

How belief in ghosts and other extraneous opinions relate to political affiliation

Cassie Noble
David Kelly

Introduction

We are attempting to train a SVM to correctly predict a person's political party affiliation based on seemingly extraneous and outrageous questions from a survey gathered by Cards Against Humanity.

Data

- 1000 people were surveyed.
- 482 identified as either Democrat or Republican.
 - This was our dataset after cleaning.
- 12 extraneous questions were utilized.
- Questions were posed as numerical or categorical.
 - 2-6 possible responses for categorical questions.

Survey Questions

- What would you say is the likelihood that your current job will be entirely performed by **robots** or computers within the next decade?
- Do you believe that **climate change** is real and caused by people, real but not caused by people, or not real at all?
- How many **Transformers movies** have you seen?
- Do you agree or disagree with the following statement: **scientists** are generally honest and are serving the public good.
- Do you agree or disagree with the following statement: **vaccines** are safe and protect children from disease.
- How many **books**, if any, have you read in the past year?
- Do you **believe in ghosts**?
- What **percentage** of the federal budget would you estimate is spent on scientific **research**?
- Is federal **funding** of scientific **research** too high, too low, or about right?
- True or false: the **earth is always farther away from the sun in the winter** than in the summer.
- If you had to choose: would you rather be **smart and sad, or dumb and happy**?
- Do you think it is acceptable or unacceptable to **urinate in the shower**?

Methods

- Utilized a 2-class SVM to classify subjects as either Democrat or Republican.
- Test various kernel functions.
- 1-in-k-coding
- 10-fold cross-validation

Experiments

- Train a binary classifier SVM to predict political affiliation.
- Find optimal kernel function for SVM.
- Identify questions that were strongly correlated with a subjects political affiliation.
- Look for trends in subjects' responses.

Results

- Overall best avg. train/test accuracy with RBF.
- Train Accuracy
 - avg: 72.4%
 - min: 70.9%
 - max: 73.7%
- Test Accuracy
 - avg: 69.3%
 - min: 62.5%
 - max: 83.3%

Results:

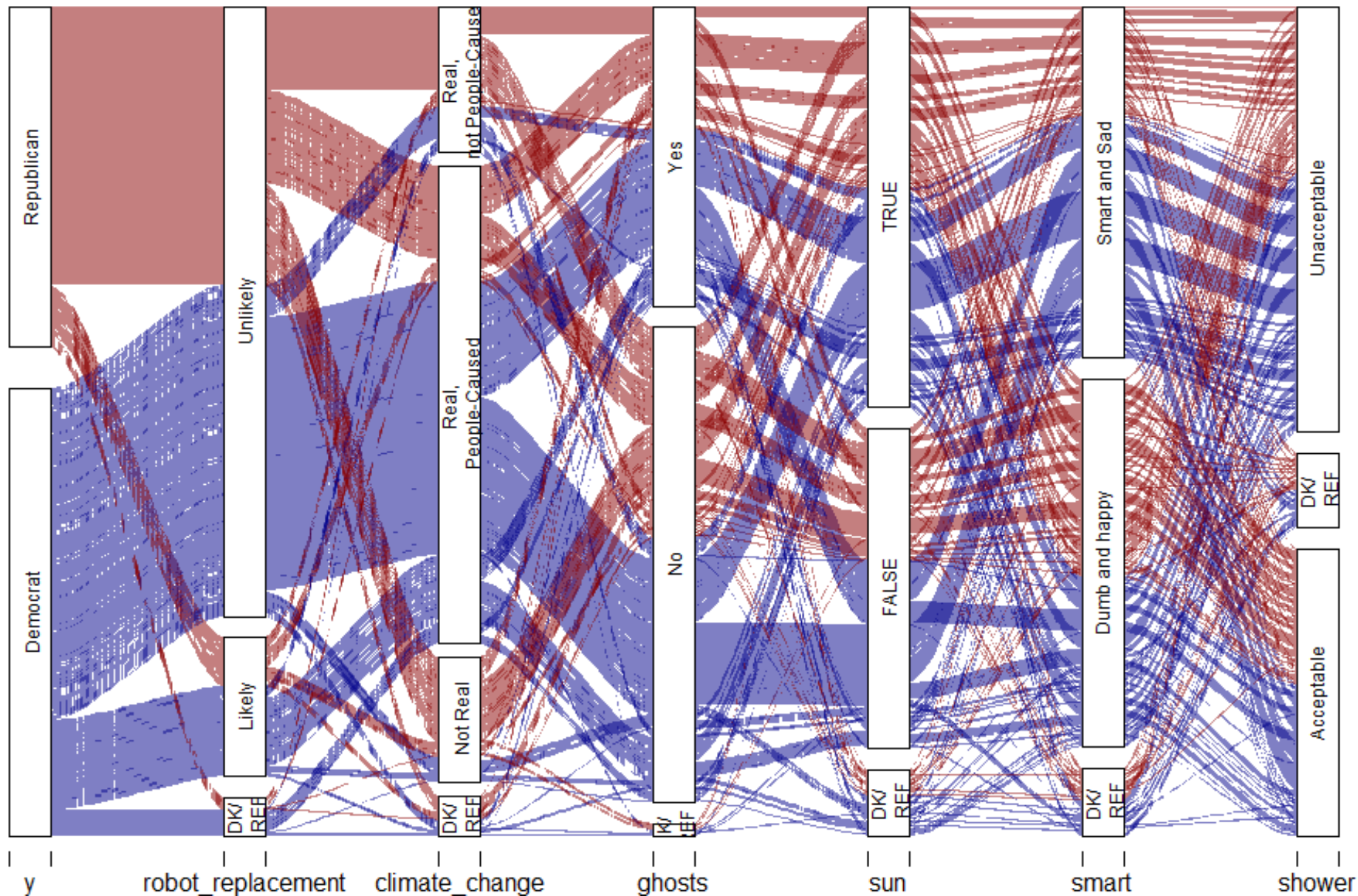
Correctly v.s Incorrectly Predicted

Predicted	True Political Affiliation	Robot Replacement	Climate Change	Transformers Movies	Scientists Good	Vaccines Safe	Books Read	Believe in Ghosts	Research %	Adequate Funds	Earth Always Farther from Sun in Winter	Smart and Sad or Dumb and Happy	Urine in Shower
Democrat	Democrat	Unlikely	Real, people-caused	0	Strongly Agree	Strongly Agree	6	No	10	Too Low	True	Smart and Sad	Not Acceptable
Republican	Democrat	Likely	Real, people-caused	1	Somewhat Agree	Strongly Agree	8	Yes	22.5	Too Low	True	Smart and Sad	Not Acceptable
Democrat	Republican	Unlikely	Not Real	0	Strongly Agree	Strongly Agree	4	No	15	About Right	False	Dumb and Happy	Not Acceptable
Republican	Republican	Unlikely	Real, people-caused	0	Somewhat Agree	Strongly Agree	5	No	12	About Right	True	Dumb and Happy	Not Acceptable

Results:

Most vs. Least Extreme Predictions

Predicted	True Political Affiliation	Robot Replacement	Climate Change	Transformers Movies	Scientists Good	Vaccines Safe	Books Read	Believe in Ghosts	Research %	Adequate Funds	Earth Always Farther from Sun in Winter	Smart and Sad or Dumb and Happy	Urinate in Shower
Most Certain Republican	Republican	Unlikely	Real, not people-caused	2	Somewhat Disagree	Strongly Agree	6	No	2	About Right	False	Dumb and Happy	Acceptable
Least Certain Republican	Republican	Likely	Real, not people-caused	1	Somewhat Agree	Strongly Disagree	20	No	30	Too high	False	Dumb and Happy	Not Acceptable
Least Certain Democrat	Republican	Likely	Real, not people-caused	1	Strongly Disagree	Somewhat Agree	20	No	30	Too high	False	Dumb and Happy	Not Acceptable
Most Certain Democrat	Democrat	Likely	Real, people-caused	0	Somewhat Agree	Somewhat Agree	20	Yes	50	Too low	True	Smart and Sad	Acceptable



Discussion

- Large number of support vectors
- Weights were all very similar
 - support vectors weighted only slightly more

Future Work

- Multi-class classification problem including independents and moderates
- Determine which survey questions were the best predictors

References

References

Nicolas Ballarini and Yi-Da Chiu. *SubgrPlots: Graphical Displays for Subgroup Analysis in Clinical Trials*, 2020. URL <https://CRAN.R-project.org/package=SubgrPlots>. R package version 0.1.3.

Cards-Against-Humanity. Pulse of the nation, Dec 2017. URL https://www.kaggle.com/cardsagainsthumanity/pulse-of-the-nation#201709-CAH_PulseOfTheNation.csv.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3): 273–297, 1995.

MATLAB. *version 9.8.0 (R2020a)*. The MathWorks Inc., Natick, Massachusetts, 2010.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL <https://www.R-project.org/>.