A Text-based HMM of Foreign Affair Sentiment

Sean M. Gerrish
Department of Computer Science
Princeton University
Princeton, NJ 08544
sgerrish@cs.princeton.edu

David M. Blei
Department of Computer Science
Princeton University
Priceton, NJ 08544
blei@cs.princeton.edu

Abstract

We present a time-series model for foreign relations, in which the pairwise sentiment between nations is inferred from news articles. We describe a model of dyadic interaction and illustrate our process of estimating sentiment using Amazon Mechanical Turk labels. Across articles from twenty years of the of the *New York Times*, we predict with modest error on heldout country pairs.

The written history of foreign relations is peppered with the bias of politics and hindsight. Because historians currently have limited resources to review thousands of relevant news sources, research may be biased by popular knowledge or even politics and culture. The goal of this work is to create a history of the relationships between the world's nations using the text of newspaper articles and political commentary. An assumption of our work is that the tension between two nations – or a warm and robust relationship between them – is reflected by the language we use to discuss them. Using this assumption, we outline a model that is able to infer the relationship between pairs of countries whose relationship has not been observed in training.

An advantage of a text-based approach to history is that we can incorporate information from all articles of a given collection with modest computational cost. This means that historians and political scientists can then search and review thousands of historical documents to identify forgotten or overlooked "blips" in history.

1 Model

Reflecting the intuitions above, we assume that each nation lies in a space of latent '"foreign sentiment". A spatial a model has two benefits. First, it provides interpretability: nations with similar positions in this latent space tend to interact more positively, while nations further apart tend to have more tension in their relationship. A spatial model also allows us to draw on ideas from multidimensional scaling, which has been used successfully in both political science [1, 2] and social network modeling [3, 4].

In the model we outline below, we assume that each country c takes a position $\bar{x}_c \in \mathbb{R}^P$ in the space of latent political sentiment. International relations are determined by the interaction of countries' positions $\bar{x}_c$.

A temporal model of interaction. Foreign relations are not static, however; nations' alliances and preferences change over time with the evolution of economies, tech-

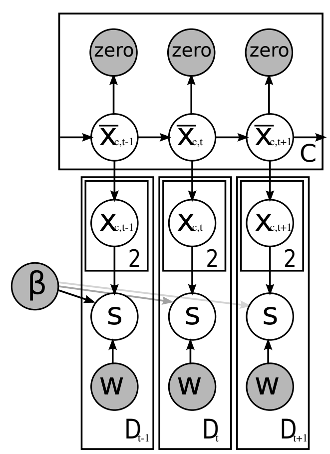


Figure 1: A time-series model of countries' interactions. Pseudo-observations of "zero" are added for regularization. Amazon Mechanical Turk labels are used to fit β, which is used to infer unobserved sentiments.

nology, and culture. We make an assumption that each piece of news d about two nations c_1 and c_2 reflects some underlying sentiment s_d between them. We make this a fully temporal model by allowing each country's mean position $\bar{x}_{c,t}$ to drift over time with the Markov transition

$$\bar{x}_{ct}|\bar{x}_{c,t-1} \sim N(\bar{x}_{c,t-1}, \sigma_K^2), \tag{1}$$

as shown in Figure 1. At any time t, state c_1 may interact with state c_2 in the following way:

$$\begin{aligned} x_{c_1,d} &\sim N(\bar{x}_{c_1,t}, \sigma_D^2) \\ x_{c_2,d} &\sim N(\bar{x}_{c_2,t}, \sigma_D^2) \\ s_d &:= x_{c_1,d}^T x_{c_2,d}, \end{aligned} \tag{2}$$

where we interpret s_d as the sentiment between c_1 and c_2 as reflected by article d. When c_1 and c_2 are similar (as measured by their inner product), their sentiment s_d will be positive; if they are dissimilar, their sentiment will be negative. More extreme values indicate stronger sentiment.

When a news source discusses the relationship between these nations, the author's choice of words $\boldsymbol{w}_d$ reflects the relationship between the countries. We model this sentiment with the text of the article d. Using text regression [5], the sentiment is modeled on wordcounts $\boldsymbol{w}_d$ of the article:

$$s_d|\boldsymbol{w}_d, \beta \sim \mathcal{N}(\boldsymbol{w}_d^T \beta, \sigma_W^2).$$

We describe how to fit β with Amazon Mechanical Turk workers in Section 4.2.

1.1 Related work

Spatial models such as Item Response Theory (IRT) have been developed over the past century by quantitative social scientists for analyzing behavior. While much of this work has been used to model parliamentary voting behavior, these techniques have also been used to model voting in the UN General Assembly. Gartzke et al., for example, use these votes and alliance models to study the nations' affinities [6].

These models have been developed for dyadic data more fully in network models such as the latent space model [3, 7], in which the probability of a link between two nodes is a function of their latent-space distance. The qualitative relationship of entities' dyadic relationships has been more fully developed with text by the relational topic model, which uses free text to model the relationship between actors in an unsupervised setting [4].

2 Inference

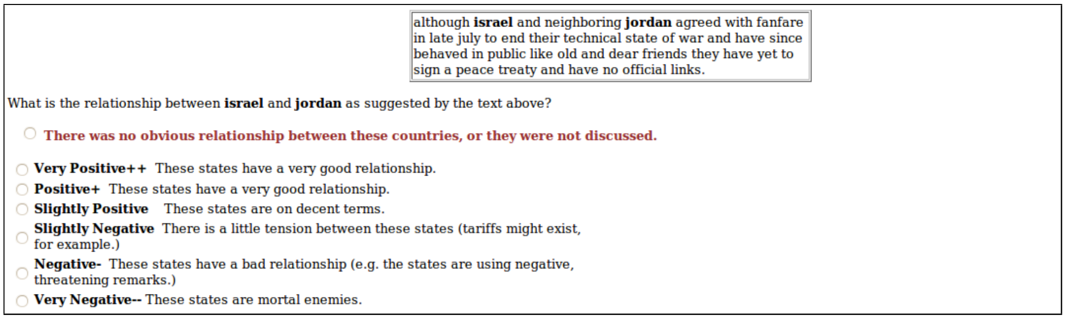
We fit the *MAP* objective of this probabilistic model. This has the benefit of both clean exposition and simple implementation, and it can be interpreted as a form of unregularized variational inference. We optimize the *MAP* objective in this model using traditional EM.

M Step. In the M step, the mean of each country's position is updated using a modified Kalman filter. This step differs from a standard Kalman filter in that we may have no or multiple observations on any given date. We also add *pseudo-observations* for each country at each day t with mean 0 and variance 10. These observations are a form of "time-series regularization" and reflect the sense that a lack of news is effectively neutral news. The prior over the ends of the chain are standard normal.

E Step. In the E-step, our goal is to infer nations' positions $x_{c,d}|\bar{x}_{c,t_d}, s_d$ with each interaction d given their means and the sentiment s_d for this interaction. We find these positions by gradient ascent on $x_{c,d}$.

3 Estimation of sentiment s

To infer the sentiment s_d between two countries, we treat the corresponding news article as a bag-of-words and use text regression [5]. In each of these cases, the sentiment s_d is first estimated on



although **israel** and neighboring **jordan** agreed with fanfare in late july to end their technical state of war and have since behaved in public like old and dear friends they have yet to sign a peace treaty and have no official links.

What is the relationship between **israel** and **jordan** as suggested by the text above?

- **There was no obvious relationship between these countries, or they were not discussed.**
- **Very Positive++** These states have a very good relationship.
- **Positive+** These states have a very good relationship.
- **Slightly Positive** These states are on decent terms.
- **Slightly Negative** There is a little tension between these states (tariffs might exist, for example.)
- **Negative-** These states have a bad relationship (e.g. the states are using negative, threatening remarks.)
- **Very Negative--** These states are mortal enemies.

Figure 2: A screenshot of a Mechanical Turk labeling task. Sometimes relationships may be complicated; both raters gave this example a score of "slightly positive".

a training set using Amazon Mechanical Turk (AMT). AMT is a crowdsourcing platform which provides access to thousands of "workers" who perform simple tasks over the Internet. These labels are then attached to all articles by predicting their sentiment.

4 Experiments

We fit and evaluate this model over news articles discussing 245 nations from twenty years of the *New York Times* (NYT). This collection spanned the years 1987 to 2007, a period which included both Gulf wars; the collapse of the Soviet Union; the reunification of Germany; September 11th, 2001; and countless other world events.

4.1 Datasets and tokenization

We made an important assumption that the scope of foreign sentiment discussion is at the level of a paragraph. We therefore used the subset of paragraphs which discuss exactly two nations as "documents" d, a set of 257,472 distinct paragraphs which came from the foreign, business, financial, and magazine desks of the newspaper during this period. We used standard stopword filters and removed the geographic tags for labeling nations from the sentiment vocabulary.

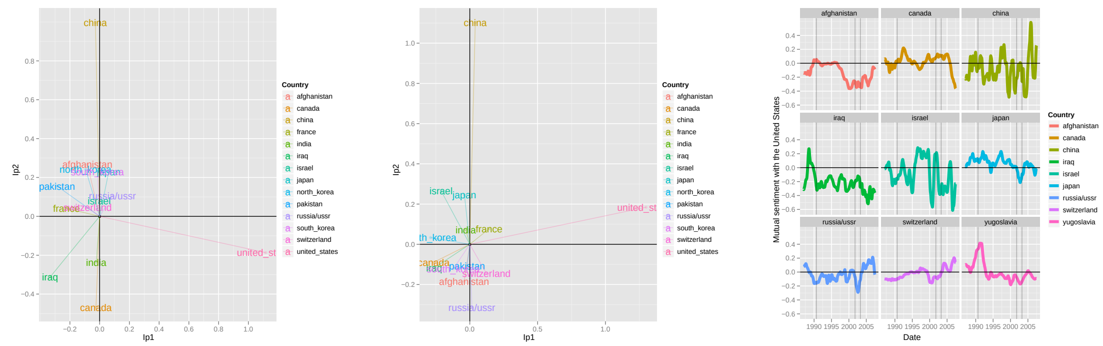
4.2 Mechanical Turk Labels

To fit the model, we asked *Amazon Mechanical Turk* raters to rate the sentiment between two nations mentioned in a paragraph in the range -5 = "mortal enemies", $\ldots, 5$ ="very good relationship". With these training examples, we fit the coefficients β of the text regression discussed in Section 1. This coefficient was then fixed in the joint model in Figure 1 to allow us to learn sentiment from paragraphs' words. We trained and fit the model with the following procedure:

1. Randomly select 3607 paragraphs discussing pairs of 245 countries.
2. Label each of these paragraphs' sentiment with two ratings from Amazon Mechanical Turk.
3. Hold out 42 random country pairs (244 paragraphs) for testing.
4. Estimate sentiment model parameters β using the training paragraphs.
5. Infer the spatial sentiment model with these parameters on all 257,472 paragraphs.
6. Evaluate model sentiment prediction on the heldout 244 test paragraphs.

4.3 Results

Heldout sentiment For two nations c_1 and c_2 mentioned together at time t, we predict their sentiment to be $\tilde{s}_{c_1,c_2} = \bar{x}_{c_1,t}\bar{x}_{c_2,t}$. Based on this estimate, the average squared error for heldout nations is 2.32 ($R^2 = 0.51$), considerably better than a baseline of text regression, which had means-squared-error 5.53; for comparison, average *inter-rater squared error* – the minimum theoretically possible –



(a) (b) (c)

Figure 3: (a) Example positions of selected countries in the latent space of national sentiment, in 1987. Sentiment is given by the inner product between two vectors. (b) Example positions of selected countries in 2007. (c) Mutual sentiment $s = \bar{x}_{c,\cdot}^T \bar{x}_{\mathrm{us},\cdot}$ with the United States over time. The two Iraq wars and September 11th, 2001 are marked.

Table 1: Average error predicting sentiment s_d between heldout nation pairs.

Model	Mean Squared Error	Mean Absolute Error
Inter-rater agreement	1.77 (7.11)	1.037 (2.07)
Text regression	5.53	1.94
Prior variance 0.1	2.36	1.09
Prior variance 1	**2.32**	**1.07**
Prior variance 10	2.32	1.08
Prior variance 100	2.34	1.09
Prior variance 1000	2.33	1.08

was 1.77, and the square of the difference *between* raters was 7.11. We also found that using "pseudo observations" with even modest observation variance improves predictive performance; these results are summarized in Table 1.

Nations' positions are illustrated in Figure 3. Figures 3(a,b) show nations' relative positions and interactions for a two- dimensional model; heldout error slightly increased for higher dimensions. The United States' relationship with Iraq (see Figure 3(c)) serves as an excellent example of this model in action; the relationship between these nations degrades during both the Gulf War and the invasion of Iraq following September 9th, 2001. Israel's relationship with the United States demonstrates one of the model's downfalls: while Israel is considered a close ally of the United States, the raters' 95 ratings of these nations' mentions had mean 0.12 and standard deviation 2.65. Because the model is only as good as its ratings, the nations appear to have a rocky relationship.

Future work

There are many ways in which we hope to expand this work. Creating a joint model of text, sentiment, and votes in bodies such as the UN General assembly is one possible direction. Another improvement on this model would assign a sentiment intercept term for each nation to Equation 2. Finally, we hope to explore more unsupervised models, in which the we characterize the relationships between nations, using an approach related to relational topic models [8].

Acknowledgments

David M. Blei is supported by ONR 175-6343, NSF CAREER 0745520, AFOSR 09NL202, the Alfred P. Sloan foundation, and a grant from Google. Yahoo! provided many of the machines on which these experiments were performed.

References

[1] Martin, A. D., K. M. Quinn. "Dynamic ideal point estimation via markov chain monte carlo for the u.s. supreme court, 1953-1999." *Political Analysis*, 10:134–153, 2002.

[2] Jackman, S. "Multidimensional analysis of roll call data via bayesian simulation: Identification, estimation, inference, and model checking." *Political Analysis*, 9(3):227–241, 2001.

[3] Hoff, P., A. E. Raftery, M. S. Handcock. "Latent space approaches to social network analysis." *Journal of the American Statistical Association*, 97:1090–1098, 2002.

[4] Chang, J., D. M. Blei. "Relational topic models for document networks." *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AIStats) 2009*, 5, 2009.

[5] Kogan, S., D. Levin, B. Routledge, J. Sagi, N. Smith. "Predicting risk from financial reports with regression." In "ACL Human Language Technologies," 272–280. Association for Computational Linguistics, 2009.

[6] Gartzke, E. "Kant we all just get along? opportunity, willingness, and the origins of the democratic peace." *American Journal of Political Science*, 42(1):1–27, 1998.

[7] Sarkar, P., A. W. Moore. "Dynamic social network analysis using latent space models." *ACM SIGKDD Explorations Newsletter*, 7(2):31–40, 2005.

[8] Chang, J., J. Boyd-Graber, S. Gerrish, C. Wang, D. Blei. "Reading tea leaves: How humans interpret topic models." In "Neural Information Processing Systems (NIPS)," 2009.