Uncovering partisanship at the national and state level by identifying unique voter blocs in congress

Evelyn Siu and Dustin Marshall
esiu@uchicago.edu,dustinmarshall@uchicago.edu

December 2, 2022

Abstract

Singular value decomposition and non-negative matrix factorization are machine learning methods that can uncover hidden patterns in large datasets. We apply these methods to detect partisanship in voting behavior in national and state (Illinois) legislatures. We find that there are two clear voting blocs at both the national and state level that align with party affiliation. Party lines clearly define voting behavior at the national level, but do not predict voting behavior as well at the state level, suggesting a more fluid party affiliation at the state than national level.

1 Introduction

It's commonly understood that partisanship has infected politics at the national level in the United States. For this reason, party affiliation should be a strong predictor of whether any given congressperson will vote for or against a measure. Researchers have used methods from machine learning, such as singular value decomposition (SVD), to track the increase in partisanship at the federal congressional level over the years (Moler 2020). It has even been used as an example to highlight the value of SVD at uncovering hidden structures in data (Martin and Porter, 2017).

However, the results of SVD are not interpretable because of the presence of negative values. The additional non-negative constraints placed on another matrix factorization technique, non-negative matrix factorization (NMF) make it easier to interpret than SVD (Fogel et al. 2013). Researchers have used NMF similarly to track and predict changes in voting decisions on the boards of technology companies (Ribeiro-Navarrete et al. 2019).

Our project builds upon existing literature to use a combination of SVD and NMF to uncover and interpret partisanship at not only the national, but also the state level. We expect party lines to be more blurred at the state than national level.

2 Methods

Our tool for detecting partisanship is matrix factorization, which is a process of splitting a single matrix into a product of multiple matrices. This process helps us identify clusters in the original matrix, which in our dataset represent unique voting blocs. Our project leverages two matrix factorization processes. First, we use singular value decomposition (SVD) to approximate the number of unique voting blocs in the data. SVD works by factoring the original matrix into a left and right matrices that can be linearly combined with a third matrix (Σ) to recreate the original matrix. The Σ matrix is a diagonal matrix of singular values in descending order. The magnitude of each singular value corresponds to the predictive value contained in each dimension in the dataset. A steep drop-off in the singular values would tell us where there might be a cut-off in the number of unique voting blocs in the US and IL legislatures. We use this information to approximate the best n-dimensional subspace to represent our original voting data.

Second, we take the number of unique voting blocs identified by SVD, and use NMF to predict the voting bloc of a certain representative, given their voting history. NMF works by factoring our original n x p (samples x features) matrix into two matrices (W and H) with a user-specified k number of dimensions. The first resulting matrix, W, is a $n\ x\ p$ matrix, and H is a $k\ x\ p$ matrix (Figure 1). The maximum value in each row of W indicates which samples in the original matrix contribute most to identifying each unique k dimension of the data (Figure 2). The maximum value in each column of the H matrix represents the predicted cluster that each feature falls into.

In our particular data, our features were each representative, and samples were each voting event. The sigma values from the SVD told us how many unique voting blocs there were, which we used to set k.

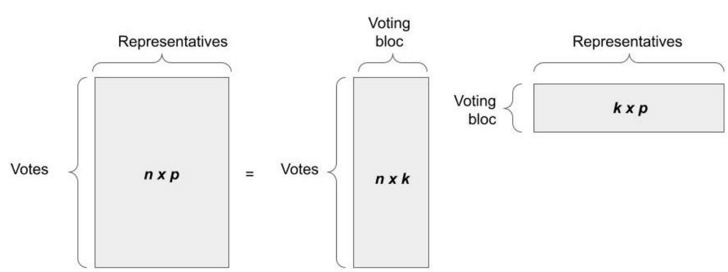

Figure 1: Non-negative matrix factorization of our original data to represent votes per voting bloc and voting bloc per representative.

	Rep 1	Rep 2	Rep 3
Bloc 1	0.2	**0.4**	**0.6**
Bloc 2	**0.55**	0.1	0.3
Predicted Bloc	Bloc 2	Bloc 1	Bloc 1

Figure 2: The maximum value for each representative indicates which voting bloc they are predicted to follow.

3 Results

The magnitude of our data's singular values dropped off steeply at both the federal and state levels after two singular values (Figure 3). This drop-off was even more pronounced in the Illinois than federal congressional dataset, suggesting there is a clearer delineation of two voting blocs at the state than federal level. Therefore, we set our number of dimensions to two for NMF.

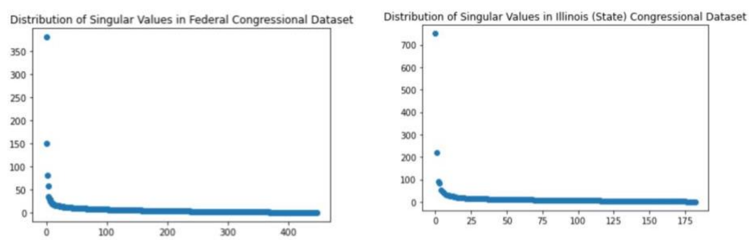

Figure 3: Graph of singular values by magnitude for Federal and State congressional votes.

Using $k = 2$ as the dimension for NMF, at the federal level, there are 333 representatives in one voting bloc and 211 in another. At the state level, 80 representatives are clustered into one bloc, and 102 in another.

The results of NMF alone do not identify which bloc is considered Democrat and which party is considered Republican. We labeled each bloc by measuring the accuracy of prediction if one bloc was considered Dem. v. Republican. Based on these results, there is a higher accuracy in prediction if we set Bloc 1 to represent Democrats, and Bloc 2 to represent Republicans. The accuracy of this labeling was 89% at the federal level and 52% at the state level.

Using these labels, we measured the accuracy of each prediction on an individual representative level, at the federal and state level. The prediction is most accurate for US congressional Democrats (100% accuracy), and worst for Illinois Republicans (44% accuracy). The results suggest federal Democrats strongly vote together, and there are 58 federal Republicans and 2 Independents that vote similar to Democrats. At a state level, there are 35 Republicans that vote with the Democrats and 52 Democrats that vote with the Republicans.

	US Overall Accuracy: 89%			Illinois Overall Accuracy: 52%		
US & Illinois Legislative Voting Blocs	Predicted Counts	Actual Counts	Accuracy	Predicted Counts	Actual Counts	Accuracy
Bloc 1: Democrats	333	275	• Accurately predicted all 275 D's • Incorrectly predicted 58 R's + 2 I's as D	102	119	• Accurately predicted 67 D's • Inaccurately predicted 35 R's as D's
Bloc 2: Republicans	211	267	• Accurately predicted 211 R's out of 267 • Incorrectly predicted 0 D's as R's	80	63	• Accurately predicted 28 R's • Inaccurately predicted 52 D's as R's
Bloc 3: Independents (Not predicted)	N/A	2	• Did not include Independents as a voting bloc	N/A	0	• Did not include Independents as a voting bloc

Figure 4: Breakdown of predicted versus actual party affiliations on a federal and state level.

4 Conclusion

These results suggest that at both the national and state level, there are two overall voting blocs. At the national level, there is strong partisanship - there is a clear delineation in voting patterns between the two voting blocs. However, at the state level, there is less partisanship - Democrats are voting with Republicans, and vice versa.

The steep drop-off of in magnitude after the first two singular values suggest that there are 2 clear voting blocs. On the federal level, there may be a weaker 3rd and 4th voting bloc. In future iterations of this experiment, we could explore NMF results when k=*4*. In this paper, we only explored k=*2*.

The results of the NMF support previous conclusions that partisanship is strong at a national level. At a state level, the NMF was unable to predict a single representative's voting bloc as accurately, suggesting weaker party affiliation during votes. This result aligns with our current understanding that local and state political identification is more fluid than at the national level.

In future work, we could continue to explore partisanship, by examining which votes were considered important to the classification of each representative as one voter bloc or another. In addition to identifying the vote events, we would have to examine the actual content of the vote to identify what issues are considered key Democrat v. key Republican issues. We could also continue our current work by looking specifically at the voting records of representatives that vote with a voting bloc different from their political identification. We could examine what makes these representatives more flexible and bipartisan than their fellow representatives.

References

[1] Carla D. Martin, & Mason A. Porter. (2012) The Extraordinary SVD. *The American Mathematical Monthly*, 119,(10) pp. 838-851. https://doi.org/10.4169/amer.math.monthly.119.10.838
[2] Fogel, Paul. Douglas M. Hawkins, Chris Beecher, George Luta, S. Stanley Young.(2013) *A Tale of Two Matrix Factorizations.* National Institute of Statistical Sciences, Technical Report. https://www.niss.org/sites/default/files/tr185.pdf
[3] Juan Pineiro-Chousa, Marcos Vizcaíno-González, Belén Ribeiro-Navarrete (2019) *Assessing the effects of disruptive events on technology firms using non-negative matrix factorization* International Review of Economics Finance, Volume 62, pp. 79-86. ISSN 1059-0560, https://doi.org/10.1016/j.iref.2019.02.017
[4] Moler, C. (2020, Aug. 23) *SVD quantifies increasing partisanship in the U. S. senate.* Cleve's Corner: Cleve Moler on Mathematics and Computing. Retrieved November 6, 2022, from https://blogs.mathworks.com/cleve/2020/08/23/svd-quantifies-increasing-partisanship-in-the-u-s-senate/
[5] Porter, M. A., Mucha, P. J., Newman, M. E., Warmbrand, C. M. (2005). it A network analysis of committees in the U.S. House of Representatives. Proceedings of the National Academy of Sciences of the United States of America, 102(20), 7057–7062. https://doi.org/10.1073/pnas.0500191102
[6] Sirovich L. (2003) it A pattern analysis of the second Rehnquist U.S. Supreme Court. Proceedings of the National Academy of Sciences of the United States of America, 100(13), 7432–7437. https://doi.org/10.1073/pnas.1132164100