

# レストランレビュー文章群からの評判とその対象の抽出 Acquisition of attribute and evaluative expressions from restaurant reviews

今平純一<sup>†</sup> 柴田喜匡<sup>†</sup> 二本木伸佳<sup>†</sup> 細川雅由<sup>†</sup> 奥村学<sup>‡</sup>

Junichi Imahira Yoshimasa Shibata Nobuyoshi Nihongi Masayoshi Hosokawa Manabu Okumura

## 1. まえがき

近年、個人を情報発信源とした CGM コンテンツが増大しており、その膨大な情報からの情報抽出の技術が注目されている。その中でもある特定の対象に対する評判を抽出するニーズは高くなっている。この評判の抽出の際には、通常なんらかの表現辞書が必要となっているが、この表現辞書は、抽出を行いたい対象の領域毎に異なり、新規に構築するには、それなりのコストがかかってしまう。そこで、本研究ではある特定の領域において、できるだけ人手を介さずに辞書構築を半自動で行う方法を実現し、その方法を実際にレストランに関する領域に適用することで評価を行う。

## 2. 辞書構築元としてのレビューサイト

辞書を構築するには構築のための資源となるなんらかの文章群が必要である。文章群の候補として blog や新聞など様々なものがあるが、構築した辞書でその後の評判抽出結果が左右されるため対象となる文章群はできるだけ良質である必要がある。そのためには文章群は以下の条件を満たす必要がある。

- (1) 良質な評判表現を多く含んでいる
- (2) 特定領域に関して一定量の文を用意できる

以上の条件を満たすものとして、レビューサイトがあげられる。レビューサイトは、評判を述べることを目的としており、文章中に評判が含まれる割合が高い。また評判の対象が特定されており、他の領域に関する評判が含まれていることが少ない。

そこで本研究では、レストランに関するレビューサイトを利用する。

## 3. 辞書構築手法

評判というものは、〈対象〉〈属性〉〈評価〉の3つの組み合わせ<sup>①</sup>と考えられる。レビューサイトでは、通常レビューのタイトルに対象が明示されていることが多く対象の抽出は容易である。本研究では、評価表現と属性表現の抽出を行い、その辞書を同時に作成していく手法を示す。

### 3.1 辞書構築概要

最初に種となる評価表現、属性表現を評価辞書、属性辞書として登録する。この種となる表現は、いずれか片方だけに登録することも可能である。そして3.2に示すプロセスを行うことで

辞書への登録候補を抽出する。その抽出された評価表現、属性表現を新たに辞書に加えて同様のプロセスを繰り返し行うことで辞書を拡充していく。

なお辞書登録候補抽出処理において、係り受け解析には cabocha<sup>②</sup>を形態素解析には mecab<sup>③</sup>を利用した。

### 3.2 辞書登録候補抽出方法

レビュー文章内からの辞書登録候補の抽出は、以下のフローで行う。

- (1) 前処理  
登録済みの評価辞書、属性辞書の読み込み、HTML タグの除去、文章の文への分割などが行われる。
- (2) 文の係り受け解析
- (3) 辞書登録表現を含む文節ペアの取得  
評価辞書または属性辞書内の表現を含む文節とその文節と係り受け関係にある文節をペアとして取得する。この時、係り受けの距離が遠すぎるペアと文節を形態素解析した結果の単語数が多いものを含むペアは除外する。  
抽出したペアの辞書に登録されていた側が評価辞書に含まれていた場合、係り受け関係にあった文節内の語を属性辞書候補とみなし、逆に属性辞書に含まれる語を含む文節のペアの場合は、係り受け関係の文節内の語を評価辞書候補とみなす。
- (4) 係り受けの形式によるフィルタリング  
係り受け関係の順番により、以下の判別を行いペアのフィルタリングを行う。

- ・ 「属性表現」 - 「評価表現」

文節ペアが、属性表現文節、評価表現文節の順で現れる場合、属性表現文節を形態素解析して、文節内の最後に出現する単語が「助詞-格助詞」「助詞-係助詞」且つそれが「が」「は」「も」のいずれかであるものを残す。

例：属性辞書に「お肉」がある場合

○お肉が/ジューシーだ

→文節「ジューシーだ」に評判語候補を含むとみなす

×お肉の/産地は

- ・ 「評価表現」 - 「属性表現」

文節ペアが、評価表現文節、属性表現文節の順で現れる場合、評価表現文節を形態素解析して、文節内の最後に出現する単語の活用形が「基本形」「体言接続」であるものを残す。

例：属性辞書に「ピザ」がある場合

<sup>†</sup>株式会社NTTPコミュニケーションズ

<sup>‡</sup>東京工業大学

○贅沢な/ピザだった

→文節「贅沢な」に評判語候補を含むとみなす

×アンチョビの/ピザです

#### (5) 辞書登録候補単語の抽出

文節内の単語を末尾から順に品詞を確認し、品詞が記号、助詞、助動詞、接続詞、以外が出現するまで削除し評価表現語、属性表現語として抽出する。ただし評価表現を含む文節の場合は、助動詞「ない」は削除対象としない。抽出語の最後の単語が活用していた場合は、基本形に変換を行う。

例：文節から抽出する例

風味でしたし、→ 風味

変わらないね。→ 変わらない

#### (6) 抽出した評価表現の品詞によるフィルタリング

抽出した語を形態素解析し、その最後の単語の品詞を利用してフィルタリングを行う。

文節の最後の単語が、評価表現語の場合は「名詞-一般」

「形容詞」「名詞-形容動詞語幹」であるものとし、属性表現語の場合は、「名詞-一般」「名詞-固有名詞」であるものとする。

## 4. 抽出実験結果

3で示した方法を複数サイクル実行することで評価辞書、属性辞書の構築を行う。抽出時には、抽出できた表現語だけでなく抽出回数も合わせて記録し、1回のサイクル終了時点でその抽出回数が10件以上となっていた語を辞書に登録して再度抽出処理を行う。これを新規に辞書に登録できる単語がなくなるまで行う。

元文章としてレストランレビューサイトの中華カテゴリに絞った3000件と和食カテゴリに絞った3000件を用意して行った。

また、初期登録状態として以下の単語を評価辞書、属性辞書に登録した。

【評価辞書】

美味しい、多い、狭い、良い

【属性辞書】

味、量、お店、雰囲気

最終的に作成された辞書を見ると評価辞書については、「最高」「リーズナブル」「おしゃれ」などのように中華と和食に共通に現れる表現が多かった。属性辞書では「メニュー」「ランチ」「価格」などのようなカテゴリを問わない共通表現とあわせて、中華カテゴリの「餃子」「杏仁豆腐」「麻婆豆腐」などや、和食カテゴリの「お酒」「お寿司」「ネタ」などのようなカテゴリに特有なものも多く取得できた。

一方、属性辞書において、「人」「目」のようにレストランという領域において登録される属性として考えると不適切と考えられるものが数件観察された。「人」に関しては、「人が多い」のような表現から抽出されている傾向が強く、「目」に関

しては、「濃い目」のような表現から抽出されていた。

また、初期に登録する種となる単語を変えて、抽出処理を行った場合でも、最終的に作成される辞書はほぼ同一のものとなり、辞書の内容は種に依存せず、元文章に依存することもわかった。

以下、表1、表2に辞書の成長過程を示す。

表1 中華カテゴリでの辞書成長過程

周回数	評判辞書登録数	属性辞書登録数	総抽出候補数	評判辞書追加単語数	属性辞書追加単語数
1	4	4	1206	28	20
2	32	24	1822	24	44
3	56	68	1490	13	13
4	69	81	1455	6	3
5	75	84	1388	1	1
6	76	85	1380	0	1
7	76	86	1380	0	0

表2 和食カテゴリでの辞書成長過程

周回数	評判辞書登録数	属性辞書登録数	総抽出候補数	評判辞書追加単語数	属性辞書追加単語数
1	4	4	899	12	17
2	16	21	1143	7	32
3	23	53	1119	19	3
4	42	56	1022	1	9
5	43	65	990	2	0
6	45	65	986	0	0

## 5. まとめ

本稿ではレストランという領域における評判抽出のための辞書の構築方法について述べ、実際に確認を行い効率的に行えることを確認した。今回最終的に登録される価値があるような表現が辞書に登録されないケースも多くあったが、辞書登録のための閾値を下げたり、元文章となるレビュー文章の数を増やすことで、もっと多くの単語を辞書に登録できることを期待できる。一方、文章中で評判表現となっているが、抽出できなかった表現があったり、抽出しても評判表現と言えないようなノイズも含まれているため、なんらかの抽出精度をあげるための対策も必要になる。

本手法を用いることでレストラン以外にレビューサイトが存在する家電や映画などのエンターテインメントなどの領域にも適用できるものと考えられる。

## 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. テキストマイニングによる評価表現の収集. 研究報告「自然言語処理」No.154, 2003
- [2]<http://chasen.org/~taku/software/cabocha/>
- [3]<http://mecab.sourceforge.net/>