

短期間データに基づく製パン新商品の受注量パターン予測

Bread Sales Pattern Forecast from Short-Term Data

工藤 大輝¹ 後藤 裕介¹

Daiki Kudo¹ Yusuke Goto¹

¹岩手県立大学

¹Iwate Prefectural University

In general, baking companies develop a large number of new products, but only some of them become established as standard products, and the production of new products may result in a large amount of waste loss due to unpredictable sales. Therefore, it is necessary to predict the future order volume based on the short-term order data of new products. In this paper, we propose a method for forecasting and evaluation using a combination of multiple explanatory variables and models, using shipment data from bakeries for a short period of time, such as one month after the start of sales. In the proposed method, the most accurate results were obtained by constructing a model with variables that took into account the characteristics of random forests and bread making, and explanatory variables that were modified for the low number of stores.

1. はじめに

製パン企業では生産時、人の勘(前回の発注数, 似たような商品の発注数 etc)で出荷予測を行い, 廃棄ロスが発生している. 一般に多数の新商品を展開しており, 新商品は特別な材料(クリームやパッケージ etc)を使用している. 新商品の一部しか定番商品として定着しないため, 新商品の製造にあたっては売行き予想が外れた結果として, 使用している特別な材料が廃棄となり, 多量の廃棄ロスが発生することがある. このため, 新商品の短期間の受注データに基づいて, 将来の受注量を予測することが必要である.

製パンの特徴として, 短命で季節性が高いことや商品毎に取り扱い店舗数が違う, 発注頻度が高いことが挙げられる. 短命で季節性が高い点については, 予測できるデータに限りがあるという問題, 取り扱い店舗数の違いや発注頻度が高い点については, 予測時のパラメータに関わってくる可能性が考えられる.

関連研究は, 従来の予測モデルとして, ARIMA^[1]や prophet^[2], 新商品の予測モデルとしては 3F アルゴリズム^[3]が挙げられる. 従来の予測モデルである ARIMA や prophet には, 短期間データでは予測が難しいことや製パン工場の出荷データは限られたデータしか使えないため, 季節性やイベント, 顧客情報などのデータを使えないと行った課題がある. 新商品の予測モデルである, 3F アルゴリズムには, 製パ

ンと取り扱い店舗数や発注頻度が違うという課題がある. このため, 製パンの特徴を考慮した変数などを用いた予測が必要であると考え.

上述の背景より, 製パン工場の出荷データを利用し, 販売開始されてから1ヶ月などの短期間のデータ使用し, 複数の説明変数とモデルの組み合わせで予測, 評価を行う手法を提案する.

2. 関連研究

これまでに商品の需要予測を行う研究は多数行われている. 本性では製パンの特徴に触れながら, 従来の手法や新商品に適応した手法と比較し課題点について述べる.

2.1 製パンの特徴

予測を行う上で, 商品の特徴を捉え, それに基づいた予測を行う必要があると考え. 製パンの特徴として, 短命で季節性が高い, 商品毎に取り扱い店舗数が違う, 発注頻度が高いの3つがあると考えた.

製パンの新商品は, 季節ごとの商品などレパートリーが豊富であり, 販売期間が約3ヶ月ほどと短く, 短命で季節性が高いことが共同研究先企業様からのヒアリングと基礎分析からわかった. データが3ヶ月以内のものしか使用できず, 本研究で使用する出荷データには出荷店舗やイベント情報が含まれていない. そのため, トレンドやイベントデータも使用できず, 予測に使用できるデータに限りがあるとい

った課題がある。

製パンの商品は、店舗の発注者が仕入れ商品に対し発注、発注がきてから製パン工場で商品を生産し、出荷を行う形となっている。店舗での商品の売れ行きや発注者のこの商品は売れそうであるなどの考えで発注の仕方が変化し、人気の高い商品は良く売れるため取り扱い店舗数が増加、人気の低い商品はあまり売れないため取り扱い店舗数が減少傾向になることが考えられることから、商品毎に取り扱い店舗数が違うことが言える。取り扱い店舗数が違うことから、店舗数の増減が商品の人気を表し、出荷数の増減と関係、受注量パターン予測時の一種のパラメータとして効果的な可能性が考えられる。

製パンの賞味期限が短く、出荷されてから約3日となっており、他の商品と比べ発注間隔が短い。そのため3日以内のサイクルで生産されていることから発注頻度が高いと言える。パンによって、生産サイクルが違い、人気が高い商品ほど生産サイクルが短いことが考えられる。発注頻度を変数の中に考慮することで製パンの売れ行きの違いが表現できるのではないかと考える。

2.2 従来の予測手法

従来の予測モデルとして、ARIMA^[1]や prophet^[2]が挙げられる。ARIMA^[1]は、自己回帰(AR)に、移動平均(MA)と差分過程(Integral)を組み合わせたモデルである。prophet^[2]は、ドメイン知識(季節性やイベント etc)を持つ人が統計知識がなくても簡単に予測できるように設計されたモデルである。これらの従来の予測手法の課題として、短期間データでは予測が難しいことや製パン工場の出荷データは限られたデータしか使えないため、季節性やイベント、顧客情報などのデータを使えないと行った点が挙げられる。

2.3 新商品の予測手法

新商品の予測モデルとしては3Fアルゴリズム^[3]が挙げられる。3Fアルゴリズム^[3]は、ファストファッションの短命で季節性の高く、リードタイムが短いといった問題に着目し、短期間のデータを使用し、短時間で予測を行う、グレイ法(GM)と拡張極端学習マシン(EELM)を組み合わせたモデルである。新商品の予測モデルの課題として、取り扱い店舗数や発注頻度が違いが考慮されていないという点が挙げられる。製パンは、規模の大きい店舗から小さい店舗までさまざまな店舗から発注を受け出荷されている。人気のある商品はさまざまな店舗が発注し、取扱店舗数の増加や発注頻度が高くなる。取扱店舗数や発注頻度が売れる売れないに関係し、予測時の変数として効果的なのではないかと考える。

3. 基礎集計

基礎集計より、新商品は販売期間が短いものが多い、初週の出荷数割合出荷数割合のパターン(伸び方)に違いがありそうという2つの特徴がわかった。

図1は商品の販売期間による生存曲線である。生存曲線は、商品毎に販売開始から販売日数が経過するにつれ、対象商品のうちどれだけ商品が残るかの割合で作成した。生存曲線より、半数の商品が90日以内で販売終了、8割の商品が221日以内で販売終了になり、1年以上販売が続く商品は約1割と販売1年以内での死亡率が高いことが見て取れる。このことから新商品は販売期間が短いことがわかった。

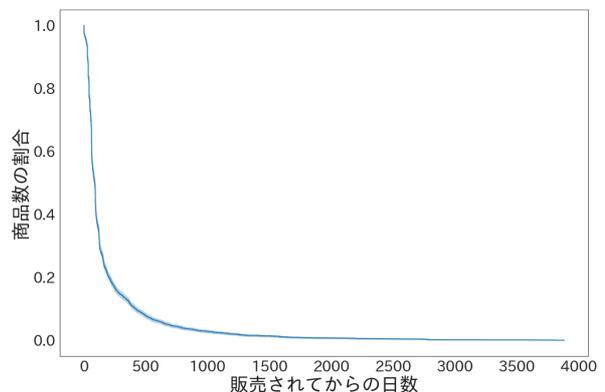


図1：商品の販売期間による生存曲線

図2は出荷数割合をクラスタリングしたセントロイドである。予測に際し、製パンの出荷数の推移での特徴理解や特徴量、基準とするため、クラスタリングを行った。クラスタリング手法としては、非階層型クラスタリング分析手法であるK-means法を使用する。クラスタ数は、エルボー法とシルエット分析を用いて推定し、クラスタ数3が適切であると判断し、クラスタリングを行った。クラスタ結果より、出荷数割合がどの日数で100%に近づくかで分類が行われ、初週の出荷数割合より出荷数割合のパターン(伸び方)に違いがありそうということがわかった。

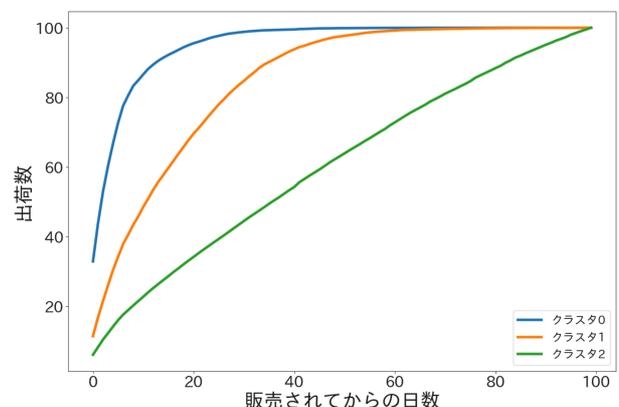


図2：出荷数割合でのクラスタリング

4. 提案する予測手法

4.1 対象データ

本研究で使用するデータは、製パン工場の出荷データである。データの概要は、(1)カテゴリー：菓子パン類、(2)期間：2012年1月～2019年12月、(3)レコード数：532,371件、(4)商品件数：1,027件、(5)店舗数：5,027件、(6)内容：商品ID、商品名、店舗ID、店舗名、出荷数、販売日である。

製パンの新商品の販売期間が約3ヶ月と言われていた点から、商品の販売開始から約3ヶ月(100日)のデータを使用して予測を行う。

4.2 受注量パターンの定義

基礎集計2)より、出荷数割合の伸び方のピークまでの期間の違いが大きく分けて3つ見られた。違いが見られた3つのパターンを受注量パターンとして定義する。

- ① 1ヶ月で出荷数割合が100%に到達
- ② 2ヶ月で出荷数割合が100%に到達
- ③ 3ヶ月以上出荷が続く

4.3 提案手法の概要

図3は提案手法の概要となっており、製パンの特徴を考慮したものや工夫していないものなど各説明変数を作成し、それらを複数の分類予測モデルでそれぞれ学習させ、一番精度の良好な組み合わせを探す。

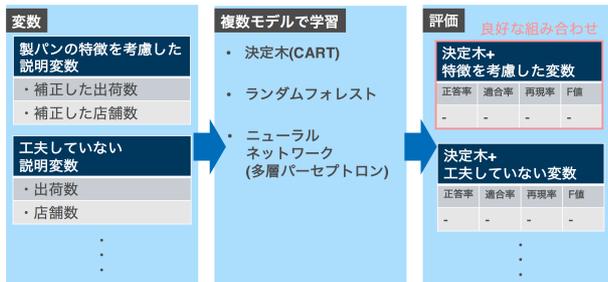


図3：提案手法の概要

5. データ作成

データ作成にあたり、製パンの商品によって出荷数の規模が異なることを考慮するため、商品毎に約3ヶ月(100)日目の出荷数を最大出荷数とし、出荷数割合と累積出荷数割合を求め、分析に使用する。

5.1 目的変数の作成

定義した受注量パターンで分類し、作成した目的変数を表1に示す。

表1：出荷数割合の伸び方によるパターン分類

Class	基準	件数
0	1ヶ月で100%	322件
1	2ヶ月で100%	378件
2	3ヶ月続く	327件

5.2 説明変数の作成

5.2.1 補正をかけた出荷数割合の作成

商品によって出荷数の規模が違うなか、出荷数のみで予測を行なった場合、似た推移の同じ受注量パターンに分類される商品が、ルール作成時に別のものと識別され、うまく学習できないことが考えられる。そこで出荷数の割合にすることで同じパターンとして、うまく学習させる。

しかし、そのまま割合にすると利用できる最終タイミングを100%として比率に変換するが、予測期間以上売れ続ける商品が考慮されず、受注量パターンの違いとしてうまく学習できないことが考えられる。そこで出荷数の傾きを変化量として考慮することで、売れ続ける商品とそうでない商品の違いを表現することで学習しやすくする。

出荷数の傾きを変化量として考慮する際に、1)基礎集計より、初週の出荷数の大きさにより、パターンの分類が地学なりそうなこと、2)発注頻度が高く、3日以内のサイクルで生産されているという2つの考えを変化量として使用する。

実際に決定木分析で予測を行い、どの日数を変化量とすることで効果的なのか比較を行った結果を表2に示す。

表2：変化量に使用する日数の比較

使用日数	正答率	適合率	再現率	F値
2日	0.80	0.80	0.80	0.80
3日	0.77	0.77	0.77	0.76
5日	0.78	0.78	0.79	0.78
7日	0.81	0.81	0.81	0.81

表2の結果より、変化量として考慮する日数として7日が最適であることがわかった。

これらのことより、補正をかけた出荷数割合の作成方法として、まず商品*i*の販売開始から7日までの出荷数の推移で伸びた場合の累積出荷数の最大値 P_i^{Max7} を7日目の累積出荷数 P_{i7} を用いて計算する。

$$P_i^{Max7} = \frac{P_{i7}}{7} \times 100 \dots \text{式1}$$

7日目の累積出荷数の傾きで推移した場合の補正した出荷数割合 PF_{in} を式1で求めた P_i^{Max7} と n 日目の商品の累積出荷数 P_{in} を用いて計算する。

$$PF_{in} = \frac{P_{in}}{P_i^{Max7}} \times 100 \dots \text{式2}$$

5.2.2 店舗数の推移の作成

取り扱い店舗数の変動が商品の人気度に関係し、出荷が続くのか、終了するかという、受注量パターンの判断ができるのではないかとこの考えから、店舗数を説明変数として使用する。

店舗数の推移をそのまま使ってしまうと、取り扱い店舗数の規模によって、同じような推移でも、違うものと認識されうまく学習できないことが考えられるため、商品*i*毎の*n*日目のユニーク店舗数の推移 S_{in} を初日のユニーク店舗数 S_{i1} で正規化したデータ SN_{in} を作成する。

$$SN_{in} = \frac{S_{in}}{S_{i1}} \dots \text{式 3}$$

5.2.3 店舗数での補正

図4はモデル構築時に見られた、予測が外れた商品の一部ランダムに抽出し、店舗数の推移をグラフとして作成したものである。図4のように、モデル構築時の外れている商品の原因分析の際に、外れている商品の中に多数、店舗数が低いまま停滞しているような推移の商品が多く存在していることがわかった。

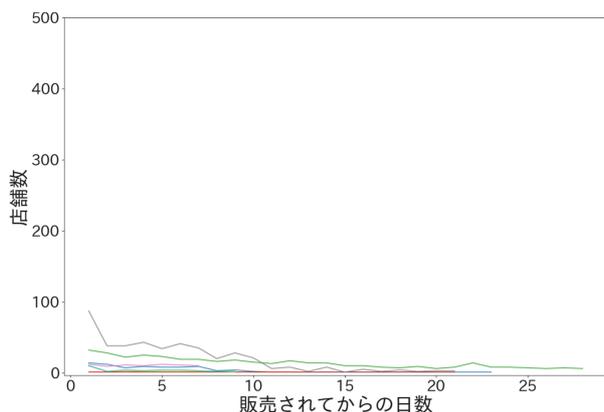


図4：一部外れた商品の店舗数の推移

そこで精度向上のため、店舗数の分布によって、どれほどユニーク店舗数を抱えている商品の中に、外れている商品が多いのか詳しく調査を行った。図5は、モデル構築時に、予測が当たった商品と外れた商品のユニーク店舗数のカウントを積み上げヒストグラムとして作成したものである。図5より、店舗数が10以下のものに外れている商品が多く存在していることがわかった。そこで、予測時の精度を向上させるため、ユニーク店舗数が10以下の商品を除外し補正を行ったものを説明変数として追加し、補正したものと比較を行う。

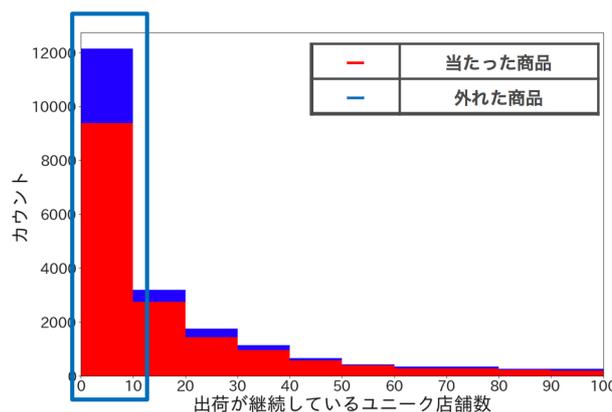


図5：店舗数の分布

6. 提案手法の評価

製パンの特徴を考慮したものや工夫していないものなどの複数の変数を、複数モデルで構築・予測、精度の比較を行い、一番精度の高いモデルの組み合わせを見つける。

6.1 使用する変数

分類予測モデルに用いる変数として、目的変数は定義した受注量パターン、説明変数は、① 製パンの特徴を考慮した説明変数と② 工夫していない説明変数、③ ①に取り扱い店舗数が低いものを補正したもの、④ ②に取り扱い店舗数が低いものを補正したものの4つの変数を使用する。

6.2 モデルの構築

今回予測を行うにあたり、分類の予測に使われるモデルを複数構築し、評価を行う。

データが少ないため、全てのデータに CV(K=3) と F 値で GridSearch を行い、最適なパラメータで学習を行った。

分類の予測に使用するモデルは以下の3つを使用した。

- (1) 決定木 (CART)
 段階的にデータを分割し、木のような分析結果を出力する。
- (2) ランダムフォレスト
 決定木により、複数の学習器を作成・統合し、汎化能力を向上させたモデル。
- (3) ニューラルネットワーク (多層パーセプトロン)
 人間の脳の神経細胞とのつながりを人工ニューロンという数式的なモデルで表現したもの。

表 3：各説明変数とモデルの組み合わせの評価

モデル	説明変数	正答率	適合率	再現率	F 値
決定木	①	0.81	0.81	0.81	0.81
	②	0.77	0.77	0.77	0.77
	③	0.83	0.86	0.85	0.85
	④	0.82	0.83	0.85	0.84
ランダムフォレスト	①	0.85	0.85	0.85	0.85
	②	0.93	0.93	0.93	0.93
	③	0.96	0.96	0.97	0.97
	④	0.94	0.94	0.95	0.95
多層パーセプトロン	①	0.79	0.79	0.79	0.79
	②	0.86	0.87	0.86	0.86
	③	0.74	0.83	0.79	0.77
	④	0.81	0.86	0.83	0.83

表 4：予測に使用する日数を短くし、各モデルで予測を行った評価

モデル	使用日数	正答率	適合率	再現率	F 値
決定木	14 日	0.66	0.63	0.61	0.62
	21 日	0.79	0.80	0.75	0.77
ランダムフォレスト	14 日	0.69	0.68	0.65	0.66
	21 日	0.66	0.66	0.61	0.60
多層パーセプトロン	14 日	0.68	0.71	0.64	0.66
	21 日	0.65	0.63	0.63	0.63

6.3 精度の評価

精度の評価にあたり、正解率・適合率・再現率・F 値の 4 つを評価指標として用いる。

6.4 予測結果

各説明変数で作成したモデルについて、各性能で比較を行った結果、表 3 のようになった。

表 3 よりランダムフォレストに③の製パンの特徴を考慮した変数と店舗数が低いものに補正をかけたモデルの構築が 4 つの指標で 0.96 以上と、一番精度が高い結果となった。

決定木やランダムフォレスト構築時に予測の分類に効果的なルールとして以下の 2 つの特徴が見られた。

1. はじめに 29～31 日目の出荷数や店舗数に関連した変数の変化量で、Class2 に属するものか、そうでないものかに分類
2. Class2 に分類されなかったものの中で、6～12 日目の出荷数や店舗数に関連した変数の

変化量で、Class0 に属するものか、Class1 に属するものか分類

効果的なルールと見られた 2 つの特徴より、製パン新商品の 1 ヶ月続くものと 2 ヶ月続くものかの分類は、6～12 日目の出荷数や店舗数、3 ヶ月以上販売されるものは、29～31 日目の出荷数や店舗数を見ることで判断できそうなことがわかった。

7. さらに短い期間での予測

製パンの予測を行う上で、1 ヶ月に販売終了してしまうような商品に対して、1 ヶ月のデータで予測を行うのでは遅く、更に短い期間(2, 3 週間)のデータで予測を行うことができると経営的判断によりいかせそうだということが、共同研究企業様からいただいたアドバイスよりわかった。

そこで、更に短い期間(2, 3 週間)のデータを使用し、予測モデルの構築・性能の評価を行う。提案手法の評価で、1 ヶ月の予測で精度が一番よかった③の説明変数を使用し、各モデルで予測を行った。

予測に使用する日数を変更し、作成したモデルについて、各性能で比較を行った結果、表 4 のようになった。表 4 より、決定木に説明変数③の使用日数

を 21 日に変更したものが 4 つの指標の中で一番精度が良い結果となった。

8. 考察

ランダムフォレストと説明変数③の組み合わせが各性能において高くなった結果について考察を行う。

ランダムフォレストと説明変数③の組み合わせ時の混同行列の結果が表 5 である。

表 5: ランダムフォレストと説明変数③で構築したモデルの混同行列

	Class	予測されたクラス		
		0	1	2
実際の クラス	0	115	1	0
	1	1	240	10
	2	1	10	215

表 5 の混同行列より、Class0 の分類精度がかなりいいことがわかった。また、他の予測手法では Class1 と Class2 の分類精度が悪く、どちらか判別できていないものが多かったが、ランダムフォレストと説明変数③の組み合わせでは、それぞれ誤差が 10 ほどと少なく分類精度がいいことがわかった。このことから、全体的に分類がうまく行っているため、各性能が高いモデルができていることがわかった。

更に短い期間での予測より、決定木に説明変数③の使用日数を 21 日に変更したものが 4 つの指標の中で一番精度が良い結果となったことについて考察を行う。モデルの精度で考えるとランダムフォレストが一番精度の向上が見込まれそうだが、決定木が一番高い結果となった。

表 6: 説明変数③(使用日数: 21 日)を用いて、決定木とランダムフォレストで構築したモデルの混同行列

実際の クラス	Class	予測されたクラス		
		0	1	2
決定木	0	66	43	7
	1	9	221	21
	2	8	35	183
ランダム フォレスト	0	37	69	10
	1	14	189	48
	2	7	51	168

表 6 の混同行列により、Class 分類を比較してみたところ、ランダムフォレストでは Class0 の 1 ヶ月続く商品の分類がうまくいっていないことがわかった。また、ランダムフォレストでは予測に効果的なルールとして、18~21 日の出荷数や店舗数が多く使用されていることが明らかになった。決定木では、Class0 の分類や他の分類もうまくいっていることがわかった。また効果的なルールとして、13 日や 4 日

などの出荷数や店舗数が使用されており、使用される日数の幅が大きく、受注量パターンの違いをうまく学習されているのではないかと考えられる。

9. 終わりに

本研究では、製パン工場の出荷データを利用し、販売開始されてから 1 ヶ月などの短期間のデータ使用し、予測の際に一番精度の高い説明変数とモデルの組み合わせを探す予測手法を提案した。

その結果、ランダムフォレストと説明変数③の製パンの特徴を考慮した変数と店舗数が低いものに補正をかけたモデルの構築が 4 つの指標で一番精度が高い結果となった。また、混同行列により高精度の考察を行った結果、Class0 の分類精度がかなり良く、他のモデルの組み合わせでは精度の悪い Class1 と Class2 の分類精度もよく、全体的に分類がうまく行っているため、各性能が高いモデルの構築ができていることがわかった。

また、更に短い期間(2, 3 週間)のデータで予測を行うことができると経営的判断によりいかせそうだという考えから、更に短い期間での予測を行った結果、決定木に説明変数③の使用日数を 21 日に変更したものが 4 つの指標の中で一番精度が良い結果となった。4 日や 13 日など幅広い期間での出荷数や店舗数で分類が行われ、受注量パターンの違いをうまく学習できたため精度が高いのではないかと考えられる。

今後の課題として、他の分類予測モデルでのさらなる実装・比較、更に短い期間での予測について精度の改善、出荷数や店舗数以外の新たな変数の検討などが挙げられる。

謝辞

分析データは、白石食品工業株式会社様より提供いただきました。心より感謝申し上げます。

参考文献

- [1] Downs G. W., Rocke D. M.: "Municipal budget forecasting with multivariate ARMA model", Journal of Forecasting, Vol. 2, pp. 377-387, 2017
- [2] Sean J.Taylor, Benjamin Letham: "Forecasting at Scale", PeerJ Preprints, 2017
- [3] Tsan-Ming Choi, Chi-Leung Hui, Na Liu, Sau-Fun Ng, Yong-Yu: "Fast Fashion Sales Forecasting with Limited Data and Time," Decision Support System, Vol. 59, pp. 84-92, 2014.