

飲食店口コミサイトを対象にした評価文抽出

川又亜弓^{†1} 立澤裕樹^{†2} 岡誠^{†3} 森博彦^{†1}

飲食店口コミサイトに掲載されている口コミは、利用する飲食店を選択する際に有効活用されているが、飲食店に対する口コミであるために店舗選択の際に重要となる料理の評価が埋もれてしまっている。そこで本稿では、料理別に掲載されている既存の口コミをコーパスとして、飲食店に対する口コミから料理に対する口コミ（評価文）を抽出する手法を提案する。料理の口コミには低頻度の特徴語が多いことから、頻度情報を抑えた tf-idf 値を用いることで料理の口コミに適用した。結果として F 値 67%の精度で評価文を抽出することができた。

Extracting sentences about dish reviews on restaurant review sites

AYUMI KAWAMATA^{†1} YUKI TACHIZAWA^{†2} MAKOTO OKA^{†3}
HIROHIKO MORI^{†1}

Customer reviews on restaurant review sites are useful in selecting a restaurant. In case a user want to enjoy dishes, he/she tries to find only review sentences about the dishes among all sentences which contains various sorts of viewpoints such as the atmosphere, the motivation and so on. In this paper, we propose a method of extracting review sentences about dishes. As each dish-specific word tends to appear in low frequency, we modified the tf-idf measure to reduce the influence of the frequency. As the results, F-measure of extraction was 67.

1. はじめに

近年、様々な対象の口コミを集めたサイトが増え、商品を利用したユーザによる商品情報や感想などの情報収集が容易になっている。例えば、多ジャンルの製品に関する口コミを集めた価格.com¹や、飲食店に対する口コミを集めた食べログ²などが挙げられ、これらは実際に利用する製品・飲食店舗を選択する際の判断材料として有効活用されている。特に飲食店口コミサイトには、飲食店に対する口コミが投稿されるが、その中には店舗を訪れた経緯や店舗の雰囲気、次回何を食べたいかなど、店舗選択の判断材料としては比較的必要な情報が含まれるため、閲覧しているユーザが最も必要としている料理に関する口コミが埋もれてしまっている。

2. 関連研究

評価要素の辞書を作成する研究として小林ら[1]は車など製品の評価を抽出するため、評価を<対象, 属性, 評価値>という3要素で定義し、3要素の共起パターンや頻度情報から抽出した表現を、辞書登録を繰り返し活用することで属性・評価値表現を抽出した。上記の半自動収集手法は人手で収集するよりも効率的に表現を収集できることを明らかにした。

製品の 口コミからの情報収集を容易にする目的で口コミ

ミの分類や文章抽出を行う研究は盛んに行われている。立澤[2]は、自由形式で書かれている製品口コミを評価視点別に分類する研究を行った。分類には評価視点ごとに、分類の手がかりとなる知識が必要となるが、それらを作成することは容易ではない。そこであらかじめ評価視点別に記述されている口コミをコーパスとすることで分類を容易にした。

飲食店の口コミを対象にした研究として、矢野ら[3]は Web 上にある飲食店の口コミから店舗の評判情報を自動抽出し、個人的嗜好に沿って順位づけする手法を提案した。評価情報の抽出部では、対象となる料理を評価する際には用いられない表現を集めた負例辞書や属性と評価値の頻出パターンを用いて、評価対象や評価値の抽出を行った。

3. 研究目的

本研究では、飲食店に対する口コミから料理に対する評価情報が含まれる文（以下、評価文）を料理別に自動抽出することを目的とする。また従来研究では製品の口コミを対象とした研究は多いが、飲食の口コミに関する研究は少ない。両ドメインにおける口コミの特徴を考慮した上で、飲食の口コミに適した辞書作成手法を探る。

4. 評価文の要素

小林らは製品の 口コミに対し<対象, 属性, 評価値>の3要素で定義したが、飲食の口コミに関しては評価文に評価対象が明記されない場合が多い。評価文例を以下に示す。

例1 「非常に辛いけど美味しいカレーです。」

例2 「皮はパリパリ、具はジューシー！」

^{†1} 東京都市大学院工学研究科
Graduate School of Engineering, Tokyo City University
^{†2} 株式会社プラスアルファ・コンサルティング
Plus Alpha Consulting
^{†3} 東京都市大学知識工学部
Faculty of Knowledge Engineering, Tokyo City University
1 <http://kakaku.com/>
2 <http://tabelog.com/>

例1の評価対象は「チキンカレー」であるが、評価文には「カレー」と明記している。また例2の評価対象は「餃子」であるが、評価文に全く記述されていない。このように評価文には評価の対象が記述されない、または対象の略称を記述する場合が多い。よって本研究では、料理名である評価対象に対する評価文は<属性, 評価値>で構成されると定義する。属性は評価対象の部分的要素や略称など、評価値の主語となり得る名詞を指し、評価値は評価対象や属性を食べた後の記述者の心的状況を表す全ての表現を指す。例1, 例2の属性, 評価値を以下に示す。

- 例1 属性：カレー
 評価値：辛い, 美味しい
- 例2 属性：皮, 具
 評価値：パリパリ, ジューシー

あらかじめ各要素の辞書を作成し、それらを手掛かりに評価文抽出を行う。

5. システム概要

5.1 要素辞書作成手法

評価文の要素となり得る表現は、「カレー」や「ラーメン」などの飲食店舗ジャンル内で特徴的な表現であると考えられる。そこでジャンル別に特徴語を抽出し、それらを各ジャンルにおける評価文の要素とすることで要素辞書の作成を行う。

辞書作成のコーパスは、ぐるなび³のロコミを用いる。ぐるなびはジャンル別に店舗を分類しており、料理別にロコミが掲載されているため、これらから特徴語を抽出することで評価文の要素となり得る表現を収集できると考えた。ぐるなびのジャンルgにおける特徴語抽出の流れを図1に示す。

ジャンルgにおける特徴語候補の特徴量を算出し、閾値以上の形態素を特徴語として抽出する。また立澤は特徴量を算出する際に比較対象となるコーパスは同じ製品のロコミを使用していたが、本研究においては同じ飲食のロコミと比較した場合、特徴語が抽出しにくいことが懸念される。そこで本研究では、特徴量算出の際の比較対象となるコーパスとして現代語コーパス⁴を用いる。飲食に関する記述がされていないコーパスと比較することで、よりジャンルの特徴語が抽出できると考えた。

まず二つのコーパスを形態素解析にかけ、更にジャンルgコーパスを品詞フィルタリングにかける。品詞フィルタリングでは、属性・評価値となり得る品詞の形態素を特徴語候補として抽出する。特徴語候補とする品詞一覧を表1に示す。

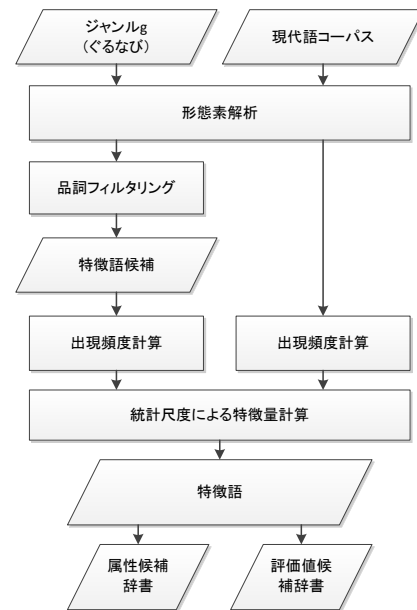


図1 特徴語抽出フローチャート

表1 特徴語候補品詞一覧

要素	品詞	細分類	例
属性候補品詞	名詞	一般	肉
		固有名詞	カツ
評価値候補品詞	名詞	サ変接続	マッチ
		形容動詞語幹	苦手
		副詞可能	たくさん
	動詞	自立 (サ変は除く)	合う
		形容詞	自立
	副詞	一般	ぷりぷり
助詞類接続		ピリッ	

特徴語候補やコーパス全体の頻出頻度を算出し、それらを用いて特徴量を算出する。閾値以上の形態素から表1に記された品詞によって各要素に分類し、辞書を作成する。

5.2 評価文抽出手法

評価文抽出処理の流れを図2に示す。評価文抽出に使用するデータは、ジャンルgに分類される店舗のロコミ、店舗の評価対象、作成したジャンルgの要素辞書である。評価対象は、評価文抽出対象のサイトに掲載されている店舗のメニューページから引用する。

まず、ロコミ内から評価対象が出現した文以降の全文を、その評価対象に対するロコミとして抽出する。抽出した文を構文解析にかけ、要素辞書を参照し条件にマッチした文をその評価対象に対する評価文として抽出する。評価文とする条件は、属性表現と評価値表現が共起し係り受け関係にある文、代名詞の係り先が評価値表現の文である。

³ <http://www.gnavi.co.jp/>
⁴ <http://www.kotonoha.gr.jp/shonagon/>

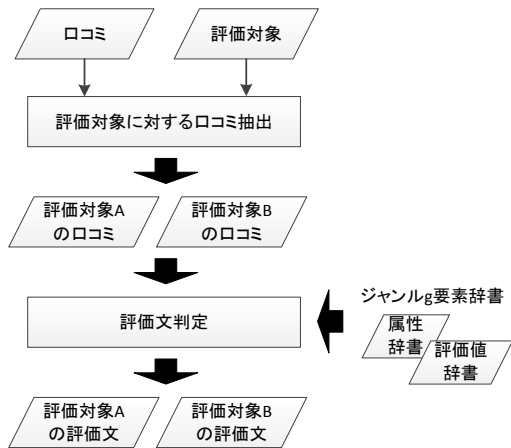


図 2 評価文抽出フローチャート

6. 特徴量算出手法

6.1 製品と飲食におけるロコミの比較調査

両ドメインのロコミにおける特徴語の傾向を知るため、比較調査を行った。立澤がコーパスとして使用した価格.comに掲載されたカメラに対するロコミと、本研究でコーパスとして使用するぐるなびに掲載された全ジャンルのロコミを用いた。製品ドメインはカメラのデザインに関するロコミ、飲食ドメインはカレーのジャンルに分類された料理のロコミでそれぞれの tf-idf 値を算出し、散布図を作成して特徴語の分布を調べた。製品のロコミと飲食のロコミにおける特徴語とそれ以外の一般語の散布図をそれぞれ図 3、図 4 に示す。

製品ロコミの特徴語はある一定範囲の tf 値にて特徴語が分布しているが、一般語は tf 値, idf 値共に広範囲で分布している。しかし、飲食ロコミにおいては特徴語と一般語共に広範囲に分布しており、また極端な低頻度の部分に集中して分布している。そのため、飲食においては製品に比べて低頻度の特徴語が非常に多いと考えられる。

6.2 頻度情報を抑えた tf-idf 値

比較調査の結果から、通常の tf-idf 値では飲食のロコミから特徴語を抽出できないと考えられる。そこで頻度情報を抑えた tf 値を用いることで飲食のロコミに適用できるのではないかと考えた。頻度情報を抑えた tf 値を式 1、idf 値を式 2、特徴量を式 3 で算出する。

G: ジャンル g	C: 現代語コーパス
G _T : 全形態素数	C _T : 全形態素数
G _U : 全ロコミ数	C _U : 全文章数
G _t : 形態素 t の出現数	C _t : 形態素 t の出現数
G _{u,t} : 形態素 t が出現するロコミ数	C _{u,t} : 形態素 t が出現する文章数

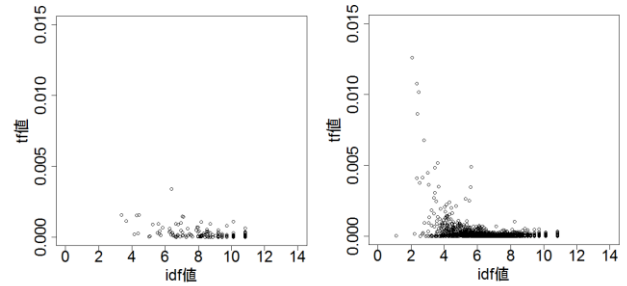


図 3 製品ロコミの特徴語 (左) と一般語 (右) の散布図

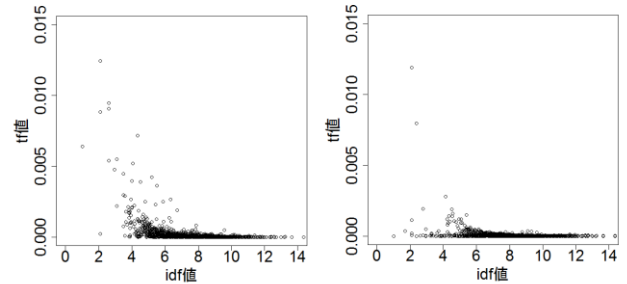


図 4 飲食ロコミの特徴語 (左) と一般語 (右) の散布図

$$tf = \frac{\log_{10} G_t}{G_T} \quad \text{式 1}$$

$$idf = \log \left(\frac{G_U + C_U}{G_{u,t} + C_{u,t}} \right) + 1 \quad \text{式 2}$$

$$F(G, t) = tf \times idf \quad \text{式 3}$$

G_t の対数を取った値を tf 値に与えることで高頻度と低頻度との tf 値の差を小さくし、低頻度語の特徴語を抽出できると考えられる。

6.3 特徴量の検証

飲食のロコミにおける頻度情報を抑えた tf-idf 値の有用性を示すため検証を行った。ぐるなびのロコミから、カレー (4351 ロコミ)、ラーメン (39499 ロコミ)、うなぎ・どじょう (3338 ロコミ) の 3 ジャンルを手で正解タグ付けを行い、比較する特徴量指標を用いて特徴語を抽出し 10 分割交差検定を行った。

比較する特徴量指標は、通常の tf-idf 値、頻度情報を抑えた改良 tf-idf 値、相互情報量[4]、ポアソン確率[5]とした。

6.3.1 結果

検証結果を表 2 に示す。3 ジャンル全てにおいて、改良 tf-idf 値の F 値が最大となった。改良 tf-idf 値は、適合率は低いが再現率は他の指標に比べ非常に高い数値であるため、特徴語を網羅的に抽出できていると言える。

6.3.2 考察

拡張 tf-idf 値では、他の特徴量では抽出できなかった低頻度の特徴語が抽出できた。例えばカレージャンルの中の「グリーンピース」や「ホウレンソウ」など、カレージャ

表 2 特徴量検証結果 (%)

ジャンル	特徴量指標	適合率	再現率	F 値
カレー	tf-idf	71.7	69.1	70.4
	相互情報量	71.9	60.6	65.8
	ポアソン確率	70.4	57.4	63.2
	拡張 tf-idf	63.1	86.5	73.0
ラーメン	tf-idf	54.3	69.9	61.1
	相互情報量	52.5	58.1	55.2
	ポアソン確率	56.0	64.1	59.8
	拡張 tf-idf	52.4	74.7	61.6
うなぎ・ どじょう	tf-idf	70.1	68.0	69.0
	相互情報量	70.8	61.7	65.9
	ポアソン確率	67.5	57.8	62.3
	拡張 tf-idf	60.6	86.1	71.1

ジャンルの中では珍しい属性が特徴語として抽出することができた。このような低頻度語は通常の tf-idf 値では tf 値の重みが強すぎるため、idf 値がいくら高くても特徴語として抽出できなかつたが、tf 値の重みを小さくすることで、idf 値の重みがより特徴量に反映されたため抽出できたと考えられる。

しかし「店」「雰囲気」など、料理ではなく飲食店の評価をする際に頻出するような形態素がノイズとして抽出されてしまった。ぐるなびには料理別に掲載されていても評価文とはいえない文が混ざっており、現代語コーパスでは頻出しないために抽出されてしまったと考えられる。

7. システム評価実験

評価文抽出の精度を調べるため、評価実験を行った。ぐるなびのカレー、ラーメン、うなぎ・どじょうの3ジャンルを用いてそれぞれの要素辞書を作成した。特徴量には改良した tf-idf 値を用いた。

評価文抽出の対象は食べログのロコミとした。食べログのカレーに属する 12 店舗 98 ロコミ、ラーメンに属する 16 店舗 108 ロコミ、うなぎ・どじょうに属する 11 店舗 100 ロコミと各店舗の評価対象名を抽出し、提案システムを用いて評価文を抽出した。抽出した評価文とあらかじめ人手で作成した正解データを比較し、適合率、再現率、F 値を求めて評価した。

7.1 結果

システムの評価文抽出精度を表 3 に示す。またシステムの入出力結果例を以下に示す。

【店舗例 1 店舗ジャンル：うなぎ・どじょう】

<入力ロコミ 1>

カウンターに座ると、おしぼりが出てきて、うな重を注文すると、ものすごく時間がかかりますので、ごゆっくりお

表 3 評価文抽出精度 (%)

	適合率	再現率	F 値
カレー	58.0	76.7	65.4
ラーメン	70.1	73.3	71.4
うなぎ・どじょう	54.8	76.2	63.4
全体	61.0	75.4	66.7

待ちください。とのこと。食べログを見て行ったので、覚悟はできていた。そして目の前で、活きたうなぎを見事な手さばきで捌いていく。それにしても、ご主人の動き、話し方を見てみると、変わってる人だなーと思う。お店の空間も不思議だと書かれていた人が多かったけど、一番目につくのはご主人。そして待つこと約 50 分。全然、苦にならなかった。口に入れると香ばしい香りがしてとろっととろける。たれは薄目だけど、物足りない感じはまったくない。うなぎの脂がうまくでてるのか、すばらしい味でした。上品なうなぎという印象。肝吸いも薄いのはうなぎに合っているのだと思います。名古屋のうなぎが好きなのですが、関東風もおいしいんだとはじめて思いました。総評 3.5 か 4 か悩んだんですが、好きな調布にあるということで、甘めの 4 にしました。今度は白焼きを食べてみたいです。

<出力結果 1>

評価対象：うな重

- (1) そして目の前で、活きたうなぎを見事な手さばきで捌いていく。
- (2) それにしても、ご主人の動き、話し方を見てみると、変わってる人だなーと思う。
- (3) お店の空間も不思議だと書かれていた人が多かったけど、一番目につくのはご主人。
- (4) 口に入れると香ばしい香りがしてとろっととろける。
- (5) たれは薄目だけど、物足りない感じはまったくない。
- (6) うなぎの脂がうまくでてるのか、すばらしい味でした。
- (7) 上品なうなぎという印象。
- (8) 肝吸いも薄いのはうなぎに合っているのだと思います。
- (9) 名古屋のうなぎが好きなのですが、関東風もおいしいんだとはじめて思いました。

【店舗例 2 店舗ジャンル：ラーメン】

<入力ロコミ 2>

ネギラーメン 650 円を注文。私が入った後にすぐにお客様 2 人。列び〜すごっ！さて、ラーメンきました。しらがネギに辛味のごま油がかかっています。良い香り〜。スー

プをまず. う～ん. 美味しいです. でもちょっとサラっとして感じが私にはちょっと物足りないかな・・・. そして, 一気に麵をいただきます. 玉ねぎは多めと言えばよかったな・・・. 美味しいです. 本当にでも～やっぱコクがないかな～. 列ぶ感じがわからない・・・. 好みもあるからな～. 御馳走様でした～.

<出力結果 2>

評価対象: ネギラーメン

- (1) 私が入った後にすぐにお客様 2 人.
- (2) しらがネギに辛味のごま油がかかっておりません.
- (3) 良い香り～.
- (4) スープをまず.
- (5) うなぎの脂がうまくでてるのか, すばらしい味でした.
- (6) 上品なうなぎという印象.
- (7) でもちょっとサラっとして感じが私にはちょっと物足りないかな・・・.
- (8) そして, 一気に麵をいただきます.
- (9) 玉ねぎは多めと言えばよかったな・・・.

7.2 考察

評価実験結果の適合率が低くなってしまった主な要因として, 評価対象が出現してから後方全てを評価文判定の対象とすると, 料理に関する記述以外の文が多く含まれてしまうことが挙げられる. そのため更に評価対象別に文書分類を行う必要がある.

また, 要素辞書にノイズと思われる表現が含まれたために, 誤って評価文と判定された事例も見られた. 出力結果 1-(2)は属性辞書に登録された「人」と評価値辞書に登録された「変わった」が係り受け関係であるために誤って抽出されてしまった. 「変わる」は例えば「変わった食材が使われて…」などで評価値表現として使われるため適切であるが, 「人」に関しては属性表現としては不適切であり, 属性辞書のノイズと言える.

また評価値辞書に追加された出力結果 1-(1)の「活きた」や出力結果 1-(3)の「つく」など, 動詞によって誤抽出されてしまった文章が多く見られた. 動詞においては活用形が未然形の場合評価値表現となる場合が多いが, 連用形の動詞が使用された場合評価値表現としては不適切であることが多い. そのため動詞に関しては活用形を考慮した評価値辞書作成が必要である.

正解タグが付いているが評価文として抽出されなかった文章においては, ほとんどが短い文章であった. 例えば入力ロコミ 2にある「美味しいです.」という評価値のみの評価文は抽出することができないように, 飲食のロコミには短い評価文が多く, 属性表現と評価値表現が共起してい

ないために評価文と判定されなかったと考えられる. そのため, 属性・評価値が共起しなかった場合の条件を新たに追加する必要がある.

全体を通して, 要素辞書に追加されたノイズによって誤って抽出された文が多くみられた. 特に動詞においてノイズと思われる評価値表現が多いため, 活用形やメニュー名との共起・N-gramなどを考慮するなど, ノイズを減らす新たな手法を考える必要がある. この問題を解決することで, 評価分抽出精度の向上につながると考えられる.

8. 結論

本研究では, 製品と飲食のロコミの傾向を探ったうえでぐるなびのロコミから特徴語を抽出し, 頻度情報を抑えた tf-idf 値を特徴量とすることで低頻度の特徴語も網羅的に抽出することができた. これらを踏まえ評価文の要素辞書を作成し, 食べログのロコミから評価文を適合率 61%, 再現率 75%, F 値 67%の精度で抽出することができた.

9. 今後の課題

要素辞書作成において, 低頻度の特徴語は抽出できたがノイズも多く抽出されてしまった. 例えば「当店」のように評価文の要素とは言えないが現代語コーパスでは頻出しなないために特徴語として抽出されてしまった語が多かった. ぐるなびには料理別に掲載はされているが, 評価文でない文章が混ざってしまっているロコミもある. そのためぐるなびのロコミから更にフィルタリングをかけて評価文の要素が含まれる文章のみをコーパスとして使用する必要がある. また頻度情報だけでなく, 共起・N-gramなどを考慮することで特徴語抽出の精度が向上すると考えられる.

評価文抽出においては, 評価文判定にかける範囲に問題が見られた. 評価対象が出現した文の後方全てを判定するのではなく, 評価対象別に文章分類を行うなどの方法で改善する必要がある. 評価文判定においては, 評価値のみ出現した場合など新しい条件を追加する. またその他のジャンルで評価文を抽出し汎用性を示したい.

参考文献

- [1] 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, 2005.07
- [2] Yuki Tachizawa: An automatic classification of Product review into given view-points, HCI International 2013, pp.598-606
- [3] 矢野宏美, 目良和也, 相沢輝昭: 嗜好を考慮した評判情報検索手法, 自然言語処理, Vol.108, No.164, pp.165-170, 2004.11
- [4] Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze: Introduction to Information Retrieval, Cambridge University Press(2009)
- [5] 廣嶋伸章, 安田宜仁, 藤田尚樹, 片岡良治: 地理情報検索におけるクエリ入力支援のための特徴語の提示, 第 26 回人工知能学会全国大会, 2012