Computational Methods for Comparing Discourses

Zachary K. Stine[1], James E. Deitrick[2], & Nitin Agarwal[1]
zkstine@ualr.edu, deitrik@uca.edu, nxagarwal@ualr.edu
[1] Department of Information Science, University of Arkansas at Little Rock
[2] Department of Philosophy and Religion, University of Central Arkansas

Overview

- Discourses around sociocultural identities comprise two levels of language usage:
  - Deliberately used language that ties the discourse to an identity (cultural lexicon).
  - Language that reflects the broader perspectives encoded within the discourse (cultural grammar).
- In comparative studies, deeper similarities and differences between discourses must be identified beneath the surface of superficial distinctions, often in a highly subjective way.
- **Goal**: use computational methods to compare these deeper, structural aspects of discourses in a rigorous way that mitigates pitfalls associated with comparative studies.
- The discourses we analyze are the discussion histories from several English-language Reddit communities devoted to discussing religious and spiritual identities.

Method 1. Comparing how discourses conceptualize each other

- Train topic models on each discourse separately and measure the shared consistency with which the discourse-specific models apply their topics to each other's text.
- Measures how similar discourses are in terms of how they organize information and allows for mappings between discourse-specific topics that suggest structural similarity despite surface-level distinctions.

r/**Buddhism** topic	r/**Christianity** topic	Pointwise Mutual Info
B16 Relationships	C15 Relationships	3.095
B24 Vegetarianism	C18 Abortion	2.797
B05 Repeated Text	C27 Repeated Text: Moderators	2.761
B05 Repeated Text	C10 Repeated Text: Verse Bot	2.743
B21 Intl. Politics, Conflict	C08 US Politics, Race	2.670
B12 Text Quotes	C23 Bible Verses	2.665
B25 Precepts	C25 Sex & Morality	2.617

Suggests that, beneath the surface-level differences between the topics of vegetarianism and abortion, these two topics share structural similarities across discourses. While these two topics are marked by distinctive identity-specific concerns, the ways in which those concerns are discussed are quite similar.

ZK Stine, JE Deitrick, N Agarwal. 2020. Comparative religion, topic models, and conceptualization. *CHR 2020* [link]

Method 2. Removing discourse-specific terms for comparisons

- Identify and remove distinguishing terms that strongly signal one discourse relative to the others using word-level information divergence. Distributional semantic structure of discourses is resilient to such terms being removed.
- Comparisons of discourses within topic spaces learned with and without distinguishing terms reveals interesting relationships between discourses beneath the surface of identity-specific language.

Examples of most distinguishing terms of discourses:
r/Buddhism: *buddhism buddha buddhist practice suffering meditation dhamma*
r/Christianity: *god jesus church bible christ christian sin christians christianity*
r/Hinduism: *hinduism hindu krishna shiva gita hindus brahman vedas vishnu*
r/Pagan: *pagan paganism gods pagans celtic wicca deities norse goddess*
r/Spirituality: *life spiritual love feel energy spirituality yourself soul self feeling*

ZK Stine, JE Deitrick, N Agarwal. 2021. Using Information Divergence to Differentiate Deep from Superficial Resemblances among Discourses. *C&C 2021*. [link]

Publications

ZK Stine, JE Deitrick, N Agarwal. 2020. Comparative religion, topic models, and conceptualization: Towards the characterization of structural relationships between online religious discourses *CHR 2020*. [link]

ZK Stine, JE Deitrick, N Agarwal. 2021. Using Information Divergence to Differentiate Deep from Superficial Resemblances among Discourses. *C&C 2021 (part of HCII 2021)*. [link]