

Weblog を対象とした評価表現抽出

鈴木 泰裕[†] 高村 大也[‡] 奥村 学[‡]

概要

Weblog に含まれる評判情報を扱うには、テキスト中の評価表現を抽出し、その表現が肯定・否定のどちらの感情を含んでいるかを特定することが不可欠である。評価表現には、「おいしい」のような、それ自体に肯定・否定の感情が込められているものもあるが、「大きい」といった形容詞のように、評価対象やどの部分が大きいのかといった点を考慮しなければ、肯定・否定の感情がわからないものもあり、それらを機械的に分析することは難しい課題となっている。そこで、文中から評価表現を抽出し、肯定・否定の分類を行うために、我々はラベル付きデータに加えて、ラベルなしデータも学習に組み込む semi-supervised 学習手法を用いることを提案する。実際に我々が収集した Weblog データに対し、実験による評価を行い、ラベルありデータのみを用いる方法で 66.7% の正解率、ラベルなしデータも併せて用いることで 70.4% の正解率が得られることがわかった。

1 はじめに

近年、MovableType¹ などの Weblog 作成支援ツールや、ココログ² などの Weblog ホスティングサービスの普及により、Web 上で Weblog を公開する人が急増している。そのため、個人が発信する有用な情報源の一つとして注目が集まっている。個人が発信する情報の中でも現在特に重要視されている情報の一つは評判情報である。Web 上に散在している評判情報を一括して扱えるようにすることは、一般の人々にとっての意思決定や、企業にとっての危機管理などの支援になる。Weblog に含まれる評判情報を扱うには、テキスト中の評価表現を抽出し、その表現が肯定・否定のどちらの感情を含んでいるかを特定することが不可欠である。これを実現するために従来手法の多くでは、人手により肯定・否定の

分類を行い、評価表現辞書を作成していた。しかし、評価表現によっては評価対象ごとに込められている肯定・否定の感情が違ってくることがあり、評価対象などの情報を含めた辞書を作ろうとすると、莫大なコストがかかる。本論文では、Weblog から評判情報を収集・分析するための技術の一つとして、評判情報の核となる評価表現を抽出しつつ、発言全体が肯定的な評価であるか、否定的な評価であるかを判定する semi-supervised な手法を提案する。

我々のアプローチは、多くの場合において評価表現はある種の周辺情報を伴って出現するという仮定に基づいている。例えば、「この店のラーメンはチャーシューが厚くて良い。」という文から、ラーメンのチャーシューが厚い=肯定という知識が得られるように、「A なので良い」という型の発言からは A が肯定的な評価となり得ることがわかる。さらに、A が肯定的な評価を表すという知識を用いることにより、「A なのでいい感じ」という例から、「いい感じ」が肯定を示唆する周辺情報だということがわかる。これ以外にも様々な周辺情報と評価表現が、それらの共起関係を考慮することにより、抽出可能になる。

このような考えを、我々は semi-supervised な EM アルゴリズムとナイーブベイズ分類器によって実現する。より具体的には、周辺情報などを素性にして、

[†]東京工業大学大学院総合理工学研究科
Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology
yasu@lr.pi.titech.ac.jp

[‡]東京工業大学精密工学研究所
Precision and Intelligence Laboratory, Tokyo Institute of Technology
{takamura, oku}@pi.titech.ac.jp

¹movabletype.org <http://www.movabletype.org/>

²blog サービス [ココログ] <http://www.cocolog-nifty.com/>

ナイーブベイズ分類器で発言が肯定的か否定的かを分類する。その際に、少量のラベル付きデータ（肯定的か否定的かあるいは評価でないかのラベルを手で付与した発言の集合）を初期データとして、大量のラベル無しデータを EM アルゴリズムで取り込むことにより、どの素性が肯定を示唆しているかといったパラメータを更新していく。

提案手法の有効性を示すために、実際に Weblog から集めたデータに対し、「肯定的評価」、「否定的評価」、「評価でない」の3カテゴリに分類する実験を行う。

以下、第2節で関連研究について、第3節で提案手法について、第4節で使用したデータについて、第5節で評価実験について、第6節でまとめを述べる。

2 関連研究

評判情報に関して行われている研究では、評判情報を扱う単位には大きく分けて3つのタイプがある。

1. 1つの文書が、肯定・否定どちらの感情を持って書かれたかを判定することを目指す
2. 1文レベルで何らかの評価を表しているかどうか、そしてそれが肯定的な評価か否定的な評価かを判定することを目指す
3. 単語単位で評価を表している個所を特定し、肯定・否定を判定することを目指す

評判情報を文書単位で扱ったものとしては、まず、Turney ら [1] の研究が挙げられる。彼らは、映画のレビューがその映画をお薦めしているのか否かを、教師なし学習で判定することを提案している。そのアルゴリズムは、与えられたフレーズと“excellent”という単語との相互情報量、そして“poor”という単語との相互情報量を計算することに基づいている。松本ら [2] は単語系列や部分依存木を素性として用いることで、映画レビューの肯定/否定分類性能が向上することを示した。

文レベルで評判情報を扱った研究としては、立石ら [3] が、あらかじめ用意しておいた辞書を用い、近接演算や構文的な特徴を利用して意見らしさを求めることにより、評判情報の抽出を行っている。

藤村ら [4] は機械学習 (C4.5) により、単語などの出現回数を素性として用い、掲示板のコメント一文

一文を評判 (良), 評判 (悪), その他に分類する手法を提案している。

表現単位で評判情報を扱った研究には、乾ら [5] の、節間の因果関係を利用して、節が表す出来事が望ましいか否かを判定するというものがある。

また、小林ら [6] は、意見の主要部分を < 対象, 属性, 評価 > の3要素から構成されるものと定義し、対象名辞書, 属性表現辞書, 評価語辞書を、対象領域毎に効率的に構築する手法を提案している。例えば、<A> の【B】が / は / も / に / を [C]

- A:対象になり得る語
- B:属性表現になり得る語
- C:評価語になり得る語

というパターンに当てはまる表現がコーパス中にあり、A, B, Cのうち2つが既に辞書に存在する語であるならば、残りの1つを辞書に加える候補として抽出する。このような共起パターンを9種用い、辞書を拡張していく。ただし、辞書に加える候補を実際に辞書に加える際には、人手によるフィルタリングが必要としている。また、表現の収集のみであり、肯定・否定の区別は行っていない。

藤村ら [7] は肯定的な評判と否定的な評判の差分をとることにより、肯定的な評価表現と否定的な評価表現を抽出している。

3 Semi-supervised 学習による評価表現抽出と評判分類

本研究で対象にしている評価表現とは、「この店のラーメンはスープが旨い。」という評判情報を含む文の中の、どの対象 (この店のラーメン) についての、どの部分 (スープ) が、どうなのか (旨い) という評判情報の核となる三要素からなるものである。

本研究の目標は、有効な評価表現を抽出しつつ、評判情報を、肯定的評価、否定的評価、非評価のいずれかに分類することである。第1節でも述べたように、我々は、評価表現はある種の周辺情報を伴って現れるという仮定を置いている。よって、評判情報の分類結果により新たな評価表現や周辺情報を発見することができ、また獲得された評価表現や周辺情報により未知の評判情報を正しく分類することができる。

このようなブートストラッピング的な発想を実現するために、我々は教師付き学習手法を EM アルゴ

リズムで補強する semi-supervised な方法を適用する。ここでは、教師付き学習手法として、ナイーブベイズ分類器を用いる。ナイーブベイズ分類器を用いたのは、EM アルゴリズムと組み合わせることにより、文書分類で高い性能を発揮することが Nigam らにより示されているからである [8].

以下、本研究で用いた semi-supervised な方法を説明する。

3.1 ナイーブベイズ分類器による評判情報分類

ここではまずナイーブベイズ分類器 (Naive Bayes classifiers) の一種である多項モデルについて説明する。多項モデル以外に頻繁に使用されるモデルとして多変数ベルヌーイモデルがあるが、多項モデルはテキスト分類において、多変数ベルヌーイモデルよりも良い結果を出すと報告されている [9].

多項モデルでは、カテゴリ c が与えられたときに、事例 \mathbf{x} が生起する確率は、

$$P(\mathbf{x}|c, \theta) = P(|\mathbf{x}|)|\mathbf{x}|! \prod_w \frac{P(w|c)^{N(w, \mathbf{x})}}{N(w, \mathbf{x})!} \quad (1)$$

となる。ここで、 $P(|\mathbf{x}|)$ は長さ $|\mathbf{x}|$ の文が生起する確率であり、 $N(w, \mathbf{x})$ は文 \mathbf{x} 中での素性 w の出現頻度である。文の生起はイベントのセットとしてモデル化され、そのイベントでは、単語は全語彙の中から選び出される。

ナイーブベイズ分類器を評判情報分類に適用した場合、各発言が事例 \mathbf{x} に相当し、カテゴリ c は、肯定的評価、否定的評価、非評価のいずれかの値をとる。使用される素性は、評価表現やその周辺に出現する単語などであり、第 4 節で詳述する。

3.2 ナイーブベイズ分類器と EM アルゴリズムの組合せ

EM アルゴリズムはいくつかの変数 (隠れ変数と呼ばれている) が観測できない状況で、モデルを最尤推定する手法である [10]. Nigam ら [8] はナイーブベイズ分類器と EM アルゴリズムを組み合わせることを提案している。

ナイーブベイズ・モデルの式において、関係ない

要素を無視すると、次の式を得る：

$$P(\mathbf{x}|c, \theta) \propto \prod_w P(w|c)^{N(w, \mathbf{x})}, \quad (2)$$

$$P(\mathbf{x}|\theta) \propto \sum_c P(c) \prod_w P(w|c)^{N(w, \mathbf{x})}. \quad (3)$$

以降、モデルのパラメータ群をまとめて θ と表す。

c を隠れ変数とし、ディリクレ分布をパラメータの事前分布とすると、対数尤度の隠れ変数に関する期待値 (Q 関数) は次のように定義できる：

$$Q(\theta|\bar{\theta}) = \log(P(\theta)) + \sum_{\mathbf{x} \in D} \sum_c P(c|\mathbf{x}, \bar{\theta}) \times \log \left(P(c) \prod_w P(w|c)^{N(w, \mathbf{x})} \right). \quad (4)$$

ここで、 $P(\theta) \propto \prod_c (P(c))^{\alpha-1} \prod_w (P(w|c))^{\alpha-1}$ であり、また、 α はハイパーパラメータ、さらに D はモデルの推定に用いられる事例の集合である。

この Q 関数より、次の EM 計算式が得られる：
E-ステップ:

$$P(c|\mathbf{x}, \bar{\theta}) = \frac{P(c|\bar{\theta})P(\mathbf{x}|c, \bar{\theta})}{\sum_c P(c|\bar{\theta})P(\mathbf{x}|c, \bar{\theta})}, \quad (5)$$

M-ステップ:

$$P(c) = \frac{(\alpha - 1) + \sum_{\mathbf{x} \in D} P(c|\mathbf{x}, \bar{\theta})}{(\alpha - 1)|C| + |D|}, \quad (6)$$

$$P(w|c) = \frac{(\alpha - 1) + \sum_{\mathbf{x} \in D} P(c|\mathbf{x}, \bar{\theta})N(w, \mathbf{x})}{(\alpha - 1)|W| + \sum_w \sum_{\mathbf{x} \in D} P(c|\mathbf{x}, \bar{\theta})N(w, \mathbf{x})}. \quad (7)$$

ここで、 $|C|$ はカテゴリ数、 $|W|$ は単語数を表す。ラベル付き事例については、式 (5) は使用されない。その代わりに、 c が事例 \mathbf{x} のカテゴリならば $P(c|\mathbf{x}, \bar{\theta})$ は 1.0 とし、そうでなければ 0 とする。

統計物理学的視点に基づいた EM アルゴリズムの変種に、Deterministic Annealing EM (DAEM) がある [11, 12]. この変種では、モデルの複雑さを調整することができる。DAEM は、E-ステップで式 (5) の代わりに次式を使用することで実現できる：

$$P(c|\mathbf{x}, \bar{\theta}) = \frac{\{P(c|\bar{\theta})P(\mathbf{x}|c, \bar{\theta})\}^\beta}{\sum_c \{P(c|\bar{\theta})P(\mathbf{x}|c, \bar{\theta})\}^\beta}. \quad (8)$$

ここで、 β はモデルの複雑さを決めるパラメータで、値が大きいほどモデルは複雑になる。

ラベルなしデータに対してラベルありデータが極端に少ないと、学習を繰り返していくうちにラベ

ラベルなしデータの影響が強くなりすぎて、結果が悪くなってしまうことがある。そのため、 $\lambda(0 \leq \lambda \leq 1)$ を用いて、ラベルなしデータの影響が小さくなるように式(4)の右辺の第2項を次式と入れ換える：

$$\sum_{\mathbf{x} \in D^l} \sum_c P(c|\mathbf{x}, \bar{\theta}) \log \left(P(c) \prod_w P(w|c)^{N(w, \mathbf{x})} \right) + \lambda \sum_{\mathbf{x} \in D^u} \sum_c P(c|\mathbf{x}, \bar{\theta}) \log \left(P(c) \prod_w P(w|c)^{N(w, \mathbf{x})} \right).$$

ここで、 D^l はラベル付きデータ、 D^u はラベルなしデータである。この式が示すように、 λ の値が小さいほど、ラベルなしデータの影響が小さくなる。

この新たな Q 関数を用いて導出したアルゴリズムを我々は使用した。Q 関数の値の変化が十分に小さくなることを終了条件とした。

3.3 辞書の構築方法について

前述した分類器を用い、大量のラベルなしデータに対してラベル付けを行い、その統計をとることにより、各三つ組について、評価表現としての使われ易さ、そして、肯定的評価としての使われ易さ、否定的評価としての使われ易さを求めることができる。これにより三つ組に対する一般的な評価の極性の知識を持った辞書を構築することができる。

4 データの準備

奥村ら [13] が開発したシステム (blogWatcher) により収集した blog 記事 (html の一部) から、文を切り出し、係り受け解析を行い、三つ組とその周辺情報を抽出する。係り受け解析には、CaboCha³を用いた。

4.1 文分割

まず、係り受け解析を行うために、blog 記事を文単位に分割する必要がある。文の切れ目となる可能性が高いいくつかの記号と html タグを用い、分割を行った。その際、小数やイニシャルなどの中にも使用されるピリオドや、固有名中に使われる句点など、分割すべきでない箇所を除外するためのヒューリスティックスを導入した。

³<http://chasen.org/~taku/software/cabocho/> において入手可能。

4.2 三つ組抽出

次に、評価表現の候補すなわち評価対象・属性・評価語の三つ組の候補を収集する。評価語の候補としては、形容詞・形容動詞・動詞の「ある」を考える。他にも「安さ爆発」というように、サ変名詞が評価語となったり、「雪のような肌」のように比喩が評価的に使われることもあるが、今回は対象外とした。そして、この評価表現の候補に係る文節の中から、評価対象、属性を見つけ、対象・属性・評価の三つ組を抽出する。ただし、対象と属性を特定することは困難であるため、経験的に対象の候補は具体名詞が多く、属性の候補は抽象名詞が多いと考えられるので、対象の候補は具体名詞のみ、属性の候補は抽象名詞のみに限定した。ある名詞が抽象名詞かどうかの判定は NTT の日本語語彙大系 [14] から作成した抽象名詞辞書 (エントリ数:51056) とのマッチングに加え、「～度」「～性」という語尾を持つ名詞も抽象名詞と判定する。対象や属性の候補が複数ある場合には、ガ格であるものを最優先し、更にもその中でも評価語候補に近いものを優先する。

4.3 打ち消し表現

助動詞「ない」は、直前の語の意味を打ち消したり逆転したりする効果がある。そのため、評価語候補に「ない」が連なる場合は、これを連結したものを一つの素性として扱う。例えば、「悪くない」は「悪い+打ち消し表現×奇数個」となり、「悪い」「悪くなくもない」は「悪い+打ち消し表現×偶数個」となる。

4.4 考慮する周辺情報

機械学習で素性として用いる周辺情報として、手がかりになり得ると考えられる以下の情報 (全て原形) を使用する。

1. 文末の文節内の形態素
2. 文末の記号 (。 . ! ? ?)
3. 文に含まれる感動詞 (「あぁ」「やれやれ」等)
4. 文に含まれる丸括弧の中身
5. 三つ組いずれかに係る文節内の形態素
6. 評価語候補に係る文節内の形態素

やれやれ、
 この
 ハードディスクは
 ギューンという
 音が
 うるさいので
 (怒)
 もう
 うんざりだ。

図 1: 例文の係り受け解析結果

7. 評価語候補と同じ文節に含まれる形態素

三つ組と係り受け関係にある文節は繋がりが強く、有用な手がかりになると考えられるために周辺情報として用いる。文末から得られる情報も加えたのは、書き手の結論としての感情が含まれることが多いためである。丸括弧の中身を用いるのは「(笑)」や「(怒)」,あるいは「(^_^)」といった顔文字等の情報を利用するためである。例えば、「やれやれ、このハードディスクはギューンという音がうるさいので(怒)、うんざりだ。」という文を係り受け解析すると図 1 のようになる。この文中には、対象:「ハードディスク」、属性:「音」、評価語候補:「うるさい」という三つ組が存在する。この三つ組の周辺情報としては、文末の形態素として「うんざり」「だ」、文末の記号として「。」、感動詞として「やれやれ」、丸括弧の中身として「怒る」が、三つ組いずれかに係る文節内の形態素として「この」、評価語候補が係っている文節内の形態素として「うんざり」「だ」、評価語候補と同じ文節に含まれる形態素として「ので」が抽出される。

評価語候補を含む文節内に逆接の効果をもつ接続助詞がある場合、係り先の文節内の形態素と文末の文節内の形態素と文末の記号に、逆接の効果をもつ接続助詞があるという属性を付与する。三つ組のいずれかに係る文節内に逆接の効果をもつ接続助詞がある場合には、その文節内の形態素に、逆接の効果をもつ接続助詞があるという属性を付与する。これにより、「この店のラーメンはおいしいけど、接客態度がイマイチだ。」というような文があったときに、逆接で連結している「ラーメンがおいしい」と「接客態度がイマイチだ」は逆の評価極性を持っているということをうまく学習に取り入れることができると考えられる。

4.5 データのフィルタリング

Web上の文書には非文が多く、口語的表現も多く、係り受け解析が困難であるので、得られたデータにノイズが多く含まれるという問題がある。係り受け解析が間違っていたり、対象・属性の抽出が間違っていたり、有効な周辺情報を持っていなかったりすると、学習に悪影響を及ぼすと考えられるため、そのような事例のフィルタリングを行った。具体的には、下記の条件のいずれかを満たす事例を削除した。

1. 既に存在する事例と全く同じ。
2. 周辺情報が全く無い。
3. 対象・属性どちらも抽出できなかった。
4. 評価語候補が名詞を連体修飾している。
(例: おいしいうどん)
5. 評価語候補の係り先の文節に、形容詞または形容動詞語幹または感動詞が入っていない。
6. 対象・属性に時期を表す語が入っている。
(例: 「今日は楽しかった」、「クリスマス前後は忙しい」)
7. 対象に名詞性述語接尾辞を含む。
(例: 活躍ぶり、焼きたて、若さ)
8. 対象・属性・評価語候補の間隔が16バイト以上あいている。(離れ過ぎている場合は係り受けが間違っていることが多い)

更に、出現回数が少ない素性は効果が無いということと、ノイズである可能性が高いため、3回以上出現した素性のみを使用することにした。

フィルタリングの結果、blog 約 30 万記事 (1 記事は 1 日分) から抽出した約 150 万の事例は、24262 に減少した。使用された素性の種類数は各素性タイプごとに表 1 の通りである。

4.6 ラベル付けの方針

ラベル付き訓練データとして、あるいは分類結果に対する評価データとして、正解のラベルがついた事例が必要である。正解のラベルを付ける際の簡単なガイドラインを以下に説明しておく。

- 各事例に、非評価、肯定的評価、否定的評価、という三種類のラベルのうち一つを付ける。

表 1: 素性の種類数

| 素性タイプ | 種類数 |
|----------------------|------|
| 文末の形態素 | 1750 |
| 文末の記号 | 13 |
| 文中に含まれる感動詞 | 74 |
| 文中に含まれる丸括弧の中身 | 701 |
| 三つ組いずれかに係る文節内の形態素 | 2427 |
| 評価語候補が係る文節内の形態素 | 1629 |
| 評価語候補と同じ文節に含まれる形態素対象 | 236 |
| 属性 | 331 |
| 評価語候補 | 1162 |
| 評価語候補の品詞 | 613 |
| | 3 |

表 2: ラベル付きデータの内訳

| クラス | データ数 (割合) |
|-------|-------------|
| 非評価 | 195 (18.8%) |
| 肯定的評価 | 409 (39.5%) |
| 否定的評価 | 431 (41.6%) |

- 文全体として書き手がどう対象を評価しているかではなく、文全体を考慮した上で、該当する三つ組が表す事象に対し、書き手が肯定的感情を抱いているか、否定的感情を抱いているか、そのどちらでもないかを推測してラベルをつける。

例：このラーメンは麺が硬いのが良い。

ラーメン/麺/硬い 肯定的評価

例：このうどんは麺が太いので、縦に切って食べればおいしい。

うどん/麺/太い 非評価

- 対象や属性が抽出されていない場合は、省略されていると考える。

例：このラーメン屋はとてもおいしかったのでまた来たい。

ラーメン屋//おいしい 肯定的評価

- その文だけ見ても判定できないものは、非評価とする。

例：このラーメンは熱くて、スープが濃い。

ラーメン//熱い 非評価

ラーメン/スープ/濃い 非評価

以上のガイドラインに従って、人手により 1035 の事例に対してラベル付けを行った。ラベルの内訳を表 2 に示す。

表 3: λ と β の値を変更したときの分類正解率

| $\lambda \setminus \beta$ | 0.005 | 0.01 | 0.1 | 0.5 | 0.75 |
|---------------------------|-------------|-------------|------|------|------|
| 0.01 | 67.3 | 67.2 | 66.5 | 63.3 | 62.9 |
| 0.1 | 70.4 | 70.4 | 68.1 | 44.6 | 44.1 |
| 0.5 | 68.5 | 68.5 | 60.2 | 41.8 | 41.7 |
| 1 | 68.0 | 68.1 | 52.3 | 39.9 | 41.1 |

5 実験

前節で述べた、事例数 1035 のデータを用いて提案手法の評価実験を行った。ナイーブベイズ分類器のパラメータ α については 2.0 に固定した。EM アルゴリズムで取り込まれるラベルなしデータのサイズは、23227 である。

ラベル付きデータ中で最も該当する事例が多いのは否定的評価のクラスであるため、全ての事例について否定的評価のクラスであると予測したときの正解率 41.6% が最も粗いベースラインと考えられる。

5.1 結果と考察

Leave-one-out による評価実験

まず、1035 のラベル付けされたデータのうち 1034 のデータを用いて学習を行い、残りの 1 つのデータで精度のテストを行うということを 1035 回繰り返す Leave-one-out 交差検定法により精度を求めた。 λ と β の値を変更して評価実験を行った結果は表 3 である。EM 反復を行う前の、ナイーブベイズ分類器だけによる分類精度は、66.7% である。 λ と β とともに、非常に小さい値のときに、分類結果が良くなっており、 $\lambda=0.1$ 、 $\beta=0.01$ の時に最も正解率が高く、70.4% である。これはラベルなしデータの影響を少しずつ取り入れているということに相当する。Collins ら [15] は、ラベルなしデータに現在のモデルを使ってラベルを付け、訓練データに加える際、ラベルなしデータを一度に全て訓練データとするのではなく、少量ずつ加えて徐々にモデルを更新していく方が効果的に学習できると述べている。本稿の実験結果はこの Collins らの主張に合ったものといえる。恐らく、Web 上のデータを使っていてノイズが多いために、ラベルなしデータの影響を慎重に扱う必要があるためだと考えられる。

ちなみに、 $\lambda=0.1$ 、 $\beta=0.01$ の場合、どの訓練データの組み合わせでも EM 反復は 4 回で終了し、EM

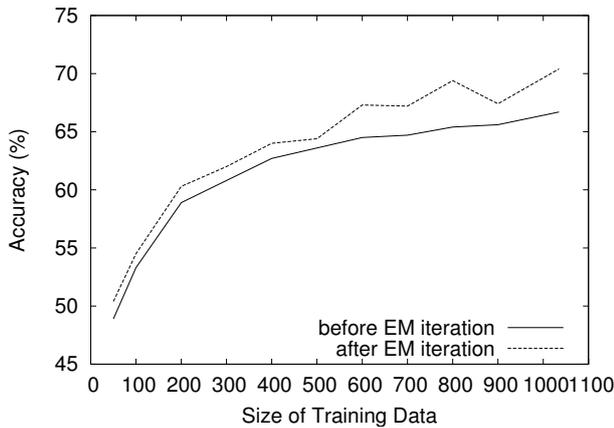


図 2: 訓練データのサイズに対する正解率

反復を 1 回行った時点では、正解率 69.6%、2 回行った時点では 70.4%、その後正解率は変化しなかった。

訓練データサイズの影響

次に、ラベル付き訓練データがさらに小さいときに、提案手法がどのような振舞いをするかを観察するため、訓練データサイズを変化させ実験を行った。各サイズについて、ランダムサンプリングで 10 回実験を行った平均値を示したのが図 2 である。ただし、各サイズについて最適なパラメータ β , λ が選択されている。図中、“before EM iteration” は、EM アルゴリズムが適用されていないナイーブベイズ分類器の正解率を示しており、“after EM iteration” は EM アルゴリズムを適用した場合に対応する。この図からわかるように、適切なパラメータが選ばれば、ラベル付き訓練データのサイズに関わらず EM アルゴリズムが効果的であることがわかる。

カバー率に対する精度と再現率

各事例は確信度を伴って非評価・肯定的評価・否定的評価に分類される。肯定的な評価であると分類された事例と否定的な評価であると分類された事例に関して、それぞれ高い確信度をもって分類された事例に限定した Precision(精度) と Recall(再現率) をグラフにしたのが図 3 と図 4 である。ここで、クラス c に関する Precision とは分類器がクラス c と予測した事例のうち、人手でもクラス c とラベル付けされていた事例の割合であり、Recall とは人手よりクラス c とラベル付けされた事例のうち、分類器がクラス c と予測した事例の割合であると定義する。

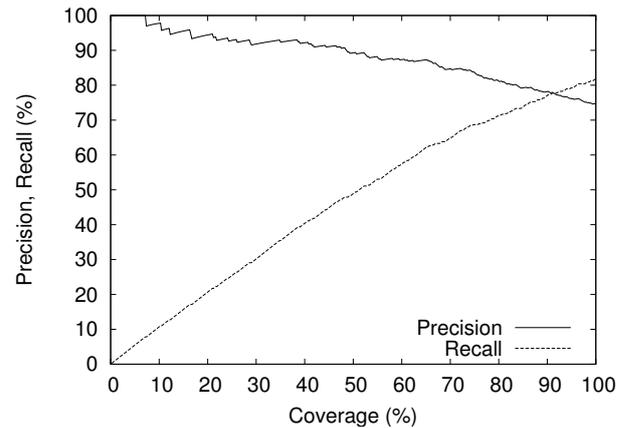


図 3: 確信度が上位 X% で肯定的評価と分類された事例に限定した Precision と Recall

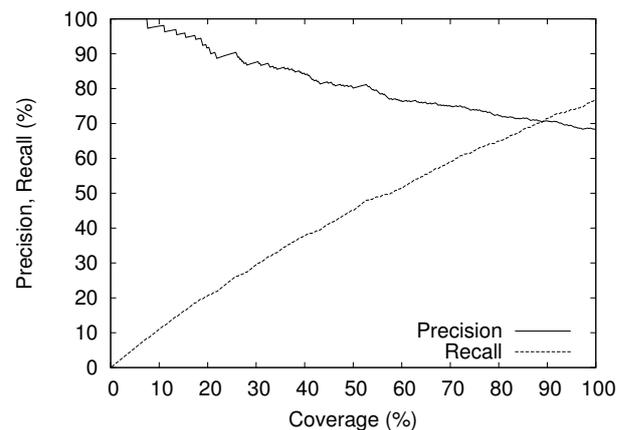


図 4: 確信度が上位 X% で否定的評価と分類された事例に限定した Precision と Recall

肯定的評価のクラス・否定的評価のクラス共に確信度が上位 30% の事例で約 90% という高い Precision を得られている。

エラー解析

以下、分類器による分類が失敗していた例についていくつか原因を考察する。⁴

- 発言：こんな体調で研究しても、生産性は悪いし、調子が良くなるってわけじゃない。
 - [属性]: 生産性, [評価語候補]: 悪い
 - [予測されたクラス]: 肯定的
 - [正解のクラス]: 否定的

⁴ただし原文をそのままここに載せるのは問題になりうるので、エラーの性質を損なわない範囲で原文を書き換えた。

- [エラーの原因]: 今回は周辺情報に関する打ち消し表現の連結は考慮しておらず、「良い」という形態素が係り先にあるが影響している。
- 発言: 優ちゃんは髪が多くて重たい印象があるから、卒業したら染めたらどうかなあと思ってたので、いい感じです。
 - [対象]: 優ちゃん, [属性]: 髪
 - [評価語候補]: 多い
 - [予測されたクラス]: 肯定的
 - [正解のクラス]: 否定的
 - [エラーの原因]: 係り受け関係にある周辺情報だけを見れば、否定的と判定できる可能性があるが、文末の形態素も素性となるため、「いい感じ」の影響を受けている。
- 発言: 市販の飲料は意外に糖分が多いので、倍に薄めて飲むと良いようです。
 - [対象]: 飲料, [属性]: 糖分
 - [評価語候補]: 多い
 - [予測されたクラス]: 肯定的
 - [正解のクラス]: 非評価
 - [エラーの原因]: 本来、「多いので」は読点以降全体に係るのだが、「良いようです」に係ると解析されている。CaboChaで扱えるのは文節単位の係り受けのみであり、句への係り受けは扱えないのが原因である。ただし、この例でのエラーは文末の形態素としても「良い」が抽出されていることの影響もある。

以上いくつか例を示したが、全体をまとめると、素性が不十分なもの、係り受け解析などの基盤技術の誤りに起因するもの、モデル自体の未熟さによるものに分けられる。

5.2 評価表現辞書構築の可能性

今回訓練した分類器を利用して、評価表現辞書を構築する場合を考える。

例えば、新たにラベル付けされた事例の中で、「クオリティ(-) + 高い」の組全てに対して肯定的評価のラベルがつけられており、「クオリティ + 高い + 打ち消し表現 × 奇数個」の組全てに対して否定的評

価のラベルがつけられていた。このことから「クオリティ(-)」が「高い」という表現は常に肯定的表現として使われると推測できる。

一方、「可能性 + 高い」に対しては、同程度の頻度で肯定的評価、否定的評価のラベルがつけられていた。よって、単に「可能性」が「高い」という表現のみから肯定・否定の判断をするのは難しい。

「レベル + 高い」に関しては、15 事例に対して肯定的評価のラベルが、5 事例に対して否定的評価のラベルがつけられ、「レベル + 高い + 打ち消し表現 × 奇数個」1 事例に対しては否定的評価のラベルがつけられていた。実際に「レベル + 高い」に否定的評価のラベルがつけられていた事例の元の文を見てみると、ライバルのレベルが高いことを良くないと思っているという文であった。このように、同じ属性・評価語の組み合わせでも、自分が好意を持つ対象に対して評価を行う場合と、好意を持たない対象に対して評価を行う場合では評価極性が反対になる。そのため、対象がきちんと抽出できることが重要であると言える。対象が省略されていたり、代名詞で置き換えられている場合に照応解析を行うことで、判定精度を向上させることができる。

6 おわりに

本論文では、Weblog から評判情報を収集・分析するための技術の一つとして、評判情報の核となる評価表現を抽出しつつ、発言全体が肯定的な評価であるか、否定的な評価であるか、あるいは非評価であるかを判定する semi-supervised な手法を提案した。提案手法は、ナイーブベイズ分類器と EM アルゴリズムの組合せから成っている。

実験では、パラメータが適切に選択された場合、EM アルゴリズムによりナイーブベイズ分類器の分類正解率が 3.7% 向上した。また、訓練データサイズを変化させても、EM アルゴリズムによる semi-supervised 学習は効果的であった。

今後の発展性として、まず、パラメータ推定の問題が挙げられる。今回は、主に適切なパラメータが選択された場合について議論した。しかし、交差検定などの手段を用いて適切なパラメータ選択が可能であるか実験的に確認する必要があるだろう。

また、我々が持つ知識をさらにモデルに入れていくことにより分類性能は向上すると思われる。例えば、周辺情報における打ち消し表現（「ない」など）

は今回はまだ取り入れていないが、それらを効果的に取り入れることで正しく分類できると思われる事例がエラー解析でいくつか見受けられた。

確率モデルという面では、単語のクラスタリング [11] が今後の重要な発展方向の一つである。例えば、「高い」と共起して肯定を表すような名詞群（レベル、解像度など）が統計的に求められれば、より多様な事例を正確に分類できるようになるだろう。

また、本稿で使用した semi-supervised な方法と Support Vector Machine などのカーネル分類器との組合せも興味深い発展方向である [16]。特に、高次のカーネル関数を用いることにより素性の組合せを考慮できる。例えば、「レベルが高い」は肯定だが、「値段が高い」は否定であることを考えると、素性の組合せが考慮できるモデルは本タスクにおいて効果的に働くものと予測される。

今回の手法を用いて blogWatcher の評判情報検索機能を拡張する方法としては、2通りある。まずは訓練した分類器をそのまま使用する方法である。そして、もう一つの方法は、大量のラベルなしデータに分類器を適用し、3つ組に対して文脈に依存しない知識を持った辞書を構築する方法である。今後は両者の可能性を考えて研究を進めていきたい。

参考文献

- [1] Peter D. Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)*, pp. 417–424, 2002.
- [2] 松本翔太郎, 高村大也, 奥村学. 単語の系列及び依存木を用いた評価文書の自動分類. FIT2004 第3回情報科学技術フォーラム (to appear), 2004.
- [3] 立石健二, 石黒義英, 福島俊一. インターネットからの評判情報検索. 情報処理学会研究報告, 2001-NL-144, pp. 75–82, 2001.
- [4] 藤村滋, 松村真宏, 石塚満. 機械学習による電子掲示板からの評判情報抽出. 情報処理学会第65回全国大会, 2003.
- [5] 乾孝司, 乾健太郎, 松本裕治. 出来事の望ましさ判定を目的とした語彙知識獲得. pp. 91–94, 2004.
- [6] 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. 情報処理学会研究報告, 2003-NL-154, pp. 77–84, 2003.
- [7] 藤村滋, 豊田正史, 喜連川優. 電子掲示板からの評価表現及び評価情報の抽出. 人工知能学会全国大会 (第18回), 2004.
- [8] Kamal Nigam, Andrew McCallum, Sebastian Thrun, and Tom Mitchell. Text classification from labeled and unlabeled documents using EM. *Machine Learning*, Vol. 39, No. 2/3, pp. 103–134, 2000.
- [9] Andrew McCallum and Kamal Nigam. A comparison of event models for naive bayes text classification. In *Proceedings of AAAI-98 Workshop on Learning for Text Categorization*, pp. 41–48, 1998.
- [10] Arthur P. Dempster, Nan M. Laird, and Donald B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B*, Vol. 39, No. 1, pp. 1–38, 1977.
- [11] Thomas Hofmann and Jan Puzicha. Statistical models for co-occurrence data. Technical Report AIM-1625, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1998.
- [12] Naonori Ueda and Ryohei Nakano. Deterministic annealing variant of the em algorithm. In *Advances in Neural Information Processing Systems 7*, pp. 545–552, 1995.
- [13] 奥村学, 南野朋之, 藤木稔明, 鈴木泰裕. blog ページの自動収集と監視に基づくテキストマイニング. 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-A401-01, 2004.
- [14] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦. 日本語語彙大系 CD-ROM 版. 岩波書店, 1997.
- [15] Michael Collins and Yoram Singer. Unsupervised models for named entity classification. In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, pp. 100–110, 1999.
- [16] Hiroya Takamura and Manabu Okumura. A comparative study on the use of labeled and unlabeled data for large margin classifiers. In *Proceedings of the 1st International Joint Conference on Natural Language Processing (IJCNLP2004)*, pp. 620–625, 2004.

Appendix

ラベル付けした 1035 のデータ全てを訓練データとし、学習を行った。事例が w を素性として含んでいたときに肯定的評価クラスに属する確率は $P(+|w)$ 、否定的評価クラスに属する確率は $P(-|w)$ と学習されたとする。このとき、 $P(+|w)$ と $P(-|w)$ が高かった素性、つまり、肯定的評価の手がかりになり易い素性と否定的評価の手がかりになり易い素性を表 4 及び表 5 に掲載する。

表 4: $P(+|w)$ の値が大きかった素性

| $P(+ w)$ | 素性 (w) | 素性タイプ |
|----------|---------|---------|
| 0.699 | 優しい | 評価語候補 |
| 0.682 | 大好き | 文末の形態素 |
| 0.678 | かっこいい | 評価語候補 |
| 0.669 | 大好き | 係り先の形態素 |
| 0.648 | * | 丸括弧の中身 |
| 0.643 | 演奏 | 属性 |
| 0.641 | 面白い | 評価語候補 |
| 0.640 | 楽しい | 文末の形態素 |
| 0.639 | * | 係り先の形態素 |
| 0.639 | 表情 | 属性 |
| 0.638 | 便利 | 文末の形態素 |
| 0.635 | 景色 | 属性 |
| 0.634 | おいしい | 評価語候補 |
| 0.632 | 映画 | 属性 |
| 0.627 | 美しい | 文末の形態素 |
| 0.623 | ラッキー | 文末の形態素 |
| 0.621 | 素晴らしい | 係り先の形態素 |
| 0.620 | 可愛い | 係り先の形態素 |
| 0.617 | 楽しい | 係り先の形態素 |
| 0.617 | うまい | 評価語候補 |
| 0.616 | 完成度 | 属性 |
| 0.616 | ・【逆接】 | 係る形態素 |
| 0.616 | 可愛い | 文末の形態素 |
| 0.615 | 美味しい | 評価語候補 |
| 0.609 | ラッキー | 係り先の形態素 |
| 0.609 | ありがたい | 係り先の形態素 |
| 0.608 | 嬉しい | 係り先の形態素 |
| 0.606 | チャット | 属性 |
| 0.606 | 仕草 | 属性 |
| 0.606 | 手間 | 属性 |
| 0.606 | 美しい | 係り先の形態素 |
| 0.601 | 【逆接】起こる | 文末の形態素 |
| 0.601 | 【逆接】起こる | 係り先の形態素 |
| 0.601 | ノ | 係る形態素 |
| 0.598 | 美しい | 評価語候補 |
| 0.597 | —【逆接】 | 係る形態素 |
| 0.597 | 眺め | 属性 |
| 0.597 | たまらない | 係る形態素 |
| 0.596 | 季節 | 係る形態素 |
| 0.596 | 色々 | 評価語候補 |
| 0.594 | 面白い | 係り先の形態素 |
| 0.593 | 楽しい | 評価語候補 |
| 0.593 | 豊富 | 評価語候補 |

表 5: $P(-|w)$ の値が大きかった素性

| $P(- w)$ | 素性 (w) | 素性タイプ |
|----------|--------|---------------|
| 0.703 | 腰 | 属性 |
| 0.697 | 悲しい | 係り先の形態素 |
| 0.651 | すみません | 文末の形態素 |
| 0.651 | すみません | 係り先の形態素 |
| 0.646 | いい加減 | 評価語候補 |
| 0.644 | すみません | 感動詞 |
| 0.637 | 悲しい | 文末の形態素 |
| 0.637 | 汚い | 評価語候補 |
| 0.636 | 残念 | 係り先の形態素 |
| 0.633 | 耐える | 係り先の形態素 |
| 0.633 | 辛い | 文末の形態素 |
| 0.633 | 忙しい | 評価語候補 |
| 0.631 | 残念 | 文末の形態素 |
| 0.619 | うるさい | 評価語候補 |
| 0.618 | つまる | 係り先の形態素 |
| 0.614 | こりゃ | 感動詞 |
| 0.614 | 下さる | 文末の形態素 |
| 0.612 | 辛い | 係り先の形態素 |
| 0.611 | たまらない | 文末の形態素 |
| 0.607 | 何 | 丸括弧の中身 |
| 0.605 | 歯切れ | 属性 |
| 0.605 | イマイチ | 文末の形態素 |
| 0.604 | 筋肉 | 対象 |
| 0.604 | 白い | 評価語候補 |
| 0.603 | 臭い | 評価語候補 |
| 0.601 | 怖い | 係り先の形態素 |
| 0.600 | 画質 | 属性 |
| 0.600 | 散々 | 評価語候補 |
| 0.600 | 実力 | 属性 |
| 0.600 | トーク | 対象 |
| 0.600 | 周 | 係る形態素 |
| 0.600 | むなしい | 文末の形態素 |
| 0.598 | 仕事 | 属性 |
| 0.595 | 方法 | 係る形態素 |
| 0.595 | 腰痛 | 属性 |
| 0.595 | 余計 | 評価語候補 |
| 0.594 | 出せる | 係り先の形態素 |
| 0.592 | ・ | 評価語候補と同文節の形態素 |
| 0.591 | こりゃ | 係り先の形態素 |
| 0.590 | どうも | 係る形態素 |
| 0.586 | こりゃ | 文末の形態素 |
| 0.586 | 意志 | 属性 |
| 0.586 | 打つ | 係り先の形態素 |