

# 可視化を目的とした非典型度と評価極性にに基づく カスタマーレビューのクラスタリング

古橋 慎之介<sup>†,a</sup>, 内田 理<sup>††,b</sup>

<sup>†</sup>東海大学大学院工学研究科情報理工学専攻, <sup>††</sup>東海大学情報理工学部情報科学科

a) 4bdrm019@mail.tokai-u.jp b) o-uchida@tokai.ac.jp

**概要** EC サイトに投稿されている製品に関する意見を直感的かつ簡潔に提示することは、ユーザが大量のレビューを閲覧する手間を軽減し、評判情報の把握を支援する点から有用であると考えられる。本研究では、評判情報を二次元空間上に可視化して提示することを目的として、評価極性及び情報の発見性や意外性の度合いを表す非典型度という指標を用いたカスタマーレビューのクラスタリング手法を提案する。また、提案手法によって得られたクラスタのレビュー文に対して、20代男性の被験者5名により非典型度、評価極性の主観評価をしてもらい、それらと提案手法の推定値との相関係数を求めることで、提案手法の有効性について検証する。

**キーワード** 非典型度, 評価極性, カスタマーレビュー

## 1 はじめに

BtoC-EC の市場規模は現在までに継続的な増加傾向にあり、今後も発展、拡大が見込まれている。平成 26 年度の経済産業省が行った調査[1]では、2014 年までの過去 5 年間の EC 市場規模は高い伸び率を示しているとともに、EC 化率(全体の商取引金額に対する EC 市場規模の割合)も上昇しており、購買手段としての EC は消費者にとってごく一般的なものであると考えられる。ところで、消費者に対して EC サービスを提供する楽天や Amazon のような EC サイトでは、一般に、誰でも製品に対する意見を容易に発信でき、かつ多くのレビューを閲覧できる機能を有している。平成 26 年版の情報通信白書[2]によると、商品購入時に参考とする情報として、「購入サイト・レビューサイトの口コミ」が 45.6%と最も高い割合を示しており、レビューなどの口コミから得られる情報はユーザの購買行動に大きく寄与していると言えるであろう。しかしながら、投稿されているレビューは製品によっては膨大な量であり、ユーザが購買の意思決定を行う際、購買対象のレビュー全てに目を通すことは困難である場合も多い。そのため、製品に関する意見を直感的かつ簡潔に提示することは、ユーザが大量のレビューを閲覧する手間を軽減し、評判情報把握を支援する点から有用であると考えられる。

評判情報の提示方法に関する研究として、打田ら[3]は HK Graph という手法による評価項目のキーワードの抽出、可視化を行うとともに評価項目の時系列変化を二次元空間上で提示している。HK Graph は、係り受け解析に基づくキーワードをグラフで可視化する手法であ

り、解析元の文章のメタデータや単語の統計的な情報などにより、階層的なキーワードグラフの構造が定義される。ユーザは、自身が把握したい内容に合わせて、グラフの構成情報を選択、フィルタリングすることでパーソナライズされた評判情報を得ることができるとしている。また、評価項目の時系列変化について、時系列ごとのレビュー集合間の評価項目の類似度に基づき、二次元マップに可視化することでユーザが製品のどの特徴に着目しているかを把握しやすくし、メーカーが自社製品の機能のアピール効果を測る上でも役立つとしている。

本研究では、評判情報を二次元空間上に可視化して提示することを目的として、従来から研究が盛んであるレビューの評価極性に加え、情報の発見性や意外性の度合いを表す非典型度という指標を導入し、類似した意見を集約する手法を提案する。提案手法では、可視化対象である製品のレビューについて、1 文単位(以降「レビュー文」と表記)で非典型度と評価極性を実数値で評価し、それらの情報を基にレビュー文のクラスタリングを行う。さらに、得られた各クラスタについて重複する内容のレビュー文が現れることが予想されるため、冗長性の削減を行う。

## 2 非典型度の算出

レビュー文の非典型度を評価するために、レビュー文が言及する評価項目を示す評価視点について内容的相違性、及び固有性という性質を考慮する。前者は、ある製品の評価視点集合において、他の評価視点とどれだけ概念的に異なっているかを示している。また、後者は、どれだけ製品固有の内容となっているかを示している。これらの性質は、ユーザが製品のレビューにおいて

記述されていることが想定しにくい内容の評価視点を抽出することを意図している. 非典型度を算出するには, これらの性質を重み付きで考慮する.

## 2.1 評価視点の抽出

評価視点の抽出には, Titov ら [4] が提案した Multi-Grain LDA (MG-LDA) を用いる. MG-LDA とは, Blei ら [5] によって提案された文書生成確率のモデル化手法として知られる Latent Dirichlet Allocation (LDA) を拡張した手法であり, レビュー文書のモデル化を目的としている. MG-LDA では, LDA で推定されるような大域的な特徴を表現するグローバルトピックと隣接センテンス T 文 (ウィンドウ幅) を 1 文書と見立てた時に推定されるような局所的な特徴を表現するローカルトピックを単語の多項分布の形で推定する. レビュー文書に対して MG-LDA を適用したときに推定されるローカルトピックは評価視点を示しており, 本研究ではこの情報を基に評価視点の内容的相違性と固有性を評価する. 具体的には, 可視化対象となる製品の製品カテゴリにおけるローカルトピックの分布から各ローカルトピックが, どれほどこの 2 つの性質を有しているかを評価する. そのため, 可視化対象となる製品のレビュー文書の他に, その製品と同カテゴリの製品のレビュー文書を加えて MG-LDA の入力文書とする. ここで, MG-LDA の素性としては, 名詞, 動詞, 形容詞, 未知語の単語を利用した (本研究では, MeCab [6] を形態素解析器として利用した). また, 「ある」や「こと」などの一部の形態素は, 評価視点を抽出する上で特徴になりにくいことから, ストップワードとして素性から除外した (ストップワードの設定は著者が手動で行なった).

## 2.2 評価視点の内容的相違性

MG-LDA によって得られたローカルトピック集合における各ローカルトピックの内容的な相違性を評価するために Brin ら [7] が提案した PageRank を用いる. PageRank とは, Web ページの重要度を Web のリンク構造から決定するアルゴリズムであり, 重要度の高いページから支持された (リンクされた) ページの重要度を高く判定する. 本研究では, ローカルトピックを Web ページとみなし, あるページからあるページへの支持度を示す推移確率の算出に, ローカルトピック間の語彙集合の異なり具合を用いることで, どのローカルトピックとも内容的に異なるローカルトピックの PageRank 値を高く判定させる. まず, ローカルトピック  $t_i$  と  $t_j$  の語彙集合の異なり具合  $Diff(t_i, t_j)$  を, 集合の類似度を測る指標である Jaccard 係数を用いて式(1)で定義する.

$$Diff(t_i, t_j) = 1 - Jaccard(t_i, t_j) \quad (1)$$

ほとんどの場合で, 2 つのローカルトピック間の Jaccard 係数は小さい値となるため, 語彙集合の異なり具合が

数値に反映されにくい. この問題に対応するために, 得られた全てのローカルトピックの組み合わせにおける  $Diff$  について, 最小値 0, 最大値 1 で正規化を行った  $Diff$  をローカルトピック間の最終的な異なり具合とする. したがって, ローカルトピック  $t_i$  の内容的相違性の評価として PageRank 値  $PR(t_i)$  は式(2)で算出される.

$$PR(t_i) = \frac{1-d}{N} + d \sum_{t_j \in L(t_i)} PR(t_j) tr(t_j, t_i) \quad (2)$$

ここで,  $N$  はローカルトピック数,  $tr(t_j, t_i)$  は推移確率であり,  $Diff$  を  $t_i$  について総和が 1 になるように正規化された値である. また,  $d$  はダンピングファクターであり, Brin ら [7] と同様に 0.85 に設定した. 算出される PageRank 値はどのローカルトピックとも語彙集合の異なり具合が大きいトピックほど大きくなる.

## 2.3 評価視点の固有性

ローカルトピックの固有性の判断に, 製品とローカルトピックとの自己相互情報量を利用する. 本研究では, 自己相互情報量の計算に, 値が  $-1 \sim +1$  の範囲に正規化される正規化自己相互情報量 [8] を用いる. 可視化対象の製品  $I$  のレビュー集合におけるローカルトピック  $t_i$  の正規化自己相互情報量  $NPMI(t_i, I)$  は式(3)で算出される.

$$NPMI(t_i) = \log_2 \frac{P(t_i, I)}{P(t_i)P(I)} / -\log_2 P(t_i, I) \quad (3)$$

ここで,  $P(t_i)$  は入力文書における  $t_i$  が割り当てられた単語の出現確率,  $P(I)$  は入力文書における  $I$  のレビュー集合が含む単語の出現確率,  $P(t_i, I)$  は  $I$  のレビュー集合における  $t_i$  が割り当てられた単語の出現確率である. なお, 計算の都合上,  $NPMI(t_i)$  が負の値を取る場合は  $NPMI(t_i) = 0$  とする. 算出される正規化自己相互情報量は, 対象製品のレビュー集合中で頻出する, または, 対象製品以外のレビューで出現が少ないローカルトピックほど大きくなる.

## 2.4 レビュー文の非典型度

評価視点の内容的相違性と固有性より, レビュー文  $s$  の非典型度  $A(s)$  を式(4)で算出する.

$$A(s) = \sum_{w \in M_s} (PR(t_w) + C_1)^\alpha (NPMI(t_w) + C_2)^{1-\alpha} P_s(t_w) \quad (4)$$

ここで,  $M_s$  は  $s$  におけるローカルトピックが割り当てられた形態素集合,  $t_w$  は  $w$  に割り当てられたローカルトピック,  $P_s(t_w)$  は  $s$  のローカルトピック分布における  $t_w$  の生起確率である. また  $\alpha$  ( $0 \leq \alpha \leq 1$ ) はパラメータである.  $C_1, C_2$  は,  $PR$  もしくは  $NPMI$  が 0 となった場合に対処するための定数項であり, 0.001 と設定している. 式(3)より,

内容的相違性あるいは固有性の大きいローカルピックが占める割合の高いレビュー文であるほど、非典型度は大きくなる。レビュー文のクラスタリングを行う際には、最小値-1, 最大値 1 で正規化した非典型度を用いる。

### 3 評価極性値の算出

レビュー文の評価極性値の算出には機械学習による回帰分析を行い、-1~+1 の範囲で極性値を推定する。レビュー文の素性ベクトルを考える上で、共起する単語により評価極性が変わる評価表現に対処する。

#### 3.1 レビュー文の素性

レビュー文を素性ベクトルに変換する上で、製品のジャンルの違いによる単語評価極性の曖昧性解消は大きな課題の一つである。例えば、自転車のレビューにおける「フレームが軽い」は肯定的な意見であるのに対して、ヘッドホンのレビューにおける「音が軽い」は否定的な意見であると捉えられる。このように同じ評価表現である「軽い」に評価極性の曖昧性が存在する。そこで、本研究では、「フレーム+軽い」のような評価表現と係り受け関係にある単語を考慮した素性ベクトルについて考える。具体的には、Mikolov ら[9]が提案した単語の概念構造を表現した分散ベクトルを得る手法である Skip-gram モデルを用いる。この手法で得られた分散ベクトルは、単語の意味的な演算を行うことができる。例えば、「king-man+woman」の演算結果に最も類似したベクトルとして「queen」を得ることができる。この性質を利用して、評価表現と、係り受け関係にある単語の組(評価表現組)をそれぞれの分散ベクトルの和として表現することで、係り受け関係を考慮したレビュー文の素性ベクトルを得ることができると考えられる。また、「画質+綺麗」と「鮮明」といった評価表現組と評価表現が同義である場合、類似したベクトルを割り当てることができ、評価表現の集約も同時に行えると考えられる。ここで評価表現の係り受け関係として認識する単語は、係り受け解析器 CaboCha[10]の結果より名詞、動詞、形容詞、副詞とし、係り受け関係にある単語が存在しない場合、単に評価表現の分散ベクトルを素性ベクトルとして利用した。また、評価表現として、小林らの評価値表現辞書[11]に登録されている評価表現の形態素を利用した。

#### 3.2 レビュー文の極性値ラベル

多様な評価表現を回帰器に学習させるためには、膨大な量のレビュー文が必要となることが想定される。しかしながら、これらのレビュー文の一つ一つに人手で極性値をアノテーションすることは、手間がかかり極めて困難である。そこで、レビューにメタデータとして付与されている 1~5 段階の評点(評点の付いていないレビューは対象外とした)を極性のラベルとして利用する。具体的

には、レビュー文の極性値として、そのレビュー文を含むレビューに付与されている 1~5 段階の評点に対して、それぞれ 0.5 刻みに-1~+1 の極性値を割り当て、レビュー文の極性値として用いる。ここで、評価表現と同じ文節または、評価表現の係り先の文節に否定を示す助動詞である「ない」、「ん」、「ぬ」が含まれる場合、極性値の符号を反転させたラベルを与える。

#### 3.3 評価極性値の推定

評価極性値の推定を推定する回帰器として、回帰や識別に用いられる機械学習アルゴリズムであるランダムフォレスト[12]を用いる。ランダムフォレストは学習データが少数である場合、過学習しやすいという問題があるが、本研究では、ラベルを自動で付与できることから、大量の学習データを用意することができる。これにより、回帰器は高い汎化性能を得られるのではないかと考えられる。極性値を推定するレビュー文に対して 3.1 の方法で素性ベクトルに変換し、ランダムフォレストによる回帰予測を行って得られた値をそのレビュー文の極性値とする。ここで、1 つのレビュー文より 2 つ以上の評価表現組が得られる場合、各評価表現組の極性値を推定後、平均をとった値をそのレビュー文の極性値とする。また、回帰結果が-1 未満、あるいは+1 を超える値となった場合は、それぞれ極性値を-1, +1 とする。

### 4 レビュー文のクラスタリングと冗長性の削減

非典型度と評価極性に基づくレビュー文の集約方法として、k-means++法[13]によるクラスタリングを行う。ここで、レビュー文間の類似度には、非典型度と評価極性から成る二次元ベクトル同士のユークリッド距離を用いる。これにより得られるクラスタには、類似文が複数出現し、内容に冗長性が生じると予想されるため、ユーザに提示する情報としては不適切である。そこで本研究では、Carbonell[14]らによって提案された情報検索における内容の網羅性と冗長性の削減を考慮した指標である Maximal Marginal Relevance (MMR) を利用する。MMR 値の大きいレビュー文のみを抽出することで、内容的に重複のないクラスタを作成する。ここで、 $R$  をクラスタのレビュー文集合、 $S$  をクラスタに残すレビュー文集合として、レビュー文  $s$  の MMR 値  $MMR(s)$  を式(5)、式(6)で算出する。

$$MMR(s) = \arg \max_{s \in R, S} [\lambda \text{Score}(s) - (1 - \lambda) \max_{s' \in S} (\text{Sim}(s, s'))] \quad (5)$$

$$\text{Score}(s) = \frac{\log_2 |M_s|}{|M_s|} \sum_{w \in M_s} \text{tfidf}(w) \quad (6)$$

ここで、 $\text{Sim}(s, s')$  は、 $s$  と  $s'$  とのコサイン類似度である。 $\text{Score}(s)$  は  $s$  の重要度を示しており、 $s$  の形態素集合

$M_s$  の長さ と tf-idf 法により算出される形態素  $w$  の重要度  $tfidf(w)$  により定義される. また,  $\lambda$  ( $0 \leq \lambda \leq 1$ ) はパラメータである. 集合  $R \setminus S$  中で計算された MMR 値が最も大きいレビュー文を集合  $S$  に加え, その都度  $R \setminus S$  中のレビュー文の MMR 値を更新する. この操作を任意の回数繰り返すことで, ユーザに提示すべきクラスタリング結果  $S$  を得る.

## 5 評価実験

### 5.1 実験方法

提案手法の有効性を検証するために, 5 名の被験者 (全て 20 代男性) による評価実験を行った. 提案手法によって得られた各クラスタで MMR 値の高い文に対して, 非典型度と評価極性値を 10 段階で評価してもらった. 今回はクラスタ数を 7, 各クラスタで MMR 値上位 5 文の計 35 文を評価対象のレビュー文とした.

### 5.2 レビューデータと対象製品

楽天データセット[15]の楽天市場において公開されている約 6400 万件のレビューの利用した. 対象製品のレビューとして折りたたみ自転車のレビュー 558 件 (うち評価表現を含むレビュー文は 917 件) を利用した.

### 5.3 非典型度算出の実験設定

MG-LDA の入力文書として, 対象製品のレビュー, 及び対象製品以外の「折りたたみ自転車」カテゴリに含まれる製品レビュー 3,6497 件を用いた. MG-LDA の設定は, グローバルトピック数を 40, ローカルトピック数を 20, ウィンドウ幅を 3 とした. また, Titov ら[4]を参考にモデルの推定には Collapsed Gibbs Sampling を用い, 繰り返し回数を 800 回とした.

### 5.4 評価極性算出の実験設定

Skip-gram モデルの単語ベクトルの学習には, 各評点で件数が均等になるように全体のレビューデータよりランダムにサンプリングしたレビュー 5,623,205 件を利用した. 今回は Skip-gram モデルにより得られる単語ベクトルの次元数は 200 とした. ランダムフォレストの学習には, 単語ベクトルの学習に用いたレビュー集合より得られた 21,513,248 件の評価表現組から各評点とも 50 万件となるように, ランダムにサンプリングしたものをを用いた (計 250 万件). Skip-gram モデルとランダムフォレストの実装には, それぞれ Python のライブラリである gensim[16] と scikit-learn[17] を利用した.

## 6 実験結果と考察

提案手法によって得られたクラスタを  $V_1 \sim V_7$  としたとき, レビュー文のプロット結果及びクラスタリング結果を図 1 に示す. ここで, 横軸は非典型度, 縦軸は評価極性を示している. また, 各クラスタを代表する MMR 値が

高くなったレビュー文を表 1 に示す. MMR 値が高い文に比較的長文が多い理由として, 文の重要度を図る際の文長による重要度の補正の影響が大きかったためであると思われる. しかし, レビュー文の情報の有用性を考慮すると, この補正は適切であると考ええる.

### 6.1 非典型度

MG-LDA により得られたローカルトピックとその代表語句を表 2 に示す. また, 得られた一部のローカルトピックについて内容的相違性, 固有性の大きい順に表 3 に示す. 表 3 を見ると, 「値段」トピックが内容的相違性と固有性がともに高くなり, クラスタ  $V_7$  の代表文にも内容が反映されていることが分かる (表 1). また, 内容的相違性の最も高い「組み立て難度」トピックは, 固有性が最も低くなっており, 非典型度中位 (クラスタ  $V_1$ ) にプロットされている (図 1).

提案手法が推定した非典型度の推定値とユーザ評価の相関係数はすべての被験者で負の相関を示した (表 4). このときのユーザ評価の一致率を示す重み付き  $\kappa$  係数 (2 次の重み) の平均は 0.586 (0.384~0.745) であった. 非典型度は評価が被験者の主観に依存する部分が大きく, 一致率は全体的にやや低めの値となった. 本実験で相関係数が負の値を示した理由として, 評価実験に用いた製品に関して適切に評価を下すための判断材料が不足していた点が挙げられる. 例として, 今回の実験では, システムにより「値段の安さ」に関する情報の非典型度が高く判定されており, 実際に対象製品の値段は, 他の同カテゴリ製品に比べ相場を大きく下回るものであり, 非点軽度の高い情報であったと言える. しかしながら, 被験者らは折りたたみ自転車の相場に関する知識を持ち合わせていなかったため, 比較的ありふれた評価視点である「値段」に対して, 非典型度を低く評価したと考えられる. このことから, 被験者が適切に主観評価を行えるような実験環境へ改善する必要である.

### 6.2 評価極性

評価極性の提案手法の推定値とユーザ評価の相関係数は全体的に高い値を示した (表 4). このときの  $\kappa$  係数 (2 次の重み) の平均は 0.796 (0.723~0.850) であり, 高い一致率を示している. これらの結果から, 提案手法における係り受けを考慮したレビュー文の素性は, 評価表現の肯否定の度合いを適切に表現できていると考えられる. しかしながら, いくつかのレビュー文で極性値を大きく誤って推定しているものが確認された. 以下にその例を示す.

- 自転車がパンクしてパンク修理するか迷っていたときに, ポチっと購入できる安さはやはり魅力です. 提案手法が推定した上記の文の評価極性値は -0.133 であり, 本来, 正の極性値が推定されるべき文に負の極性値を推定してしまっている. この例では, 購入背景の

ような製品に対する評価ではない部分に「迷う」という否定的な評価表現が含まれており、極性値推定誤りの原因となっている。このことからレビュー文の前処理として、製品に対する言及のみを抽出することが必要である。

### 7 まとめと今後の課題

本研究では、レビュー文の非典型度と評価極性を実数値で評価する手法について提案した。また、それらの特徴を基にレビュー文をクラスタリングする手法について提案した。

今後の課題として、非典型度の評価に関して、より実環境を想定した評価実験方法について検討する必要がある。具体的には、被験者に対して、対象製品以外の同カテゴリ製品に関する情報(レビューデータや商品情報)を知識として与えた状態で、それらの情報に対して対象製品レビューを相対的に評価してもらうなどの工夫が必要である。評価極性の推定では、製品に対する評価部分以外の評価表現に対しても極性値を推定してしまう問題に対処するために、レビュー文の内容が製品に対する言及か判定する方法について検討する。また、レビュー文の具体的な可視化方法についても検討していきたい。

### 謝辞

本研究では、楽天株式会社と国立情報学研究所が提供するレビューデータを利用した。ここに深い感謝の意を表す。

### 参考文献

[1] 経済産業省:電子商取引に関する市場調査, 2014.  
 [2] 総務省:平成 26 年版 情報通信白書, 2014.  
 [3] 打田裕樹, 吉川大弘, 古橋武ほか:Web ユーザレビューにおける評価情報の時系列変化の可視化, 知能と情報, Vol. 22, No. 3, pp.377-389, 2010.  
 [4] Titov, I., and McDonald, R.: Modeling online reviews with multi-grain topic models, Proc. of WWW, pp. 111-120, 2008.  
 [5] Blei, D. M., Ng, A. Y., and Jordan, M. I.: Latent dirichlet allocation, Journal of Machine Learning Research, Vol.3, pp. 993-1022, 2003.  
 [6] Kudo, T., Yamamoto, K., and Matsumoto, Y.: Applying conditional random fields to Japanese morphological analysis, Proc. of EMNLP, pp. 230-237, 2004.  
 [7] Brin, S., and Page, L.: The anatomy of a large-scale hypertextual web search engine, Proc. of WWW, pp.107-117, 1998.  
 [8] Gerlof, B.: Normalized (pointwise) mutual information in collocation extraction, Proc. of GSCL, pp.31-40, 2009.  
 [9] Mikolov, T., Chen, K., Corrado, G., et al: Efficient

estimation of word representations in vector space, Proc. of ICLR, 2013.

[10] 工藤拓, 松本裕治:チャンキングの段階適用による係り受け解析 Cabocha, 情報処理学会論文誌, Vol. 43, No. 6, pp1834-1842, 2002.  
 [11] 小林のぞみ, 乾健太郎, 松本裕治ほか:意見抽出のための評価表現の収集, 自然言語処理, Vol.12, No. 2, pp.203-222, 2005.  
 [12] Breiman, L.: Random forests, Machine Learning, Vol.45, pp.5-32, 2001.  
 [13] David, A.: k-means++: The advantages of careful seeding, Proc. of ACM/SIAM SODA, pp.1027-1035, 2007.  
 [14] Carbonell, J., and Goldstein, J.: The use of MMR, diversity-based reranking for reordering documents and producing summaries, Proc. of ACM/SIGIR, pp.335-336, 1998.  
 [15] 楽天データセット, <http://rit.rakuten.co.jp/rdr/index.html>  
 [16] gensim, <http://radimrehurek.com/gensim/>  
 [17] scikit-learn, <http://scikit-learn.org/>

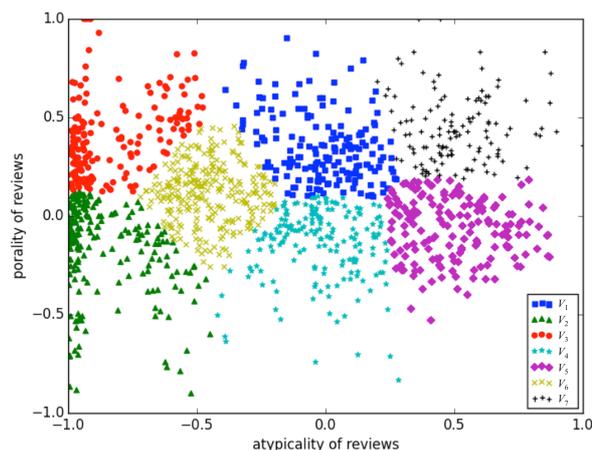


図1 レビュー文のプロット結果

表2 ローカルトピックと代表語句

トピック	代表語句
配送	届く, 商品, 早い, 対応, 到着, 発送, 迅速, 配送, 梱包, 配達, 無事
可搬性	コンパクト, 便利, 乗せる, 積む, サイズ, 持ち運び, 入れる
乗り心地	乗り, 感じ, サスペンション, 快適, 小回り, 利く, 衝撃, 吸収, 安定
色	色, 気に入る, カラー, ブラウン, 可愛い, ピンク, かわいい, 目立つ
値段	値段, 買い物, 安い, 割, 価格, 手頃, 品質, 安価, リーズナブル
デザイン	デザイン, かわいい, 可愛い, 写真, カッコいい, おしゃれ, 実物, 形
ギア・ブレーキ	調整, ブレーキ, 必要, ギア, 変速機, 効く, 調節, 状態, 変速, グリップ

表 1 クラスタのレビュー文

クラスタ	レビュー文
$V_1$ (青, □)	折りたたみタイプは初めてだったのですが、コツを掴めば非常に簡単に畳むことができ、乗り心地も良好なので、非常にいい買い物を下と思います。 他の方のレビューを参考にシールを全て剥がすとマットブラックの印象がぐっと良くなり満足感がアップしました。
$V_2$ (緑, △)	制限外だといえ、かなりきつい体制になったので、大人の男性は厳しいと思います。 ハンドルが格好いい分、畳んだ時の横幅が若干広く、通勤に使っていますが、電車から降りるときに人にぶつからないか心配です。
$V_3$ (赤, ○)	東京内を江東区から秋葉、新宿、赤坂、六本木、芝公園、新橋、日本橋を通り 7 時間掛けて江東区まで走りましたが折りたたみ式とは思えないほどよく走ります。 非常にオシャレな一品ですし、車載も簡単ですので、家族みんなで楽しめると思います。
$V_4$ (シアン, ※)	姿勢が前かがみになる点とブレーキの角度が地面と平行なので慣れるまで怖かったです。 付属の工具でブレーキワイヤーを緩めて調整しようとしたのですが、ワイヤーが短くて、途中のワイヤー保護金具を外さないと調整ができず大変でした(^_^;)
$V_5$ (紫, ◇)	自転車のほうも簡単に組立でき満足してますが、ブレーキが片利きしたりホイールがぶれていたり、自転車屋さんで買えば完璧に調整されている部分の調整が不十分でしたし、塗装も弱くすぐに傷が入ってしまいます。 ハンドルが遠くて前かがみにならないといけないのが体勢的にしんどいです。
$V_6$ (黄, ×)	こぐとズボンの裾が金具に引っかかることがあるそうで、少し金具を内側に倒すなど工夫が必要なようです。 折りたたみ時にロック解除に手こずりましたがコツをつかめば大丈夫かな。
$V_7$ (黒, +)	2 台目の折り畳み自転車ですが、値段の割には高級感があって良いです。 一万円以下でシマノ製の変速機付だったのでとてもリーズナブルだと思って買いました。

表 3 トピックの性質

順位	内容的相違性	固有性
1	組み立て難度	値段
2	色	サドル
3	デザイン	ギア・ブレーキ
4	値段	塗装
5	乗り心地	デザイン
⋮	⋮	⋮
18	ギア・ブレーキ	大きさ
19	重量	可搬性
20	不明	組み立て難度

表 4 評価実験結果

被験者	非典型度	評価極性
A	-0.427	0.722
B	-0.421	0.704
C	-0.364	0.648
D	-0.394	0.671
E	-0.220	0.728