

大規模 EC サイトの商品レビュー傾向分析

林 驍[†] 伊東栄典[‡] 廣川佐千男[‡]

消費者による商品レビューは、購入時判断に重要である。広告には売り手の意思が入るのに対し、レビューは購入した客の正直な感想であるため信用できる。ネット上の通信販売では情報提供として消費者による商品レビューを記載している。我々は、ネット上の EC サイトにおける商品レビューからの知識発見について研究している。やらせやステマ、誹謗などのノイズを除去し、かつ消費者に有用な情報が無いレビューを除去したいと考えている。本論文で、楽天市場で販売される商品についてのレビュー文を対象に、レビュー投稿行為の不自然さや、レビュー文の特徴分析を行った。また、購入していない商品に対するレビュー文を調べることで、楽天市場における特殊な商品レビューの傾向を理解した。これらの結果について報告する。

Analysis of customer reviews for items on a large-scale electric commerce site

Xiao Lin[†] Eisuke Ito[‡] Sachio Hirokawa[‡]

Customer reviews are important for other customers. Although advertisement may include seller's will, customer reviews may express honest impressions by customers who bought items. In case of online shopping site, customer reviews and reputation points are important to gain customer satisfaction. For these days, stealth marketing pretending customer review become serious problem on online shopping. Slander review is also problem. The authors want to discover knowledge from reviews of a large scale EC site. They want to remove noise reviews stealth marketing reviews and slander reviews, and also classify information less reviews. In this paper, they researched un-natural review posting like spam mail sending. They extract feature words in reviews which were posted by not-bought customers. They found some tendencies of reviews by not-bought customers.

1. はじめに

消費者による商品についての口コミ評価やレビューは、商品購入時の重要な判断材料として用いられてきた。テレビや雑誌に掲載される宣伝広告には、製造元やサービス提供側などの売り手側の意思が入るため素直に信用できない。一方、消費者レビューは実際に商品やサービスを購入した客の正直な感想であるため信用でき、そのため商品購入の際の重要な判断材料になる。

近年、電子商取引（EC, Electric Commerce）が普及している。ネット上の通信販売サイトでは、実店舗と異なり、商品の実物を見たり触ったりできない。そのため、販売サイトでは情報提供として、消費者による商品への評価点と商品レビューを記載できるようにしている。

消費者による商品への評価点は、後から購入する消費者の重要な判断材料になる。多くの消費者から良い評価点を与えられた商品は、良い商品であろうと期待できる。多くの消費者が与えた評価点は、協調フィルタリングを用いた商品推薦にも活用できる。

しかし最近、商品への評価点や、商品へのレビューは有用であるため、それらの悪用が行われている。宣伝や販売促進のため、売り手側の意思が入った「やらせ」や「ステ

マ」などが発生している。実際、本研究が対象とする楽天市場でも、架空注文による販売数の偽装や、ウソの口コミを投稿によるステマが発覚しており、通販への不信がさらに広がりかねない状況が発生している[1]。また、商品の評価とは関係ない、販売会社への誹謗中傷の文章がレビュー欄に記載される場合もある。他にも、スマートフォン用のアプリ提供・販売サイトでは、一時的に高い順位にするために、大人数から多数ダウンロードされ、かつ評価点が高い状況にすることも行われている。このような、商品評価点や商品レビューの悪用は、健全な商行為を妨害するため、これらを防止する必要がある。

我々は、ネット上の EC サイトにおける商品レビューからの知識発見について研究している。やらせやステマ、誹謗などのノイズとなるレビューの除去を行いたい。また、有用な情報の低いレビューについての除去も行いたいと考えている。内容の薄いレビューのような、有用度の低いレビュー文の除去についても検討している。

今回、楽天市場で販売される商品についてのレビュー文を対象に、レビュー投稿行為の不自然さや、レビュー文の特徴分析を行った。また、購入していない商品に対するレビュー文を調べることで、楽天市場における特殊な商品レビューの傾向を理解した。これらの結果について報告する。

本論文の構成を述べる。第2節では、本研究で用いた楽天市場データセットについて述べる。第3節では、楽天市場データセットに含まれる、不自然なレビュー投稿についての傾向を述べる。第4節で、購入していない商品に対する

[†] 九州大学大学院 システム情報科学府
Faculty of Information Science and Electrical Engineering, Kyushu University.
[‡] 九州大学 情報基盤研究開発センター
Research Institute for Information Technology, Kyushu University.

レビュー（未購入レビュー）の傾向を説明する。また第 5 節で、商品評価点毎の未購入レビューの傾向について述べる。最後に第 6 節でまとめと今後の課題を述べる。

2. 楽天市場データセット

本研究が対象とするデータは、楽天株式会社が国立情報学研究所の「情報学研究リポジトリ」[2]を通じて研究者に提供している楽天市場データセット[3]である。このデータセットには、楽天市場で販売される商品の情報と、商品へのレビューが含まれている。

2.1 データ数

楽天データセットに含まれるデータ数を表 1 に示す。商品は約 1 億 5600 万種類ある。利用者から投稿された商品へのレビューは約 6400 万件ある。期間は、2010 年 1 月から 2012 年 12 月までの 3 年間（36 ヶ月）である。

表 1 楽天市場データセットの件数

	件数	サイズ
商品	約 156,000,000	約 48GB
レビュー	63,553,087	約 42GB

2.2 レビューされた商品の数

2010 年 1 月から 2012 年 12 月までの 36 ヶ月間、各月におけるレビューされた一意な商品数を図 1 に示す。レビューされる商品が、徐々に増加していることが分かる。

2.3 レビューアの割合

レビューを投稿する人（レビューア）について、一意なレビューア数と、レビューアの性別（男、女、匿名）を調査した。図 2 にその結果を示す。

図 2 を見ると、全期間において性別を指定しない匿名の割合が多いことが分かる。男女を明らかにしているレビューアについては、女性が多いことがわかる。男性の割合はあまり変化しないのに対し、女性の割合は徐々に減少して、代わりに匿名の投稿者が増えている。

2.4 出現頻度と単語数

次に、レビュー文に出現する全単語について、単語の出現頻度を調べた。レビュー文からの単語抽出にはオープンソースの日本語形態素解析エンジン MeCab [4]を用いた。

図 3 に、横軸に単語の出現頻度を、縦軸にその出現頻度の単語数を描画した散布図を示す。両軸とも対数尺度になっている。

出現頻度が低い単語数は、両対数尺度で直線の分布を示しており冪分布となる。一方、高頻度の部分はゆるい曲線を描いている。

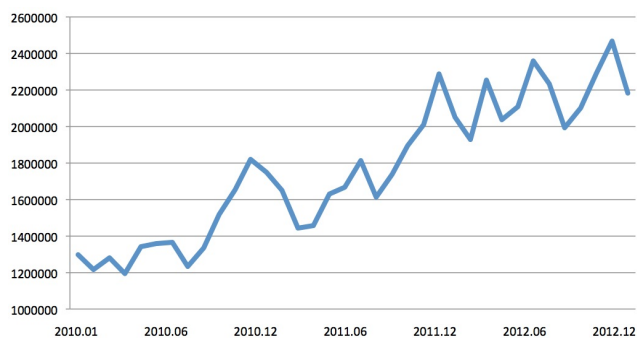


図 1 レビューされた商品の数
(2010 年 1 月～2012 年 12 月)

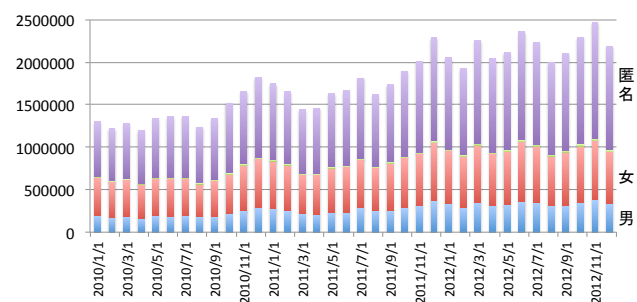


図 2 レビューア数の推移

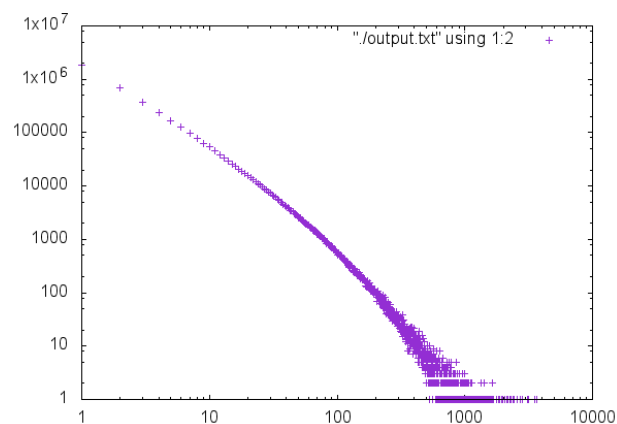


図 3 頻度の順位-単語の出現頻度

3. 不自然なレビュー投稿

最初の傾向分析として、不自然なレビュー投稿行為について調査した。ここでは一人の利用者が短期間に多数のレビュー投稿を行う行為に着目した。電子メールで、短期間に同一送信元から多数のメール送信が行われた場合、そのメールは迷惑メールである可能性が高い。これと同様に、一人の利用者が、短期間に多数の商品へレビュー投稿を行った場合、そのレビュー投稿には不自然さがある。

レビュー投稿数の多い利用者について、投稿の傾向を見た。図 4 にレビュー投稿数が上位 10 人のレビューアについて、毎月の投稿数を示した。

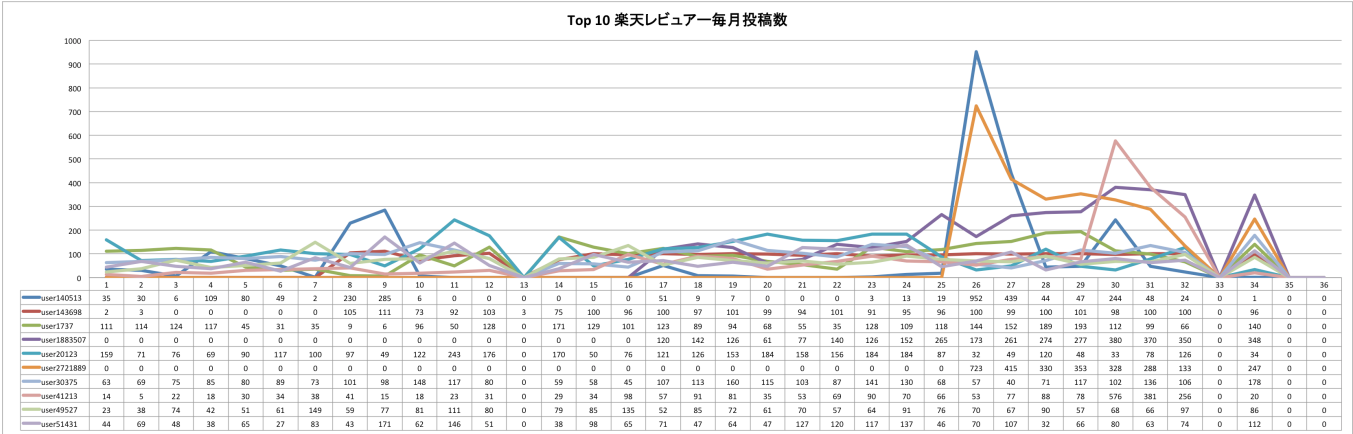


図 4 レビュー投稿数上位 10 人の、各月のレビュー投稿数

全期間でレビュー投稿数が最も多い利用者 user140513 について述べる。図 4 でオレンジ色の線で示すように、利用者 user140513 のレビュー投稿は、2011 年 12 月から 2012 年 3 月の間が最も多い。その期間における各月のレビュー投稿数を表 2 に示す。表 2 にあるように、一ヶ月間に 952 回もレビューを投稿するのは普通ではない。

表 2 user140513 のレビュー投稿数
(2011 年 12 月～2012 年 3 月)

時期	2011 年	2012 年		
	12 月	1 月	2 月	3 月
レビュー投稿数	13	19	925	436

多数投稿されたレビュー文を詳細分析することで、迷惑メール排除と同様に、不自然なレビューを排除することが可能になる。

4. 未購入商品へのレビュー投稿

多くの EC サイトでは、購入していない商品へレビューを投稿できない。一方、本研究が対象としている楽天市場データセットには、商品を購入していない利用者から投稿された商品レビューが含まれている。ここで、商品を購入した利用者からの商品レビューを購入レビュー、商品を購入していない利用者からの商品レビューを未購入レビューと呼ぶことにする。この節では未購入レビューの傾向を述べる。

4.1 調査対象ジャンル

楽天市場データセットにはレビューが 6300 万件以上含まれている。これら全てを調査するのは大変であるため、表 3 に示す三つの商品ジャンルだけを対象とすることにした。

表 3 調査対象ジャンルとレビュー件数

ラベル	ジャンル	レビュー数
Health	ダイエット・健康	2,319,165
Game	おもちゃ・ホビー・ゲーム	1,470,270
PC	パソコン・周辺機器	1,264,808

「ダイエット・健康」は、やらせ・ステマや、有用でないレビューなどが多く想定して選択した。他の「おもちゃ・ホビー・ゲーム」と「パソコン・周辺機器」は、我々が詳しい分野であるため、内容の判断が容易な分野として選択した。以降、表 3 のラベル文字「Health, Game, PC」で、これら三つのジャンルを表記する。なお、全商品のジャンルは楽天市場ディレクトリ ID の Web サイト [5] で公開されている。

4.2 未購入レビューの割合

表 3 に示した三つの商品ジャンルにおけるレビューについて、購入レビューと未購入レビューとの割合を調べた。結果を表 4 に示す。表 4 から、ほとんどのレビューは購入レビューであることが分かる。

表 4 購入レビューと未購入レビューとの割合

ジャンル	Health	Game	PC
レビュー件数	2,319,165	1,470,270	1,264,808
購入	2,312,875	1,465,269	1,261,172
未購入	6,290	4,999	3,636
未購入率	0.27%	0.34%	0.29%

次に、商品への評価点をごとに未購入レビューの割合調べた。楽天市場データセットには、レビュー毎に、レビュー投稿者が与えた商品的评价点が含まれている。評価点は 1～5 点の 6 段階である。5 点が最も良く、1 点が最も悪い評価点である。なお、データセットには評価が

0 点のレビューが 9 件含まれていたけれども、以降の分析には影響しないため省略している。

表 5 評価点毎の未購入レビュー率 (Health)

	5 点	4 点	3 点	2 点	1 点
購入	1,219,624	742,707	258,434	49,370	42,734
未購入	2,973	2,071	750	103	393
割合(%)	0.24	0.28	0.29	0.21	0.91

表 6 評価点毎の未購入レビュー率 (Game)

	5 点	4 点	3 点	2 点	1 点
購入	938,747	374,262	107,394	22,718	22,148
未購入	2,659	1,426	434	69	411
割合(%)	0.28	0.38	0.40	0.30	1.82

表 7 評価点毎の未購入レビュー率 (PC)

	5 点	4 点	3 点	2 点	1 点
購入	733,801	371,233	109,493	19,970	26,674
未購入	1,778	1,090	354	77	337
割合(%)	0.24	0.29	0.32	0.38	1.25

表 5, 6, 7 を見ると、三つのジャンルにおいて、評価点が 1 点の場合に、未購入レビューの割合が最も多い。特に Game で最も高い割合を示している。

5. 未購入レビューの傾向分析

前節で、評価点が 1 点の場合に未購入レビューの割合が最も多いことを述べた。そこで、評価点が 1 点の未購入レビューを分析した。5 点を付けているのに、結局買わない原因も知りたいです。そこで、評価点が 5 点の未購入レビューを分析した。

5.1 評価 1 点未購入レビューの傾向分析

レビュー文に出現する名詞、動詞、形容詞の単語について出現頻度を数えた。単語と出現頻度から、SVM で購入レビューか未購入レビューかを判別する線形カーネルを作成した。SVM カーネル作成に用いたレビューの件数を表 8 に示す。

表 8 SVM カーネル作成用のレビュー件数

ジャンル	Health	Game	PC
未購入レビュー件数	310	334	273
購入レビュー件数	68	29	30

SVM カーネルが判別に用いる単語を特徴語として、名詞、動詞、形容詞を上位からそれぞれ 10 個取得した。

表 9 に評価 1 点未購入レビューの特徴語を、表 10 に評価 1 点購入レビューの特徴語を示す。

表 9 評価 1 点未購入レビューの特徴語

名詞	動詞	形容詞
定価	れる	ない
メール	送る	ひどい
キャンセル	下さる	酷い
連絡	かかわる	悪い
電話	問い合わせる	腹立たしい
最低	受け付ける	おかしい
在庫	待つ	低い
問い合わせ	経つ	無い
返事	売る	高い
最悪	書く	ほしい

表 10 評価 1 点購入レビューの特徴語

名詞	動詞	形容詞
残念	送る	もったいない
汚れ	振込む	遅い
返品	まつ	弱い
最悪	騙す	悪い
お金	閉める	うまい
無駄	いらっしゃる	酷い
認識	返る	難しい
状況	とどく	ほしい
ダメ	経つ	黒い
入荷	見せる	茶色い

表 11 評価 1 点の未購入レビュー例

普通、発送が 1 か月先なら、売り切れか予約って書いてあるはずなのに在庫あるって嘘。発送は 8 月 30 日だからキャンセルメールしたのに全然返事なし。
在庫ありになってるが、実際はないようです。急いで必要でしたが、一週間すぎても届かず、しまいにはメーカーの都合でキャンセルのメールが来ました。だったら在庫有りにしなきゃいいのに…
購入して次の朝には注文のお礼メールが来たのに夕方になって商品がないとの連絡がきました。お金だけはしっかりとられて何が何か意味がわかりません。在庫がないなら他のお店で買ったのに!!! マラソンも 10 店舗買い回りしたのに・・・最悪です。早くお金を返してください!!!

表 9 に示した、評価 1 点の未購入レビューの特徴語だけを関連グラフ[6]として表示したものを図 5 に示す。図 5 を見ると、未購入レビューには、商品の品質や使い勝

手についての苦情ではなく、商品を販売する店への苦情で有ることが読み取れる。

表 11 に、評価 1 点の未購入レビューの例を三つ示す。実際のレビューを読んでも、評価 1 点の未購入レビューには、商品を販売する店への苦情である事が多い。

5.2 評価 5 点未購入レビューの傾向分析

前節で、評価点が 1 点の場合に未購入レビューについて分析結果を述べた。続いて、評価点が 5 点の未購入レビューを分析した。SVM カーネル作成に用いたレビューの件数を表 12 に示す。

表 12 SVM カーネル作成用のレビュー件数

ジャンル	Health	Game	PC
未購入レビュー件数	2,530	2,148	1,223
購入レビュー件数	3,196	2,911	1,966

SVM カーネルが判別に用いる単語を特徴語として、名詞、動詞、形容詞を上位からそれぞれ 10 個取得した。表 13 に評価 5 点未購入レビューの特徴語を、表 14 に評価 5 点購入レビューの特徴語を示す。

表 13 に示した、評価 5 点の未購入レビューの特徴語だけを、関連グラフで表示したものを図 6 に示す。図 6 を見ると、良い評価単語が並んでいるが、未購入レビューについては、傾向を断定出来ない。

表 15 に、評価 5 点の未購入レビューの例を五つ示す。実際のレビューを読んでも、評価 5 点の未購入レビューには、商品はまだ届かない商品への期待が多い。また他人へのプレゼントも多い。

表 13 評価 5 点未購入レビューの特徴語

名詞	動詞	形容詞
楽しみ	探す	待ち遠しい
プレゼント	せる	ほしい
納期	気に入る	すごい
コメント	決める	早い
お客様	頂く	嬉しい
最高	いただく	欲しい
丁寧	勧める	うれしい
おすすめ	見つける	可愛い
5つ	遊ぶ	かわいい
さん	楽しむ	楽しい

表 14 評価 5 点購入レビューの特徴語

名詞	動詞	形容詞
満足	喜ぶ	安い
子供	くれる	良い
綺麗	助かる	美味しい
プール	入る	可愛い
きれい	もらえる	大きい
水晶	買える	かわいい
石	頼む	うれしい
浄化	遊ぶ	小さい
プレスレット	気に入る	嬉しい
さざれ	あげる	欲しい

表 15 評価 5 点の未購入レビュー例

期待を込めて、★5で。息子のクリスマスプレゼントに購入しました。
まだ届いてませんが、金額の安さに星 5 つで`^`v 届くのを楽しみにしています。
注文しました。まだ先ですが届くのが楽しみです。マシン本体の改造が楽しみです。
まだ届いてないけど楽しみ！早く届くのが待ち遠しいです_o`^`o
早く届くのが楽しみです。息子も楽しみにしています。

5.3 考察

世の中には多くの小売店やサービス業がある。多くの商店が、顧客が「買わない理由」を切実に知りたいと考えている。買わない理由の把握は困難であるため、利用者アンケートによる調査を行うことが多いようである。

本研究では、楽天市場データセットの未購入レビューの傾向を分析することで、楽天市場の EC サイトにおける「買わない理由」を一部説明することができた。

6. おわりに

我々はネット上の EC サイトにおける商品レビューからの知識発見を目指している。本論文では、最初に楽天市場で販売される商品についてのレビュー文を対象に、レビュー投稿行為の不自然を調査した。一ヶ月に千件に近い商品レビューを投稿する利用者があることをみつ得た。次に、購入していない商品に対するレビュー文を調べた。評価点ごとに未購入レビューの投稿数を調べ、評価が 1 点の場合に、最も未購入レビューが多いことが分かった。また、評価 1 点のレビュー集合について、購入レビューと未購入レビューを調べたところ、未購入レビューには、商品への不満ではなく、販売店の対応の悪さに不満がある文章が多いことが分かった。

今後は、SVM 等の機械学習手法を用いて、未購入レビューと購入レビューを判別する手法を検討したい。また、やらせや誹謗レビュー、多重投稿レビュー、有用な情報が少ないレビューなどについて、より詳細に調査する予定である。

参考文献

1) J-Cast ニュース, 「やらせ」のロコミや架空の注文なんと11 万件超! 「楽天市場」で大掛かりなステマが発覚,
<http://www.j-cast.com/2015/03/21231013.html>
(accessed at Oct.8.2015).

2) NII, 情報学研究リポジトリ,
<http://www.nii.ac.jp/dsc/idr/> (accessed at Oct.8.2015).
3) NII, 楽天株式会社, 楽天データセット,
<http://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html>
4) 形態素解析エンジン MeCab,
<http://taku910.github.io/mecab/>(accessed at Oct.8.2015).
5) 楽天市場, 楽天ディレクトリ ID,
<https://www.rakuten.co.jp/com/inc/rms/navi/directory.html>,
(accessed at Nov. 19, 2015).
6) Brendan Flanagan, Yohei Inokuchi, Chengjiu Yin, Sachio Hirokawa, Using Automatically Generated Mind Maps to Promote Initial Communication, Proc. Of AECT International Conference on the Frontier in e-Learning Research 201, pp.332-333, 2013.

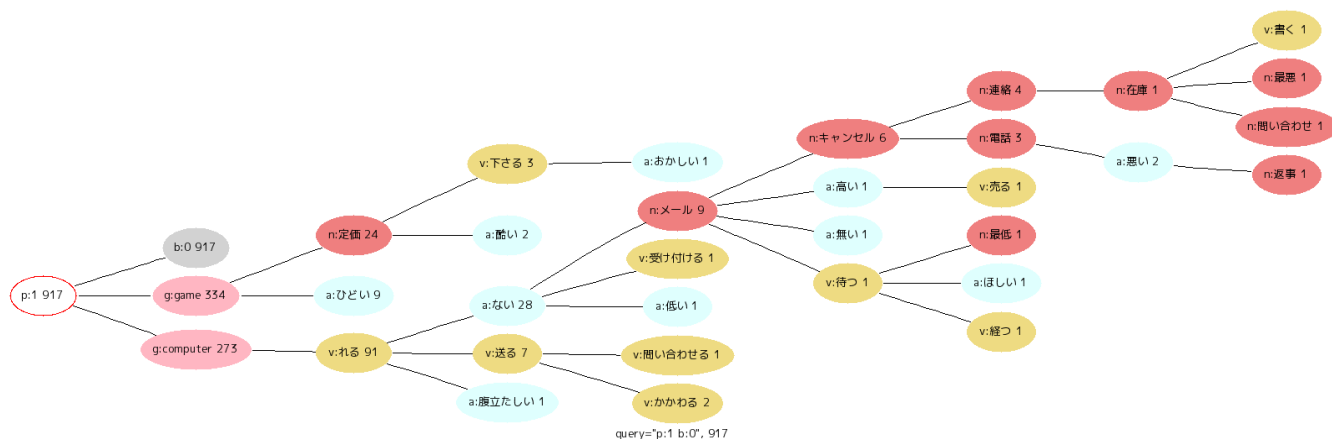


図 5 評価 1 点未購入レビューにおける特徴語の関連グラフ

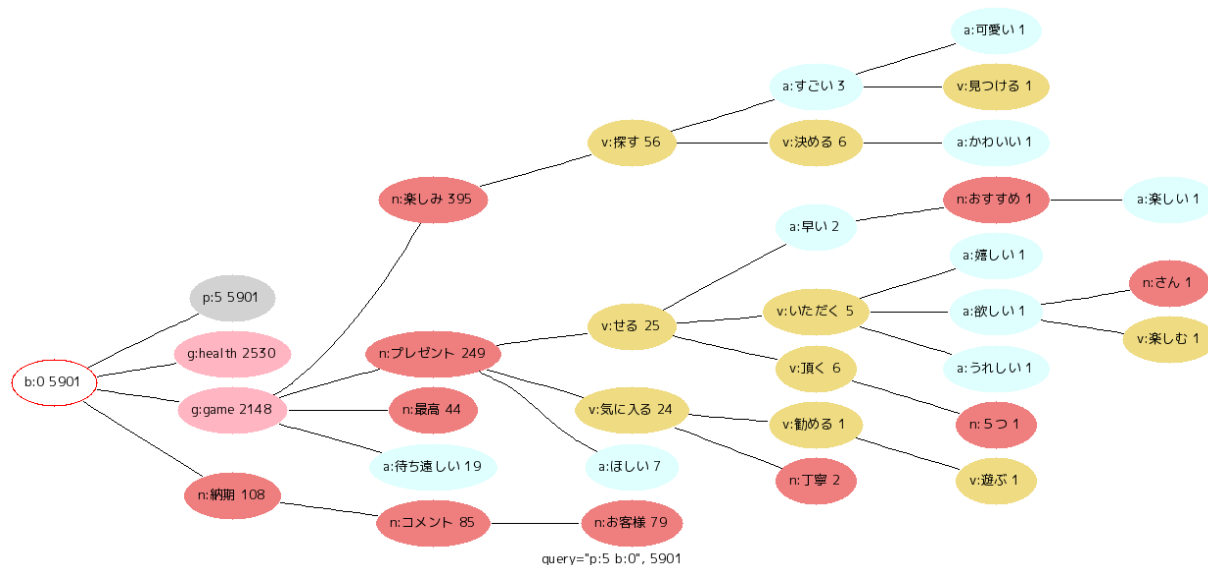


図 6 評価 5 点未購入レビューにおける特徴語の関連グラフ