Draft version May 1, 2023
Typeset using LaTeX default style in AASTeX631

A Statistical Exploration of the Hypothesized Partition of the Books of Genesis and Exodus into Priestly and Non-priestly Components

Gideon Yoffe [1,2] Axel Bühler [3,4] Thomas Römer [3] Nachum Dershowitz [5] Eli Piasetzky [6]
Israel Finkelstein [7] and Barak Sober [1]

[1] Dept. of Statistics and Data Science, the Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel
[2] Dept. of Archaeology and Ancient Near Eastern Cultures, Tel Aviv University, Ramat Aviv 6997801, Israel
[3] Collège de France, 11 Pl. Marcelin Berthelot, 75231 Paris, France
[4] Faculté de théologie, Université de Genève, 1205 Geneva, Switzerland
[5] School of Computer Science, Tel Aviv University, Tel Aviv University, Tel Aviv 6997845, Israel
[6] School of Physics and Astronomy, Tel Aviv University, Tel Aviv University, Tel Aviv 6997845, Israel
[7] Dept. of Archaeology and Ancient Near Eastern Cultures, Tel Aviv University, Tel Aviv University, Tel Aviv 6997845, Israel

1. INTRODUCTION

We examine the hypothetical distinction between texts of priestly (P) and non-priestly (nonP) origin in the books of Genesis and Exodus, for which exists a surprisingly large agreement amongst biblical scholars (e.g., Knohl 2007; Römer 2014; Faust 2019). Examining this distinction with an independent, unsupervised computational methodology would establish a measure of confidence therein and encourage its application to additional instances of biblical texts, especially those of greater controversy, where our approach could help tilt the scale in favor of one hypothesis over another.

2. METHODOLOGY

We intertwine descriptive and inferential statistics. The first is used in text classification and interpretability analyses, whereas the latter quantifies uncertainty through hypothesis testing. While descriptive statistics were successfully applied to specific texts (e.g., Kestemont et al. 2016; Verma 2017), we are unaware of similar studies where uncertainty quantification was considered. Furthermore, identification of literary features *responsible* for the classification, as opposed to cluster-wise significant feature detection (e.g., Hruschka and Covoes 2005; Cai et al. 2010; Zhu et al. 2015), is novel for stylometry.

2.1. *Corpus*

We use STEPBible[1] (digitized Leningrad codex), with its morphological and semantic tags for all words, prefixes, and suffixes. We consider two representations of the text: word-wise and a grammatical representation by phrase-dependent parts-of-speech (pdps).

We obtained a scholarly labeling assigning each verse in Genesis and Exodus as P/nonP.

2.2. *Parameterization and Embedding*

Our underlying assumption is that significant literary differences between texts manifest in simple linguistic parameters. Therefore, we consider three parameters, distinct combinations of which result in different classifications. These are: (**1**) word-/pdp-wise representations, (**2**) n-gram size, the length of sequences of consecutive words/pdps, and (**3**) running-window size, the number of verses surrounding the original, providing additional context.

We use tf-idf to encode each verse, assigning a relevance score to each feature in the context (Aizawa 2003). The critical consideration behind choosing this traditional embedding is that it allows interpretability of the results, unlike neural-net-based language models, which are convoluted (e.g., Chakraborty et al. 2017; Devlin et al. 2018).

2.3. *Optimization*

[1] https://github.com/STEPBible/STEPBible-Data

We use k-means to classify the embedded verses and use an unbalanced accuracy measure to quantify the goodness of classification. We perform cross-validated grid-search on a range of running-window and n-gram sizes for words/pdps, identifying the combination that yields the highest accuracy (Fig. 1, upper left panel).

2.4. *Testing and Validating*

To establish the statistical significance of our results, we introduce a hypothesis-testing cyclic label-shift test, that preserves the structure of the hypothesized partition and overcomes the fact that label-permutation tests do not consider correlations between units of text, which affect their likelihood of being clustered together (see Fig. 1, upper right panel).

2.5. *Feature Importance*

Minimizing k-means loss is equivalent to maximizing *inter-cluster* variances. Leveraging this, we extract a vector of feature-wise importance that maximizes the inter-cluster variance found by 2-means. This vector allows us to trace the features most responsible for the classification (Fig. 1, bottom panels).

3. CONCLUSIONS

We examined the hypothetical distinction between texts of priestly (P) and non-priestly (nonP) origin in the books of Genesis and Exodus, which we explored with a novel unsupervised pipeline for text stylometry. We sought a combination of literary parameters that optimized the overlap between the unsupervised and hypothesized partitions. We established the statistical significance of our results using a cyclic-shift test, which we show to be more adequate for text stylometry problems than a naive permutation test. Finally, we extracted n-grams that contribute the most to the classification, their respective proportions, statistical robustness, and correlation to other features. We achieve optimal, statistically significant overlap values of 73% and 90% for the books of Genesis and Exodus, respectively.

We find the discrepancy in optimal overlap values between the two books to stem from two factors: (**1**) A more sporadic distribution of P texts in Genesis, as opposed to a more formulaic one in Exodus. (**2**) The sensitivity of our pipeline to a distinct semantic field manifested in two large P blocks in Exodus, comprising the majority of the P-associated text therein.

Through complementary exegetical and statistical analyses, we show that our methodology differentiates the unique generic style of the Priestly source, characterized by lawgiving, cult instructions, and streamlining a continuous chronological sequence of the story through third-person narration. This observation corroborates and hones the stance of most biblical scholars.

REFERENCES

A. Aizawa. An information-theoretic perspective of tf–idf measures. *Information Processing & Management*, 39(1): 45–65, 2003.

D. Cai, C. Zhang, and X. He. Unsupervised feature selection for multi-cluster data. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 333–342, 2010.

S. Chakraborty, R. Tomsett, R. Raghavendra, D. Harborne, M. Alzantot, F. Cerutti, M. Srivastava, A. Preece, S. Julier, R. M. Rao, et al. Interpretability of deep learning models: A survey of results. In *2017 IEEE Smartworld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (Smartworld/SCALCOM/UIC/ATC/CBDcom/IOP/SCI)*, pages 1–6. IEEE, 2017.

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

A. Faust. The world of P: The material realm of priestly writings. *Vetus Testamentum*, 69(2):173–218, 2019.

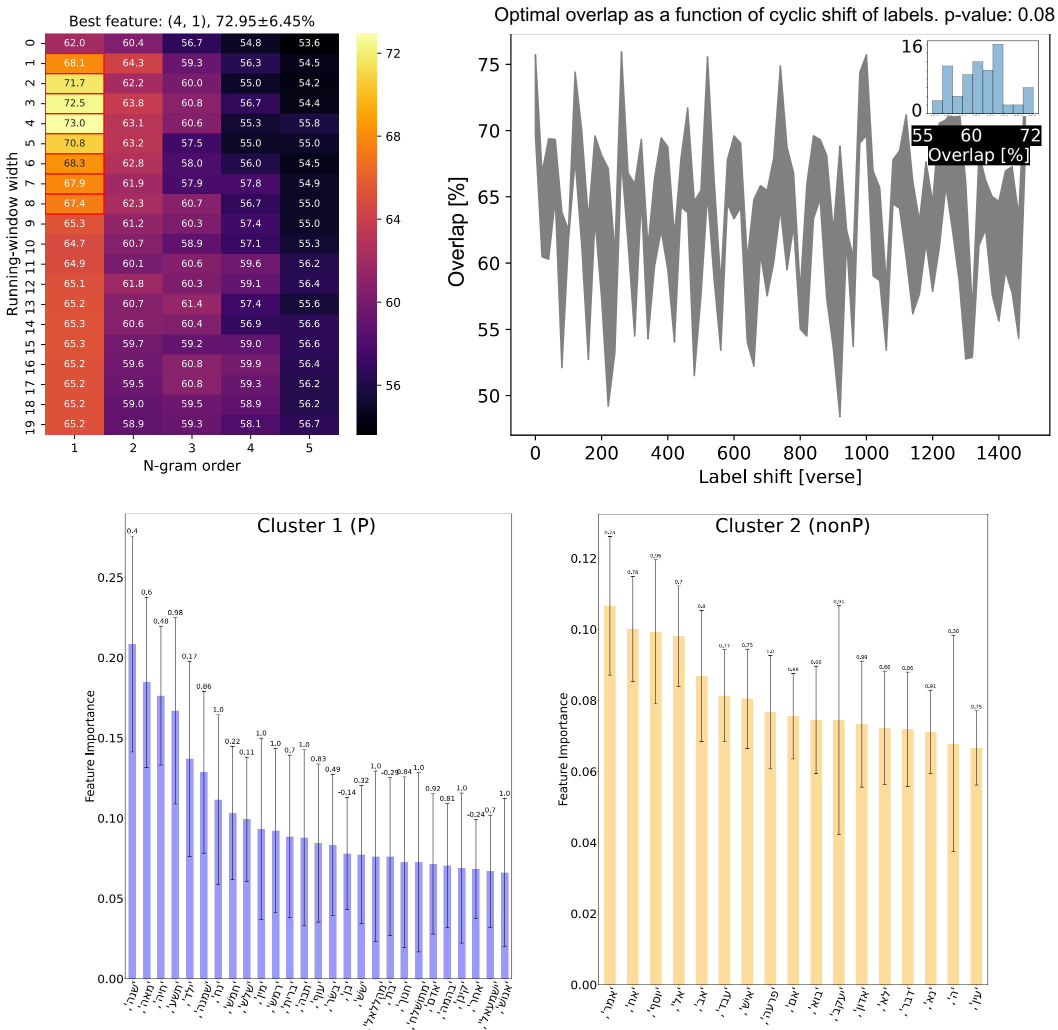

Figure 1. A statistical exploration of the hypothesized partition of the books of Genesis and Exodus into Priestly and non-Priestly constituents: results for the book of Genesis (lexemes representation). **Upper Left Panel – Optimization**: Cross-validated grid-search over verse running-window widths and n-gram sizes to identify the combination yielding an optimal overlap. The combination, value, and uncertainty of the optimal overlap are plotted on top of the panel, and all combinations whose overlap value is within 1σ of the optimal overlap value are marked in red cells. **Upper Right Panel – Hypothesis Testing**: Cross-validated hypothesis testing, where we simulate the null hypothesis distribution through a cyclic shift of the hypothesized P/nonP labeling. The p-value is measured with respect to the optimal overlap value minus its standard deviation. **Bottom Panels – feature importance Analysis**: Cross-validated important feature analysis (running-window 4 and n-gram size 1) and their statistical stability, displaying features bearing 75% of the explained variance. Features in the left and right panels are important to the P-associated and nonP-associated clusters, respectively. The small numbers above each error bar indicate the cluster-wise abundance of the feature.

E. R. Hruschka and T. F. Covoes. Feature selection for cluster analysis: an approach based on the simplified silhouette criterion. In *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, volume 1, pages 32–38. IEEE, 2005.

M. Kestemont, J. Stover, M. Koppel, F. Karsdorp, and W. Daelemans. Authenticating the writings of Julius Caesar. *Expert Systems with Applications*, 63:86–96, 2016.

I. Knohl. *The Sanctuary of Silence: The Priestly Torah and the Holiness School.* Eisenbrauns, 2007.

T. Römer. From the call of Moses to the parting of the sea: Reflections on the priestly version of the Exodus narrative. In *The Book of Exodus*, pages 121–150. Brill, 2014.

M. Verma. Lexical Analysis of Religious Texts using Text Mining and Machine Learning Tools. *International Journal of Computer Applications*, 168(8):39–45, June 2017. https://doi.org/10.5120/ijca2017914486. URL https://doi.org/10.5120/ijca2017914486.

G.-N. Zhu, J. Hu, J. Qi, J. Ma, and Y.-H. Peng. An integrated feature selection and cluster analysis techniques for case-based reasoning. *Engineering Applications of Artificial Intelligence*, 39:14–22, 2015.