Fact Checking Congressional Voting Claims

Brett Walenz, Junyang Gao, Emre Sonmez, Yubo Tian, Yuhao Wen, Charles Xu, Bill Adair, Jun Yang
Duke University
Computational Journalism Group
bwalenz@cs.duke.edu

ABSTRACT

Congressional voting records have historically been used both to attack and defend legislators during elections. Media spend a significant amount of effort fact checking claims on voting records, but manual checking requires considerable expertise and time and is difficult to scale up. As an aid to fact checkers and researchers, we have built a system called iCheck*. iCheck enables in-depth exploration of congressional voting records along multiple dimensions: time, political parties, special interest groups, and Presidential support. iCheck helps check claims and find interesting claims using a computational procedure called* perturbation analysis*, which automatically considers how a claim's conclusion changes in a large number of different contexts.*

1 Introduction

Congressional voting records are routinely dissected for various purposes, and often highlighted during election cycles. Consider the following two examples.

In a bid for reelection, Scott Brown, a Massachusetts Senator during 2010-2013, claimed he was the most bipartisan Senator during his time in the Senate. An analysis by PolitiFact [2] rated the claim as "Mostly True." It considered numbers from different news organizations, with some sources claiming Brown voted with Democrats 46% of the time, up to 62% of the time.

As another example, in the run for a North Carolina Senate seat, Thom Tillis used Kay Hagan's voting record with Barack Obama as an attack during the previous election cycle. Multiple news organizations [1, 4] devoted significant effort to fact checking this claim. Data for the analysis was limited and had to be collected and tabulated from multiple sources, with records of presidential support sourced from *Congressional Quarterly*.

Fact checking such claims is challenging because of the complexity of congressional voting records and nuances in the claims. These claims are often factually correct but nonetheless misleading. For example, how often a legislator agrees with a party or the President can change over time or over different subsets of the votes (e.g., all roll calls vs. key votes identified by a special interest group), sometimes significantly. In addition, what appears in the eye of public to be a high percentage of agreement (or disagreement) may turn out to be not so high once we put this number into the context of how other legislators have voted. The current practice of manually checking congressional voting claims demands a great deal of expertise and effort. It is difficult to cover all such claims, let alone extending coverage to state legislatures as well.

In response, we have designed the *iCheck* system with the intention of mitigating these pain points for journalists and other researchers using computational techniques.

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

2 Overview

The key computational technique underlying iCheck is *perturbation analysis*. We can view a claim as a parameterized query over the underlying data, where the parameters encode what specific (and often partial) view of the data that the claim presents to its audience. By perturbing these parameters and reevaluating the query, iCheck obtains a "bigger picture" with which we can assess the qualities of the claim such as robustness and uniqueness. Using the same analysis, iCheck can also find counterarguments that contradict the original claim, or just interesting claims in the first place.

Many human fact checkers implicitly apply this perturbation analysis, but formalizing this analysis as a computational problem and automating it allows iCheck to help humans tackle bigger data sets, perform more comprehensive analysis, and check more claims. We had a few guiding principles in the development of the iCheck system. First, it should be easy to research a specific claim. Second, it should provide an intuitive, visual representation of the "bigger picture" around the specific claim. Third, it should allow the user to "perturb" a claim easily by hand. Finally, iCheck should also automatically highlight interesting perturbations (counterarguments or otherwise).

Perturbation analysis can be computationally challenging too. We refer interested readers to [5, 3] for details on implementation and optimization.

3 Data

For this effort, data is integrated from a wide variety of sources. Voting records are collected from *govtrack.us* containing individual vote listings, as well as party, state, and committee information about each legislator. Presidential support during 2008-2014 is based on *Congressional Quarterly*. Key vote listings are obtained from various organizations, including the American Conservative Union, Americans for Democratic Action, AFL-CIO, and U.S. Chamber of Commerce. We also collect two metrics of "recognizability" for determining people of interest: their presence in traditional broadcast media as well as social media. Appearances in traditional broadcast media are tracked by *Congressional Quarterly*'s *Face Time* (http://media.cq.com/facetime/) on a weekly basis. Social media presence is measured by scores based on aggre-

gate statistics from the website *klout.org*. Lastly, we obtain metrics of how *safe* a House or Senate seat is from the Cook Political Report.

4 User Interface

4.1 Person-Centric View

The person-centric view is the primary means of analyzing claims with respect to a specific legislator. The initial claim is *fixed*, meaning the parameters are already set. From here, the user can perturb the parameters to find counterarguments or other viewpoints. The visualization is a trend line which shows vote agreement with the target over time. In Figure 4, the trend plot shows the time-varying average agreement percent with President Obama for each Senator as a line (Hagan shown in orange, near the top).

The color background in this plot shows the time-varying density distribution of agreement percentages; darker shade means higher density. Density contribution from Democrats (or Republicans) is shown in blue (or red). Purple indicates when a subgroup of particular Democrats and Republicans have similar agreement percent with President Obama. But for the most part, the two clusters have little overlap in Figure 4. In fact, Figure 4 shows an interesting division: Democrats are strongly clustered at around 95%, whereas Republicans are scattered throughout the rest of the plot.

The `Did you know?` section shows that Hagan does not rank that high among the Democrats: 49th percentile, right in the middle. Furthermore, this section shows that if we just look at *Congressional Quarterly* key votes in which Pres. Obama also had a position on in 2014, Hagan does not rank high either: 44th percentile among all senators (both Democrats and Republicans). Contrast this to 26th percentile when compared on all votes, and an argument can be made that Hagan tends to agree less with President Obama on key issues.

4.2 Population View

Population views compare and contrast how a group of legislators (e.g., by state or party) vote with a target: Republican majority, Democrat majority, or President. The comparison can be done over a different sets of key votes. The view illustrates an overall breakdown of how the population votes with the target. iCheck features controls so the user can manipulate the window of comparison manually, which can reveal trends—such as the 2013 government shutdown which shows the change in how the Republican party in the House votes with the President.

Rank The rank view shows each legislator as a bar, sorted by percent agreement vote with the target of comparison.

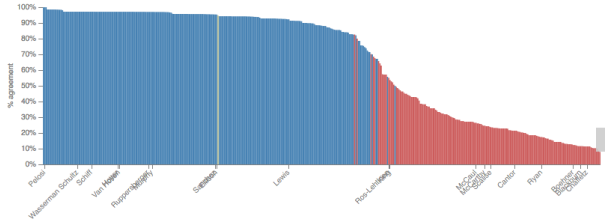


Figure 1: Comparing House of Representatives voting with President from 2008-2009.

Trend The trend view shows each legislator as a line. Most recognizable (see Section 3) legislators are highlighted, while the others are summarized as density plots (similar to the person-centric view).

2013 Government Shutdown We close this section with an illuminating example. The United States government shutdown of

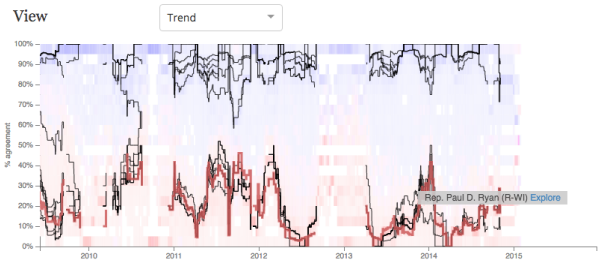


Figure 2: Comparing House of Representatives voting with President in trend view.

2013 was preceded by a lengthy political battle between House Republicans and President Obama, and the results can be seen clearly in iCheck. In Figure 1, we constrain the window of comparison to be 2009-2010, while in Figure 3, we constrain the window to be 2013-2014, during the government shutdown. Both charts show how each member voted with President Obama's supported position. When you compare the two figures, it is obvious how drastically lower the average Republican votes with President Obama's position.

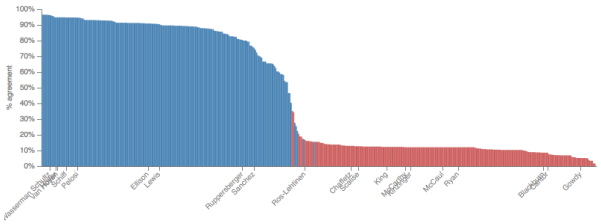


Figure 3: Rank view of House members vs. President from 2013-2014.

5 Conclusion

Current fact checking of claims regarding congressional voting records is limited by the difficulty of performing analysis. iCheck is designed as a tool to aid fact checkers and researchers in checking such claims. We collect data from sources that are traditionally difficult to integrate. Visualizations let users perform analysis on their own, while automated perturbation analysis aids the user by highlighting interesting facts and counterarguments.

References

[1] Mark Binker. Fact check: Does hagan vote with obama 95 percent of time? 2014. URL http://www.wral.com/fact-check-does-hagan-vote-with-obama-95-percent-of-the-time-/13930106/.

[2] Jim Haddadin. Scott brown says he was the most bipartisan senator, 2014. URL http://www.politifact.com/new-hampshire/statements/2014/mar/27/scott-brown/scott-brown-says-he-was-most-bipartisan-senator/.

[3] Brett Walenz and Jun Yang. Perturbation analysis of database queries. *Proceedings of the VLDB Endowment*, 9(14), 2016.

[4] Tim Waller. Truth check: Does kay hagan 'vote with obama 96 percent of time'? 2014. URL http://www.wyff4.com/news/truth-check-does-kay-hagan-vote-with-obama-96-percent-of-time/29162956.

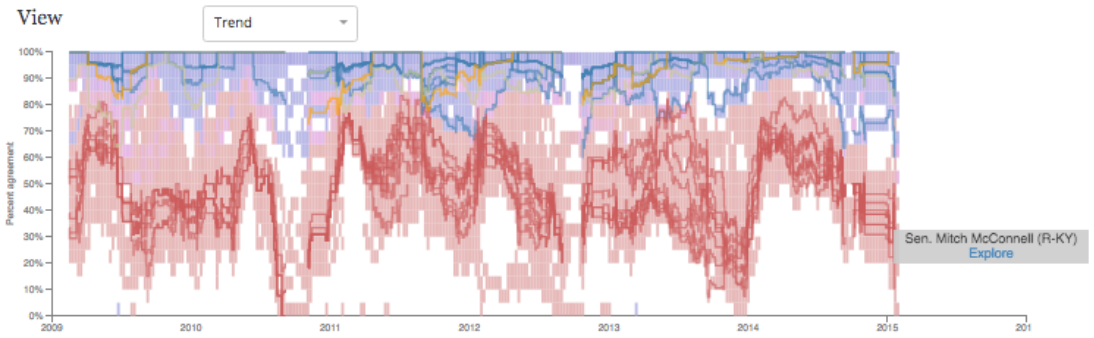
[5] You Wu, Pankaj K. Agarwal, Chengkai Li, Jun Yang, and Cong Yu. Toward computational fact-checking. *Proceedings of the VLDB Endowment*, 7(7):589–600, 2014.

Hagan voted with President 96% during Jan. 1, 2009-Jan. 1, 2016, over all 2178 votes.

Among Senators, this record ranked the 59th (top 42nd percentile) out of 139 (27 excluded for insufficient history).

Breakdown

The **rank** visualization shows vote agreement with President between Jan. 1, 2009 - Jan. 1, 2016. The **trend** visualization shows vote agreement over time, and shows the most popular members in the group, with the other members represented by shaded regions.



Compare and Contrast

You can compare the current analysis to other settings, such as: different time periods, populations (state vs. national), or along specific key vote listings.

Compare and Contrast

How They Stack Up

- Among Democratic Senators, this record ranked the 57th (bottom 25th percentile) out of 75 (5 excluded for insufficient history)
- Among Senators of NC, this record ranked the 1st out of 2 (2 excluded for insufficient history)
- Among female Senators, this record ranked the 14th (bottom 43rd percentile) out of 23 (4 excluded for insufficient

Figure 4: Vote Correlation Analysis for Kay Hagan