

季節変動のある購買履歴のフーリエ解析による販売予測手法の 提案と精度検証

Proposal and Accuracy Verification of Sales Forecast Method Using Fourier Analysis on Sales History Data with Seasonal Variation

吉澤 貴拓^{*1}
Takahiro Yoshizawa

善甫 啓一^{*2*3}
Keiichi Zempo

中林 紀彦^{*4*2}
Norihiko Nakabayashi

^{*1} 筑波大学 社会工学類
College of Policy and Planning Sciences, Univ. Tsukuba

^{*2} 筑波大学 システム情報系
Faculty of Engineering, Information and Systems, Univ. Tsukuba

^{*3} 産業技術総合研究所 人工知能研究センター
Artificial Intelligence Research Center, AIST

^{*4} 株式会社オプトホールディングス
データサイエンスラボ
OPT Holding Inc. Data Science Lab.

In fashion industry, sales forecasting plays an important role in profit rate. This research aims to forecast sales with seasonal variation using historical data. In this paper, we propose a new method for sales forecast using Fourier transform and Lasso to verify the prediction accuracy, we analyzed sales data for 6years in a Japanese fashion company's store and forecast sales in the next year. As a result, the proposed method is superior to the accuracy of forecasting in SARIMA model. Therefore, the proposed method is useful for forecasting sales with seasonal variation.

1. はじめに

アパレル業界では機会損失の低減や過剰在庫削減などのコスト削減や利益率向上のために、販売予測がとても重要な役割を担っている。しかしアパレル業界の売上は、気候や商品の持つ季節性、流行、経済に至る様々な要因に影響を受けるとも変化しやすいものであり、その予測は難しいと言われている。

ファッション業を対象にした販売予測の研究は、大きく2つに分けられる。1 つ目は ARIMA モデルやベイズ的アプローチなどの伝統的な統計的予測手法であり、2 つ目はニューラルネットワークなど AI を用いた予測手法である。統計的手法の特徴として、実装が簡単かつ計算が高速というメリットがある一方で、統計的な知識が必要な上にその精度は低いなどのデメリットがあることが挙げられる[Liu *et al.*, 2013]. その中で、周期性を持つデータに対しフーリエ級数を基底変数とする非線形回帰モデルが良い推定を行うことが知られており[Kawano, 2010], 統計的一手法として[Fumi *et al.*, 2013]ではファッション産業におけるフーリエ変換を用いる予測手法が提案された。しかし、この論文で提案された予測手法は、予測モデル選択にあたって予測期間における実測値を使用するため、予測手法として不完全なものと言える。

以上を踏まえ、本論文の目的を 2 つとし、1 つ目は、学習データのみから予測を行うことができるフーリエ変換を応用した予測アルゴリズムを構築することである。2 つ目は、このアルゴリズムをアパレル産業の購買履歴データに対して適用し SARIMA モデルの予測値と比較することで、季節変動のある購買履歴デ

ータに対してフーリエ変換を用いるアルゴリズムの有用性を検証することである。

2. 手法

2.1 予測手法

本稿の分析に用いた予測アルゴリズムの手順とその計算について説明する。この予測アルゴリズムは、周期性のあるデータに対して良い曲線推定をするフーリエ解析と過剰適合を防ぎロバストなモデルを作成するための Lasso を組み合わせたものである。

とあるデータ長 N の時系列データ $y(t)$ はフーリエ級数展開することによって以下の式に書き換えることが可能である。

$$y(t) = \sum_{i=1}^N (a_i \cos(2\pi f_i \cdot t) + b_i \sin(2\pi f_i \cdot t)) \quad (1)$$

ここで f_i は任意の周波数を表しており、対応する a_i, b_i はそれぞれの周波数成分の $y(t)$ への寄与度を表している、元の時系列データを書き換えることができ、またこの(1)式は以下のように書き換えが可能であり、これによって元の時系列データを周期性成分に分解することが可能である。

$$y(t) = \sum_{i=1}^N (A_i \cos(2\pi f_i \cdot t + \theta_i)) \quad (2)$$

ここでの A_i, θ_i はそれぞれの周波数成分に対する寄与度と位相成分を表している。

Lasso は、L1 型正則化法の一つであり、目的変数を \mathbf{y} , 説明変数を \mathbf{X} とする回帰分析において、残差2乗和にパラメータの絶対値を用いる正則化項を含む誤差関数(3)の最小化によって、線形回帰のパラメータ $\boldsymbol{\beta}$ の推定を行う手法である。[Kawano, 2010]

Contact: 吉澤 貴拓, 筑波大学理工学群社会工学類,
〒305-8573 茨城県つくば市天王台 1-1-1,
takahiro.yoshizawa@sk.tsukuba.ac.jp,
zempo@iit.tsukuba.ac.jp

$$S_{\alpha}(\beta) = {}^t(\mathbf{y} - \mathbf{X}\beta)(\mathbf{y} - \mathbf{X}\beta) + \alpha \sum_{j=1}^n |\beta_j| \quad (3)$$

この特徴として、幾つかのパラメータを 0 と推定することが可能であり、結果としてパラメータ推定と変数選択を同時に行うことが可能である。この 2 つを組み合わせることによって学習データの周期性に着目した予測モデルを構築する。

今回用いた予測アルゴリズムは以下の通りであり、図 1 はその概略図である。

- データ前処理
学習データの対数変換を行い、トレンド成分を取り除く。
- 周波数成分分析
フーリエ級数展開を行い $i:1 \sim N$ に対応する周波数成分へ分解する。
- Lasso 回帰: モデル決定
データ前処理後の学習データを目的変数、周波数 f_i ごとに以下の(4)式で逆変換を行ったものを説明変数とする Lasso 回帰を行う。

$$y_i(t) = |A_i| \cos(2\pi f_i \cdot t + \theta_i) \quad (4)$$

Lasso はパラメータを 0 と推定することで変数選択を行う事が可能であり、これを用いてスパースなモデルを作成する。今回は一定の範囲内でパラメータスキャンを行い、 α を 0.0001 として推定を行う事とした。

- フーリエ逆変換: 予測
周波数成分と Lasso で算出されたそれぞれの寄与率を用いて、予測を行う。
- データ後処理: 予測値算出
前処理と逆の手順を行う。トレンド成分 Z を再度適用し、その後、指数化を行う。そしてこれを予測値として精度検証に用いる。したがって、予測値 $F(t)$ の算出式は次の(5)式のようになる。

$$F(t) = \exp \left(\left(\sum_{i=1}^N \beta_i (|A_i| \cos(2\pi f_i \cdot t + \theta_i)) \right) + Z \right) \quad (5)$$

2.2 精度評価方法

本論文では、一般的によく用いられる SARIMA モデルでの予測値を用いて今回の予測値を比較評価する。また、評価基準として式(6)に示される平均絶対誤差率(MAPE), C_{MAPE} , と(7)式に示される二乗平均平方根誤差(RMSE), C_{RMSE} の二つを使用する。二つの評価関数はそれぞれ次の通りである。

どちらの評価指標も 0 に近ければ近いほど、優れた予測モデルである事を示す。

$$C_{MAPE} = \sum_{t=1}^N (|F(t) - Y(t)| / Y(t)) / N * 100 \quad (6)$$

$$C_{RMSE} = \sqrt{\frac{1}{N} \sum_{T=1}^N (F(t) - Y(t))^2} \quad (7)$$

3. 予測分析・検証結果

3.1 分析データ

本研究に用いる購買履歴データは、とある日本のファッションブランドの実際の店舗における商品カテゴリ別の 2007 年～2013 年の 7 年分の販売個数であり、そのうち全 20 商品カテゴリを対象に分析を行う。全 7 年のうち、前 6 年分を学習データとし、後 1 年を検証データとした。

3.2 精度比較・検証

本論文で提案したアルゴリズムを用いて予測を行い同条件下での SARIMA モデルの予測値と比較したところ、どちらの評価指標においても 11 の商品カテゴリで高い精度の予測となった。その予測精度は C_{MAPE} において、20%以下のものが 7 カテゴリ、50%以上のものが 3 カテゴリとなった。一方 SARIMA モデルでは、20%以下のものが 5 カテゴリ 50%以上のものが 3 カテゴリとなった(表 1)。また、 C_{MAPE} では SARIMA モデルよりも悪い評価であったが、 C_{RMSE} では良い評価となったカテゴリが 2 カテゴリ存在した。

図 2-1 は予測精度の高かったパンツのカテゴリの予測値である。予測結果の特徴として 0 週付近と 20～30 週付近にある大きなピークを時期のずれなく予測でき、かつ小さな売上の上下も予測できた。一方で SARIMA モデルは年間に 2 回現れているピークをどちらも追うことができていない。図 2-2 は、パンツの学習データであり、同様の波形が周期的に観察される。

C_{MAPE}	20%	20%～50%	50%～
Fourier Transform	7	10	3
SARIMA	5	12	3

表 1: C_{MAPE} 値とカテゴリ数

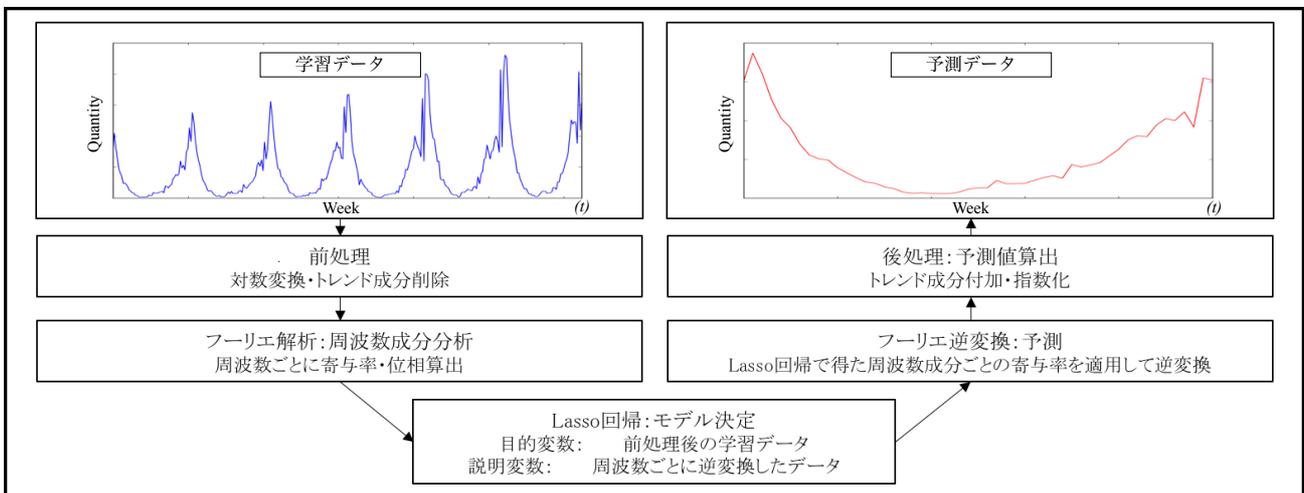


図 1: 予測アルゴリズム 概略図

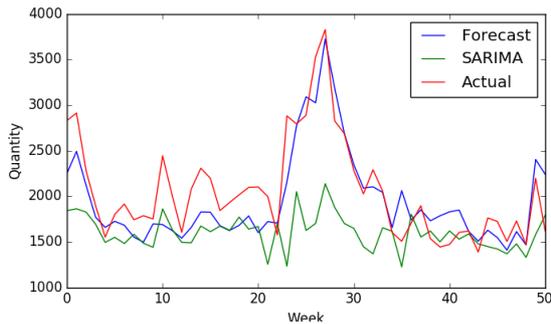


図 2-1 : パンツの予測値

表 2 : パンツの予測評価値

	$C_{MAPE; \text{Pants}}$	$C_{RMSE; \text{Pants}}$
Fourier Transform	11.82	309.01
SARIMA	20.30	644.01

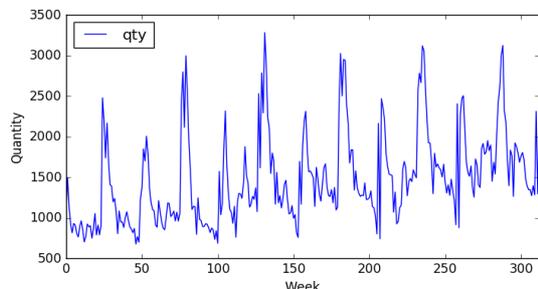


図 2-2 : パンツの学習データ

表 4 : 学習データ変更後 ニット予測評価値

	$C_{MAPE; \text{Knit}}$	$C_{RMSE; \text{Knit}}$
Fourier Transform	12.85	420.23

図 3-1 は予測精度の低いニットのカテゴリの予測プロットであり、予測値は 0~10 週にかけて大きく下回っている一方で、30 週目以降は大きく上回っている。図 3-2 はニットの学習データであるが、毎年 50 の倍数付近に存在するピークが 0 週付近には観察されない。これは発売初期であったことに起因して売上が少なくなっていたことがわかったため、1 年目のデータを外的要因による売上の異常であることから学習データのノイズとして扱うこととした。ニットの学習データにおいて毎年同様の周期性を確認できる後ろ 5 年間に学習期間を短くして予測を行った結果、大幅に精度が改善された(図 3-3,表4)。これより、1 年目のデータによって誤った学習をしたと考えられる。同様の理由で計測期間内に売上の少ない時期が存在したカテゴリにおいて、学習期間を同様に短くして予測したところ、20%以上の精度改善がなされたものは 2 カテゴリ、10%以上の精度改善がされたものは 2 カテゴリであり、これらはどれも 6 年の学習データにおいて精度が 50%以上のものであった。

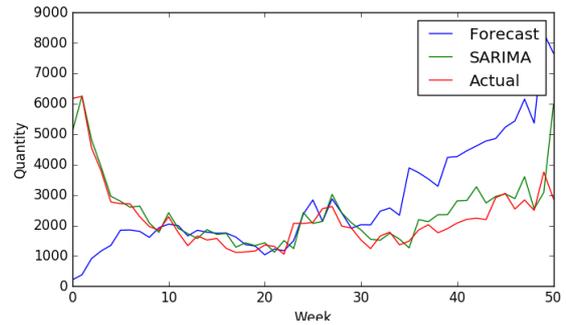


図 3-1 : ニットの予測値

表 3 : ニットの予測評価値

	$C_{MAPE; \text{Knit}}$	$C_{RMSE; \text{Knit}}$
Fourier Transform	56.51	2080.92
SARIMA	17.53	588.43

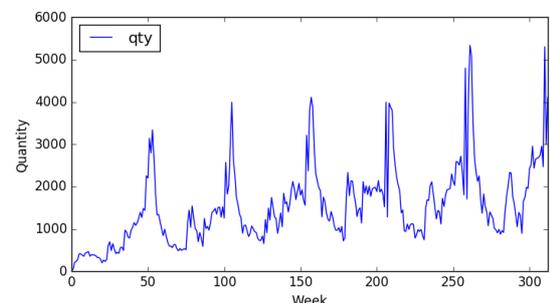


図 3-2 : ニットの学習データ

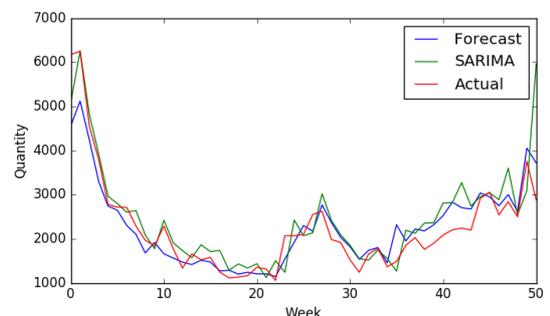


図 3-3 : 学習データ変更後 ニットの学習データ

4. 結論

本稿では、年間を通した周期による変動を持つ購買履歴データの販売予測手法としてフーリエ変換を応用した予測手法を提案し、その予測結果と SARIMA モデルによる予測結果との比較検証をおこなった。

提案された手法は、すべて学習データからモデルを作成することができ、[Fumi et al., 2013]によって提案されていた予測手法の不完全な点を補った。また、このアルゴリズムによる予測は半数以上の商品カテゴリにおいて SARIMA モデルより高い精度での予測が可能であり、1 年や半年の大きな周期だけでなく数週間周期の小さな波形も予測することができた。よって、本稿で提案された予測アルゴリズムの季節変動を持つアルゴリズムの有用性は十分にあると言える。さらに、SARIMA モデルに比べ

パラメータチューニングの必要性が少なく、計算時間も短いことから実装においてもより有用な予測手法である。

商品カテゴリの季節性の有無と精度には顕著な関係性は見られず、またニットカテゴリのように売り始め期間を取り除くことによる学習期間の変更によって精度が改善されたことから、学習データの周期の安定性が精度に影響を与えていることが確認された。

今後の課題として安定した周期性がないデータや突発的な変化を持つデータへ対応によって、売り始めやセールスなどの時期における挙動への対応のモデル化などがあげられる。

謝辞

ジェイモードエンタープライズ株式会社の久保氏、中根氏には多岐に亘りご指導・ご協力いただいた。

参考文献

- [Liu *et al.*, 2013] Na Liu, Shuyun, Tsan-Ming Choi, Chi-Leung Hui, and Sau-Fun Ng : Sales Forecasting for Fashion Retailing Service Industry: Review , Mathematical Problems in Engineering, Hindawi Publishing Corporation, 2013.
- [Fumi *et al.*, 2013] Andrea Fumi, Arianna Pepe, Laura Scarabotti and Massimiliano M. Schiraldi: Fourier Analysis for Demand Forecasting in a Fashion Company, International Journal of Engineering Business Management, 2013.
- [Kawano, 2010] 川野 秀一, 廣瀬 慧, 立石 正平, 小西 貞則: 回帰モデリングと L_1 型正則化法の最近の展開, 日本統計学会誌, 2010.