

## スピンモデルを用いた単語の感情極性抽出における 精度改善法

後藤拓馬

東京工業大学総合理工学研究科 知能システム科学専攻

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一般の個人が情報を発信する機会が増加している。個人が発信するこれらの情報は、ある対象（サービスなど）に関するその人の評価などの個人の意見が記述される場合が多い。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近年活発化しており [1]、実際に、ブログやレビュー文などを対象とした評価情報の抽出システムが提案され [2]、また、サービスとしても提供されている [3]。

評価情報を抽出し、評価するシステムを構築する際の要素技術に、評価表現辞書の構築がある。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。構築された辞書の極性データを用いることで文の評価情報を抽出することが可能となる。例えば、池田ら [4] は評価表現辞書を基に単語極性反転モデルを用いることで高精度で評価文分類ができると報告している。

評価表現辞書を構築する方法の一つとして、語彙ネットワークを利用した手法がある [5, 6]。この手法では語釈文<sup>1</sup>、シソーラス<sup>2</sup>、コーパス<sup>3</sup>などの情報に基づいて、語彙をノードとする語彙ネットワークを構築する。その上で、評価極性情報をネットワークで伝搬させることにより、語彙ネットワーク上の全てのノード（語彙）の感情極性を求める。

高村ら [5] は語彙ネットワークを1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評価することで、単語の感情極性を求めている。熱平均値の計算には平均場近似が用いられている。

高村らの手法では、“温度”がシステムを特徴づける制御変数として導入されている。温度を十分高温から下げていった場合、この手法により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システムは強磁性相に転移し、感情極性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。

スピンシステムにおける相転移現象は相互作用行列の第1固有ベクトル方向へ熱揺らぎが不安定化することにより引き起こされる。語彙ネットワークモデルについて相互作用行列の第1固有ベクトルを数値的に評価してみると各成分は同程度の大きさでほぼ同符号であることが確認された。これらのことから、この第1固有ベクトルが高村らの既存

<sup>1</sup>辞書にある見出し単語の説明文

<sup>2</sup>上位/下位関係、部分/全体関係、同義関係、類義関係などによって単語を分類し、体系付けた辞書

<sup>3</sup>新聞記事、小説、ウェブテキストなどの実際の文章データを収集し、蓄積した言語資源

手法を用いた際にモデルが強磁性相に転移し精度があったした要因であると推測される。そこで、本研究では、既存手法のモデルから第1固有ベクトルの要素を取り除き、システム全体が強磁性相にならないようなモデルに再構成し、精度の改善を試みた。その結果、同程度の必要計算時間で高村らの既存手法より高い鮮度で単語の感情極性が抽出でき、評価表現辞書が構築することが可能となった。

本講演では、スピンモデルを用いた評価表現辞書構築法の概略を説明し、その後、我々の提案手法を紹介する。

## 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] Sasha Blair-Goldensohn, Kerry Hannan Ryan McDonald Tyler Neylon George A. and Reis Jeff Reynar, “Building a Sentiment Summarizer for Local Service Reviews”, In *WWW Workshop on NLP Challenges in the Information Explosion Era(NLPIX)*, 2008.
- [3] goo ブログ評判分析, <http://blog-hyoban.goo.ne.jp>
- [4] 池田 大介, 高村 大也, 奥村 学, “単語極性反転モデルによる評価文分類”, 人工知能学会論文誌, vol.25, no.1, pp.50–57, 2010.
- [5] Hiroya Takamura, Takashi Inui and Manabu Okumura, “Extracting Semantic Orientations of Words using Spin Model”, In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL2005)*, pp.133–140, 2005.
- [6] Lenid Velikovich, Sasha Blair-Goldensohn, Kerry Hannan, Ryan McDonald, “The Viability of Web-derived Polarity Lexicons”, In *Proceedings of The Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp.777–785, 2010.