Draft version December 3, 2022
Typeset using LaTeX default style in AASTeX631

An Independent Unsupervised Examination of the Distinction Between Texts of Priestly and Non-priestly Origins in the Books of Genesis and Exodus

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., 8; 9; 5). 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., 7; 10), 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., 6; 2; 11), 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 (1). 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., 3; 4).

2.3. *Optimization*

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).

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

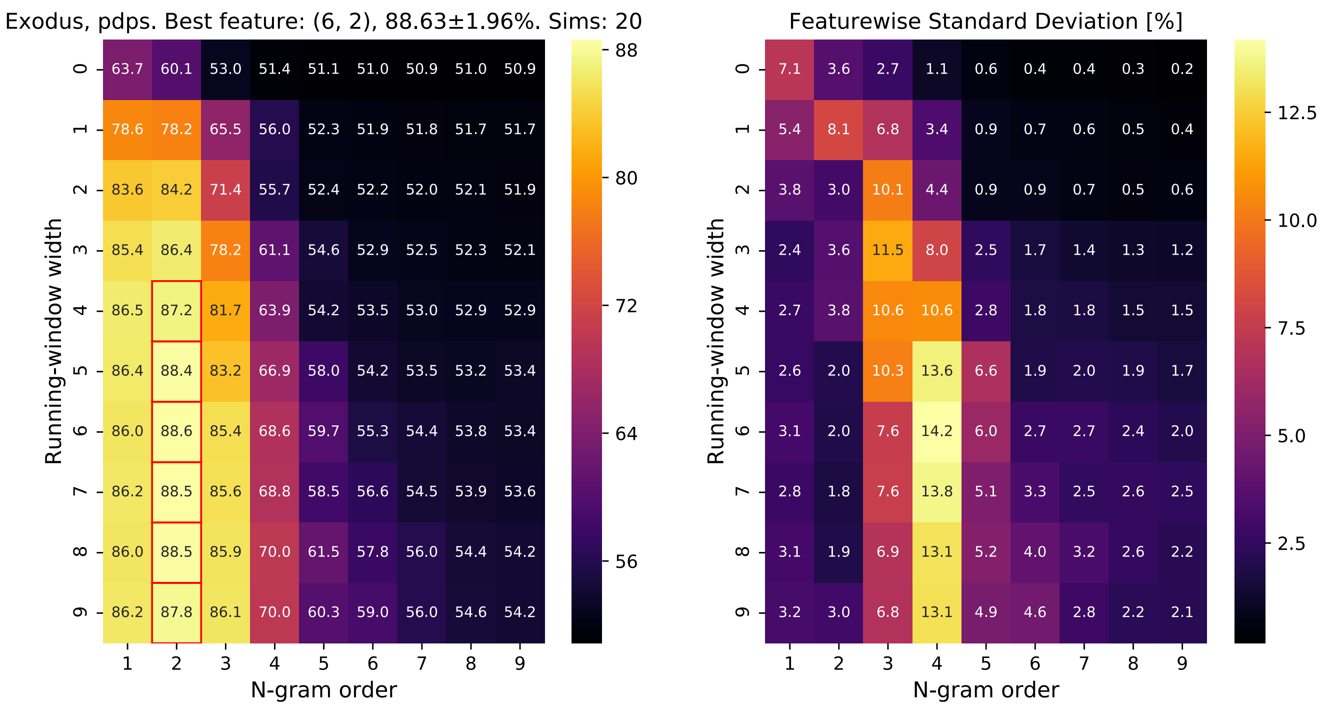

Figure 1. Cross-validated grid-search optimization for Exodus (pdps). **Left Matrix**: averaged accuracy over 10 simulations with respect to combinations of window sizes (y-axis) and n-gram sizes (x-axis). The best-fit combination yields $\approx 89.51.96\%$ accuracy, and the cells of feature combinations within 1σ thereof are marked with red. **Right Matrix**: standard deviation of the left matrix.

2.4. *Testing and Validating*

Through hypothesis testing, we aim to establish statistical significance of the distinction. We perform two tests, under the null hypothesis that our labels were randomly assigned: (**1**) arbitrary permutations, and (**2**) *cyclic* permutations, where we generate the null by shifting the labeling cyclically, seeking to conserve implicit correlations between consecutive verses.

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. 2).

3. CONCLUSIONS

We examined the hypothesized P/nonP distinction in Genesis and Exodus and introduced a novel computational and statistical methodology for text stylometry that is essentially independent of-, but in synergy with- established philological practices. We sought an optimal single feature—a combination of running-window and n-gram sizes, and extracted features that contribute most to classification and their respective proportions. We achieve a 73% and 90% (balanced) accuracy for Genesis and Exodus. The difference in accuracy between the two seems to arise from the more sporadic distribution of P in Genesis, as opposed to a more formulaic one in Exodus.

REFERENCES

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

[2] Deng Cai, Chiyuan Zhang, and Xiaofei 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.

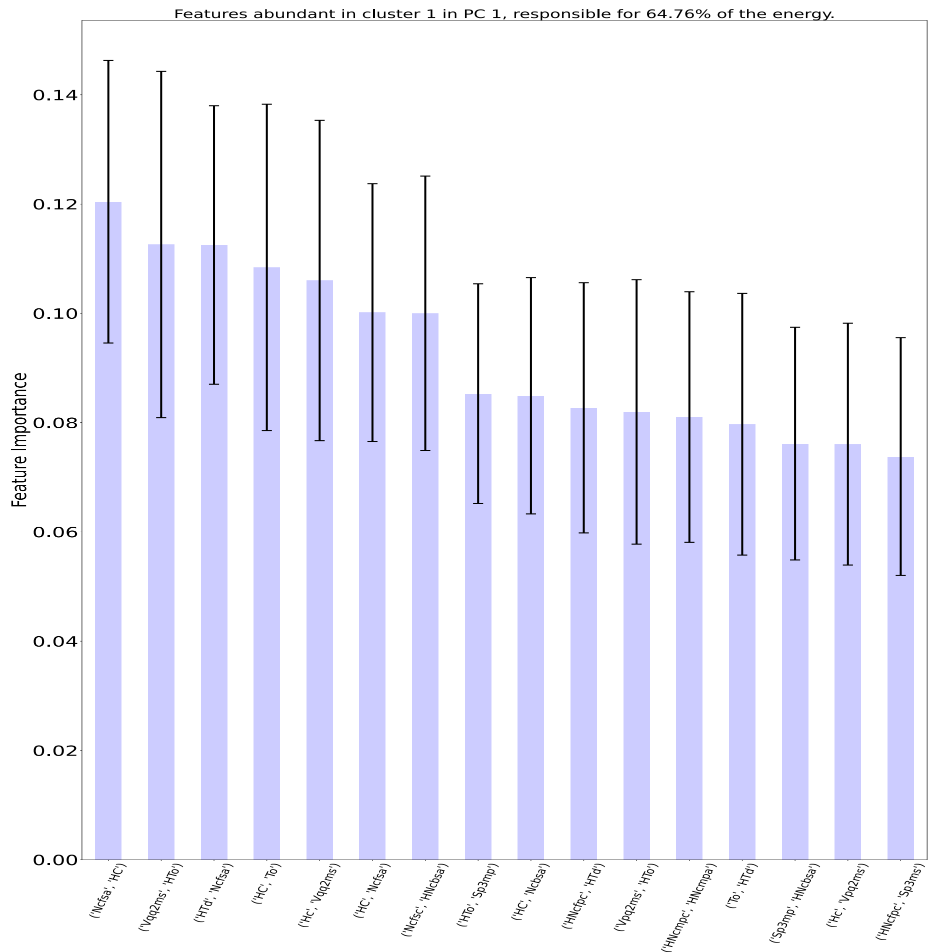

Figure 2. Feature importance bar plot for Exodus embedded as bigrams of pdps with window width 4. Here, 100% of the distinction is made by features that are abundant in the nonP cluster.

[3] Supriyo Chakraborty, Richard Tomsett, Ramya Raghavendra, Daniel Harborne, Moustafa Alzantot, Federico Cerutti, Mani Srivastava, Alun Preece, Simon Julier, Raghuveer 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.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

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

[6] Eduardo R. Hruschka and Thiago 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.

[7] Mike Kestemont, Justin Stover, Moshe Koppel, Folgert Karsdorp, and Walter Daelemans. Authenticating the writings of Julius Caesar. *Expert Systems with Applications*, 63:86–96, 2016.

[8]Israel Knohl. *The Sanctuary of Silence: The Priestly Torah and the Holiness School*. Eisenbrauns, 2007.

[9]Thomas 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.

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

[11]Guo-Niu Zhu, Jie Hu, Jin Qi, Jin Ma, and Ying-Hong Peng. An integrated feature selection and cluster analysis techniques for case-based reasoning. *Engineering Applications of Artificial Intelligence*, 39:14–22, 2015.