

商品レビューを用いたプレゼント支援の検討

Support system for present selection using product reviews

田口 拓明*¹
Hiroaki Taguchi

田村 哲嗣*¹
Satoshi Tamura

速水 悟*¹
Satoru Hayamizu

*¹ 岐阜大学工学部
Faculty of Engineering, Gifu University

There are many web sites to select presents. In these sites, attributes of the receiver are labeled for each item on sale and prospective senders can search new items as presents using the information. But not all the items are labeled. This paper proposes a classifier if the item is suitable for a present. The classifier is trained using the review articles with labeled items and their attributes. Using this classifier, attributes for the item without labels are estimated. Subjective evaluations are conducted for the labeled items using the proposed method.

1. 序論

日本には贈答文化という独特の習慣があり、お中元、お歳暮など、誕生日プレゼントも含めると、1年の内でも贈り物をする機会は頻繁にある。また、交友関係が増えると、親しい友人の誕生日を祝う時にプレゼントを贈る、会社の上司にお歳暮を用意するなど、相手に贈り物を用意する機会も増える。

近年、インターネットショッピング利用者が増加し続ける中で、この贈答文化は、昔とは随分と様変わりしてきている。贈り物を取り扱ったインターネットショッピングサイトに、例えば、楽天市場の「おくりもの市場」や AEON の「オヤコム」、Twitter や mixi 等と提携する「ギフトナウ」等があり、多数存在する。

これらのサービスでは、運営者がプレゼントに適した商品を選んでる。しかし、インターネットショッピングサイト上に商品はたくさんあり、例えば Amazon の商品数は 5000 万点、楽天市場であれば 1 億 5000 万点である。その中からプレゼントに適した商品を見つけるのは非常に手間である。

そこで、本研究では商品がプレゼントに適しているかを自動で判定し、プレゼント選びを支援するシステムを構築する。

2. 関連研究

[堀 2011]は、喜ばれるプレゼントを、相手の趣味や嗜好を考慮して推薦するサービスを提案している。相手の趣向を知る手段として、独自に実装した SNS からユーザの行動の解析をしてプレゼントを推薦している。しかし、この推薦手法では SNS サイトによる情報がなければ商品がプレゼントに適しているかを推定できない。本研究ではプレゼントを推薦するために、商品のレビューだけを使用しているため、多くの商品に対してプレゼントに適しているかを推定できる。

検索および推薦にレビュー情報を用いる研究はいくつか行われている。[倉敷 2007]は、比較評価情報として評価対象、比較対象、属性、評価を抽出し、対象の優劣をランキングで可視化している。対象を映画としており、評価として「良い」「優れている」という比較の方法をしている。本研究では、商品すべてを対象としていることと、評価の優劣ではなく、プレゼントとして適しているかどうかで比較している点で異なる。

3. 提案システム

本システムの構成図を図 1 に示す。本システムがプレゼントに適した商品抽出する手法は、以下の 2 つの部分に分かれる。

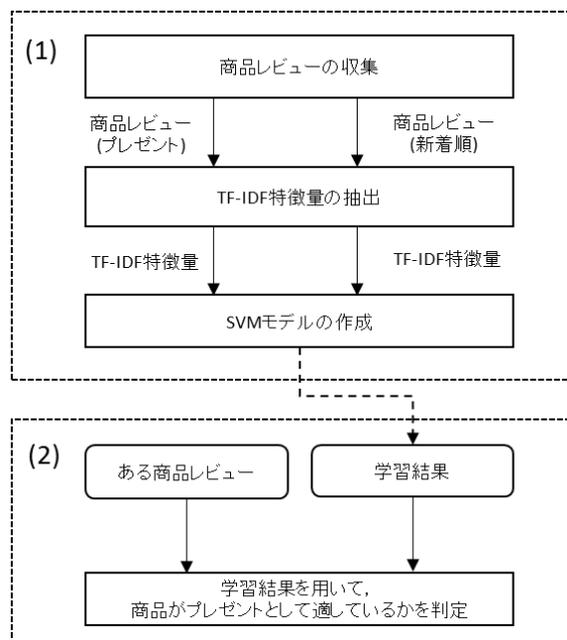


図1 システムの構成図

- (1) はじめに商品のレビューを収集する。次に各レビューに対し、単語の TF-IDF 特徴量を抽出する。そして、プレゼントに適しているかどうかを判定する SVM(Support vector machine)を学習する。
- (2) 学習したモデルを使用し、ある商品のレビューがプレゼントに関係した内容かどうかを判定する。そして、レビューがプレゼントに関係した内容の商品を、プレゼントに適している商品とする。

3.1 商品のレビューの収集

商品のレビューは、楽天の「みんなのレビュー・クチコミ」から取得する。みんなのレビュー・クチコミでは様々な商品のレビューを扱っており、そのレビューを様々な項目を基に検索することができる。例えば、「使いみち」が「プレゼント」の時の商品を検

索することができ、その商品レビューも分かる。本システムでは、“使いみち”が“プレゼント”の条件で検索し、収集した商品のレビューを、プレゼントとして用いられた商品のレビューとして使用した。また、SVM の学習モデルの作成用に、商品のレビューを新着順に取得した。

3.2 TF-IDF 特徴量の抽出

プレゼントとして用いられた商品のレビューと、新着順に取得した商品のレビューをそれぞれ形態素解析し、TF-IDF 値を計算する。本研究では、形態素解析エンジン MeCab[Kudo 2004] を利用し、「名詞」「形容詞」「動詞」の単語を取得する。「名詞」「形容詞」「動詞」を取得したのは、レビューの評判分析では、よく使用される品詞だからである。[白石 2012]は、インターネット上の口コミサイトにおいて化粧品の評判分析をする際に、レビューから「名詞」「形容詞」「動詞」を抽出している。また、レビュー・クチコミなどのテキストを形態素解析すると、MeCab の単語辞書にはない未知語が多く出現すると考えられるため、未知語の対策として[黒澤 2011]の手法を参考に MeCab の単語辞書 392,126 語に、Wikipedia のタイトルの単語と、はてなキーワードの単語の計 1,449,225 語を追加した。

プレゼントとして用いられた商品のレビューの TF-IDF 値を計算するとき、IDF は、新着順に取得した商品レビューを利用した。これはプレゼントとして用いられた商品のレビュー内の単語に対し、商品のレビュー全体で一般性の高い単語の TF-IDF 値を減少させるためである。

3.3 単語のフィルタリング

商品レビューの単語の TF-IDF 値は、値が 0 のものが多く存在するため、SVM の学習モデルの作成に使用する単語をプレゼントに特徴的な単語だけに限定する。プレゼントに特徴的な単語の抽出は以下の流れで行う。

- (1) プレゼントとして用いられた商品のレビューを取得し、全ての文章を結合して、TF 値を計算する。このとき TF 値を $tf_{present}$ とする。
- (2) 新着順に複数件取得した商品レビューの IDF 値を計算する。IDF 値を idf_{all} とする。
- (3) プレゼントとして用いられた商品のレビュー内の全ての単語 w に対し、(1)、(2)で計算した $tf_{present}$ 値と idf_{all} 値を掛け合わせ、 $S(w)$ を求める。 $S(w)$ は式(1)で算出される。

$$S(w) = tf_{present} \cdot idf_{all} \quad (1)$$

そして、 $S(w)$ が上位の単語を抽出する。

3.4 SVM の学習

単語ベクトルから SVM の学習モデルを作成する。SVM は、分類対象が 2 種類だけの 2 値データに対する分類器であり、本研究では SVM により商品レビューがプレゼントに関連した内容であるかを判定する。SVM の学習のために、正例のデータとしてプレゼントとして用いられた商品のレビューを利用し、負例として新着順に複数件取得した商品レビューを利用する。

3.5 プレゼントに適した商品の判定

3.4 項で作成した SVM の学習モデルを利用し、ある商品がプレゼントに適しているかを判定する。まず、ある商品に対し、レビューを取得し形態素解析して TF-IDF 値を算出する。次に、

SVM を用いてプレゼントに関連したレビューかどうか判定する。最後に、判定された商品をプレゼントに適している商品として抽出する。

4. 評価実験

本手法の有用性を確認するために実験を行った。

4.1 実験条件

本実験では、3.3 項で述べた(1)単語のフィルタリング、(2)学習データ数により、プレゼントに適切な商品の判定の識別率がどのように変化するか調査した。

SVM の学習では、R の keanlab ライブラリ*1を利用した。学習モデルを作成する条件を表 1 に示す。

表 1 実験条件

学習データ	・プレゼントとして扱われた商品レビュー ・新着順に取得した商品レビュー
形態素	名詞・形容詞・動詞
特徴量	TF-IDF 値
識別関数	SVM
カーネル	ガウシアンカーネル
評価	オープン条件
評価尺度	適合率, 再現率, F 値

評価は、全ての商品のレビューからランダムで半数を抽出して SVM を用いて学習し、抽出していない半数について識別するオープン条件により行った。また評価尺度は、適合率, 再現率, f 値を採用した。

(1) 単語のフィルタリング

単語のフィルタリングでは、プレゼントとして用いられた商品のレビュー 10,000 個から、特徴的な単語に限定した。本実験では、限定した単語数を変化させることで、SVM の識別率がどう変化するか調査した。限定した単語数を表 2 にまとめる。また、特徴的な単語の例を表 3 に示す。表 3 は、単語の TF-IDF 値の高い順に 15 個示したものである。また、単語の左側には、その順位を記述した。なお、学習データは、プレゼントとして扱われた商品レビューと新着順に複数件取得した商品レビューともに 1000 個を利用した。

表 2 フィルタリングした単語数

商品レビュー内の 総単語数との 割合(%)	20%	40%	60%	80%	100%
単語数(個)	2,354	4,707	7,061	9,414	11,767

表 3 特徴単語の例

1	プレゼント	6	くれ	11	友達
2	喜ん	7	もらえ	12	ラッピング
3	誕生	8	用	13	商品
4	購入	9	友人	14	母
5	日	10	思い	15	祝い

*1 <http://cran.r-project.org/web/packages/kernlab/index.html>

(2) 学習データ数

学習データ数を 2,000 個・4,000 個・6,000 個・8,000 個・10,000 個と変化させたときの SVM の学習モデルの識別を行った。また学習データ数は、プレゼントとして扱われた商品レビューと新着順に取得した商品レビューが、全体の商品レビューの半数ずつになるように設計した。学習データ以外の実験条件は、表 1 に示した条件で行った。

4.2 実験結果

(1) 単語のフィルタリング・(2) 学習データ数の違いによる実験結果を、それぞれ表 4・表 5 に示す。また、学習データ数による違いの実験結果を折れ線グラフで表したものを図 2 に示す。

表 4 単語のフィルタリングによる実験結果

フィルタリングした単語の割合	適合率	再現率	f 値
20%	0.809	0.950	0.874
40%	0.800	0.942	0.865
60%	0.800	0.935	0.865
80%	0.800	0.942	0.863
100%	0.798	0.950	0.867

表 5 学習データ数を変更したときの実験結果

学習データ数	適合率	再現率	f 値
2,000	0.831	0.928	0.877
4,000	0.852	0.945	0.897
6,000	0.865	0.954	0.907
8,000	0.874	0.952	0.911
10,000	0.871	0.955	0.911

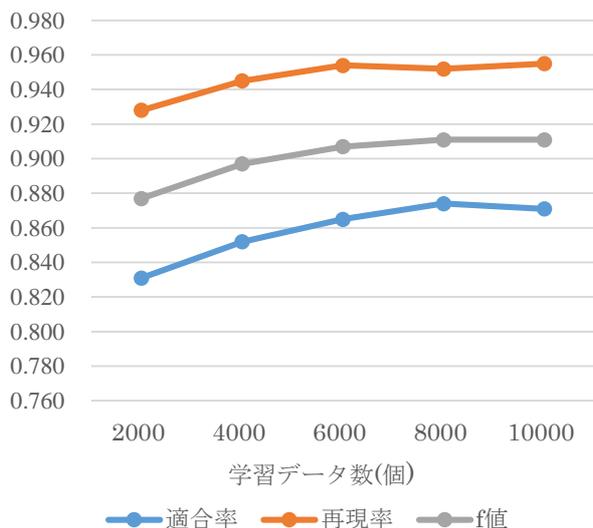


図 2 学習データ数を変更したときの実験結果

4.3 考察

(1) 単語のフィルタリングでは、フィルタリングした単語の割合を 20%・40%・60%・80%・100% と変更させても、適合率・再現率・f 値にあまり大きな変化は見られなかった。そこで、本研究ではフ

ィルタリングした単語の割合を減らすことで、SVM の学習モデルの作成時に計算時間を減らせることから、フィルタリングする単語の割合を 20% とすることにした。(2) 学習データ数の変更では学習データが増加するほど、学習モデルの識別精度が増加することが確認できた。また、図 2 を見ると学習データ数が 2,000・4,000・6,000 個では、適合率・再現率・f 値に大きな変化があるが、6,000・8,000・10,000 個では、あまり変化がないため、これ以上学習データ数を大きくしても識別精度はあまり変わらないと考えられる。そのため、データ数 10,000 個を使用するのが適していると推察できる。

5. 被験者実験

本システムが、被験者に対してプレゼントに適した商品を提供することができるかを調査するため、被験者実験を行った。

5.1 実験条件

評価実験は男女 10 人の被験者に対して、Yahoo!ショッピングの商品のレビューを用いて行った。Yahoo!ショッピングの商品のレビューは API を基に取得することができ、例えば商品のジャンルを指定して商品検索を行なうことができる。また、検索したい商品の単語をクエリとして、商品検索を行なうこともできる。本実験では、6 つの商品のジャンルと 4 個の検索したい商品の単語を指定した。表 6、指定したジャンルと単語のリストを示す。そのジャンル・単語により検索された商品を、それぞれ 1,000 個取得し、SVM によりプレゼントに適している商品として抽出した商品リスト A (商品数 10 個) と、商品からランダムで抽出した商品リスト B (商品数 10 個) を比較した。被験者には、商品リスト A と B を比べ、以下の 2 つの質問をした。

- 2 つの商品リストの中で、どちらがプレゼントに適した商品を多く含んでいますか？
- なぜその商品リストを選びましたか？

本研究では、被験者が提案手法で抽出した商品リスト A を選んだ割合を算出し、評価した。被験者が商品リスト A を選んだ数を x 、被験者の回答数を k としたとき、提案手法により抽出した商品リストを選んだ割合 $f(x)$ は式(2)で計算される。

$$f(x) = \frac{x}{k} \times 100 \quad (2)$$

なお、SVM の学習は表 7 の条件で行なった。

表 6 検索した商品ジャンルと単語

商品ジャンル	DIY, 工具, 文具・インテリア・楽器, 趣味, 学習・ドリンク, お酒, 家電, 家庭用品
検索単語	カエル・熊・ディズニー

表 7 実験条件

学習データ	・プレゼントとして扱われた商品レビュー (5,000 個) ・新着順に複数件取得した商品レビュー (5,000 個)
形態素	名詞・形容詞・動詞
特徴量	TF-IDF 値
限定した単語数	全体の 20% (2,354 単語)
識別関数	SVM
カーネル	ガウシアンカーネル
評価	オープン条件
評価尺度	適合率, 再現率, F 値

5.2 実験結果

実験結果を図3に示す。

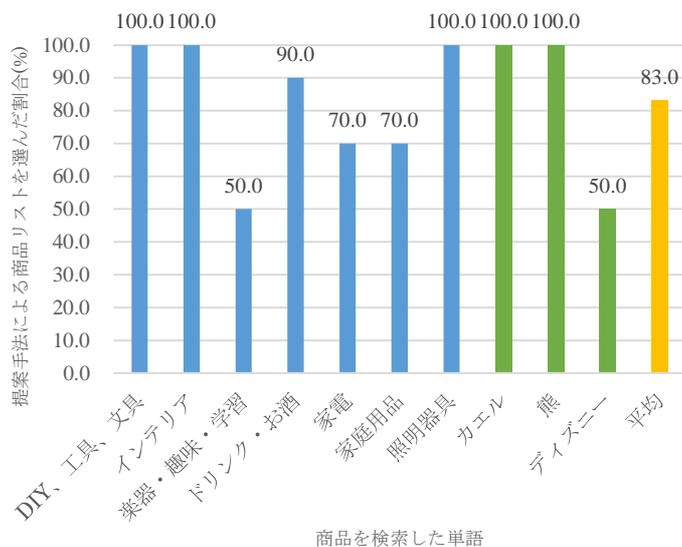


図3 被験者実験の実験結果

6. 考察

本研究の提案手法を用いて選んだ商品リスト A と商品からランダムに抽出した商品リスト B を被験者に比較してもらったところ、商品リスト A を「プレゼントに適した商品を多く含んでいる」と評価した回答の平均が 83%であった。そのため、提案手法による商品の抽出の有用性が確認できた。

実験結果では、DIY, 工具, 文具・インテリア・照明器具・カエル・熊に関連した商品リストに対し、提案手法で抽出した商品のリストを選んだ割合は 100%と高い値であった。以下に提案手法による抽出方法の利点を述べる。

- ・「DIY, 工具, 文具」では、その商品群の中からプレゼントに適する文具を抽出できていた。
 - ・「インテリア」では、プレゼントに適さない家具などを抽出せず、加湿器やマッサージクッションなどの小物を抽出できていた。
 - ・「照明器具」では、蛍光灯などの日常的に使用する商品などは抽出せず、おしゃれなランプやろうそくなどを抽出できていた。
 - ・「カエル」では、カエルの飼育用品などではなく、カエルをキャラクターとした雑貨が抽出できていた。
 - ・「熊」では、「カエル」と同様に、雑貨を抽出できていた。
- 一方、趣味, 楽器, 学習・ディズニーに関連した商品リストに対し、提案手法で抽出した商品のリストを選んだ割合は 50%と低い値であった。以下にその理由を述べる
- ・「趣味, 楽器, 学習」では、子供向けの商品が多く抽出された。今回の実験では、贈る人は指定せず、万人にプレゼントしたときに適切なものとしていたため、万人にプレゼントを贈るには不適切なものとされた。
 - ・「ディズニー」では、「趣味, 楽器, 学習」と同様に、子供向けの商品が多かったため、万人にプレゼントを贈るには不適切なものとされた。

実験結果から、提案手法による商品抽出では、小物やおしゃれな雑貨を抽出でき、プレゼントに適した商品を選べていることが分かった。また、今回の実験では、贈る人を限定しておらず、

万人にプレゼントを贈るには不適切なものが抽出されていた。今後は、プレゼントする贈り相手や季節を考慮して、商品を選ぶことが課題である。

7. 結論

本研究では、商品がプレゼントに適しているかを自動で判定するシステムの構築を行った。

本システムでは、商品がプレゼントに適しているかを判定するために、プレゼントとして扱われた商品のレビューの情報を利用した。プレゼントとして用いられた商品のレビューからTF-IDF特徴量を抽出し、SVMを作成した。このSVMを用いて、ある商品のレビューがプレゼントに関連した内容か判定し、判定された商品をプレゼントに適している商品として抽出した。評価実験として、本手法の有用性を確認するため、学習モデルの識別精度を確認した。そして、提案手法により、プレゼントに適した商品を判定できているか調べるため、被験者による評価実験を行った。評価実験では、提案手法による商品の抽出の有用性が確認できた。提案手法による商品の抽出では、小物やおしゃれな雑貨を抽出でき、プレゼントに適した商品を選べていることが分かった。また、今回の実験では、贈る人を限定しておらず、万人にプレゼントを贈るには不適切なものが抽出されていた。

本システムを利用すると、商品カテゴリーの中からプレゼントに適した商品を抽出できるため、例えば「カエル」関連のプレゼントを相手にあげたい時に、インターネットショッピングサイト上で取り扱っている「カエル」関連の商品の中でプレゼントに適したものが抽出でき、プレゼントを選ぶ時の支援になる。また、プレゼントに適した商品を選ぶときに商品のレビューのみを使用しているため、様々なインターネットショッピングサイトの商品レビューのある商品の中から、プレゼントに適した商品を取得できる。一方で、本システムはプレゼントを贈る対象者やプレゼントするシーンに適した商品の抽出は行っていないため、今後の課題として贈り相手やシーンを考慮した商品推薦を行ってきたい。

参考文献

- [堀 2011] 堀 瑛子, 田中 悠一郎, 高瀬 友彬, 杉浦 秀幸, 陸 西, 北 栄輔: ユーザーの嗜好に基づく贈り物レコメンドシステムの提案, 第4回楽天研究 開発シンポジウム, 2011.
- [倉敷 2007] 倉敷 健, 別所 克人, 戸田 浩之, 内山 俊郎, 片岡 良治, 奥 雅博: 比較評価情報に基づくランキング手法, 日本データベース学会論文誌(DBSJ Letters) Vol. 6, No. 1, pp. 5-8, 2007.
- [黒澤 2011] 黒澤義明, 竹澤寿幸: マイクロブログサービスの返信行動に着目した投稿及びユーザーの分類, 言語処理学会第17回年次大会 発表論文集, 2011.
- [Kudo 2004] Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto: Applying Conditional Random Fields to Japanese Morphological Analysis, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp.230-237, 2004.
- [白石 2012] 白石 由香利, 橋本 隆子, 久保山 哲二: インターネット上の口コミサイトにおける化粧品の評判分析, 学習院大学 計算機センター, Vol. 33, 2012.