

## Good debt or bad debt: Detecting semantic orientations in economic texts

Pekka Malo , Ankur Sinha , Pekka Korhonen , Jyrki Wallenius , Pyry Takala 

First published: 26 November 2013

<https://doi.org/10.1002/asi.23062>



PDF

### Abstract

The use of robo-readers to analyze news texts is an emerging technology trend in computational finance. Recent research has developed sophisticated financial polarity lexicons for investigating how financial sentiments relate to future company performance. However, based on experience from fields that commonly analyze sentiment, it is well known that the overall semantic orientation of a sentence may differ from that of individual words. This article investigates how semantic orientations can be better detected in financial and economic news by accommodating the overall phrase-structure information and domain-specific use of language. Our three main contributions are the following: (a) a human-annotated finance phrase bank that can be used for training and evaluating alternative models; (b) a technique to enhance financial lexicons with attributes that help to identify expected direction of events that affect sentiment; and (c) a linearized phrase-structure model for detecting contextual semantic orientations in economic texts. The relevance of the newly added lexicon features and the benefit of using the proposed learning algorithm are demonstrated in a comparative study against general sentiment models as well as the popular word frequency models used in recent financial studies. The proposed framework is parsimonious and avoids the explosion in feature space caused by the use of conventional n-gram features.

### References

Antweiler, W., & Frank, M. (2004). Is all that talk just noise? The information content of Internet stock message

This website utilizes technologies such as cookies to enable essential site functionality, as well as for analytics, personalization, and targeted advertising. You may change your settings at any time or accept the default settings. You may close this banner to continue with only essential cookies. [Privacy Policy](#)

Manage Preferences

Accept All

Reject Non-Essential

---

Barber, B., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies*, 21, 758-818.

[Web of Science®](#) | [Google Scholar](#)

---

Bollen, J., Gonçalves, B., Ruan, G., & Mao, H. (2011). Happiness is assortative in online social networks. *Artificial Life*, 17(2), 237-251.

[PubMed](#) | [Web of Science®](#) | [Google Scholar](#)

---

Chomsky, N. (1957). *Syntactic structures*. The Hague, The Netherlands: Mouton & Co.

[Google Scholar](#)

---

Das, S., & Chen, M. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science*, 53(9), 1375-1388.

[Web of Science®](#) | [Google Scholar](#)

---

Davis, A., Piger, J., & Sedor, L. (2012). Beyond the numbers: Measuring the information content of earnings press release language. *Contemporary Accounting Research*, 29(3), 845-868.

[Web of Science®](#) | [Google Scholar](#)

---

Demers, E., & Vega, C. (2008). Soft information in earnings announcements: News or noise? SSRN Working Paper.

[Google Scholar](#)

---

Dzielinski, M., Rieger, M., & Talpsepp, T. 2011. Volatility asymmetry, news, and private investors. In G. Mitra & L. Mitra (Eds.), *The Handbook of news analytics in finance*. Hoboken, NJ: John Wiley & Sons.

[Google Scholar](#)

This website utilizes technologies such as cookies to enable essential site functionality, as well as for analytics, personalization, and targeted advertising. You may change your settings at any time or accept the default settings. You may close this banner to continue with only essential cookies. [Privacy Policy](#)

Manage Preferences

Accept All

Reject Non-Essential

[Web of Science®](#) 

[Google Scholar](#) 

---

Gabrilovich, E., & Markovitch, S. (2010). Wikipedia-based semantic interpretation for natural language processing. *Journal of Artificial Intelligence Research*, 34, 443–498.

[Web of Science®](#) 

[Google Scholar](#) 

---

Hatzivassiloglou, V., & McKeown, K. (1997). Predicting the semantic orientation of adjectives. In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL-97) (pp. 174–181). Madrid, Spain.

[Google Scholar](#) 

---

Hsu, C.-W., & Lin, C.-J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(2), 415–425.

[PubMed](#) 

[Google Scholar](#) 

---

Li, F. (2009). The information content of the forward-looking statements in corporate filings: A naïve Bayesian machine learning approach. Working paper.

[Google Scholar](#) 

---

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries and 10-Ks. *Journal of Finance*, 66(1), 35–66.

[Web of Science®](#) 

[Google Scholar](#) 

---

Maks, I., & Vossen, P. (2010). Annotation scheme and gold standard for Dutch subjective adjectives. In Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC-10). European Language Resources Association (ELRA). Malta.

[Google Scholar](#) 

---

Malo, P., Siitari, P., & Sinha, A. (2013). Automated query learning with Wikipedia and genetic programming.

This website utilizes technologies such as cookies to enable essential site functionality, as well as for analytics, personalization, and targeted advertising. You may change your settings at any time or accept the default settings. You may close this banner to continue with only essential cookies. [Privacy Policy](#)

[Manage Preferences](#)

[Accept All](#)

[Reject Non-Essential](#)

Manning, C., & Schütze, H. (2003). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

[Google Scholar](#) 

Mishne, G. (2005). Experiments with mood classification in Blog posts. In Proceedings of the 1st Workshop on Stylistic Analysis of Text for Information Access at SIGIR 2005. Salvador, Brazil: ACM.

[Google Scholar](#) 

Mishne, G., & de Rijke, M. (2006). Capturing global mood levels using Blog posts. In Proceedings of the AAAI Spring Symposium on Computational Approaches to Analysing Weblogs (pp. 145-152). Palo Alto, CA: AAAI Press.

[Google Scholar](#) 

Mitra, L., & Mitra, G. (2010). Applications of news analytics in finance: A review. Middlesex, United Kingdom: OptiRisk Systems. Retrieved from <http://optirisk-systems.com/papers/Opt0014.pdf>.

[Google Scholar](#) 

Mitra, L., & Mitra, G. (2011). *The handbook of news analysis in finance*. Hoboken, NJ: John Wiley & Sons.

[Google Scholar](#) 

Mitra, L., Mitra, G., & diBartolomeo, D. (2009). Equity portfolio risk (volatility) estimation using market information and sentiment. *Quantitative Finance*, 9(8), 887-895.

[Web of Science®](#)  | [Google Scholar](#) 

Mitrovic, M., Paltoglou, G., & Tadic, B. (2011). Networks and emotion-driven user communities at popular blogs. *The European Physical Journal B*, 77(4), 597-609.

[CAS](#)  | [Web of Science®](#)  | [Google Scholar](#) 

This website utilizes technologies such as cookies to enable essential site functionality, as well as for analytics, personalization, and targeted advertising. You may change your settings at any time or accept the default settings. You may close this banner to continue with only essential cookies. [Privacy Policy](#)

Manage Preferences

Accept All

Reject Non-Essential

[Google Scholar](#) 

---

O'Hare, N., Davy, M., Bermingham, A., Ferguson, P., Sheridan, P., Gurrin, C., & Smeaton, A.F. (2009). Topic-dependent sentiment analysis of financial blogs. In Proceedings of the 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion (pp. 9–16). Hong Kong, China: ACM.

[Google Scholar](#) 

---

Pang, B., Lee, L., & Vaithyanathan, S. (2002, July). Thumbs up? Sentiment classification using machine learning techniques. In Conference on Empirical Methods in Natural Language Processing (EMNLP 2002) (pp. 79–86). Philadelphia, PA.

[Google Scholar](#) 

---

Polanyi, L., & Zaenen, A. (2004). Contextual valence shifters. In working notes of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications (pp. 106–111). Dordrecht, The Netherlands: Springer.

[Google Scholar](#) 

---

Riloff, E., & Wiebe, J. 2003. Learning extraction patterns for subjective expressions. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2003) (pp. 105–112). Sapporo, Japan: ACL.

[Google Scholar](#) 

---

Robinson, W. (1957). The statistical measurement of agreement. *American Sociological Review*, 22, 17–25.

[Web of Science®](#)  | [Google Scholar](#) 

---

Shrout, P., & Fleiss, J. (1979). Intraclass correlation: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428.

[CAS](#)  | [PubMed](#)  | [Web of Science®](#)  | [Google Scholar](#) 

This website utilizes technologies such as cookies to enable essential site functionality, as well as for analytics, personalization, and targeted advertising. You may change your settings at any time or accept the default settings. You may close this banner to continue with only essential cookies. [Privacy Policy](#) 

Manage Preferences

Accept All

Reject Non-Essential

---

Tetlock, P., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance*, 63, 1437-1467.

[Web of Science®](#) | [Google Scholar](#)

---

Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social Web. *Journal of the American Society for Information Science and Technology*, 63(1), 163-173.

[Web of Science®](#) | [Google Scholar](#)

---

Thelwall, M., Buckley, K., Paltoglou, G., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544-2558.

[Web of Science®](#) | [Google Scholar](#)

---

Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39, 165-210.

[Web of Science®](#) | [Google Scholar](#)

---

Wilson, T. (2008). *Fine-grained subjectivity analysis (Unpublished doctoral dissertation)*. University of Pittsburgh, Pittsburgh, PA.

[Google Scholar](#)

---

Wilson, T., Wiebe, J., & Hoffman, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, 35(3), 399-433.

[Web of Science®](#) | [Google Scholar](#)

---

Young, L., & Soroka, S. (2012). Affective news: The automated coding of sentiment in political texts. *Political Communication*, 29, 205-231.

[Web of Science®](#) | [Google Scholar](#)

This website utilizes technologies such as cookies to enable essential site functionality, as well as for analytics, personalization, and targeted advertising. You may change your settings at any time or accept the default settings. You may close this banner to continue with only essential cookies. [Privacy Policy](#)

Manage Preferences

Accept All

Reject Non-Essential

**ABOUT WILEY ONLINE LIBRARY**

Privacy Policy

Terms of Use

About Cookies

Manage Cookies

Accessibility

Wiley Research DE&I Statement and Publishing Policies

**HELP & SUPPORT**

Contact Us

Training and Support

DMCA & Reporting Piracy

Sitemap

**OPPORTUNITIES**

Subscription Agents

Advertisers & Corporate Partners

**CONNECT WITH WILEY**

The Wiley Network

Wiley Press Room

This website utilizes technologies such as cookies to enable essential site functionality, as well as for analytics, personalization, and targeted advertising. You may change your settings at any time or accept the default settings. You may close this banner to continue with only essential cookies. [Privacy Policy](#)



**Manage Preferences**

**Accept All**

**Reject Non-Essential**