

テキストからの評判分析と機械学習

鍛治伸裕

東京大学 生産技術研究所

kaji@tkl.iis.u-tokyo.ac.jp

近年の CGM¹の爆発的な普及に伴って、テキストデータから特定の製品や組織などに対する評判を自動的に抽出して集約するための技術が注目を集めている。こうした技術は評判分析や Sentiment Analysis などと呼ばれ、人工知能や自然言語処理などの分野において盛んに研究が行われている。この数年の間に、AAAI や ACL といった最高峰の国際会議のシンポジウムやワークショップとして評判分析を扱う会議が開催されていることから、このトピックへの関心の高まりを見て取ることができる [4, 1]。また、国内においても、論文誌に解説論文が掲載されるなど注目を集める分野となっている [7]。

評判分析における初期の研究としては、Hatzivassiloglou と McKeown の論文が有名である [2]。これは、単語が持つ評価極性²という性質を議論した先駆的な論文であり、語彙ネットワークを用いて形容詞の評価極性を判定する方法が提案されている。この5年後、レビュー文書を肯定的内容と否定的内容に分類する方法を提案した論文が、複数の著者によって続けて発表された [6, 3]。これをきっかけに、評判分析を分類問題として定式化して、機械学習を用いて解くという考え方が広く知られることになった。自然言語処理全般における機械学習の成功や、冒頭で述べたような社会的背景の後押しもあり、これ以降、評判分析は自然言語処理におけるメイントピックの一つへの成長した。

近年では、評判分析が単純な分類問題として議論されることは少なくなってきたおり、研究者の興味はより複雑で現実的な問題設定へと移りつつある。例えば、複数レビュー文書からの要約生成のために、評判の aspect を抽出する問題などはその一例である [5]。これに伴って、そこに適用される機械学習手法も、構造予測や分野適応やベイズモデルなど多様化を見せてい

る。本講演では、こうした評判分析の研究の流れを概観しながら、機械学習を応用したアプローチを中心として、最近の評判分析の研究動向の紹介を行う。

参考文献

- [1] Michael Gamon and Anthony Aue, editors. *Proceedings of the Workshop on Sentiment and Subjectivity in Text*. Association for Computational Linguistics, Sydney, Australia, 2006.
- [2] Vasileios Hatzivassiloglou and Katherine R. McKeown. Predicting the semantic orientation of adjectives. In *Proceedings of ACL*, pp. 174–181, Madrid, Spain, 1997.
- [3] Bo Pang, Lillian Lee, and Shivakumar Vaihyathan. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the EMNLP*, pp. 79–86, Philadelphia, Pennsylvania, USA, 2002.
- [4] James G. Shanahan, Yan Qu, and Janyce Wiebe, editors. *Computing Attitude and Affect in Text: Theory and Applications*. Springer, 2006.
- [5] Ivan Titov and Ryan McDonald. A joint model of text and aspect ratings for sentiment summarization. In *Proceedings of the ACL*, pp. 308–316, Columbus, Ohio, USA, 2008.
- [6] Peter D. Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the ACL*, pp. 417–424, Philadelphia, Pennsylvania, USA, 2002.
- [7] 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.

¹Consumer Generated Media のこと。ここではアマゾンなどのレビューデータやブログデータのことを指している。

²肯定的, 否定的という評価に関する軸のこと。英語では Semantic Orientation または Polarity と呼ばれる。