

スピンモデルを用いた評価表現辞書構築における精度改善法

後藤 拓馬[†] 樺島 祥介^{††} 高村 大也^{†††}

東京工業大学 総合理工学研究科^{†††}, 精密工学研究所^{†††}

[†]takuma@sp.dis.titech.ac.jp, ^{††}kaba@dis.titech.ac.jp,

^{†††}takamura@pi.titech.ac.jp

1 はじめに

近年のインターネット環境の整備やブログなどの普及によって評価情報が蓄積されてきている。これらの評価情報を抽出し、整理し、提示することは、評価対象の提供者である企業や、評価対象を利用する立場の一般の人々の双方にとって有益となる。このため、評価情報を扱う研究が活発に行われている [1]。

テキストの評価分析に用いられる要素技術に、評価表現辞書の構築 [2, 3] がある。評価表現辞書とは、評価表現とその表現がもつ評価極性の対 (例: 良い - 肯定) の集合である。この集合の要素を感情 (評価) 極性とよぶ。

高村ら [2] は語彙ネットワークを一種のスピンシステムとしてモデル化することにより、感情極性を求め、評価表現辞書を構築する手法を提案した。この手法では適当な温度での各スピンの熱平均値を平均場近似を用いて評価することで、単語の感情極性を求めている。このシステムでは温度を十分高温から下げていった場合、得られる感情極性の精度はある臨界的な温度までは単調に改善される。ところが、温度を臨界値以下に下げると強磁性相 (ほぼ全てのスピンの向きが同じ方向を向く状態) に転移し、分類精度が急激に悪化する、という特徴的な振る舞いが観察される。

そこで、提案手法では低温でシステムを強磁性相にする要素をモデルから取り除くことにより、システム全体が強磁性相にならないような工夫を導入した。その結果、実験により、提案手法は既存の手法 [2] より高い精度で評価表現辞書を構築できることがわかった。

2 関連研究

評価表現辞書の自動構築に関する技術についてはこれまでにもさまざまな研究が行われてきた [2, 3, 4, 5, 6]。

Velikovich ら [4] は、web 上から収集した大量のコーパスデータを基に重み付ネットワークを構築し、種となる単語からの重み和を用いてネットワーク上の単語の極性評価を行う手法を提案した。

Rao ら [5] は Label Propagation [7] と呼ばれるアルゴリズムを WordNet [8] 上の類義語と上位・下位語を

連結したネットワークに適用することで単語の極性評価を行う手法を提案した。

Kamps ら [6] は WordNet [8] 上の類義語を連結したネットワーク上の任意の形容詞への “good” からの最短距離と “bad” からの最短距離を比較して、単語の極性を評価する方法を提案した。

3 スピンモデルと平均場近似

イジングスピンとよばれる $S \in \{\pm 1\}$ をとる N 個の構成要素からなるシステムを考える。このシステムの状態を N 次元ベクトル $S = (S_1, S_2, \dots, S_N)$ によって表す。平衡統計力学では、温度 $T = \beta^{-1}$ の熱平衡状態においてシステムがある状態 S に見出される確率 $P(S)$ は以下の分布関数 (ボルツマン分布) によって表現されると仮定する:

$$P(S) = \frac{e^{-\beta E(S)}}{Z}. \quad (1)$$

$E(S)$ はエネルギー関数であり、この分布関数の形を特徴付ける関数である。相互作用下にあるイジングスピンモデルのエネルギー関数はしばしば

$$E(S) = -\frac{1}{2} \sum_{ij} J_{ij} S_i S_j \quad (2)$$

の形に仮定される。ここで、 J_{ij} は二つのスピン (i, j) のリンクの重みであり、スピン間の相互作用の強さを表す。 β は逆温度とよばれ、この値が大きいほどエネルギー関数の確率分布への寄与が大きくなる。また、 Z は分布関数の規格化因子 $Z = \sum_S e^{-\beta E(S)}$ であり、分配関数とよばれる。

この分布関数を用いてエネルギー $E(S)$ で特徴付けられるシステムのさまざまな統計量を評価することができる。例えば、あるスピン i の熱平均値は

$$\langle S_i \rangle = \sum_S \frac{S_i e^{-\beta E(S)}}{Z} \quad (3)$$

と評価される。

実際にさまざまな統計量をボルツマン分布に対して計算するのは、ベクトルの次元数 N が大きい場合、現

実は計算量的に不可能である．簡単に表記される上述のスピンモデルでさえ，式(3)からわかるようにスピンの熱平均値の計算に 2^N の計算量を必要とする．この困難を現実的に解決するため，平均場近似を用いてシステムの統計量を近似的に求める．平均場近似では次元の同時確率分布である $P(S)$ をそれぞれの変数の積で書ける関数 $Q(S) = \prod_i Q_i(S_i)$ で近似する．この Q 関数は以下の変分自由エネルギー F を最小化する条件で求まる：

$$\beta F[Q] = \beta \sum_S Q(S) E(S) - \left(- \sum_S Q(S) \log Q(S) \right). \quad (4)$$

この F を最小化することは P と Q のKL距離を最小にすることと同値であることが知られている[9]．変分自由エネルギーの極値条件より，以下の平均場方程式が得られる：

$$m_i = \tanh \left\{ \beta \left(\sum_j J_{ij} m_j \right) \right\}. \quad (5)$$

ただし， m_i は分布 $Q(S)$ でのスピン S_i の熱平均値 $m_i = \sum_S S_i Q(S)$ を表している．この方程式を反復的に適用し収束解を求めることで各スピンの熱平均値の近似値 m_i が評価できる．

我々のスピンシステムでは磁化という秩序変数を定義することができる：

$$m = \frac{1}{N} \sum_i m_i. \quad (6)$$

ここで， N はシステムの要素の数である．この秩序変数を用いてシステムのマクロな状態(相と呼ばれる)が議論できる． $m \neq 0$ であれば，システムは正の方向，あるいは負の方向に空間的に秩序だっており(強磁性相)， $m = 0$ であれば，システムは空間的には無秩序の状態(常磁性相)にある．システムの温度を少しずつ変化させていくと，ある温度で空間的に秩序だったスピンの突然無秩序状態になったりその逆の現象を観察することがある．この様に，ある温度を境にシステムの質的な振る舞いが大きく変化することを相転移とよぶ．

4 評価表現辞書構築

本節では，高村ら[2]によって提案されたスピンモデルによる評価表現辞書構築法の概略を説明する．

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．極性が同じと思われる単語同士(SL:same-orientation links)，反対の極性を持つと思われる単語

同士(DL:different-orientation links)を連結する．次に，シソーラスで与えられた類義語ペア(SL)，反義語ペア(DL)，上位語下位語ペア(SL)をリンクで連結する．さらにコーパスに出現する接続表現を利用し，コーパスの情報を加える[3]．

このようにして連結されたネットワークに対し，各リンクの重み J_{ij} を次のように定める：

$$J_{ij} = \begin{cases} \frac{1}{\sqrt{d(i)d(j)}} & (l_{ij} \in SL) \\ -\frac{1}{\sqrt{d(i)d(j)}} & (l_{ij} \in DL) \\ 0 & otherwise \end{cases} \quad (7)$$

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 がリンクされている単語数を表す．

4.2 感情(評価)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．これは，少数の単語の感情極性がすでにわかっている状態であることに対応する．このラベル付データは以下のようにして式(2)のエネルギー関数を書き換えることでシステムに取り入れることができる：

$$E(S) = -\frac{1}{2} \sum_S J_{ij} S_i S_j - \sum_{i \in L} a_i S_i. \quad (8)$$

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)，はラベル付データに対する重み(正の定数)を表す．物理的には，初期単語集合に強さの外場を a_i の方向にかけたことに対応する．このエネルギー関数は既存手法[2]のモデル化と多少異なるが，初期単語集合に強力なバイアスを加えることには変わりなく，本質的には同等である(バイアスに温度依存性があるか否かが異なる)．このエネルギー関数に対応する平均場方程式は

$$m_i = \tanh \left\{ \beta \left(\sum_j J_{ij} m_j + a_i \right) \right\} \quad (9)$$

となる．

収束判定には変分自由エネルギーを用いる．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，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．

4.3 数値計算

高村ら[2]の数値実験の条件設定とその結果を示す．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，WordNet[8]を用いる．語釈文に対する品詞タグ付けには，TreeTagger[10]を用いる．

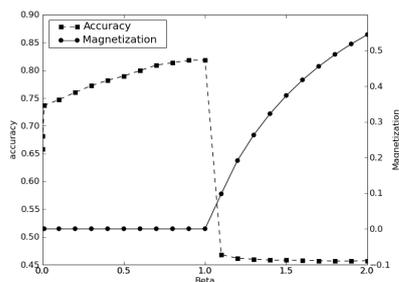


図 1: 14 初期単語における分類正解率と磁化

シソーラスとしては WordNet[8] の持つ類義語、反義語、上位下位語を利用する。また、コーパスからは、PennTree Bank[11] の Wall Street Journal corpus と Brown corpus (両コーパスとも約 100 万語程度) から 804 個の表現を抽出する。結果、約 88,000 語からなるネットワークが得られる。

General Inquirer[12] の単語リストから “Positiv” もしくは “Negativ” のラベルがついている単語を抽出し、正解データとして使用する。この正解データは 1616 のポジティブな語、1980 のネガティブな語からなる 3596 語のラベル付単語集合である。これら 3596 語はそれぞれ対応するノードを上記ネットワーク上にもつ。

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、 β の値は $1.0 \cdot 10^3$ とする。

実験で用いる初期単語集合は以下の通りである：

ポジティブ {good, nice, excellent, positive, fortunate, correct, superior};

ネガティブ {bad, nasty, poor, negative, unfortunate, wrong, inferior};

図 1 に 14 初期単語での各温度における分類精度と磁化を示す。これより、 $\beta = 1.0$ 付近で強磁性相への相転移が起こっており、その直前の β で最も良い正解率となっていることがわかる。

5 モデルの分析と再構築

我々が考察しているモデルの J_{ij} 行列は以下のように直交基底を用いて固有値展開を行うことができる：

$$J = \sum_{\mu} \lambda_{\mu} \mathbf{x}^{\mu} (\mathbf{x}^{\mu})^T. \quad (10)$$

ここで、 λ_{μ} は J_{ij} 行列の第 μ 固有値、 \mathbf{x}^{μ} は J_{ij} 行列の長さが 1 に規格化された第 μ 固有ベクトルである。

システムを低温にするに従い、平均場方程式の解であるスピンの熱平均値は J_{ij} 行列の第 1 固有ベクトルの成分に近づいていくことが期待される。すなわち、システムを低温にして強磁性相に相転移しているということは、モデルの第 1 固有ベクトルが強磁性なベクトル（ほぼ全ての成分が正）であるということが推測できる。

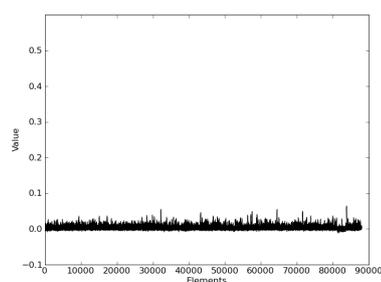


図 2: J_{ij} 行列の第 1 固有ベクトル

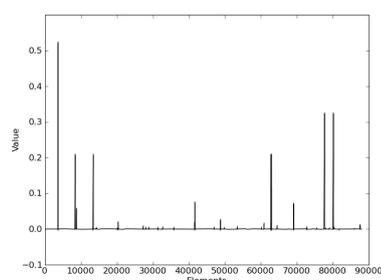


図 3: J_{ij} 行列の第 2 固有ベクトル

ここで、既存のモデルの J_{ij} 行列の第 1 固有ベクトル、第 2 固有ベクトルを図 2, 3 に示す。これより、第 1 固有ベクトルはほとんどの成分で正の値を持っており、実際に強磁性なベクトルであることがわかる。第 2 固有値は少数の成分のみが有限値をもち、それ以外の成分はほぼ 0 に近い値となっている。このことは、これらの有限値をもっている成分が強く連結し、それ以外の成分とはほとんど関わりを持たないでクラスタを作っていることを意味している。

これらの考察、結果により、第 1 固有値の要素がシステムを低温に下げたときに強磁性相に相転移する主な要因であると考えられる。すなわち、 J_{ij} 行列から第 1 固有ベクトル成分を取り除くことによって相転移の発現が抑えられ、感情極性の抽出精度が向上することが期待される。

そこで、提案手法では J_{ij} 行列を以下のように修正する：

$$J' = J - \lambda_1 \mathbf{x}^1 (\mathbf{x}^1)^T. \quad (11)$$

ここで、 λ_1 は J_{ij} 行列の規格化された第 1 固有値、 \mathbf{x}^1 は第 1 固有ベクトルである。第 2 項で J_{ij} 行列に対する第 1 固有ベクトルの構成成分を引いている。

6 評価実験

改良したモデルの感情極性の抽出精度を確かめるために、数値実験を行った。少数の固定初期単語集合を用いて極性推定を行う以外に、10 分割の交差検定 (3200 語余りの初期単語集合を用いたことに相当) に

表 1: 最適の β を用いた場合の異なる初期単語集合での分類正解率 (%)。括弧内にその時の β の値を記載した。交差検定では、毎回異なる β が採用されるので値は省略した。

seeds	Takamura	提案手法
交差検定	91.5(-)	91.8(-)
14	81.9(1.0)	86.2(1.2)
4	74.4(0.6)	83.5(1.2)
2	75.2(0.8)	84.5(1.2)

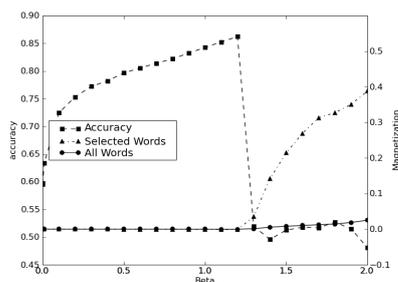


図 4: 14 初期単語集合の分類正解率と磁化。磁化は正解ラベルを持っている単語についてのみの磁化 (Selected Words) と全ての単語についての磁化 (All Words)

よる評価も行った。これは、多くの初期単語集合を用いた際に得られる各手法の性能限界を確認するためである。既存のモデルとの比較の為に、モデルの構成以外は既存のモデルと同様に行った。

6.1 実験結果

提案手法の 3 種類の初期単語集合と 10 分割交差検定に対する、単語の感情極性分類を既存手法 [2] と比較した結果を表 1 に示す。全ての場合で分類精度が向上した。特に、初期単語集合が小さい場合の向上率が大きい。

図 4 に本実験による磁化と正解率の変化のグラフを示す (14 初期単語)。これより、改良されたモデルでは低温にしてもシステム全体では強磁性相に相転移していないことがわかる。しかし、正解率を評価している単語のみの磁化 (3,596 語の磁化) に注目すると、これらの単語のみでは磁化が発生している。すなわち、低温領域においては正解率を評価している単語のみに注目すると、それらは正の方向に秩序だっている。それゆえ、 $\beta = 1.3$ より低温領域では正解率が下がっている。これは、低エネルギー状態を実現する各スピンの向きが感情極性を表現していないことを意味する。

もっとも高い正解率は相転移する直前の臨界状態において観察できる。臨界状態では外場の影響に非常に敏感であることから、低エネルギーを実現するスピン

状態の情報ではなく、外場 (種となる単語) からの影響の受けやすさの情報が感情極性分類のタスクに対して有用な情報をもっていることが示唆される。

7 おわりに

本稿ではスピンモデルを用いた評価表現辞書構築における、語彙ネットワークからの単語の感情極性抽出法を提案した。提案手法では、既存の手法におけるモデルの特徴的な振る舞いに着目し、そのモデルを解析し、モデルの改良を行った。その結果、既存の手法と比較して、より良い精度で評価表現辞書を構築することが可能となった。

今後の課題としては、より多くの言語データの使用、実際の評価文分析への適用、さらなるモデルの改良などがあげられる。

参考文献

- [1] 乾 孝司, 奥村 学, “テキストを対象とした評価情報の分析に関する研究動向”, *自然言語処理*, vol.13, no.3, pp.201–241, 2006.
- [2] Hiroya Takamura, Takashi Inui and Manabu Okumura, “Extracting Semantic Orientations of Words using Spin Model”, *ACL2005*, pp.133–140, 2005.
- [3] Vasileios Hatzivassiloglou and Kathleen R. McKeown, “Predicting the semantic orientation of adjectives”, *35thACL*, pp.174–181, 1997.
- [4] Lenid Velikovich, Sasha Blair-Goldensohn, Kerry Hannan, Ryan McDonald, “The Viability of Web-derived Polarity Lexicons”, *NACL*, pp.777–785, 2010.
- [5] Delip Rao and Deepak Ravichandran, “Semi-Supervised Polarity Lexicon Induction”, *12th EACL*, pp.685–682, 2009.
- [6] Jaap Kamps, Maarten Marx, Robert J. Mokken, and Maarten de Rijke, “Using wordnet to measure semantic orientation of adjectives”, *LREC 2004*, vol. 4, pp.1115–1118, 2004.
- [7] Xiaojin Zhu and Zoubin Ghahramani, “Learning from labeled and unlabeled data with label propagation”, Technical Report CMU-CALD-02-107, Carnegie Mellon University, 2002.
- [8] Christiane Fellbaum, *WordNet: An Electronic Lexical Database*, Language, Speech, and Communication Series, MIT Press, 1998.
- [9] Manfred Opper and David Saad, *Advanced Mean Field Methods: Theory and Practice*, The MIT Press 2001.
- [10] Helmut Schmid, “Probabilistic part-of-speech tagging using decision trees”, *NeMLaP*, pp.44–49, 1994.
- [11] Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz, “Building a large annotated corpus of english: The Penn Treebank”, *Computational Linguistics*, vol.19, no.2, pp.313–330, 1993.
- [12] Philip J. Stone, Dexter C. Dunphy, Marshall S. Smith, and Daniel M. Ogilvie, *The General Inquirer: A Computer Approach to Content Analysis*, The MIT Press, 1966.