Can Natural Language Processing Help Identify the Author(s) of the Book of Isaiah?

Reva Freedman
Department of Computer Science
Northern Illinois University
DeKalb, IL
rfreedman@niu.edu

Abstract

Many historians believe that the Biblical book of Isaiah was written by two authors approximately two hundred years apart, generally called First Isaiah and Second Isaiah. Some even believe that the second part was itself written by two or more authors. In this paper we use natural language processing techniques to study this hypothesis. We used the Stanford parser to parse the book of Isaiah. Using Student's t and two measures of text complexity, average sentence length and average tree height, we were able to differentiate the second part of Second Isaiah, commonly called Third Isaiah, from the rest of the book. We then used MALLET's implementation of LDA to identify ten topics in the book. Using ANOVA, we were able to find two topics that could differentiate selected parts of Isaiah. We then successfully used MALLET's implementation of the Naive Bayes algorithm to find differences between First Isaiah and Second Isaiah and also to differentiate the two parts of Second Isaiah. Finally, we showed that the same technique could be used to easily differentiate Isaiah from another prophetic book of the Bible, I Samuel.

Introduction

The authorship of the book of Isaiah is an open question in Biblical text studies. The book of Isaiah contains 1291 verses divided into 66 chapters. (Some translations contain 1292 verses, obtained by splitting the last verse of chapter 63. In these editions the second half of this verse is numbered 64:1.) Many scholars believe that Isaiah must have had more than one author because the later chapters, starting with chapter 40, appear to describe incidents that happened approximately 200 years later than the earlier chapters. These putative authors are usually called First Isaiah and Second Isaiah. Some scholars go further and divide Isaiah into three sections: chapter 1-39, 40-55 and 56-66, calling the author of the final section Third Isaiah (Ward, 1991).

We used both syntactic and bag of words approaches to see whether these sections could be identified automatically. In our syntactic experiment, we used the Stanford parser to parse each verse of the text. We used Student's t to see if statistics derived from the parse trees could be used to differentiate between sections of the text.

In our first bag of words experiment, we used MALLET[1] to identify ten topics in the book of Isaiah. We then used ANOVA to see whether any of them could be used to differentiate the sections of the text. In our second bag of words experiment, we used MALLET's implementation of the Naive Bayes algorithm to attempt to separate First Isaiah and Second Isaiah and also to do a three-way comparison of First Isaiah and the two sections of Second Isaiah. To verify that the approach is workable, we concluded by comparing the book of Isaiah against another prophetic book of the Bible that we judged would be stylistically very different from Isaiah, namely I Samuel.

Data Source

The Jewish Publication Society edition of 1917 is a variant of the well-known King James Bible. Within the confines of English grammar, the King James Bible is a literal translation (Sarna and Sarna 1988, Greenspoon 2003). We wanted a literal translation in order to stay as close as possible to the original Hebrew source text so that we would be analyzing a text closer to the Hebrew syntax rather than a translator's syntax. As a practical matter, this version has been made freely available by the publisher and both the Hebrew and the translation are freely available online from Mechon Mamre, a non-profit,

[1] http://mallet.cs.umass.edu/.

Section	Chapters	Verse Count	Avg Words/ Sent – Mean	Avg Words/ Sent – s.d.	Avg Tree Height – Mean	Avg Tree Height – s.d.
A	1-39	766	26.0	10.0	12.7	3.4
B	40-66	525	27.2	10.5	13.0	3.3
B1	40-55	333	26.4	10.3	12.8	3.3
B2	56-66	192	28.5	10.8	13.2	3.3

Table 1: Descriptive statistics for sections of Isaiah.

non-denominational institute that provides curated versions of the Bible and related texts.[2]

Syntactic Approach

Data Preparation and Feature Identification

We ran the Stanford parser[3] (Klein and Manning 2003) on each verse, requiring approximately one half hour on a 3.2 GHz Intel machine. The actual number of sentences was slightly larger because, as in many books of the Bible, the verse numbering of Isaiah does not always line up with sentence breaks in English. Figure 1 shows an example of a verse containing two English sentences.

> **On what part will ye yet be stricken, seeing ye stray away more and more? The whole head is sick, and the whole heart faint;**

Figure 1: JPS 1917 translation of Isaiah 1:5.

The output of the Stanford parser is a tree structure using Penn Treebank tags (Santorini 1995). Figure 2 shows a sample tree from the first clause in Figure 1.
The Stanford parser tags each constituent in the tree using an extensive set of tags, including a fine-grained set of parts of speech, phrases headed by each of the main parts of speech, including *wh*-phrases, and sentence-level tags, including questions and subordinate clauses.

The parser uses a finer grained system than we needed, for example, subcategorizing nouns by number and verbs by tense. We rolled up the parts of speech into the categories generally used by linguists, including noun, verb, adjective, adverb and preposition. We looked at phrases headed by each of these parts of speech.

We calculated 138 features for each verse. Since some verses were longer than others, we normalized all of our counts two ways, per sentence and per 100 words of text. We added two additional fields, average words per sentence and average sentence tree height. Since the height of the parse tree is a rough measure of the amount of subordination in a sentence, these two features are measures of complexity. We also rolled up the verse data to the chapter level so we could do analysis at both levels.

Results

To test our hypotheses, we divided the book of Isaiah into the two sections commonly suggestions by historians, consisting of chapters 1-39 and 40-66 respectively, which we labeled A and B. We then divided the second section into chapters 40-55 and 56-66, labeled B1 and B2.

Most of our significant results involved the sentence complexity measures. Table 1 shows the descriptive statistics for these two features for the three sections. We used the two-tailed t-test with unequal variances to see if there were significant differences between the sections using either of these features.

Comparing sections A and B with regard to average words per sentence, we get $t = -1.95$ ($df = 1092$), $p = .05$. Comparing the same two sections with regard to tree height, we get $t = -1.68$ ($df = 1155$), $p = .09$. Both of these results indicate a trend.

In a three-way division of Isaiah, Ward (1991) believes the last section is the most distinct. Hence we chose A vs. B2 as the comparison to look at. We obtained $t = -2.92$ ($df = 280$), $p = .004$, for average sentence length, which is

```
(ROOT
  (SBARQ
    (WHPP (IN On)
      (WHNP (WDT what) (NN part)))
    (SQ (MD will)
      (NP (PRP ye))
      (ADVP (RB yet))
      (VP (VB be)
        (VP (VBN stricken) (, ,)
          (S
            (VP (VBG seeing)
              (S
                (NP (PRP ye))
                (VP (VB stray)
                  (ADVP (RB away)
                      (RBR more)
                  (CC and)
                  (RBR more)))))))))
    (. ?)))
```

Figure 2: Parse of the first sentence of Isaiah 1:5.

[2] http: //www.mechon-mamre.org. The name means Text Institute in Hebrew.
[3] http://nlp.stanford.edu/software/lex-parser.shtml.

Section	Count	Topic 3 Mean	Topic 3 Variance	Topic 8 Mean	Topic 8 Variance	Topic 9 Mean	Topic 9 Variance
A	39	0.4299	0.0178	**0.1085**	**0.0134**	0.0121	0.0005
B1	16	0.3725	0.0144	0.2273	0.0473	**0.2500**	**0.0361**
B2	11	0.4130	0.0181	0.2758	0.0488	0.0205	0.0007

Table 2: Descriptive statistics from chapter-wise LDA/ANOVA experiment.

highly significant, and $t = -1.96$ ($df = 300$), $p = .05$ for average tree height, which indicates a trend.

Bag of Words Approaches

Latent Dirichlet Allocation

We used MALLET, version 2.0.8 (McCallum 2002), for our bag of words experiments. We used the default stoplist for English, which included 524 common words.

In our first bag of words experiment, we used MALLET's implementation of LDA (Blei, Ng, and Jordan 2003) to identify ten topics from the words of the book of Isaiah, considering it as 66 documents of one chapter each. We set the parameters to do hyperparameter optimization every ten iterations.

#3 1.35768 Lord hath people God day make hand saith man behold earth nations land Israel hosts men pass glory made Jerusalem

#8 0.49171 thy thou thee hast shalt behold art Lord thine holy peace children earth didst surely sake things truth daughter sons

#9 0.16797 saith God Israel yea servant things Jacob declare thereof formed maketh created set heavens image declared graven ashamed make tree

Figure 3: Top topics obtained by LDA.

As a sample of the output, Figure 3 shows the top 20 words for the three topics with the largest Dirichlet parameters. Due to the hyperparameter optimization option, these parameters are approximately proportional to the percent of the data assigned to each topic. Capitalization has been used for readability; the implementation is case-independent. In terms of human comprehensibility, the other seven topics are similar. It is interesting to note the number of words that are obsolete forms (*thy*, *saith*, etc.) of common words which otherwise would have been suppressed by the stoplist.

For each of these three topics, we used a single factor ANOVA to analyze whether that factor could be used to separate the three sections. Although topic #3 had the highest use, it was not useful for identifying sections of the book ($p = 0.34$). However, topic #8 was clearly able ($p < .01$) to separate First Isaiah from Second Isaiah, and topic #9 was able ($p < .001$) the first part of Second Isaiah from the other two sections. Table 2 shows some of the descriptive statistics for these three topics, showing the sections that were significantly different in bold.

Naive Bayes

Our second bag of words experiment consisted of three studies using the Naive Bayes algorithm, using 10-fold cross validation each time. We did a two-way comparison, comparing First Isaiah to Second Isaiah, and a 3-way comparison, comparing sections A, B1 and B2. We did these comparisons at the sentence level instead of at the chapter level. We balanced the number of sentences in each category by proportionally selecting sentences randomly from the larger categories. To have something to contrast these results against, we added a comparison of the book of Isaiah to another book that we felt the system would have no trouble differentiating. For this purpose we chose the book of I Samuel, another prophetic book from the Bible but one which we judged would be as different from Isaiah as possible. In addition to the fact that Samuel was a different type of prophet from Isaiah, I Samuel contains a larger percent of narrative than Isaiah.

Table 3 shows a summary of our results, including the training and test accuracy and the F1 values for every category. We obtained good results (training accuracy .95, test accuracy .80) comparing First Isaiah and Second Isaiah. When comparing all three categories, we obtained a similar training accuracy of .95 but a much lower test accuracy of .58. Finally, when comparing Isaiah to I Samuel, we obtained a training accuracy of .97 and a test accuracy of .91, thus validating both our experimental design and our belief that it would be easier to differentiate Isaiah from Samuel.

Related Work

There is a small but growing body of work using syntactic analysis to identify stylistic differences. Coh-Metrix (Graesser et al. 2004) uses parts of speech, word frequency statistics and other features to measure cohesion. Wang et al. (2014) used syntactic features to differentiate conference papers from workshop papers. Freedman and Krieghbaum (2014) used syntactic features to differentiate early essays from revised ones by the same student authors, while Freedman and Krieghbaum (2015) used syntactic features to differentiate two faculty authors.

Experiment	Train Acc	Train F1 (per catg)	Test Acc	Test F1 (per catg)
A vs. B	.95	.95 .95	.80	.79 .81
A / B1 / B2	.95	.96 .94 .95	.58	.59 .57 .56
Isaiah vs. Samuel	.97	.97 .97	.91	.91 .91 .91

Table 3: Results of Naive Bayes experiments.

Coeckelbergs and van Hooland (2016) have made an initial attempt to identify topics in the Hebrew Bible using the MALLET implementation of LDA. Rafael Alvarado of the University of Virginia has implemented the Topic Modeling Workbench[4], which permits interactive topic modeling of preprocessed texts, including the King James Bible, using a JavaScript implementation of LDA.

Conclusions and Future Work

Historians commonly divide the book of Isaiah into either two or three sections. In this paper we did three experiments to try to identify those sections automatically.

Using syntactic measures, we were able to differentiate sections of Isaiah based on average words per sentence, and we were able to show a trend using the height of the parse trees. Using bag of words approaches, we were able to differentiate sections of the book using both LDA and Naive Bayes, and we were easily able to differentiate Isaiah from another prophetic book, I Samuel. These results are consistent with Ward's (1991) theory that the main differences are semantic.

In addition to the null hypothesis that the book of Isaiah was written by one person, there are several other factors that must be considered. It cannot be ruled out that the later sections of the book were written by other authors in the style of the first Isaiah, or simply that the prophetic style of the time was consistent among authors. With regard to the use of chapters, it must also be remembered that the chapter divisions are a medieval invention, i.e., approximately two millennia after the redaction of the original text.

Our planned studies include one using only function words, as is frequently done for author identification. We eventually hope to redo this analysis using the original Hebrew text. In addition to the types of analysis performed here, the Hebrew text offers the possibility of morphological analysis and the analysis of lexical roots.

Acknowledgements

Dr. Elliot Lefkovitz, Department of History, Loyola University Chicago, introduced me to this problem. Douglas M. Krieghbaum wrote the core Python code, using NLTK[5] (Bird, Klein, and Loper 2014) tree handling code.

References

Bird, S., Klein, E., and Loper, E. 2014. *Natural Language Processing in Python: Analyzing Text with the Natural Language Toolkit.* Available at http://www.nltk.org/book.

Blei, D., Ng, A., and Jordan, M. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3: 993-1022.

Coeckelbergs, M., and van Hooland, S. 2016. Modeling the Hebrew Bible: Potential of Topic Modeling Techniques for Semantic Annotation and. Historical Analysis. In Proceedings of the Second International Workshop on Semantic Web for Scientific Heritage (SW4SH 2016). Available from http://ceur-ws.org/Vol-1595/paper5.pdf.

Freedman, R., and Krieghbaum, D. 2014. Effects of Rewriting Essays on Linguistic Measures of Complexity. In 25th Annual Meeting of the Society for Text and Discourse.

Freedman, R., and Krieghbaum, D. 2015. Comparison of Expert Tutors through Syntactic Analysis of Transcripts. In Proceedings of 17th International Conference on Artificial Intelligence in Education (AIED 2015).

Graesser, A., McNamara, D., Louwerse, M., and Cai, Z. 2004. Coh-Metrix: Analysis of Text on Cohesion and Language. *Behavior Research Methods, Instruments and Computers*, 36: 193-202.

Greenspoon, L. 2003. Jewish Translations of the Bible. In *The Jewish Study Bible*, ed. A. Berlin and M. Brettler, 2005-2020. New York: Oxford.

Klein, D., and Manning, C. 2003. Accurate Unlexicalized Parsing. In Proceedings of the 41st Meeting of the Association for Computational Linguistics, 423-430.

McCallum, A. 2002. MALLET: A Machine Learning for Language Toolkit. Available at http://mallet.cs.umass.edu.

Santorini, B. 1995. Part-of-Speech Tagging Guidelines for the Penn Treebank Project. University of Pennsylvania, Natural Language Processing Project. Technical Report MS-CIS-90-47, LINC LAB 178.

Sarna, J., and Sarna, N. 1988. Jewish Bible Scholarship and Translations in the United States. In *The Bible and Bibles in America*, ed. E. Frerichs, 83-116. Atlanta: Scholars Press.

Wang, H., Di Eugenio, B., Lin, S., and Yu, C. 2014. Distinguishing Different Types of Conference Submissions: The ACL Case Study. Chicago Colloquium on Digital Humanities and Computer Science. Available at http://www.nlp.cs.uic.edu/PS-papers/dhcs_2014_manuscript.pdf.

Ward, J. 1991. *Thus Says the Lord: The Message of the Prophets*. Nashville: Abingdon.

[4] http://tmw.drupal.shanti.virginia.edu/.

[5] http://www.nltk.org/.