

# 非典型度と評価極性に基づくカスタマーレビューのクラスタリングとその可視化

古橋 慎之介<sup>1</sup>, 内田 理<sup>2</sup>

<sup>1</sup> 東海大学大学院工学研究科情報理工学専攻  
4bdrm019@mail.tokai-u.jp

<sup>2</sup> 東海大学情報理工学部情報科学科  
o-uchida@tokai.ac.jp

## 1. はじめに

BtoC-EC の市場規模は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 年度の経済産業省による調査[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おり、2020 年までに 20 兆円台への成長が見込まれている。ところで、消費者に対して EC サービスを提供する EC サイトでは、一般に、誰でも商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、商品購入時の参考情報として、「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、商品レビューはユーザの購買行動に大きく寄与していると言える。しかしながら、投稿されているレビューは商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、商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。

本稿では、評判情報を直感的に分かりやすい形式で提示することを目的として、情報の非典型性を表す非典型度と意見の肯否定の度合いを表す評価極性という指標を導入し、類似した特徴を持つ意見を集約する手法を提案する。提案手法の流れは、基本的に先行研究[3]と同様であるが、非典型度の算出、レビュー文の集約精度向上のため、それぞれのアプローチに変更を加えている。また、先行研究[3]では対象商品のレビューが約 500 件と小規模であったため、よりレビュー数の多い商品を利用した場合での提案手法の有効性について検証する。さらに、推定された非典型度、評価極性を基にレビュー文を二次元空間上に可視化し、評判情報を閲覧できるシステムのプロトタイプの実装を行った。

## 2. 非典型度の算出

レビュー文の非典型度の評価に、その文が言及する評価項目である評価視点について内容的相違性、及び固有性という二つの性質を考慮する。前者は、ある商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。後者は、どれだけ商品固有の情報となっているかを示している。これらの性質は、ユーザが商品レビューにおいて記述されていることが想定しにくい希少な評価視点を持つ情報の抽出を意図している。

### 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、商品レビューといった評判文書の生成確率のモデル化手法であり、レビュー一件を文書とした時に推定されるような大域的特徴を持つ

グローバルトピックと隣接センテンスを文書と見立てた時に推定されるような局所的特徴を持つ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、レビューにおける「画質」や「音質」のような評価項目が表現されており、このようなトピックを評価視点として扱うことで、その内容的相違性と固有性を評価する。具体的には、可視化対象となる商品の商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、この 2 つの性質をいかに有しているかを評価する。そのため、可視化対象となる商品のレビュー文書の他に、その商品と同カテゴリの商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、形態素解析器として MeCab[5] を利用し、ユーザ辞書として mecab-ipadic-NEologd[6]を用いた)。また、著者がストップワードとして定義した単語は素性から除外した。

### 2.2. 評価視点の内容的相違性

ローカルトピックの内容的相違性の評価に 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手法であり、重要度の高いページから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 ページに見立てることで、PageRank 値の算出を行う。先行研究[3]では、PageRank 値算出のための推移確率を語彙集合の異なりに基づき定義していた。しかしながら、このアプローチでは、PageRank 値に幅が生じにくく、非典型度の算出時に内容的相違性の影響力が弱まってしまう問題があった。そこで今回は、あるページからあるページへのリンク(ローカルトピック間の有向リンク)の有無を、ローカルトピック間の語彙集合の異なり具合により決定し、それに基づき推移確率を定義する。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)で算出する。

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

そして、得られた全ての  $Diff$  (自身との組み合わせを除く)について、最小値 0, 最大値 1 で正規化を行った  $Diff$  に対して、閾値以上でリンクを持たせることで推移行列を作成する(本研究では、 $Diff(t_i, t_j) > 0.6$  の時、 $t_i$  から  $t_j$  へリンクを持たせた)。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、 $L(t_i)$  は  $t_i$  へリンクするローカルトピック集合である。 $tr(t_j, t_i)$  は推移行列を基に求められる  $t_j$  から

ら  $t_i$  への推移確率であり, 1 を発リンクの数で除算した値となる. また,  $d$  は減衰係数であり, 0.85 に設定した.

### 2.3. 評価視点の固有性

ローカルピックの固有性の評価に, 商品とローカル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. 対象商品  $I$  のレビュー集合におけるローカルピック  $t_i$  の正規化自己相互情報量  $NPMI(t_i, I)$  は式(3)で算出される.

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

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, 対象商品のレビュー集合中で頻出する, または, 対象商品以外のレビューで出現が少ないローカルピックほど大きくなる.

### 2.4. レビュー文の提示

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.

$$A(s) = \sum_{w_l \in MI_s} (PR(t_{w_l}) + C_1)^\alpha (NPMI(t_{w_l}) + C_2)^{1-\alpha} P_s(t_{w_l}) \quad (4)$$

ここで,  $MI_s$  は  $s$  におけるローカルピックが割り当てられた形態素集合,  $t_{w_l}$  は  $w_l$  に割り当てられたローカルピック,  $P_s(t_{w_l})$  は  $s$  のローカルピック分布における  $t_{w_l}$  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 と設定する. 式(4)より, 内容的相違性あるいは固有性の高いローカルピックが占める割合の大き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法による回帰分析を行い,  $-1 \sim +1$  の範囲で極性を推定する. 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化す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, 商品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「軽い」という同様の評価表現を用いても, 共起する単語が「フレーム」と「音」では評価極性が変わってしまう問題がある. そこで, 本研究では, 「フレーム+軽い」のような評価表現と共起する単語を考慮した素性ベクトルを構築する. 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単語ベクトルは, 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単語ベクトルとして「queen」を得ることができる. この性質を利用して, 評価表現と, 共起する単語の組(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, 評価表現とその共起単語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. ここで評価表現の共起関係にある単語として認識する対象は, 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, 単に評価表現の単語ベクトルをレビュー文の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を利用した.

### 3.2. 学習データの極性値ラベル

多様な評価表現を学習させるためには, 膨大な量のレビュー文が必要である. しかし, これらのレビュー文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, 製品に対する総合評価としてレビューに付与される 1~5 段階の評点を極性値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文に付与されている 1~5 段階の評点に対して, それぞれ 0.5 刻みに  $-1 \sim +1$  の値を割り当てる. 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」「無い」「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. レビュー文の極性値推定

極性値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法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, レビューへのラベル付与方法の性質上, 大規模な学習データが利用可能である. 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. 極性値の推定には, 評価表現を含む対象製品の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(3.2 と同様に否定を示す助動詞を含む場合は, 推定値の符号を反転させる). ここで, 一つのレビュー文より二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, 平均の値をそのレビュー文の極性値とする. また, 回帰による推定結果が  $-1$  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削減

ユーザが目を通すレビュー文書量を減らすためには, 似た特徴を持つ評判情報を集約し, 整理してユーザに提示することが望ましい. 本研究では, 評判情報の集約方法として, 非典型度と評価極性から成るレビュー文に対して k-means++法[13]を適用してクラスタリングを行う. こうして得られたクラスタには, 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, 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, 内容的重複のないクラスタを

作る。  $R$  をクラスタのレビュー文集合、  $S$  を提示するレビュー文集合として、レビュー文  $s$  の MMR 値  $MMR(s)$  を式(5)、(6)で算出する。

$$MMR(s) = \operatorname{argmax}_{s \in R \setminus S} \left[ \lambda \operatorname{Score}(s) - (1 - \lambda) \max_{s' \in S} (\operatorname{Sim}(s, s')) \right] \quad (5)$$

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

ここで、  $\operatorname{Sim}(s, s')$  は、  $s$  と  $s'$  のコサイン類似度であり、  $\lambda$  ( $0 \leq \lambda \leq 1$ ) はパラメータである。また、  $\operatorname{Score}(s)$  はクラスタにおける  $s$  の重要度を示しており、  $b$  を各クラスタにおける最大文長として 0-1 で正規化された  $s$  の長さ、  $\operatorname{tf-idf}$  法により算出される形態素  $w$  の重要度  $\operatorname{tfidf}(w)$  (各クラスタにおける L2 ノルムの値を採用)により定義される。これは、先行研究[3]での文のスコアリング方法において、文長や  $\operatorname{tf-idf}$  値に正規化を行っていなかったために、長すぎる文や重要語句を含む短い文の重要度が高く評価されやすいという問題に対処している。集合  $R \setminus S$  中の 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。提案手法により得られた各クラスタの代表文に対して、非典型度と評価極性を10段階で評価してもらった。被験者の評価値と提案手法の推定値との相関係数を測ることで精度評価を行った。

実験のレビューデータとして、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。可視化対象となる商品として折りたたみ自転車(レビュー4785件、うち評価表現を含むレビュー文12133文)を利用した。今回はクラスタ数を7、各クラスタでMMR値が上位5文の計35文を評価対象のレビュー文とした。

### 5.2. 非典型度算出の実験設定

MG-LDAの入力文書として、対象商品のレビュー、及び対象商品以外の「折りたたみ自転車」カテゴリに含まれる商品レビュー49,301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。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。

### 5.3. 評価極性値算出の実験設定

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文書約560万件を利用した。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約2000万件の評価表現組から各評点とも100万件となるように、ランダムにサンプリングしたものをを用いた(計500万件)。

## 6. 実験結果と考察

5名の被験者の評価値と提案手法の推定値との相関係

数を表1に示す。また、評価実験に用いたレビュー文の一部とその非典型度、評価極性の算出結果を表2に示す。

### 6.1. 非典型度の算出

非典型度の被験者評価の重み付き一致率(一次の重み)の平均は0.277(0.042~0.545)であった。一部の被験者で著しく一致率が下がった原因として、被験者の評価方法(相対評価と絶対評価)の違いが挙げられる。これについては全ての被験者で評価方法を統一するなどの対応が必要である。表1より、非典型度の相関係数は全ての被験者で無相関または負の相関を示した。このことや一致率の平均の低さから、非典型性は、被験者個々の背景知識に依存する傾向にあり、主観性が強い感情であるため、被験者ごとに非典型度を高く評価する評価視点が大きく異なることが分かる。また、今回の評価実験に用いたレビュー文のうち複数の評価視点を含むレビュー文の非典型度を低く評価する傾向にあり、推定精度を下げる原因になっていることが分かった。そのようなレビュー文の例を以下に示す。

- 組み立ては簡単、自転車はきれいで、値段も安いし購入は正解でした。

提案手法のMMRに基づくクラスタの代表文の選択方法では、文の重要度の評価に  $\operatorname{tf-idf}$  法による単語スコアが大きく関わっている。そのため、評価視点に関連する単語のスコアは比較的高く評価されやすく、同文長であれば複数の評価視点を含む文が選択されやすくなっている。しかしながら、そうように選択されたレビュー文は、単体の評価視点のみを持つレビュー文に比べ、情報量が少なく、被験者は非典型性を感じにくい傾向にあったのではないかと考えられる。このことから、クラスタの代表文の選択方法にも改善が必要であることがわかる。

### 6.2. 評価極性値の算出

評価極性の被験者評価の重み付き一致率の平均は0.806(0.708~0.881)であった。評価極性は、非典型度に比べ主観的な感情が入りにくいため、一致率は高い値となった。表1より評価極性の相関係数は、全ての被験者で正の相関を示しており、高い推定精度を得られていることが分かる。しかしながら、商品特有の評価表現では極性値の推定誤りが起きる場合が確認された。その例を以下に示す。

- 一緒についてきたライトは自転車に固定するベルトの箇所が若干緩くしっかり固定されません。

このレビュー文例は否定的な意見であるが、0.244という肯定的な極性値が推定されていた。このように、「ベルト+緩い」といった折りたたみ自転車の特殊性が強い評価表現は、今回の全体よりランダムにサンプリングした学習データのみでは十分に網羅できないことが推定誤りに繋がったのではないかと考えられる。

### 6.3. 可視化システム

レビュー文の可視化システムのイメージを図1に示す。システムの実装には、HTML5、JavaScript、jQuery、及びグラフ描画ライブラリであるHighcharts[16]を用いた。図1は、今回の実験で利用した対象商品のレビューをWEBブラウザ上で可視化したときの表示例である。画面左には、縦軸を評価極性、横軸を非典型度としたときのレビュー文のプロット、及びクラスタリング結果が可視化されている。各レビュー文の内容はプロット点にカーソルを合わせることで表示されるようになっている。

また、画面右には、クラスタ数を任意に変更することができるスライダーと各クラスタの代表文(MMR 値上位 5 文)が表示されている。クラスタ数の変更例として、図 1(a)にクラスタ数 7(代表文 35 文)のとき、図 1(b)にクラスタ数 4(代表文 20 文)のときの表示例を示す。これを見ると図 1(b)は図 1(a)に比べ、レビュー文の集約粒度が粗くなっており、右画面で閲覧できるクラスタの代表文が減少することが分かる。このように、ユーザは左のプロット、及びクラスタリング結果を参照しながら、読みたい文書量や評判情報の集約粒度を容易に選択することができるのではないかと考える。

## 7. まとめと今後の課題

商品レビューの非典型度、評価極性を推定し、それらから評判情報をクラスタリング・可視化する手法について提案した。

今後の課題として、ユーザの背景知識を考慮した非典型度の算出方法の検討が挙げられる。また、評価極性推定のための学習データについて、効率よく評価表現を網羅する方法について検討していきたい。さらに、可視化システムについては、被験者によるユーザビリティなどの検証実験を実施したい。

## 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] 古橋慎之介, 内田理:可視化を目的とした非典型度と評価極性に基づくカスタマーレビューのクラスタリング, ARG 第7回 Web インテリジェンスとインタラクション研究会, 2015.
- [4] Titov, I., and McDonald, R.: Modeling online reviews with multi-grain topic models, Proc. of WWW, pp. 111-120, 2008.
- [5] Kudo, T., Yamamoto, K., and Matsumoto, Y.: Applying conditional random fields to Japanese morphological analysis, Proc. of EMNLP, pp. 230-237, 2004.
- [6] mecab-ipadic-NEologd, <https://github.com/neologd/mecab-ipadic-neologd>
- [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, pp.1834-1842, 2002.
- [11] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: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.

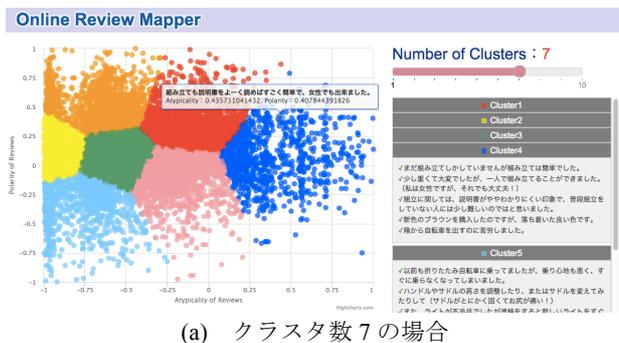
- [15] 楽天データセット, <http://rit.rakuten.co.jp/rdr/index.html>
- [16] Highcharts, <http://www.highcharts.com/>

表 1 評価実験結果

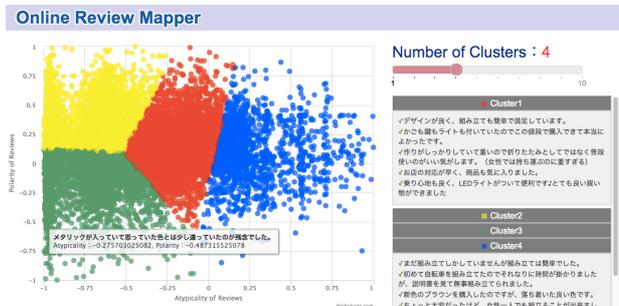
被験者	非典型度	評価極性
A	-0.153	0.232
B	-0.284	0.383
C	-0.232	0.438
D	-0.017	0.326
E	-0.060	0.375
平均	-0.149	0.351

表 2 非典型度と評価極性の算出結果例

レビュー文	非典型度	評価極性
デザインが良く、組み立ても簡単で満足しています。	-0.073	0.558
組立に関しては、説明書がややわかりにくい印象で、普段組立をしていない人には少し難しいのではと思う。	0.271	-0.048
新色のブラウンを購入したのですが、落ち着いた良い色です。	0.225	0.294
作りがしっかりしていて重いので折りたたみとしてではなく普段使いのいい気がします。(女性では持ち運ぶのに重すぎる)	-0.202	-0.216



(a) クラスタ数 7 の場合



(b) クラスタ数 4 の場合

図 1 可視化システムの画面例