Playing with Time

Jeri Wieringa
jwiering@gmu.edu
George Mason University, United States of America

The question of change over time is central for historians as they trace the contours of people, communities, organizations, and states. But computationally tracing changes in discourse over time is not a straightforward process with our current algorithmic tools and methods. While charting topics over time can give a general picture of the prevalence of ideas or discourses at various points, Benjamin Schmidt and others have raised concerns regarding the ways existing tools and methods model change over time. Most text-mining algorithms assume that time is linear and progressive, is experienced as such, and that the data is generally consistent over time. As Schmidt notes, even models such as Topics over Time, which was developed to account for shifts in language, are problematic because of the assumptions they make regarding how language changes (Schmidt 2012, see also Underwood, 2012, and Nowviskie, 2016).

If we allow that time is relative and that the experience of time shapes the ways a community and its discourses develop, we are faced with the challenge of how to incorporate varied experiences of time into our computational models. This poster will present my preliminary results applying various periodization schemes to track the development of Seventh-day Adventist discourses around salvation and health over the first 70 years of the denomination's existence.

Seventh-day Adventism is an apocalyptic and millennialist belief system, in that followers anticipate the imminent return of Jesus Christ and the corresponding end of the world. Born out of William Miller's teaching that 1843 (later 1844) would be the date of Christ's return, early Seventh-day Adventists reinterpreted the date in the wake of the Great Disappointment (the period after October 22, 1844, the date the Millerites believed Jesus would return. to signify the start of a new, final, and assumedly short phase in the work of salvation). They also adopted Sabbatarianism, holding Saturday, rather than Sunday, as the proper day for Christian observance. As such, early adherents operated within a changing temporal imaginary, organizing their weeks and years in contrast to their religious neighbors and anticipating, with varying degrees of urgency at different points in their history, the second coming.

The changing constructions of time within Seventh-day Adventism provide an instructive case study for examining how alternative structures and experiences of time might be modeled computationally. With a corpus of approximately 13,000 periodical issues split into nearly 200,000 pages, I am using Latent Dirichlet Allocation as implemented in Gensim to cluster the pages according to five different periodization schemes: no periodization (the whole corpus, spanning 70 years); periodization by decade (the corpus split by decade increments); cumulative periodization by decade (the corpus split by decade increments, but with each subsequent period added to the previous - i.e., 1844-1850, 1844-1860, 1844-1870, etc.); historical periodization (the corpus split according to "crisis" points in the denomination's history); cumulative historical periodization (the corpus split according to the historical periodization scheme above but with each subsequent period added to the previous).

Comparing these different schemes will enable me to explore whether and how periodization influences which discursive patterns I am able to surface computationally. To compare between the resulting models, I will evaluate on the following aspects: how well the documents are described by the topic labels assigned; the coherence of the topics, both internally and in relation to each other; and the visibility of change and development of topics over time.

My hypothesis is that the cumulative historical periodization will provide the richest picture of the shifts in the community's discourse. I anticipate that this will be visible through the percentage of documents labelled as relating to different topics, the appearance of new related topics over time, and different topic compositions at different points in time. However, I may conclude that the different periodization schemes provide little additional information. I propose that a negative conclusion is also important for the field and either result will provide productive information for developing computational methods that address the complexities that are at the heart of the humanities.

Bibliography

Schmidt, B. M. (2012) "Words Alone: Dismantling Topic Models in the Humanities" in *Journal of Digital Humanities* 2.1 http://journalofdigitalhumanities.org/2-

1/words-alone-by-benjamin-m-schmidt/ (accessed 01 Nov. 2016).

Underwood, T. (2012). "What Can Topic Models of PMLA Teach Us About the History of Literary Scholarship?" in *Journal of Digital Humanities* 2.1 http://journalofdig-italhumanities.org/2-1/what-can-topic-models-of-pmla-teach-us-by-ted-underwood-and-andrew-gold-stone/ (accessed 01 Nov. 2016)

Nowviskie, B. (2016) recent discussions of the logics of time and archival practices in "Speculative Collections," Nowviskie.org, 27 October 2016. http://nowvis-kie.org/2016/speculative-collections/ (accessed 01 Nov. 2016).