スピンモデルの感情情報処理への応用

高村 大也 †　乾 孝司 †　奥村 学 ‡
東京工業大学 精密工学研究所
{takamura,oku}@pi.titech.ac.jp, tinui@lr.pi.titech.ac.jp

1　序論

文書データにおける感情 (意見や態度を含む) の判別は, 様々な応用において重要になってきている. 例を挙げると, インターネットの電子掲示版上に書き込まれたある製品に関する消費者の意見が特定・判別できると, その製品に関する広範囲なサーベイが可能になる. また, すべての回答に人間が目を通すことなくアンケートの概要を把握することも可能になる. そのような自動特定・判別を実現するための重要なリソースとして, 各単語の感情極性 (ポジティブ「望ましい」, もしくはネガティブ「望ましくない」) リストを挙げることができる. 最もシンプルには, ある文書においてポジティブな単語が多く出現している場合, その文書全体もポジティブであると推測ができる.

本研究の目的は, そのような感情極性リストを自動的に作成する方法を提案することである. そのため, ここではスピン系のモデルを用いる. 電子のスピンが方向 (上向きあるいは下向き) を持っているように, 各単語は感情極性 (ポジティブあるいはネガティブ) を持っている. 電子のスピン系において隣り合ったスピンは同じ方向を持ち易いという傾向がある. 我々は, 辞書の見出し語とその語釈文にある単語とは, 同じ感情極性を持ち易いと仮定し, 語釈文で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. スピン系で使用されている平均場近似を用いて系の状態を算出し, 各単語の感情極性を判定する. さらに, 系の状態に影響を与えるハイパパラメータの予測方法も同時に提案する. ここでは, 統計物理の基本的概念である平均磁化を用いる.

非常に小規模の初期単語集合 (種となる単語とその感情極性) を与えるだけで, 多くの単語の感情極性が高い正解率で判定できることを, 実験的に示す. 我々の実験では, 英語語彙については 81.9%, 日本語語彙については 79.8%の正解率を得た. また, 獲得された単語の感情極性を用いて, 文の感情極性を判定する実験も行い, 提案手法が優れていることを示した.

2　関連研究

Turney ら (2003) はウェブの検索エンジンを用いて種となる単語との関連性の強さを測っている. 具体的には, "*word* NEAR good" などのクエリを検索エンジンに投げ, そのヒット数から感情極性の指標を算出している. 実験結果は, General Inquirer の語彙 (Stone et al., 1966) で評価している.

Hatzivassiloglou (1997) は, "simple and well-received" あるいは "simplistic but well-received" などの接続表現に注目した. 前者は両単語の極性は等しく, 後者では異なると予想される. 彼らは, まず接続表現を収集し, それらを同極性か異極性か判定した. その判定結果を用いて単語を二つのグループに分け, 片方をポジティブと判定している.

小林ら (2001) は, 単語の語釈文に対するルールを人手で 52 個作成し, そのルールを繰り返し適用することにより, 単語の感情極性を決定した. この手法は我々の提案手法のもととなっているが, 再現率の低さと言語依存性という二つの欠点を持つ.

Kamps ら (2004) は, WordNet (Fellbaum, 1998) の同義語ネットワーク上での, 初期単語 ("good" と "bad") への最短距離を用いて感情極性を判定した.

Hu ら (2004) は, 初期単語集合から始めて, シソーラス上で同義関係にある単語を取り込んでいくブートストラッピング方式を提案している. 彼らの手法の本質的な部分は, Kamps らの手法と同じである.

3　スピンモデルと平均場近似

ここでは, スピンモデルと平均場近似 (Mackay, 2003) を簡単に紹介する. これらは, 統計物理や機械学習などの分野で盛んに研究されてきた.

スピン系とは, N 個の電子の成す系であり, 各電子には "$+1$ (上向き) " もしくは "-1 (下向き) " のどちらかの値をとるスピンが付随している. 隣接した電子は, お互いに同じ向きを取り易いことが知られている. このモデルをスピンモデルと呼ぶことにする[1]. スピン系のエネルギー関数は次のように表される：

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

ここで, x_i と x_j はそれぞれ電子 i と j のスピンの方向を, 行列 $W = \{w_{ij}\}$ は電子間の重みを表す.

スピン系においては, 変数ベクトル $\mathbf{x}$ はボルツマン分布に従う：

$$P(\mathbf{x}|W) \quad = \quad Z(W)^{-1} \exp(-\beta E(\mathbf{x}, W)). \qquad (2)$$

ここで, $Z(W) = \sum_{\mathbf{X}} \exp(-\beta E(\mathbf{x}, W))$ は正規化項であり, $\beta(= 1/T)$ は逆温度定数と呼ばれる定数である. なお, T は物理の世界では温度に対応する定数である. この分布関数が示すように, エネルギーが小さな状態ほど確率が大きくなる.

確率分布は既知だが, この系の状態は実は依然としてわからない. その理由は, 2^N 個の異なる状態が存在するので, 確率値の計算量が非常に大きいからである.

それゆえ, スピン系においては, しばしば平均場近似と呼ばれる近似計算が用いられる. 以下, スピン系における平均場近似を説明する. $P(\mathbf{x}|W)$ を $Q(\mathbf{x};\theta)$ で近似することを考える. θ は Q を決定するパラメータの集合であり, $P(\mathbf{x}|W)$ にできるだけ近くなるように決定される. P と Q の近さの指標としては, 変分自由エネルギー F を用いる. F は, 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：

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

この F を最小にするパラメータ θ が選ばれる. 実は, F の最小化は, P と Q の間のカルバック・ライブラー情報量の最小化と等価である (Mackay, 2003).

平均場近似では, 次に $Q(\mathbf{x};\theta)$ の形を, 計算量的に扱いやすいもの $\prod_i Q(x_i;\theta_i)$ に制限する. 簡単な代入計算により, 変分自由エネルギーは次のように表される：

[1] より正確には, イジングスピンモデルなどと呼ばれる.

$$F(\theta) = -\frac{\beta}{2}\sum_{ij} w_{ij}\bar{x}_i\bar{x}_j - \sum_i\big(-\sum_{x_i} Q(x_i;\theta_i)\log Q(x_i;\theta_i)\big). \quad (4)$$

静止条件より, 次のような平均場方程式が得られる：

$$\bar{x}_i = \frac{\sum_{x_i} x_i \exp\big(\beta x_i \sum_j w_{ij}\bar{x}_j\big)}{\sum_{x_i}\exp\big(\beta x_i \sum_j w_{ij}\bar{x}_j\big)}. \quad (5)$$

この方程式は, 現在の値で右辺を計算し左辺に代入するという繰り返し更新計算により, 解くことができる.

4 スピンモデルによる感情極性の判定

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取りやすい. 我々は, 各単語を電子とみなし, 単語の感情極性をスピンの向きとみなす. もしある単語が別の単語の語釈文に出現した場合にその二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. このようなモデル化の背後には, ある単語とその語釈文に出現する単語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.

4.1 語彙ネットワークの構築

ある単語が別の単語の語釈文に出現したらその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. 各リンクは二つのグループに分けられる：SL (同極性リンク集合) と DL (逆極性リンク集合). 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.

リンクの重み $W = (w_{ij})$ を次のように決める：

$$w_{ij} = \begin{cases} 1/\sqrt{d(i)d(j)} & (l_{ij} \in SL) \\ -1/\sqrt{d(i)d(j)} & (l_{ij} \in DL) \\ 0 & otherwise \end{cases}. \quad (6)$$

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. 連結されていなかった単語同士は, 重み 0 のリンクで連結されているとみなすことができる. 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, シソーラス上の同義語, 反義語, 上位語をリンクで連結することにより別のネットワーク *gloss-synonymy network (GS)* を構築する. 反義語のリンクは DL 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 節で述べたように, Hatzivassiloglou (1997) はコーパスに出現する接続表現を利用した. 同様に我々も二つの単語が and による接続表現で出現したらそのリンクを SL に, また, but による接続表現で出現したら DL に加える. このようにしてできたネットワークを *gloss-synonymy-corpus network (GSC)* と呼ぶことにする.

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. この少量のラベル付きデータを以下のように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：

$$H(\beta, \mathbf{x}, W) = -\frac{\beta}{2}\sum_{ij} w_{ij}x_i x_j + \alpha\sum_{i\in L}(x_i - a_i)^2, \quad (7)$$

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, 次の更新式が得られる：

$$\bar{x}_i^{new} = \frac{\sum_{x_i} x_i \exp\big(\beta x_i s_i^{old} - \alpha(x_i - a_i)^2\big)}{\sum_{x_i}\exp\big(\beta x_i s_i^{old} - \alpha(x_i - a_i)^2\big)}. \quad (8)$$

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. 新しい更新式は, Inoue (2001) の画像修復方法に従っている.

平均 $\bar{x}_i$ の初期値は, 初期単語についてはそのラベルに従い, 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充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し, 負だった単語は, 感情極性がネガティブと判定する.

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ので, その値を適切に決定する必要がある. ここでは, そのための基準として平均磁化 (magnetization) $m = (1/N)\sum_i \bar{x}_i$ を用いることを提案する.

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. これを相転位といい, その時の温度を臨界点という. 臨界点より僅かに高温であるとき, スピン系は局所的に秩序を持つが, 全体的には秩序的でない状態になる.

我々が望む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, 我々は臨界点より僅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 を用いる.

5 文の感情極性判定

前節で説明した手法により獲得された単語の感情極性を用いて, 文の感情極性を判定することを試みる. ここで試す手法は, 文を単語の集合とみなし, 各単語の感情極性値を足し合わせるというシンプルなものである. より厳密には以下のように定式化できる.

文 r の感情極性値 $g(r)$ を, 次のように計算する：

$$g(r) = \sum_{i\in r} u(i,r)t(i). \quad (9)$$

ここで, $i \in r$ は r に出現する各単語を表す. $t(i)$ は単語 i の感情極性を表す関数で, ここでは前節の手法で求めたスピンの平均値を用いて単純に $t(i) = \bar{x}_i$ とする. また, $u(i,r)$ は x の文 r における文脈を数値化したものである. ここでは単純に, 否定語のあとに出てきた場合は -1, そうでなければ 1 とする.

$g(r)$ が正ならば, 文 r はポジティブであると判定し, $g(r)$ が負ならば, 文 r はネガティブであると判定する.

ここで提案する手法は, 教師無し分類手法であることに注意されたい.

6 実験

提案手法の性能を, 実験により示す. まず単語の感情極性判定の実験について述べ, 最後に文の感情極性判定の実験について述べる.

英語の語彙ネットワークの構築には, WordNet (Fellbaum, 1998) の語釈文, 同義語, 反義語, 上位語を用いた. 語釈文に対する品詞タグ付けと原形化には, TreeTagger (Schmid, 1994) を用いた. 動詞, 名詞, 形容詞,

表 1: 予測された β を用いた場合の分類正解率 (%). 括弧内にはそのときの β の値が付与されている.

# seeds	GSC	GS	G
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)

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 33 個になった. ネットワークは約 88,000 単語から成る. また, コーパスからの接続表現としては, Wall Street Journal と Brown corpus から 804 個の表現を抽出した.

ラベル付きデータとして, Turney らの実験 (2003) に従い, General Inquirer (Stone et al., 1966) の語彙を使用した. "Positiv" もしくは "Negativ" のラベルがついている単語を General Inquirer から抽出し, 多義性により複数のエントリがある語は一つにまとめた. その結果, 3596 語のラベル付き単語集合 (ポジティブは 1616 語, ネガティブは 1980 語) が得られた.

実験は, 0.1 間隔で 0.1 から 2.0 までの異なる β の値で行い, 4.3 節で述べた方法を用いて最も良い値を予測した. α は, 充分大きな値 (1.0×10^3) に固定した.

初期単語集合としては Turney らが使用したものと同じものを用いる：

ポジティブ初期単語集合 = {good, nice, excellent, positive, fortunate, correct, superior},

ネガティブ初期単語集合 = {bad, nasty, poor, negative, unfortunate, wrong, inferior}.

岩波国語辞典 (西尾ら, 1994) を用いて, 日本語の語彙に対する実験も行った. 形態素解析には ChaSen を用いた. 英語の場合と同様に内容語のみを用いた. また, ここでの否定語は「ない」と「ぬ」のみである. 日本語に対しては, ネットワーク G のみを試した. 孤立した単語を削除すると, 58185 単語からなるネットワークが得られた.

正解として人手で 5632 語にタグ付けした. このうち, 2491 語がポジティブであり, 3141 語がネガティブである. 日本語の初期単語集合は「良い」と「悪い」の二語のみとして試してみた.

6.1 実験結果

まず, 英語の実験結果について述べたあと, 日本語の実験結果について簡単に付け加える. 結果は予測された β を用いた場合のものである.

表 1 は, 三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", そして "inferior" に対応する.

多くの場合, シソーラス及びコーパスが正解率を上昇させていることがわかる. 例外は, サイズ 2 の初期単語集合において G が GS を上回っているケースである. この理由は調査中であるが, 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.

この提案手法の結果を Turney らの結果と比較してみる. サイズ 14 の初期単語で彼らは, 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%の正解率を報告している. 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, 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, 我々の手法は大きなコーパスの Turney らの結果と同等な結果を出している.

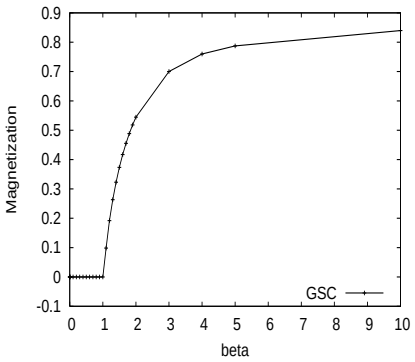

図 1: 平均磁化の例 (14 初期単語の場合).

提案手法が言語依存でないことを示すために, 我々は日本語での補足実験も行った. その結果, 79.8%の正解率を得ることに成功した.

6.2 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ために, 表 1 と表 2 を比較する. 表 2 では最適な値を用いた結果が示されている. 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. 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.

平均磁化の変化を表した図 1 も, 参考として加える. $\beta = 1.0$ 付近で相転位が起こっているのがわかる.

6.3 確信度が高い分類結果に対する精度

次に, 高い確信度をもって分類された単語に対する分類精度を算出する. 確信度としては, 平均値 $\bar{x}$ の絶対値を用いた. 絶対値の降順に単語を並べ, 上位の単語のみで分類精度を測った.

結果は図 2 に示す. 横軸は選ばれた上位単語数, 縦軸は分類精度である. 上位 1000 語で 92%を越える精度が得られた. 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.

Kamps ら (2004) は, 彼らの手法で "good" と "bad" を初期単語として用い, 667 語の形容詞について 67.32% という精度を報告している. 比較のために我々も上記確信度を用いて 667 語の形容詞を選び, 精度を測った (表 3). 提案手法が Kamps らの手法を上回っていることが示されている.

6.4 エラー解析

提案手法を用いた場合の分類誤りを調査し, 主に三種類の原因があることがわかった.

一つは語義の曖昧性の問題である. 例えば, "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関わらず, 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. 例えば, "arrogance" の語釈文は

表 2: 最も良い β を用いた場合の分類正解率 (%). 括弧内にはそのときの β の値が付与されている.

# seeds	GSC	GS	G
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)

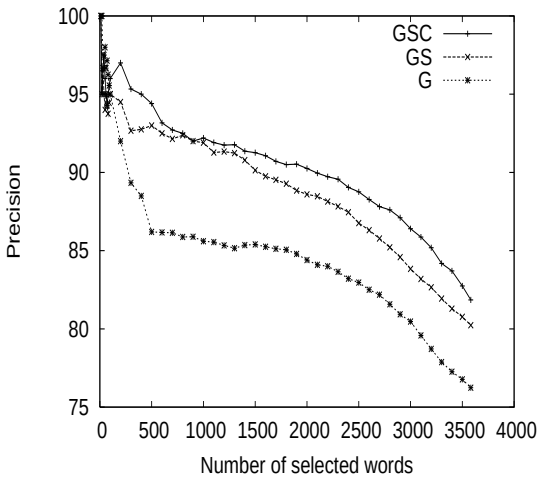

図 2: 14 初期単語のときの分類精度 (%).

表 3: 上位 667 語の形容詞に対する分類精度 (%).

# seeds	GSC	GS	G
2	94.5 (1.0)	76.3 (1.0)	86.7 (1.0)

"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. 例えば, "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, 正しく分類ができていなかった.

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予定である.

6.5 実験：文の感情極性判定

用いたデータは Hu ら (2004) の提供している製品レビューデータである[2]. これは, 5 種類の電化製品に対する消費者の感想などを Amazon.com と C|net.com から収集したものである. 14 初期単語の場合について, G, GS, GSC の三種類のネットワークをそれぞれ用いて求めた単語の極性によって文の感情極性を判定した. 比較のため Hu らの手法を実装し, 同じ条件で実験を行った. 結果は, 表 4 の通りである.

提案手法はどれも平均値において Hu らの手法を上回っている. 簡単な解析により正解率上昇の理由を調べてみると, 形容詞しか扱えない Hu らの手法は動詞を利用できないが, recommend などのポジティブな単語が我々の手法では効果的に働いていることがわかった.

表 4: 文の感情極性判定正解率 (%)

製品の種類	14 seed words			
	Hu's	G	GS	GSC
Digital camera 1	70.8	69.5	71.6	72.9
Digital camera 2	75.5	73.0	81.8	82.4
Cellular phone	66.7	67.4	73.6	74.8
MP3 player	61.0	65.7	70.3	72.8
DVD player	49.7	59.6	61.4	66.1
平均	64.7	67.0	71.7	73.8

[2] http://www.cs.uic.edu/~liub/FBS/FBS.html で入手可能.

7 結論と今後の課題

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みなし, 語釈文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. また, 提案手法が言語に依存しない一般的な手法であることも示した. また, 獲得された単語の感情極性を用いて, 文の感情極性を判定する実験も行い, 提案手法が優れていることを示した.

今後の課題としては, 語釈文の構文構造や語順情報の取り入れ, 多値のスピンモデルへの拡張, 能動学習による初期単語集合作成, より大きなコーパスの使用, Turney らの手法との融合, 辞書にない単語に対するオンデマンド極性付与方法の開発などが挙げられる. また, ここでの感情極性は Osgood (1957) の三因子のうちの evaluation 因子に対応するが, 残りの二つ (potency 因子と activity 因子) についても提案手法は応用可能である. それ以外にも, スピン系に類似した構造を背後に持つ概念は多く存在すると思われる.

参考文献

Fellbaum, C., 1998. *WordNet: An Electronic Lexical Database, Language, Speech, and Communication Series*. MIT Press.

Hatzivassiloglou, V., and McKeown, K., 1997. Predicting the semantic orientation of adjectives. *35th ACL*, pp. 174–181.

Hu, M., and Liu, B., 2004. Mining and summarizing customer reviews *KDD-2004*, pp. 168–177.

Inoue, J., and Carlucci, D., 2001. Image restoration using the q-ising spin glass. *Physical Review E*, 64:036121–1 – 036121–18.

Kamps, J., Marx, M., Mokken, R., and de Rijke, M., 2004. Using wordnet to measure semantic orientation of adjectives. *4th LREC, volume IV*, pp. 1115–1118.

Mackay, D., 2003. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H., 1957. *The Measurement of Meaning*. University of Illinois Press.

Riloff, E., Wiebe, J., and Wilson, T., 2003. Learning subjective nouns using extraction pattern bootstrapping. *7th CoNLL*, pp. 25–32.

Schmid, H., 1994. Probabilistic part-of-speech tagging using decision trees. *International Conference on New Methods in Language Processing*, pp. 44–49.

Stone, P., Dunphy, D., Smith, M., and Ogilvie, D., 1966. *The General Inquirer: A Computer Approach to Content Analysis*. The MIT Press.

Turney, P., and Littman, M., 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems*, 21(4):315–346.

小林のぞみ, 乾孝司, 乾健太郎. 2001. 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.

西尾実, 岩淵悦太郎, 水谷 静夫. 1994. 岩波国語辞典. 岩波書店.