

コールセンターを対象とした業務量予測に関する研究

Towards Improvement of Call Arrival Prediction

田原 琢士*¹ 王 軼謳*¹ 山浦 佑介*¹ 大西 健司*¹
 Takuji Tahara Yiou Wang Yusuke Yamaura Takeshi Onishi

*¹富士ゼロックス株式会社 研究技術開発本部
 Research & Technology Group, Fuji Xerox Co., Ltd.

In order to manage staffing and scheduling effectively and efficiently, accurate forecasting of call arrival volumes is a key challenge for all call centers. In this paper, we propose a method to improving prediction performance by automatic feature generation and increment based on error-driven learning. We evaluate the usefulness of our method in the real call center data set, demonstrate that new features automatically generated by our method provides substantial performance gains in call arrival prediction, and also yield a few important open questions.

1. はじめに

企業のコールセンターは単なる応答窓口ではない。コールセンターは顧客との直接の対話を通して、新規顧客の獲得や顧客の満足度向上、顧客の維持に関わっている重要な拠点である[伊藤 05]。このため、コールセンターでは各エージェントの応対品質の向上及び顧客の問い合わせに耐えうる十分なエージェント数の配置が必要になる。しかしながら、これらはコールセンターの運営上コストがかかる行為でもある。特に後者については、配置するエージェントが少な過ぎればお客様の待ち時間が長くなって顧客満足度の低下につながるが、逆に多過ぎればお客様の待ち時間が少なくなるもののコールセンター側の費用がかかさんでしまう。この両者はトレードオフの関係にあり、バランスをとるためには正確にコール量予測を行うことが必要となる。こういった背景から、企業のコールセンターにおけるコール量の予測は重要なタスクである。

これまでコール量予測のために様々な手法が提案されてきた。一般にコール量は時系列に大きく依存しており、日、週、月、年単位で周期性或いは季節性を示す。この性質を利用して、ARIMA や指数平滑といった一般的な時系列解析モデルをコール量予測に適用する研究が行われてきた[Ibrahim 16]。これらの手法は周期性のあるデータに大きな効果を発揮するが、反面外れ値などに弱いことがわかっている。コールセンターでは通常営業日のコール量の周期性・季節性とは別に、祝日明けや年度始めといったコール量に影響を及ぼす特異な要素(特異日)が存在し、上で述べた一般的な時系列解析モデルにこれらの特異な要素を吸収させることは難しい。この問題の解決法として、特異日を細かくモデリングしたり[Barrow 16, Barrow 16]、特異な値そのものを学習の前段階で人手で全て取り除いたりといったことが考えられる。しかしながら、以上に挙げた手法は確かに精度の向上につながるものの、人手で過去のデータを見ながら外れ値の除去・特異日のモデリングを行わなければならない。またこの作業は時間がかかると共に経験に依存するものとなるため、コールセンターでの予測モデルの構築及び運用が、熟練した個人に依存することになる。

本稿ではこの問題を解決するべく、自動的に予測に影響を及ぼす特異な要素を発見し、予測モデルに素性として追加する手

法を提案する。提案手法では、まず過去のコール量データを学習データと検証データに分割し、予測日の基本属性を素性として学習データから予測モデルを構築、検証データを予測する。次に、予測日の属性に階層性を与えたものから親子関係にある組み合わせ(例えば、月と週など)を生成し、これらの組み合わせごとに予測値と実測値の平均誤差を算出、誤差の大きな組み合わせを新たに素性として追加する。以上の手順を全体の平均誤差が小さくなるまで繰り返すことにより、予測モデルを改善していく。本手法でユーザが与えるのは、過去のコール量データと対応する日の基本的な属性(例:年、月、曜日、休日・祝日か否かなど)と、時間の階層性であり、そのコールセンターに特有の事前知識を持っている必要性はない。すなわち、予測モデルの運用を熟練した個人が行う必要が無くなる。またこの手法はコール量に限らず、様々な需要予測に用いることが可能である。

実験では、実在するコールセンターの5年分のコール量のデータに対して提案手法を適用した。実験の結果、本手法によってお盆や年末といったコール量に大きな影響を及ぼす特異な期間を抽出することができ、それらに対応する素性を自動で生成・追加することによって予測誤差が6.4%減少することを確認し、提案手法の有効性を確認することができた。

2. ベースライン手法

本稿では、提案手法の有効性を検証するためのベースラインモデルに用いるアルゴリズムとして、アンサンブル学習の一種であるGBDT(Gradient Boosting Decision Tree)を選択した。GBDTは他のアンサンブル学習に比べて性能が優れているのが知られており、また素性を追加するのが容易なモデルである。実験では、このアルゴリズムに日付の基本的な属性である年、月、曜日を素性としたものをベースラインとし、提案手法で生成した素性を加えていくことで性能評価を行う。

3. 提案手法

提案手法は以下の二つの部分で構成されている(図1)。

1. 誤差算出部

過去のコール量データを学習データと検証データに分割し、現在保持している素性情報と学習データを用いて検証データを予測、検証データに対する予測誤差を算出する。

連絡先: 田原琢士, 富士ゼロックス(株) 研究技術開発本部,
 横浜市西区みなとみらい6丁目1番,
 tahara.takuji@fujixerox.co.jp

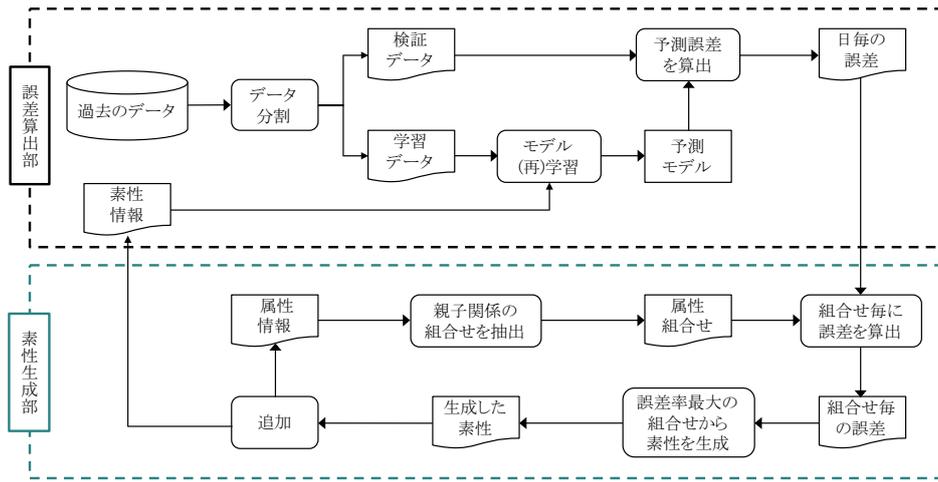


図 1: 提案手法の構成

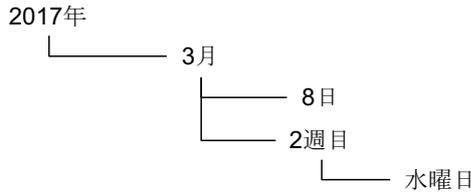
2. 素性生成部

予測日の属性について、親子関係にある属性の組み合わせを追加素性の候補として生成し、平均誤差率が最大となるような組み合わせを現在保持している素性情報と属性情報に追加する。

以上を繰り返すことで、予測誤差に影響のある素性を自動で生成・追加する。以下では本手法で重要な素性生成部の詳細について述べる。

3.1 親子関係にある属性の組み合わせ

素性生成部では、まず、親子関係にある属性の組み合わせを現在保持している属性情報から生成する。ここで、ある日 d の属性の親子関係は時系列の階層性で決定される。例えば 2017 年 3 月 8 日 (水) の持つ属性は以下のような階層性を考えることができる。



このとき、3 月 8 日の属性の親子関係にある組み合わせは (2017 年, 3 月), (3 月, 8 日), (3 月, 2 週目), (2 週目, 水曜日) の 4 つである。このような、ある日 d の属性の親子関係にある組み合わせの集合を以降では $C_{p-c}(d)$ と表す。

3.2 属性の組み合わせごとの平均誤差率の計算

次に、誤差算出部で計算された日毎の誤差を用いて、前節で説明した親子関係にある属性の組み合わせごとに平均誤差率を計算する。ある属性の組み合わせ (f_p, f_c) ごとの平均誤差率とは、組合せ (f_p, f_c) が C_d に含まれているような日 d に対する予測値 P_d と実測値 A_d の平均絶対誤差率のことを指し、以下の式で表される。

$$MAPE(f_p, f_c) = \frac{100}{|D_{(f_p, f_c)}|} \sum_{d \in D_{(f_p, f_c)}} \frac{|P_d - A_d|}{A_d} \quad (1)$$

ここで $D_{(f_p, f_c)}$ は素性の組み合わせ (f_p, f_c) を C_d に含むような d の集合、すなわち

$$D_{(f_p, f_c)} = \{d \mid (f_p, f_c) \in C_{p-c}(d)\} \quad (2)$$

である。

3.3 追加素性の決定及び素性情報への追加

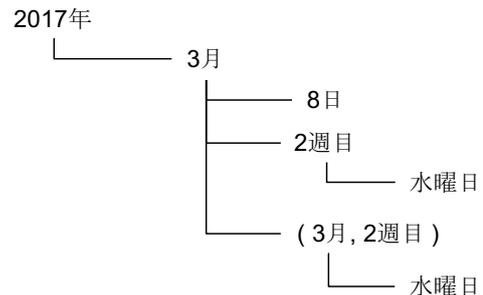
各属性組み合わせに対して平均誤差率を計算したのち、その値が最大となる属性の組、

$$(f_p, f_c) = \arg \max_{f_p, f_c} MAPE(f_p, f_c) \quad (3)$$

を素性として追加する。具体的には、 $\{0, 1\}$ の値をとる新たな素性 f_{new} を生成する。各日 d に対する f_{new} のとる値 $f_{new}(d)$ は、素性の組み合わせ (f_b, f_a) が C_d に含まれるかどうかで決定される。

$$f_{new}(d) = \begin{cases} 1 & (f_p, f_c) \in C_{p-c}(d) \\ 0 & (otherwise) \end{cases} \quad (4)$$

このように生成した素性 f_{new} を素性情報と各日の属性情報に追加したのち、再度誤差算出部で新しい素性情報を用いてモデルの再学習を行い、検証データの予測を行う。ただし追加した素性によって全体の誤差率が上昇してしまう場合は、追加した素性を取り除いてやり直すものとする。またこの新しく追加する素性は、この素性を構成する組合せのうち子の方の属性と全く同じ親と子を持つように属性情報に追加される。例えば、仮に (3 月, 2 週目) という素性が追加される場合、2017 年 3 月 8 日の属性の階層性は以下のように更新され、(3 月, 2 週目) は 3 月を親、水曜日を子に持つ属性となる。



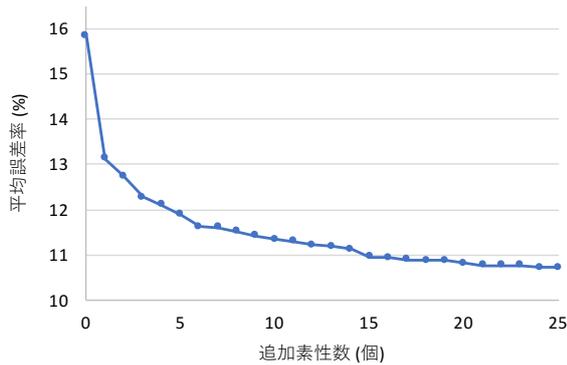


図 2: 素性追加時の検証データの平均誤差率の推移

このように更新を行う理由は、追加した素性を更に組み合わせに用いることで、属性の組み合わせを網羅できることにある。上の例では、(3月, 2週目)と水曜日を組み合わせることで(3月, 2週目, 水曜日)と言う素性の候補を作り出すことが可能となる。このように全ての組み合わせを考えるのではなく、平均誤差率に基づいて生成された属性との組み合わせを考えることで効率化を図っている。

4. 実験

本章では手法の有効性を確認するための実験について述べる。

4.1 データセット

実験に使用するデータは、ある企業の実在するコールセンターの2012年4月から2017年1月までの約5年分の日毎のコール量データである。このコールセンターは通常平日に営業しており、土曜日は限定的に営業、日曜日は完全に休日となっている。本実験では、2012年4月から2016年3月までの4年分のデータを素性生成用のデータとして使用し、残りの10カ月のデータを評価データに使用して性能評価を行う。素性の生成を実行する際には素性生成用の4年分のデータを学習データ3年分と検証データ1年分に分割し、学習データを用いて予測データを予測するが、検証データのある年度に固定するのではなく、各年度ごとのデータを検証データとした場合の予測誤差をそれぞれ計算し、それらを全て用いて属性の組み合わせに対する平均誤差率を計算する。

4.2 実験設定

コールセンターでコール量予測を行うのは、その予測値を用いて適切な人員配置を決定するのが大きな目的である。データ受領先のコールセンターでは月ごとに人員配置を決めており、1か月前には配置を決定する必要がある。このため、本実験では予測月ごとに予測モデルを作成して日毎の予測を行い、その際の学習データには予測月の前月のデータを用いないという制約を与えて性能評価を行う。また、素性を生成する際には誤差算出と素性生成を繰り返し行うことになるが、今回の実験では繰り返し回数の上限を100回とした。

4.3 結果

図2及び図3は、本手法を適用して生成された素性を追加した際の検証データ及び評価データに対する平均誤差率を、素性を追加した順に左から並べたものである(一番左(0)は素性を追加する前の状態である)。図2からわかるように、提案手法による素性の追加によって検証データに対する予測誤差率

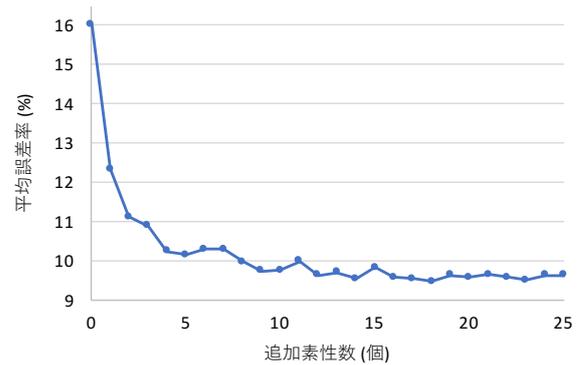


図 3: 素性追加時の評価データの平均誤差率の推移

が追加前より5.1%減少している。一方評価データに対する予測誤差率も、図3からわかるように生成された素性の追加によって減少しており、最終的には追加前より6.4%減少している。この結果から、本手法で生成された素性の追加による予測モデルの精度向上が確認できた。

表 1: 提案手法で生成された素性の詳細

ID	素性の詳細	該当期間
1	{12月, 30日}	年末
2	{12月, 29日}	年末
3	{8月, 14日}	お盆
4	{8月, 15日}	お盆
5	{1月, 4日}	年始
6	{5月, 8日}	GW明け
7	{12月, 月末から1週目, 週末から1日目の営業日}	年末
8	{8月, 16日}	お盆
9	{8月, 12日}	お盆
10	{8月, 11日}	お盆
11	{4月, 28日}	GW前
12	{5月, 月初から1週目}	GW
13	{1月, 5日}	年始
14	{6月, 16日}	不明
15	{12月, 月末から2週目}	クリスマス付近
16	{12月, 22日}	クリスマス付近
17	{3月, 30日}	年度末
18	{10月, 20日}	不明
19	{5月, 月初から1週目, 週末から2日目の営業日}	GW
20	{5月, 19日}	不明
21	{2月, 月初から3週目, 週初めから2日目の営業日}	不明
22	{2月, 月初から3週目, 週初めから4日目の営業日}	不明
23	{2月, 月初から3週目, 週初めから1日目の営業日}	不明
24	{7月, 5日}	不明
25	{10月, 6日}	不明

4.4 考察

表1は、提案手法によって追加された組合せ素性の詳細を示しており、IDは素性が追加された順に割り当てている。追加順が若い組み合わせ、即ち、組み合わせに対する平均誤差率が特に高かった組み合わせを見ると、対象コールセンターでコール量が普段の10分の1以下になる年末やお盆期間、逆にコール量が通常より増加するゴールデンウィーク明けやクリスマス付近の期間を素性として追加していることがわかる。これらは、提案手法によって対象コールセンターでの一年の中での特異な期間をうまく捉えられている例であり、追加によって評価データに対する誤差が減少している。

一方で、追加順が後の方になるにつれて、一年の中での特異な期間とは考えられないものが追加されている。該当期間を”不明”としているもののうち、IDが21, 22, 23の素性は2月の3週目が特異な期間であると捉えているが、これは一年の中での特異性ではなく2012年度の2月のみの特異性であることが実データの確認によってわかっている。それ以外の該当期間を不明としているものは、素性生成に用いたデータ中にこれらの属性を持つものが2日間しかなかったものである。これらの素性は一部の特異な値によって平均誤差率が大きくなり、素性として追加されてしまったと考えられる。

このように、どの素性が生成されるかは誤差の指標に依存しており、例えば平均絶対誤差(MSE)を用いた際には、祝日明けといった今回生成されなかった素性が生成された。誤差の指標に何をを用いるかは今後の検討課題である。また本手法では、検証データの予測誤差が減少するか否かを素性の追加基準としているが、結果として、評価データに対する予測誤差を増加させてしまう素性も追加されている。この追加基準についても、改善の余地があると考えられる。

5. まとめと今後の課題

本稿では、予測に用いるための素性を生成する手法を提案した。具体的には、基本的な属性のみを素性に用いて過去のデータの予測を行い、その誤差に基づいて属性の組み合わせからなる素性を生成・追加し、追加された素性を用いて再度予測誤差を計算するのを繰り返す手法である。組み合わせを考える際には、単純に全ての組み合わせを考えるのではなく、属性間の時系列性にに基づいた階層性を考慮することで効率化を図っている。実験では、生成された素性を追加することで評価データに対する予測誤差が減少し、生成された素性は対象コールセンターにおいてコール量が大きく変動する期間(年末やお盆等)を捉えることができた。これらのことから、予測誤差を用いた素性生成の有効性を確認できたと考えられる。一方で、一部データに大きく影響されたために予測に有効でない素性も生成されてしまっているが、素性の追加基準や指標を改善することで解決できると考えている。

提案手法では時系列の階層性に基づいて属性の組み合わせを考えたが、時系列データにおいてはデータの系列性を考慮することも必要である。日毎の予測であれば、前後の日の属性に影響を受けることは十分に考えられるので、これらを考慮できるように手法を改善する予定である。また、今回はコールセンターのデータを用いて実験を行ったが、本稿で提案した手法はコールセンターだけに特化したものではない。今後、様々なデータに用いることで有効性の検証を行っていく予定である。

参考文献

- [伊藤 05] 伊藤 稔: コールセンターにおけるインバウンド予測, UNISYS TECHNOLOGY REVIEW 第87号 (2005)
- [Ibrahim 16] Rouba Ibrahim, Han Ye, Pierre L'Ecuyer, Haipeng Shen: Modeling and forecasting call center arrivals, A literature survey and a case study, International Journal of Forecasting (2016)
- [Barrow 16] Devon K. Barrow: Forecasting intraday call arrivals using the seasonal moving average method, Journal of Business Research (2016)
- [Barrow 16] Devon Barrow, Nikolaos Kourentzes: The impact of special days in call arrivals forecasting: A neural network approach to modelling special days, European Journal of Operational Research (2016)