Exploring U.S. Congressional Voting Behavior Using Bayesian Models

Ana Mateos Mata

January 2025

Introduction

- Machine learning revolutionizes political science.
- Study focuses on the voting behavior of the U.S. Congress.
- Uses 1984 U.S. Congressional Voting Records dataset.
- Goal: Analyze voting patterns and predict behavior using Bayesian models.

Dataset Overview

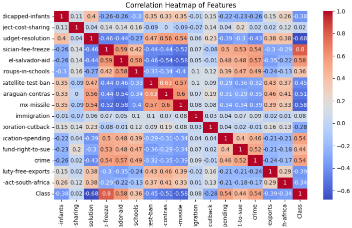
- Source: 1984 Congressional Voting Records dataset.
- Includes 435 members of Congress, 16 critical issues.
- Features voting outcomes: Yea, Nay, and Unknown.
- Party affiliation: Democrat (267) and Republican (168).

Software Used

- Python library: `pgmpy`.
- Tailored for probabilistic graphical models.
- Facilitates Bayesian Network construction and inference.
- Chosen for ease of integration and robust features.

Exploratory Data Analysis

- Class imbalance: Democrats overrepresented.
- Strong correlations: Key features like physician-fee-freeze and education-spending.
- Weak correlations: Features like water-project-cost-sharing.



Data Preprocessing

- Missing data: Less than 7% for most features.
- Strategy:
  - Abstentions treated as a separate category.
  - Maintains political context of abstentions.

Structure Planning

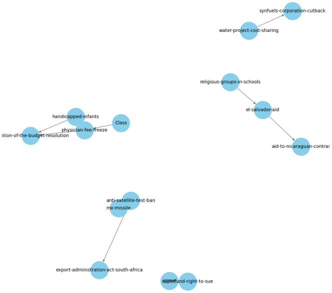
- Compared PC algorithm and Hill Climb Search.
- Evaluation metrics: BIC, K2, BDeu, Fisher's C, etc.
- Trade-off: Simplicity (PC) vs. complexity (K2).

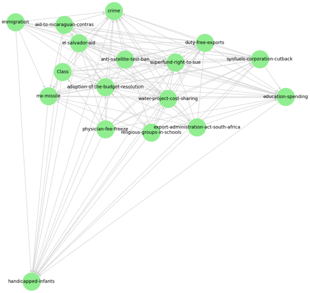
	Score Method Model	Number of Edges	Log-Likelihood Scores	Correlation Scores	Fisher's C	k2	bdeu	bds	bic
0	BIC	17	-4348.815021	0.000000	0.000000e+00	-4.574580e+03	-4582.117516	-4.688169e+03	-4.649545e+03
1	K2	117	-2651.431347	0.000000	5.831104e-04	2.982926e+07	-7322.512028	-1.478398e+08	-2.615133e+08
2	BDeu	30	-4089.303682	0.000000	2.341967e-08	-4.498432e+03	-4479.501010	-4.614306e+03	-4.833534e+03
3	BDs	29	-4099.487793	0.000000	1.087064e-11	-4.492308e+03	-4492.758625	-4.610651e+03	-4.807266e+03
4	PC	9		0.092308	0.000000e+00	-5.310152e+03	-5325.158266	-5.425665e+03	-5.362571e+03

Figure: Evaluation of Model Structures

Model Evaluation

- K2: Highest log-likelihood (-2651.43) but more complex (117 edges).
- PC: Simpler model with 9 edges, better interpretability.
- Chose K2 for inference due to its better fit.





Parameter Training

- Conditional Probability Distributions (CPDs) estimated.
- Used MaximumLikelihoodEstimator.
- Parameters derived from the learned network structure.

CPD for water-project-cost-sharing:

Class	...	Class(republican)
adoption-of-the-budget-resolution	...	adoption-of-the-budget-resolution(y)
aid-to-nicaraguan-contras	...	aid-to-nicaraguan-contras(y)
anti-satellite-test-ban	...	anti-satellite-test-ban(y)
el-salvador-aid	...	el-salvador-aid(y)
export-administration-act-south-africa	...	export-administration-act-south-africa(y)
immigration	...	immigration(y)
physician-fee-freeze	...	physician-fee-freeze(y)
religious-groups-in-schools	...	religious-groups-in-schools(y)
water-project-cost-sharing(n)	...	0.6666666666666666
water-project-cost-sharing(p)	...	0.0
water-project-cost-sharing(y)	...	0.3333333333333333

Inference Results

- Prior probabilities: Democrat (61.38%), Republican (38.62%).
- Advanced queries reveal:
  - Impact of economic sanctions on missile defense.
  - Education spending linked to religious groups in schools.

Advanced Query Example

- Query: Probability of `mx-missile` given votes on economic sanctions.
- Result: Sanctions increase missile defense probability to 81.6%.

prob$_m x_m issile_g iven_s anctions = infer.query(variables = [' mx - missile'],$
evidence='export-administration-act-south-africa': 'y', 'duty-free-exports': 'y')

Key Insights

- Bayesian Networks capture interdependencies between policies.
- Military policies influence social policy outcomes.
- K2 provides better fit but at the cost of complexity.
- PC is more interpretable, suited for causal insights.

Conclusion

- Bayesian models offer robust tools for analyzing voting behavior.
- Findings reveal how party affiliation and policies interact.
- Future work: Extend analysis to newer datasets or include additional policy features. Absent analysis, better comparison between models, K2 is the best?