協調フィルタリングを用いたクチコミ文書の極性判定

坂村　華怜† 　三浦　孝夫†

† 法政大学理工部創生学科　〒 184-0002 東京都小金井市梶野町 3-7-2
E-mail: †karen.sakamura.6t@stu.hosei.ac.jp, ††miurat@hosei.ac.jp

あらまし　本稿では化粧品クチコミ文書サイトのクチコミ文書に対し協調フィルタリングを用いて未出現の単語の値を推定し極性判定を行う．具体的にはクチコミ文書に対して形態素解析を行い, 拡張した極性辞書内の単語の重みにより, クチコミ文書のベクトル化を行う．協調フィルタリングを用いてそれらの値から未出現の単語の重みを推定し, テストデータと類似度の高いクチコミ文書の極性をもとに極性判定を行う．以上を協調フィルタリング前と比較し提案手法の有効性を検証する．

キーワード　協調フィルタリング, 極性判定

1. 前　書　き

インターネットの普及によりソーシャルネットワーク (SNS) が急速に普及し，利用者は簡単に意見を表明することができる．このような個人的な情報は多様で大量に生じ，これらを素早く解析することで新たなビジネスチャンスの場を生み，また市場動向の分析等に寄与する場を生む [1].

本研究では，SNS データのうち，とくに商品の評判情報 (”レビュー”情報，”クチコミ文書”情報) を対象として，未出現の単語の出現頻度などを推定し，その極性分析手法を提案する．クチコミ文書情報は，商品ブランドごとに存在し，”ふてぶてしい” などの特徴的な表現で自らの意見を象徴的に記述し，極端な場合，商品イメージを固定してしまうことある．

SNS データでは，短い文章が多く，単語の出現数が少なく，未出現語も多い．このため，出現分布に一般性がなく偏りが大きいため分析が容易でないことから，単純に（支持/不支持など）極性判定を行うことが多い．しかし，極性辞書に出現する単語をはじめ，特徴的な未出現の単語を推定することができれば，短い文章でも極性判定に必要な情報量を得ることができる．マスカラの”フィルム” (お湯で落とせる型の名称) のように，特徴語・特徴表現は商品に依存して対応することが多く，極性判定（この例では否定）に有効な辞書を体系的・自動的に構築することは容易ではない．

本研究では，極性辞書に新たにカテゴリに依存した単語を追加し，協調フィルタリングを用いて未出現の極性語や特徴語の出現頻度を推定し，短い文章で極性を判定を可能にする手法を提案する．

2. クチコミと極性判定

商品の評判情報 (”レビュー”情報，”クチコミ文書”情報) は，商品ブランドごとに存在し，利用者は自らの使用状況をもとに個人意見を感想として要約する．この様な情報は，多数の汎用を生む場合，支持あるいは不支持の表明を伴って集約され，商品イメージを判定する要因となる．このような極性 (支持・不支持, 好評価・悪評価などの正負判定）評価による評判情報の要約で，商品傾向を素早く判断できることになる．

通常，評価を特徴づけるものは，単語またはその組合わせ (語句) であり，名詞・動詞・形容詞・副詞の自立語であると考えることができる．”美しい” などのように予め定まった評価を持つ者は辞書として集約でき，これを極性辞書という．

極性語が文章に含まれている割合により文書の極性を調べることができる．実際，文章に含まれるポジティブ語，ネガティブ語の総数が一定量の差を有すれば，当該極性を仮定してよいであろう．例えば乾，岡崎研究室の極性辞書 [5] には

（ポジティブ）あこがれる, あじわう, かなう, したう, すがすがしい

（ネガティブ）あがく, あきらめる, あきる, あきれる, あせる

などが存在し辞書内には 5,279 個（“ポジティブ”が 2,107 個，“ネガティブ”3,172 個）登録されている．評価極性を持つ（複合）名詞，約 8,500 表現に対して，評価極性情報を人手によって付与したデータである．

本研究では，コロケーションについては考慮せず，極性辞書のうち，単語であるものを利用する．極性が反転することを懸念し，付属語“ない”はそれを含め一単語とする．この場合，極性語は“ポジティブ”が 547 語，“ネガティブ”が 758 語となる．

我々は，極性辞書とクチコミ文書から同時関係抽出を用いてカテゴリに依存した極性語抽出を提案している [1]．この抽出には，ある集合の中の要素間の関連性について分析を行う．頻出するアイテムの組み合わせの規則を漏れなく抽出し，その中から興味深い結果を探し出す．この結果，カテゴリに依存した特徴的な単語も辞書に追加する．”フィルム”, ”ロング”, ”カール”, ”セパレート”などの特徴語が新たに辞書に追加される．

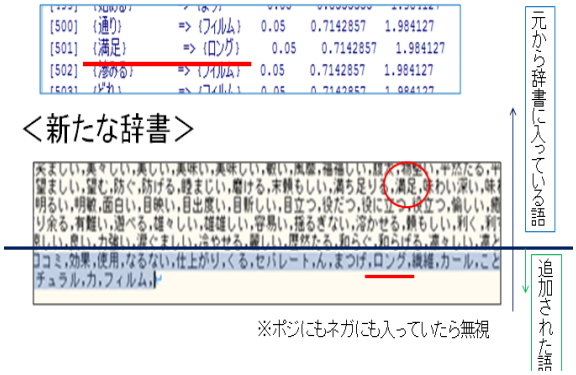

図 1　R を用いた同時関係抽出による辞書拡大

3.　辞書を用いた協調フィルタリング

本章では，拡大した極性辞書 [5] に基づいて協調フィルタリングを利用し単語の出現頻度を推定する手法を提案する．協調フィルタリングでは，多くのユーザの嗜好情報を蓄積し，あるユーザと嗜好の類似した他のユーザの情報を用いて自動的に推論を行うもので，趣味の似た人からの意見を参考にするというクチコミ文書の原理に基づいている．実装にはユーザ同士の類似度を，同じアイテムにつけた評価の相関係数など類推に利用する．協調フィルタリングにはコサイン類似度を利用する．

本研究では協調フィルタリングで定義される横軸のアイテムを極性語と見なし，上記で説明したユーザにあたる縦軸をクチコミ文書とする．値は各クチコミ文書に出現する極性語の値である．極性語の出現具合から類似した未出現の単語の出現頻度を推定する．クチコミ文書 d と極性語の出現頻度 w j によって d= w1,..,wn で表わす．クチコミ文書 d の集合 D を学習データとしクチコミ文書とそのほかのクチコミ文書とのコサイン類似度を求め，重み平均をそのクチコミ文書の未出現の極性語の推定値とする．この操作により短い文書のように単語の出現頻度が少ない場合でも極性判定に十分な単語量が期待される．　表 1 に協調フィルタリング動作の例を示す．横軸に示した極性語が縦軸のクチコミ文書に生じ，出現頻度をセル内の数値で表している．このとき，文書 1 の単語「しまう」の出現頻度を推定したい．現時点で分かっている値を基に，クチコミ 1 を基準として各クチコミ文書とのコサイン類似度を求め，sim として右側に表記する．sim の値を基に図 2 に示した式からコサイン類似度によって文書 1 の単語「しまう」の出現頻度を推定する．

表 1　辞書を用いた協調フィルタリング

					極	性	語					
		1	2	3	4	5	6	7	8	9	10	
		良い	力	くる	いい	しまう	出る	こちら	仕上がり	伸びる	目元	sim(1.m)
口	1	1		3						5		1
コ	2			5	4			4			2	-0.18
ミ	3	2	4		1	2		3		4	3	0.41
文	4		2	4		5			4			-0.1
書	5			4	3	4	2					-0.31
	6	1		3		3			2			0.59

北山ら [2] は，賛否が分かれた属性は商品利用者の印象に強く残った属性である，と定義し商品を構成する属性への肯定，

(式)　$$r_{i,x} = \frac{\sum_{j \in N(i;x)} S_{ij} * r_{ix}}{\sum S_{ij}}$$

$$r_{1,5} = (0.41 * 2 + 0.59 * 3) / (0.41 + 0.59) = 2.6$$

図 2

否定評価表現を含む文をレビューから抽出し，評価の対象となった各属性を商品の特徴語に設定して，商品の特徴ベクトルとして他の商品の特徴ベクトルとのコサイン類似度を計算することで類似する商品をユーザに推薦する手法を提案している．評価表現文を抽出する際に乾，岡崎らの極性辞書 [4] の主観的な評価の部分を利用している点，レビュー文から特徴語を抽出し，商品を特徴語でベクトル化している点で関連があるが，本研究では単語抽出の際に名詞以外にも動詞と形容詞なども抽出している点，形態素間距離を考慮していない点などで本研究と異なっている．

松波らはレビュー文から評価表現を抽出し，キーワードの共起に基づく評価表現辞書を構築し，”うるおい効果”や”美白効果”などの項目を評価項目として設定した上で，各評価項目の評価値を算出 [6]，奥田らは，松波らの辞書 [6] をもとに，真に有用なレビューを推薦するためには，価値観を共有しうるユーザーによるレビューを発見することが重要だと考え，[6] で得点づけした評価項目別スコアを用いて類似ユーザーを判定する手法 [7] を提案している．アットコスメという化粧品レビューサイトのクチコミレビューをデータとして扱っている点では似ているが本研究では乾，岡崎らの極性辞書 [4] をクチコミレビューと共に同時関係抽出を用いてカテゴリに依存した拡大辞書 [8] を基に極性語を抽出し，単語の分布から類似したクチコミレビューを検出し極性判定を行っている点で異なっている．

4.　実　　験

4.1　実験準備

本研究では化粧品の“クチコミ文書”サイトである，“アットコスメ”を利用する．データの収集期間は 2016/12/9〜12/15 であり，カテゴリは“マスカラ”，“最新クチコミ文書ランキング”(図 3) の 1 位を対象として，“クチコミ文書”と“レビュー”情報であるおすすめ度 (図 4) を，多く人が参考になると判断した場合にポイントが増される“LIKE”という値の高い順に各商品 105 件収集する．

形態素解析ソフト，“Mecab”を利用し，文書の形態素解析を行う．クチコミデータ例を図 5 に示す．この中から名詞，動詞，形容詞を抽出し，極性が反転することを考慮し，付属語“ない”まで含めて一語として取得する．

H クチコミ文書 90 件を学習データとして文書解析ツール“R”を用いて辞書を用いた協調フィルタリングを行い，未出現の単語の出現頻度を推定する．クチコミ文書の極性語集合の要素はその極性語の tf，idf，tfidf，deltatfidf の 4 種類で行う．tf はその文書内のその極性語の頻度を正規化しタ値であり，idf はその極性語が出現する文書数の逆数を対数化したもので，tf-idf

図 3　最新クチコミ文書ランキングの選択

図 4　実際のクチコミ文書例

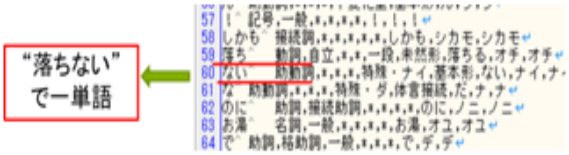

図 5　形態素解析後のデータ

は tf と idf をかけあわせたものである．

deltatfidf は図 6 と図 7 の数式で表される，実際に出現する極性から影響された単語出現頻度をもとにした値である．[3]

$C_{t,d}$　：d の文書内で t の単語の出現頻度
P_t　：単語 t が出現していてポジティブに分類された文書数
N_t　：単語 t が出現していてネガティブに分類された文書数
$|P|$:ポジティブな文書数
$|N|$:ネガティブな文書数
$V_{t,d}$:文書 d 内の単語 t の特徴量

図 6

$$V_{t,d} = C_{t,d} * \log_2(\frac{|P|}{P_t}) - C_{t,d} * \log_2(\frac{|N|}{N_t})$$
$$= C_{t,d} * \log_2(\frac{|P|}{P_t}\frac{N_t}{|N|}) = C_{t,d} * \log_2(\frac{N_t}{P_t})$$

図 7

良い悪いを判断する基準として“おすすめ度”を用いる．各商品には“レビュー”情報の平均である，“おすすめ度平均”（図 8）が存在し，クチコミ文書 115 件の各“おすすめ度”と“おすすめ度平均”を比較する．“おすすめ度平均”の小数点以下は無視し，“おすすめ度平均”＜“おすすめ度”の場合，クチコミ文書は商品に対して良い評価，“おすすめ度平均”＞“おすすめ度”の場合，クチコミ文書は，商品に対して悪い評価と定義する．

本研究では再現率，適合率及び F 値を用いる．再現率は実際に正しいデータ中で正しいと予測した数，適合率は正しいと予測したなかで，実際に正である数と定義する．クチコミ文書データ 90 件を学習データとして協調フィルタリングを行うとき，ベクトルのバランスを整えたあと，テストデータ 15 件との類似度を計算し，閾値を 0.4 としてテストデータと最も類似しているクチコミ文書を学習データ 90 件の中から抽出する．抽出したクチコミ文書の評価と同じ評価をテストデータもしていると定義する．

協調フィルタリング前も同様に“クチコミ文書”の極性分析を行い，協調フィルタリング前と後で再現率，適合率及び F 値を基に比較する．F 値とは，再現率 a と適合率 b の調和平均 $1/\{(1/a + 1/b)/2\} = 2ab/(a + b)$ をいう．

図 8　データ収集のサイト例

4.2　実 験 結 果

クチコミ文書のベクトル内で値を持たない“NA”の協調フィルタリング後の数は全 7,830 件中 7,830 件が推定できる．

表 2　NA の変化

	フィルタリング後	フィルタリング前
	NA の数	NA の数
tf	6656/7830	7830/7830
idf	6656/7830	7830/7830
tf-idf	6656/7830	7830/7830
deltatf-idf	6820/7830	7830/7830

協調フィルタリングによって未出現の単語の値がうまった tf の例を示す．クチコミ文書 1〜10 それぞれ値を持たなかった NA が値を持ったことが分かる．

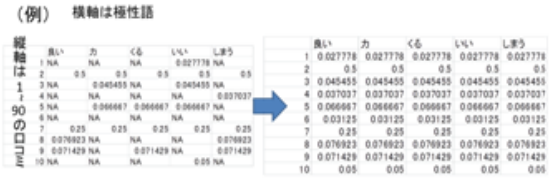

図 9

極性判定結果を再現率を用いて示す．ｔｆに関してはすべての極性語の値が同じになってしまったため，判定不能．類似度の平均，分散は idf で，0.34，0.0059，tf-idf で 0.37，0.0082，deltatf-idf で 0.54，0，024 である．特に tf-idf では 32%も再現率が向上する．

表 3　再　現　率

	協調フィルタリング前	協調フィルタリング後
ｔ F	4/15(27 %)	各クチコミ文書の単語の頻度が同値になり判定不可
idf	5/7(71 %)	4/4(100 %)
tf-idf	5/11(8 %)	2/5(40 %)
deltatf-idf	6/15(25 %)	3/12(25 %)

極性判定結果を適合率を用いて示す．ｔｆに関してはすべての極性語の値が同じになってしまったため，判定不能．特に idf では 8%適合率が向上する．deltatf-idf では 1%適合率が下がる．

表 4　適　合　率

	協調フィルタリング前	協調フィルタリング後
ｔ F	56/176(32%)	各クチコミ文書の単語の頻度が同値になり判定不可
idf	69/195(35%)	32/74(43.2%)
tf-idf	12/38(31%)	103/218(32%)
deltatf-idf	123/348(35%)	59/172(34%)

極性判定結果を F 値を用いて示す．ｔｆに関してはすべての極性語の値が同じになってしまったため，判定不能．ここでは deltatf-idf では適合率が下がるが，idf,deltatf-idf では値が向上する．

表 5　F　　値

	協調フィルタリング前	協調フィルタリング後
ｔ F	29%	各クチコミ文書の単語の頻度が同値になり判定不可
idf	47%	60%
tf-idf	13%	36%
deltatf-idf	29%	29%

それぞれ協調フィルタリングが有効に作用している．

4.3　考　　察

協調フィルタリングにより単語全ての単語の出現頻度が推定できる．tf の場合は頻度がすべて 1 のために各クチコミ文書ごとにすべての単語が同じ値になる．ｔｆによって未出現語でも idf 値を用いれば，極性判定に利用可能となる．協調フィルタリング前と比較して閾値以上の類似度の高いクチコミ件数は減ったものもあるが，適合率，再現率，およびＦ値も精度があがる．Idf，tf-idf は協調フィルタリング後，反対の極性を取ることはなく Idf では協調フィルタリング前は 2 件あったものが，協調フィルタリング後は 0 件になり，Tfidf では協調フィルタリング前 1 件あったものが，協調フィルタリング後は 0 極性のもの 1 件に減る．協調フィルタリング前は，単純に出現単語が含まれるかどうかによって大きく類似度が左右されていたが，協調フィルタリング後はすべての単語が値を持ち類似度の算出に影響を与えるため，より精度が高く類似したクチコミ文書の推定ができる．この idf の二つの例を以下に示す．

表 6　協調フィルタリング前ｉｄｆで行ったテストデータ 15 件の判定結果

おすすめ度	テストデータ	類似度最大のクチコミ No.	類似度	おすすめ度	正解
6	1	4	0.343448241	4	0
5	2	1	0.321707584	4	0
7	3	3	0.423645258	7	1
7	4	4	0.21669717	4	0
5	5	4	0.390944351	4	0
6	6	6	0.349603764	6	1
6	7	1	0.49118757	4	0
5	8	1	0.272271696	4	0
7	9	14	0.433455106	7	1
6	10	1	0.500267147	4	0
3	11	4	0.479292436	4	1
4	12	1	0.465132727	4	1
1	13	1	0.274618918	4	1
4	14	1	0.269586556	4	1
2	15	1	0.403983653	4	1

表 7　協調フィルタリング後ｉｄｆで行ったテストデータ 15 件の判定結果

おすすめ度	テストデータ	類似度最大のクチコミ No.	正解数	類似度	正解
6	1	2	6	0.31573268	1
5	2	2	6	0.227429482	0
7	3	2	6	0.308652528	1
7	4	2	6	0.2756977	1
5	5	2	6	0.349189483	0
6	6	2	6	0.460984625	1
6	7	2	6	0.480650308	1
5	8	2	6	0.295292988	0
7	9	15	6	0.436117321	1
6	10	2	6	0.328230733	1
3	11	13	4	0.441551084	1
4	12	2	6	0.390683892	0
1	13	2	6	0.265584507	0
4	14	2	6	0.260081284	0
2	15	2	6	0.331338719	0

テストデータのクチコミ 7 と 10 が協調フィルタリング前は誤った極性判定によって 1 を表示しているが，協調フィルタリング後は正しい極性判定，”2”，”類似性のあるクチコミなし”の判定ができていることが分かる．協調フィルタリング後の単語の頻度の変化 (表 8) と，実際にクチコミ文書に出現している単語（表 9）を次のページに示す．表 8 では，”美しい”，”

表 8　クチコミ 1，2 の協調フィルタリング前と後の単語の値の変化

	クチコミNO.	極性語																					
	+	良い	力	くる	いい	しまう	出る	こちら	仕上がり	伸びる	目元	もの	度	感じる	塗り	サンプル	後	気に入る	満足	出来る	嬉しい	勧める	美しい
協調フィルタリング前	1	1.327902142	1.356547324	1.431798276	0	1.455931956	1.468521083	1.468521083	1.508638306	0	1.568636236	1.568636236	1.585026652	1.619788758	1.638272164	0	1.698970004	0	0	0	0	2.045757491	0
	2	0	0	0	0	0	0	0	0	1.552841969	0	0	0	0	0	0	0	0	1.886056648	0	0	0	0
協調フィルタリング後	1	1.327902142	1.356547324	1.431798276	1.646710656	1.455931956	1.468521083	1.468521083	1.508638306	1.676602979	1.568636236	1.568636236	1.585026652	1.619788758	1.638272164	1.696342795	1.698970004	1.740122104	1.731418189	1.744029867	1.739242993	2.045757491	1.735309133
	2	1.635574276	1.649173253	1.692599337	1.694191748	1.701545777	1.706331528	1.690653674	1.71308995	1.552841969	1.7200489	1.710170511	1.718091051	1.728553639	1.72250889	1.7300403	1.739638392	1.742853337	1.886056648	1.768487643	1.765502517	1.748686726	1.724964983
	+	合う	柔らかい	素晴らしい	程よい	保つ	ぴったり	イイ	うまい	助かる	慣れる	目立つ	出来上がる	楽しむ	頑張る	馴染む	面白い	明るい	手早い	好む	落ち着く	上手い	信じる
協調フィルタリング前	1	0	0	2.397940009	0	0	2.522878745	0	2.522878745	0	0	0	0	0	0	0	0	0	0	0	2.698970004	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
協調フィルタリング後	1	1.725400816	1.76046526	2.397940009	1.730704557	1.752847681	2.522878745	1.692317392	2.522878745	1.692317392	1.754123398	1.692317392	1.692317392	1.722602607	1.767924332	1.721674753	2.692317392	1.747333505	1.722361372	1.769717135	2.698970004	1.692317392	1.692317392
	2	1.723676915	1.727982729	1.753302265	1.719449309	1.719449309	1.719449309	1.794185522	1.719449309	1.719449309	1.719449309	1.719449309	1.735195981	1.728184357	1.719449309	1.719449309	3.719449309	1.719449309	1.750624673	1.719449309	1.719449309	1.726672803	1.719449309
	-	カール	塗る	方	よう	まつげ	タイプ	いい	しまう	こちら	時	気	度	言う	何	求める	塗り	フィルム	悪い	物足りない	悩む	汚い	遅い
協調フィルタリング前	1	1.259637311	1.301029996	1.318758763	1.327902142	1.387216143	1.408935393	0	1.455931956	1.468521083	1.48148606	0	1.585026652	1.619788758	1.619788758	1.638272164	1.638272164	1.657577319	0	2.045757491	0	2.301029996	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
協調フィルタリング後	1	1.259637311	1.301029996	1.318758763	1.327902142	1.387216143	1.408935393	1.646710656	1.455931956	1.468521083	1.48148606	1.652251766	1.585026652	1.619788758	1.619788758	1.638272164	1.638272164	1.657577319	1.773529409	2.045757491	1.725210818	2.301029996	1.717784334
	2	1.561557055	1.624030877	1.613769649	1.633109692	1.696301279	1.69016398	1.694191748	1.701545777	1.690653674	1.705253683	1.704358259	1.718091051	1.719568702	1.728161168	1.735662428	1.72250889	1.7214954	1.722766566	1.726623976	1.724737243	1.726762237	1.73876986
	-	迷う	ゆるい	忙しい	痛い	汚れる	乏しい	損なう	崩れる	欠く	だれる	忘れる	慌てる	面倒臭い	しつこい	もったいない	欠ける	あきらめる	間違う	飽きる	困る	ばっちい	
協調フィルタリング前	1	0	0	0	0	2.522878745	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
協調フィルタリング後	1	1.737482564	1.692317392	1.742459661	1.75533782	2.522878745	1.731948274	1.692317392	1.754548124	1.692317392	2.728966333	1.692317392	1.777244992	1.747623099	1.765405553	1.692317392	1.692317392	1.768716178	1.692317392	1.718152731	1.729129362	1.729129362	
	2	1.766357437	1.754922218	1.726658984	1.719449309	1.746389776	1.719449309	1.719449309	1.728184357	1.749326547	3.731773926	1.719449309	1.760241017	1.726672803	1.719449309	1.719449309	1.763820145	1.731773926	1.719449309	1.728214861	1.719449309	1.719449309	

表 9　実際にクチコミ文書に出現している単語

idf	テストデータ	協調フィルタリング前		協調フィルタリング後	idf	テストデータ	協調フィルタリング前		
極性	p	n		p	極性	p	n		
クチコミ番号	7	1	→	2	クチコミ番号	10	1	→	×
+	出来る,1	こちら,1		満足,1		しまう,1	こちら,1		
	こちら,1	感じる,1		伸びる,1		ぴったり,1	感じる,1		
	感じる,1	しまう,1				力,1	しまう,1		
	珍しい,1	度,1					度,1		
	しまう,1	素晴らしい,1					素晴らしい,1		
	度,1	もの,1					もの,1		
	後,1	勧める,1					勧める,1		
	仕上がり,1	後,1					後,1		
	合う,1	落ち着く,1					落ち着く,1		
	くる,1	仕上がり,1					仕上がり,1		
	力,1	うまい,1					うまい,1		
	出る,1	ぴったり,1					ぴったり,1		
	伸びる,1	塗り,1					塗り,1		
		くる,1					くる,1		
		良い,1					良い,1		
		目元,1					目元,1		
		力,1					力,1		
		出る,1					出る,1		
-	こちら,1	こちら,1				まつげ,1	こちら,1		
	求める,1	時,1				求める,1	時,1		
	時,1	嘆く,1				しまう,1	嘆く,1		
	しまう,1	しまう,1				何,1	しまう,1		
	度,1	汚れる,1				カール,1	汚れる,1		
	カール,1	度,1				方,1	度,1		
	気,1	カール,1				フィルム,1	カール,1		
	方,1	汚い,1				タイプ,1	汚い,1		
	悩む,1	フィルム,1					フィルム,1		
	塗る,1	タイプ,1					タイプ,1		
	タイプ,1	求める,1					求める,1		
		まつげ,1					まつげ,1		
		何,1					何,1		
		言う,1					言う,1		
		よう,1					よう,1		
		塗り,1					塗り,1		
		方,1					方,1		
		塗る,1					塗る,1		
		物足りない,1					物足りない,1		

だれる ” など，136 語の単語が値を持ったことが分かる．

このように協調フィルタリング前は同じ単語が出現しているだけで類似度が高くなるが，協調フィルタリング後は他にも様々な単語が影響していることが分かる．

5. 結　　論

協調フィルタリングを利用して出現していない極性語の推定可能（NA の減少）である．協調フィルタリングを利用して，利用する前に比べ全体の再現率，適合率，F 値が上がる．再現率は最高で idf の時約 29%，平均で 20%上がる．適合率は最高で idf の時約 8.2%，平均で 2.7%上がる．極性語の出現頻度の少ないもの出現していない文書でも判定可能である．

文　　献

[1] Han, Jiawei, Jian Pei, and Micheline Kamber. Data mining: concepts and techniques. Elsevier, 2011.
[2] 吉田, 北山：商品レビューの極性分析に基づく特徴語抽出手法の評価, 電子情報通信学会信学技法 DE2016-5, 2016
[3] Justin Martineau, Tim Finin: Delta TFIDF: An Improved Feature Space for Sentiment Analysis
[4] 乾, 岡崎：日本語評価極性辞書-東北大学 乾・岡崎研究室
[5] Xujuan Zhou,Xiaohui Tao,Md Mostafijur Rahman,Ji Zhang：Coupling Topic Modeling in Opinion Mining for Social Media Analysis
[6] 松波友稀, 上田真由美, 中島伸介：コスメアイテムに対する評価項目別レビュー自動スコアリング方式の開発,DEIM Forum 2017 B5-3
[7] 奥田麻美, 松波友稀, 上田真由美, 中島伸介：コスメレビュー共有システムのための類似ユーザ判定手法,DEIM Forum 2017 P8-5
[8] 坂村, 白井, 三浦: データマイニングを用いたクチコミの極性分析,2017 年電子情報通信学会総合大会学生ポスターセッション,2017