隠れ変数モデルによる複数語表現の感情極性分類

高村大也[†] 乾 孝司[††] 奥村 学[†]

複数語から成る評価表現のモデルおよびそれに基づいた分類手法を提案する．複数語から成る評価表現の感情極性は，その構成語の感情極性を単純に足し合わせるだけでは算出できないことが多い．極性の出現や反転が頻繁に起こる．そのような複数語表現の特性に対応するために，我々はモデルに隠れ変数を導入する．実験により，提案した隠れ変数モデルは複数語から成る評価表現分類において，約 82%という高い分類正解率を得ることに成功した．

Latent Variable Models for Semantic Orientations of Phrases

Hiroya Takamura,[†] Takashi Inui[††] and Manabu Okumura[†]

We propose models for semantic orientations of phrases as well as classification methods based on the models. Although each phrase consists of multiple words, the semantic orientation of the phrase is not a mere sum of the orientations of the component words. Some words can invert the orientation. In order to capture the property of such phrases, we introduce latent variables into the models. Through experiments, we show that the proposed latent variable models work well in the classification of semantic orientations of phrases and achieved nearly 82% classification accuracy.

1. 序 論

テキストにおける感情情報処理技術が，産業界を含む多くの場所で注目を集めている．そのような技術は，レビューの解析による新製品のサーベイ，アンケート処理など様々な応用の場を持つ．たいていの応用においては大量のデータを処理するので，感情情報処理の自動化は，高速で包括的な調査のためには必要不可欠である．

テキストの感情情報処理における最も基礎的な技術は，単語の感情極性の獲得であるといえる．ここで感情極性とは，ポジティブ（望ましい）かあるいはネガティブ（望ましくない）かを表す．たとえば，“美しい” はポジティブだが，“汚い” はネガティブである．また，“長い” などのようにニュートラル極性を考えることもできる．このタスクについては，いくつかの研究があり良い結果が出ている[4),8),9),17),20)]．次に解くべき問題の 1 つとして，複数語から成る表現の感情極性をいかにして扱うかという問題があげられる．これまでは，単語の感情極性のために開発された手法をそのまま複数語表現に適用した研究や，人手で作成した規則に基づく手法はあったが，複数語の特性を考慮に入れた計算モデルは提案されていない．本稿の目的は，複数語から成る評価表現のモデルおよびそれに基づいた分類手法を提案することである．

複数語表現の感情極性は，その構成語の極性の単純な和ではない．たとえば，“ノートパソコンが軽い” という表現はポジティブであるが，“軽い” も “ノートパソコン” もそれら自体はポジティブではない．このように極性の発現が頻繁に起こる．また，極性を反転させる作用を持つ単語も存在する．たとえば，“リスクが低い” においては，“リスク” のネガティブ極性が “低い” によって反転させられている．表現の感情極性がその構成語の感情極性から容易に予測できないこのような性質（非構成性と呼ぶことにする）がモデルに取り入れられる必要がある．上であげた非構成性は，「名詞＋述語」のように 2 語から成る単純な複数語表現にも十分に見られ，このような単純な複数語表現の感情極性分類問題に対する効果的な解決方法は，3 語以上から成る複雑な複数語表現を扱ううえでの基盤になると考えられる．また，2 語から成る複数語表現が，文書の感情極性分類に有効であるとの報告もある[19)]．

† 東京工業大学精密工学研究所
Precision and Intelligence Laboratory, Tokyo Institute of Technology
†† 東京工業大学統合研究院
Integrated Research Institute, Tokyo Institute of Technology

このような理由から本研究では「名詞＋述語」の場合を扱う．述語としては形容詞と形容動詞をここでは考える．

名詞に対応する確率変数，述語に対応する確率変数を用意して，確率モデルを考えていく．ここで複数語表現の非構成性をとらえるために，我々は隠れ変数をモデルに導入する．このモデルを用いて，“リスク”と“感染率”のように感情極性という観点で類似した（どちらも“低い”とポジティブである）語がグループ化されるようなクラスタリングを実現する．これにより，単語対の感情極性分類を高精度に行うことができるようになる．このような，分類問題に適切なクラスタの抽出は，より複雑な複数語の感情極性分類にも必要となる技術であり，本稿における提案手法は一般的な問題を解くうえでの基盤になると期待できる．また，極性が既知の単語とのコーパスにおける共起に基づいた方法[1),19)]が提案されているが，これらはニュートラル極性を取り込めない．なぜなら，極性が既知の単語として，ポジティブな語（たとえば，good，excellent）やネガティブな語（たとえば，bad，poor）は考えられるが，典型的なニュートラルな語は考えにくいからである．一方で，我々の手法はニュートラル極性を自然に取り込めるという利点がある．我々の手法は，単語ペアと感情極性の共起データのみを用いているので，言語非依存である．

本稿は以下のような構成である．まず 2 章で関連研究について説明する．3 章では本稿で使用する隠れ変数モデルについて説明する．4 章では，実験について述べる．最後に 5 章に結論を述べる．

2. 関 連 研 究

本研究の関連研究として，一般的な単語対の分類問題という側面と，感情極性分類という側面の両方から見ていくことにする．

2.1 単語対の分類

Torisawa[18)] は，格が未知である名詞と動詞の対が与えられたときにその格を推定するという問題において，確率モデルを利用した．彼らの確率モデルは，2 つの確率変数の同時分布モデルであり，確率的潜在意味解析（Probabilistic Latent Semantic Indexing, PLSI）モデル[5)] と等価である．Torisawa の手法は，隠れ変数モデルを単語対分類に利用しているという点で，我々のモデルに類似しているが，Torisawa の目的は格推定であり，感情極性分類という我々の目的と異なる．また，我々のモデルは，複数語の感情極性分類というタスクに合うように PLSI を拡張したものである．

Fujita ら[3)] は，自動的に言い換えられた文における誤った格割当てを検出するというタスクにおいて解決策を提案している．彼らはこのタスクを，動詞と名詞の対の分類問題として定式化している．彼らはまず，PLSI で隠れ変数を獲得し，その隠れ変数を素性として k–近傍法に類似した手法を用いた．格割当てにおける誤り検出という彼らの目的も我々のものと大きく異なる．また，彼らは確率モデルを素性抽出に利用しているという点で，我々の手法とも Torisawa の手法とも異なる．

2.2 感情極性分類

単語の感情極性分類についてはいくつかの研究があり，良い成果が出ている[4),8),9),17),20)]．具体的には，実験設定に依存するが，ポジティブもしくはネガティブの二値分類問題において 80 から 90%程度の値が報告されている．しかし，複数語表現の感情極性分類に関しては，これまでは，単語の感情極性のために開発された手法をそのまま複数語表現に適用した研究や，人手で作成した規則に基づく手法はあったが，複数語の特性を考慮に入れた計算モデルは提案されていない．

文書の感情極性分類において単語の出現パターンを使おうという試みはなされている．Pang ら[14)] は bigram を素性として文書の感情極性分類を行った．Matsumoto ら[11)] および松本ら[22)] は，シーケンシャル・パターンや依存木の部分木パターンを素性とすることを提案している．複数語から成るそのようなパターンは，文書の感情極性分類において有用であるということは示されたが，パターン自体の極性についてはまったく言及されていない．

鈴木ら[23)] は，Expectation-Maximization（EM）アルゴリズムとナイーブベイズ分類器を組み合わせることにより，ラベルなしデータを三つ組評価表現（対象，属性，評価語）の分類に取り込んだ．Turney[19)] は，単語の感情極性分類のために開発した手法を複数語表現にも適用している．彼らの手法は，種となる極性が既知の単語と複数語表現から成るクエリ（たとえば，“*phrase* NEAR good”）をウェブの検索エンジンに投げ，そのヒット数を用いて極性を決定する．Baron ら[1)] は，まず Xtract[16)] を用いてコーパスからコロケーションを抽出し，周辺の単語の極性によりコロケーションの極性を決定した．Baron らの手法は，種となる単語との共起を用いているという点で Turney の手法に非常に類似している．これら 3 つの手法は，コーパス中で複数語表現の周辺に出現する語句などのような周辺情報を利用しようとしたものである．一方，我々の手法は複数語表現に対し，その構成語の意味ク

表 1 複数語表現の感情極性分類における関連研究のまとめ
Table 1 Related work on phrase classification according to semantic orientations.

手法	アプローチ	使用するコーパス	分類に使用する主な情報	ニュートラル	未出現語
鈴木ら[23]	機械学習	教師付きおよび教師なし	周辺情報	△	×
Turney[19]	頻度	ウェブ	周辺情報	×	○
Baron ら[1]	頻度	教師なし	周辺情報	×	△
Inui[7]	人手	—	構成語の属性	×	×
Wilson ら[21]	人手	—	構成語の属性	○	×
提案手法	機械学習	教師付き	構成語の属性	○	×

ラスタを通して感情極性の生成をモデル化しようとする．また，Turney の手法[19] と Baron らの手法[1] はニュートラル極性を取り込めない．なぜなら，極性が既知の単語として，ポジティブな語（たとえば，good, excellent）やネガティブな語（たとえば，bad, poor）は考えられるが，典型的なニュートラルな語は考えにくいからである．しかし，我々の手法はニュートラル極性を自然な形で取り込むことができる．

Inui[7] は，複数語表現の感情極性分類において，各単語に *plus*/*minus* のどちらかの値をとる属性を考え，その属性値と構成語の感情極性に基づいた極性決定規則を提案している．たとえば，[negative+minus=positive] という規則は “リスク (negative)+低い (minus)” がポジティブであると決定する．また Wilson ら[21] は，*plus*/*minus* の属性とほぼ等価である polarity shifter という概念を導入し，複数語表現の感情極性を扱っている．どちらの研究においても，*plus*/*minus* 属性（もしくは polarity shifter）は人手で用意している．しかし，*plus*/*minus* 属性（もしくは polarity shifter）では複数語表現のすべてを適切に分類することはできない．たとえば，「ノートパソコン＋軽い」のように極性発現が起こる場合は，構成語がどちらも極性を持たないので，*plus*/*minus* 属性ではうまく対応できない．また，単語に属性を付与するという作業は専門的知識を要するものであり，包括的なリソースを作るのは困難である．本稿で提案する手法は，非常に一般的な視点からは，Inui や Wilson らのアイデアを確率モデルを用いて自動化し拡張した手法であるととらえることができ，このような問題を解決できる．

複数語表現の感情極性分類における関連研究について**表 1** にまとめる．左端の列は手法を示す．左端から 2 番目の列は用いられたアプローチを表す．極性が既知の単語との共起頻度を利用する方法は，“頻度” としてある．3 番目の列は，必要とされるコーパスデータの種類（教師付きコーパスか教師なしコーパスかなど）を表している．ウェブコーパスは，教師なしコーパスの一種と考えられるが，Turney の手法はウェブの利用を特長としているので，それを明示した．頻度依存の手法は，ウェブのような大規模なデータが利用可能な場合に高い性能が期待できる．ウェブを利用した Turney の手法は，教師付きデータや学習アルゴリズムが不要であるという点において初期導入コストが低いが，実際に適用する場合は，新しい複数語表現が見つかるごとにウェブにアクセスする必要があり，運用コストは高い．一方，機械学習や人手によるアプローチは，教師付きデータや規則作成などが必要という点において初期導入コストは高いが，運用コストは比較的低い．4 番目の列は分類に使用した主な情報を表し，5 番目の列はニュートラルの分類が可能かどうかを表す．○ならば可能，×ならば不可能を表す．鈴木らの手法が△なのは，モデルとしては分類可能であるが，評価データのニュートラルな事例が少数であり，ニュートラルの分類に関しては性能が不明だからである．右端の列は，未出現語（訓練データや作成した規則に出現しない単語）から成る複数語表現を分類できるかどうかを表す．Turney の手法は，ウェブの検索エンジンを使っているので，分類可能（○）とした．Baron らの手法はウェブのような大規模なコーパスを利用できた場合に可能となるので，条件付きで分類可能（△）としておいた．鈴木らの手法は，ウェブのような大規模なコーパスを学習することは非現実的であるので分類不可能（×）とした．我々の手法は，機械学習を用いて，構成語の属性をクラスタという形で抽出して確率モデルを構築するものであり，ニュートラル事例の分類にも対応できる．しかし，未出現語から成る複数語表現は分類不可能であるという弱点を持っている．

3. 複数語表現の感情極性分類のための隠れ変数モデル

1 章で述べたように，複数語表現の感情極性は，その構成語の極性の単なる和ではない．複数語表現の感情極性は，より複雑な計算により決定されると考えられる．たとえば，“リスクが低い” がポジティブである

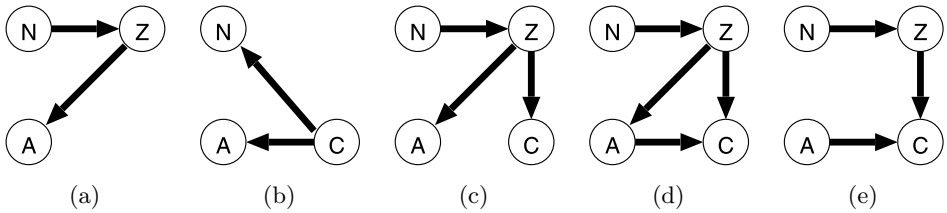

図 1 モデルのグラフ表現：(a) PLSI，(b) ナイーブベイズ，(c) 3-PLSI，(d) 三角形，(e) U 字型；ノードは確率変数を表す．有向線分は確率変数間の統計的依存性を表す．N，A，Z，C はそれぞれ名詞，述語，隠れ変数，感情極性に対応する

Fig. 1 Graphical representations of the models: (a) PLSI, (b) naive bayes, (c) 3-PLSI, (d) triangle, (e) U-shaped; Each node indicates a random variable. Arrows indicate statistical dependency between variables. N, A, Z and C respectively correspond to nouns, adjectives, latent clusters and semantic orientations.

こと，および “感染率” が “リスク” と（ある種の）同じ意味クラスタに属していることを我々が知っているとしよう．このとき我々は “感染率が低い” がポジティブであると推論することができる．それゆえ，我々は隠れ変数モデルを用いて，そのような隠れた意味クラスタをとらえ，複数語表現の高精度な分類を実現する（本稿では 2 つの単語から成る表現を考える）．提案モデルでは，量が減ることによりポジティブとなるような名詞（たとえば，“リスク”，“感染率”）が，クラスタを形成することが期待される．我々が本稿で扱うモデルのうちいくつかは，Hofmann[6] により協調フィルタリングに適用されている．

図 1 は，本稿で扱うモデルにおける確率変数間の統計的依存関係をグラフで表現したものである．図中，確率変数 N，A，Z，C はそれぞれ名詞，述語（形容詞もしくは形容動詞），隠れ変数，感情極性に対応する．図 1 (a) は，PLSI モデルを表しているが，これは感情極性に対応する確率変数を持たないので今回のタスクには利用できない．ここでは参考のために載せた．図 1 (b) は，ナイーブベイズモデルを表す．図 1 (c) は，PLSI を観測変数 3 つの場合に拡張したもので，本稿ではこれを 3-PLSI モデルと呼ぶことにする．ナイーブベイズモデルと 3-PLSI モデルの 2 つは，以下で説明する提案モデルに対する比較用のモデルである．図 1 (d) は，確率変数が三角形を形成していることから，本稿では三角形モデルと呼ぶことにする．三角形モデルにおいては，Z と N，Z と A がそれぞれ直接連結しているので，名詞と述語の両方に対してクラスタが形成される．また，図 1 (e) は，U 字型モデルと呼ぶことにする．一般的に，述語と比較して名詞は非常に種類が多いので，ここでは N と Z を連結することにより，名詞に対してクラスタが形成されるようなモデルのみを考える．三角形モデルと U 字型モデルでは，確率 $P(c|az)$ や $P(c|a)$ を通して，述語が感情極性に直接影響を与える．これら以外に，複数の隠れ変数を用いるようなモデルも考えられるが，隠れ変数の個数を増やすと，可能な隠れ変数の状態数が組合せ的に増大し，適切な状態数の予測に多大な計算量がかかることになる．よって，ここでは上記のシンプルなモデルのみを考えることにする．

我々は，図 1 の各モデルを複数語表現の感情極性分類に適用する．ここでは特に，我々が本タスクへの適用を提案する三角形モデルと U 字型モデルについて詳しく説明する．

3.1 三角形モデル

D を，名詞 n と述語 a とその極性 c のタプルの集合であるとする：

$$D = \{(n_1, a_1, c_1), \cdots, (n_{|D|}, a_{|D|}, c_{|D|})\}, \quad (1)$$

ここで，$c \in$ { ネガティブ，ニュートラル，ポジティブ } とする．本稿ではこのように極性の状態数として 3 値を考えるが，$c \in$ { 非常にネガティブ，ネガティブ，ニュートラル，ポジティブ，非常にポジティブ } のようなより細かい分類も考えることができる．我々の目的は未知のペア n と a に対して，その極性 c を予測することである．

図 1 (d) より，n，a，c，z の生成確率は，

$$P(nacz) = P(z|n)P(a|z)P(c|az)P(n) \quad (2)$$

となる．ただし，紙面の節約のため，確率変数間のカンマを省いて表記する．

モデル推定には，Expectation-Maximization（EM）アルゴリズム[2] を用いる．Q 関数（隠れ変数の事後確率に関する完全データの対数尤度の期待値）は

$$Q(\theta) = \sum_{nac} f_{nac} \sum_{z} \bar{P}(z|nac) \log P(nazc|\theta) \quad (3)$$

と表される．ここで，θ はパラメータの集合を表し，

f_{nac} はタプル $<n, a, c>$ の訓練データ中での頻度を表す．$\bar{P}$ は，更新前のパラメータを用いて計算された確率値であることを示す．

E ステップ（expectation ステップ）は，単純な事後確率の計算に帰着する：

$$\bar{P}(z|nac) = \frac{P(z|n)P(a|z)P(c|az)}{\sum_z P(z|n)P(a|z)P(c|az)}. \quad (4)$$

M ステップ（maximization ステップ）における更新式の導出には，ラグランジュの未定乗数法が用いられる．ただし，これは制約付き（$\sum_z P(z) = 1$, $\forall z$, $\sum_n P(n|z) = 1$, $\forall z$, $\sum_a P(a|z) = 1$, $\forall a, z, \sum_c P(c|az) = 1$）の最適化問題であることに注意されたい．よって，以下の更新式を得る：

$$P(n) = \frac{\sum_{ac} f_{nac}}{\sum_n \sum_{ac} f_{nac}}, \quad (5)$$

$$P(z|n) = \frac{\sum_{ac} f_{nac}\bar{P}(z|nac)}{\sum_{ac} f_{nac}}, \quad (6)$$

$$P(a|z) = \frac{\sum_{nc} f_{nac}\bar{P}(z|nac)}{\sum_{nac} f_{nac}\bar{P}(z|nac)}, \quad (7)$$

$$P(c|az) = \frac{\sum_{n} f_{nac}\bar{P}(z|nac)}{\sum_{nc} f_{nac}\bar{P}(z|nac)}. \quad (8)$$

この 2 つのステップは収束するまで交代しつつ繰り返される．Q 関数の変化が十分に小さくなったときに，収束したと見なされる．

極性が未知の単語ペア n，a に対し，確率値

$$P(c|na) = \frac{\sum_z P(z|n)P(a|z)P(c|az)}{\sum_{cz} P(z|n)P(a|z)P(c|az)} \quad (9)$$

を計算し，この値が最大になるような c を，求める極性の予測値として出力する．

3.2 U 字型モデル

U 字型モデル（図 1 (e)）においては，n，a が与えられたときの c，z の条件付き確率は，

$$P(cz|na) = P(c|az)P(z|n) \quad (10)$$

と表される．

三角形モデルの場合と同様に，EM アルゴリズムを用いてモデル推定を行う．Q 関数（式 (3)）に対しラグランジュの未定乗数法などを用いることで，以下の EM ステップが得られる：

E step

$$\bar{P}(z|nac) = \frac{P(c|az)P(z|n)}{\sum_z P(c|az)P(z|n)}, \quad (11)$$

M step

$$P(c|az) = \frac{\sum_n f_{nac}\bar{P}(z|nac)}{\sum_{nc} f_{nac}\bar{P}(z|nac)}, \quad (12)$$

$$P(z|n) = \frac{\sum_{ac} f_{nac}\bar{P}(z|nac)}{\sum_{ac} f_{nac}}. \quad (13)$$

分類には，次の式を用いればよい：

$$P(c|na) = \sum_z P(c|az)P(z|n). \quad (14)$$

3.3 比較のためのその他のモデル

図 1 (c) に対応する 3-PLSI モデルも考える．n，a，c，z の生成確率は，

$$P(nacz) = P(z|n)P(a|z)P(c|z)P(n) \quad (15)$$

となる．EM アルゴリズムの更新式は，三角形モデルの更新式において $P(c|az)$ の代わりに

$$P(c|z) = \frac{\sum_{na} f_{nac}\bar{P}(z|nac)}{\sum_{nac} f_{nac}\bar{P}(z|nac)}.$$

を用いることで得られる．

隠れ変数モデルに加え，次のような単純な確率モデルを用いたベースライン分類器を用意しておく：

$$P(c|na) \propto P(n|c)P(a|c)P(c). \quad (16)$$

このベースライン分類器は，素性が 2 つのナイーブベイズ分類器[13] と等価である．ベースラインモデルのグラフ表示は図 1 (b) である．パラメータは，

$$P(n|c) = \frac{1 + f_{nc}}{|N| + f_c}, \quad (17)$$

$$P(a|c) = \frac{1 + f_{ac}}{|A| + f_c} \quad (18)$$

と推定すればよい．ここで $|N|$ と $|A|$ はそれぞれ n と a に対応する単語の種類数である．これは，ナイーブベイズ分類器でよく使用される推定方法で，ディリクレ分布を事前分布とした事後確率最大化推定[12] である．

結局我々は，ベースラインモデル，3-PLSI モデル，三角形モデル，U 字型モデルの 4 つのモデルを用意したことになる．

3.4 モデルや計算についての考察

実際の EM の計算では，通常の EM アルゴリズムでなく，tempered EM アルゴリズム[5] を用いる．このアルゴリズムでは，正の値をとるハイパーパラメータ β が導入される．この値を調整することにより，計算途中の隠れ変数の事後確率値をどの程度信頼するかを調整することができる．具体的には，この値が小さいほど，計算途中の隠れ変数の事後確率値を信用しないことになる[6]．通常の EM アルゴリズムの E ステップにわずかな変更を加えるだけで，tempered EM アルゴリズムが実現できる．たとえば，U 字モデルの場合は，

$$\bar{P}(z|nac) = \frac{\big(P(c|az)P(z|n)\big)^{\beta}}{\sum_z \big(P(c|az)P(z|n)\big)^{\beta}}, \quad (19)$$

となる．他のモデルに関しても同様に tempered EM アルゴリズムが導出できる．

また，隠れ変数の可能な状態数を M で表すことにする．結局我々は，β と M の 2 つのハイパーパラメータを決定する必要がある．これらの値の決定は，ヘルドアウト法を用いて行う．すなわち，与えられた訓練データのうち 90%を一時的な訓練データとして学習を行い，残りの 10%を一時的なテストデータとして評価を行う．これを様々な β と M の組について行い，最も正解率が高かったハイパーパラメータの組を選ぶ．選ばれたハイパーパラメータを用いて改めて訓練データ全体で学習を行うことにより，確率モデルを求める．

また，ある確率変数（たとえば Z）に N と A から同時に有向線分が入ってくるようなモデルは，現実的でない．これは，$P(z|na)$ という $M|N||A|$ のオーダのパラメータを推定する必要が出てくるからである．

極性の間には，実は計量が導入されるべきである．つまり，ネガティブとポジティブの違いは，ネガティブとニュートラルの違いよりも大きいと考えるのが自然であろう．しかし，ここまで説明してきたようなモデルでは，異なる極性間の計量は考慮されていない．Hofmann[6] は，協調フィルタリングにおいて極性間に計量を持たせるためにパラメータ $P(c|az)$ を一次元ガウス分布でモデル化した．しかし，我々はここではガウス分布を導入しない．なぜなら，我々のデータセットでは c はポジティブ，ニュートラル，ネガティブのわずか 3 種類の値しかとりえないので，ガウス分布が真の分布の適切な近似にならないことが予想されるからである．実際，ガウス分布を用いて予備実験を行ったところ，モデルの予測性能はガウス分布を用いないモデルと比較して非常に悪かった．クラス変数 c がより多種の値をとりうるようなデータにおいては，ガウス分布によるモデル化が有効になるだろう．

次章で述べる本稿での実験では，極性が一定と考えられるような語（たとえば，“良い”，“悪い” など）を述語とする単語対も訓練データに含まれている．実際に応用においては，既存研究[4),8),9),17),20)] で得られた単語の感情極性を利用するなどして，極性が一定な語を述語とする単語対についてはその単語極性を用い，それ以外の単語対のみでモデルを構築するなどの工夫も可能である．これにより教師付きデータ作成の手間の軽減，計算速度の向上などが期待できる．

4. 実　　験

4.1 実験設定

まず，データセットについて述べる．毎日新聞記事[10)] から，主語となる名詞とその述語となる形容詞もしくは形容動詞の対を抽出し，各対にポジティブ，ニュートラル，ネガティブのいずれかの感情極性タグを付けた．得られたデータセットのサイズとその内訳を**表 2** に示す．名詞の種類数は 4,770 であり，形容詞もしくは形容動詞の種類数は 384 である．今回利用したデータに関し，正解タグ付け作業者間の一致度について述べておく．2 人の作業者間の κ 値は 0.640 だった．この値は高い値ではないが，許容範囲であるといえる．実は，ポジティブとネガティブが入れ替わっているような不一致は，データのわずか 0.7%であった．このことは，ニュートラルを判定することの本質的な難しさを表している．実際，ニュートラルな表現を分類するような研究はこれまでほとんどなされていない．

評価には 10 分割の交差検定を用い，その平均正解率を算出した．ただし，訓練データとテストデータに同じ単語対が出現しないように分割した．

また，未出現語（訓練データに出現しない単語）については，そもそも手がかりとなる統計情報がない．よって，別のデータやリソースを使用しない限り，未出現語から成る対の分類は不可能であり，そのような分類問題は本稿で扱う範囲を越える．つまり，未出現語から成る対をテストデータに使用すると，正確な評価値が算出できない．このような理由から，テストデータ中に出現する対の 2 単語のうち少なくとも片方が未出現語であったら，その対は評価には使用しない．結局，対としては訓練データに入っていないが，各単語は訓練データに出現しているような対のみを評価に使用していることになる．

隠れ変数モデルの有効性がより明確に分かるように，名詞と組み合わせられたときの極性が一定でないと思われる 17 語の形容詞を述語として持つような対を元のデータセット（標準データセットと呼ぶ）から抜き出すことにより，新しいデータセットを作成した．17 語は以下のとおりである：

表 2 データセットのサイズ
Table 2 Statistics on the dataset.

	のべ	異なり
ポジティブ	3,355	2,074
ニュートラル	4,252	2,647
ネガティブ	4,459	2,695
合計	12,066	7,416

高い，低い，大きい，小さい，重い，軽い，強い，弱い，多い，少ない，ない，すごい，激しい，深い，浅い，長い，短い．

この新しいデータセットを，極性不定形容詞データセットと呼ぶことにする．これらの極性不定形容詞に対しては，極性出現や極性反転などの現象が見られやすいと思われるので，このデータセットに対する分類性能によって，複数語表現の性質がうまくとらえられたかどうかが分かる．極性不定形容詞データセットは4,787の異なる対を含み，標準データセットの部分集合となっている．極性不定形容詞データセットは，評価データとしてのみ使用した．訓練には，つねに標準データセットを用いた．

ハイパーパラメータ β の値としては，0.1, 0.2, ···, 1.0を試した．また，ハイパーパラメータ M の値としては，2, 3, 5, 7, 10, 20, 30, 40, 50, 70, 100, 200, 300, 500を試した．適切なハイパーパラメータの値を予測する場合は，これらの値の組の中から，3.4節で述べたヘルドアウト法を用いて最も高い予測正解率を出す組を選んだ．また，EMアルゴリズムの計算におけるパラメータの初期値は乱数を用いて決定した．

4.2 結　　果

表3 に，ヘルドアウト法で決定した β と M を用いたときの4手法の分類正解率を示す．ただし，ベースライン分類器については，β と M は関係ない．また，β と M は，交差検定の各分割で異なる値が予測されうるので，表中の β と M の数値は10分割の交差検定の結果を平均したものである．この表から分かるように，三角形モデルとU字型モデルに関しては80%を上回る正解率が得られ，他と比較して良い性能を示している．また，極性不定形容詞データセットに対しても，提案手法は他と比較して良い性能を示している．この結果は，隠れ変数を通して複数語表現の感情極性の非構成性をとらえることに成功したことを示唆している．実験設定や対象言語が異なるので直接の比較はできないが，参考までにこれまでに報告された正解率をあげると，Wilsonら[21]による複数語表現の感情極性判定の正解率は65.7%であった．また，単語の感情極性判定の正解率は，実験設定に依存するが，ポジティブもしくはネガティブの二値分類問題において80から90%程度の値が報告されている．

3-PLSIモデルはうまく働かなかった．Hofmann[6]は，協調フィルタリングには3-PLSIモデルは制限が強過ぎる（モデルの自由度が低過ぎる）としており，複数語表現の感情極性判定タスクにおいても同様のことがいえることが実験的に示された．

表3 予測された β および M を用いたときの分類正解率
Table 3 Accuracies with predicted β and M.

	標準			極性不定形容詞		
	正解率	β	M	正解率	β	M
ベースライン	73.40	—	—	65.93	—	—
3-PLSI	67.02	0.73	91.7	60.51	0.80	87.4
三角形	81.39	0.60	174.0	77.95	0.60	191.0
U字型	81.94	0.64	60.0	75.86	0.65	48.3

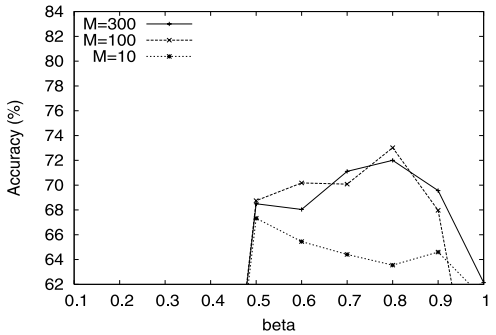

図2 3-PLSIモデル，標準データセット
Fig. 2 3-PLSI model with standard dataset.

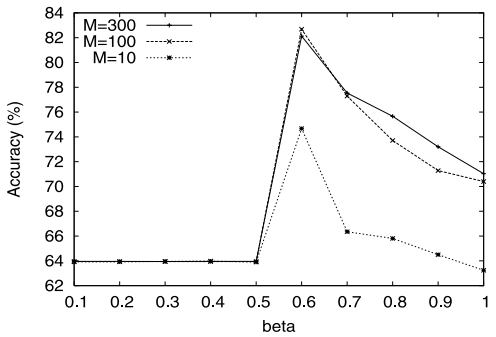

図3 三角形モデル，標準データセット
Fig. 3 Triangle model with standard dataset.

次に，ハイパーパラメータの値の影響を見る．**図2**, **図3**, **図4** は，それぞれ3-PLSIモデル，三角形モデル，U字型モデルを用いた場合の，交差検定による平均正解率の β に対する変化を，いくつかの M についてプロットしたものである．つまり，ここではハイパーパラメータの予測は行われていない．図から分かるように，分類性能は β の値に大きく影響を受けている．大きめの M の値（$M = 100$, $M = 300$）の方が，小さめの M の値より良い結果を出している．しかし，これは分類性能と学習時間とのトレードオフであり，M が大きくなれば学習に多大なコストがかかる．そのような観点から，三角形モデルと比較して，U字型モデルは小さな M（$M = 10$）でも良い分類性能を示しており，実際の応用に有用であると考えら

表 4 予測された β と M の標準偏差および最適値との差の平均値
Table 4 Standard deviations of predicted β and M.

	標準				極性不定形容詞			
	標準偏差		最適値との差		標準偏差		最適値との差	
	β	M	β	M	β	M	β	M
3-PLSI	0.06	105.5	0.13	122.9	0.11	108.5	0.13	156.6
三角形	0.00	142.6	0.00	206.0	0.00	134.3	0.00	213.3
U 字型	0.12	56.2	0.05	60.0	0.12	57.8	0.09	41.7

表 5 予測された β と M を用いたときの U 字型モデルによる分類結果の分割表
Table 5 Confusion matrix of classification result by the U-shaped model with predicted β and M.

		U 字型モデル			
		ポジティブ	ニュートラル	ネガティブ	合計
正解	ポジティブ	1,856	281	69	2,206
	ニュートラル	292	2,021	394	2,707
	ネガティブ	102	321	2,335	2,758
	合計	2,250	2,623	2,798	7,671

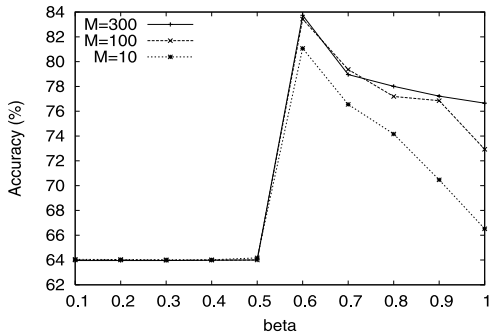

図 4 U 字型モデル，標準データセット
Fig. 4 U-shaped model with standard dataset.

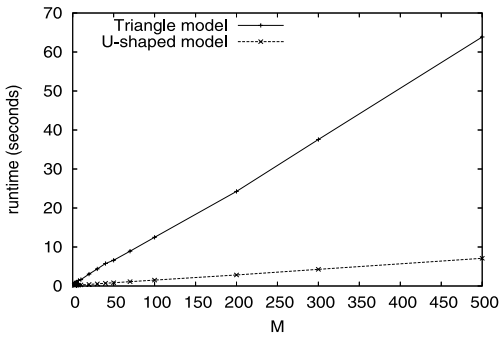

図 5 三角形モデル（triangle model）と U 字型モデル（U-shaped model）における隠れ変数の状態数と学習に要した時間との関係（β は予測平均値に近い 0.6 に固定した）
Fig. 5 Number of states of the latent variable versus time required for training for the triangle model and the U-shaped model (the value of β was set to 0.6).

れる．実際，三角形モデルと U 字型モデルについて，隠れ変数の状態数と学習に要する時間の関係をグラフにすると，**図 5** のようになる．学習時間を算出するにあたっては，12,066 事例すべてを訓練に用いた．今回の実験設定では状態数を 500 などとしても学習時間は約 1 分であり，学習時間に関する性質が高い重要性を持っているとはいえない．しかし，実際の応用などでさらに大規模なデータを扱う場合には，学習時間が非常に大きくなることも考えられ，学習時間を考慮することも必要であろう．

また，ハイパーパラメータの予測についてより詳しく調べるために，予測されたハイパーパラメータの値の標準偏差，および最適値との差の平均値を**表 4** に示す．この表から分かるように，β に関しては予測値に大きな変動はないが，M の方は交差検定の各分割において大きく異なる値が予測され，また最適値との差も大きい．つまり，より高精度な最適状態数予測手法が使用できれば，分類性能はさらに上がるものと思われる．

一般に EM アルゴリズムの性能は初期値に依存する．初期値を変えても実験結果が大きく変わらないことを示すため，乱数を用いて異なる初期値で 20 回計算を行った．モデルとしては，U 字型モデルを用い，ハイパーパラメータの値は，$\beta = 0.6$ と $M = 100$ とした．その 20 回中の最大正解率と最小正解率の差は 1.5 ポイントであり，結論には影響を与えないと考えられる．

さらに，全体的なエラーの傾向を見るために，予測されたハイパーパラメータを用いたときの U 字型モデルでの分類結果の分割表を**表 5** に示す．ただし，この表内の数値は，実際に評価に使われた事例に対するも

のである．つまり，表 1 に示した 12,066 事例中 7,671 事例が評価に使用されたことになる．この表から分かるように，エラーのほとんどはニュートラルを適切に分別できなかったものであり，ポジティブをネガティブに，あるいは逆にネガティブをポジティブに間違えて予測した例は全体の 2.23%にすぎない．つまり，提案モデルは，極性を逆に予測してしまうような大きな間違いをすることは非常に少ないことが分かる．

次に，極性を逆に予測してしまったような少数の例を観察し，簡単にエラー分析をしてみる．

“食品＋高い” のように，実際には “食品の価格が高い” ことを意味しているが，“価格” の部分が解釈されていないと思われるものが多く見られた．このような事例に対しては，たとえば対象と属性を前処理で正確に特定するような枠組が必要である．たとえば，Popescu ら[15] などは，そのような方向性の研究を行っている．

我々は，隠れ変数を導入することによりデータスパースネス問題を軽減したが，この問題は依然として存在する．たとえば，“手詰まり感＋色濃い” のように，低頻度語に対して判定を誤る例があった．これらはデータを大きくすることで対応できると思われる．極性ラベル付きデータの準備が困難な場合は，半教師付き学習によりラベルなしデータを有効に利用する必要があるだろう．

4.3 得られたクラスタの例

定性的に結果を見るために，得られたいくつかのクラスタ z に対して，名詞 n を $P(z|n)$ の値の降順でソートし，上位 50 語に入っている名詞 n のうちデータセット中で 3 回以上出現しているようなものを示す．ここでは例として，$\beta = 0.6$，$M = 60$ なる設定の下で U 字型モデルが算出したクラスタ群からの抜粋を紹介する．

クラスタ 1	トラブル，反対意見，病気，苦情，心配，既往症
クラスタ 2	リスク，死亡率，感染率，発症率
クラスタ 3	縁，意見，愛着，意味合い，あこがれ，意志
クラスタ 4	得票，応募，話題，支持者
クラスタ 5	弊害，悪化，ショック，衝撃，負担
クラスタ 6	悪化，差別，負荷，弊害
クラスタ 7	比重，影響度，数字，ウェート，帰属意識，波，呼び声

クラスタの例を見ると分かるように，人間の直観に合ったモデルが得られている．たとえばクラスタ 2 には，“高い” と対になってネガティブになり，“低い” と対になってポジティブになるような名詞が集まっている．実際，クラスタ 2 に対して，感情極性の事後確率値を計算してみると，

$$P(C = \text{ネガティブ} \,|A = \text{高い}, Z = \text{クラスタ 2}) = 0.995,$$
$$P(C = \text{ポジティブ} \,|A = \text{低い}, Z = \text{クラスタ 2}) = 0.973$$

である．単純な共起情報に基づいたクラスタリングでは，クラスタ 2 に “成功率” のような極性が逆になるようなものが含まれてしまうことが多い．極性クラス c をモデルに組み込んだ結果，このような感情極性判定という目的に合致したクラスタを獲得することができたといえる．

5. 結　　論

複数語から成る評価表現のモデルおよびそれに基づいた分類手法を提案した．複数語から成る評価表現の特質を考慮し，モデルに隠れ変数を導入した．実験により，提案した隠れ変数モデルは複数語から成る評価表現分類において，正解率で 82%という高い性能を持つことを示した．今回の実験では日本語のデータを用いたが，手法自体は言語非依存であり，汎用性を持っている．

今後の発展としては，まず訓練データにおける低頻度語や未出現語への対応があげられる．4.2 節のエラー解析でも述べたが，半教師付き学習の利用で適切に対応できる可能性がある．また，3 単語以上から成る表現への本手法の適用がある．モデルとしては容易に拡張可能であるが，分類器としての有効性は調査する必要がある．また，Fujita ら[3] が隠れ変数を素性として k–近傍法を利用したように，我々のモデルが抽出した隠れ変数を他の分類器の素性として用いることもできる．また，他の研究から得られた単語の感情極性との融合という課題もある．そのような異なるレベルの知見を融合することにより，より高性能なモデルが構築できるだろう．

参 考 文 献

1) Baron, F. and Hirst, G.: Collocations as cues to semantic orientation, *AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications* (2004).
2) Dempster, A.P., Laird, N.M. and Rubin, D.B.: 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).
3) Fujita, A., Inui, K. and Matsumoto, Y.: Detection of incorrect case assignments in automatically generated paraphrases of Japanese sentences, *Proc. 1st International Joint Conference on Natural Language Processing* (*IJCNLP*), pp.14–21 (2004).
4) Hatzivassiloglou, V. and McKeown, K.R.: Predicting the semantic orientation of adjectives, *Proc. 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Chapter of the Association for Computational Linguistics*, pp.174–181 (1997).
5) Hofmann, T.: Unsupervised learning by probabilistic latent semantic analysis, *Machine Learning*, Vol.42, pp.177–196 (2001).
6) Hofmann, T.: Latent semantic models for collaborative filtering, *ACM Trans. Information Systems*, Vol.22, pp.89–115 (2004).
7) Inui, T.: Acquiring Causal Knowledge from Text Using Connective Markers, Ph.D. thesis, Graduate School of Information Science, Nara Institute of Science and Technology (2004).
8) Kamps, J., Marx, M., Mokken, R.J. and de Rijke, M.: Using wordnet to measure semantic orientation of adjectives, *Proc. 4th International Conference on Language Resources and Evaluation* (*LREC 2004*), Vol.IV, pp.1115–1118 (2004).
9) Kobayashi, N., Inui, T. and Inui, K.: Dictionary-based acquisition of the lexical knowledge for p/n analysis (in Japanese), *Proc. Japanese Society for Artificial Intelligence, SLUD-33*, pp.45–50 (2001).
10) Mainichi: Mainichi Shimbun CD-ROM version (1995).
11) Matsumoto, S., Takamura, H. and Okumura, M.: Sentiment classification using word subsequences and dependency sub-trees, *Proc. 9th Pacific-Asia Conference on Knowledge Discovery and Data Mining* (*PAKDD-05*), pp.301–310 (2005).
12) McCallum, A. and Nigam, K.: A comparison of event models for naive bayes text classification, *Proc. AAAI-98 Workshop on Learning for Text Categorization*, pp.41–48 (1998).
13) Mitchell, T.M.: *Machine Learning*, McGraw Hill (1997).
14) Pang, B., Lee, L. and Vaithyanathan, S.: Thumbs up? sentiment classification using machine learning techniques, *Proc. Conference on Empirical Methods in Natural Language Processing* (*EMNLP'02*), pp.79–86 (2002).
15) Popescu, A.-M. and Etzioni, O.: Extracting product features and opinions from reviews, *Proc. joint conference on Human Language Technology/Conference on Empirical Methods in Natural Language Processing* (*HLT/EMNLP'05*), pp.339–346 (2005).
16) Smadja, F.Z.: Retrieving collocations from text: Xtract, *Computational Linguistics*, Vol.19, No.1, pp.143–177 (1993).
17) Takamura, H., Inui, T. and Okumura, M.: Extracting semantic orientations of words using spin model, *Proc. 43rd Annual Meeting of the Association for Computational Linguistics* (*ACL'05*), pp.133–140 (2005).
18) Torisawa, K.: An unsuperveised method for canonicalization of Japanese postpositions, *Proc. 6th Natural Language Processing Pacific Rim Symposium* (*NLPRS 2001*), pp.211–218 (2001).
19) Turney, P.D.: Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, *Proc. 40th Annual Meeting of the Association for Computational Linguistics* (*ACL'02*), pp.417–424 (2002).
20) Turney, P.D. and Littman, M.L.: Measuring praise and criticism: Inference of semantic orientation from association, *ACM Trans. Information Systems*, Vol.21, No.4, pp.315–346 (2003).
21) Wilson, T., Wiebe, J. and Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis, *Proc. joint conference on Human Language Technology/Conference on Empirical Methods in Natural Language Processing* (*HLT/EMNLP'05*), pp.347–354 (2005).
22) 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，第 3 回情報科学技術フォーラム (FIT 2004) 講演論文集，第 2 分冊，F-006, pp.213–214 (2004).
23) 鈴木泰裕，高村大也，奥村 学：Semi-Supervised な学習手法による評価表現分類，言語処理学会第 11 回年次大会，pp.668–671 (2005).

(平成 17 年 11 月 22 日受付)
(平成 18 年 9 月 14 日採録)

高村 大也（正会員）

1974 年生．1997 年東京大学工学部計数工学科卒業．2000 年同大学大学院工学系研究科計数工学専攻修了（1999 年はオーストリアウィーン工科大学にて研究）．2003 年奈良先端科学技術大学院大学情報科学研究科博士課程修了．博士（工学）．2003 年より東京工業大学精密工学研究所助手．自然言語処理，特に学習理論等の応用に興味を持つ．ACL 会員．

乾 孝司（正会員）

1976 年生．1999 年九州工業大学情報工学部卒業，2001 年同大学大学院情報工学研究科修士課程修了，2004 年奈良先端科学技術大学院大学情報科学研究科博士課程修了．同年東京工業大学 21 世紀 COE ポスドク研究員，2005 年日本学術振興会特別研究員，2006 年東京工業大学統合研究院助手，現在に至る．博士（工学）．主に自然言語処理の研究に従事．言語処理学会，ACL 各会員．

奥村 学（正会員）

1962 年生．1989 年東京工業大学大学院情報理工学研究科計算工学専攻博士後期課程修了．1989 年より東京工業大学大学院情報理工学研究科助手．1992 年より 2000 年北陸先端科学技術大学院大学助教授．1997 年より 1998 年トロント大学客員助教授．2000 年より東京工業大学精密工学研究所助教授．自然言語処理，自動テキスト要約，コンピュータによる語学学習支援，テキストデータマイニングに関する研究に従事．工学博士．AAAI，ACL，JSAI，JCSS 各会員．