

Challenges in Interpreting the Outcome of Political Text Classification; A study on the US Congress

Pooya Oladazimi
Master of Web Science
poladazimi@uni-koblenz.de
Summer 2020

Introduction

- Liberal/Conservative conflicts increased in US [1] → shows tension in the society and between politicians
- Conflicts lead to change in vocabulary [2]
- Ideology classification: [3] → Ideology shapes our behavior, a blueprint for our actions
 - Understand the controversial times and issues
 - Predict future outcomes
- Computer science can help → Text Classification to understand political conflicts and differences

Introduction

- Bei Yu et.al [3] study:
 - Party affiliation for Ideology classification
 - US House is more partisan than Senate: House-trained gives higher accuracy than Senate-trained
 - US Congress partisanship increased in recent years

Motivation

- Hypothesis:
 - H1: US House more partisan than Senate
 - H2: US Congress partisanship is increasing
 - H3: Party = Ideology since only ideology shapes our choice of vocabulary
 - Not in parliamentary systems [4]
 - Vocabulary difference does not always mean Ideology conflict

Approach

- Compare House-trained with Senate-trained over years (H1)
- Reduce House data size (H1) → US House has an advantage: Four times more members than Senate
- Pick optimized train data and test on all Senates (H2)
- Extract top features for each party in each Congress [4], Compare based on party position (Incumbent/Challenger) [6] (H3)

Data

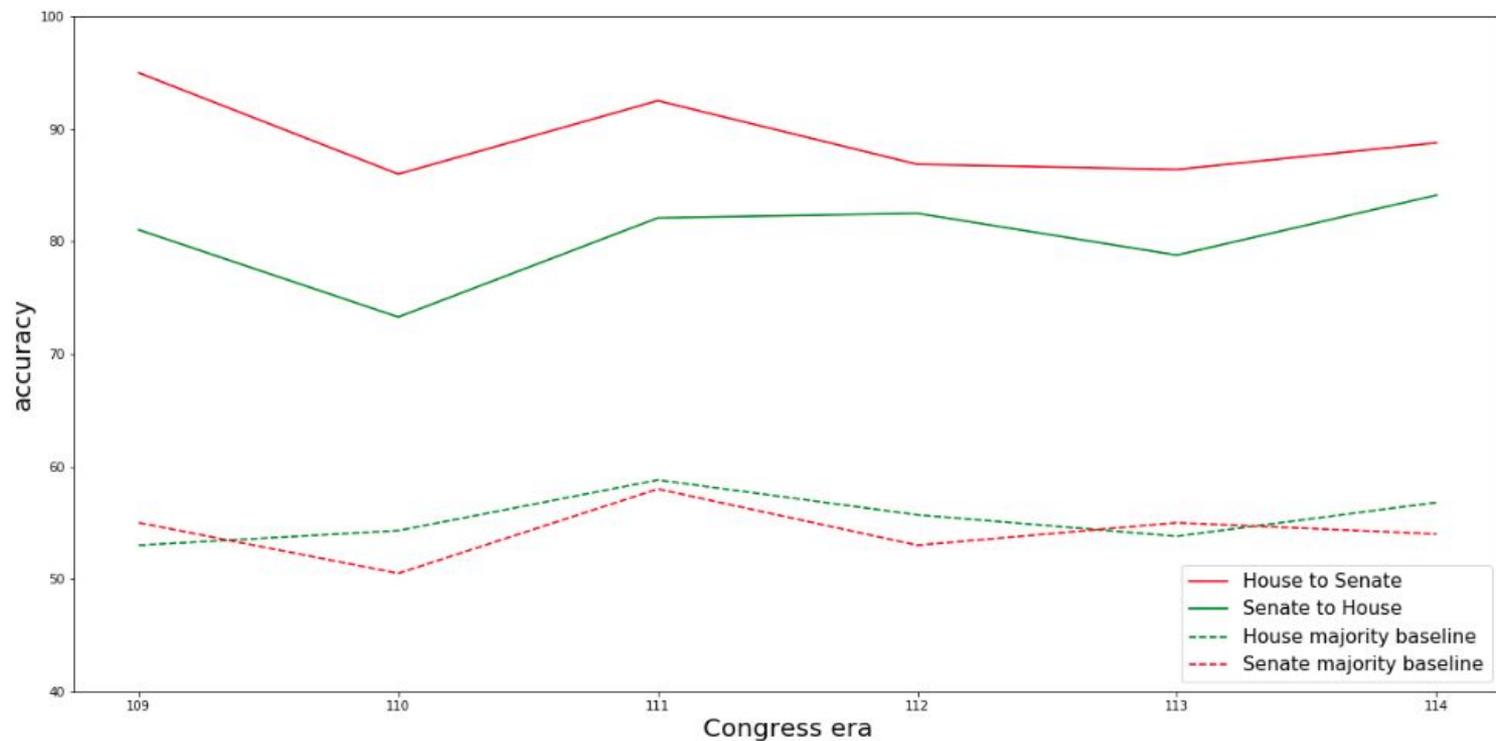
- Data source: Stanford Congressional Records 2006-2017 [5] (hein-daily)

Speeches.txt	SpeakerMap.txt
<ul style="list-style-type: none">- Speech_id- speech	<ul style="list-style-type: none">- Speaker_id- Speech_id- Lastname- Firstname- Chamber- State- Gender- party

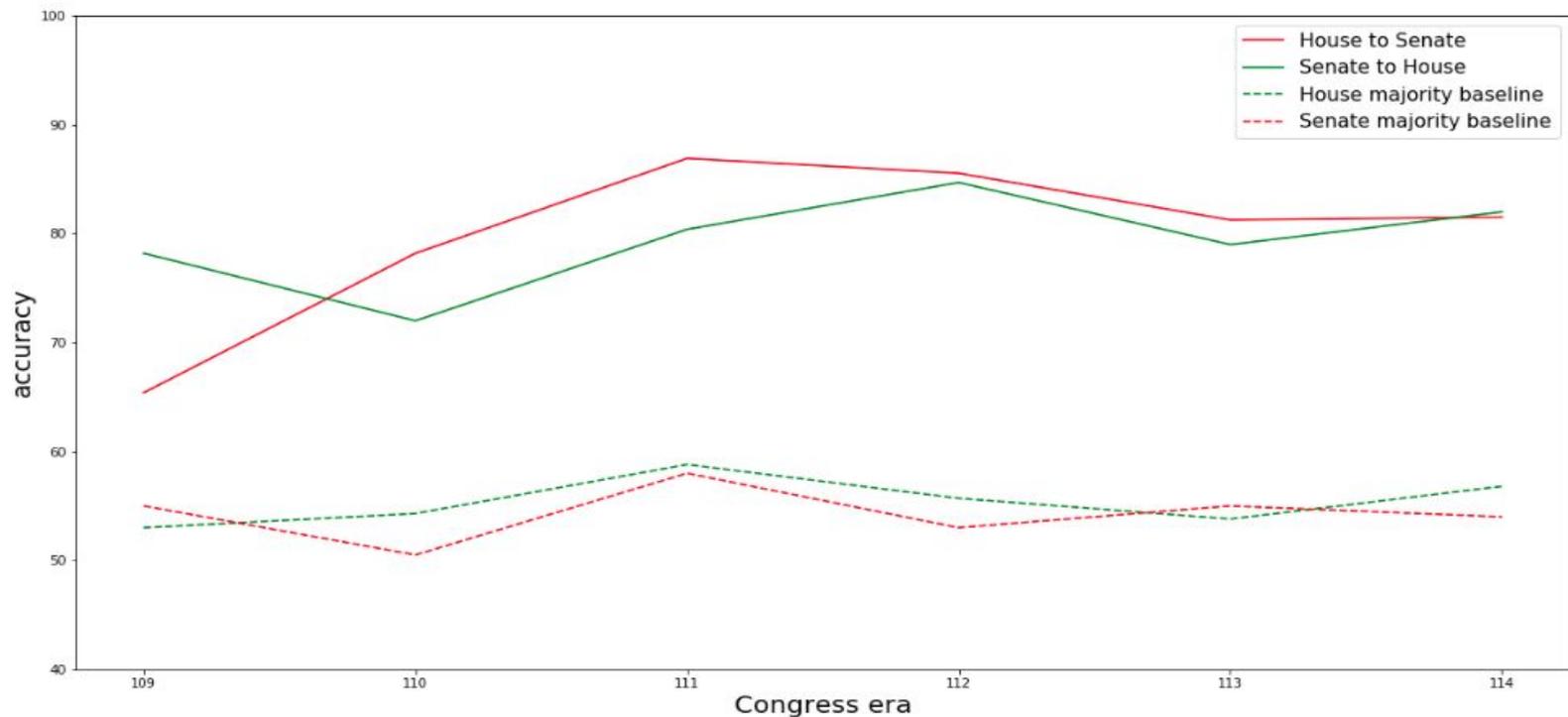
Method

- Classification: Support Vector Machine (SVM) [7][2]
- Vectorization: TF-IDF [3]
- Bag-of-Words [3][4]
- Remove State/People names [3] → avoid possible correlation between names and parties

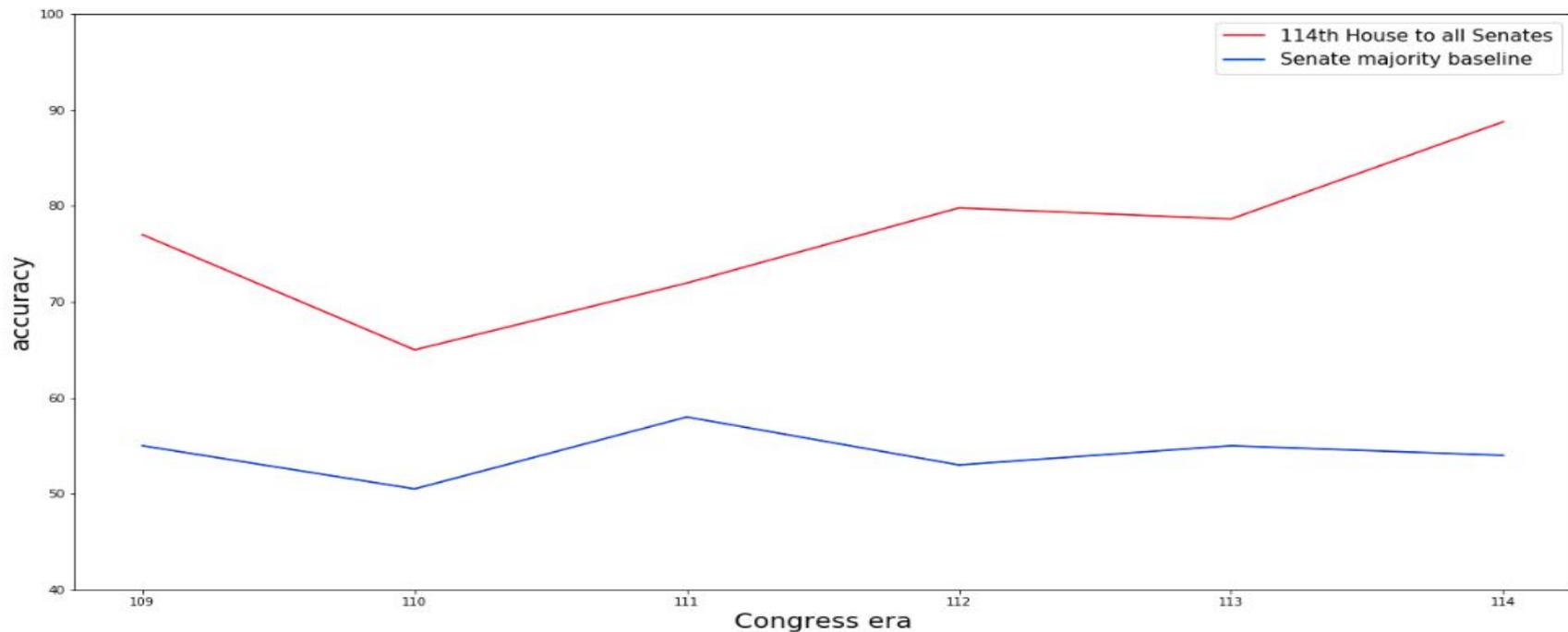
Result: House-trained vs Senate-trained



Result: House-trained (100 random) vs Senate-trained



Result: 114th House-trained, test on all Senate



Result: Top 20 features for Republican (114th) and Democrats (110th)

114th House: Obama administration

110th House: Bush administration

Republican (114th)	Democrats (110th)
Taxpayers, space, unborn, legislative , mental Days, ask , obamacare, remarks , members, extend , Iran, spending, actually, suspend Government, gentleman , extraneous , revise , administration	administration , revise , ask , suspend , consent Bush, unanimous, extraneous , material, remarks , move, war, blue, extend , legislative Thank, presidents, consideration, public gentleman

Result: Context example for “administration” and “ask”

Republican (114th)	Democrats (110th)
I ask the question to this administration and to the world: How many more Christians have to be blown up	were going to ask administration officials some tough questions so that we can make government work again for all Americans
I ask that this administration take this very seriously and consider religious freedom for all around the world	I ask the current administration to present this Congress with a comprehensive plan for Iraq
This administration asked us to trust Iran	The administration asks for this extension
but it seems the Obama administration is asking us to trust Iran and then trust some more	the administration has been unhappy when we have voted here more money than they have asked for
we repeatedly asked the administration to simply start over with a process that works with stakeholders to achieve the goals of the Clean Water Act	This administration has asked us to cut funding for children
but asking us to trust an administration	the same administration who brought us this disaster are now asking us to trust them again
We are asking the administration to use that money	But I would ask the administration to please look into trying some diplomatic efforts in the Middle East

Conclusion

- Political parties use vocabulary to identify themselves
- Machine Learning makes political speeches a rich source for party affiliation
- Partisanship is increasing in between US major parties in Congress
- Higher classification accuracy does not always indicates higher partisanship:
 - Data characteristic can also influence the result
- Ideology is **Not** the only factor which shapes our behavior, Political power relations also affects us.
 - To classify Ideology, other measures are also needed beside party affiliation

References

1. Brewer, Mark D. "The rise of partisanship and the expansion of partisan conflict within the American electorate." *Political Research Quarterly* 58, no. 2 (2005): 219-229.
2. Bayram, Ulya, John Pestian, Daniel Santel, and Ali A. Minai. "What's in a Word? Detecting Partisan Affiliation from Word Use in Congressional Speeches." In *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-8. IEEE, 2019.
3. Yu, Bei, Stefan Kaufmann, and Daniel Diermeier. "Classifying party affiliation from political speech." *Journal of Information Technology & Politics* 5, no. 1 (2008): 33-48.
4. Hirst, Graeme, Yaroslav Riabinin, and Jory Graham. "Party status as a confound in the automatic classification of political speech by ideology." In *Proceedings of the 10th International Conference on Statistical Analysis of Textual Data (JADT 2010)*, pp. 731-742. 2010.
5. Gentzkow, M., J. M. Shapiro, and M. Taddy. "Congressional Record for the 43rd-114th Congresses: Parsed speeches and phrase counts. Palo Alto, CA: Stanford Libraries." (2018).
6. Helms, Ludger. "Five ways of institutionalizing political opposition: lessons from the advanced democracies." *Government and opposition* 39, no. 1 (2004): 22-54.
7. Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." *Machine learning* 20, no. 3 (1995): 273-297.

Thanks for your attention!