

WEB の評判情報を用いた評価モデルの作成とシステム化

八田 衆導 , 石崎 俊

{ shudo , ishizaki }@sfc.keio.ac.jp
慶応義塾大学 政策・メディア研究科

概要

近年, Web 上で人の意見が評判情報として記述された情報が増大しており, 多様な活用が可能であるとして重要視されている. 本研究では, 宿泊施設の意見について答えた自由回答文を利用して, 宿泊施設に関する評価表現として自動的に抽出し, サポートベクターマシンを用いて宿泊施設の総合評価を行う評価モデルの構築を行った. 評価表現の抽出では, 肯定および否定表現の分類を中心に解析した. この評価表現の抽出から宿泊施設の特徴を示す特徴ベクトルと, 多数の評価アンケート(5 段階での評価)によって得られた総合評価値を用いて評価モデルを作成して分析した結果などを報告する.

1. はじめに

近年, Web アプリケーション(Microsoft .Net Framework など)や Web コミュニケーションツール(ウェブログなど)の開発とともに, Web 上で評判情報が増大し続けている. 従って, これらの評判情報が商品購入の意思決定のための広範で有用な資源となる可能性が高い.

そのため, これらの情報から評価表現を抽出, 集約することの必要性が高まっている. ある商品に対し定量的で信頼性のあるデータを取得することができれば利用が促進されると思われる. しかし現状ではそのデータを用いて商品进行评估する分析者が必要であるという問題が生じる. 近年ではナイーブベイズ(以下, NB)やサポートベクターマシン(以下, SVM)など非線形データに対して頑強な学習モデルが開発され, 学習モデルを利用した様々な分析の自動化することが可能となりつつある.

そこで, 本研究では, 商品の評判情報から自動的に総合評価するシステムを検討する. 事例として, 宿泊施設の評判情報(宿泊施設予約サイト「じゃらん」*)を用いて, その宿泊施設を総合評価する学習モデルを作成し, その評価を行う. 学習モデルには SVM を利用した. ナイーブベイズのような分類ルールによる学習アルゴリズムより, SVM などの連続値を扱える学習モデルの方が精度の高い分類を実現できる可能性がある. また, 多値分類における学習精度の比較を行った研究[1]において, SVM と NB に大きな差は見られないことが分かっている.

以下, 第 2 章では, 全体のシステムフローについて述べ, 第 3 章では評判情報から評価表現の抽出

し, 肯定的な評価表現(肯定評価表現)と否定的な評価表現(否定評価表現)へ分類する方法を紹介する. 第 4 章では SVM のアルゴリズムについて述べ, SVM による評価モデルの生成, かつ, その精度について報告する.

2. システムフロー

システムの流れは以下のようになる.

1. 評価語データベースの作成
2. 評判情報の解析(係り受け解析, CaboCha*)
3. 評価表現の抽出
4. 否定・願望の処理
5. 評価学習のためのデータ作成
6. 評価学習

評価語データベースは, 商品のキーワードとそれら进行评估する語または記号(評価語)をリンクさせたものとする. 例えば, 「風呂」というキーワードに対しては「広い」「きれい」などの評価語がリンクされている. この評価語ではさらに肯定評価語, 否定評価語に分類し, 評価語データベースを設定する.

次に, このデータベースと評判情報の係り受け解析によって評価表現を抽出する. 文にキーワードがある場合に係り受けのノードをたどることによる評価語探索を行うことで評価表現の抽出を行う.

評価表現として, キーワードと評価語の対を抽出するが, 実際には「料理はあまりおいしいとは思わなかった」というように評価表現に対する否定や願望の表現が行われている場合がある. そのため, ここでは, 否定と願望のパターンを分析することにより, これらの処理モデルを作成し, 評価表現

*1 「じゃらん」 <http://www.jalan.net/>における口コミ情報を利用した

*2 <http://chasen.org/~taku/software/cabocha/>

を最終的に肯定評価表現,否定評価表現と分類する.評価表現の抽出,否定・願望の処理をもとに評価表現分類モジュールを作成する.

SVM による評価学習を行うために,評価表現分類モジュールを用いて評判情報を特徴ベクトル化する.また,多数の被験者によるアンケートにより,被験者は評判情報を読んで宿泊施設を 5 段階で表現する.

この評判情報の特徴ベクトル化と総合評価値による学習データを作成し,SVM により評価学習を行い,宿泊施設の評価モデルを生成する.

3. 評価表現の分類と評価

3.1 評価表現の抽出方法

評価表現を抽出するうえで評価語データベースが不可欠である.もちろん,評価表現の抽出精度はそのデータベースのデータ量にも左右されることが考えられる.鈴木らは,ナイーブベイズ分類器とEMアルゴリズムを用いて評価表現を自動収集する研究[2]を行っており,将来には評価語データベースが自動的に蓄積されることが考えられる.そのため,本研究における評価表現の抽出では,宿泊施設に関する評価語データベースは人手で作成し,将来は自動的に作成されることを期待している.

このデータベースと係り受け構造を用いて評価表現を抽出する.また,評価表現をもつ文 2190 文でキーワードと評価語をもつ文節間の距離を分析した結果,距離が 6 以内である文が全体の 97.2%であった.このような評価表現をもつ文の特性を踏まえて,キーワードと評価語をもつ文節間の距離が 6 以内の範囲で,それらが係り受けのノードに存在する場合のみ評価表現として抽出する.

3.2 否定・願望の処理

抽出した評価表現は,評価語が肯定・否定のいずれかに判別されることで,肯定または否定評価表現に分類される.しかし,その分類段階では,評価表現をもつ文の文意を反映していない場合があるため,以下のような否定・願望の処理を行う.

評価表現が否定となる例として,「料理は,量は多かったけど,そんなにおいしいとは思わなかった」という文について考えると,評価表現としてまず「料理 おいしい」「量 多い」といった評価表現が抽出できる.しかし,「料理 おいしい」の評価表現は評価語が肯定であるため,肯定評価表現とされるが,評価表現自体に否定が行われているため,否定評価表現として分類される必要があ

る.

次に,評価表現に対して願望が表現されている例として,「温泉はもう少し大きかったらよかった」という文について考える.この文では,「温泉 大きい」という肯定評価表現が抽出される.しかし,文意として「温泉は大きくはない」ということを含んでいるため,否定評価表現に分類する必要がある.

パターン処理の観点から否定のパターンと願望のパターンまとめると,それぞれ以下のように集約できる.

否定のパターン

1. 評価語と否定辞が同じ文節で接続する
2. 評価語をもつ文節の後に否定辞のみ,または否定辞と助詞が接続した文節がある
3. 評価語をもつ文節の後に思考動詞,知覚動詞と否定辞が接続した文節がある

願望のパターン

1. 評価語に助動詞「た」の仮定形,または接続助詞「ば」「と」が接続し,その後の文節に標準評価語が存在する
2. 願望を表す形容詞「ほしい」または助動詞「たい」が評価語に同じ文節で接続する

ここで,標準評価語はすべてのキーワードに対して考えられる評価語としている.

これらのパターン処理には,文節内の品詞とその接続情報,係り受け関係の三つの情報を用いて処理を行っている.上記のパターンのみの処理により,最終的に評価表現を肯定評価表現と否定評価表現に分類する.

3.3 評価表現の分類精度

評価表現の分類の処理は,肯定,否定の分類のいずれにおいても同様の処理過程である.そのため,ここでは否定評価表現へ分類精度を算出する.否定評価表現を含む文を宿泊施設の口コミ情報から人手により取り出し,正解データを作成した.正解データは表 1 のようになる.「評価語:否定」は評価表現における評価語が否定的意味である場合である.「評価語:肯定および評価表現:否定」と「評価語:肯定および評価表現:否定」は,それぞれ評価語が肯定的意味であり,前者は評価表現が否定される場合,後者は評価表現に対する願望がある場合である.それぞれ,順に P1,P2,P3 と表現する.

作成した評価表現分類モジュールにより正解データから否定評価表現の抽出を行った.その結果を表 2 に示す.

表 1:否定評価表現をもつ文の正解データ数

P1	評価語:否定	122 文
P2	評価語:肯定および評価表現:否定	82 文
P3	評価語:肯定および評価表現:願望	80 文

表 2:評価表現分類による抽出率

P1	評価語:否定	96.72%
P2	評価語:肯定および評価表現:否定	87.21%
P3	評価語:肯定および評価表現:願望	86.25%

P1 の文では高い精度で評価表現の抽出を行うことができた。この処理においては、係り受けのノードを辿りながら評価表現を抽出するため、分類精度は評価語データベースにキーワードまたは評価語が格納されているか否かにより変わってくる。つまり、P1 においては、データベースの補完によりさらに精度を向上させることができると考える。P2,P3 においては、否定や願望の表現の曖昧さを含め、分類することができない文が存在した。P2 では、評価語が肯定であると判断されたにも関わらず、肯定、または否定評価表現に分類できなかった誤分類のパターンとして、次のような場合が多く存在した。

1. 否定辞を含まず、評価語が他の品詞により否定される (ex.従業員は親切さに欠ける)
2. キーワードが一般化し、分析対象のキーワードではない (ex.ここまで親切な旅館はなかなかない)

ここで、1 の誤抽出については、否定辞の他に語彙的否定形式をとる品詞によっても否定の処理を行うことで誤分類を解消できる。しかし、2 の例では、「旅館」がその宿泊施設を示さず、一般的な宿泊施設を示しており、それに対する否定により、対象とする「旅館」に対し肯定的な評価を行っている。このような他との比較による評価を行った文の分析は検討の余地があると考えられる。P3 における誤分類は、例えば「温泉はもうちょっと広かったらなあ」のように、評価語とチャンクした助動詞「た」の仮定形、または、接続助詞「ば」の後に標準評価語が存在しない場合であった。

4. 評価モデルの作成

4.1 SVM

SVM は 1995 年に Vapnik らにより開発された非線形データに対して頑強な 2 値識別器である [3]。SVM の学習モデルで特徴的であるのは、カーネルトリックとマージン最大化と呼ばれる方法である。

カーネルトリックは非線形データを扱う場合に有効な方法で、非線形となる特徴ベクトルをカーネル関数を用いて、はるかに高次元の空間へ写像し、線形分離可能となるように特徴ベクトルを変換する方法である。この高次元空間で 2 つの空間に線形分離することにより 2 値識別を行う。

ここで線形分離を行うときに用いられるのがマージン最大化である。マージン最大化とは、2 つのクラス間の距離を最大にするような線形関数を抽出する方法である。この線形関数は、クラスが異なる最も近い距離にあるデータ集合 (特徴ベクトル) により求められる。

SVM は、この二つの手法による過程を繰り返し行うことにより、高い精度で分離する最適なカーネル関数を選択する。これにより、頑強な 2 値識別を行うことを可能としている。

SVM を多値識別に応用することが求められるような事例が多くある。この場合の手法として、Multi-class SVM が考えられている。これは、クラス対での学習 (one-versus-one)、または、あるクラスと全ての他のクラスでの学習 (one-versus-rest) を行うモデルである。両者での学習では、学習時間や識別時間の差はあるが、認識率に差はない。本研究では図 1 のようにクラス対での学習により段階的に識別する one-versus-one での学習を適用した。

しかし、Multi-class SVM による学習では、2 クラスに識別されないクラスに対しても、その 2 クラスのいずれかに識別するようにしているため、識別する課題によっては識別精度が低下する問題がある。次節で行う評価学習では、評価の程度を表す離散値をクラスとしており、また、特徴ベクトルの各要素値の大小がクラスの識別を左右すると考えられるため、この Multi-class SVM での学習が有効であるといえる。

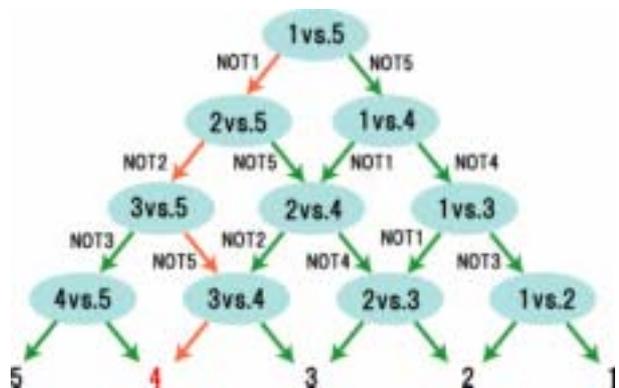


図 1:SVM における one-versus-one による識別

4.2 評価モデルの作成と精度

宿泊施設の評判情報をもとにその特徴ベクトルを作成するうえで、まず、風呂、料理、料金、サービス、立地、施設の6つのカテゴリを設定した。さらに、各カテゴリを肯定表現、否定表現に分けることで、12次元の特徴ベクトルを作成する(表3)。

表3:特徴ベクトルの要素群

風呂肯定	風呂否定	料理肯定	料理否定	料金肯定	料金否定
サービス肯定	サービス否定	立地肯定	立地否定	施設肯定	施設否定

要素値は、各要素での肯定・否定評価表の出現数を評判情報の記述者数で割った平均とした。

次に、クラスを設定するために、宿泊施設の評価アンケートを行った。各宿泊施設の評判情報を読んでもらい、評価3を「普通」の基準とした5段階での総合評価を行った。特徴ベクトルで設定したカテゴリを踏まえた評価とするために、各カテゴリについても評価してもらった。

以上を踏まえ、学習データを作成し、Multi-class SVMでの評価学習を行った。これにより、評価モデルを生成し、テストデータによる認識率を計算した。評価学習は10 fold cross validationで行った。これは、学習データをランダムに10分割した10個のデータセットを作成し、9個のデータセットで学習を行い、残りの1個で評価を行う過程を繰り返す方法である。この方法により最適なモデルが生成される。

まず、評価値対での認識率を算出した(表4)。各評価値対での学習データ・テストデータ数はそれぞれ一定の80個、25個としている。これらの認識率を比較すると、評価3 vs 評価4での認識率が非常に低いことがわかった。

表4:評価値対での認識率(%)

	1	2	3	4	5
1		71.4	90.0	91.7	95.0
2	71.4		70.3	70.4	87.9
3	90.0	70.3		54.1	91.0
4	91.7	70.4	54.1		85.9
5	95.0	87.9	91.0	85.9	

ここで、3 vs 4 の評価値対の識別モデルにより5段階での全体評価モデルの精度が低下していることが分かる。そのため、評価モデルとして、5段階評価値、(1,2,4,5)の評価値、(1,2,5)の評価値を識別する3つの評価モデルを作成し、その認識率を算出した。その結果を表5に示す。

表5:各評価モデルの精度

評価値	学習データ	テストデータ	認識率
5段階	600個	128個	62.2%
1,2,4,5	397個	85個	81.6%
1,2,5	250個	50個	90.2%

5段階での評価モデルは、62.2%と低い認識率であり、評価3,4を省いた評価モデルでは高い認識率となった。また、学習データ数、テストデータ数を全ての評価モデルで一定にした場合においても、同様の認識率であった。

4.3 考察

1,2,5 の評価値での評価モデルの認識率から、「良い」または「悪い」といった明確な評価の識別は可能であることが理解できた。一方で、3vs4の評価値対では認識率が低い。これはアンケートにおいて「普通」という評価と「まあまあ良い」という評価に揺れが生じていること、もしくは、「普通」という評価自体が多少肯定的な意味を含んでいるということが考えられる。

評判情報から学習モデルを用いて自動評価を行うシステムを作成する場合には、学習させる評価値の設定、かつ、アンケートの作成法を考える必要がある。本研究では連続的な離散値による評価学習を行ったが、「買いたい」、「買いたくない」といった項目による評価モデルなどの検討が必要であると考えられる。

5. おわりに

本研究では、評価表現に対する否定、願望を理解することにより、詳細な肯定・否定評価表現への分類を行った。しかし、評価を表現した文には評価表現を含んだ文以外のものも存在するため、それらの文を分析するモデルを作成する必要があると考える。また、評判情報から商品を自動評価するシステムを構築し、その可能性を検討した。自動評価モデルとしてSVMを適用したが、今後は他のモデルとの比較、検討を行う必要があると考える。

参考文献

- [1] Zadrozny, B., Elkan, C, Transforming Classifier Scores into Accurate Multiclass Probability Estimates, ACM, pp694-699, 2002
- [2] 鈴木泰裕, 高村大也, 奥村 学, "Weblog を対象とした評価表現抽出", 人工知能学会, セマンティックウェブとオントロジー研究会, SIG-SWO-A401-02, 2004.
- [3] Valdimir N. Vapnik, Statistical Learning Theory, John Wiley & Sons inc, 1998