Supplement to
Analyzing Framing through the Casts of Characters in the News

Dallas Card[1] **Justin H. Gross**[2] **Amber E. Boydstun**[3] **Noah A. Smith**[4]

[1]School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[2]Department of Political Science, University of Massachusetts, Amherst, MA 01003, USA
[3]Department of Political Science, University of California, Davis, CA 95616, USA
[4]Computer Science & Engineering, University of Washington, WA 98195, USA

dcard@cmu.edu   aboydstun@ucdavis.edu   jhgross@polsci.umass.edu
nasmith@cs.washington.edu

A   Framing Dimensions

The 15 framing dimensions defined in the Media Frames Corpus are reproduced in Figure 1.

Economic: costs, benefits, or other financial implications

Capacity and resources: availability of physical, human or financial resources, and capacity of current systems

Morality: religious or ethical implications

Fairness and equality: balance or distribution of rights, responsibilities, and resources

Legality, constitutionality and jurisprudence: rights, freedoms, and authority of individuals, corporations, and government

Policy prescription and evaluation: discussion of specific policies aimed at addressing problems

Crime and punishment: effectiveness and implications of laws and their enforcement

Security and defense: threats to welfare of the individual, community, or nation

Health and safety: health care, sanitation, public safety

Quality of life: threats and opportunities for the individual's wealth, happiness, and well-being

Cultural identity: traditions, customs, or values of a social group in relation to a policy issue

Public opinion: attitudes and opinions of the general public, including polling and demographics

Political: considerations related to politics and politicians, including lobbying, elections, and attempts to sway voters

External regulation and reputation: international reputation or foreign policy of the U.S.

Other: any coherent group of frames not covered by the above categories

---

Figure 1: Framing dimensions used in the MFC.

B   Model Details

B.1   Dirichlet Process Mixture Model

There are a number of equivalent formulations of Dirichlet process mixture models. Here we present the formulation based on the stick-breaking process. According to this perspective, each mixture component is drawn from an (infinite) set of mixture components (equivalently, clusters), each of which is drawn from a base measure, H. The conditional probability of a cluster assignment is distributed according to an (infinite) multinomial distribution, generated according to the stick-breaking process, with hyperparameter λ. In particular,

$$\{\pi_k'\}_{k=1}^{\infty} \sim \text{Beta}(1, \lambda) \tag{1}$$

$$\{\pi_k\}_{k=1}^{\infty} \sim \pi_k' \prod_{l=1}^{k-1}(1 - \pi_l') \tag{2}$$

$$s_i \sim \boldsymbol{\pi} \tag{3}$$

$$\theta_{s_i} \sim H \tag{4}$$

In our model, we take H to be a symmetric Dirichlet distribution with hyperparameter α. Given a cluster assignment for the ith document, each entity's persona is then drawn according to $p_e \sim \theta_{s_i}$, where s_i indexes the cluster assignment of the ith document.

B.2   Collapsed Gibbs Sampling

The probability of a document being assigned to an existing cluster (mixture component) is proportional to the number of documents already assigned to that cluster times the likelihood of the document's current personas being generated from that cluster's dis-

tribution over personas (θ_{s_i}). The probability of the document being assigned to a new cluster is proportional to λ times the likelihood of the document's personas being generated from a new draw from the base distribution. Integrating out θ and π gives:

$$p(s_i = s' \mid \mathbf{s}_{-i}, \mathbf{p}, \alpha, \lambda) \propto n_{s',*}^{(-i)} \times f(s') \quad (5)$$

$$p(s_i = s^{new} \mid \mathbf{s}_{-i}, \mathbf{p}, \alpha, \lambda) \propto \lambda \times f(s^{new}) \quad (6)$$

$$f(s) = \prod_{j=1}^{J} \frac{\alpha + n_{s,p_j}^{(-i)} + \sum_{j'=1}^{j-1} \mathbb{I}_{\{p'_j = p_j\}}}{P\alpha + n_{s,*}^{(-i)} + (j-1)} \quad (7)$$

Here, s' is an existing cluster, J ranges over the entities in document i, and p_j is the persona of the jth entity. $n_{s,p_j}^{(-i)}$ is the number of entities in documents of type s with persona p_j, excluding those in i. $n_{s,*}^{(-i)}$ is the total number of entities in this set, and $\mathbb{I}_{\{\cdot\}}$ is the indicator function.

The equation for sampling personas is similar, and can be shown to be

$$p(p_e = p \mid \mathbf{p}_{-e}, \mathbf{z}, s_e, \alpha, \beta) = (\alpha + n_{s_e,p}^{(-e)})$$

$$\times \prod_{r=1}^{R} \prod_{t=1}^{T_{e,r}} \frac{\beta + n_{p,r,k_t}^{(-e)} + \sum_{t'=1}^{t-1} \mathbb{I}_{\{k_{t'} = k_t\}}}{K\beta + n_{p,r,*}^{(-e)} + (t-1)} \quad (8)$$

where $n_{s_e,p}^{(-e)}$ is the number of entities with persona p in documents with cluster s_e, excluding entity e. R is the number of categories of relations (agent, patient, attribute), $T_{e,r}$ is the number of tuples for entity e with relation r, k_t is the topic of the tth tuple in $T_{e,r}$, $n_{p,r,k_t}^{(-e)}$ is the number of tuples with relation r and topic k_t for entities with persona p, excluding entity e, and $n_{p,r,*}^{(-e)}$ are these counts summed over topics.

The equation for sampling the topic of a tuple attached to entity e is:

$$p(z_t = k \mid \mathbf{z}_{-t}, \mathbf{p}, \mathbf{w}, \mathbf{r}, \beta, \gamma) =$$

$$(\beta + n_{p_e,r_t,k}^{(-t)}) \times \frac{\gamma + n_{k,w_t}^{(-t)}}{V\gamma + n_{k,*}^{(-t)}} \quad (9)$$

where V is the size of the vocabulary, $n_{p_t,r_t,k}^{(-t)}$ is the number of tuples with relation r_t and topic k attached to entities with persona p_e, excluding tuple t. $n_{k,w_t}^{(-t)}$ is the number of tuples with w_t assigned to topic k, excluding t, and $n_{k,*}^{(-t)}$ is the sum of these counts across topics.

C Extraction Patterns

The extraction patterns used to extract $\langle w, r \rangle$ tuples from dependency trees are given in Table 1.

Relation type	Neighbor	Arc type	POS
Attribute	parent	nsubj	JJ
	parent	nsubj	NN*
	child	amod	JJ
Agent	parent	agent	VB*
	parent	nsubj	VB*
	child	acl	VB*
Patient	parent	dobj	VB*
	parent	nsubjpass	VB*
	parent	iobj	VB*

Table 1: Extraction patterns.