Supplementary Material for

"A Topic Model For Movie Choices and Ratings"

Timothy N. Rubin
Mark Steyvers

Below we describe how parameters are learned for the Ratings Topic Model. We then describe how we use observations about choices (m) to improve estimation of user parameters θ when ratings are not observed.

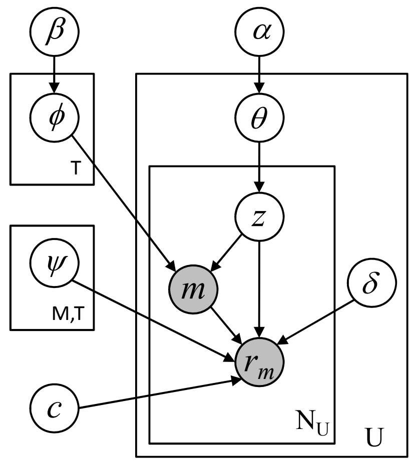

Figure S.1: Ratings Topic Model shown using graphical model notation

Section 1: Parameter Estimation

The full joint distribution for ratings, topics, movies and users in the model can be factored into two primary components:

$$p(r,m,z,u) = p(r \mid m,z,u)\,p(m,z,u) \qquad (\textit{eq. 1})$$

where $p(r \mid m,z,u)$ is the conditional probability of a rating given a movie, topic and user, and $P(m,z,u)$ is the joint density of movies, topics and users. The conditional distribution over ratings is modeled using the ordered-logit model, while the joint distribution of movies, topics and users is modeled using LDA.

The ordered-logit and LDA components of the model were estimated in alternation using MCMC methods in order to sample from the full posterior joint density over ratings, movies, topics and users. LDA parameters θ and ϕ were learned using Gibbs sampling. Ordered-logit parameters $\psi_{m,t}$, δ_u and c were learned using a sequential Metropolis Hastings algorithm.

Ordered-Logit Inference The ordered-logit model specifies the probability of observing a rating as a function of utility (U) and the rating-cutoffs (c), where utility is defined as the sum of the user rating-bias δ and a movie's preference value ψ:

$$U_{u,m} = \psi_{t,m} + \delta_u + \varepsilon$$

The probability of observing the rating user u gave to movie m is defined as:

$$p(r = r_{u,m} \mid \psi_{t,m}, \delta_u, c) = p(c_i < U_{u,m} < c_{i+1}) = \frac{1}{1+e^{\psi_{t,m}+\delta_u-c_{i+1}}} - \frac{1}{1+e^{\psi_{t,m}+\delta_u-c_i}} \qquad (\textit{eq. 2})$$

where $c_i = r_{u,m}$.

Using a Metropolis-Hastings algorithm, we sample from the posterior density of Ordered-Logit parameters $\psi_{m,t}$, δ_u and c conditional on LDA parameters θ and ϕ and the observed ratings. To draw samples from the posterior, it is sufficient to specify the relative probability of observing the data ($r_{i \ldots N_r}$):

$$p\left(r_{i \ldots N_r} \middle| \psi_{t,m}, \delta_u, c\right) \propto p\left(\psi_{t,m}, \delta_u, c \middle| r_{i \ldots N_r}\right) = \prod_{m,t,u,r} \left[\frac{1}{1+e^{\psi_{t,m}+\delta_u - c_{i+1}}} \frac{1}{1+e^{\psi_{t,m}+\delta_u - c_i}}\right] \qquad (\textit{eq. 3})$$

For each complete iteration of the Metropolis-Hastings algorithm, we sequentially evaluate proposal values for individual parameter values (this allows the parameter space to be explored more rapidly given limited time). Proposals were drawn from normal distributions, with $\sigma_\delta = .4$, $\sigma_\psi = .3$, $\sigma_c = .1$. Two iterations of the Metropolis-Hastings algorithm were run between each cycle of the Gibbs sampler. After the Gibbs sampler has converged, LDA parameters were fixed and 400 samples were drawn for each of the ordered-logit parameters. The means of these samples were used to produce posterior predictions.

LDA Inference Estimates of the posterior density $P(m,t,u)$ was performed using a modified form of the collapsed-Gibbs sampler (Griffiths and Steyvers, 2004). Griffiths and Steyvers (2004) demonstrated that estimates of θ and ϕ can be obtained by sequentially sampling assignments of each observation z to a topic. Here, topic assignments are made conditional the observed rating value r_i in addition to parameters θ and ϕ:

$$p(z_i = j | m_i, u_i, r_i) \propto \sum_{j=1}^{T} p(r_i \,| m_i, u_i, z_i = j\,)\, p(m_i | z_i = j)\, p(z_i = j | u_i) \qquad (\textit{eq. 4})$$

where z_i refers to the topic-assignment for the i^{th} rating. Estimates of the probabilities of movies given topics, ϕ, and the probabilities of topics given users, θ, are computed as:

$$p(m|t) = \frac{n_{m_{i,j}}^{MT} + \beta}{\sum_{k=1}^{M} n_{m_{i,j}}^{MT} + M\beta}$$

$$p(t|u) = \frac{n_{u_{i,j}}^{UT} + \alpha}{\sum_{t=1}^{T} n_{u_{i,j}}^{UT} + T\alpha}$$

Using these estimates and the ordered-logit parameters, the probability that the current rating should be assigned to topic j, conditioned on all other topic assignments is:

$$P(z_i = j | z_{i-1}, m_i, u_i, r_i) \propto P(r_i |\, \psi_i, \delta_i, c)\, \frac{n_{m_{-i,j}}^{MT} + \beta}{\sum_{k=1}^{M} n_{m_{-i,j}}^{MT} + M\beta}\, \frac{n_{u_{-i,j}}^{UT} + \alpha}{\sum_{t=1}^{T} n_{u_{-i,j}}^{UT} + T\alpha} \qquad (\textit{eq. 5})$$

Here, the subscript notation "-i" refers to the fact that the topic assignment for the current rating z_i is removed from the counts. Parameters α and β are parameters for the dirichlet priors of the multinomial distributions. (Steyvers and Griffiths, 2007).

Section 2: Parameter Estimation for Choices

To infer parameters for new users, the Ratings Topic Model can first be trained on a database of users in order to learn parameters $\psi_{m,t}$ and ϕ, as well as the rating-cutoffs c (if they are set globally across all users). These parameters are then fixed.

The inference problem for the new users' training ratings and choices corresponds to the two graphical models depicted in Figure S.2. Since the database parameters have been fixed, topic parameters $\psi_{m,t}$ and ϕ_t and the ordered-logit parameter c are observed. For both cases, δ parameters are sampled as described in eq. 3, with the caveat that in the absence of any training ratings, δ is sampled directly from the prior. Inference of θ_u is performed using collapsed-Gibbs sampling.

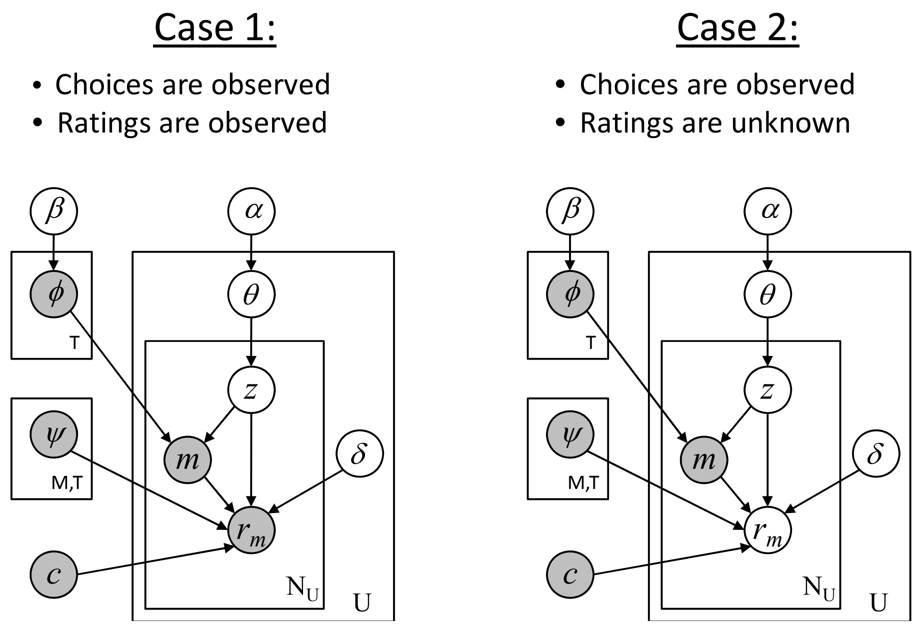

Figure S.2: Graphical model illustrating inference problem for new users given fixed database-parameters. "Case 1" corresponds to the situation in which ratings are observed, and "Case 2" corresponds to the situation in which only choices are observed.

To make a topic assignment for an observed rating we use eq. 5. However, sampling a topic assignment when only a choice is observed requires a modified procedure. To make a topic assignment, we first sample a rating from the posterior predicted distribution of rating $r_{u,m}$:

$$p(r_i = j|m_i, u_i) \propto \sum_{j=1}^{T} p(r_i\,|m_i, u_i, z_i = j\,)\; p(m_i|z_i = j)\; p(z_i = j|u_i) \qquad (eq.\ 6)$$

and then sample a topic conditional on this rating *as if* it were observed, using eq. 5.

Section 3: Inference details

To obtain the parameter estimates for the model comparison shown in Figure 4, s5 chains were run for T=1, 10, 20, 25 and 50. For each chain, 400 iterations of the Gibbs sampler were run, with 2 Metropolis-hastings iterations run between each iteration of the Gibbs Sampler. After 400 iterations, topic-assignments were fixed and 400 samples of each of the ordered-logit parameters were drawn to produce a point estimate for each parameter. Rating and choice predictions were computed for each chain and averaged over chains.

To obtain estimates for the results shown in Figure 5, a single data-base chain was run to generate estimates for database parameters. Then, for each combination of observations of choices and ratings, 15 chains were run. Estimates of user parameters were averaged over chains and a single rating prediction was computed across for these averaged parameter estimates. This was repeated for each of the five validation/training datasets, and accuracy of the predictions was averaged across all five datasets.

References

Griffiths, T. L., Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences, 101,* 5228-5235

Steyvers, M. and Griffiths, T.L. (2007) *Probabilistic topic models*. In Latent Semantic Analysis: A Road to Meaning (Landauer, T. et al., eds), Erlbaum.