

## 自動車購入における検索行動の先行期間を考慮した販売台数の予測

山口 太一† 角田 孝昭†† 吉田 光男††† 津川 翔†††† 山本 幹雄††††

†筑波大学大学院システム情報工学研究科 ††株式会社サイバーエージェント秋葉原ラボ

†††豊橋技術科学大学情報・知能工学系 ††††筑波大学システム情報系

### 1 はじめに

ある月の自動車販売台数をその何ヶ月も前に予測できれば、メーカーや販売店は大きなメリットを享受できる。予測値を参考にすることで生産コスト削減やキャンペーン開催が可能になるからである。将来の販売台数は、毎月の販売台数を時系列データとして捉えて状態空間モデルを適用することで、ある程度予測できる。しかしデータ中に急激なピークやトレンド転換が発生すると、高精度な予測は難しくなる。

我々はこれまで予測精度向上のために、自動車に対しどの程度情報収集が行われているかを数値化した検索行動量というデータを併用した予測モデルを検討してきた [1]。これは自動車に関する検索行動量が、実際の販売台数の変動に対し先行する傾向を持つからである。どの程度先行した（先行期間と呼称）検索行動量を用いれば予測精度が高くなるかは自動車ごとに異なるが、これまでは全ての自動車に対して同一の先行期間を設定していた。本稿ではこれを拡張し、複数の先行期間の中から、自動車ごとに最適な期間を自動で選択するモデルを提案し、予測精度について評価を行う。

### 2 検索行動の先行性

自動車に関する検索行動量は、実際の販売台数の変動に先行して同様の変動を示す傾向にある。本研究では予測対象の22種類の自動車について、検索行動量の先行期間ごとに販売台数と検索行動量のピアソン相関係数を算出し、検索行動量を先行させる場合とさせない場合のどちらが最も相関係数の値が高くなるか予備調査を行った。表1に結果を示す。表1より、検索行動量を先行させる場合の方が、相関係数が最も高くなる自動車が多いことが分かる。これより、検索行動量に先行性が認められる場合に、モデルに検索行動量を与えることで予測精度の向上が期待できる。

表1: 検索行動量を先行させる場合とさせない場合に販売台数との相関係数が最も高くなる自動車の種類数

検索行動量	先行させない	先行させる
Google Trends	6	16
Wikipedia ページ閲覧数	7	15

### 3 提案モデル

本研究では状態空間モデルを用いて予測モデルを構築する。我々のモデルでは時点  $t$  における販売台数  $y_t$  を以下の式 (1) で表現する。

$$y_t = \mu_t + s_t + \beta_t \mu_t^{sv} + v_t \quad v_t \sim N(0, V) \quad (1)$$

式 (1) は、販売台数のトレンド傾向を示す  $\mu_t$ 、季節的な傾向を示す  $s_t$ 、更に検索行動量のトレンド傾向を示す  $\mu_t^{sv}$  に時変の係数  $\beta_t$  を掛けた値、誤差項  $v_t$  で構成される。 $\mu_t$ ,  $s_t$ ,  $\mu_t^{sv}$ ,  $\beta_t$  は式 (2) から式 (5) で表現する。

$$\mu_t = 2\mu_{t-1} - \mu_{t-2} + w^\mu \quad w^\mu \sim N(0, W^\mu) \quad (2)$$

$$s_t = -\sum_{l=1}^{11} s_{t-l} + w^s \quad w^s \sim N(0, W^s) \quad (3)$$

$$\mu_t^{sv} = 2\mu_{t-1}^{sv} - \mu_{t-2}^{sv} + w^{\mu^{sv}} \quad w^{\mu^{sv}} \sim N(0, W^{\mu^{sv}}) \quad (4)$$

$$\beta_t = \beta_{t-1} + v^\beta \quad v^\beta \sim N(0, V^\beta) \quad (5)$$

本研究で考慮する検索行動量の先行期間は、0ヶ月（先行させない）、1ヶ月、2ヶ月、3ヶ月である。どの程度先行した検索行動量を用いるかにより式 (1) の  $\mu_t^{sv}$  は  $\mu_{t-1}^{sv}$ ,  $\mu_{t-2}^{sv}$ ,  $\mu_{t-3}^{sv}$  と変化する。選択する先行期間は、それぞれの先行期間の検索行動量を組み込んだモデルにおいてディベロップメントデータを用いてRMSE (Root Mean Square Error) を算出し、その値が最も小さくなるものとする。また本研究では、Google Trends を用いるモデル、Wikipedia ページ閲覧数を用いるモデル、両方を併用するモデル、どちらの検索行動量を用いるかも選択するモデルの4パターンを考える。

### 4 実験と評価

#### 4.1 実験に用いるデータ

実験では我々のこれまでの研究 [1][2] で用いた新車登録車台数、Google Trends、Wikipedia ページ閲覧数の

Car Sales Prediction Considering the Interval between Search and Purchase

†Taichi YAMAGUCHI ††Takaaki TSUNODA †††Mitsuo YOSHIDA ††††Sho TSUGAWA ††††Mikio YAMAMOTO

†Graduate School of Systems and Information Engineering, University of Tsukuba ††Akihabara Lab., CyberAgent, Inc

†††Department of Computer Science and Engineering, Toyohashi University of Technology ††††Faculty of Engineering, Information and Systems, University of Tsukuba

3種類のデータを用いた。新車登録車台数は日本自動車販売協会連合会と全国軽自動車協会連合会が毎月公表しているデータから、22種類の自動車について月ごとの値を取得した。Google Trends は22種類の自動車について「メーカー名+自動車名」をクエリとして月ごとの値を取得した。Wikipedia ページ閲覧数は22種類の自動車それぞれの記事への1時間ごとの接続リクエスト数を取得し、月ごとに集計した。3種類のデータは全て2010年1月から2014年12月までの値を用いた。

#### 4.2 実験条件

取得した3種類のデータを学習データ（2010年1月から2012年12月まで）、適した先行期間を決定するためのディベロップメントデータ（2013年1月から2013年12月まで）、テストデータ（2014年1月から2014年12月まで）に分割し、1ヶ月先の販売台数を予測する1期先予測を以下の手順で行った。

1. 2012年12月までの販売台数とその*i*ヶ月前 ( $i = 0, 1, 2, 3$ ) までの検索行動量を用いてパラメータを学習し、ディベロップメントデータを用いて予測を行い、RMSEを算出する
2. RMSEにより選択する先行期間 ( $=i$ ) を決定する
3. 2013年12月までの販売台数と先行期間*i*ヶ月の検索行動量を用いて改めてモデルのパラメータを学習し、テストデータを用いて予測を行い、最終的なRMSEを算出する

モデルの予測精度はRMSEを算出して評価した。但し自動車ごとに販売規模に差があることを考慮し、RMSEは実際の販売台数に対する予測値の比率を求め、相対的な値として評価した。モデルの予測精度は22種類の自動車のRMSEの平均値とした。なお学習するパラメータの初期値は  $V, W^{\mu}, W^s, W^{\mu^v}, V^{\beta}$  を1、状態の初期値が従う正規事前分布の平均を0、状態の初期値が従う正規分布の分散を  $10^7$  に設定した。

#### 4.3 実験結果と考察

実験の結果算出された各モデルのRMSEの比較を図1に示す。図1は左から、検索行動量を組み込まないモデル、自動車ごとに最もRMSEが小さくなる先行期間を選択した理想的なモデル（オラクルと呼称）、Google Trendsを組み込んだ提案モデル、Wikipedia ページ閲覧数を組み込んだ提案モデル、両方を組み込んだ提案モデル、組み込む検索行動量も選択する提案モデル、のグラフである。オラクルにおけるRMSEは提案モデルが取りうるRMSEの下限であるため、提案モデルのRMSE

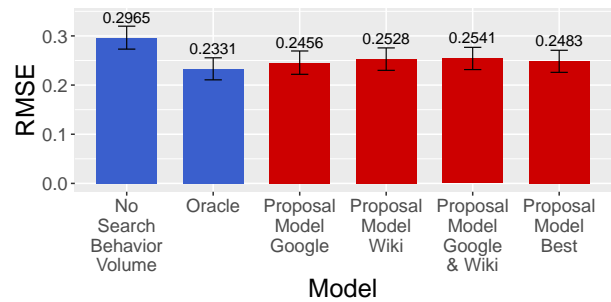


図1: (左から) 検索行動量無し, オラクル, 提案モデル (4パターン) のRMSEの比較

がこの値に近いほど優れた予測モデルであることを意味する。

結果に対してはブートストラップ法を用いた有意差検定を行い、4種類の提案モデル全てにおいて、検索行動量を組み込まないモデルに対し有意に予測精度が向上したことが、オラクルとの予測精度に有意差が無いこと、の2点を確認した。これより提案モデルは、検索行動量を組み込まない場合より有意に予測精度が向上するだけでなく、モデル自身に自動で先行期間を選択させても理想的なモデルであるオラクルに匹敵するだけの予測精度を実現できるモデルだと結論付けられる。

## 5 おわりに

本研究では検索行動量を組み込んだ状態空間モデルを用いて将来の自動車販売台数を予測するモデルを提案した。実際の販売台数の変動に対して先行性を持つ検索行動量を時変の係数を掛けて組み込み、更に自動車ごとに適した先行期間をモデルに選択させることで、検索行動量を用いない場合と比較して有意に予測精度を向上させることができた。更に提案モデルは理想的なモデルであるオラクルに匹敵するだけの予測精度を持つことを確認し、自動で先行期間を選択させても高精度な予測が可能なモデルを実現できた。

## 参考文献

- [1] 角田孝昭, 吉田光男, 津川翔, 山本幹雄. 状態空間モデルを用いた検索トレンドとページビューからの自動車販売台数の予測. 人工知能学会全国大会論文集, Vol. 29, pp. 1-4, 2015.
- [2] 吉田光男, 角田孝昭, 山本幹雄, 荒瀬由紀, 津川翔. トレンドと検索頻度とWikipedia ページビューの三角関係. 人工知能学会インタラクティブ情報アクセスと可視化マイニング研究会 (第10回), 2015.