

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,p1-4,2015

角田孝昭, 吉田光男, 津川翔, 山本幹雄

蘆田 誠人

背景

- 自動車は高額商品である
 - ▶ 購入前に綿密に検索するため, 「検索行動量」が販売台数に反映される
 - ✓ 検索行動量: Webでどれだけ情報収集をしたか
- 検索行動量を用いた売上予測を行う研究がある

関連研究

- 自動車及び自動車部品の売上予測^[1]
 - 広告クリック率の予測^[2]
 - 通話料収入の予測^[3]
- ▶ 上記三つの予測は、時系列データの将来予測に対して強力である状態空間モデルを用いる

[1]Choi, H. and Varian, H.: Predicting the present with google trends, Economic Record, Vol. 88, No. s1, pp. 2-9 (2012)

[2]本橋 永至, 磯崎 直樹, 長尾 大道, 樋口 知之:状態空間モデルによる インターネット広告のクリック率予測, オペレーションズ・リサーチ:経営の科学, Vol. 57, No. 10, pp. 574-583 (2012)

[3]矢田 健, 井上 正之, 北川 源四郎:カルマンフィルタによる通話料収入予測, 電子情報通信学会技術研究報告. IN, 情報ネットワーク, Vol. 93, No. 23, pp. 43-50 (1993)

状態空間モデル

- もとは物理システムの記述に用いられていた
- 観測方程式と状態方程式に分類
 - 観測方程式：時系列モデルと同様
 - 状態方程式：関連する何らかの事象についての方程式

研究目的

- 状態空間モデルを自動車販売台数の予測へ適用するとともに、検索行動量を活用し予測精度の向上を試みる

データセット

- 販売台数データ
 - ▶ 日本の自動車(軽も含む)22車種
 - ▶ 2010年1月～2015年2月の販売台数データ
- 検索行動量
 - ▶ Google Trends, Wikipedia閲覧数

モデル構築

- ベースラインモデル
 - 回帰モデルとほぼ同じ

$$y_t = \mu_t + \gamma_{1,t} + v_t$$

v_t : 時点tの誤差項

y_t : 時点tの自動車販売台数

μ_t : 時点tのトレンド

$\gamma_{1,t}$: 時点tの季節成分

検索行動量と販売台数の相関

- 8車種の相関関係

車種	Google Trends			Wikipedia PV		
	shift0	shift1	shift2	shift0	shift1	shift2
Demio	0.65	0.67	0.67	-0.12	-0.10	-0.10
eK	0.84	0.77	0.69	0.62	0.58	0.54
Freed	0.81	0.85	0.83	0.50	0.44	0.35
Mira	0.64	0.62	0.57	0.79	0.84	0.89
Move	0.88	0.84	0.80	0.67	0.66	0.64
Note	0.97	0.94	0.90	0.90	0.92	0.92
Passo	0.90	0.82	0.74	0.27	0.22	0.16
WagonR	0.72	0.71	0.69	0.19	0.28	0.36

- 相関が高い関係についての予測

- 検索行動量を用いると精度が良くなるのでは？

モデル構築

- 提案モデル
 - ベースラインモデルに検索行動量を考慮した予測モデル

$$y_t^{s1} = \mu_t^{s1} + \gamma_{1,t}^{s1} + \alpha_0 \mu_t^{s2} + \alpha_1 \mu_{-1,t}^{s2} + \alpha_2 \mu_{-2,t}^{s2} + v_t^{s1}$$

v_t : 時点tの誤差項

y_t : 時点tの自動車販売台数

μ_t : 時点tのトレンド

$\gamma_{1,t}$: 時点tの季節成分

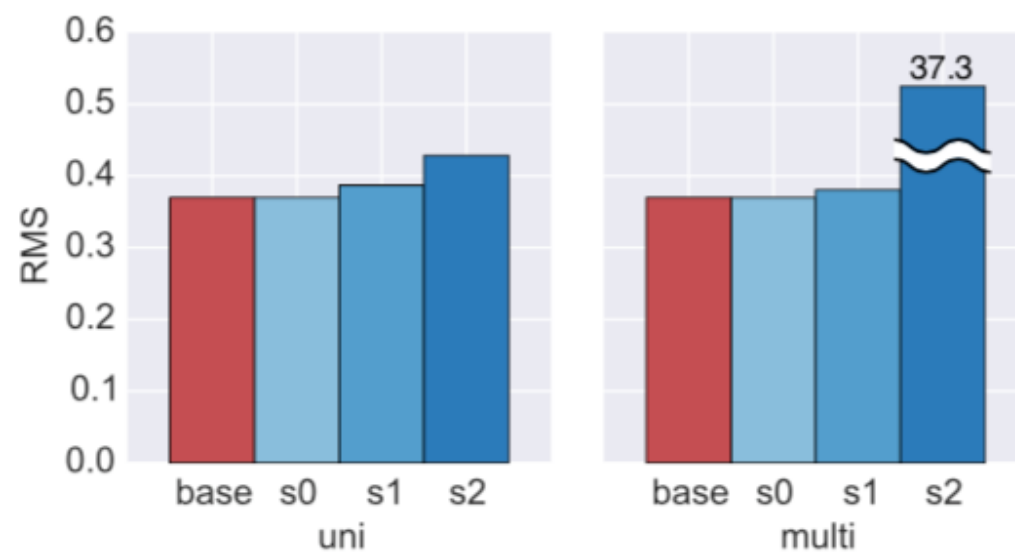
$\alpha_{0,1,2}$: {0,1,2}期前のトレンドの重み

実験方法

- 新車販売台数と検索行動量の比較
 - 予測した値と実測値の相対誤差で評価

実験結果

- uniモデル：一定期間シフトさせた検索行動量トレンドを用いる
- multiモデル：その期間までの検索行動量トレンドを加味する

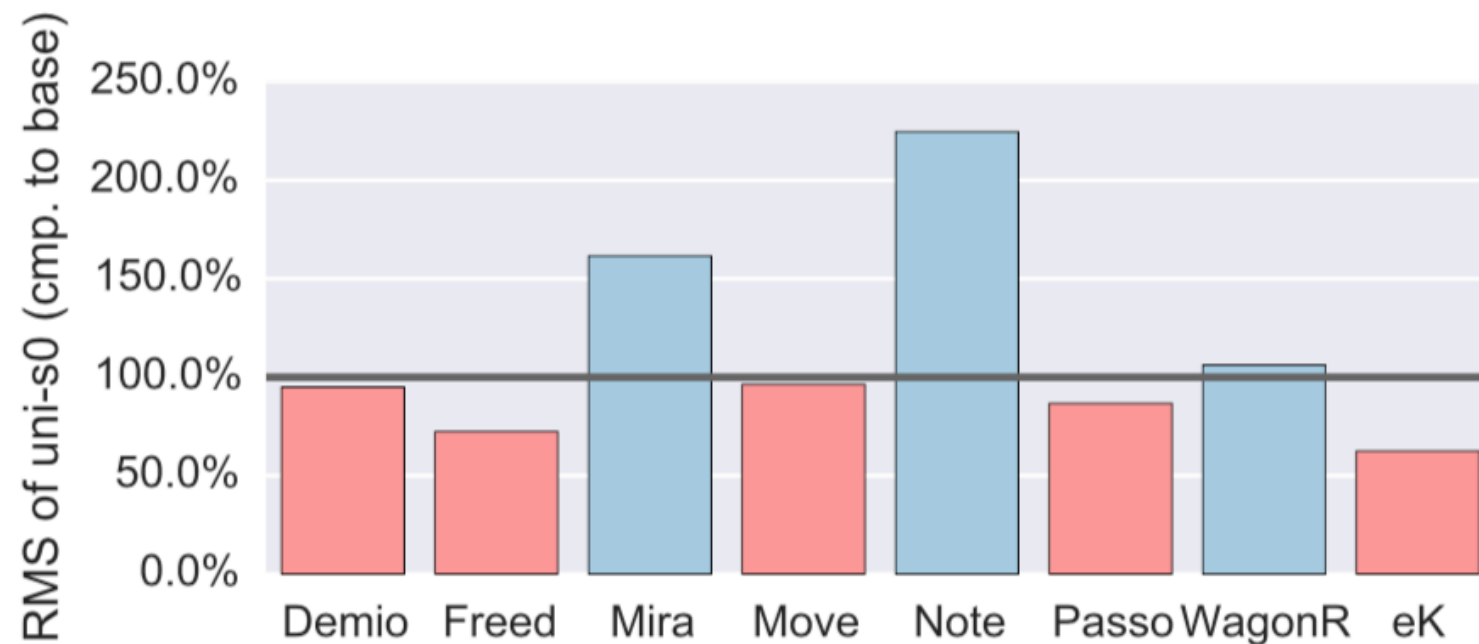


(a) シフト期間固定・1 期先予測 (Google Trend)

- シフトを考慮しないモデルの方が予測精度はよい
 - 車種によってシフト期間が異なるため，固定すると精度がぶれる

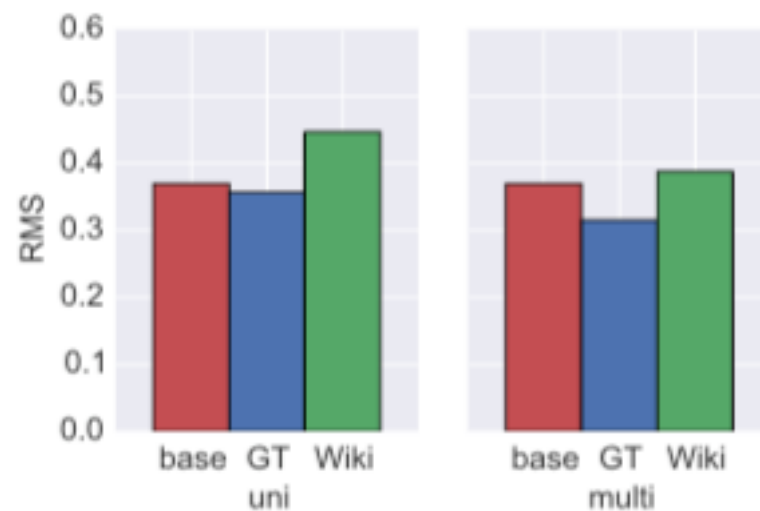
実験結果

- シフト0で固定した時の予測モデルの比較
 - フリード, パッソ, eKでは軽減
 - × ミラ, ノート, ワゴンRは悪化
- 結果, 平均すると予測精度はプラマイゼロ

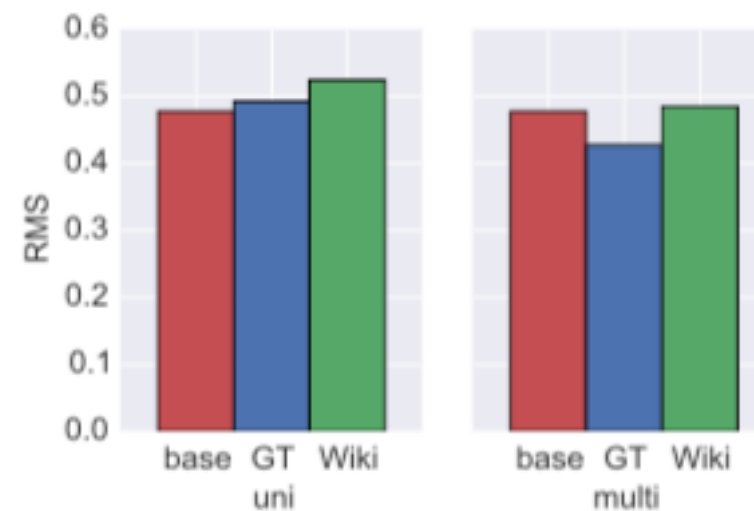


実験結果

- 車種ごとに最適なシフトを行った場合の比較
 - ▶ Google Trendを用いたモデルが比較的精度が良い
 - ▶ 一期先のmultiモデルは約15%精度が向上
 - Wikipediaを考慮したモデルは精度を下げた



(b) シフト期間可変・1期先予測



(c) シフト期間可変・2期先予測

まとめ・所感

- まとめ

- ▶ 状態空間モデルを用いて自動車販売予測を行った
- ▶ 検索行動量の一つであるGoogle Trendsと実際の売れ行きを考慮したモデルを用いると精度が向上する

- 所感

- ▶ 自分にとってまた新しい時系列モデルなので、更に勉強したいと思った
- ▶ 最適なシフト期間を決定した方法を知りたかった