Jeri E. Wieringa

Ph.D. candidate, Department of History & Art History, George Mason University

@jerielizabeth | jeri.elizabeth@gmail.com | jeriwieringa.com

Research Questions

- 1. What is the best strategy for managing time when topic modeling historical documents?
- 2. Can we use topic modeling to see periods of change in a community? To see changes in their conception of time?

Narrative

The question of change over time is central for historians as they trace the changing 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. Benjamin Schmidt and others have raised concerns about 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 relatively consistent over time.[1] 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 presents 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.

Methodology

I downloaded 13,000 PDF files from the Seventh-day Adventist Office of Archives, Statistics, and Research, split by page, and extracted the text into 197,943 txt files.

I cleaned all of the documents to address standard OCR errors, including lineendings, split words, and burst words. I did not standardize spelling.

I organized the corpus according to five periodization schemes and then preprocessed and modeled all five sets using Gensim.

To evaluate the models, I compared them across one time slice, as demonstrated in this poster. I selected topics related to health and religious beliefs, and examined the ease of labeling, the match between the topic identification and the document, and the distribution of topics over time.

Findings

- 1. The language of the early Seventh-day Adventists poses a challenge for the topic modeling algorithm, as the pervasive use of metaphor and discussion of abstract topics makes it difficult to arrive at "clean" topics.
- 2. Comparing across the periodization schemes, the topics and documents identified are similar, with the greatest variance in the Full Corpus Model.
- 3. The Historical Period Model provides the best correspondence between topics and documents and the greatest ease of labeling.
- 4. All models show an increase in health reform focused pages around 1866 and a relative decline in pages presenting Seventh-day Adventist theology.
- 5. The models provide a useful overview of general trends in the corpus, revealing periods of change in the dominant discourses. However, they are too generalized to reveal shifts in the denomination's conception of time.

Questions for Further Research

- 1. Would better results be achieved using only one title at a time?
- 2. In the cumulative models, would there be a noticeable difference if the documents were weighted by distance from the concluding year?
- 3. What methods can be used in conjunction with topic modeling to surface periods of heightened millennial expectation?

Additional Information

An expanded discussion of methods, references, and the source notebook files are available at https://github.com/jerielizabeth/DH2017.

Playing with Time



Full Corpus Model	body, blood, death, pain, suffering, danger bread, milk, meat, supper, salt, sugar character, influence, knowledge, principle, condition, nature commandment, seventh, holy, sanctuary, covenant, rest disease, health, physician, treatment, patient, doctor food, animal, water, fish, fruit, meal history, century, rome, bishop, horn, empire husband, drink, wine, drinking, thief, mourn nation, government, europe, england, american, party president, united, congress, washington, national, american russia, russian, italy, italian, fleet, naval scripture, sign, point, second, view, text soldier, army, france, french, court, officer sunday, religious, rest, labor, league, liberty tobacco, temperance, liquor, drink, habit, smoking water, room, hour, minute, cold, foot	Percentage of Top 30 Docs from Full Corpus Model that "Match" Topic 36 Percentage of Top 30 Docs from Full Corpus Model that "Match" Topic 0 Percentage of Top 30 Docs from Full Corpus Model that "Match" Topic 18 Percentage of Top 30 Docs from Full Corpus Model that "Match" Topic 18 Percentage of Top 30 Docs from Full Corpus Model that "Match" Topic 18 Percentage of Top 30 Docs from Full Corpus Model that "Match" Topic 11	Topic 36 Words: scripture, sign, point, second, view, text, dead, subject, meaning, nature, record, seen, evidence, statement Hypothesis: The second coming and the Seventh-day Adventist interpretation of scripture. Final Label: Theological debates, particularly around the nature of the soul. Topic 0 Words: character, influence, knowledge, principle, condition, nature, study, youth, individual, effort, form, purpose, evi Hypothesis: Development of character and healthy practices. Final Label: Health through obedience to natural law; importance of education and parenting in achieving health. Topic 18 Words: body, blood, death, pain, suffering, danger, exclaimed, trouble, brain, poison, sleep, cell, caused, effect, germ Hypothesis: Descriptions of consequences of breaking the natural law. Final Label: Principles of Health Reform; Diseases and their remedies. Topic 11 Words: commandment, seventh, holy, sanctuary, covenant, rest, keeping, fourth, kept, adventist, change, observance Hypothesis: Explanations and arguments for Seventh-day Adventist beliefs about the Sabbath and the Sanctuary. Final Label: Arguments concerning the Sabbath, the Ten Commandments, and God's covenants.
Decade Model, 1861-1870	Decade Model, 1861-1870 beast, nation, wrath, judgment, wicked, plague disease, bath, fever, treatment, cold, patient dress, disease, physician, woman, drug, patient food, meal, eating, meat, bread, flesh gospel, sinner, obedience, principle, justice, paul prophecy, event, daniel, paul, period, verse religion, crime, peace, country, religious, society rome, pope, france, country, religious, society rome, pope, france, country, religious, seriety rome, pope, france, country, rese, wicked, judgment, text satan, devil, cain, darkness, spiritualism, serpent sunday, argument, doctrine, position, institution, claim tobacco, habit, country, effect, disease, force tract, cloth, exposition, prophecy, chart, showing	Percentage of Top 30 Docs from Decade Model 1861-1870 that "Match" Topic 19 Percentage of Top 30 Docs from Decade Model 1861-1870 that "Match" Topic 22 Percentage of Top 30 Docs from Decade Model 1861-1870 that "Match" Topic 4 Percentage of Top 30 Docs from Decade Model 1861-1870 that "Match" Topic 4 Percentage of Top 30 Docs from Decade Model 1861-1870 that "Match" Topic 11	Words: prophecy, event, daniel, paul, period, verse, vision, apostle, prophet, king, history, gospel, chapter, language Hypothesis: Discussions of prophecies and their interpretation. Final Label: Defense of Christianity; Arguments concerning Seventh-day Adventist theology. Topic 22 Words: sunday, argument, doctrine, position, institution, claim, error, change, adventist, observance, prove, statement Hypothesis: Arguments in favor of Saturday Sabbath. Final Label: Arguments in defense of Seventh-day Adventism, particularly regarding the Sabbath. Topic 4 Words: religion, crime, peace, country, religious, society, money, character, conscience, influence, public, moment Hypothesis: Signs of the second coming, particularly failures of the churches. Final Label: Signs of the second coming, particularly war; Development of or lack of Christian character. Topic 11 Words: food, meal, eating, meat, bread, flesh, animal, fruit, diet, stomach, appetite, sugar, drink, vegetable, flour, wheat Hypothesis: Discussion of the dietary causes of diseases and the virtues of Health Reform. Final Label: Description of and case for the Health Reform Diet.
Historical Period Model, 1859-1875	Historical Period Model, 1859-1875 animal, blood, action, food, organ, substance country, nation, government, rome, pope, catholic disease, tobacco, treatment, bath, patient, fever doctrine, resurrection, verse, judgment, throne, king food, diet, disease, meat, meal, fruit instructor, youth, chart, samuel, ofthe, paradise mother, woman, dress, school, girl, parent pray, message, faithful, trial, satan, trust prophecy, beast, sanctuary, message, daniel, vision room, morning, line, large, window, sleep sunday, argument, institution, observance, statement, prove wicked, righteous, wrath, immortal, judgment, matt	Percentage of Top 30 Docs from Historical Period 1859-1875 Model that "Match" Topic 9 Percentage of Top 30 Docs from Historical Period 1859-1875 Model that "Match" Topic 2 Percentage of Top 30 Docs from Historical Period 1859-1875 Model that "Match" Topic 18 Percentage of Top 30 Docs from Historical Period 1859-1875 Model that "Match" Topic 18 Percentage of Top 30 Docs from Historical Period 1859-1875 Model that "Match" Topic 18 Percentage of Top 30 Docs from Historical Period 1859-1875 Model that "Match" Topic 19 Partial Polypartial 3.33%	Words: pray, message, faithful, trial, satan, trust, overcome, dear, promise, experience, feeling, seek, mercy, salvation Hypothesis: Descriptions of Adventist piety, particularly persevering despite delay of the second coming. Final Label: Letters from readers with testimony and encouragement. Topic 2 Words: country, nation, government, rome, pope, catholic, europe, religion, history, united, public, england, france Hypothesis: Interpretations of historical events, framed as a struggle with Catholicism. Final Label: Interpretation of historical events; Commentary on news and signs of the end of time. Topic 18 Words: doctrine, resurrection, verse, judgment, throne, king, event, text, peter, idea, future, language, reign, raised Hypothesis: Discussion of conditional immortality and/or the sanctuary doctrine and Christ's work in heaven. Final Label: Conditional immortality of the soul and the events of the Final Judgement. Topic 19 Words: food, diet, disease, meat, meal, fruit, eating, appetite, drink, medicine, drug, physician, salt, wine, flesh Hypothesis: Discussion of the causes of disease and treatment of disease through embrace health reform.

Cumulative Decade Model, 1849-1870 faithful, overcome, trial, writes, praise, trying Cumulative health, disease, food, water, meal, stomach immortality, idea, existence, sense, immortal, punishment nation, government, pope, rome, france, europe Decade principle, moral, object, religious, character, article prophecy, sign, event, prophet, fulfillment, fulfilled Model, sanctuary, israel, covenant, verse, moses, priest satan, enemy, peace, minister, religion, strength 1849-1870 sunday, century, history, ancient, council, festival text, moses, fourth, argument, disciple, institution tobacco, habit, appetite, temperance, drink, flesh woman, dress, mother, lady, husband, wife

Cumulative

Historical

1849-1875

Model,

Cumulative Historical Model, 1849-1875 adam, resurrection, existence, immortality, dust, immortal country, york, society, catholic, dollar, million disease, dress, patient, cold, treatment, physician food, bread, fruit, meat, meal, salt king, verse, babylon, nation, iron, dream prophecy, beast, sign, event, period, vision religion, principle, character, moral, religious, woman rome, government, europe, france, nation, england sunday, argument, institution, statement, keeping, observance thee, satan, rich, peace, sorrow, trial tobacco, disease, habit, food, appetite, poison wicked, salvation, righteous, righteousness, sinner, apostle

Words: satan, enemy, peace, minister, religion, strength, influence, sinner, devil, mercy, face, cross, deep, danger, turn **Hypothesis**: Ambiguous — sense of persevering through some trouble.

Final Label: Case for Hygienic Medicine; Description of Health Reform Diet.

Final Label: Unclear. The Christian Life (broadly construed) and the threats of Satan.

Words: faithful, overcome, trial, writes, praise, trying, felt, present_truth, trust, striving, truly, mercy, keeping, rejoice Hypothesis: Letters from readers offering testimony and expressing belief. Final Label: Letters from readers with testimony and encouragement.

Percentage of Top 30 Docs from Cumulative Decade 1849-1870 Mode that "Match" Topic 11

Percentage of Top 30 Docs from Cumulative Decade 1849-1870 Model that "Match" Topic 22

Percentage of Top 30 Docs from Cumulative Historical Period 1849-1875 Model that "Match" Topic 22

Percentage of Top 30 Docs from Cumulative Historical Period 1849-1875 Model that "Match" Topic 1

Percentage of Top 30 Docs from Cumulative Historical Period 1849-1875 Model that "Match" Topic 9

Words: text, moses, fourth, argument, disciple, institution, change, proof, creation, prove, new_testament, resurrection Hypothesis: Arguments for Seventh-day Adventist theology, particularly Seventh-day Sabbath. **Final Label**: Debates regarding the Sabbath.

Words: health, disease, food, water, meal, stomach, physician, patient, diet, eating, meat, treatment, bread, cure, fever **Hypothesis**: Description of the principles and effectiveness of Health Reform. Final Label: Applied Hygienic Medicine: descriptions of diseases and their treatments

Topic 22

Words: wicked, salvation, righteous, righteousness, sinner, apostle, jesus_christ, hath, glorious, everlasting, darkness **Hypothesis**: Description of the Second Coming. **Final Label**: Sermons and lectures on the Judgment and Second Coming.

Words: religion, principle, character, moral, religious, woman, self, wrong, action, object, conscience, rule, crime **Hypothesis**: Development of Christian virtues.

Final Label: Warnings against immortal behavior and prescriptions for the development of virtue.

Words: thee, satan, rich, peace, sorrow, trial, seek, self, temptation, pleasure, pray, treasure, suffering, fruit, lead **Hypothesis**: Letters from readers offering testimony and stating belief. Final Label: Exhortations on Christian character and perseverance.

Topic 5

Percentage of Top 30 Docs from Cumulative Historical Period 1849-1875 Model that "Match" Topic 5 Words: tobacco, disease, habit, food, appetite, poison, psalm, stomach, animal, effect, physical, medicine, action, blood **Hypothesis**: Description of the principles and effectiveness of Health Reform. **Final Label**: Case for and Principles of Hygienic Medicine.