

Annotated Bibliography

Anany Sachan

September 4, 2025

References

- [1] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, EMNLP '02, pages 79–86, USA, 2002. Association for Computational Linguistics.

Pang, Lee, and Vaithyanathan (2002) investigate sentiment classification, aiming to categorize documents by overall opinion rather than topic. Using a large dataset of movie reviews, they compare Naive Bayes, maximum entropy, and support vector machines, finding that these algorithms significantly outperform human-selected word lists but still achieve lower accuracy than in topic-based classification. The paper highlights challenges unique to sentiment analysis, including subtle language cues, rhetorical contrasts, and the need for contextual understanding (e.g., negation). Published in the *Proceedings of EMNLP* and widely cited, the work is highly credible and foundational to sentiment analysis research. While the paper’s models may seem simple by today’s standards, they illustrate enduring concepts such as feature representation (bag-of-words, unigrams vs. bigrams) and classification tradeoffs. As a cornerstone in the field, this source highlights the early challenges of sentiment analysis and sets the stage for evaluating modern neural and transformer-based approaches.