Report: Replicating "A Timely Intervention: Tracking the Changing Meanings of Political Concepts with Word Vectors"

PPOL 6801, Spring 2024 Prof. Tiago Ventura

Zhiqiang Ji, Minh Quach

Summary

This replication study reproduced (Rodman, 2020). The original paper, "A Timely Intervention: Tracking the Changing Meanings of Political Concepts with Word Vectors," tackles two significant challenges for political text analysis: tracing the evolving cultural meaning of political concepts over time and leveraging the capabilities of Word Vector method with limited text volume. Through testing four time-aware word2vec implementations against a gold standard from 161 years of newspaper articles centrally related to the theme "equality," Rodman's study identifies an iterative training workflow that aligns with human judgments on the evolving meaning of equality in America's public discourse. It also utilizes bootstrap resampling to overcome the constraints of limited text volumes and enhances the reliability and applicability of word vector models in political science research. Since the manual coding is not replicable, this replication study used the coded text data to create the gold standard, then produced the four diachronic implementations of word2vec models (the naive time model, the overlapping time model, the chronologically trained model, and the post-training aligned model) from scratch. Consequently, this replication reproduced the original study's findings.

Process

The author provided all the necessary data, well-documented code, and comprehensive documentation, significantly facilitating our replication effort. Furthermore, as the study focused on handling limited text volumes, we didn't face the challenges of acquiring, cleaning, and preprocessing extensive datasets for replication (Rodman, 2019). Nonetheless, replicating research that's nearly five years old brought its own set of challenges, particularly due to changes in the software environment. This required us to carefully modify the original packages and scripts to ensure compatibility with the updated syntax and functionalities of newer software versions. Specifically, we had to update the Python scripts, especially `word2vec_functions.py`, to work in the modern environment of Python 2.7 and Gensim 3, as used by the original authors for training the word2vec models.

Additionally, the original study utilized the ReadMe package (Hopkins & King, 2010), an excellent choice for constructing the gold standard to evaluate the four word2vec models. This package analyzes the corpus by dividing it into predefined topics, comparing word

stem frequency distribution patterns between a manually annotated training set and an unannotated test set, to infer the textual distribution across different periods. However, the challenge arose because the package, initially released in 2010, had a significantly updated version 2 released in 2019, which adopted pre-trained GloVe vectors for text representation. Despite the active maintenance of the newer version's GitHub repository, the original ReadMe V1 hadn't been updated in seven years. At the beginning of our replication process, we were tempted to use the updated ReadMe package. Yet, understanding the operational principles of both versions and their compatibility with Rodman's research objectives was crucial. After a thorough evaluation, we opted to closely replicate Rodman's methodology, aiming to see if word2vec could achieve results comparable to the more traditional, yet resource-intensive, manual labeling methods.

Differences

The subsequent replication process proceeded smoothly, and we generated charts very similar to those of the original authors. This confirmed the authors' conclusion that among the four diachronic implementations of word2vec models, the Chrono model, which employs an iterative training workflow using the training results of one time slice as the starting point for the next, performed the best. However, we also noted that the advantage of the Chrono model over the second-best, the Overlap model, seems to be smaller than reported in the original paper, as shown in the comparison tables below.

Table 1: Rodman's ANOVA

	Model	Deviance	Squared.Deviance	Correlation
1	naive	26.613	25.589	0.555
2	overlap	25.163	25.602	0.564
3	**chrono**	22.689	21.860	0.612
4	aligned	25.385	25.775	0.515

Table 2: Replication ANOVA

	Model	Deviance	Squared.Deviance	Correlation
1	naive	27.856	27.182	0.532
2	overlap	25.558	25.563	0.520
3	**chrono**	24.179	23.312	0.568
4	aligned	27.508	26.797	0.536

Our findings closely mirrored those presented in the original paper, with the figures we obtained showing a high degree of similarity. The distinctions between our study and the original research are more apparent in the numerical data than in the visual plots. Despite these minor discrepancies, whether concerning the gold standard or the models themselves, the observed trends in the changing similarities between the concept of Equality and thematic words were identical to those reported in the original study. The minor variations

in our replication can be attributed to the random sampling process employed in generating the Gold Standard, as well as differences in the version of the Gensim package used.

From the results of our replication effort, we reaffirm the original author's conclusion that the Chrono model is the most aligned with human judgment and adequately demonstrates the validity of using the unsupervised word2vec method to trace the evolution of political concepts' meanings. However, it's worth noting that, according to our observations, the advantage of the Chrono model over the second-best Overlap model seems to be less pronounced than what was reported in the original paper.

Autopsy

According to the author's report, the estimated time to run all the scripts from beginning to end was about 46 hours. Fortunately, five years later, significant improvements in personal computer processing speeds, along with upgrades to Python and Gensim, have optimized computational performance. This enhancement allowed us to reduce the time required to run all scripts to approximately 12 hours, enabling us to execute each script at least twice to verify the validity of our results.

Also, as previously mentioned, this paper stands out with its greatly detailed documentation of the author's choices and assumptions, no matter how minute. Thanks to such precisely defined scope and caveats of research, the paper is highly replicable, and the replication results are consequently very comparable, which we certainly appreciate.

Extension

An interesting future direction to explore is the use of updated tools, such as ReadMe 2 (Jerzak et al., 2023), to develop more efficient or more generalizable methods based on word vectors as representations of text. The use of word vectors is expected to provide richer semantic information about words and the relationships between them, compared to traditional approaches. These methods could be particularly useful in capturing the evolving meaning of words, especially political (and sociological) concepts, over time. The approach can be applied to a myriad of thematically significant terms such as 'power', 'rich', 'justice', 'wrong', just to name a few.

Moreover, with the emergence of models based on the Transformer architecture, which have shown promise in capturing the complex relationships between words and their contexts, there is significant potential for these models to be applied more extensively in analyzing diachronic changes in language.

Reference

Hopkins, D. J., & King, G. (2010). A Method of Automated Nonparametric Content Analysis for Social Science. *American Journal of Political Science*, *54*(1), 229–247. https://doi.org/10.1111/j.1540-5907.2009.00428.x

Jerzak, C. T., King, G., & Strezhnev, A. (2023). An Improved Method of Automated Nonparametric Content Analysis for Social Science. *Political Analysis*, *31*(1), 42–58. https://doi.org/10.1017/pan.2021.36

Rodman, E. (2019). *Replication Data for: A Timely Intervention: Tracking the Changing Meanings of Political Concepts with Word Vectors* [dataset]. Harvard Dataverse. https://doi.org/10.7910/DVN/CGNX3M

Rodman, E. (2020). A Timely Intervention: Tracking the Changing Meanings of Political Concepts with Word Vectors. *Political Analysis*, *28*(1), 87–111. https://doi.org/10.1017/pan.2019.23