

Text Mining for the Social Sciences Reading List

Textbooks

- Manning, Raghavan, and Shütze (2009) (MRS)
 - Murphy (2012) (KM)
-

General overview

- Grimmer and Stewart (2013)

Text processing

- MRS 2.2

Count-based methods

- MRS 1, 6.1-6.3
- Baker, Bloom, and Davis (2013)
- Tetlock (2007)
- Tuckett, Ormerod, Nyman, and Smith (2015)
- Loughran and McDonald (2011)
- Hoberg and Phillips (2010)

Latent Semantic Analysis

- MRS 18
- Deerwester, Dumais, Furnas, Landauer, and Harshman (1990)
- Acosta (2014)
- Schonhardt-Bailey (2013)

Probability models for discrete data

- KM 2.5.4, 3.3-3.4, 27.1-27.3.2

LDA and collapsed Gibbs sampling

- KM 27.3.1-27.3.6
- Heinrich (2009)

Variational Inference

- Blei, Ng, and Jordan (2003)
- KM 21

References

- ACOSTA, J. M. (2014): “FOMC Responses to Calls for Transparency: Evidence from the Minutes and Transcripts Using Latent Semantic Analysis,” Mimeograph, University of Stanford.
- BAKER, S. R., N. BLOOM, AND S. J. DAVIS (2013): “Measuring Economic Policy Uncertainty,” Mimeograph, Stanford University.
- BLEI, D. M., A. Y. NG, AND M. I. JORDAN (2003): “Latent Dirichlet Allocation,” *Journal of Machine Learning Research*, 3, 993–1022.
- DEERWESTER, S., S. T. DUMAIS, G. W. FURNAS, T. K. LANDAUER, AND R. HARSHMAN (1990): “Indexing by latent semantic analysis,” *Journal of the American Society for Information Science*, 41(6), 391–407.
- GRIMMER, J., AND B. STEWART (2013): “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts,” *Political Analysis*, pp. 1–31.
- HEINRICH, G. (2009): “Parameter estimation for text analysis,” Technical report, vsonix GmbH and University of Leipzig.
- HOBURG, G., AND G. PHILLIPS (2010): “Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis,” *Review of Financial Studies*, 23(10), 3773–3811.
- LOUGHRAN, T., AND B. MCDONALD (2011): “When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10Ks,” *Journal of Finance*, 66(1), 35–65.
- MANNING, C. D., P. RAGHAVAN, AND H. SHÜTZE (2009): *An Introduction to Information Retrieval*. Cambridge University Press.
- MURPHY, K. P. (2012): *Machine Learning: A Probabilistic Perspective*, Adaptive Computation and Machine Learning. MIT Press.
- SCHONHARDT-BAILEY, C. (2013): *Deliberating Monetary Policy*. MIT Press, Cambridge.
- TETLOCK, P. C. (2007): “Giving Content to Investor Sentiment: The Role of Media in the Stock Market,” *Journal of Finance*, 62(3), 1139–1168.
- TUCKETT, D., P. ORMEROD, R. NYMAN, AND R. E. SMITH (2015): “Improving Economic Prediction: A New Method for Measuring Economic Confidence and its Impact on the Evolution of the US Economy,” unpublished manuscript, University College London.