシステムの提供を目的としている。昨今のビッグデータ活用時代において、個人のデータを蓄積し利活用することの重要性が増してきている。個人が頻繁に利用しデータ取得可能な業態として飲食店が挙げられ、そこで蓄積されたデータを分析することで CRM に有用な示唆を得ることができる。本分析では、顧客の購買行動の特徴を明らかにし、在庫管理と販売政策について検討を行った。

1. はじめに

昨今のビッグデータ活用時代においては、個人のデータを蓄積し利活用することの重要性が増してきている。特に、個人が頻繁に利用しデータ取得可能な業態として店舗の IT 化が進む飲食店が挙げられ、顧客や売上のデータを参考にしようにはなったが、これらのデータを顧客の分析に活用するまでには至っていない。今回、分析の対象とする飲食店(寿司)は生魚を多く扱うため、保存期間が短く在庫管理が難しく、正確な販売予測が課題とされていた。各商品の売上データを分析することで、商品ごとの販売個数を予測することができると考える。

本研究は、飲食店の POS データを分析し、一日毎の各商品の販売予測を目的としている。先行研究において、店舗全体の売上予測に関する分析が多く、各商品の販売個数を対象とした分析は限定的である。

2. 先行研究

実店舗の売上予測に関する研究が多い中、在庫管理の観点から材料発注量に着目した報告として、星野・谷崎ら(2018)が挙げられる。当研究では、店舗運営高度化支援の研究の一環として、機械学習を主体とした来店客数と発注量の予測を行っており、来店客数の予測率は 80% と良好な結果であったが、発

注量の予測に関しては、発注量が最も多い 3 種類の商品に限定したにも関わらず、予測率は 60% であり実用レベルではないとの報告を行っている。

また、辛・瞿ら(2019)は、本研究と同一飲食店のデータを用いて、飲食店での顧客購買データ及び会員データを分析し、新規客から常連客に進むためにはどのような特徴があるのかを分析した。新規顧客が再来店する行動を説明する要素を検出するために、データとの関連性を明らかにし、機械学習を用いて来店予測や購買予測など、販売施策の検討を行っている。

3. データ

POS データ及び会員データについては、該当飲食店の共同研究の了承のもと、ソフト・シアター株式会社から取得した。本研究は、2015 年 1 月 1 日~2018 年 12 月 31 日までの 1,455 日間 (4 年間) において、該当飲食店 1 店舗の POS データから 1 会計ごとのオーダーデータを抽出し分析を行った。表 1 は上記の POS データであり、表 2 はそのサンプルを示したものである。

表 1 POS データ

項目	数量	解釈
日数	1,455日	2015/1/1~2018/12/31
会計件数	83,237件	1テーブルごと
注文件数	1,574,126件	1注文ごと
商品種類	824種類	4年間

表 2 POS データサンプル

店舗コード	顧客コード	来店日	年	伝票番号	行	商品名	単価	数量	金額	売上部門名	イベント名	来店時間
000	10000000	2018/1/1	2018	0000	1	XX	000	1	000	テーブル	一般	00:00:00

4. 分析方法

本研究では、POS データから一日毎の各商品の販売数と単価を抽出し、各商品の販売数を日付毎に集計したデータをもとに分析する。

表3 日付別商品の統計データサンプル

来店日	商品名	販売数	単価
2019/1/5	ウーロン茶	5	199円
2019/1/5	カルピス	2	199円

飲食店において、4年間という長い期間のPOSデータを分析する上では、各商品の販売が終了したり、新商品が登場したりすることを考慮しなければならない。したがって、分析方法には汎化性能が高いXGBoostを用いて、各商品の販売量の予測を試みる。XGBoostはアンサンブル学習と決定木を組み合わせた手法で非常に高い汎化能力を誇る。予測モデルを作るために、2015年1月1日~2017年12月31日のPOSデータをもとに訓練データを作成し、2018年1月1日~2018年12月31日の各商品の販売量を日毎に予測した。

被説明変数を各商品が同日に受けた注文数とし、表3のような日付別商品の統計データを用いた。内部のPOSデータと外部のマクロデータ(時系列、曜日、天候など)から得られるデータをもとに説明変数を抽出した。以下の表はそれらをまとめたものである。

表4 被説明変数と説明変数

被説明変数	解釈	変数属性
販売数	同日に各商品が注文を受けた数	連続
説明変数	解釈	変数属性
商品名	該当商品かどうか	ダミー
価格	各商品の価格	連続
年	該当日は何年か	連続
月	該当日は何月か	連続
日	該当日は何日か	連続
曜日	該当日は何曜日か	ダミー
祝日	該当日は祝日かどうか	ダミー

5. 分析結果

全商品の中から、販売数が多い主要な品目を分析するために、販売数上位23品目(2015年1月1日~2017年12月31日の3年間で販売数7000以上)を選出し予測値を算出した。図1は予測対象である2018年1月1日~2018年12月31日の販売実績値と算出された販売予測値の年間平均を商品ごとに

示したものである。

実績値と予測値はほぼ同じ傾向を示している。年間を通して全体を包括すると精度が高い予測モデルのように見えるが、RMSE(root mean squared error)は5.8であり、日毎に各商品の実績値と予測値を比較すると、実績値と予測値の乖離が大きい日もあるため、予測モデルの改善が必要である。

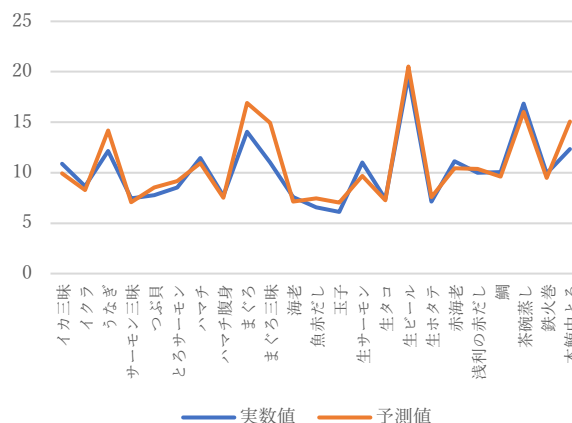
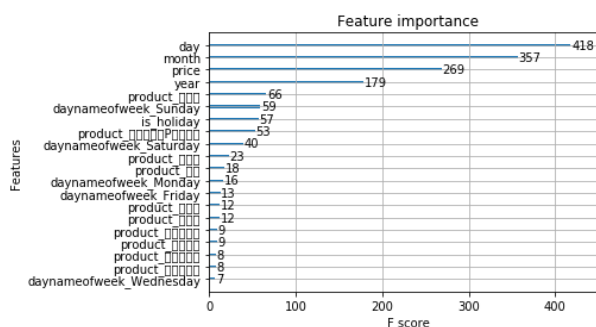


図1 商品別販売数予測値

予測モデルの特徴を可視化するために、各説明変数の重要度を算出した。表5は予測モデルの特徴量の重要度を示したものである。重要度を見ると、相対的に「日・月・価格・年」の重要度が高いことが分かる。XGBoostはアンサンブル学習と決定木を組み合わせた非常に複雑な手法であるため、モデルの概要が把握しにくいのが、特徴量の重要度を計算することで、予測結果に影響する重要な要素を把握できた。

表5 特徴量の重要度



6. まとめ

本研究は、飲食店のPOSデータを対象とした分析を行うことで、機械学習を用いて、商品ごとの販売量を予測する事を試みた。分析の結果、販売量が多

い商品に関しては高い精度を得られる一方、販売量が少ない商品に関しては改善の余地が大きいことを確認した。今後の課題としては、天候や気温などの外部要因、及び店側が操作可能な商品販促やメニュー構成などの内部要因を要素として取り込むことなどが挙げられる。

参考文献

- [1] 辛郷孝・瞿雪吟・菅愛子・高橋大志：高級寿司割烹店における CRM を目的とした顧客行動研究, SIG-BI #12, (2019)
- [2] 竹村遼・鈴木達哉・宮本道子・飯塚佳代：外食産業の売上予測について - 外食チェーンのある店舗における分析結果より -, 2012 年春季全国研究発表大会, (2012)
- [3] 星野智洋・谷崎隆士・新村猛・竹中毅：機械学習を用いた飲食店運営の効率化へのアプローチ, 2018 年度人工知能学会全国大会, (2018)
- [4] 新美潤一郎・星野 崇宏：顧客行動の多様性変数を利用した購買行動の予測, 人工知能学会論文誌, (2017)
- [5] Tianqi Chen・Carlos Guestrin：XGBoost: A Scalable Tree Boosting System, (2016), <https://arxiv.org/abs/1603.02754>