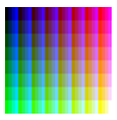
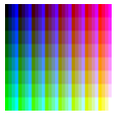
An Adaptable Legislative Analysis and Advocacy Framework

Matt Weiden
Advised by Xifeng Yan
University of California, Santa Barbara
2012



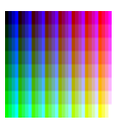
Problem Description

1. Political advocacy serves a special purpose in governance
2. Two core advocacy challenges
    1. Quantifying legislator or constituent support of a cause
    2. Motivating these groups to take action
3. What can be improved?
    1. Generating this analysis is time consuming, biased
    2. Current methods of quantifying support can be improved upon



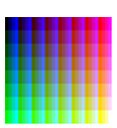
Addressing this Problem

1. A Web & Machine Learning-Based Framework for Advocates
    1. Adaptable - can be used for any state legislature or Congress
    2. Advocacy - Uses statistical methods to rank legislators latent support, future actions
    3. Accountability - Information published to web applauds those supportive, holds accountable those who are not



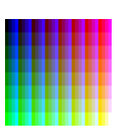
Presentation Outline

1. Method
    1. Workflow
        1. Bill Review
        2. Dataset Completion
        3. Crosscheck
        4. Prediction
    2. Stuff under the hood
        1. Missing Values (k-NN)
        2. Prediction (Random Forests)
2. Experiment
    1. Design
    2. Procedure
    3. Results / Method Logic
3. Discussion



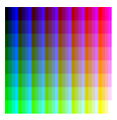
Method

1. Two Major Components
    1. web interface - publishing and workflow for data input, output
    2. stuff under the hood - machine learning, run as scheduled jobs

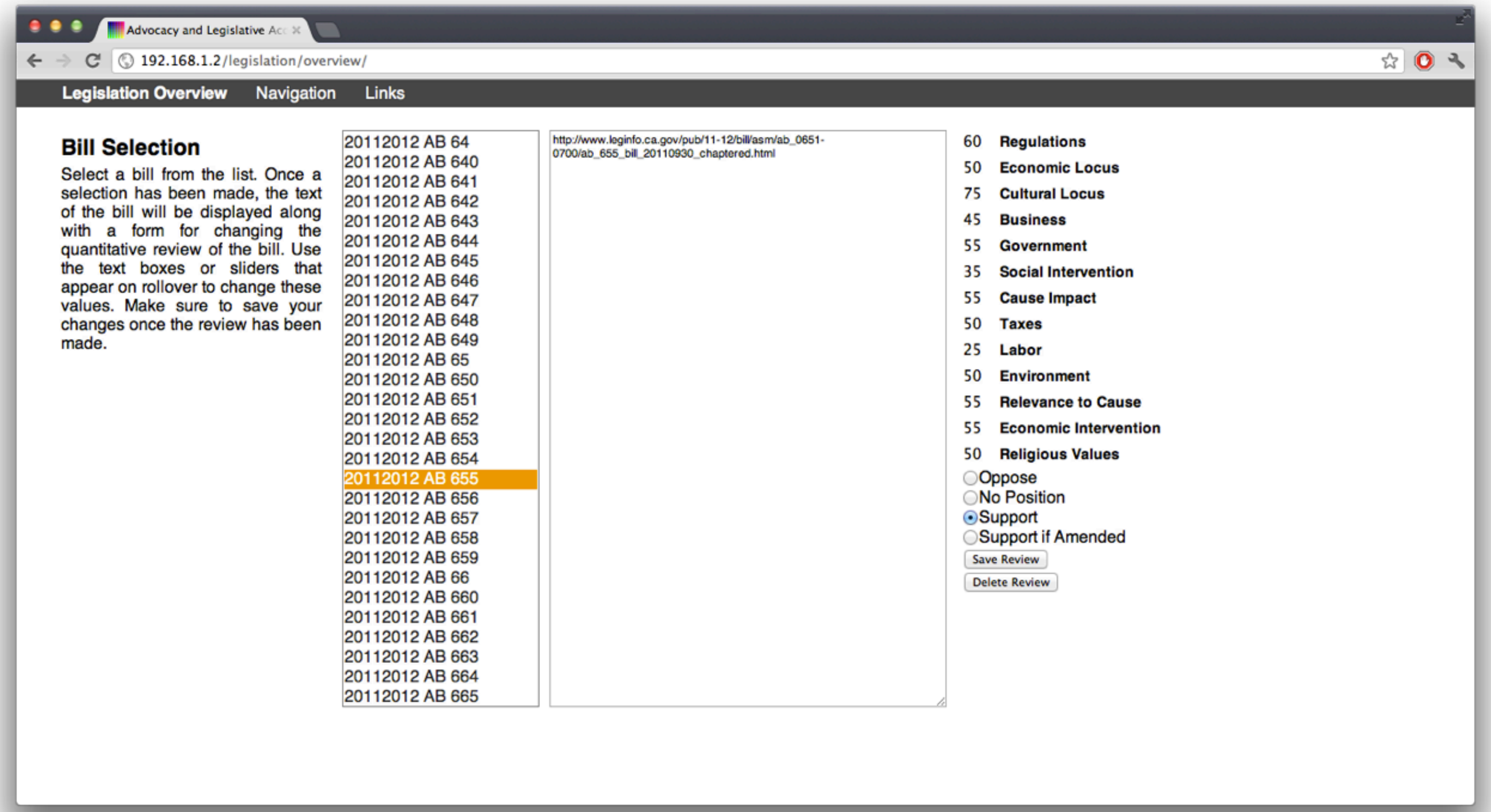


Bill Review

1. Motivation
    1. How do we determine the stance of the advocate, legislators?
2. Solution
    1. "Bill Reviews" - Quantitative profiles of bills
        1. Format Machine Learning algorithms can understand
    2. Voting Records
        1. Ideal vote types give information stance of the advocate
        2. Publicly available legislative voting information gives position of legislators

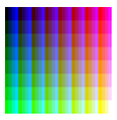


Bill Review


Advocacy and Legislative Acc
192.168.1.2/legislation/overview/
Legislation Overview
Navigation
Links
Bill Selection
Select a bill from the list. Once a selection has been made, the text of the bill will be displayed along with a form for changing the quantitative review of the bill. Use the text boxes or sliders that appear on rollover to change these values. Make sure to save your changes once the review has been made.
20112012 AB 64
20112012 AB 640
20112012 AB 641
20112012 AB 642
20112012 AB 643
20112012 AB 644
20112012 AB 645
20112012 AB 646
20112012 AB 647
20112012 AB 648
20112012 AB 649
20112012 AB 65
20112012 AB 650
20112012 AB 651
20112012 AB 652
20112012 AB 653
20112012 AB 654
20112012 AB 655
20112012 AB 656
20112012 AB 657
20112012 AB 658
20112012 AB 659
20112012 AB 66
20112012 AB 660
20112012 AB 661
20112012 AB 662
20112012 AB 663
20112012 AB 664
20112012 AB 665
http://www.leginfo.ca.gov/pub/11-12/bill/asm/ab_0651-0700/ab_655_bill_20110930_chaptered.html
60 Regulations
50 Economic Locus
75 Cultural Locus
45 Business
55 Government
35 Social Intervention
55 Cause Impact
50 Taxes
25 Labor
50 Environment
55 Relevance to Cause
55 Economic Intervention
50 Religious Values
Oppose
No Position
Support
Support if Amended
Save Review
Delete Review

Bill Review: Dimensions

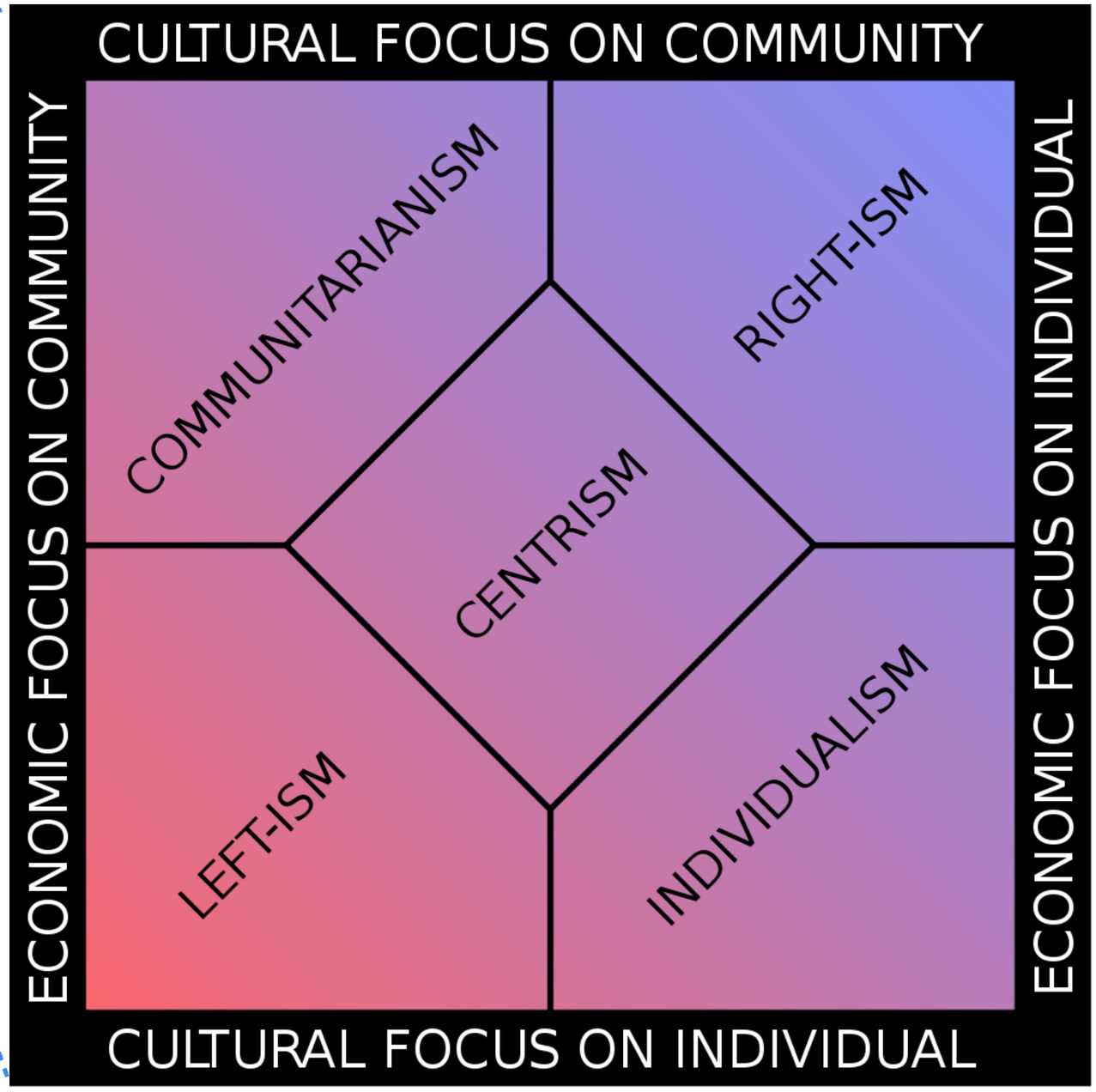
1. Customizable
2. Prototype Dimensions
    1. Chosen based on preliminary research & folk-political science
    2. Chosen for discriminative power



Bill Review: Dimensions

Prototype Review Dimensions

1. Cultural Locus
2. Economic Locus
3. Cultural Intervention
4. Economic Intervention
5. Government
6. Regulations
7. Taxes
8. Environment
9. Business
10. Labor
11. Religious Values
12. Cause Impact
13. Relevance to Cause



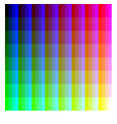
The Nolan Chart: Used by many libertarians as a political spectrum.

Bill Review: Dimensions

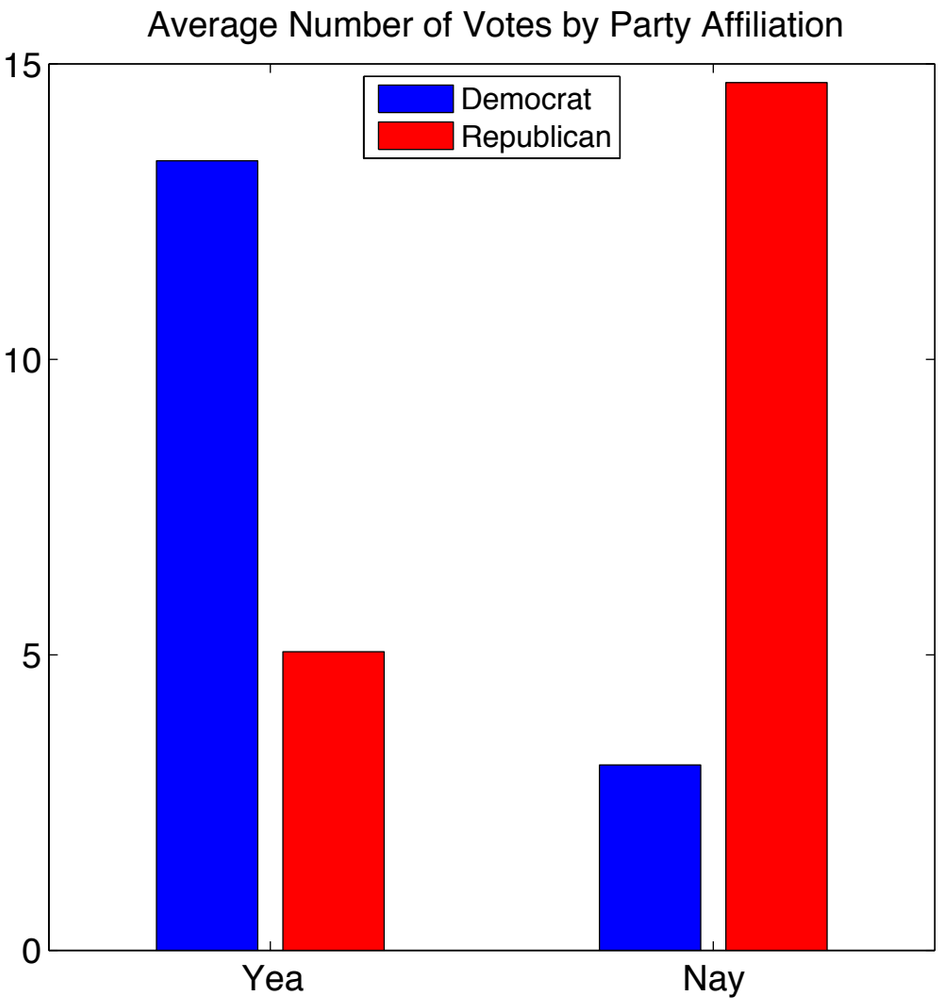
1. Further research would be necessary to select more discriminative dimensions

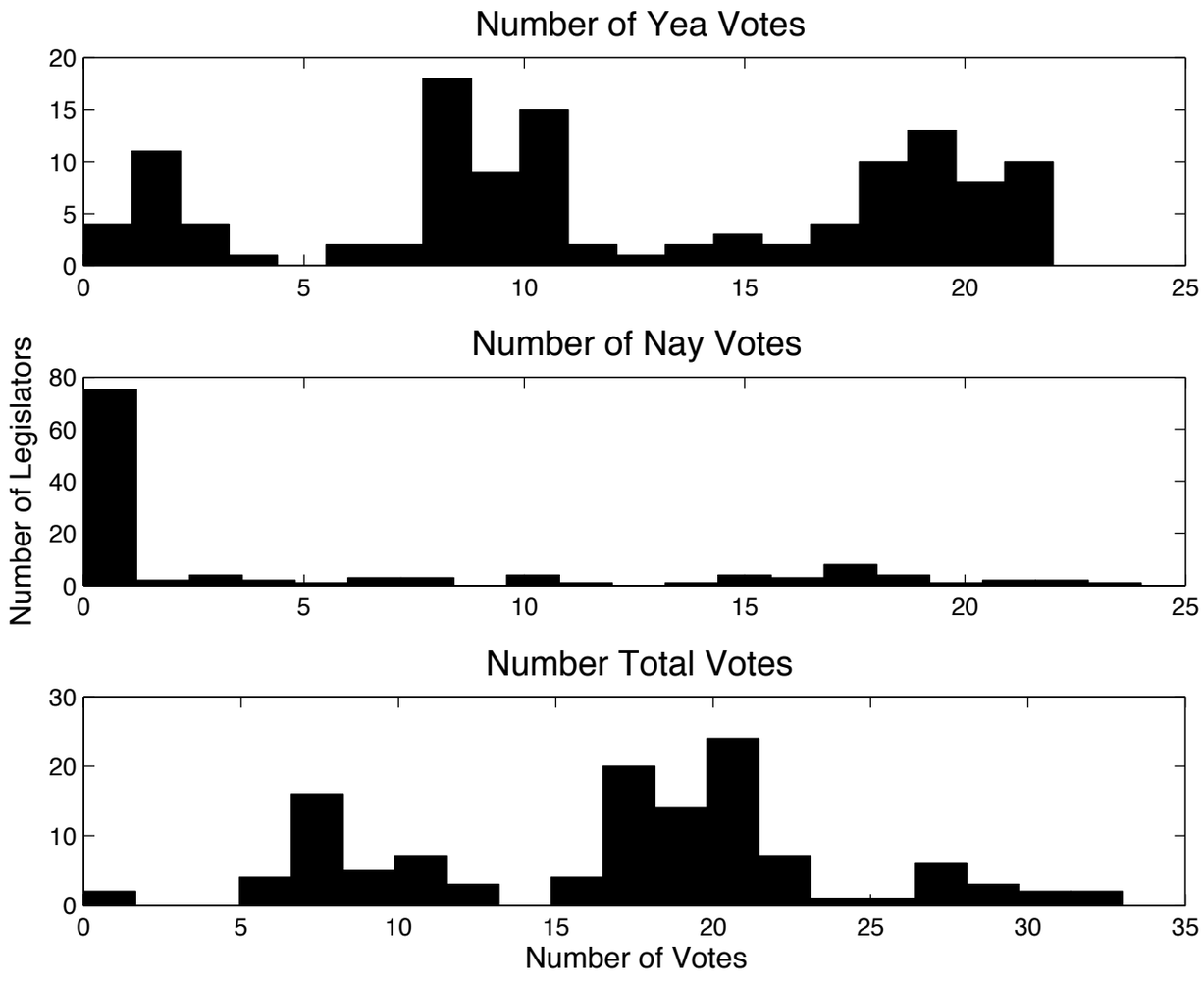
2. Limitations

    1. Only takes into account “position-issues” not “valence-issues”

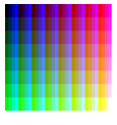


Dataset Completion

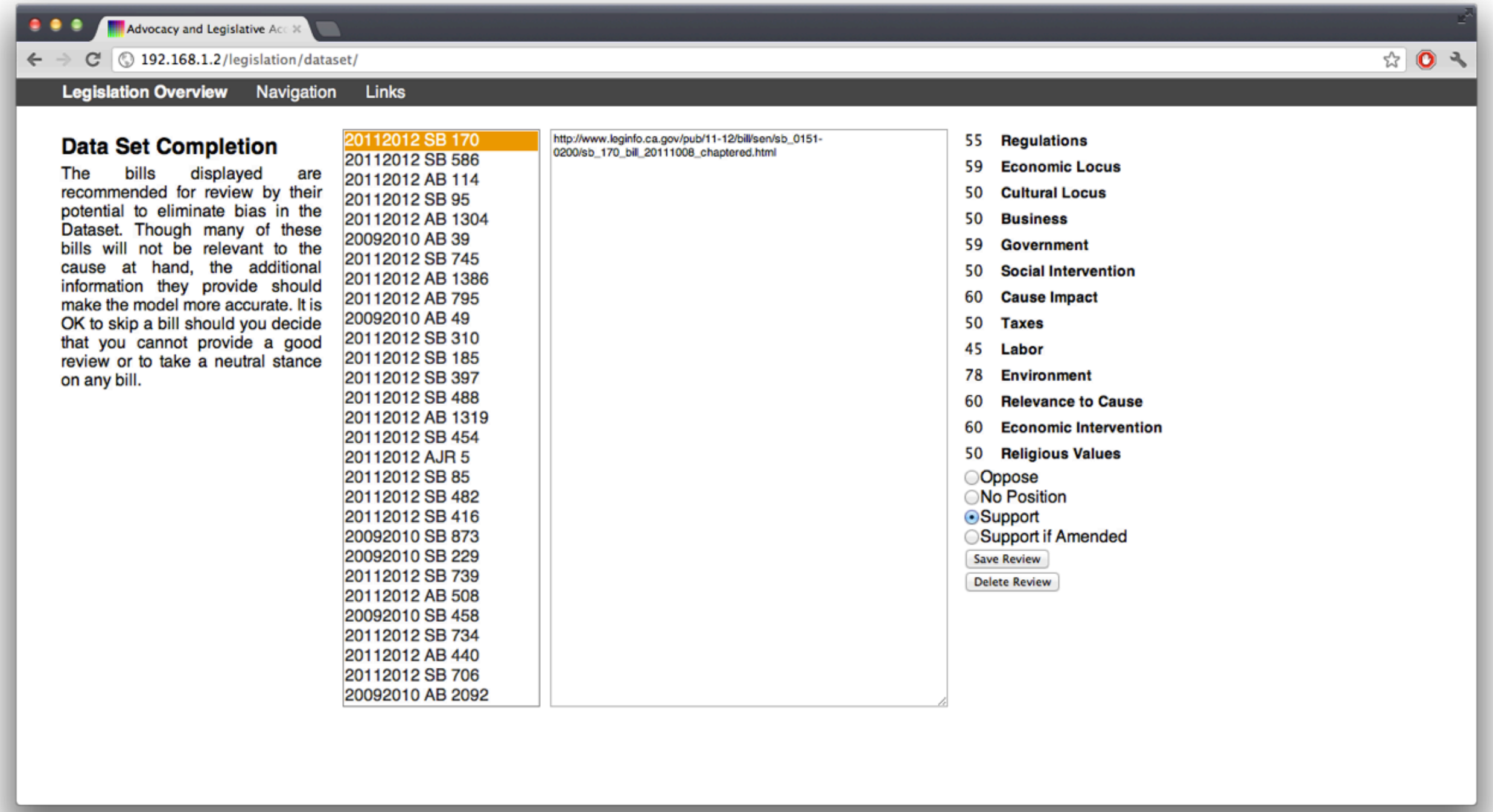




1. Motivation
    1. Voting record data is very biased, has many missing values
2. Solution
    1. Intelligently introduce more data to reduce bias.



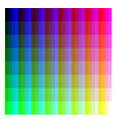
Dataset Completion



Dataset Completion

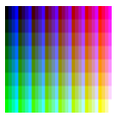
1. Looks the same as Bill Review, but it *isn't*
    1. Suggested by ability to increase vote-type entropy
    2. Improves classifier performance

$$C(V) = \arg\max_{X} - \sum_{i=1}^{n} p(v_i \cup x_i) \log_2 p(v_i \cup x_i)$$

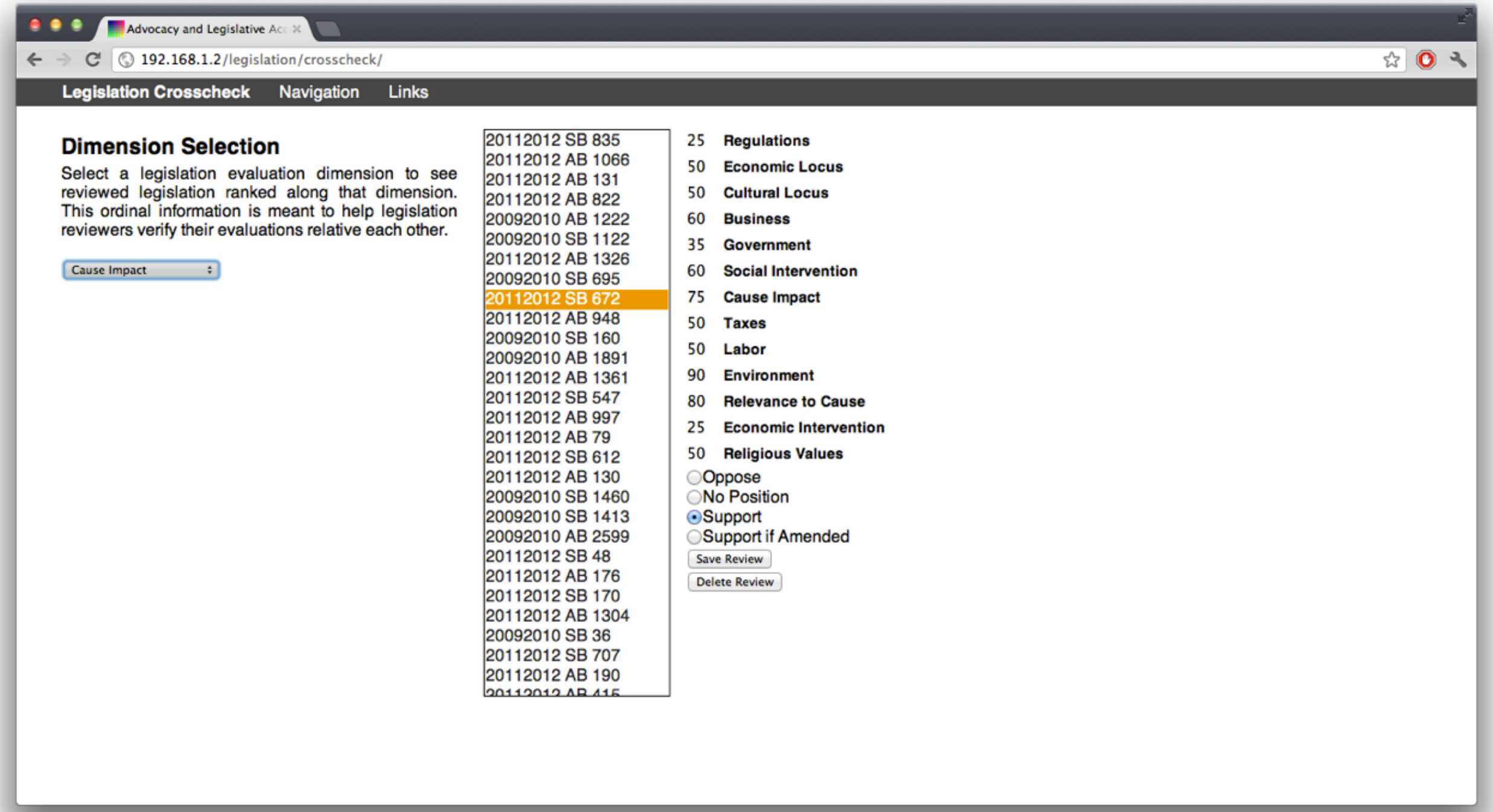


Crosscheck

1. Motivation
    1. What if advocate's scaling on dimensions isn't consistent between bills?
    2. What if different advocates are reviewing bills?
2. Solution
    1. Sort bills by a selected dimension
        1. Advocate can compare reviews, check for consistency.
        2. Make adjustments if necessary.

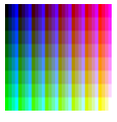


Crosscheck

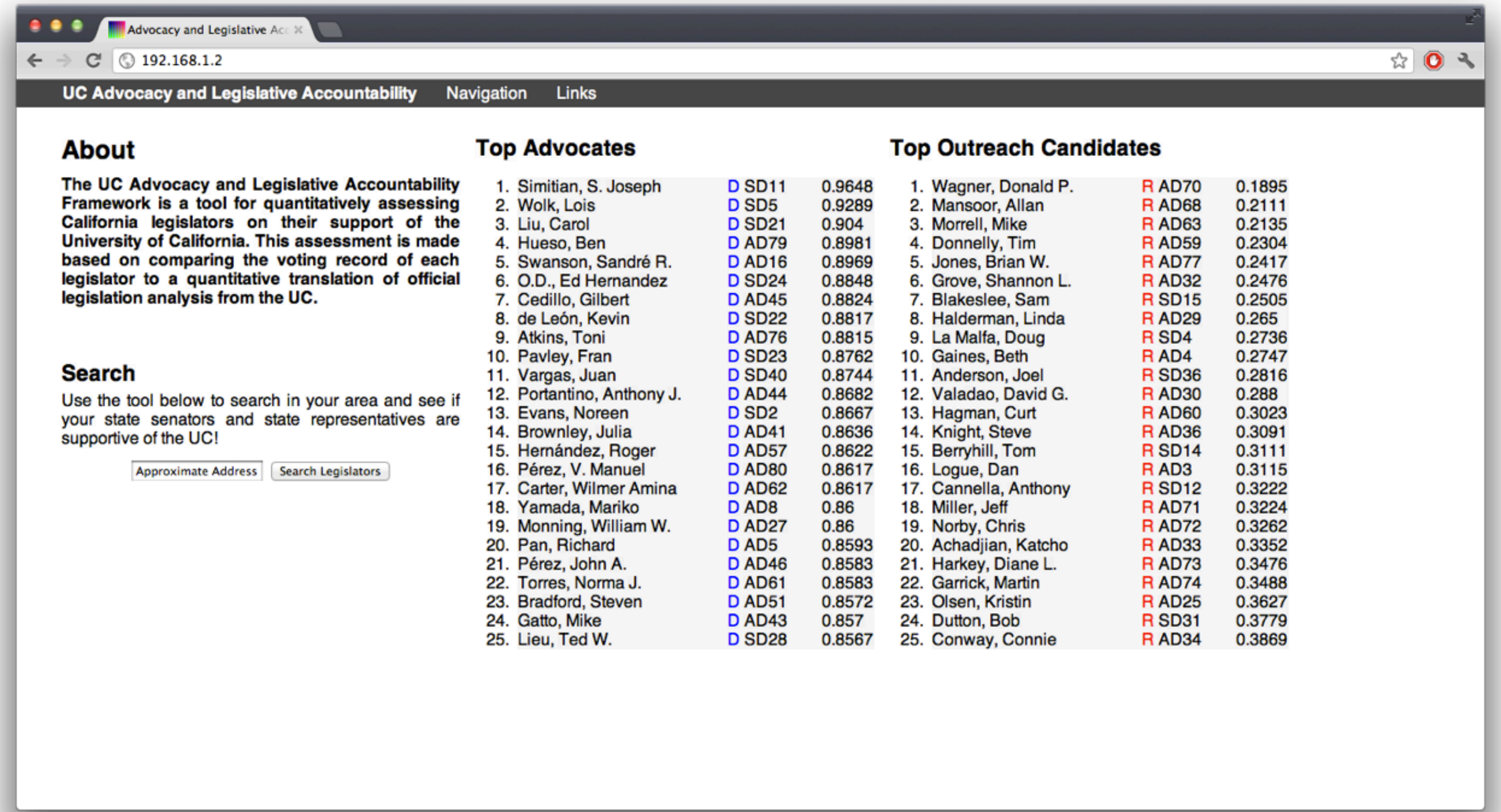

Advocacy and Legislative Acc
192.168.1.2/legislation/crosscheck/
Legislation Crosscheck
Navigation
Links
Dimension Selection
Select a legislation evaluation dimension to see reviewed legislation ranked along that dimension. This ordinal information is meant to help legislation reviewers verify their evaluations relative each other.
Cause Impact
20112012 SB 835
20112012 AB 1066
20112012 AB 131
20112012 AB 822
20092010 AB 1222
20092010 SB 1122
20112012 AB 1326
20092010 SB 695
20112012 SB 672
20112012 AB 948
20092010 SB 160
20092010 AB 1891
20112012 AB 1361
20112012 SB 547
20112012 AB 997
20112012 AB 79
20112012 SB 612
20112012 AB 130
20092010 SB 1460
20092010 SB 1413
20092010 AB 2599
20112012 SB 48
20112012 AB 176
20112012 SB 170
20112012 AB 1304
20092010 SB 36
20112012 SB 707
20112012 AB 190
25 Regulations
50 Economic Locus
50 Cultural Locus
60 Business
35 Government
60 Social Intervention
75 Cause Impact
50 Taxes
50 Labor
90 Environment
80 Relevance to Cause
25 Economic Intervention
50 Religious Values
Oppose
No Position
Support
Support if Amended
Save Review
Delete Review

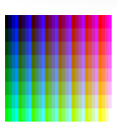
Prediction

1. Motivation
    1. What about novel legislation?
    2. How can we use the framework to target advocacy efforts?
    3. Can we improve upon current methods of quantifying support for a cause? Ranking?
2. Solution
    1. Machine Learning model and novel bill can predict a legislature's voting behavior.
    2. Introduce a new ranking metric: Average Probability of Agreement

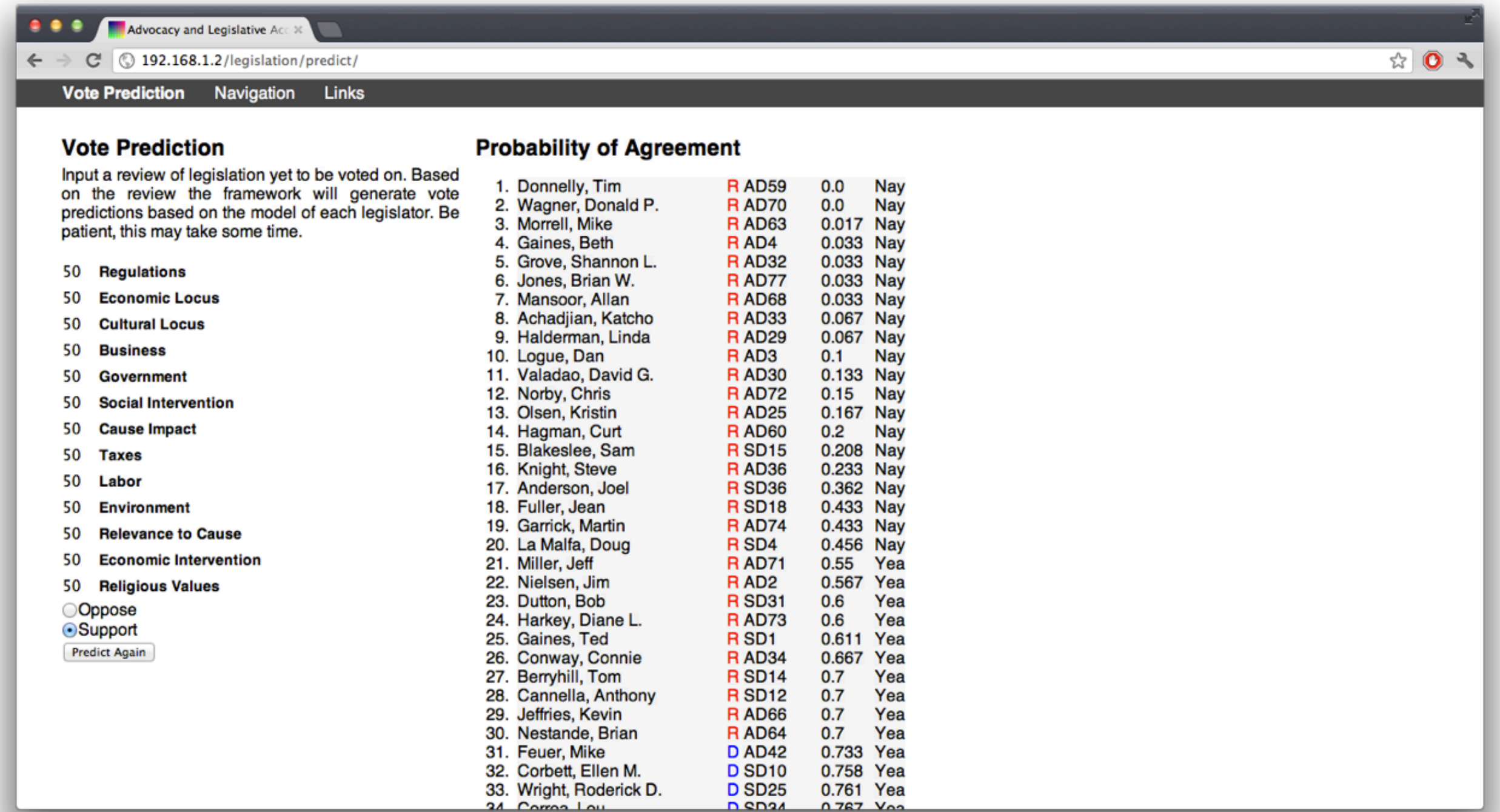


Prediction


Advocacy and Legislative Acc
192.168.1.2
UC Advocacy and Legislative Accountability
Navigation
Links
About
The UC Advocacy and Legislative Accountability Framework is a tool for quantitatively assessing California legislators on their support of the University of California. This assessment is made based on comparing the voting record of each legislator to a quantitative translation of official legislation analysis from the UC.
Search
Use the tool below to search in your area and see if your state senators and state representatives are supportive of the UC!
Approximate Address
Search Legislators
Top Advocates
1. Simitian, S. Joseph D SD11 0.9648
2. Wolk, Lois D SD5 0.9289
3. Liu, Carol D SD21 0.904
4. Hueso, Ben D AD79 0.8981
5. Swanson, Sandré R. D AD16 0.8969
6. O.D., Ed Hernandez D SD24 0.8848
7. Cedillo, Gilbert D AD45 0.8824
8. de León, Kevin D SD22 0.8817
9. Atkins, Toni D AD76 0.8815
10. Pavley, Fran D SD23 0.8762
11. Vargas, Juan D SD40 0.8744
12. Portantino, Anthony J. D AD44 0.8682
13. Evans, Noreen D SD2 0.8667
14. Brownley, Julia D AD41 0.8636
15. Hernández, Roger D AD57 0.8622
16. Pérez, V. Manuel D AD80 0.8617
17. Carter, Wilmer Amina D AD62 0.8617
18. Yamada, Mariko D AD8 0.86
19. Monning, William W. D AD27 0.86
20. Pan, Richard D AD5 0.8593
21. Pérez, John A. D AD46 0.8583
22. Torres, Norma J. D AD61 0.8583
23. Bradford, Steven D AD51 0.8572
24. Gatto, Mike D AD43 0.857
25. Lieu, Ted W. D SD28 0.8567
Top Outreach Candidates
1. Wagner, Donald P. R AD70 0.1895
2. Mansoor, Allan R AD68 0.2111
3. Morrell, Mike R AD63 0.2135
4. Donnelly, Tim R AD59 0.2304
5. Jones, Brian W. R AD77 0.2417
6. Grove, Shannon L. R AD32 0.2476
7. Blakeslee, Sam R SD15 0.2505
8. Halderman, Linda R AD29 0.265
9. La Malfa, Doug R SD4 0.2736
10. Gaines, Beth R AD4 0.2747
11. Anderson, Joel R SD36 0.2816
12. Valadao, David G. R AD30 0.288
13. Hagman, Curt R AD60 0.3023
14. Knight, Steve R AD36 0.3091
15. Berryhill, Tom R SD14 0.3111
16. Logue, Dan R AD3 0.3115
17. Cannella, Anthony R SD12 0.3222
18. Miller, Jeff R AD71 0.3224
19. Norby, Chris R AD72 0.3262
20. Achadjian, Katcho R AD33 0.3352
21. Harkey, Diane L. R AD73 0.3476
22. Garrick, Martin R AD74 0.3488
23. Olsen, Kristin R AD25 0.3627
24. Dutton, Bob R SD31 0.3779
25. Conway, Connie R AD34 0.3869



Prediction

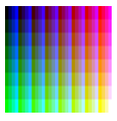

Advocacy and Legislative Ac
192.168.1.2/legislation/predict/
Vote Prediction
Navigation
Links
Vote Prediction
Input a review of legislation yet to be voted on. Based on the review the framework will generate vote predictions based on the model of each legislator. Be patient, this may take some time.
50 Regulations
50 Economic Locus
50 Cultural Locus
50 Business
50 Government
50 Social Intervention
50 Cause Impact
50 Taxes
50 Labor
50 Environment
50 Relevance to Cause
50 Economic Intervention
50 Religious Values
Oppose
Support
Predict Again
Probability of Agreement
1. Donnelly, Tim R AD59 0.0 Nay
2. Wagner, Donald P. R AD70 0.0 Nay
3. Morrell, Mike R AD63 0.017 Nay
4. Gaines, Beth R AD4 0.033 Nay
5. Grove, Shannon L. R AD32 0.033 Nay
6. Jones, Brian W. R AD77 0.033 Nay
7. Mansoor, Allan R AD68 0.033 Nay
8. Achadjian, Katcho R AD33 0.067 Nay
9. Halderman, Linda R AD29 0.067 Nay
10. Logue, Dan R AD3 0.1 Nay
11. Valadao, David G. R AD30 0.133 Nay
12. Norby, Chris R AD72 0.15 Nay
13. Olsen, Kristin R AD25 0.167 Nay
14. Hagman, Curt R AD60 0.2 Nay
15. Blakeslee, Sam R SD15 0.208 Nay
16. Knight, Steve R AD36 0.233 Nay
17. Anderson, Joel R SD36 0.362 Nay
18. Fuller, Jean R SD18 0.433 Nay
19. Garrick, Martin R AD74 0.433 Nay
20. La Malfa, Doug R SD4 0.456 Nay
21. Miller, Jeff R AD71 0.55 Yea
22. Nielsen, Jim R AD2 0.567 Yea
23. Dutton, Bob R SD31 0.6 Yea
24. Harkey, Diane L. R AD73 0.6 Yea
25. Gaines, Ted R SD1 0.611 Yea
26. Conway, Connie R AD34 0.667 Yea
27. Berryhill, Tom R SD14 0.7 Yea
28. Cannella, Anthony R SD12 0.7 Yea
29. Jeffries, Kevin R AD66 0.7 Yea
30. Nestande, Brian R AD64 0.7 Yea
31. Feuer, Mike D AD42 0.733 Yea
32. Corbett, Ellen M. D SD10 0.758 Yea
33. Wright, Roderick D. D SD25 0.761 Yea

Stuff Under the Hood

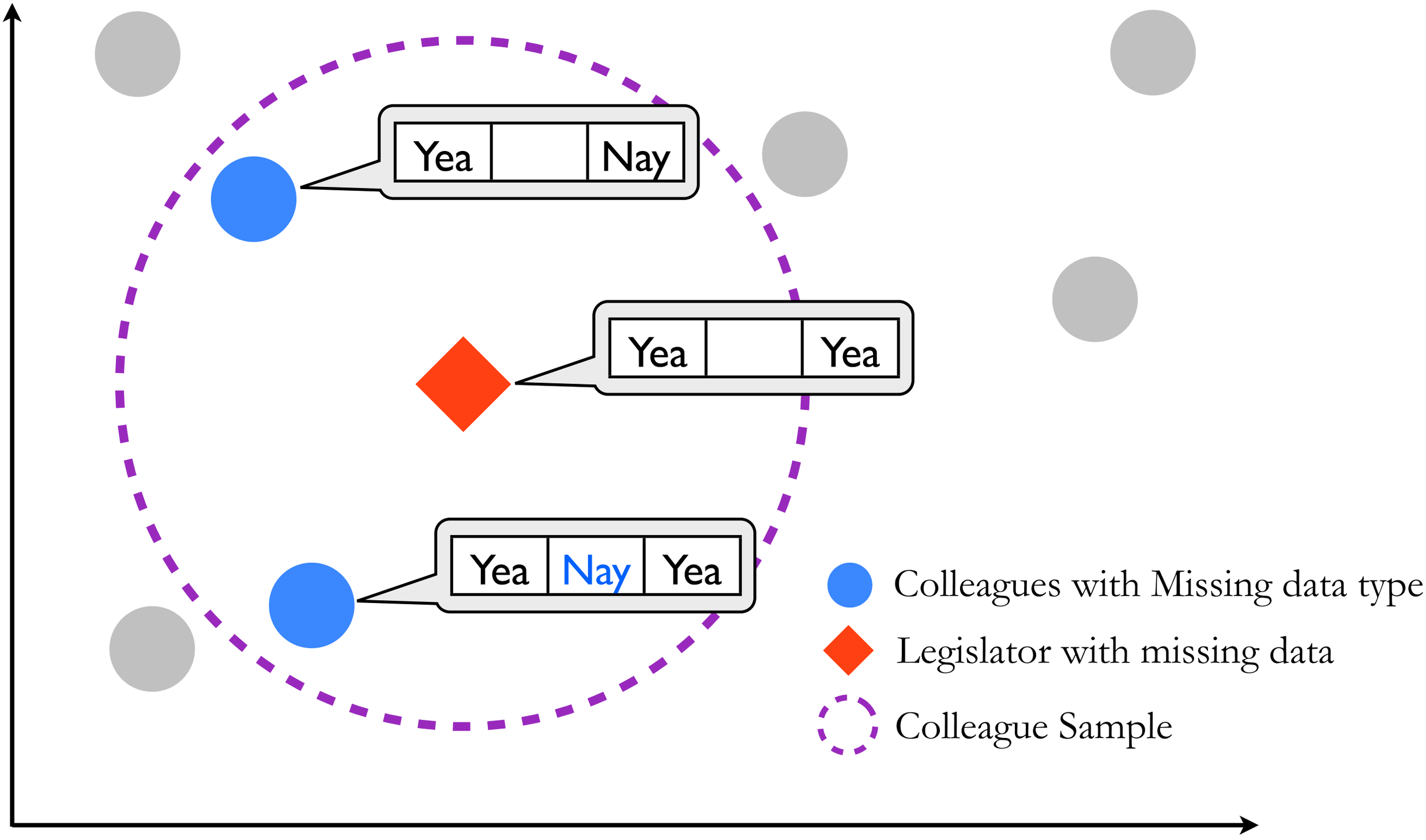
1. Major Challenges to Approach

    1. How do we fill in missing vote types for biased legislators? How do their colleagues vote?

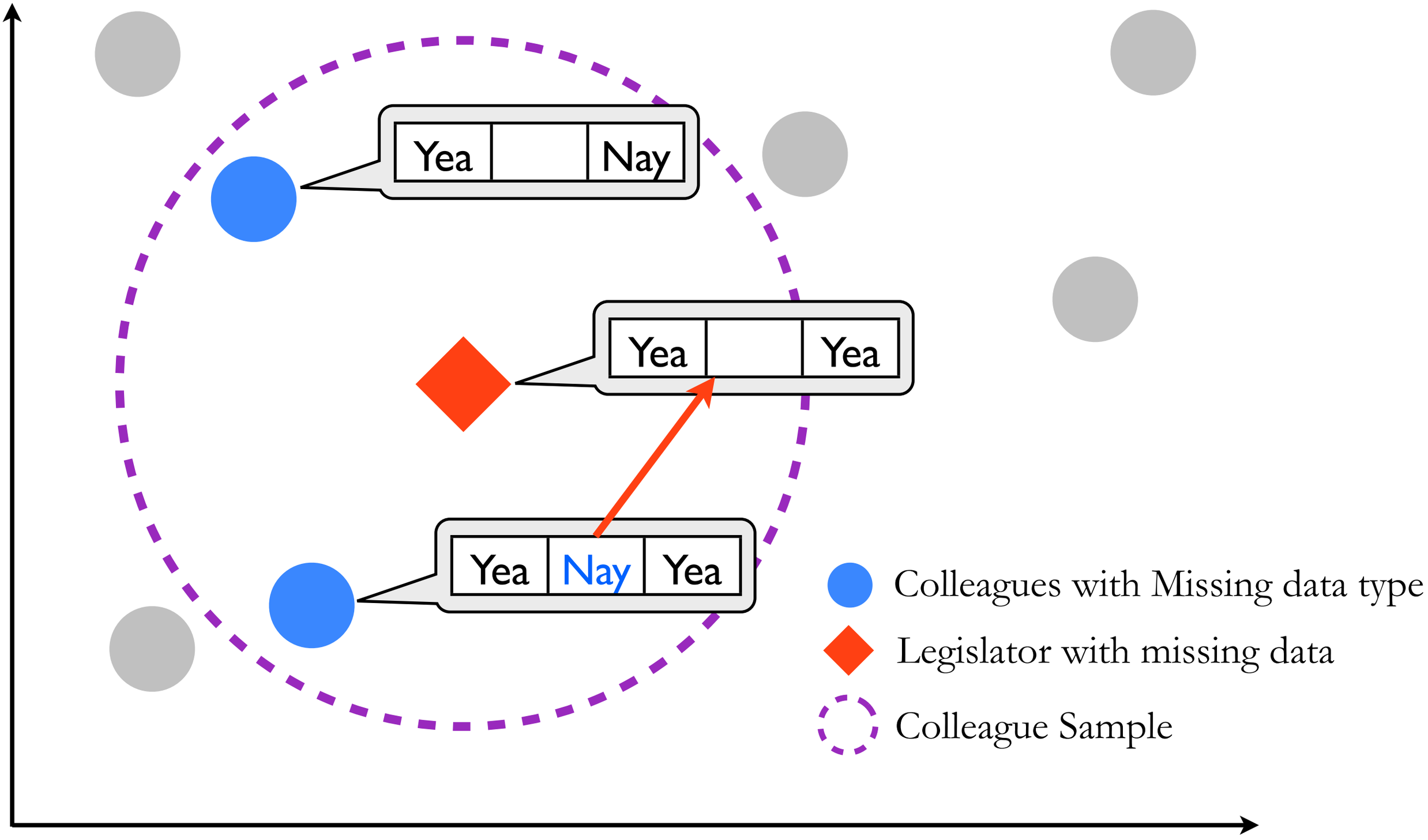
        1. k-Nearest Neighbor vote replacement.



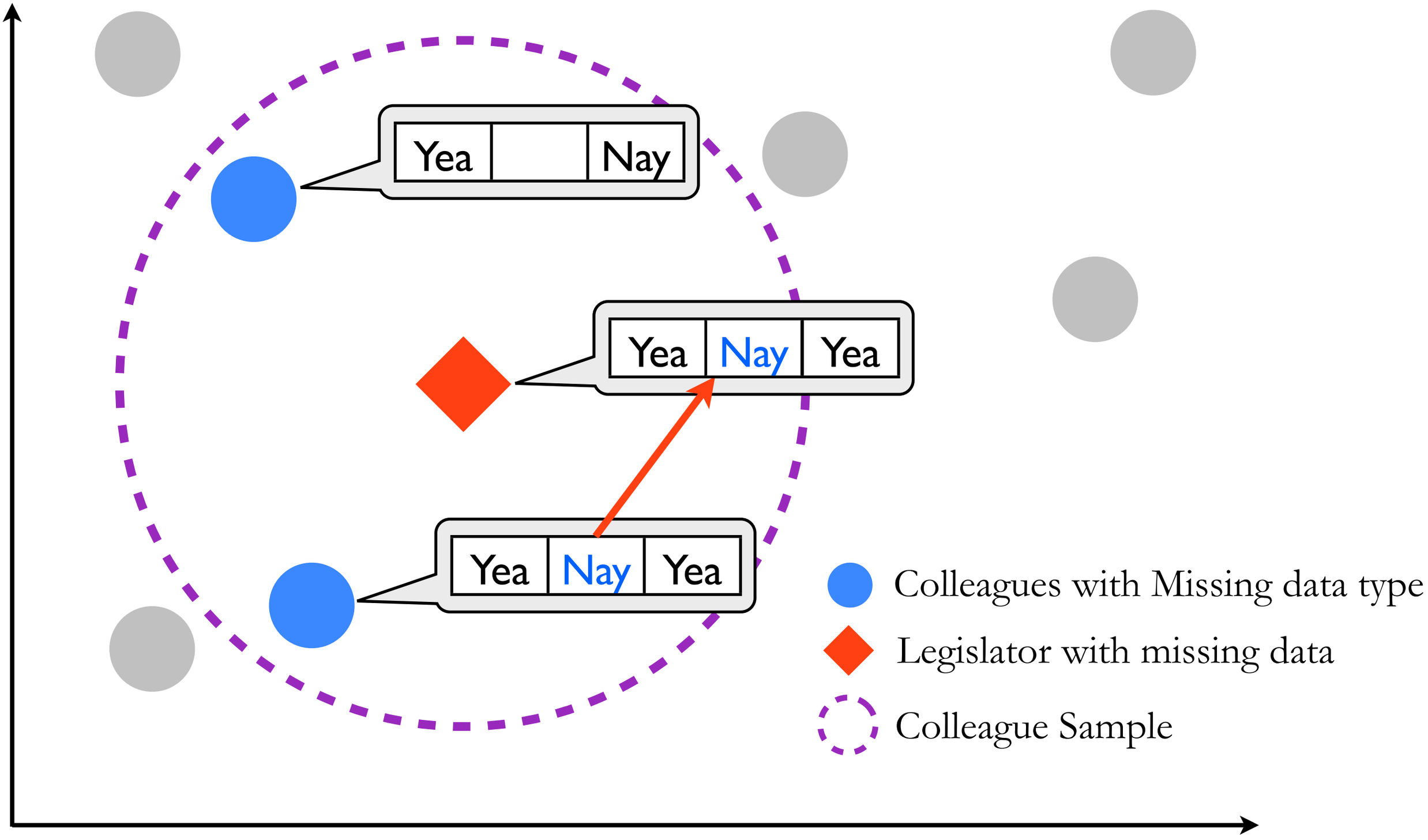
Filling in Missing Data



Filling in Missing Data



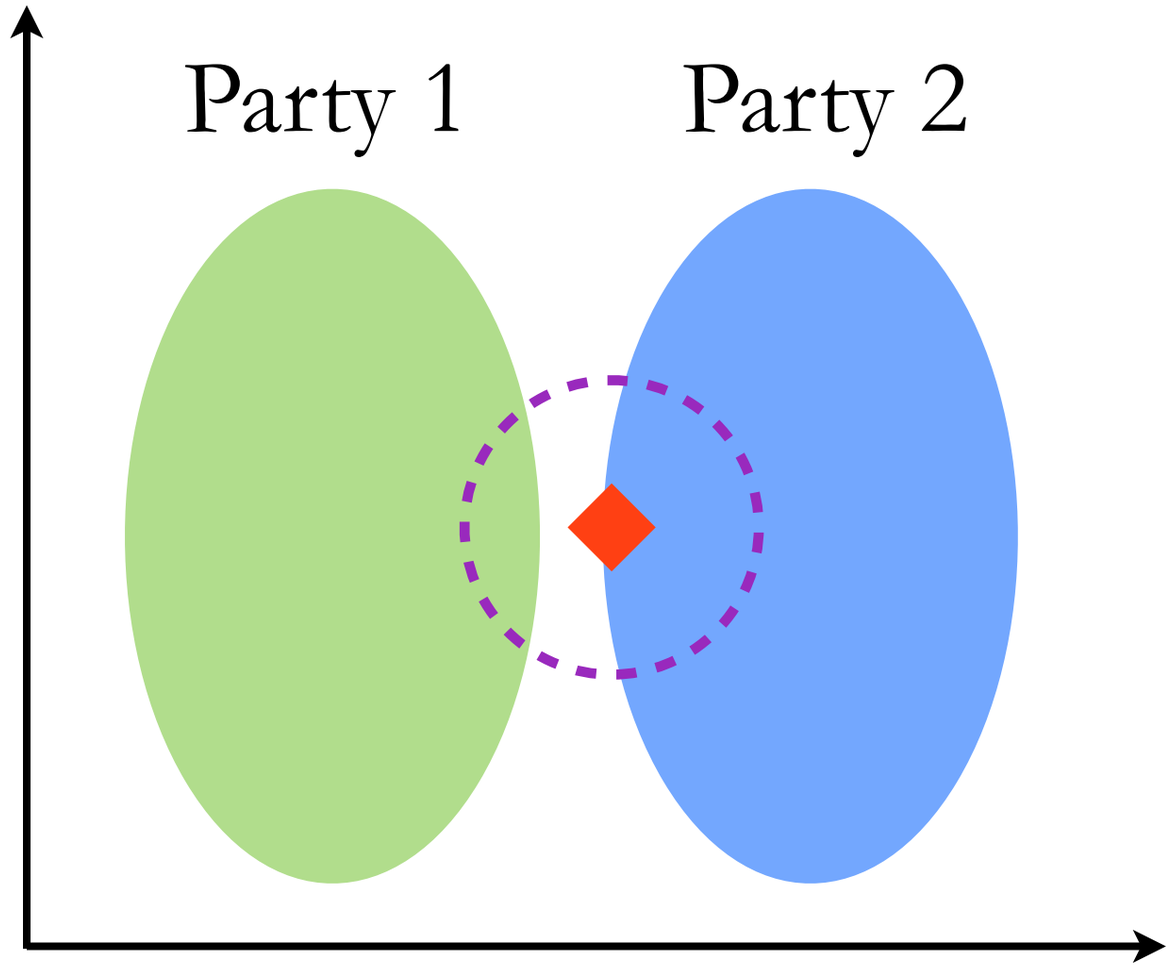
Filling in Missing Data



Filling in Missing Data

1. Fill a small percentage of abstaining votes with colleagues' non-abstaining votes of the missing type.

2. Use k-Nearest Neighbor (k=30) in identifying colleagues to take moderates into account.



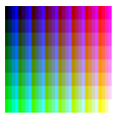
Legislator with missing data

Colleague Sample

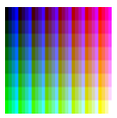
Who are the colleagues of this legislator given these dimensions? The fringes of his/her own party or close neighbors in the other party?

Prediction

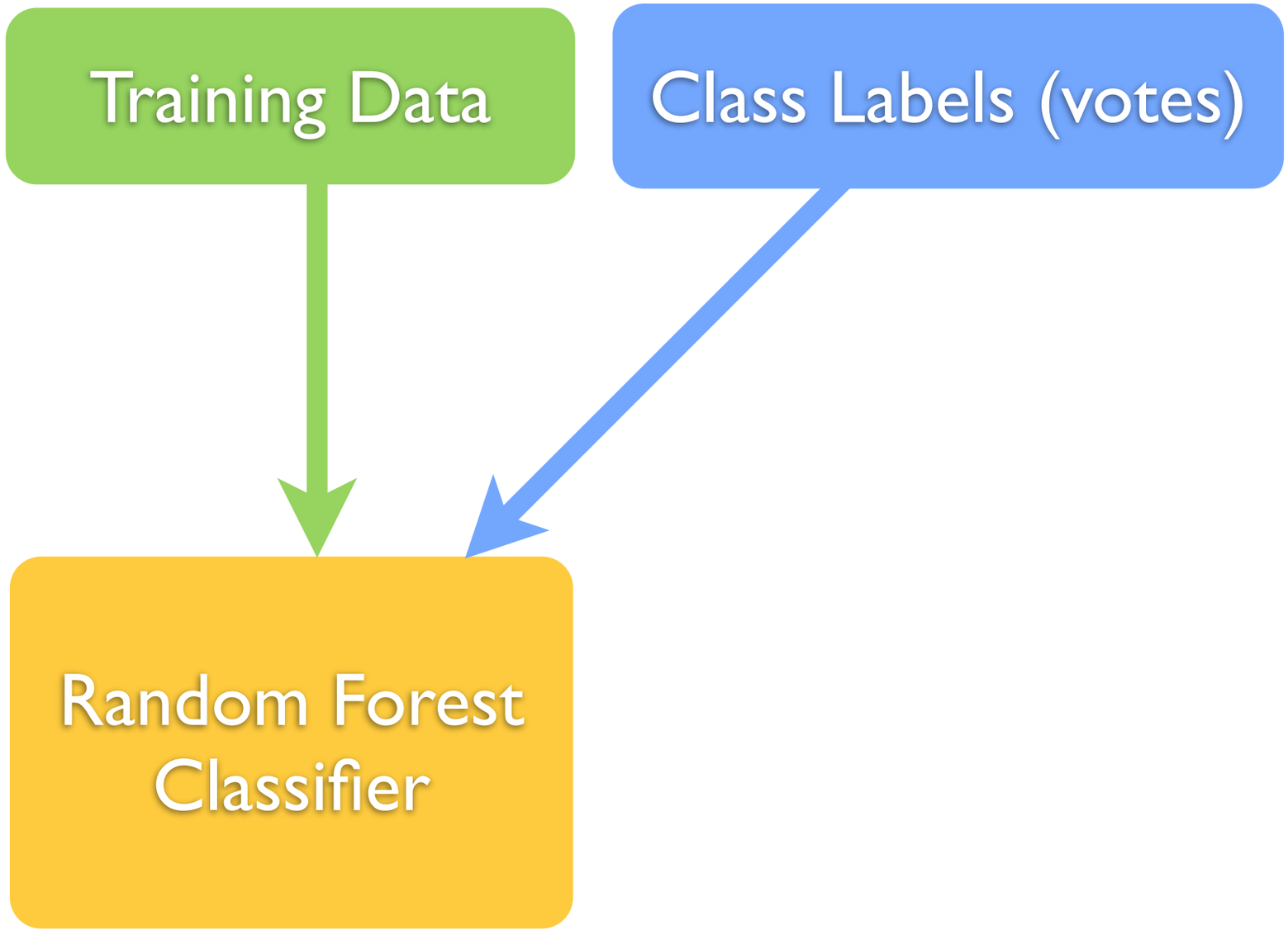
1. Random Forest Classifier (t=30) predicts voting behavior.
    1. Collection of decision trees that vote, each trained on decision feature sequences selected from randomized subsets of features.
    2. Generates class probabilities (Probability of "Yea" or "Nay")
    3. **Average Probability of Agreement (APoA)** mean of legislator vote probabilities for ideal vote-types.



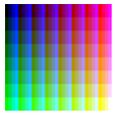
Average Probability of Agreement (APoA)



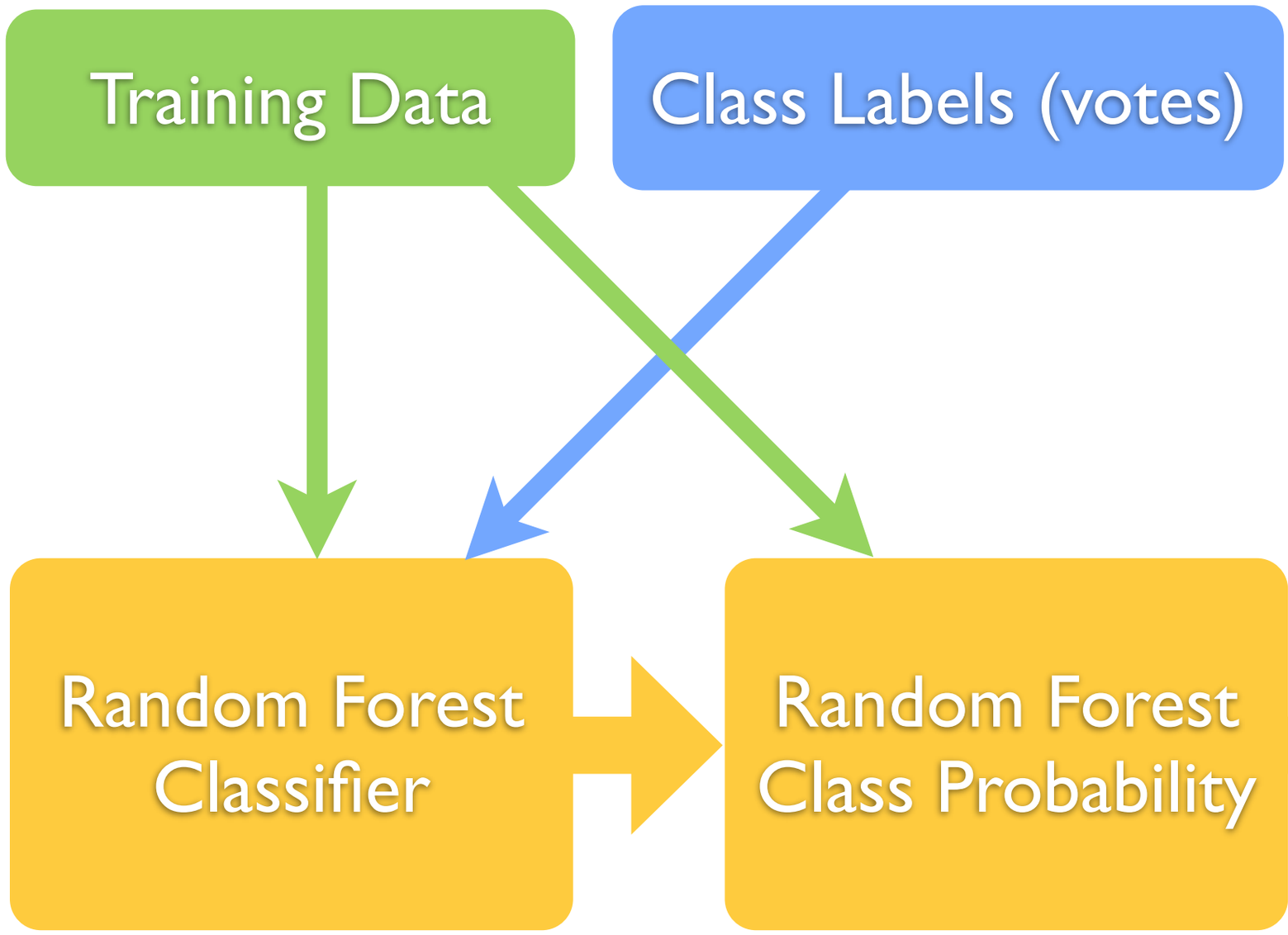
Average Probability of Agreement (APoA)



1. Train

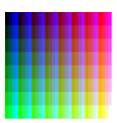


Average Probability of Agreement (APoA)

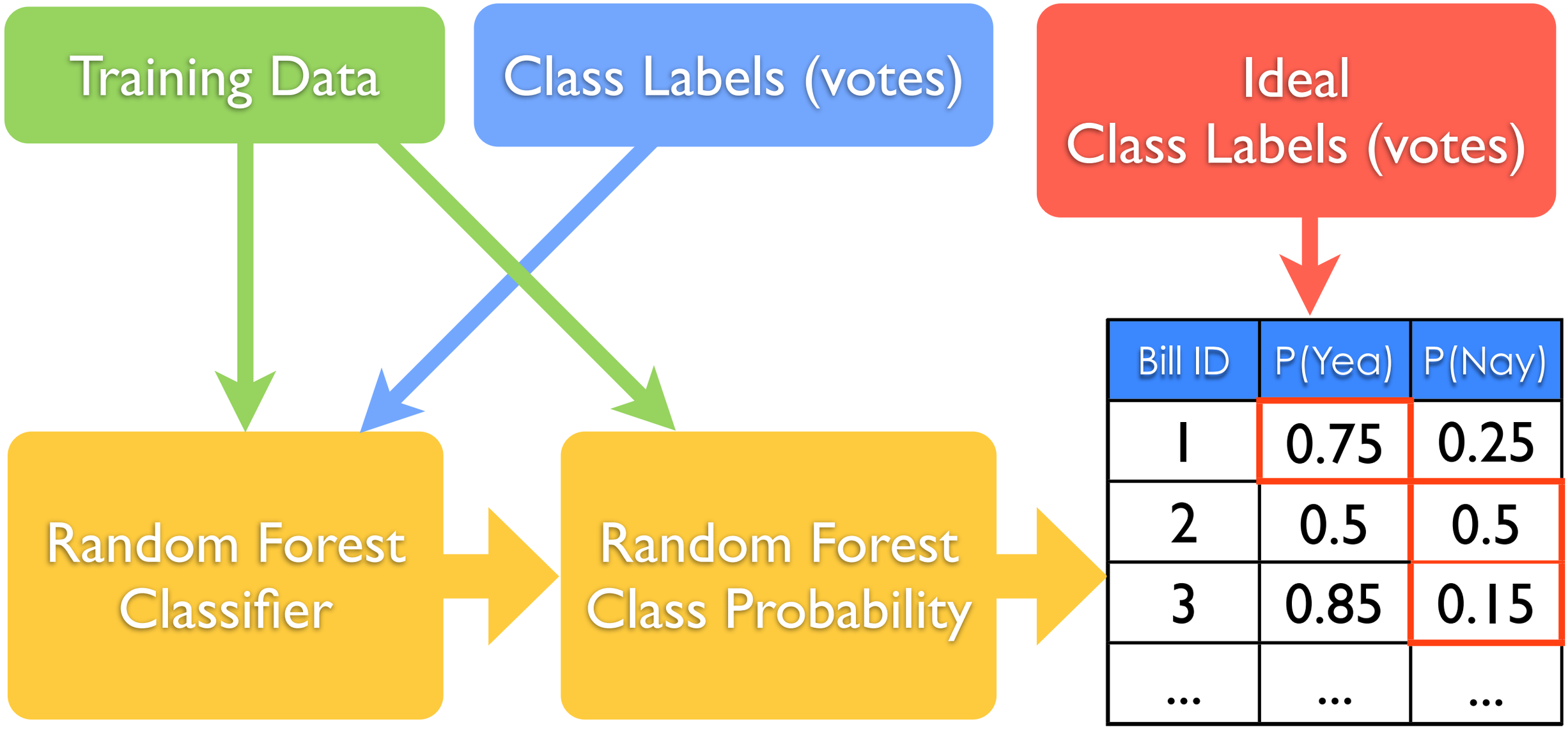


Bill ID	P(Yea)	P(Nay)
1	0.75	0.25
2	0.5	0.5
3	0.85	0.15
...	...	...

1. Train

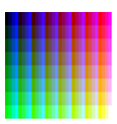


Average Probability of Agreement (APoA)



1. Train

2. Average Idea Vote Probabilities

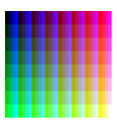


Relative Frequency of Agreement (RFoA)

SIERRA CLUB CALIFORNIA **SENATE REPORT CARD**	VOTE COUNT	SCORE	SB 1x 2	AB 1112	AB 724	AB 1319	AB 376	SB 790
			Renewable Portfolio Standard	Oil Spill Prevention	Clean E Jobs	Bisphenol A	Shark Fins	Electricity: Com choice
Alquist, Elaine (D-13)	9/13	69%	+	+	+	+	+	+
Anderson, Joel (R-36)	3/13	23%	-	-	-	-	+	-
Berryhill, Tom (R-14)	3/13	23%	NV-	-	-	-	+	+
Blakeslee, Sam (R-15)	7/14	50%	+	+	-	-	+	+
Calderon, Ronald (D-30)	7/13	54%	+	+	-	-	+	+
Canella, Anthony (R-12)	1/13	8%	-	-	-	-	-	-
Corbett, Ellen (D-10)	12/14	86%	+	+	+	+	+	+
Correa, Lou (D-34)	4/14	29%	-	-	-	+	NV-	+
de Leon, Kevin (D-22)	9/13	69%	+	+	+	+	NV-	+
DeSaulnier, Mark (D-7)	11/13	85%	+	+	+	+	+	+
Dutton, Bob (R-31)	1/13	8%	-	-	-	-	NV-	-
Emmerson, Bill (R-37)	2/14	14%	-	NV-	-	NV-	+	-
Evans, Noreen (D-2)	12/13	92%	+	+	+	+	+	NV-
Fuller, Jean (R-18)	2/13	15%	NV-	-	-	-	NV-	-
Gaines, Ted (R-1)	3/13	23%	-	-	-	-	+	+

4. Current methods of prediction and quantifying support can be improved upon.

    1. Call this **Relative Frequency of Agreement**



Relative Frequency of Agreement (RFoA)

SIERRA CLUB CALIFORNIA

SENATE REPOR CARD

VOTE

Sierra Club California supported "ay

			SCORE	SB 1x-2	AB 1113	AB 724	AB 1219	AB 376	SB 790
								Shark Fins	Electricity: Com choice
Alquist, Elaine (D-13)								+	+
Anderson, Joel (R-36)								+	-
Berryhill, Tom (R-14)								+	+
Blakeslee, Sam (R-15)								+	+
Calderon, Ronald (D-30)								+	+
Canella, Anthony (R-12)								-	-
Corbett, Ellen (D-10)								+	+
Correa, Lou (D-34)								NV-	+
de Leon, Kevin (D-22)								NV-	+
DeSaulnier, Mark (D-7)								+	+
Dutton, Bob (R-31)								NV-	-
Emmerson, Bill (R-37)								+	-
Evans, Noreen (D-2)	12/13	92%	+	+	+	+	+	+	NV-
Fuller, Jean (R-18)	2/13	15%	NV-	-	-	-	NV-	NV-	-
Gaines, Ted (R-1)	3/13	23%	-	-	-	-	+	+	+

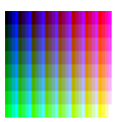
Report Card Legend & Notes

- **+** means pro-environment vote
- **–** means anti-environment vote
- **NV–** means legislator was present, but chose not to cast a vote in support of a pro-environment bill
- **NV+** means legislator was present, but chose not to cast a vote on an anti-environment bill
- **E** means excused absence (does not count toward total score)

Scores are based on the number of "+" and "NV+" votes cast versus the total number of possible votes (excused votes do not count against a score, but NV- does).

4. Current methods of prediction and quantifying support can be improved upon.

    1. Call this **Relative Frequency of Agreement**



Relative Frequency of Agreement (RFoA)

SIERRA CLUB
CALIFORNIA

SENATE REPOR

VOTE

Representative	District	Party	%	HB 88	HB 218	HB 442	HB 1261	HB 1438	SB 461	SB 464	SB 526	SB 763	SB 920
ns, Alma	58	D	70	●	●	●	○	○	●	●	●	●	○
ander, Kelly	107	D	70		●	●	○	○	●	●	●	●	○
ander, Martha	106	D	60	●	●	●	○	○	A	●	●	●	○
n, Lucy	49	D	70	●		●	○	○	●	●	●	●	○
d, Cary[1]	64	R		●	●	●	○	A	NV	NV	NV	NV	NV
n, Marilyn	40	R	40	○	●		○	○	○	●	○	●	○
hart, Jeff	82	R	20	○	●	A	○	○	○	○	A	●	A
Larry	21	D	70	●	●	●	○	○	●	●	●	●	○
kwell, Hugh	86	R	30	○	●	○	○	○	○	●	○	●	○
kwood, Curtis	68	R	40	○	●	●	○	○	○	●	○	●	○
Dan[2]	33	D		●	●	●	○	○	NV	NV	NV	NV	NV
t, John	62	R	10	○	○	○	○	○	○	○	○	●	○
s, James	52	R	20	○	●	○	○	○	○	○	○	●	○
sen, Alice	63	D	70	●	●	●	○	○	●	●	●	●	○
son, William	10	D	50	○	●	●	○	○	●	●	A	●	○
vn, Larry	22	D	30	○	●	●	○	○	○	○	○	●	○
aker, Harold	78	R	30	○	●	●	○	○	NV	○	○	●	○
nt, Angela	7	D	70	●	●	●	○	○	●	●	●	●	○
Justin	67	R	20	○	○	○	○	○	○	●	○	●	○
is-Floyd, Pearl	110	R	40	○	●	●	○	○	○	●	○	●	○
ey, Becky	102	D	40	A	A	A	A	A	●	●	●	●	○
eland, George	14	R	10	○	○	○	○	○	○	○	○	●	○
es, Lorene	77	D	70	●	●	●	○	○	●	●	●	●	○
Nelson	65	D	70	●	●	●	○	○	●	●	●	●	○
am, Tricia	100	D	70	●	●	●	○	○	●	●	●	●	A

& Notes

e

te

nt, but chose not to
ro-environment bill

nt, but chose not to
onment bill

oes not count

+" and "NV+" votes
ble votes (excused
ut NV- does).

Sierra Club California supported "ay

SCOPE	SB 1x 2	AB 1113	AB 724	AB 1219	AB 376	SB 790
					Shark Fins	Electricity: Com choice
					+	+
					+	-
					+	+
					+	+
					+	+
					-	-
					+	+
					NV-	+
					NV-	+
					+	+
					NV-	-
					+	-
+	+	+	+	+	+	NV-
-	-	-	-	-	NV-	-
-	-	-	-	-	+	+

fying support can be improved upon.

ement

Method: Under the Hood

Vote Percent 92%

Title	Key Votes	Vote
Full Repeal of Obamacare	H.R. 2 fully repealed the 2010 government takeover of the private healthcare system, otherwise known as Obamacare.	✓
Spending Cut, Full $100 Billion in Promised Cuts	The Blackburn amendment cut the overall level of nonsecurity discretionary spending in the FY 2011 "continuing resolution" appropriations legislation to achieve the full $100 billion in nonsecurity sp ...	✓
Spending Cut, 2006 Spending Levels	The Mulvaney amendment cut the overall level of nonsecurity discretionary spending in the FY 2011 "continuing resolution" appropriations legislation back to the level in FY 2006.	✓
Defund Davis-Bacon Requirements	The King amendment barred any funds being made available to administer Davis-Bacon prevailing wage requirements.	✓
Obamacare Expansion	H.R. 525 expanded an Obamacare grant program by making veterinarians eligible for public health workforce grants and loan repayment.	✓
Short-Term "Do Nothing" Continuing Resolution	H.J. Res. 48 funded the federal government's operations at far too high a level and failed to include either necessary spending cuts or important policy riders to defund Obamacare, Planned Parenthood ...	✗
NPR Funding Ban	H.R. 1076 prohibited the use of funds to support National Public Radio (NPR) and its often-biased programming.	✓
Obama Foreclosure Assistance	H.R. 839 terminated authority for the Obama Administration to provide market-distorting government assistance under the Home Affordable Modification Program (HAMP).	✓
DC Opportunity Scholarships	H.R. 471 authorized the Secretary of Education to award five-year grants on a competitive basis to nonprofit organizations to carry out a program to provide expanded school choice opportunities to stu ...	✓
National Mediation Board and Union Expansion	The LaTourette amendment to the FAA reauthorization bill struck a provision that overturned a National Mediation Board (NMB) ruling allowing a majority of voting workers (as opposed to a majority of t ...	✓
Prevent the EPA Energy Tax	H.R. 910 amended the Clean Air Act to block the EPA's global warming regulations on green house gases designed to impose a national tax on energy consumers.	✓
Disapproval of Internet Regulation, "Net Neutrality"	H.J. Res. 37 disapproved the FCC's "net neutrality" rules designed to regulate the internet and broadband industry practices. Net neutrality refers to the principle that networks providing internet a ...	✓

ns, Alma
ander, Kelly
ander, Martha
n, Lucy
d, Cary[1]
n, Marilyn
hart, Jeff
Larry
kwell, Hugh
kwood, Curtis
, Dan[2]
t, John
s, James
sen, Alice
on, William
n, Larry
aker, Harold
nt, Angela
, Justin
is-Floyd, Pear
ey, Becky
eland, George
es, Lorene
, Nelson
am, Tricia

Vote Percent 92%

Title	Key Votes	Vote
Full Repeal of Obamacare	H.R. 2 fully repealed the 2010 government takeover of the private healthcare system, otherwise known as Obamacare.	✔
Spending Cut, Full $100 Billion in Promised Cuts	The Blackburn amendment cut the overall level of nonsecurity discretionary spending in the FY 2011 "continuing resolution" appropriations legislation to achieve the full $100 billion in nonsecurity sp ...	✔
Spending Cut, 2006 Spending Levels	The Mulvaney amendment cut the overall level of nonsecurity discretionary spending in the FY 2011 "continuing resolution" appropriations legislation back to the level in FY 2006.	✔
Defund Davis-Bacon Requirements	The King amendment barred any funds being made available to administer Davis-Bacon prevailing wage requirements.	✔
Obamacare Expansion	H.R. 525 expanded an Obamacare grant program by making veterinarians eligible for public health workforce grants and loan repayment.	✔

	Puppy Mills Cosponsor	Chimps in Labs Cosponsor	Horse Slaughter Cosponsor	Ag Subsidies Vote	Funding Letter	Leaders	Score
Alabama							
Sessions, Jeff (R)				✓			20
Shelby, Richard (R)				✓			20
Alaska							
Begich, Mark (D)	✓	✓	✓	✓		✓	100
Murkowski, Lisa (R)				✓			20
Arizona							
Kyl, Jon (R)				✓			20
McCain, John (R)				✓			20
Arkansas							
Boozman, John (R)				X		✓	20

Key to Senate Chart

SP Prime Sponsor

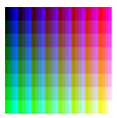
✓ Took pro-animal position through cosponsorship of a bill, a vote, signing a letter, or leading on pro-animal legislation

Experiment

1. Cause

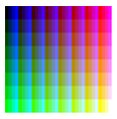
1. University of California Advocacy (www.ucforcalifornia.org)

2. Tested on real data from the California State Legislative Sessions from 2010-Present (Sunlight Foundation, DC)



Experimental Design

1. Independent Variable - Methods used (Framework, classifiers...)
2. Dependent Variables
    1. Evaluating Prediction - Receiver Operating Characteristic Area Under the Curve (Az) in determining “Yea” or “Nay”
    2. Evaluating Ranking - Spearman’s Footrule



Experimental Procedure

1. Followed workflow of the framework described in Method section.

2. Adapted reviews from UC Office of the President State Governmental Relations correspondences with legislators.

UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO • SANTA BARBARA • SANTA CRUZ

OFFICE OF THE PRESIDENT

DANIEL M. DOOLEY
Senior Vice President – External Relations

Office of State Governmental Relations
1130 K Street, Suite 340
Sacramento, California 95814
(916) 445-9924
Steve Juarez, Associate Vice President and Director

September 2, 2011

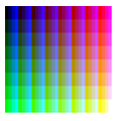
The Honorable Edmund G. Brown, Jr.
Governor, State of California
State Capitol
Sacramento, CA 95814

RE: SB 611 (Steinberg) - REQUEST FOR SIGNATURE

Dear Governor Brown:

On behalf of the University of California (UC), we respectfully request that you sign SB 611 (Steinberg). This measure expands and codifies in state law the University of California Curriculum Integration Institute (UCCII), an entity working at the cutting edge of high school course design. The UCCII brings together high school teachers, university faculty and other experts to develop new model courses that provide students with rigorous academic content that is linked to real world applications.

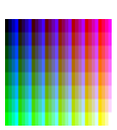
Established in 2008 and administered by UC, the UCCII connects K-12 educators with faculty and instructors in a number of disciplines from UC, the California State University, California Community Colleges, private higher education institutions, and statewide career technical education associations to develop high school curriculum based on career-oriented, integrated academic and technical education content. These newly designed, fully integrated, "a-g" approved courses will serve to enhance pupil prospects for postsecondary education and/or employment. By establishing the UCCII in state statute, SB 611 recognizes the importance and benefits to California from providing specific State authorization for this program. The University is pleased to continue and expand its work in this area to develop model career-oriented, integrated academic and technical education courses that fulfill University "a-g" admissions subject requirements, in accordance with the availability of additional funding from federal, state, or private sources.



Classification Accuracy with Abstains

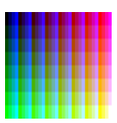
CLASSIFIER	Yeas	Nays	Abstains	Mean
SVM: K=RBF	27%	55%	78%	53%
SVM: K=Poly	28%	53%	73%	51%
LinearSVM	39%	57%	60%	52%
Decision Tree	38%	53%	57%	49%
Random Forest	34%	58%	67%	53%
Prior	29%	16%	55%	

Hit rates of several classifiers is charted here. The first, second, and third best performances per class label are highlighted in green, yellow, and orange respectively.



Data Notes

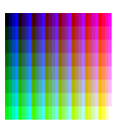
1. Yeas are hardest to classify.
    1. Can't blame it on data shortage since nays are fewer and yet are easier to classify.
    2. Probably many abstain profiles that are similar to Yea profiles.
2. Solution: throw out abstains!
    1. Results in the probability that a legislator will vote *given that they vote!*
    2. This means we cannot predict vote outcomes.



Classification Accuracy with no Abstains

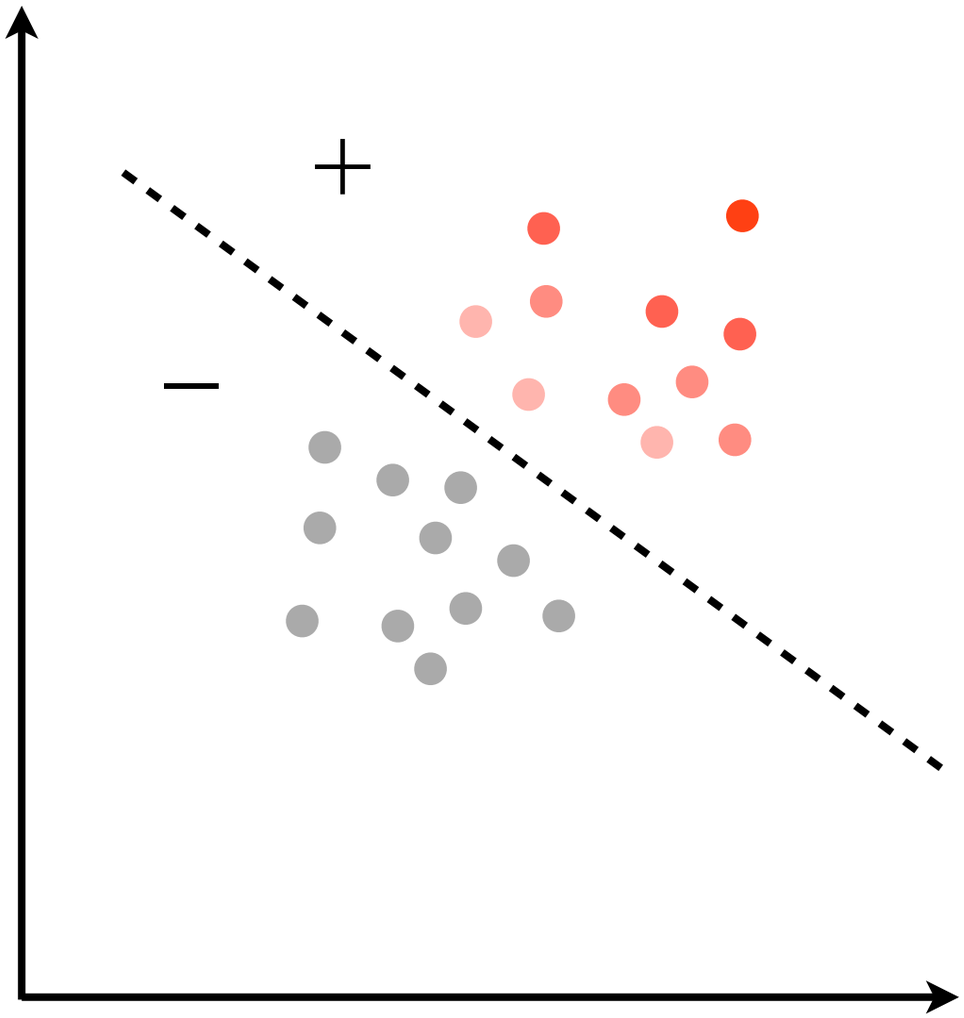
CLASSIFIER	Yeas	Nays	Mean
SVM: K=RBF	90.01%	73.69%	84.87%
SVM: K=Poly	87.52%	74.27%	83.35%
LinearSVM	85.73%	66.96%	79.81%
Decision Tree	80.44%	66.39%	76.03%
Random Forest	88.52%	73.30%	83.73%
Gradient Boosting	80.44%	67.47%	76.35%
Prior	68.65%	31.35%	

Hit rates of several classifiers is charted here. The first, second, and third best performances per class label are highlighted in green, yellow, and orange respectively.



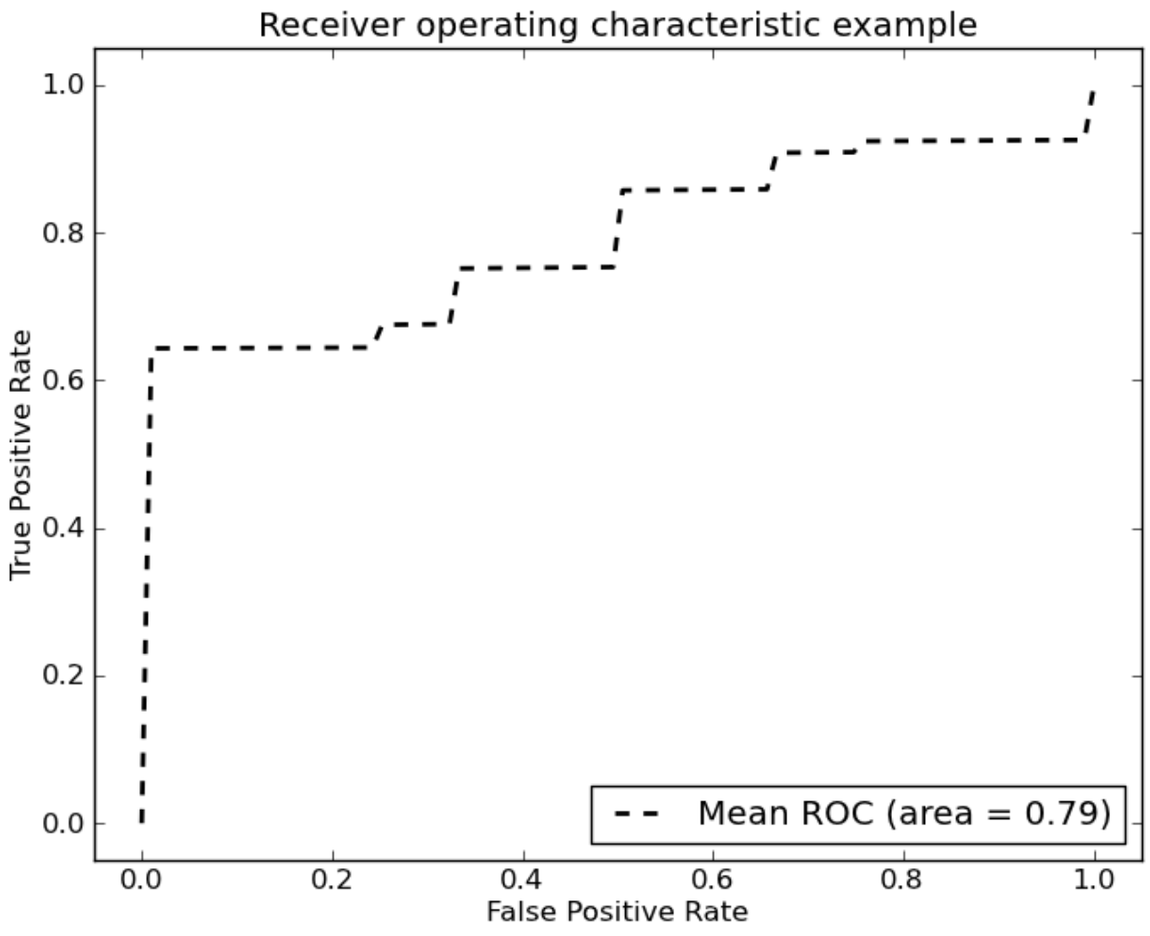
Prediction: Classifier Notes

1. Why not SVM?
    1. SVM does not generate probabilities naturally.
        1. Calculated with five-fold cross-validation in SVMLIGHT.
        2. The max probability and discriminating component decisions can be different!
        3. Discriminating component cannot be used as likelihood estimate.
2. Random Forests generate probabilities naturally, consistently.

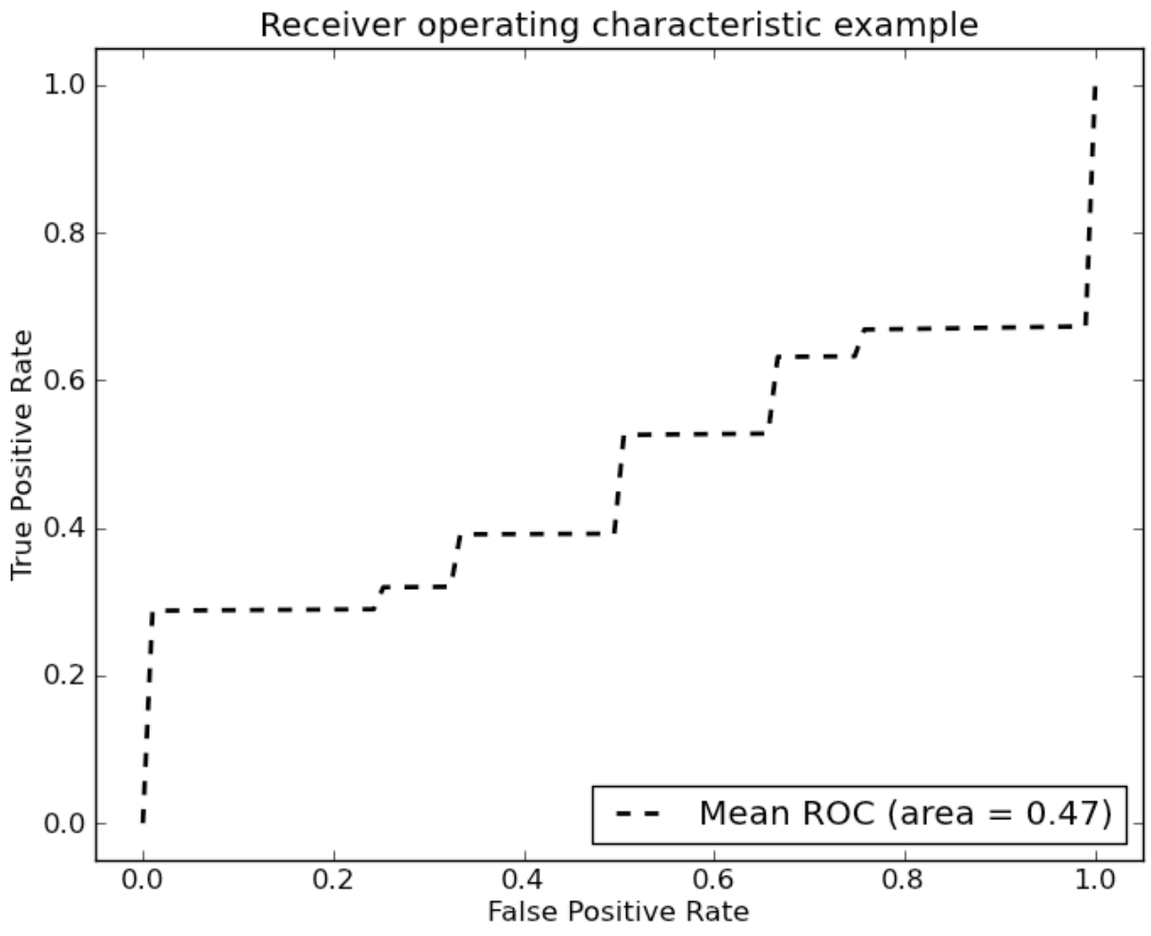


Results

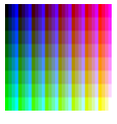
1. With Random Forest Classifier, leave-five-out cross-validation with 1,000 Monte Carlo iterations



Random Forest

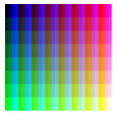


SVM: k=RBF



Ranking & Scoring

1. Why rank or assign scores?
    1. Identify legislators for outreach... *it's predictive!*
    2. Central question: What is a legislator's *latent* support?



Ranking

1. Compare ranking by **Average Probability of Agreement** to **Relative Frequency of Agreement**

Ranked by **APoA**

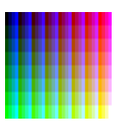
Ranked by **RFoA**

?

Ranked by Future RFoA

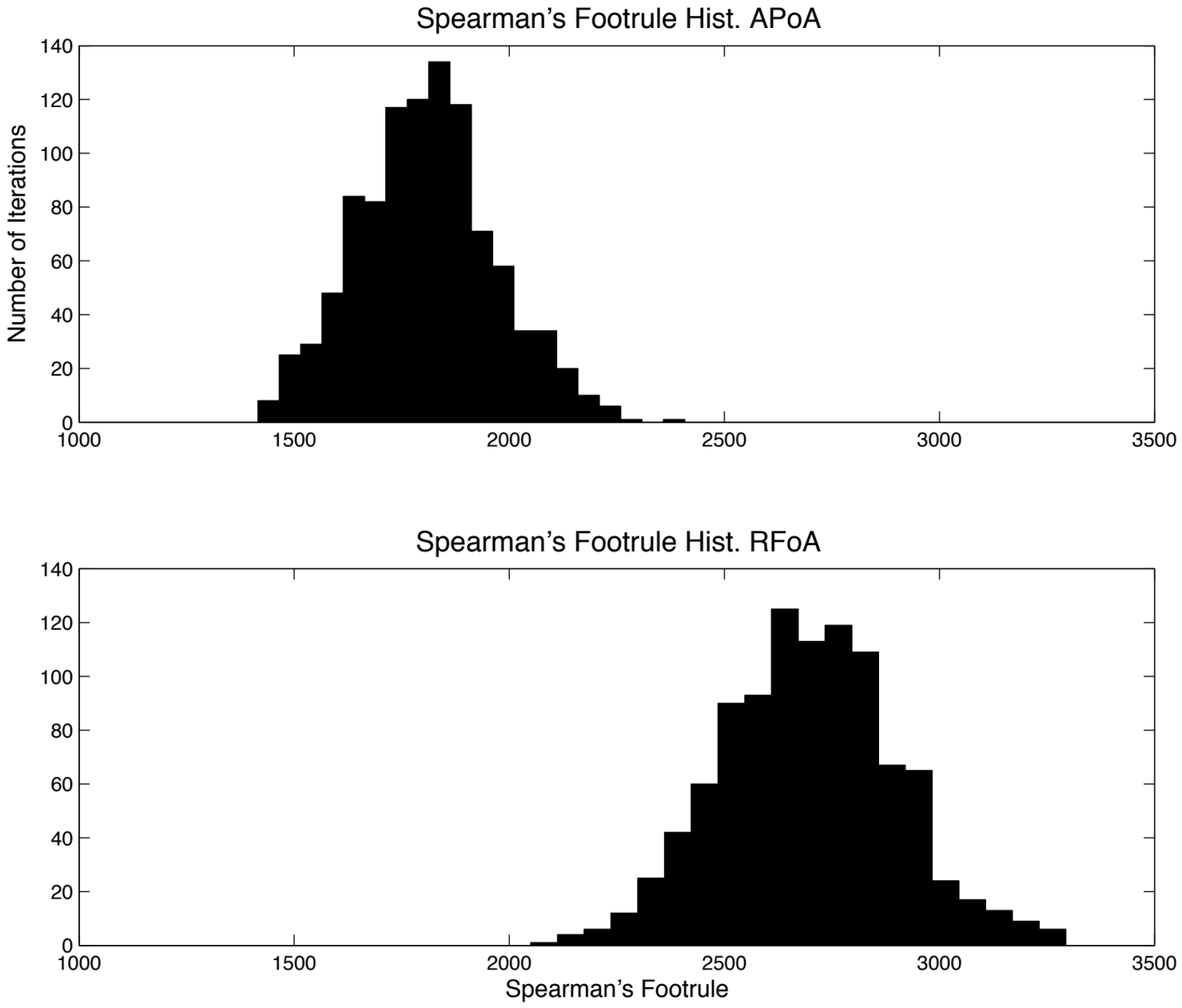
How many places is a given legislator off by?
Use Spearman's Footrule:

$$Fr^{|S|}(\sigma_1, \sigma_2) = \sum_{i=1}^{|S|} |(\sigma_1(i) - \sigma_2(i))|.$$



Ranking

1. Tested over 1,000 iterations of leave-five-out cross-validation



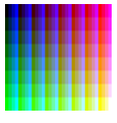
$$Fr^{|S|}(\sigma_1, \sigma_2) = \sum_{i=1}^{|S|} |(\sigma_1(i) - \sigma_2(i))|.$$

$|S| = 118$

How many places is a given legislator off by?

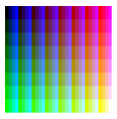
On Average:

APoA	RFoA
15.33	22.85



Discussion

1. Confirmed the effectiveness of the framework
    1. In predicting / quantifying level of support for a cause
    2. Improved upon current method of ranking legislators
2. Future work
    1. Scalability / Design
    2. Adding a layer of abstraction so anyone can make a custom page
    3. Improving classifier performance
    4. Account system



Questions?

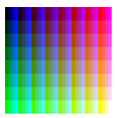
Thank you

Xifeng Yan, Ambuj Singh

Greta Carl-Halle, Katie Ellis

Kirsten Deshler, Monica Solorzano

Everardo Diaz



Matt Weiden, 2012