

スピンモデルによる単語の感情極性抽出

高村 大也[†] 乾 孝司[†] 奥村 学[†]

単語の感情極性（望ましいか否か）を判定する方法を提案する．提案手法では，単語の感情極性を電子のスピン方向と見なし，語釈文，シソーラス，コーパスによって構築された語彙ネットワークをスピン系でモデル化する．平均場近似を利用してスピン系の状態を近似的に求めることにより，単語の感情極性を判定する．また，系の状態に影響を与えるハイパーパラメータの予測方法も同時に提案する．提案手法を用いて WordNet に収録されている語彙に対して実験を行い，14 語という少数の単語を種とした場合は約 80% の正解率で，3,000 語あまりの単語を種とした場合は約 90% の正解率で単語の感情極性判定が実現できることを示した．

Extracting Semantic Orientations Using Spin Model

HIROYA TAKAMURA,[†] TAKASHI INUI[†] and MANABU OKUMURA[†]

We propose a method for extracting semantic orientations of words: desirable or undesirable. We construct a lexical network out of glosses in a dictionary, a thesaurus and a corpus. Regarding semantic orientations of words on the network as spins of electrons, we use the mean field approximation to compute the approximate probability function of the system instead of the intractable actual probability function and determine the semantic orientations of words. We also propose a criterion for hyper-parameter selection. We conducted experiments with English lexicon in WordNet and obtained approximately 80% accuracy with 14 seed words and approximately 90% accuracy with about 3,000 seed words.

1. 序 論

文書内に含まれる感情（意見や態度を含む）の発見，特定，および分析は，様々な応用可能性を有する重要な課題である．たとえば，電子掲示板の意見が自動的に発見できれば，インターネットから新製品に関する意見を効率的に収集することができる．また，アンケートのすべての回答に目を通すことなく，回答群全体の傾向をつかむことが可能になる．

このような感情分析における重要なリソースの 1 つとして，単語の感情極性があげられる．単語の感情極性とは，ある単語が良い印象を持つ（ポジティブ）か，それとも悪い印象を持つ（ネガティブ）かを示す二値変数である．たとえば，“美しい”はポジティブであり，“汚い”はネガティブである．単語の感情極性は，その後の感情・意見分析で基礎的な役割を果たす．たとえば，与えられた文書中でのポジティブな単語の出現

数とネガティブな単語の出現割合によって文書そのものの感情極性が推定できるという報告がある^{4),9),25)}．立石ら²⁴⁾は意見要約に単語の感情極性を利用している．文書を読者に提示する際に，感情極性を持つ単語をハイライトすることにより，文書の感情極性が直観的に一目で分かるようなインタフェースを提供している実際のサービスなどもある²³⁾．

本稿の目的は，辞書の語釈文，シソーラス，コーパスなどから，単語の感情極性を自動的に抽出する手法を提案することである．そのために我々は，スピンを持つ複数の電子の系であるスピン系のモデルを利用する．各電子がスピンと呼ばれる方向（上向きあるいは下向きをとる）を持つように，各単語は感情極性と呼ばれる方向（ポジティブあるいはネガティブ）を持つと考えられる．それゆえ我々は，単語をノードとする語彙ネットワークを構築し，これをスピン系でモデル化する．語彙ネットワークの構築においては，感情極性が等しいと思われる単語ペアを辞書の語釈文，シソーラス，コーパスから抽出して正の重みで連結し，同様に感情極性が異なると思われる単語ペアを負の重みで連結することにより，単語をノードとする語彙ネットワークを構築する．さらに，統計物理の分野で開発さ

[†] 東京工業大学
Tokyo Institute of Technology
現在，日本学術振興会特別研究員
Presently with research fellow of the Japan Society for
the Promotion of Science

れてきた平均場近似を利用して系の状態を計算し、その結果から単語の感情極性を判定する。これにより、語彙ネットワークにおける情報伝搬が実現できる。語釈文、シソーラス、コーパスから抽出された同極性/異極性の情報はつねに正しいわけではなくノイズを含んでいる。しかし、このように系全体を見ながら大域的に系の状態を推定することにより、衝突がなるべく少ない極性割当てが実現できると考えられる。このような性質を持たない既存手法では、一部の情報に偏って精度の低い極性割当てになってしまったり、信頼できる同極性/異極性の情報しか利用できなかつたりするので、カバレッジが低くなってしまふなどの問題点があった。また、系全体のスピンの方向に関する傾向を表す平均磁化という概念を用いて、系の確率分布のハイパーパラメータを決定する手法も提案する。

実験により、提案手法は少数の単語の感情極性を種として与えるだけで、大量の単語の感情極性を決定できることが分かった。

2. 関連研究

Turney ら²⁰⁾ は、単語の感情極性判定のための2種類のアルゴリズムを提案している。彼らの手法では、まず“word NEAR good”や“word NEAR bad”などの検索クエリにより、極性を調べたい単語 word とすでに極性が分かっている“good”や“bad”などの単語がインターネット上で近接して出現する回数を検索エンジンのヒット数という形で調べる。そのヒット数を用いて、ポジティブな語とのつながりの強さ、およびネガティブな語とのつながりの強さを計算し、それらの差を感情極性判定の指標としている。また、潜在意味解析⁵⁾を用いて“good”や“bad”などとの共起性の強さを測ることにより感情極性を判定する手法も提案している。実験の評価は、内容分析のために収集された General Inquirer¹⁸⁾ に載っている 3,596 単語に対して行われた。General Inquirer は、単語に対しその属性を記述したもので、本研究で用いる“Positive (望ましい)”や“Negative (望ましくない)”をはじめ、“Strong (意味的に強い)”や“Weak (意味的に弱い)”など、多くの属性が付与されている。

Hatzivassiloglou⁸⁾ は、“simple and well-received”や“simplistic but well-received”などの接続表現に注目した。前者は、この2単語が同じ感情極性を持っていることを示唆しているが、後者はこの2単語が異なる感情極性を持っていることを示唆している。彼らの手法では、まず、様々な接続表現を同極性か異極性かに分類し、その分類結果を用いて単語をクラスタリ

ングして2種類に分け、その片方をポジティブ、もう片方をネガティブとする。評価は彼ら自身が作成したデータセットで行われ、手法の性質上、形容詞のみを対象としている。

小林ら²⁶⁾ は、ブートストラッピング的な手法を用いて単語の感情極性を抽出する手法を提案している。人手で作成した52個のルールを語釈文に繰り返し適用し、極性を決定していく手法である。小林らはこのタスクに対し詳細な調査を行い、また語釈文を利用しているという点で我々の手法に大きな影響を与えている。しかし、小林らの手法は、ルールの適用条件が厳しく、また日本語特有の表現を利用しており、被覆率の低さと言語依存性という2つの欠点を持つ。

Kamps ら¹²⁾ は、WordNet⁶⁾の類義語ペアを連結することにより語彙ネットワークを構築し、“good”および“bad”から任意の形容詞へのネットワーク上での最短距離の差を用いて感情極性を決定している。彼らの手法の欠点としては、類義語データが必要とされること、反義語の情報を取り入れることが不可能であること、類義語ネットワークに依存していることで異なる品詞間で極性情報が伝達しないことなどがあげられる。Hu ら⁹⁾の手法は、Kamps らの最短距離に基づいた手法に類似している。Hu らの手法は、種となる初期単語集合が与えられ、その初期単語集合と隣接する単語を同極性として取り込んでいくことにより初期単語集合をブートストラッピング的に拡大していく。そのような意味では、小林ら²⁶⁾の手法にも類似しているといえる。

単語の感情極性とは少し離れるが関連している研究として、主観性に関する研究がある。主観的な単語は、多くの場合感情極性を持つと考えられる。Wiebe²²⁾は機械学習を用いて主観的な形容詞をコーパスから収集する手法を提案している。Riloff ら¹⁶⁾は、主観的な名詞を収集することに焦点を当てている。

本稿では後の評価実験において、提案手法と、Turney らの手法、Kamps らの手法、Hu らの手法との比較を行う。実験設定に関する情報が足りないことや言語依存性の問題から、小林らの手法との比較はできなかった。

それ以外の上記関連研究についてはその目的が異なるので、我々の手法との直接的な比較は行わない。

3. スピンモデルと平均場近似

スピンモデルと平均場近似について簡単に紹介する。これらは、統計物理や機械学習の分野でさかんに研究されてきている話題である^{7),11),13)}。

スピン系とは、複数の電子のなす系であり、各電子はスピンと呼ばれる方向を持っており、スピンは +1 (上向き) もしくは -1 (下向き) のどちらかの値をとる。隣り合った電子どうしはエネルギー的な理由により同じ方向を向きやすいことが知られている。このモデルはイジングスピンモデル、もしくは簡単にスピンモデルと呼ばれる²⁾。

スピン系のエネルギー関数は、

$$E(\mathbf{x}, W) = -\frac{1}{2} \sum_{ij} w_{ij} x_i x_j, \quad (1)$$

と表される。ここで、 x_i と x_j は、それぞれ電子 i, j のスピンを表し、 $\mathbf{x} = (x_1, \dots, x_N)$ である。 N は電子の個数である。また、行列 $W = \{w_{ij}\}$ は、2つの電子間のリンクの重みを表す。

スピン系では、変数ベクトル \mathbf{x} はボルツマン分布

$$P(\mathbf{x}|W) = \frac{\exp(-\beta E(\mathbf{x}, W))}{Z(W)} \quad (2)$$

に従うことが知られている。ここで、 $Z(W) = \sum_{\mathbf{x}} \exp(-\beta E(\mathbf{x}, W))$ は分配関数と呼ばれる正規化因子であり、 β は逆温度定数と呼ばれる定数である。この分布関数が示すように、エネルギーが高い状態が実現する確率は小さくなる。

確率分布関数がかかったにもかかわらず、系の様々な統計量を計算することはしばしば困難をとまなう。問題は分配関数の計算であり、これは各スピンの“+1”と“-1”の2状態をとりうるので 2^N 個の異なる状態について和をとる必要がある。

上記の問題に対し、ここでは平均場近似と呼ばれる手法を用い、 $P(\mathbf{x}|W)$ をより簡単な関数 $Q(\mathbf{x}; \theta)$ で近似することを考える。関数 Q のパラメータ集合 θ は、 $Q(\mathbf{x}; \theta)$ がなるべく $P(\mathbf{x}|W)$ に近くなるように決定される。 P と Q の距離を測る指標として、 Q に関する平均エネルギーと Q のエントロピーの差として定義される変分自由エネルギー F を使用する：

$$F(\theta) = \beta \sum_{\mathbf{x}} Q(\mathbf{x}; \theta) E(\mathbf{x}; W) - \left(- \sum_{\mathbf{x}} Q(\mathbf{x}; \theta) \log Q(\mathbf{x}; \theta) \right). \quad (3)$$

この変分自由エネルギーを最小にするパラメータ θ が選ばれる。また、この F を最小にすることは、 P と Q の間のカルバック・ライブラ距離を最小にすることと等価であることが示されている¹³⁾。

次に、 $Q(\mathbf{x}; \theta)$ が、

$$Q(\mathbf{x}; \theta) = \prod_i Q(x_i; \theta_i) \quad (4)$$

と、各 x_i の関数の積で表されるとする。簡単な代入と式変形により、次のような具体的な変分自由エネルギーの形

$$F(\theta) = -\frac{\beta}{2} \sum_{ij} w_{ij} \bar{x}_i \bar{x}_j - \sum_i \left(- \sum_{x_i} Q(x_i; \theta_i) \log Q(x_i; \theta_i) \right) \quad (5)$$

が得られる。ここで、 \bar{x}_i は x_i の平均値、

$$\bar{x}_i = \sum_{x_i} x_i Q(x_i; \theta_i) \quad (6)$$

である。

ラグランジュ乗数法を用いることにより、平均場方程式を得る：

$$\bar{x}_i = \frac{\sum_{x_i} x_i \exp(\beta x_i \sum_j w_{ij} \bar{x}_j)}{\sum_{x_i} \exp(\beta x_i \sum_j w_{ij} \bar{x}_j)}. \quad (7)$$

この方程式は、次の更新式を繰り返し適用することにより解くことができる：

$$\bar{x}_i^{new} = \frac{\sum_{x_i} x_i \exp(\beta x_i \sum_j w_{ij} \bar{x}_j^{old})}{\sum_{x_i} \exp(\beta x_i \sum_j w_{ij} \bar{x}_j^{old})}. \quad (8)$$

各スピンについて1つずつこの更新を行っていく。この場合、変分自由エネルギーは増加しないことが保証される。また、変分自由エネルギーの第1項は $-(\beta/2) \sum_{ij} w_{ij}$ で下界値が存在し、第2項はネガティブ・エントロピーであるから下界値が存在する。よって変分自由エネルギーは下界値が存在する。それゆえ、変分自由エネルギーの変化がある閾値より小さくなったら計算を終了するとしておけば、計算は収束する。

4. スピンモデルによる感情極性の抽出

ここで、単語の感情極性判定にスピンモデルを用いるという我々の提案手法の説明をする。

各電子のスピンは、上向きと下向きのうちどちらかの値をとり、隣り合ったスピンは同じ値をとりやすい。我々は、各単語を電子と見なし、単語の感情極性をスピンの向きと見なす。関連する単語ペアを連結することにより語彙ネットワークを構築し、これをスピン系と見なす。下でより詳細に説明するが、関連する単語ペアとしては、ある単語とその語釈文に出現する単語、シソーラスでの類義語ペア、反義語ペア、上位語下位語ペア、コーパス中で“and”などの接続詞で連結されて出現する形容詞ペアなどをここでは考える。このようなモデル化の背後には、そのような単語ペアは同じ感情極性を持ちやすいであろうという直観が存在する(ただ

し反義語ペアなどは異なる感情極性を持つであろう)。

以下、語彙ネットワークの構築方法、近似関数の計算方法、感情極性の抽出方法について順に説明していく。最後に、提案モデルの性質について理論的な考察を加える。

4.1 語彙ネットワークの構築

まず、ある単語とその単語の語釈文内の各単語を連結することにより語彙ネットワークを構築する。各リンクを、 SL (same-orientation links, 同極性リンク集合), DL (different-orientation links, 逆極性リンク集合) という2つのグループに分ける。これらはもちろん、必ず同極性、あるいは必ず逆極性という意味ではなく、そうなりやすいと予測されるという意味である。もし、少なくとも片方の単語がもう片方の単語の語釈文で否定語(英語ならば“not”など)の後に出現しているならば、両単語をつなぐリンクは DL に分類される。そうでなければ SL に分類される。

次にリンクに対する重み $W = (w_{ij})$ を次のように決める:

$$w_{ij} = \begin{cases} \frac{1}{\sqrt{d(i)d(j)}} & (l_{ij} \in SL) \\ -\frac{1}{\sqrt{d(i)d(j)}} & (l_{ij} \in DL) \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

ここで、 l_{ij} は単語 i と j の間のリンクを表し、 $d(i)$ は単語 i の次数、すなわち単語 i が連結している単語数を表す。連結されていない単語どうしは、重み0のリンクで連結されていると見なすことができる。このネットワークを *gloss network* (G) と呼ぶことにする。

我々は、ネットワーク G に加えて、シソーラスで与えられた類義語ペア、反義語ペア、上位語下位語ペアをリンクで連結することにより別のネットワーク *gloss-thesaurus network* (GT) を構築する。反義語のリンクは DL に属するとする。それ以外のリンクは SL に属するとする。

さらにコーパスから得られた情報を加える。2章で述べたように、Hatzivassiloglou ら⁸⁾ はコーパスに出現する接続表現を利用した。同様に我々も2つの形容詞が“and”による接続表現で出現したらそのリンクを SL として、また、“but”による接続表現で出現したら DL として加える。このようにしてできたネットワークを *gloss-thesaurus-corpus network* (GTC) と呼ぶことにする。

上記の語彙ネットワーク拡張はリンク数を増加させるので、より円滑な情報の伝達が期待される。またコーパスによる拡張は、リンク数を増加させるだけで

なく、将来的に未知語(辞書に載っていない語)に対する本手法の適用も可能にする。

本来ならば、リソース(語釈文、シソーラス、コーパス)の違いによるリンクの重みの違い、あるいは否定表現の否定の強さによる重みの違いなどを考慮すべきだが、そのような違いをモデルに取り込む効果的な方法は考案されていないので、今回はそのような重みの違いは考えない。

4.2 感情極性判定

小規模な初期単語集合が与えられているとする。つまり、少量の単語の感情極性がすでに分かっている状態であるとする。この少量のラベル付きデータを以下のようにエネルギー関数を書き換える形でスピン系に取り入れる。

式(2)において、 $\beta E(\mathbf{x}, W)$ の代わりに次の関数 $H(\beta, \mathbf{x}, W)$ を用いる:

$$H(\beta, \mathbf{x}, W) = \frac{-\beta}{2} \sum_{ij} w_{ij} x_i x_j + \alpha \sum_{i \in L} (x_i - a_i)^2. \quad (10)$$

ここで、 L は初期単語集合、 a_i は初期単語 i の感情極性(1または-1)、 α はラベル付きデータに対する重み(正の定数)を表す。つまり、 x_i ($i \in L$) が a_i と異なる場合、ペナルティが与えられる。

この関数 H を用いると、初期単語 x_i ($i \in L$) に対して次の更新式

$$\bar{x}_i^{new} = \frac{\sum_{x_i} x_i \exp(\beta x_i s_i^{old} - \alpha(x_i - a_i)^2)}{\sum_{x_i} \exp(\beta x_i s_i^{old} - \alpha(x_i - a_i)^2)} \quad (11)$$

が得られる。ここで、 $s_i^{old} = \sum_j w_{ij} \bar{x}_j^{old}$ である。また、 \bar{x}_i^{old} と \bar{x}_i^{new} は、それぞれ更新前と後の x_i の平均である。ここでの新しい更新式は、Inoue ら¹¹⁾ の画像修復方法に従っている。初期単語集合に属さない単語に関しては式(8)の更新式を用いる。

初期値としては、初期単語の \bar{x}_i については、そのラベルに従って a_i とする。それ以外は0と設定しておく。

更新の前後で変分自由エネルギーの変化が十分に小さくなったら、収束したものと見なして計算を終了する。最終的に平均値が正だった単語については、感情極性がポジティブと判定する。逆に負だった単語は、感情極性がネガティブであると判定する。3章での収束性の議論と同様の議論がここでも成り立ち、計算の

収束性は保証される。

4.3 ハイパーパラメータの推定

提案手法の性能は、ハイパーパラメータ β の値に大きく依存する。本手法を実際に使用可能なものにするためには、 β の値を適切に決定しなくてはならない。

そこでまず疑似 leave-one-out エラー率

$$\frac{1}{|L|} \sum_{i \in L} [a_i \bar{x}'_i] \quad (12)$$

を導入する。これは、収束後の各平均値を用いて簡単に計算することができる。ここで、 $[t]$ は t が負のときに 1 となり、それ以外の場合 0 をとるものとする。また、 \bar{x}'_i は、式 (8) のペナルティ項 $\alpha(\bar{x}_i - a_i)^2$ を無視して計算する。初期単語集合が大きい場合は、この値が信頼できる、すなわち評価データに対するエラー率に近いと考えられるので、そのような場合はこの疑似 leave-one-out エラー率を最小にするようなハイパーパラメータ β を選べばよい。

しかし、大きな初期単語集合が与えられていない場合、疑似 leave-one-out エラー率の値は信頼できない。このような場合は、平均磁化 m を用いる：

$$m = \frac{1}{N} \sum_i \bar{x}_i. \quad (13)$$

高温度においては、スピンの方向には秩序がない（常磁性相、 $m \approx 0$ ）。低温度においては、スピン系は秩序を持っており同方向を向こうとする（強磁性相、 $m \neq 0$ ）。統計物理学の世界では、ある温度において突然、常磁性相と強磁性相が入れ替わることが知られている（次章で述べる我々の実験においても、図 1 の magnetization の曲線が示すように、このような現象が起こっている）。これを相転位といい、そのときの温度を臨界点という。臨界点よりわずかに高温であるとき、スピン系は部分的に秩序を持つが、全体的には秩序的でない状態になる。

直観的に、語彙ネットワークにおいては、“good” と同極性リンクで連結された単語の多くはポジティブであり、“bad” と同極性リンクで連結された単語の多くはネガティブであると考えられる。よって、単語を正しく分類するような望ましい語彙ネットワークの状態は、局所的に秩序立っていると予想される。それゆえ、我々は臨界点よりわずかに高い値を用いる。具体的には、一定間隔で β を大きくしていき、 m がある閾値より大きくなる直前の β を最良の値であると予測する。

4.4 モデルについての考察

我々のモデルでは、平均場近似により語彙ネットワーク上の単語の感情極性が大域的に決定される。このよ

うな大域的な最適化を用いるからこそ、語釈文やコーパスのような、シソーラスと比べてノイズが含まれやすい（すなわち、隣り合っているも同じ極性を持たないことが起こりやすい）リソースを取り入れることが可能になるのである。最短距離を利用した手法や単純なブートストラッピングを利用した手法のような既存手法では、そのようなリソースを取り入れることはできない。

また我々のモデルでは、収束後のスピンの平均値によって感情極性を決定する。平均値の利用は、いくらかヒューリスティックな印象を与えるかもしれないが、実は周辺化事後確率最大化推定、いい換えると有限温度推定という理論的背景を基にしている^{10),15)}。平均値 \bar{x}_i は、周辺化事後確率を最大化する推定値であると見なすことができる。このような有限温度推定は、アニーリングを基にした推定手法とは根本的に異なる。アニーリングを基にした推定手法（ゼロ温度推定とも呼ばれる）では、逆温度定数 β を徐々に無限大まで大きくしていく（すなわち、温度を 0 に近づけていく）。これは、自然言語処理の他のタスクでもしばしば利用されており³⁾、事後確率最大化推定と等価である。単語の感情極性抽出において、有限温度推定がうまく働き、アニーリングがうまく働かない理由は、タスクの特徴とモデルの性質による。語彙の中には、感情極性という点で曖昧な単語、あるいは本質的に極性を持たない単語が多く存在する。現モデルは、アニーリングを用いた場合これらの単語についても極性をポジティブあるいはネガティブのどちらかに決定しようとするため、正しい極性割当てが実現できない。しかし、有限温度推定を用いると、曖昧な語や無極性な語の極性は、確率的な意味で曖昧なまま保たれる。いい換えると、ポジティブおよびネガティブのどちらの可能性も保持しようとする。それゆえ、有限温度推定がうまく働くのである。もし、曖昧な語や無極性な語をうまく取り入れられるようなモデルを将来的に考案することができれば、そのようなモデルに対してはアニーリング手法もうまく機能する可能性がある。提案手法もアニーリング手法も様々な β に関して計算することになるので、計算にかかるコストはほとんど変わらない。

また、モデル推定のアルゴリズムは単純な更新式の繰返し計算に帰着されたので、語彙ネットワークに対しても利用されてきた（たとえば語義の曖昧性解消²¹⁾）活性伝搬モデルとの関連性についても議論する必要がある。実際、提案手法は、二値のモデル（2-イジングモデル）を使用している限り、特殊な更新式を用いた活性伝搬モデルと見なすことができる。しかし、一般

的な概念である活性伝搬モデルと比較して、我々のモデル化にはいくつかの長所がある。特筆すべきは、理論的背景である。スピン系としてモデル化したため、我々は目的関数を持ち、その近似手法を持ち、それに基づいた更新式を持っている。それゆえ、モデル推定の良さを表す指標を持っており、またベータ近似¹⁹⁾などの別の近似手法を用いることも可能である。また、本稿でハイパーパラメータの推定に使うことを提案している平均磁化という概念もある。それ以外にも、統計物理の分野で発展してきた様々な手法を今後取り入れていくことができる。特に、活性伝搬モデルの枠組みではとらえきれない多値スピンへの拡張は容易である(Q-イジングモデル)。

また別の考察点としては、自然言語処理でさかんに利用されている最大エントロピー法¹⁾との関連性がある。我々のモデルは、エネルギー関数に関する制約の下での分布関数 $Q(x)$ のエントロピーの最大化という観点でもとらえることが可能である。

5. 実 験

提案手法の性能を、実験により示す。

語彙ネットワークの構築について説明する。語釈文データとしては、WordNet⁶⁾の語釈文を用いた。語釈文に対する品詞タグ付けと原形化には、TreeTagger¹⁷⁾を用いた。動詞、名詞、形容詞、副詞などの内容語のみを用い、さらに“be”や“have”などのストップワードは削除した。否定語としては、“not”や“never”などだけでなく、“free from”や“lack of”などの表現も加え、すべて合わせると39表現(主辞となる単語の種類数は33)になった。シソーラスとしては、同じくWordNet⁶⁾の持つ類義語、反義語、上位下位語を利用した。また、コーパスからの接続表現としては、PennTree Bank¹⁴⁾のWall Street Journal corpusとBrown corpus(両コーパスとも約100万語程度)から804個の表現を抽出した。孤立した単語を削除する

ストップワードは以下のとおりである: be, do, have, go, come, make, give, take, become, anyway, anyone, anybody, anywhere, anything, here, where, when, whereby, sometime, somewhere, someone, somebody, something, every, everywhere, act, very, off, word, as, thing.

否定表現は以下のとおりである: free of, free from, rid of, short of, devoid of, incapable of, unable to, lack(名詞) of, deficiency of, decrease of, reduction of, lack(名詞) in, deficiency in, decrease in, reduction in, absense of, absense from, loss of, avoidance of, failure to, damage to, compensation for, lack(動詞), destroy, compensate, reduce, relieve, prevent, remove, deprive, abstain from, refrain from, deprive of, rid of, not, never, less, no, without.

と、約88,000単語からなるネットワークが得られた。平均次数はGでは12.73, GTでは17.93, GTCでは17.94となった。リンク全体における負のリンクの割合は、Gで3.52%, GTで5.63%, GTCで5.63%であった。

Turneyらによる実験²⁰⁾に従い、General Inquirer¹⁸⁾の語彙を正解データとして使用した。“Positiv”もしくは“Negativ”のラベルが付いている単語をGeneral Inquirerの単語リストから抽出し、多義性により複数のエントリがある単語は、1つにまとめた。その結果、1,616のポジティブな語、1,980のネガティブな語からなる3,596語のラベル付き単語集合が得られた。これら3,596語はそれぞれ対応するノードを上記語彙ネットワーク上に持つ。

実験は、0.1間隔で0.1から2.0までの異なる β の値を用いて行い、4.3節で述べた方法を用いて最も良い β の値を予測した。ハイパーパラメータの予測のための平均磁化の閾値は、 1.0×10^{-5} に設定した。すなわち、平均磁化がこの閾値を超えないような β のうち最大のもを予測値とする。予備実験により臨界点の検出に相当であると考えられたので、この値に設定した。 α の値は、 1.0×10^3 とした。今回は初期単語集合のラベルが十分信頼できるので、 α の値はこのように十分大きくとっておけばよく、たとえば 1.0×10^2 などとしても結果は変わらない。もしノイズを含むような初期単語集合しか手元にない場合は、 α の値の調整も必要になってくるだろう。

我々は、10分割の交差検定による評価実験(3,200語余りの初期単語集合を用いたことに相当する)、および、少数の固定初期単語集合を用いて極性推定し、残りの単語で評価するという実験を行った。交差検定については、現存する言語資源を十分に利用した場合に提案手法はどれくらいの性能を発揮するのかを確認するために行った。一方、固定初期単語集合での実験はTurneyらの手法との比較のために行った。よって、初期単語集合は彼らが使用したものと同じものを用いた:

- 14 初期単語
ポジティブ {good, nice, excellent, positive, fortunate, correct, superior};
ネガティブ {bad, nasty, poor, negative, unfortunate, wrong, inferior};
- 4 初期単語 {good, superior, bad, inferior};

1 つの単語に異なる極性が与えられている場合は、正解データに含めなかった。

表 1 予測された β を用いた場合の、各語彙ネットワークに対する異なる初期単語集合での分類正解率 (%)。括弧内には、予測された β の値が記載されている。交差検定 cv については、10 回の異なる β の値が予測されるので、その値は省略する

Table 1 Classification accuracy (%) with various networks and four different sets of seed words. In the parentheses, the predicted value of β is written. For cv, no value is written for β , since 10 different values are obtained.

seeds	GTC	GT	G
cv	90.8 (—)	90.9 (—)	86.9 (—)
14	81.9 (1.0)	80.2 (1.0)	76.2 (1.0)
4	73.8 (0.9)	73.7 (1.0)	65.2 (0.9)
2	74.6 (1.0)	61.8 (1.0)	65.7 (1.0)

● 2 初期単語 {good, bad}.

5.1 実験結果：分類正解率

表 1 は、3 種類の初期単語集合に対する、それぞれのネットワークによる単語の感情極性分類の実験結果を示す。2 個の初期単語は “good” と “bad” に対応し、4 個の初期単語は “good”, “bad”, “superior”, “inferior” に対応する。表中、cv は 10 分割交差検定の結果に対応する。交差検定においては、ハイパーパラメータ推定は疑似 leave-one-out エラー率を用いて行った。それ以外の場合は、平均磁化を用いて行った。

多くの場合、シソーラスおよびコーパスの情報を取り入れることで正解率が上昇していることが分かる。例外は、サイズ 2 の初期単語集合において GT が G を下回っているケースである。定式化した最適化問題としての局所解が多く存在するが、我々が実際に欲しいのはそのうちの一部の解である。初期単語が少数である場合は解空間の自由度が大きく、最適化の過程で計算が期待される解と異なる局所解に陥りやすくなっている可能性がある。これは、本来の目的と定式化とのギャップが問題であり、近似性能などの問題とは種類が異なる。

提案手法の結果を Turney らの結果と比較する。14 初期単語で彼らは、小さなコーパス (約 1×10^7 単語) を用いて 61.26%, 中程度のサイズのコーパス (約 2×10^9 単語) を用いて 76.06%, 大きなコーパス (約 1×10^{11} 単語) を用いて 82.84% の正解率を得ている。ただし、サイズの大小は相対的なものである。

我々は、コーパスもシソーラスも使用することなく (しかし辞書の語釈文を使用して) 正解率 76.2% に到達し、中程度のサイズのコーパスでの Turney らの結果と同等の結果を得た。さらに、コーパスやシソーラスで語彙ネットワークを拡充した場合、正解率 81.9% に到達し、我々は大きなコーパスでの Turney らの結果

表 2 最適の β を用いた場合の、各語彙ネットワークに対する異なる初期単語集合での分類正解率 (%)。括弧内にそのときの β の値が記載されている (ただし cv の場合は除く)

Table 2 Actual best classification accuracy (%) with various networks and four different sets of seed words. In the parentheses, the actual best value of β is written, except for cv.

seeds	GTC	GT	G
cv	91.5 (—)	91.5 (—)	87.0 (—)
14	81.9 (1.0)	80.2 (1.0)	76.2 (1.0)
4	74.4 (0.6)	74.4 (0.6)	65.3 (0.8)
2	75.2 (0.8)	61.9 (0.8)	67.5 (0.5)

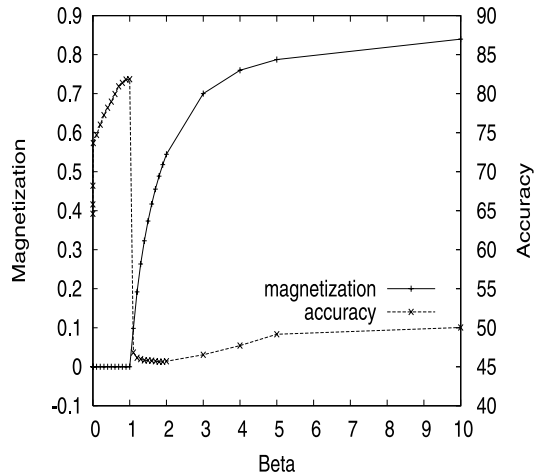


図 1 平均磁化と正解率の変化 (14 初期単語)

Fig. 1 Example of magnetization and classification accuracy (14 seed words).

と同等な結果を得た。

5.2 ハイパーパラメータ β の予測

ハイパーパラメータの予測手法の性能を調べる。表 2 に最適な値を用いた結果を示す。

表 1 と表 2 を比較することにより、特に 14 初期単語の場合において、我々の予測手法は良い性能を示していることが分かる。その他の場合も、いくらか大きな値を予測する傾向があるものの、正解率としては遜色ない値が出ているといえる。予測値が大きくなる正確な理由は分からないが、平均磁化が閾値を超えたときの値を用いるというシンプルな方法を用いているので、将来的に変化点検出のような技術を考案し使用することにより解決できる可能性はある。

β の値を変化させたときの、平均磁化と正解率の変化を表した図 1 も、参考のために付け加える。図 1 においては、 $\beta = 1.0$ 付近で相転位が起こっており、その直前に正解率がピークを迎えている。我々の提案した β の予想方法が期待どおりに機能していることが分かる。

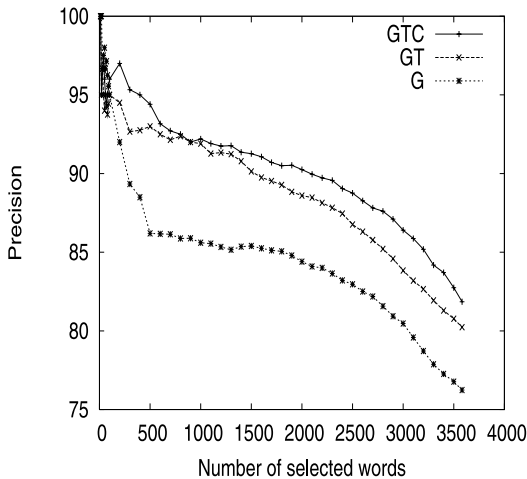


図 2 14 初期単語の場合の精度 (%)

Fig. 2 Precision (%) with 14 seed words.

5.3 確信度が高い分類結果に対する精度

次に、高い確信度を持って分類された単語に対する分類正解率（すなわち、精度）を算出する。確信度としては、スピンの平均値の絶対値 $|x|$ を用いた。この絶対値の降順に単語を並べ、上位の単語のみで精度を測った。

初期単語数 14 の場合を例にとり、結果を図 2 に示す。横軸は選ばれた上位単語数、縦軸は精度である。上位 1,000 語では 92% を超える精度が得られた。この結果から、平均値の絶対値が分類の確信度として機能することも確認できた。

5.4 他の手法との詳細な比較

提案モデルについてさらに詳しく調べるために、限定された条件でいくつかの実験を行う。

5.4.1 Kamps らの手法との比較

Kamps ら¹²⁾ による最短距離に基づく手法と、我々の用いた大域的最適化に基づく手法の違いを明確にするために、計算方法以外の設定を等しくして比較実験を行う。Kamps らの手法は負のリンクを計算に取り入れることができないので、彼らが用いた類義語関係だけからなる語彙ネットワークを使う。同じネットワークを利用しているので、計算手法の違いが結果の違いに直接的に現れるはずである。また、評価対象を 697 形容詞に制限した。これは、General Inquirer から抽出した単語のうち Kamps らの手法が 0 でない極性値を出力したものの数である。また、Kamps らの手法はもともと 2 初期単語に対して作られたものであるので、4 初期単語や 14 初期単語の場合は、ポジティブ初期単語集合、ネガティブ初期単語集合からの平均最短距離の差を用いて極性値を算出した。表 3 に結果

表 3 類義語ネットワークでの分類正解率 (%)。提案手法と Kamps らの手法との比較

Table 3 Precision (%) with the synonym network. Comparison between the proposed method and the shortest-path method.

seeds	提案手法	Kamps らの手法
14	73.4 (1.0)	70.8
4	71.0 (1.0)	64.9
2	68.2 (1.0)	66.0

表 4 類義語・反義語ネットワークでの分類正解率 (%)。提案手法と Hu らの手法との比較

Table 4 Precision (%) with the synonym and antonym network. Comparison between the proposed method and the bootstrapping method.

seeds	提案手法	Hu らの手法
14	83.6 (0.8)	72.8
4	82.3 (0.9)	73.2
2	83.5 (0.7)	71.1

を示す。この結果より、スピンモデルの大域的な最適化が、感情極性抽出問題においてうまく働いていることが分かる。

GTC や GT に対しても、Kamps らの手法を試したが、非常に低い正解率を出した。これは、4.4 節の議論でも述べたように、語彙ネットワーク上の最短距離を用いるような方法はノイズに敏感であり、語釈文やコーパスのようにノイズが含まれやすいリソースはうまく活用できないからである。

5.4.2 Hu らの手法との比較

次に、Hu ら⁹⁾ の単純なブートストラッピングに基づく手法との比較を行う。ここでは、類義語関係と反義語関係を用いて語彙ネットワークを構築する。一般のシソーラスの類義語関係および反義語関係は同品詞の単語間のみ存在する。Hu らの手法はシソーラスに依存しているため、初期単語と等しい品詞の語しか、感情極性を予測できない。それゆえ、評価は形容詞 (1,470 語) だけに限定する。結果を表 4 に示す。スピンモデルの大域的な最適化が、感情極性抽出問題においてうまく働いていることを示している。

GTC や GT に対しても Hu らの手法を試したが、Kamps らの手法の場合と同様に、非常に低い正解率を出した。これは、Hu らの手法は、Kamps らの手法と同様にノイズに弱いからである。

5.5 出力例とエラー解析

ネットワークとして GTC を用い、14 初期単語を用いた場合の提案手法の出力を例として下に示す。ここでは、平均値が高いものから 30 単語 (ポジティブであると予測されたもの)、逆に平均値が低いものから 30 単語 (ネガティブであると予測されたもの) を載せる。

ポジティブ first-class, topping, top-hole, top-flight, spiffing, fantabulous, best, out-and-outer, excellently, well-off, better, heaven-sent, well-to-do, well-situated, well-heeled, well-fixed, nicely, ace, tops, topnotch, tiptop, a-one, crack, first-rate, brag, superb, auspicious, jimhickey, jim-dandy, crackerjack

ネガティブ piteous, hapless, pitiful, wretched, deplorable, grotty, ill-starred, ill-omened, poorest, ill-fated, lousy, regrettable, pitiable, miserable, pathetic, punk, crummy, cheesy, bum, unlucky, dispossessed, chintzy, roofless, woefully, lamentably, deplorably, tinny, sadly, sleazy, castaway

またたとえば、ネガティブな極性を持つ “gullible” は、語彙ネットワーク上でポジティブな極性を持つ “trust” とともに、ネガティブな極性を持つ “fool” とともに重みでリンクされており、極性の衝突がここで起こっているが正しく分類されている。Hu ら⁹⁾ の手法のような単純なブートストラッピングによる手法では、“gullible” は正しく分類できない可能性が高い。

また、提案手法を用いた場合の分類誤りを調査し、主に 3 種類の原因があることが分かった。

1 つは語義の曖昧性の問題である。たとえば、“costly” の語釈文は “entailing great loss or sacrifice” である。しかし、ここでの “great” は程度が大きいことを示しているにもかかわらず、ポジティブな極性を持つ「偉大」という意味の用法の影響を受けて、結果として “costly” が誤ってポジティブに分類されている。

また、構文構造が取り入れられていないことによる分類誤りも存在する。たとえば、“arrogance” の語釈文は “overbearing pride evidenced by a superior manner toward the weak” であり、正しくはネガティブな単語であるが、“superior” という単語の影響でポジティブに分類されてしまっている。“superior” であるのは “manner” であることが構文構造から推測できるようになると、このような分類誤りは避けることができる。

さらに別のタイプの誤りとしては、熟語的表現がある。たとえば、“brag” は、“show off” の意味でネガティブなのだが、“show” と “off” のどちらもネガティブな意味を持っていないので、正しく分類ができなかった。

現在のモデルにおいては、これらの分類誤りをうまく扱うことはできない。このような問題の解決は、今後の課題として取り組んでいく。

6. 結 論

単語の感情極性を判定する手法を提案した。提案手法では、感情極性を電子のスピンと見なし、語釈文、シソーラス、およびコーパスを用いて語彙ネットワークを構築した。さらに平均場近似を用いて近似確率分布関数を計算し、各単語の極性を調べた。実験により、提案手法は少数の初期単語を用いて高い精度で感情極性分類を判定できることが示した。

今後の発展性としてはたとえば以下のようなものが考えられる。

- 語釈文をなす単語の重要性はその構文的役割に依存するので、構文情報の利用は分類に役立つことが期待される。
- 能動学習も望まれる発展方向の 1 つである。初期単語集合の作成の手間を軽減するためには、少数の “良い” 初期単語が自動的に選択されるような方法が必要とされる。
- 今回は式 (9) のように重みを決定したが、この決め方はヒューリスティックであり、他の式も多数考えられる。また、各リソース（語釈文、シソーラス、コーパス）からの情報を異なる扱いにすることによる分類性能の向上も期待できる。このような重みの決め方に関する研究も重要である。
- 今回は、“and” もしくは “but” で並列につながれた形容詞ペアのみをコーパスから抽出して利用したので、GT と GTC の平均次数はほとんど違わなかった。しかし、感情極性が等しくなりやすい言語パターンはほかにも存在すると思われる。そのような言語パターンの発見と利用も今後考えていきたい。
- 我々のモデルは容易に多クラスモデルへ拡張可能であるが、そのような拡張の有効性はまだ示されていない。
- 提案モデルは、同じ極性を持ちやすいという傾向を利用しているだけである。それゆえ、コーパスには出現するが辞書に載っていない語に対しても、感情極性を推定可能である。そのような未知語に対する利用可能性も追求していく価値がある。
- 今回使用したコーパスは小さいものであった。大きなコーパスを利用することにより、分類性能はさらに向上する可能性がある。特に、Turney ら²⁰⁾ のようなウェブを用いた手法との融合は、大きな可能性を含む。たとえば、エネルギー関数にバイアス項 $-\sum_i h_i x_i$ を付け加え、Turney らの手法を用いて h_i の値を推定するような方法が考えら

れる．このときはハイパーパラメータ β の予測方法などにも拡張が必要であることが予想されるので，注意しなくてはならない．

- 語彙ネットワーク上の大域的最適化は汎用的な枠組みであり，提案手法を他のタスクへ適用していくことも今後期待できる．
- 獲得された単語の感情極性を文あるいは文章の極性分類に利用することも考えている．その際は，文脈に応じて単語の感情極性にある種の演算を施すことにより，ここで得られた極性情報を有効に利用できるであろう．

参 考 文 献

- 1) Berger, A.L., Pietra, S.D. and Della Pietra, V.J.: A maximum entropy approach to natural language processing, *Computational Linguistics*, Vol.22, No.1, pp.39–71 (1996).
- 2) Chandler, D.: *Introduction to Modern Statistical Mechanics*, Oxford University Press (1987).
- 3) Cowie, J., Guthrie, J. and Guthrie, L.: Lexical disambiguation using simulated annealing, *Proc. 14th conference on Computational linguistics*, Vol.1, pp.359–365 (1992).
- 4) Dave, K., Lawrence, S. and Pennock, D.M.: Mining the peanut gallery: Opinion extraction and semantic classification of product reviews, *Proc. 12th International World Wide Web Conference (WWW-2003)* (2003).
- 5) Deerwester, S.C., Dumais, S.T., Landauer, T.K., Furnas, G.W. and Harshman, R.A.: Indexing by latent semantic analysis, *Journal of the Society for Information Science*, Vol.41, No.6, pp.391–497 (1990).
- 6) Fellbaum, C.: *WordNet: An Electronic Lexical Database, Language, Speech and Communication Series*, MIT Press (1998).
- 7) Geman, S. and Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.6, pp.721–741 (1984).
- 8) Hatzivassiloglou, V. and McKeown, K.R.: Predicting the semantic orientation of adjectives, *Proc. 35th Annual Meeting of the Association for Computational Linguistics and the Eighth Conference of the European Chapter of the Association for Computational Linguistics*, pp.174–181 (1997).
- 9) Hu, M. and Liu, B.: Mining and summarizing customer reviews, *Proc. 2004 ACM SIGKDD international conference on Knowledge discovery and data mining (KDD-2004)*, pp.168–177 (2004).
- 10) Iba, Y.: The nishimori line and bayesian statistics, *Journal of Physics A: Mathematical and General*, pp.3875–3888 (1999).
- 11) Inoue, J. and Carlucci, D.M.: Image restoration using the q-ising spin glass, *Physical Review E*, 64:036121-1–036121-18 (2001).
- 12) Kamps, J., Marx, M., Mokken, R.J. and deRijke, 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).
- 13) Mackay, D.J.C.: *Information Theory, Inference and Learning Algorithms*, Cambridge University Press (2003).
- 14) Marcus, M.P., Santorini, B. and Marcinkiewicz, M.A.: Building a large annotated corpus of English: The Penn Treebank, *Computational Linguistics*, Vol.19, No.2, pp.313–330 (1993).
- 15) Marroquin, J.L.: Optimal bayesian estimators for image segmentation and surface reconstruction, Technical Report A.I. Memo 839, Massachusetts Institute of Technology (1985).
- 16) Riloff, E., Wiebe, J. and Wilson, T.: Learning subjective nouns using extraction pattern bootstrapping, *Proc. 7th Conference on Natural Language Learning (CoNLL-03)*, pp.25–32 (2003).
- 17) Schmid, H.: Probabilistic part-of-speech tagging using decision trees, *Proc. International Conference on New Methods in Language Processing*, pp.44–49 (1994).
- 18) Stone, P.J., Dunphy, D.C., Smith, M.S. and Ogilvie, D.M.: *The General Inquirer: A Computer Approach to Content Analysis*, The MIT Press (1966).
- 19) Tanaka, K., Inoue, J. and Titterton, M.: Probabilistic image processing by means of the bethe approximation for the q-ising model, *Journal of Physics A: Mathematical and General*, 36:11023–11035 (2003).
- 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) Veronis, J. and Ide, N.M.: Word sense disambiguation with very large neural networks extracted from machine readable dictionaries, *Proc. 13th Conference on Computational Linguistics*, Vol.2, pp.389–394 (1990).
- 22) Wiebe, J.M.: Learning subjective adjectives

from corpora, *Proc. 17th National Conference on Artificial Intelligence (AAAI-2000)*, pp.735-740 (2000).

- 23) 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ロジー研究会 (SIG-SWO-A401-01) (2004).
- 24) 立石健二, 福島俊一, 小林のぞみ, 高橋哲朗, 藤田篤, 乾健太郎, 松本裕治: Web 文書集合からの意見情報抽出と着眼点に基づく要約生成, 情報処理学会自然言語処理研究会 (NL-163-1), pp.1-9 (2004).
- 25) 藤村 滋, 豊田正史, 喜連川優: 電子掲示板からの評価表現および評判情報の抽出, 人工知能学会第 18 回全国大会 (2004).
- 26) 小林のぞみ, 乾 孝司, 乾健太郎: 語釈文を利用した「p/n 辞書」の作成, 人工知能学会言語・音声理解と対話研究会 SLUD-33, pp.45-50 (2001).

(平成 17 年 5 月 17 日受付)

(平成 17 年 11 月 1 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)

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. 博士 (工学). 現在, 日本学術振興会特別研究員. 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. AAI, ACL, JSAI, JCSS 各会員.