

オンラインレビューから抽出した消費者の感情に寄与する素性を用いた自動車販売予測

野中 尚輝^{1,a)} 中山 浩太郎¹ 松尾 豊^{1,†1,b)}

受付日 2015年3月4日, 採録日 2015年8月1日

概要: 企業にとって、自社の販売する商品の販売予測といった長期的なトレンド情報を把握することは重要である。近年、インターネットやスマートフォンの普及によりオンラインレビューサイトが利用されており、消費者の口コミ情報を収集することが容易になった。このような背景のもと、将来予測においてウェブ上のデータを用いる研究が行われている。株価の予測では、ウェブ上の記事の中から専門家が指定した単語と株価が相関することが示されているが、この手法には単語の選択におけるコストという問題があった。一方、先行研究から口コミ情報に対して感情分析を行った結果を用いることで株価や商品の販売予測を行えることが示されている。そこで本研究では、感情分析を行うことで口コミ情報に含まれるユーザ感情の予測に寄与する単語素性を選別し、選別された単語素性を用いて販売数の予測を行う手法を提案した。具体的には、レビューサイトから消費者の商品に対する感情を抽出し分析を行い、有用な特徴量を販売数予測問題に適用する。提案手法の検証実験は、代表的な耐久消費財である自動車の販売台数を対象として行った。

キーワード: 感情分析, オンラインレビュー, 販売数予測

Automobile Sales Prediction using Features Extracted from Online Reviews that Contributes to Consumer Sentiment

NAOKI NONAKA^{1,a)} KOTARO NAKAYAMA¹ YUTAKA MATSUO^{1,†1,b)}

Received: March 4, 2015, Accepted: August 1, 2015

Abstract: For companies, it is important to grasp long-term trend information such as sales prediction of products sold by the company. In recent years, the online review site has been used due to the spread of the Internet and smart phones, making it easy to collect consumers' word-of-mouth information. Under these circumstances, research using data on the web is carried out in the future prediction. In the prediction of stock price, it is shown that the word specified by experts correlates with the stock price from articles on the web, but this method had a problem of cost in word selection. On the other hand, it has been shown that stock price and product sales prediction can be done by using the result of sentiment analysis on word of mouth information from previous research. In this research, we propose a method to predict sales numbers by selecting word features that contribute to prediction of user sentiments included in word-of-mouth information by conducting sentiment analysis. Specifically, we extract and analyze sentiments of consumers toward products from the review site, and apply useful features to sales prediction task. As a verification experiment on the proposed method, sales of automobiles, which are representative durable consumer goods, were conducted.

Keywords: Sentiment analysis, Online review, Sales prediction

¹ 情報処理学会
IPSJ, Chiyoda, Tokyo 101-0062, Japan

^{†1} 現在, 東京大学
Presently with The University of Tokyo

^{a)} nonaka@weblab.t.u-tokyo.ac.jp

^{b)} matsuo@weblab.t.u-tokyo.ac.jp

1. はじめに

企業にとって自社の販売する商品の販売予測といった長期的なトレンド情報を迅速かつ正確に把握することは、マーケティング戦略を立案する上で重要な課題である [1]. 将来予測を行う、株価の予測や商品の販売予測といった研究では、これまで多くの場合、過去の変動や季節的な変動が用いられてきた。最近では、過去の変動パターンに加えて、消費者の意見や会社・製品に対する感情をはじめとする口コミ情報を考慮することで、商品の販売予測の精度を向上できることが示されている [2].

近年、インターネットやスマートフォンが普及したことにより、ソーシャルメディアへの投稿、ニュース記事、ユーザレビューといった様々な情報がウェブ上に大量に存在するようになった。こうしたウェブ上のデータは、収集や解析が比較的容易であるため将来予測における素性として用いられており、多くの研究でその有用性が示されている [2, 3]. 株価の予測では特定の語の出現頻度 [4] やソーシャルメディアにおける特定のハッシュタグの出現頻度をはじめとする素性 [5] が有用であることが知られている。[4] では、ウェブ上の記事に含まれる特定の単語の出現回数と株価が相関することが示されている。しかしながら、株価との相関が考えられる単語の選別は専門家によってなされており、辞書の構築コストという問題点があった。

一方、[6] や [7] において、特定のタスク（元タスク）で学習された特徴表現を別のタスク（目標タスク）に用いることで、単体で学習したときと比較して学習の性能が向上することが示されている。特に [7] では、別タスクで学習された特徴表現を目標タスクで用いることで、目標タスクにおける予測精度を向上できることが示されている。また、商品の販売予測においては、レビューをはじめとする口コミ情報から得られるユーザの感情を過去の変動パターンと組み合わせて用いることで、対象とする商品の販売予測精度が向上することが知られている [8, 9]. 口コミ情報に対する感情分析タスクでは、感情を予測する際の素性としてレビューに含まれる単語を用い、感情の予測に寄与する単語素性を特定することが可能である。以上のことから、感情分析タスクにおいて学習された単語素性の中から、感情予測に対する寄与度を指標として単語素性を選別し、販売予測タスクに用いることで予測精度を向上できると考えられる。

そこで、本研究では、ウェブ上のデータを利用した商品の販売数予測の精度向上を目指し、ユーザの感情予測に有用であった単語素性を利用する方法を提案する。本手法は、図 1 に示すように、口コミ情報に対する感情分析において学習された素性を販売数予測タスクで利用する。具体的には、レビューに対して行った感情分析から得られた素

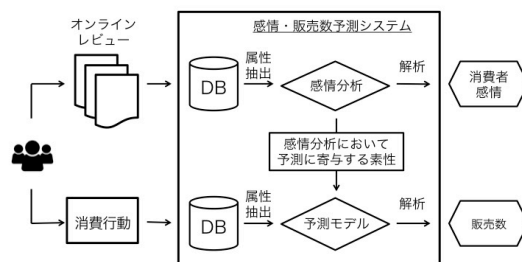


図 1 提案手法の全体像

性と季節性を考慮した過去の販売情報を用いて、代表的な耐久消費財である自動車の月次の販売数を予測する。自動車に関するレビューサイトからユーザの投稿を収集し、28 の車種について 2011 年 10 月から 2015 年 7 月までの期間のデータを用いて販売予測を行った。

本論文の貢献は、自動車のレビューについての感情分析において消費者感情の予測に寄与する素性が、自動車の販売数予測において有用であることを示したことである。

本論文は以下のように構成される。2 章では、本研究との関連性の高い感情分析と販売数予測に関する研究について紹介する。3 章では、提案するモデルについて説明する。4 章にて、モデルの検証のために行った実験とその結果について記述する。5 章に、得られた結果に関する考察を置き、最後に 6 章にて、研究の総括と今後の展望について述べる。

2. 関連研究

2.1 感情分析

感情分析は、特定の対象物に対する書き手の意見を見つけるタスクと定義される [10]. 商品の購入をはじめとして、人々は意思決定を行う過程において、他者の意見に影響される。そのため商品に対して消費者が抱く感情を知ることが、売り手にとって重要である。こうした背景から感情分析に関連する研究は、近年大きな注目を集めてきた。

感情分析のアプローチとしては、辞書を用いる手法とラベル付きデータから学習する手法に大別される。辞書を用いる手法では、あらかじめポジティブおよびネガティブな語を登録しておき、それらを用いて与えられた文章の感情を測定する [11]. 既存の辞書を用いる場合、収集したデータに対して容易に適用できることが多いが、そのような辞書には一般的な語しか含まれないため、適用するドメインに特異的な語を評価することはできない。適用するドメインに特異的な語を考慮して感情分析を行う場合、タスクに適した形で辞書を作成する必要がある。そのコストは非常に大きい。一方、ラベル付きデータから学習する手法では、収集したデータに対して機械学習を行い、感情分析を行う [12-15]. 機械学習により感情分析を行う場合、学習を行

う際に感情を表すラベルが必要となり、問題設定によってはラベルデータの入手が困難な場合もある。しかしながら、多くのレビューサイトではユーザの投稿するレビュー文章と合わせてユーザによる対象物に対する採点結果が公開されている。ユーザによる採点は、商品に対する評価を数値化したものであるため、それを感情分析における教師ラベルとすることが可能であると考えられる。そこで、本研究ではレビューサイトから収集したデータに含まれるユーザ採点をラベルとして機械学習により感情分析を行った。

感情分析を販売予測や株価の予測に応用した研究として、映画のレビューに対して感情分析を適用し各映画の売上を予測する研究 [9] や、Twitter の文章に対して感情分析を適用し、その結果をもとに株価を予測する研究 [3]、自動車に関するウェブ上の情報を収集し、収集したデータに対して感情分析を適用した後、販売予測に用いる研究 [8] が存在する。これらの先行研究では、あらかじめ学習された分類器またはあらかじめ作成された辞書を用いて感情分析を行った後、販売数予測を行っている。しかしながら、機械学習により感情分析を行う場合と比較して辞書を用いる手法の精度は必ずしも優れているわけではない [12]。また分析対象のドメインに特異的な語を含む辞書の作成には、コストがかかるといった問題点がある。そこで本研究では、レビューデータに含まれるユーザ採点の結果をラベルとして機械学習による感情分析を行い、販売予測を行った。

2.2 販売数予測

商品の販売予測は、企業が意思決定を行う際に重要となる。予測を行う際には、過去の販売データをはじめとする統計指標のみでなく、消費者の意見や会社・製品に対する感情といった口コミ情報を考慮することが重要である。すでに複数の研究において、口コミ情報が商品の販売と関連することが示されている。口コミ情報の収集は、アンケートをはじめとする手法が一般的であったが、インターネットやスマートフォンの普及とともに、ウェブ上からデータを取得することが可能になっている。中でも、オンラインレビューサイトや各種 SNS が普及したことから、それらの情報を用いて商品の販売数を予測する研究が数多く試みられている。

販売予測を行う際に用いられるデータとしては、ブログや Twitter、レビューなどの個人の発信する情報、検索クエリの回数といったユーザの行動履歴などがある。このようなデータを用いて販売数予測を行う研究としては、特に豊富なデータ数と明確な利用目的という観点から、映画および株価のデータを利用した研究が活発に行われてきた。Twitter のデータを用いる研究 [16]、ブログから得た情報を用いた研究 [17] が知られており、レビューの情報を用いた研究としては [3] がある。映画以外のドメインにおけるウェブ上のデータを用いた販売予測の研究としては、検索

クエリの回数から自動車の販売数をはじめとする経済動向を推測した [18] の論文やブログにおける商品の言及回数を素性として、その商品の販売数を予測する研究が存在する。株価の予測を行う研究としては、企業の有価証券報告書をはじめとする公式発表された書類をデータソースとする研究が存在する [19]。

ウェブ上から取得したデータを直接販売予測に用いるのではなく、感情分析を行った後に販売予測に用いる研究も存在する。[9] では、映画の売上予測を行う際にブログに対して行った感情分析の結果を素性として加える事で予測精度が向上することを示している。また、レビューにおける商品の言及回数といった単純な素性だけでなく、消費者の感情を素性として加えることで販売数予測の精度を向上できることが知られている [8]。このように感情分析により消費者の感情を定量的に評価し、販売予測に用いることで予測の精度が向上することが先行研究にて示されている。感情分析の結果が販売予測に有効であることから、感情分析において予測に対する寄与度の高い素性は販売予測においても有効であると考えられる。しかしながら、先行研究における感情分析はあらかじめ用意した辞書を用いていたため、予測に用いられる素性は一般的な語でしかなかった。本研究では、ウェブ上から収集したレビューデータに対して機械学習を用いた感情分析を行い、感情の予測に寄与する素性を用いて販売予測を行ったことがこれまでの研究と異なっている。

2.3 属性選択

本研究では、レビュー文に対する感情分析を行い、感情の予測において有用であった単語素性を用いて販売数予測を行なっている。本節では予測に用いる素性の選択と感情分析の関わりについて述べる。

文章分類における属性選択では、[20] らによるサーベイにおいて様々な属性選択の手法と文章分類精度の関係が調べられている。文章分類の中でも、特に感情分析に関わる素性の選択の研究には、[21] や [22] が存在する。[21] では、情報利得と遺伝的アルゴリズムを組み合わせた手法により感情分析に用いる素性の選択を行なっている。また、[22] では、レビュー文に含まれる主観的な表現に注目し、素性の選択を行なっている。これらの研究は、感情分析を行う際に有用な単語素性の選択を行う手法を提案している。

一方、[6] では、特定のタスク（元タスク）で学習された特徴表現が別タスク（目標タスク）にて用いる枠組みを示されている。また [7] は、実際にニューラルネットワークの中間層における特徴表現が別タスクにおいて予測に有用であることを示した。本研究では、感情分析タスクにおいて有用だった素性を別タスクである販売数予測タスクにおいて用いる点がこれまでの研究と異なっている。

3. 提案手法

この章では、商品の販売数と相関関係にある口コミ情報に対して感情分析を行い、その結果を用いて販売数を予測する提案手法について説明する。口コミ情報を用いることで販売予測の精度が向上する [2] ため、口コミ情報には商品の販売を予測する上で有用な素性が含まれることが考えられる。提案手法では、感情分析において予測に対する寄与度の高い因子を抽出し、商品の販売数を予測する。

3.1 感情分析

まず、レビューデータに対する感情分析について述べる。感情分析とは特定の対象物に対する書き手の意見を分けるタスクであり [10]、感情分析を行うことにより大規模なレビューデータから消費者が商品に対して抱いている感情を明らかにすることが可能である。対象とする商品を絞って感情分析を行う場合、対象とした商品に関連する特徴的な単語が感情分析において重要になると考えられる。そこで提案手法では、与えられたレビュー中に含まれる情報を教師ラベルとして分類器を学習するアプローチを採用する。提案手法では、感情分析において予測に対する寄与度の高い因子を販売数予測問題において入力として与える。そこで、ラベルの予測に対する各入力素性の寄与度が算出される分類器を用いた機械学習の手法を採用する。本研究では単純ベイズ分類器を用いて感情分析を行う。

対象とするレビューに含まれる単語 $\{w_1, w_2, \dots, w_{|V|}\}$ の集合を V とする。全レビュー数を N とし、 j 番目のレビューを d_j とし、 $n_i(d_j)$ をレビュー d_j に単語 w_i が出現する回数とする。以上を用いて、レビュー d の bag-of-words 表現 $\mathbf{d} := (n_1(d), n_2(d), \dots, n_{|V|}(d))$ を得る。単純ベイズ分類器では、ある文章 d が与えられた場合のクラスラベル $c^* = \arg\max_c P(c|d)$ を求める。ベイズの定理より、

$$P(c|d) = \frac{P(c)P(\mathbf{d}|c)}{P(\mathbf{d})} \quad (1)$$

が得られる。各単語 w_i の出現は独立であるという仮定をおくことで、 d のクラス分類を以下のように表わせる。

$$P(c|d) = \frac{P(c)(\prod_{i=1}^{|V|} P(w_i|c)^{n_i(d)})}{P(\mathbf{d})} \quad (2)$$

ユーザ採点をもとにレビュー d_j が肯定的か否定的かを表すラベル $y_j \in (0, 1)$ を得る。 $X = \{d_1, d_2, \dots, d_N\}$ および $y = \{y_1, y_2, \dots, y_N\}$ として、分類器に (X, y) を入力として与える。

学習された分類器は、入力として与えられた文章の bag-of-words 表現 \mathbf{d} に含まれる単語の出現回数 $(n_1(d), n_2(d), \dots, n_{|V|}(d))$ をもとに感情を表すクラスラベル y を予測する。分類器において学習されるクラスラベル c が与えられた際の各単語についての条件付確率 $P(w_i|c)$ から対数オ

Algorithm 1 Feature Extraction From Classifier

Input: trained classifier F , vocabulary set $V = [w_0, \dots, w_{|V|}]$
Output: selected words W_s
for $i = 0$ to $|V|$ **do**
 $p_i \leftarrow$ extract coefficient values for w_i from F
end for
 $W_s \leftarrow$ top K words order by p value
return W_s

ズ比を算出する。対数オッズ比の大きい単語素性は、ユーザの肯定的感情に付与されたラベルである正例の予測に対する寄与度が大きい。本研究では、これら対数オッズ比の大きい単語素性を、消費者の感情を示す素性であると考え、販売予測に用いる。

3.2 有用な素性の抽出

本節では、感情分析タスクにおいて学習した分類器から予測に有用な素性を抽出する方法について説明する。

入力として与えられたレビューを肯定的と否定的の二値に分類する分類器 F を用いる。アルゴリズム 1 に示すように、学習に用いた全語彙の集合 V の中から、学習済みの分類器 F において予測に対する寄与度の大きい単語を K 個選択し集合 W_K を得る。続いて、ある月 $t \in T$ における対象とする商品のレビュー中に存在する $u \in W_K$ の出現回数を算出し、 $c_u^{(t)}$ とする。 W_K に含まれる単語 u それぞれについて $c_u^{(t)}$ を算出し、販売数予測問題における入力とする。

感情分析において肯定的なレビューの予測に対する寄与度の高い因子は、ユーザの肯定的な感情を表す因子であると考えられる。したがってこれらの因子がより頻繁に現れる場合、ユーザはその商品に対してより強く肯定的な感情を抱いていると考えられる。ユーザによる評価が肯定的な商品はよく売れると考えられるので、抽出した因子の出現回数を販売数予測問題における素性として用いることは効果的であると考えられる。そこで感情分析タスクにおいて学習した分類器 F の中で予測に対する寄与度の大きさをスコア付けし、スコア上位 K の単語を選択し、販売数予測問題に用いる素性とする。

3.3 販売数予測

続いて、対象とする商品の販売数予測について述べる。商品の売り手にとって、商品のその後の売れ行きを予測することは、マーケティング戦略の決定をはじめとする意思決定における重要な判断材料となる。したがって、商品の販売数予測は重要な問題である。一般的に、販売数の推移は連続的であり、将来の販売数を予測するタスクにおいては直近の販売数が重要な素性になる。そこで、予測対象とする商品についての直近の販売数 S を素性として加える。また商品によっては、時期に応じた販売数の変動が見られる場合もあるため、月情報を示す素性 M を追加する。あ

る月 t における直近の販売数 $S^{(t)}$ および月情報 $M^{(t)}$ 、感情分析における寄与度の大きい素性上位 K 件の情報 $C_K^{(t)}$ を入力として与え、 t から i ヶ月後の販売数 $S^{(t+i)}$ を予測する。

$$S^{(t+i)} = f(S^{(t)}, M^{(t)}, C_K^{(t)}) \quad (3)$$

$S^{(t)}$ は実際の販売数を用い、月情報 M は 12 次元の 1-of-k ベクトルにて表す。 $C_K^{(t)}$ は、ある月 t に投稿されたレビューに含まれる感情分析における寄与度の大きい単語上位 K 件、それぞれについての出現回数 $c_n^{(t)}$ を要素とするベクトル $C_K^{(t)} = [c_1^{(t)}, c_2^{(t)}, \dots, c_K^{(t)}]$ とする。ある月 t における $S^{(t)}$ 、 $M^{(t)}$ 、 $C_K^{(t)}$ を結合し入力素性とし、 i ヶ月後の販売台数 $S^{(t+i)}$ を正解ラベルとして学習モデルに与える。

4. 実験・結果

本章では、提案手法の有効性を示すための実験を行う。本実験では、代表的な耐久消費財である自動車を対象とする。自動車のレビューに対する感情分析から得られた素性を販売数予測問題に適用し、素性を用いない場合と比較することで提案手法の有効性を示す。

4.1 データ

4.1.1 自動車販売数データ

自動車についての月次の販売台数データは、日本自動車販売協会連合会（自販連）^{*1}より、毎月の販売台数上位 30 位以内の自動車について取得した。また自販連サイトにて入手できない車種については、car-lineup.com に記載されている販売台数を用いた。なお、両サイトにおいて情報が入手可能な例については、両サイトに記載されている販売台数が一致することを確認した。販売台数のデータは、2011 年 11 月から 2015 年 8 月までの期間のデータを収集した。なお、各月の販売台数は、翌月の上旬に公開されるため、例えば 2011 年 11 月のデータは 2011 年 10 月 1 日から 31 日までの期間に販売された当該車種の合計台数となる。

4.1.2 オンラインレビュー情報

自動車のレビュー情報は、国内最大規模の自動車レビューサイトである GooNet^{*2}から取得した。レビューデータは、サイトにて閲覧可能な 2011 年 11 月からデータ収集を行った 2015 年 7 月末日までの期間にユーザにより投稿された 57,466 件を対象とした。

図 2 に投稿されるレビューの具体例を示す。ユーザの投稿には、レビュー対象の車種名、ユーザ名、ユーザにより付与されたタグ、投稿日時、ユーザによる採点結果、ユーザによる車に対する感想が含まれる。本研究では、ユーザによる採点結果とユーザによる車に対する感想のデータを用いた。



図 2 レビューの例

4.2 データに対する前処理

4.2.1 解析対象とする車種の選択

本研究で販売数予測問題における解析の対象とした車種は、28 車種である。選択する基準として以下の 3 つを考慮した。

- 分析の対象とした期間の始めである 2011 年 11 月時点ですでに販売されており、かつ分析期間の終わりである 2015 年 8 月時点でも新車として継続的に販売されている車種
- 分析期間における月次の販売台数の平均が 1,000 台以上の車種
- 予測モデルの精度検証を行う期間（2015 年 1 月から 2015 年 8 月）にモデルチェンジを行っていない車種

これは分析対象期間にモデルチェンジが行われた場合、販売数の傾向が大きく変化し、またモデルチェンジは販売メーカー側の都合に大きく左右される現象であるため、今回の研究で検証する対象の範囲外であると考えたためである。最後に、分析期間内におけるレビューの総投稿数が 300 件以上ある車種であること。これは、レビューの投稿数が少ない場合、今回の問題設定において十分に検証が行えないと判断したためである。これら 3 つの条件に照らし合わせた結果、解析の対象となる車種は、ALPHARD, COROLLA, CROWN, CUBE, DEMIO, eK WAGON, ELGRAND, ESTIMA, FIT, FREED, LEGACY, MARCH, MIRA, MOVE, NOAH, NOTE, ODYSSEY, PASSO, PRIUS, SERENA, SIENTA, SWIFT, TANTO, VITS, VOXY, WAGON R, WISH, X-TRAIL であった。

4.2.2 レビューデータに対する前処理

レビューデータに対して行った前処理の概要を説明する。収集したレビュー文に対して MeCab により、形態素解析を行い、名詞および形容詞のみを抽出し、bag-of-words 形式でレビュー文を表現した。続いて、[23] にならい、あらかじめ設定した文章頻度に基づき出現回数が閾値以下の語を除いた。閾値は、[24] を参考に使用される語彙数が同程度となるように決定した。閾値を 30 とし文章頻度の低い

^{*1} <http://www.jada.or.jp/>

^{*2} <http://www.goo-net.com/index.html>

表 1 感情分析に寄与する素性

rank	word	score	rank	word	score
1	プレミアム	11.04	10	考え	5.91
2	従来	9.07	10	基準	5.91
3	サイト	8.28	10	おとなしく	5.91
4	cx	7.88	10	音響	5.91
5	辺り	7.10	10	久しぶり	5.91
6	記録	6.70	10	中止	5.91
6	無限	6.70	17	タコメーター	5.72
6	デジタル	6.70	18	周囲	5.52
9	アクア	6.11	18	携帯	5.52
10	no	5.91	18	機敏	5.52

語を除いた結果、全語彙集合 V に含まれる語数は 2,265 であった。

4.3 感情分析

4.3.1 感情分析の精度

レビューデータに対する感情分析について説明する。収集したレビューデータは前処理を行った後、本研究で販売予測の対象期間とした 2014 年 12 月より前までに投稿されたレビューを車種による区別なく学習に用いた。

前処理を行ったレビューデータをユーザによる採点が高い順に並べ、採点の上位 3 割および下位 3 割のレビューを選択し、感情分析用のデータセットとした。ユーザ採点の結果が上位のデータには正のラベルを付与し、下位のデータには負のラベルを付与した。その後、データセットを訓練データとテストデータに 7:3 の割合で分割した。感情分析に用いる分類器としては、先行研究にて用いられてきた分類器のうち、入力として与えられた素性の予測に対する寄与度を算出できる単純ベイズ分類器を用いた。訓練データを用いて単純ベイズ分類器を学習した後、テストデータに適用し分類精度を検証した。分類の精度の評価は F 値にて行い、分類精度は 0.89 であった。得られた結果は、その後の販売数予測タスクを行う上で十分な精度であると考え、感情分析において有用であった素性を販売数予測に用いた。なお、単純ベイズ分類器の実装は scikit-learn [25] を用いて行い、パラメータは [25] の値をそのまま使用した。

4.3.2 予測に寄与する単語素性の抽出

感情分析に用いた分類器から正例の予測に寄与度の大きい素性を抽出した。予測に対する寄与度は、正例および負例が与えられた際の素性ごとの条件付き確率から対数オッズ比を算出し用いた。正例を予測に対する寄与度の高い単語を表 1 に示す。寄与度の大きな単語は「プレミアム」、「従来」であった。それぞれが使われていた代表的な事例を列挙する。「プレミアム」については、「プレミアム感のある外装や装備は満足度が高いです。」といった肯定的な

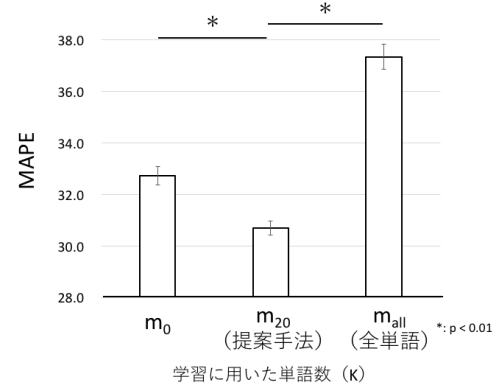


図 3 単語素性を用いない場合、全単語素性を用いる場合と提案手法の比較

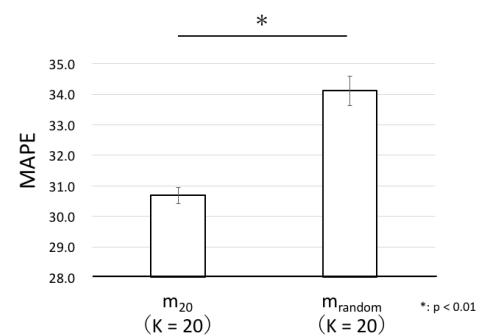


図 4 ランダムに選択された単語素性を用いる手法と提案手法の比較

感情を直接的に表現している場合が多かった。「従来」との場合は、「従来よりも燃費と経済性が優れていると感じます。」というようにこれまでのモデルや車種と比較して優れているという文脈で用いられ、レビュー文全体として肯定的になっていた。感情分析と予測に寄与していた因子についての分析から、消費者の感情を強く表すと推測される素性が、感情分析において正例の予測に強く寄与しており、それらの素性を特定・抽出することができると示唆された。

4.4 販売数予測

抽出された素性を用いて販売数の予測を行い、提案手法の有用性を検証した。予測モデルの構築は対象とした 28 車種それぞれについて行い、車種ごとに予測期間における実際の販売台数と予測される販売台数の誤差を平均絶対パーセント誤差 (Mean Absolute Percent Error; 以下 MAPE) にて評価した。

$$MAPE = \frac{1}{T} \sum_{t=1}^T \frac{|Pred_t - True_t|}{True_t} \quad (4)$$

T は学習期間に含まれる月の数であり、 $Pred_t$ は学習モデルから出力される予測値、 $True_t$ は実測値である。MAPE は時系列データに対して用いる精度の指標の一つである。

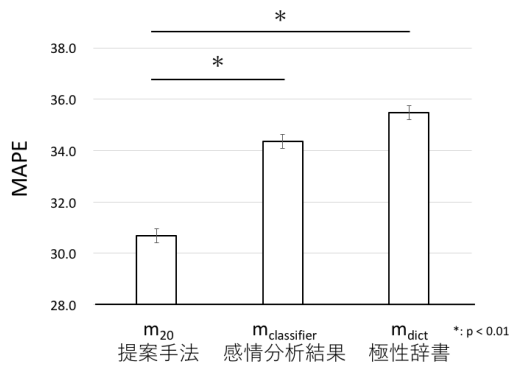


図 5 極性辞書を用いる手法，感情分析の結果を直接用いる手法と提案手法の比較

表 2 開発セットにおける精度

word num (K)	10	20	30	40	50
Average MAPE	35.37	34.57	34.70	34.79	34.93

予測値と実測値の差の絶対値を実測値で除し，予測期間での合計値を算出した後，予測期間の長さで除すことで算出される．これによりパーセント誤差が算出され，車種ごとの誤差を比較できるようになる．

予測期間は，データが入手可能であった期間のうち最後の 8 ヶ月間である 2015 年 1 月から 2015 年 8 月までとした．予測は直近 24 ヶ月間のデータを用いて学習したモデルにより，1 ヶ月後の販売台数を予測することを繰り返すスライディングウィンドウ方式で行った．つまり，まず 2012 年 12 月から 2014 年 12 月までのデータを用いて学習を行い，2015 年 1 月の販売数を予測する．続いて，2013 年 1 月から 2015 年 1 月までのデータを用いて，2015 年 2 月の販売数を予測する．このようにして 2015 年 1 月から 2015 年 8 月までの予測を行う．

予測に用いる学習器としては，ランダムフォレスト回帰 [26] を用いた．ランダムフォレスト回帰では学習に用いた各素性の重要度が算出でき，その値を用いることで利用者が各素性の重要度を知り，販売数に寄与する要素を推測しやすくなるためである．学習器の実装には，scikit-learn [25] を用い，用いる木の数を 50 とした以外，全てのパラメータで [25] の RandomForestRegressor におけるデフォルトの値をそのまま用いた．

まず，提案手法において用いる単語数 K の決定を行なった．ランダムフォレスト回帰に対して，当月の販売台数，1-of-k ベクトルにて表される 12 次元の月情報および感情分析における予測への寄与度の大きい上位 K 単語の出現回数を与え，翌月の販売数を予測した．学習に用いる 36 ヶ月分のデータにおける最初の 24 ヶ月分を学習セット，残りの 12 ヶ月分を開発セットとし，用いる単語数 K の決定を行なった． K の値としては，10, 20, 30, 40, 50 を試した．ランダムシードを固定し，学習セットでの学習を 3 回行

い，開発セットにおける平均の予測精度を算出した．算出された平均の精度を表 2 に示す．開発セットにおける精度は $K = 20$ とした場合に最も高かったため，以後の実験では $K = 20$ の場合の結果を示す．

続いて，提案手法の有効性を示すため，ベースラインとの比較実験を行なった．当月の販売台数および 1-of-k ベクトルにて表される 12 次元の月情報を与えて学習するモデルを m_0 とし， m_0 および m_0 に素性を加えた条件をベースラインとした．具体的には，単語素性を用いない条件 m_0 および m_0 に対して感情分析に用いた全語彙を加える条件 m_{all} ，極性辞書に含まれる語の出現回数を加える条件 m_{dict} ，感情分析器による予測結果を用いる条件 $m_{classifier}$ ，ランダムに選択された単語を用いる条件 m_{random} を置いた．これらのベースラインと m_0 に感情分析の結果に対する寄与度の大きい上位 20 単語の出現回数を加えた m_{20} による予測精度を比較した．実験は 3 つの異なるランダムシードを用いて行い，3 回の試行により得られた結果を平均し最終的な結果とした．

まず，用いる単語数による精度の変化を調べるため，単語素性を用いない条件 m_0 ，全単語素性を用いる条件 m_{all} と m_{20} の精度を比較した．結果を図 3 に示す．全単語を用いる条件 m_{all} における MAPE 値および単語をまったく用いない条件 m_0 における MAPE 値は，それぞれ 32.73 および 37.34 であった．一方，提案手法 m_{20} における MAPE 値は 30.69 であった．また， t 検定による検定の結果，予測の精度は有意水準 1% で有意に向上していた．

続いて，感情分析の結果得られた単語素性を用いることの重要性を示すため，ランダムに選択された語のレビューにおける出現回数を用いる m_{random} と m_{20} の比較を行なった．結果を図 4 に示す．全語彙集合に含まれる語からランダムに提案手法と同数の単語 ($K = 20$) を選択し，選択された語の出現回数を素性として予測を行なった．その結果， m_{random} における MAPE 値は 34.11 となり， m_0 での予測精度は m_{random} と比較して有意水準 1% で有意に向上していた．

最後に，感情分析器から得た単語素性を用いることの有用性を示すため，極性辞書に含まれる語の出現回数 m_{dict} および感情分析の結果をそのまま用いる $m_{classifier}$ と m_{20} を比較した． m_{dict} では，レビューにおける感情分析に用いられる極性辞書に含まれる単語の出現を素性とした．極性辞書は [27] のものを用いた．極性辞書に登録された単語のうち感情極性実数値が正の単語を選択し，その出現回数を予測における素性とした． $m_{classifier}$ では本研究で用いた感情分析器に対して，レビュー文章を与え，その分類結果におけるポジティブなレビューの結果とネガティブなレビューの比を素性として用いた．得られた結果を図 5 に示す． m_{dict} における MAPE 値は 35.48 であり， $m_{classifier}$ による予測についての MAPE 値は 34.35 であった． t 検定

による検定の結果、提案手法 m_{20} は両ベースラインと比較して有意水準 1% で有意に予測精度が向上していた。

5. 考察

5.1 感情分析

感情分析における評価の高いレビューと低いレビューの分類精度は、F 値で 0.89 であった。分類の対象としたレビューは、得点が上位 3 割および下位 3 割のレビューとした。これはユーザごとに採点基準が異なるため全てのレビューを用いた場合に、レビュー文に記されたユーザの評価が大きく異なる一方点数が同じであるという状態が生じ、感情分析の評価が低下すると考えたためである。特に車に対する評価が高くない場合の点数（辛口採点の度合）は、ユーザ個人の感覚に依存するため精度が上がらない原因となる。具体的には、あるユーザは「満足度は中程度」という文章と合わせて「4 点」と投稿するが、他のユーザは「3 点」と投稿するといった状況が考えられる。高得点のレビューについては全体的に評価の高さを連想させる語が並ぶが、ユーザ採点が全体の中央値付近のレビューでは文章が高得点のレビューと類似するものも低い得点のレビューに類似するものの両方が混在する。

本研究における感情分析では、bag-of-words 形式で表現したレビュー文章を入力として学習器に与えた。そのため、レビュー本文中で各単語が現れる文脈を考慮できていない。例えば、「燃費が悪い」と「悪い部分はない」という 2 つのレビューが与えられたとき、前者は否定的な感情を表し、後者は肯定的な感情を表すと考えられる。しかし、形態素解析の結果を bag-of-words 形式で表現した場合、前者は「燃費:1, 悪い:1」となり、後者は「悪い:1, 部分:1, ない:1」となり、「悪い」という単語がどちらの例にも含まれることとなる。そのため、bag-of-words 形式では一部の単語の予測に対する寄与度を過小評価してしまう可能性がある。また、表 1 において上位にある「従来」といった語は、「従来より〇〇」といった形で用いられることがほとんどであった。こちらについても、語順を考慮しない bag-of-words 形式で表現した場合、失われる情報があると考えられる。文脈を考慮するモデルや n-gram を用いることにより、より感情分析の精度を高めることができる可能性がある。ただし n-gram を用いる場合、語数が膨大になるためその点を考慮する必要がある。

感情分析の結果、正例の予測に寄与する単語は表 1 のような単語であった。正例の予測には、「プレミアム」や「無限」といった評価の高さを連想させるような単語の寄与度が高かった。また、「アクア」や「cx」といった個別の車種を指す単語も含まれていた。これらは、当該車種についてのレビューに含まれるだけでなく、他の車種のレビューにも出現していた。例えば、「アクア」という単語は、FIT や PRIUS といった類似の車種のレビューに多く出現して

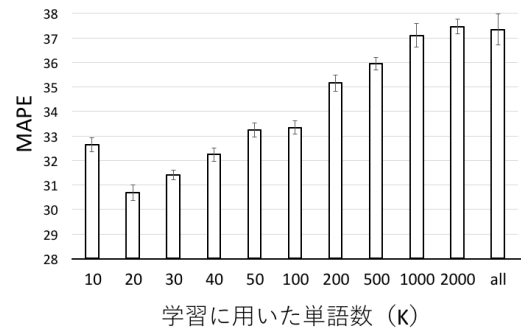


図 6 単語数に対する平均 MAPE 値

おり、しばしばレビュー中でこれらの車種と比較されていた。また「no」という単語は、「no1」といった用いられ方が多かったため、評価の高いレビューの予測に対する寄与度が大きくなったと考えられる。

5.2 販売数予測

販売数予測を行い、提案手法とベースラインを比較した結果および提案手法において用いられる単語数と精度の関係について考察する。

開発セットによるパラメータサーチの結果、本実験における最適な単語数 K の値は 20 となった。 K と精度の関係を調べるため、その他の単語数 ($K = 10, 30, 40, 50, 100, 200, 500, 1000, 2000$) についてもテストセットでの精度を検証したところ、図 6 に示す結果を得た。得られた図から、今回実験で用いたデータにおいて単語数 K は 20 が最適であること、および $K = 20$ 以降では K が増加するとともに MAPE 値は悪化していることが明らかとなった。この結果から、適切な単語数を選択することで精度の向上が期待できる一方、用いる単語が過剰になると予測精度が低下することが明らかとなった。

続いて、ベースラインとの比較を行なった。その結果から、適切に設定された単語数のもとでの提案手法による予測精度が有意に向上することが明らかとなった。ベースラインとしては、単語素性を用いない条件 m_0 および m_0 に対して感情分析に用いた全語彙を加える条件 m_{all} 、極性辞書に含まれる語の出現回数を加える条件 m_{dict} 、感情分析器による予測結果を用いる条件 $m_{classifier}$ 、ランダムに選択された単語を用いる条件 m_{random} を置いた。単語素性を用いない条件 m_0 および全語彙を用いる条件 m_{all} との比較結果から、適切な数の単語を選択することで精度が向上することが示された。テストセットにおける MAPE 値は m_0 と比較して 6.6%、 m_{all} と比較して 21.68% 改善した。 t 検定を行なった結果、 p 値は m_0 に対して 1.4×10^{-3} 、 m_{all} に対して 3.1×10^{-5} であり、統計的に有意に精度が向上していることが示された。

また、ランダムに選択された単語素性を用いる条件 m_{random} との比較結果からは、感情分析を用いて単語を選択する重要性が示唆された。提案手法と m_{random} において用いる単語素性の数 K を同一の値 ($K = 20$) に設定し、それぞれの手法における予測結果の統計的有意差を検証した結果、有意水準 1% 以下で提案手法による予測の精度が向上していた。この結果から、ランダムな単語素性を追加するだけでは予測精度を向上させることが難しく、感情分析により選択した結果を用いることが重要であることが示唆された。

最後に、極性辞書を用いる条件 m_{dict} および感情分析の結果を用いる $m_{classifier}$ と提案手法の比較結果から、レビューにおける感情分析タスクから得られる単語素性を用いることの重要性が示唆された。極性辞書を用いた条件 m_{dict} との比較結果から、予測に用いる単語素性は極性辞書に記載された一般的な語ではなく、用いるデータから得られる語の方が効果的であることが考えられる。また、 $m_{classifier}$ との比較結果から、感情分析の結果を素性として予測に用いる手法と比較して、感情分析タスクにおいて予測に寄与する単語素性を抽出し、販売予測に用いる提案手法の精度が有意に向上することが明らかとなった。

以上の結果から、感情分析タスクにおいて予測に寄与する単語素性を適切に選択された単語数 K のもと用いることで、有意に予測精度を向上できることが示された。

6. おわりに

本研究では、オンラインレビューにおける消費者の感情を分類するタスクにて予測への寄与度の大きい素性を用いることで、販売数予測タスクの精度を向上する手法を提案した。提案手法の有用性についての検証は、国内における自動車の販売台数予測問題にて行なった。まず、レビューデータに対する感情分析タスクを行い、ユーザの感情の予測に対する寄与度の大きい単語素性を特定した。続いて、感情分析タスクにおいて特定された単語素性を用いて自動車の販売数予測を行なった。販売予測における有用性の検証では、レビューに対する感情分析により選択した単語素性が予測において有用であることを示すためのベースラインとして、単語素性を用いない条件、レビューに含まれる全単語の出現回数を用いる条件、ランダムに選択された単語素性を用いる条件、極性辞書に含まれる単語の出現回数を用いる条件および感情分析タスクにて学習された分類器による分類結果を用いる条件を置いた。その結果、適切に設定された単語数のもとでは、提案手法がいずれのベースラインに対しても有意水準 1% で有意に精度向上していることが明らかとなった。また単語数と予測精度の関係から、単語数がある一定以上の数になると精度が低下することが示唆された。

本研究では、これまでの研究で知られていた、口コミ情

報が商品の売上に重要であること [2] に着想を得て販売数予測モデルを提案した。感情分析によりレビューから抽出した単語素性は、商品に対する消費者の感情を内包していると考えられ、それらを用いることで予測の精度が向上することは先行研究からの知見に合致する。本研究の提案手法は、自動車におけるモデルチェンジといったイベントを考慮して予測することが難しいといった側面があるものの、商品に対するレビューが十分に存在しユーザの感情が抽出できるという条件を満たせば、自動車のみならず幅広く用いることが可能であると考えられる。

謝辞 本研究は JSPS 科研費 JP25700032 の助成を受けたものです。

参考文献

- [1] Kuo, R. J. and Xue, K.: A decision support system for sales forecasting through fuzzy neural networks with asymmetric fuzzy weights, *Decision Support Systems*, Vol. 24, No. 2, pp. 105–126 (1998).
- [2] Dellarocas, C., Zhang, X. M. and Awad, N. F.: Exploring the value of online product reviews in forecasting sales: The case of motion pictures, *Journal of Interactive marketing*, Vol. 21, No. 4, pp. 23–45 (2007).
- [3] Bollen, J., Mao, H. and Zeng, X.: Twitter mood predicts the stock market, *Journal of Computational Science*, Vol. 2, No. 1, pp. 1–8 (2011).
- [4] Wuthrich, B., Cho, V., Leung, S., Permuntilleke, D., Sankaran, K. and Zhang, J.: Daily stock market forecast from textual web data, *Systems, Man, and Cybernetics, 1998. 1998 IEEE International Conference on*, Vol. 3, IEEE, pp. 2720–2725 (1998).
- [5] Ruiz, E. J., Hristidis, V., Castillo, C., Gionis, A. and Jaimes, A.: Correlating financial time series with microblogging activity, *Proceedings of the fifth ACM international conference on Web search and data mining*, ACM, pp. 513–522 (2012).
- [6] Caruana, R.: Multitask learning, *Learning to learn*, Springer, pp. 95–133 (1998).
- [7] Silver, D. L. and Mercer, R. E.: Sequential inductive transfer for coronary artery disease diagnosis, *2007 International Joint Conference on Neural Networks*, IEEE, pp. 2635–2641 (2007).
- [8] Ahn, H.-I. and Spangler, W. S.: Sales Prediction with Social Media Analysis, *2014 Annual SRII Global Conference*, IEEE, pp. 213–222 (2014).
- [9] Yu, X., Liu, Y., Huang, X. and An, A.: Mining online reviews for predicting sales performance: A case study in the movie domain, *IEEE Transactions on Knowledge and Data engineering*, Vol. 24, No. 4, pp. 720–734 (2012).
- [10] Feldman, R.: Techniques and applications for sentiment analysis, *Communications of the ACM*, Vol. 56, No. 4, pp. 82–89 (2013).
- [11] Taboada, M., Brooke, J., Tofiloski, M., Voll, K. and Stede, M.: Lexicon-based methods for sentiment analysis, *Computational linguistics*, Vol. 37, No. 2, pp. 267–307 (2011).
- [12] Pang, B., Lee, L. and Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques, *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume*

- 10, Association for Computational Linguistics, pp. 79-86 (2002).
- [13] Gamon, M.: Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis, *Proceedings of the 20th international conference on Computational Linguistics*, Association for Computational Linguistics, p. 841 (2004).
- [14] Yessenalina, A. and Cardie, C.: Compositional matrix-space models for sentiment analysis, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, pp. 172-182 (2011).
- [15] Sharma, A. and Dey, S.: A comparative study of feature selection and machine learning techniques for sentiment analysis, *Proceedings of the 2012 ACM Research in Applied Computation Symposium*, ACM, pp. 1-7 (2012).
- [16] Asur, S. and Huberman, B. A.: Predicting the future with social media, *Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 2010 IEEE/WIC/ACM International Conference on, Vol. 1, IEEE, pp. 492-499 (2010).
- [17] Mishne, G., Glance, N. S. et al.: Predicting Movie Sales from Blogger Sentiment., *AAAI spring symposium: computational approaches to analyzing weblogs*, pp. 155-158 (2006).
- [18] Choi, H. and Varian, H.: Predicting the present with Google Trends, *Economic Record*, Vol. 88, No. s1, pp. 2-9 (2012).
- [19] Nakatoh, T., Amano, H. and Hirokawa, S.: Prediction of growth rate of operating income using securities reports, *Advanced Applied Informatics (IIAIAI)*, 2013 IIAI International Conference on, IEEE, pp. 84-88 (2013).
- [20] Forman, G.: An extensive empirical study of feature selection metrics for text classification, *Journal of machine learning research*, Vol. 3, No. Mar, pp. 1289-1305 (2003).
- [21] Abbasi, A., Chen, H. and Salem, A.: Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums, *ACM Transactions on Information Systems (TOIS)*, Vol. 26, No. 3, p. 12 (2008).
- [22] Duric, A. and Song, F.: Feature selection for sentiment analysis based on content and syntax models, *Decision Support Systems*, Vol. 53, No. 4, pp. 704-711 (2012).
- [23] Yang, Y. and Pedersen, J. O.: A comparative study on feature selection in text categorization, *Icml*, Vol. 97, pp. 412-420 (1997).
- [24] Huang, A.: Similarity measures for text document clustering, *Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008)*, Christchurch, New Zealand, pp. 49-56 (2008).
- [25] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. et al.: Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research*, Vol. 12, No. Oct, pp. 2825-2830 (2011).
- [26] Breiman, L.: Random forests, *Machine learning*, Vol. 45, No. 1, pp. 5-32 (2001).
- [27] 高村大也, 乾孝司, 奥村学ほか: スピンモデルによる単語の感情極性抽出, 情報処理学会論文誌, Vol. 47, No. 2, pp. 627-637 (2006).

野中 尚輝

2012 年東京大学工学部化学生工学科卒業. 2014 年 3 月同大学大学院メディカルゲノム修士課程修了. 2014 年 4 月より同大学大学院技術経営戦略学専攻松尾研究室にて研究活動に従事.

中山 浩太郎 (正会員)

2007 年大阪大学大学院情報科学研究科博士号取得. 博士(情報科学). (株) 関西総合情報研究所代表取締役, 大阪大学研究員, 東京大学知の構造化センター特任助教などの勤務を経て, 2014 年より, 東京大学大学院工学系研究科技術経営戦略学専攻 特任講師. 情報処理学会山下記念研究賞, 情報処理学会 CS 専攻賞, 日本データベース学会論文賞など受賞. 専門は, 人工知能, Deep Learning, スケーラビリティ, Web Mining.

松尾 豊 (正会員)

1997 年 東京大学工学部電子情報工学科卒業 2002 年 同大学院博士課程修了. 博士(工学)産業技術総合研究所, スタンフォード大学を経て, 2007 年より, 東京大学大学院工学系研究科 技術経営戦略学専攻 准教授. 2012 年より人工知能学会理事・編集委員長, 2014 年より倫理委員長. 人工知能学会論文賞, 情報処理学会会長尾真記念特別賞, ドコモモバイルサイエンス賞など受賞. 専門は, Web 工学, Deep Learning, 人工知能.