Latent Community Detection for Modeling Legislative Roll Call Votes

Eli Ben-Michael, Runjing Liu, Jake Soloff

Department of Statistics, UC Berkeley

Objectives

We analyze voting data in the House of Representatives during the 110th Congress (2007-2008). We extend an ideal point model to incorporate caucus membership data via a stochastic block model. In doing so, we aim to

- Use caucus membership data to infer latent communities among representatives.
- Exploit this community structure to inform an estimate for each representative's ideal point.
- Model a representative's voting behavior.

Motivation

We model caucus membership because we found that caucus memberships influence legislative behavior. For example, the more caucuses two legislators share, the more likely they are to vote the same way on a bill.

We used neighborhood selection on roll call vote data to represent interactions among members of the House as an undirected graphical model; we found that subgraphs corresponding to caucuses were denser than the graph of the whole House.

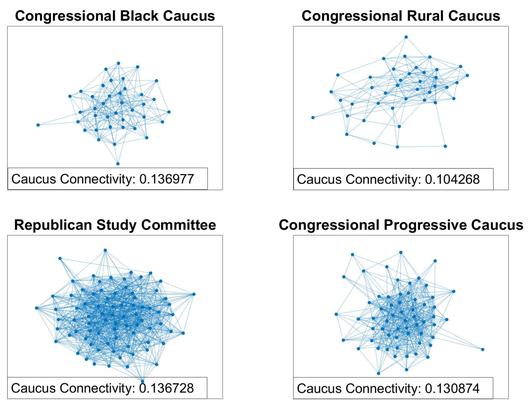


Figure: The connectivity is measured by the fraction of total edges present; connectivity of the whole House is 0.064.

Model

Our model assumes each representative u has an ideal point $x_u \in \mathbb{R}^S$, where S is a free parameter. It also assumes each representative belongs to one of K latent communities. The generative process is:

- Sample community proportions $\pi \sim \text{Dir}(\gamma 1_K)$ and each community ideal point $\nu_k \sim \mathcal{N}_S(\varpi, \sigma_\nu^2)$
- Draw representative u's community $M_u \overset{\text{iid}}{\sim} \text{Cat}(\pi)$ and ideal point $x_u \mid M_u = k, \nu \sim \mathcal{N}_S(\nu_k, \sigma_x^2)$
- Draw coexpression rates $P_{kl} \overset{\text{iid}}{\sim} \text{Gamma}(\lambda_0, \lambda_1)$
- Observe the number of common caucuses $R_{uv} \mid P, M_u = k, M_v = l \sim \text{Poisson}(P_{kl})$
- Draw a discrimination $a_d \sim \mathcal{N}_S(\eta_a, \sigma_{ad}^2)$ and a difficulty $b_d \sim \mathcal{N}_S(\eta_b, \sigma_{bd}^2)$ for each bill d
- Observe the votes $V_{ud} \mid x_u, a_d, b_d \sim \text{Bern}(\sigma(a_d \cdot (x_u - b_d)))$

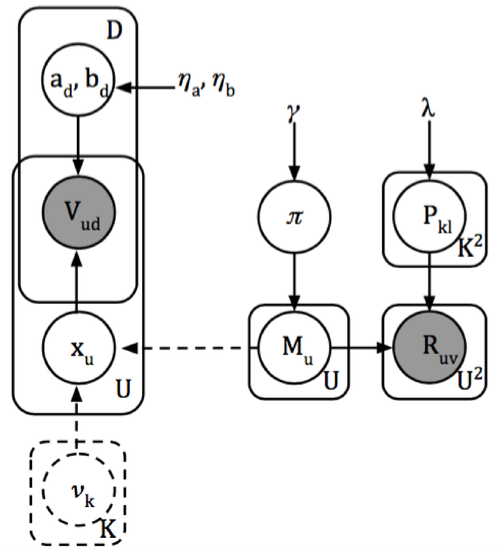


Figure: Latent Community Ideal Point Model (LC-IPM)

Variational Inference

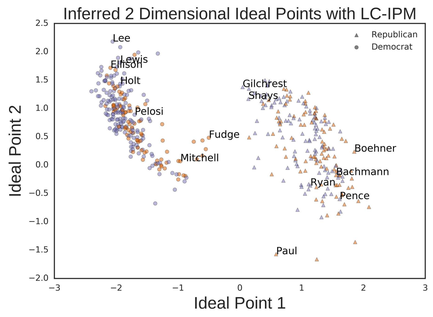
Upon observing votes $V = (V_{ud})$ and caucus comembership counts $R = (R_{uv})$, computing the posterior distribution of the latent variables given the observations is intractable. We employ *mean field variational inference*, finding the distribution q closest in KL divergence to the posterior among all fully factorized distributions. Since the graphical model for LC-IPM joins SBM and IPM via only one edge, the mean field factorization yields the same variational updates for π, P, a_d, and b_d as in SBM and IPM, and we exploited this modularity.

Conclusion

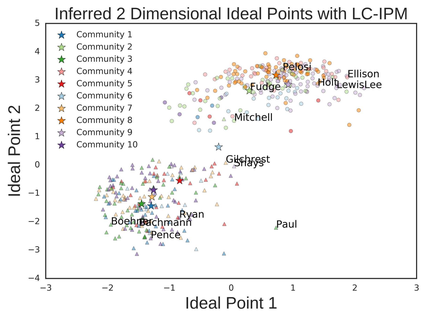
LC-IPM offers similar predictive performance to the standard ideal point model, but we gain interpretability and can make predictions for junior representatives. Our model also applies to more general collaborative filtering settings with relational data.

Model	Acc	AUC
Yea	67.123	50.000
Logistic Reg	78.340	85.105
IPM ($S = 1$)	94.769	98.418
IPM ($S = 2$)	95.451	98.998
LC-IPM ($S = 2, K = 2$)	95.405	99.013
LC-IPM ($S = 2, K = 10$)	95.415	99.015

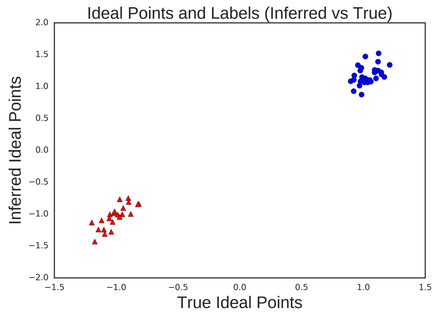
Results



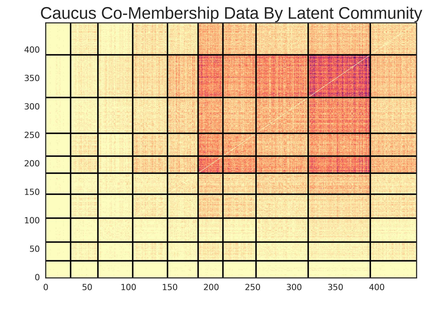
(a) With $K = 2$ latent communities, the inferred ideal points are well-separated by party affiliation. Shape denotes party affiliation; color denotes community.



(b) With $K = 10$ latent communities, the inferred ideal points cluster by community, but are still well-separated by party affiliation. Shape denotes party affiliation; color denotes community.



(c) On toy data generated according to the graphical model, we accurately infer both the true ideal points and the community memberships. Shape denotes true community; color denotes inferred community.



(d) Each entry denotes the number of caucuses shared by a pair of representatives. Red denotes many (up to thirty) common caucuses; yellow denotes close to no common caucuses.

Future Directions

Following the example of [2], we hope to combine LC-IPM with *supervised topic modeling* to incorporate text data such as bill text and speeches. We are also interested in the composability of hierarchical models for multiple data sources.

References

[1] Wainwright, M. J. & Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning.*

[2] Gerrish, S. M. & Blei, D. M. (2011). Predicting legislative roll calls from text. *Proceedings of the 28th International Conference on Machine Learning.*

[3] Blei, D. M., Kucukelbir, A. & McAuliffe, J. D. (2016). Variational inference: a review for statisticians. *arXiv:1601.00670.*

[4] Hasite, T. J., Tibshirani, R. & Wainwright, M. J. (2015). Statistical learning with sparsity: the Lasso and generalizations. *CRC Press.*

Berkeley
UNIVERSITY OF CALIFORNIA