Computational Comparative Religion: Methods for Cross-Cultural Analyses

Zachary K. Stine[0000-0001-5211-0111]

University of Arkansas at Little Rock, Little Rock AR 72204, USA
zkstine@ualr.edu

1 Motivation

In recent years we have seen the widespread adoption of various computational methods to further the study of human cultural and social systems. Typically, such research has been taken up by those in the social sciences as "computational social science," by humanities scholars as "digital humanities," and by researchers of various fields as "cultural analytics." The latter phrase has gained increasing usage as a broader, transdisciplinary term for quantitative studies of culture via the analysis of large cultural datasets [1]. While the emergence of cultural analytics has seen the implementation of computational methods to the study of human cultures in a variety of contexts, such analyses have not seen widespread adoption within the context of religious studies [2].

The development of computational methods that are informed by the theoretical lenses of religious studies are important for two reasons. First, the concerns of religious studies deal directly with issues important for our time. Religion is not disappearing from the modern world, but rather is undergoing significant changes [3]. This is especially observable in the digital world of social media. Understanding the formation of social identities, adoption of beliefs, etc. is incredibly important in our increasingly digital worlds. Second, the computational operationalization of religious studies concepts has broad application to the study of social and cultural systems beyond purely religious dimensions.

We will develop methodologies in order to carry out comparisons among a set of religious communities, each represented by a corpus of discussion text, along three dimensions: (1) a thematic dimension identifying the discursive concerns which bridge and distinguish the communities; (2) a dynamic dimension characterizing how each community explores an ideological space over time; and (3) a structural dimension that aims to uncover cultural correspondences among communities that may exist despite superficial differences between them. The latter dimension is based on the distinction between cultural lexicons and grammars (see [4] and [5]), which we aim to operationalize.

2 Significance

Given the lack of computational methods being used to address questions originating from a comparative religion context, this work is significant by attempting to develop

methods specifically tailored for this type of questioning. Additionally, the operationalization of lexical and grammatical cultural differences represents the excavation of deeper layers of meaning within text than what is available through current methods. Such deeper layers, if successful, would represent something closer to latent worldviews than latent topics.

3 Background

The primary natural language processing algorithm we will use is topic modeling: an unsupervised machine learning approach to text that reduces high-dimensional textual data into meaningful lower-dimensional latent language patterns or "topics." We use the popular topic modeling algorithm, latent Dirichlet allocation (LDA) [6]. The topics inferred by LDA may be viewed as operationalizing several important concepts related to culture including framing, polysemy, and a relational approach to meaning [7].

Because the representation of text after training a topic model takes the form of a topic distribution, various measures from information theory have been shown to be useful in interrogating the lower-dimensional text representations. Information theoretic measures have previously been used alongside topic models to compare the ideas explored over time between academic conferences [8], describe the information-seeking behavior of Charles Darwin [9], and characterize political innovation during the French Revolution [10].

4 Approach

The primary data source used consists of discussion text from a wide array of online religious communities from the popular discussion platform, Reddit [11]. Currently, we have collected discussion text from twenty religion-oriented Reddit communities, called subreddits, and are continuing to collect more data.

Topic models will be trained on various combinations of the religious communities' discussion text, each motivated by the nature of the comparison being made. For two communities that have significant lexical overlap (e.g., two communities that identify with one of the Buddhist traditions), the text of each may be combined into a single corpus for training a topic model. Topic distances can then be computed between the two communities using the Jensen-Shannon divergence. The discourses which most distinguish and bridge the two communities may then be identified based on which topics contribute the most and least to this distance.

In order to compare the ideological breadth or focus of each community, the Kullback-Leibler divergence will be computed for each post relative to the posts which precede and follow it over time (see [9] and [10]). Using a post's score (assigned via community members) will help further determine how much a community values the introduction of novel ideas versus the repetition of primary themes.

Multiple candidate methods for calculating differences along the third, structural dimension are currently being explored. Our goal is to find correspondences between

dimensions of different topic models trained on discussion texts from different communities. Direct comparisons cannot be made because each corpus defines its own semantic space, therefore necessitating indirect comparisons.

Due to the abstract nature of the concepts being operationalized, the evaluation of this work will require sophistication and subject matter expertise. Ultimately, the methods we propose must be evaluated in terms of how useful they are, whether their interpretation leads to theoretically justifiable insights, and through the careful construction of special null cases that provide a sense of relative baseline for all comparisons made. Thus, this work is transdisciplinary, requiring collaboration from scholars fluent in the theoretical frameworks of religious studies. Close readings of the data through case studies will also provide qualitative validation of meaningful relationships between quantitative results and the concepts they purport to measure.

References

1. Manovich, L.: The science of culture? Social computing, digital humanities and cultural analytics. CA. (2016). https://doi.org/10.22148/16.004.
2. Lövheim, M., Campbell, H.A.: Considering critical methods and theoretical lenses in digital religion studies. New Media & Society. 19, 5–14 (2017). https://doi.org/10.1177/1461444816649911.
3. Campbell, H.: Making space for religion in Internet studies. The Information Society. 21, 309–315 (2005).
4. Prothero, S.: Henry Steel Olcott and "Protestant Buddhism." Journal of the American Academy of Religion. 63, 281–302 (1995).
5. Deitrick, J.E.: Engaged Buddhist Ethics: Mistaking the Boat for the Shore. In: Queen, C.S., Prebish, C.S., Keown, D. (eds.) Action Dharma: New Studies in Engaged Buddhism. pp. 252–269. Psychology Press (2003).
6. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. Journal of Machine Learning Research 3, 993–1022 (2003).
7. DiMaggio, P., Nag, M., Blei, D.: Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. Poetics. 41, 570–606 (2013). https://doi.org/10.1016/j.poetic.2013.08.004.
8. Hall, D., Jurafsky, D., Manning, C.D.: Studying the History of Ideas Using Topic Models. In: Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing. pp. 363–371. Association for Computational Linguistics, Honolulu, Hawaii (2008).
9. Murdock, J., Allen, C., DeDeo, S.: Exploration and exploitation of Victorian science in Darwin's reading notebooks. Cognition. 159, 117–126 (2017). https://doi.org/10.1016/j.cognition.2016.11.012.
10. Barron, A.T.J., Huang, J., Spang, R.L., DeDeo, S.: Individuals, institutions, and innovation in the debates of the French Revolution. PNAS. 115, 4607–4612 (2018). https://doi.org/10.1073/pnas.1717729115.
11. Reddit, https://www.reddit.com, last accessed 2019/5/16.