

Latent Dimensions of Religiousness and Spirituality: A Longitudinal Correlated Topic Model

S.H. Kim¹, N.R. Strenger², and N. Lee³

¹Fuller Graduate School of Psychology, Pasadena, California, USA; shkim@fuller.edu

²Pastoral Counseling Center, Dallas, Texas, USA

³Fuller Graduate School of Psychology, Pasadena, California, USA

Abstract

The current research focuses on the variability and correlations of topic proportions in psychological text data. Topic proportions in text data are typical compositional data in terms that their components are “strictly positive real numbers and carry only relative information” (p. 8, Pawlowsky-Glahn, Egozcue, and Tolosana-Delgado, 2015). We collected and analyzed thousands of items in about 230 psychological measures of religiousness and spirituality (R/S) to identify the latent domains (i.e., topics) of R/S. We used structural topic modeling (STM; Roberts and others, 2013), an evolved development over correlated topic modeling (CTM; Blei & Lafferty, 2007). Topic models are used to “discover the hidden thematic structure in large archives of documents” (p. 1; Blei, 2012) and was originally invented in the fields of machine learning and computer science as a way of locating latent themes in a collection of text data. CTM, while based on latent Dirichlet allocation (LDA; Blei and others, 2003), directly model the variability and potential correlation between topic proportions (i.e., compositional data) in text data by adopting the Aitchison’s logistic normal distribution (Aitchison, 1982). LDA assumes that topic proportions across documents are randomly drawn from a Dirichlet distribution, which is inadequate to model the variability of topic proportions (Aitchison, 1986).

The R statistical environment (R Core Team, 2016) was used for data analysis. The collected text data was first lemmatized using the TreeTagger software (Schmid, 1994) and the koRpus package (Michalke, 2016) in R. Next, the ‘tm’ package (Feinerer and Hornik, 2015) was used to pre-process (e.g., removing punctuations and stop words) the lemmatized text documents and to create a document-term matrix as an input for topic modeling analysis. The document-term matrix as well as the time covariate of year was used as the input for the ‘stm’ package (Roberts, Stewart, and Tingley, 2016) to fit a structural topic model. The topic proportions were further plotted and analyzed with the ‘compositions’ package (van den Boogaart, Tolosana, & Bren, 2014).

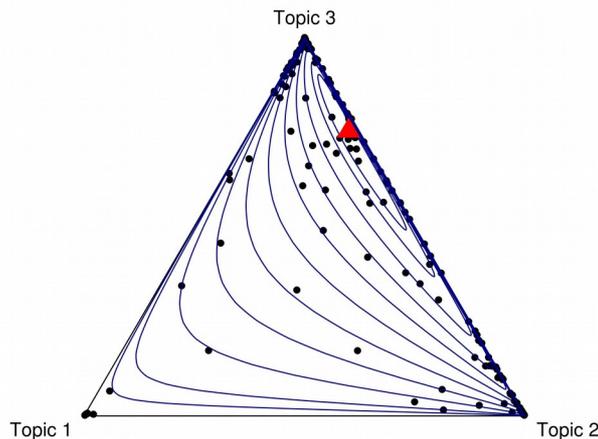


Figure 1. A ternary diagram of the three topics. Each dot reflects the topic proportion of each document (213 documents in total). The means (as represented with a red triangle in the plot) of the topic proportions are .024, .224, and .752 for Topics 1, 2, and 3 respectively, indicating the prevalence of Topic 3 over the other two.

References

- Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 139–177.
- Aitchison, J. (1986). *The statistical analysis of compositional data*. Chapman Hall, Ltd.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77–84.
- Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of Science. *The Annals of Applied Statistics*, 1(1), 17–35. <https://doi.org/10.1214/07-AOAS114>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(1), 993–1022.
- Feinerer, I. & Hornik, K. (2015). tm: Text Mining Package (Version 0.6-2) [Computer software]. Retrieved from <https://CRAN.R-project.org/package=tm>.
- Michalke, M. (2016). koRpus: An R Package for Text Analysis (Version 0.06-5) [Computer software]. Retrieved from <https://CRAN.R-project.org/package=koRpus>.
- Pawlowsky-Glahn, V., Egozcue, J. J., & Tolosana-Delgado, R. (2015). *Modeling and analysis of compositional data*. West Sussex, UK: Wiley.
- R Core Team (2016). R: A language and environment for statistical computing (Version 3.3.2) [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>.
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2016). stm: R Package for Structural Topic Models (Version 1.1.3) [Computer software]. Retrieved from <http://www.structuraltopicmodel.com>.
- Roberts, M., Stewart, B., Tingley, D., & Airoldi, E. (2013). The structural topic model and applied social science.
- Schmid, H. (1994). *Probabilistic part-of-speech tagging using decision trees*. Presented at the International Conference on New Methods in Language Processing, Manchester, UK.
- van den Boogaart, K. G., Tolosana, R., & Bren, M. (2014). compositions: R Package for Compositional Data Analysis (Version 1.40-1) [Computer software]. Retrieved from <https://cran.r-project.org/web/packages/compositions/index.html>