MIS Quarterly

SPECIAL ISSUE: BUSINESS INTELLIGENCE RESEARCH

A HIDDEN MARKOV MODEL FOR COLLABORATIVE FILTERING

Nachiketa Sahoo
School of Management, Boston University, 595 Commonwealth Avenue, Boston, MA 02215 U.S.A.
and iLab, Heinz College, Carnegie Mellon University, Pittsburgh, PA 15213 U.S.A. {nachi@bu.edu}

Param Vir Singh
David A. Tepper School of Business and iLab, Heinz College, Carnegie Mellon University,
Pittsburgh, PA 15213 U.S.A. {psidhu@cmu.edu}

Tridas Mukhopadhyay
David A. Tepper School of Business and iLab, Heinz College, Carnegie Mellon University,
Pittsburgh, PA 15213 U.S.A. {tridas@cmu.edu}

Appendix A

Prediction Distribution over Items

The probability of a user *u* selecting an item *i* in time period *t* is

$$P(i \in I^t_u) = \sum_k P(Z^t_u = k) P(i \in I^t_u; a_k, b_k, \theta_k) \tag{1}$$

where, I^t_u is the set of items selected by the user *u* in time period *t*.

Conditional on the user being in state the probability of the item being observed is

$$P(i \in I^t_u; a_k, b_k, \theta_k) = \sum_{N^t_u = 0}^{\infty} P(N^t_u; a_k, b_k) \times P(i \in I^t_u | N^t_u; \theta_k) \tag{2}$$

This follows from considering the probability of each possible number of items (N^t_u) selected by the user and for each of those number of items selected the probability of the item being included in the selected set.

Since the distribution over the items is a multinomial, the probability that an item *i* is observed in the N^t_u items that are observed in time period *t* is

$$P(i \in I^t_u | N^t_u; \theta_k) = 1 - P(i \notin I^t_u | N^t_u; \theta_k) = 1 - (1 - P(i|\theta_k)^{N^t_u} = 1 - (1 - \theta_{ki})^{N^t_u} \tag{3}$$

The last equality of equation (3) follows from the fact that the probability of any item *i* for in a multinomial is equal to the parameter of the multinomial specific to that item. Substituting equation (3) in Equation (2) we get

$$\begin{aligned} P(i \in I_u^t; a_k, b_k, \theta_k) &= \sum_{N_u^{t+1}=0}^{\infty} P(N_u^t; a_k, b_k) \times \left(1 - (1-\theta_{ki})^{N_u^t}\right) \\ &= 1 - \sum_{N_u^{t+1}=0}^{\infty} P(N_u^t; a_k, b_k) \times (1-\theta_{ki})^{N_u^t} \end{aligned} \tag{4}$$

This summation, which is expectation of the exponential function of $N_u^t \ln(1-\theta_{ki})$, can be obtained using the identity for moment generating function for the negative binomial distribution.

$$\left(e^{Xt}\right)_{x\sim P_{NDBD(a,b)}} = \left\{1 + b(1-e^t)^{-a}\right\} \tag{5}$$

Substituting $\ln(1-\theta_{ki})$ for *t* in equation (5) and using the results in equation (4) after some algebraic reduction, we obtain

$$P(i \in I_u^t; a_k, b_k, \theta_k) = 1 - (1 + b_k\theta_{ki})^{-a_k} \tag{6}$$

Substituting equation (6) in equation (1), we obtain the following expression for the probability of observing the item *i*

$$P(i \in I_u^t) = \sum_k P(Z_u^t = k)\left(1 - (1 + b_k\theta_{ki})^{-a_k}\right) = 1 - \sum_k P(Z_u^t = k)(1 + b_k\theta_{ki})^{-a_k} \tag{7}$$

For ordering items by their probability of occurrence we can drop the unity and order the items by $-\sum_k P(Z_u^t = k)(1 + b_k\theta_{ki})^{-a_k}$.

The distribution over the states for a user can be calculated from the user's distribution in the previous time period as $P(Z_u^t = k) = \sum_{l=1}^{K} P(Z_u^{t-1} = l) P(Z_u^t = k | Z_u^{t-1} = l)$, where *K* is the number of possible states.

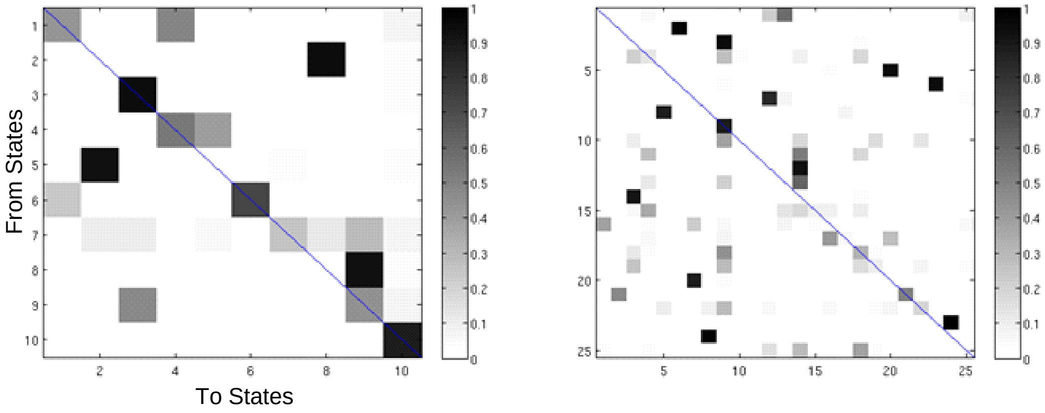
Appendix B

Latent Classes and Transitions in Blog Data

Upon examination of the transition probability matrix we find that, although there are classes from which the users do not move, there are many classes from which the users tend to switch to other classes in the subsequent time period. This behavior is illustrated in Figure B1.

The latent classes can be distinguished by their different intensity of reading and by the items that are the most popular in the latent class. These are reported in Table B1 for the HMM with 10 latent classes. As we can see, they differ in how much a user reads when the user is under a given latent class. Although there is some overlap in the top articles read by the users in each class, they have several differences in their selection of favorite articles to read.

Note that the class under which the users read the most, class 7, is also the one from which they switch away to a less active class, such as class 8. This suggests that it is improbable that the users will stay in a highly active class for long. On the other hand, the users tend to stay longer in a class that is less active, for example, class 10.



Notes: The transition probability matrix shown as gray scale images for HMM with 10 latent classes (left) and 25 latent classes (right). The darker the cell color is, the larger the transition probability. Although there are a few large probabilities on the diagonal, there are several large off diagonal probabilities suggesting a class switching behavior for the user.

Figure B1. Transition Probability Matrix

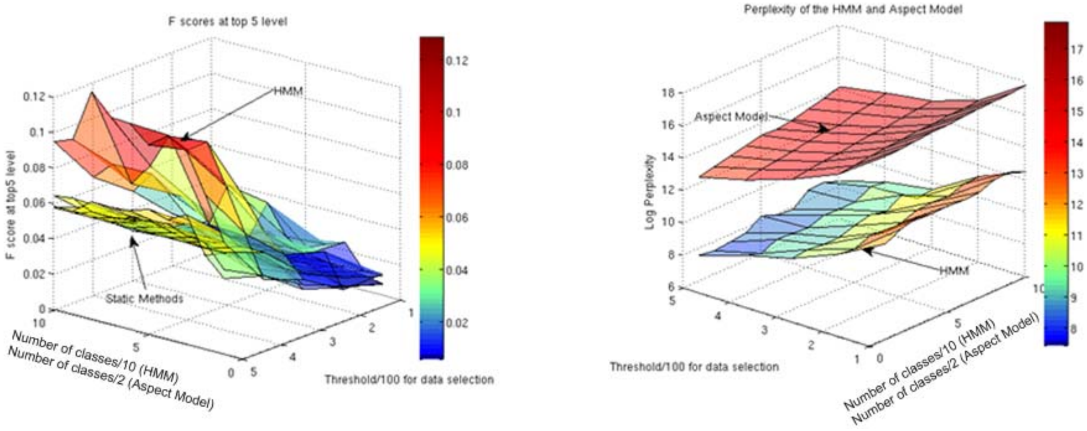
Table B1. Latent Classes[†]

Latent Class	Average number of articles read in a month	Index of the top 5 most probable articles to be read				
		#1	#2	#3	#4	#5
1	3.9121	233	107	113	223	126
2	6.6621	21	39	52	230	46
3	1.3722	233	223	107	167	80
4	3.4601	233	39	253	42	178
5	6.8933	39	42	150	36	230
6	1.7879	126	223	233	113	107
7	17.5808	39	31	156	3	187
8	5.4393	39	102	52	156	126
9	2.6996	39	52	107	126	205
10	0.0396	126	233	108	113	104

[†] The latent classes are characterized here by the average number of posts read by a user in the class in a month and the top articles read in the class.

Appendix C

Sensitivity Analysis with Blog Data



Notes: Sensitivity analysis is performed by varying the data densities and the number of classes and factors. The number of classes used for HMM was varied from 10 to 100 and from 2 to 20 for the static models. The data density was controlled by varying the minimum number of articles a user must visit and the minimum number of users an article must be read by from 100 to 500. The “Number of Classes” axis shows (number of classes/10) used for the HMM and (number of classes/2) used for static models. The “Threshold” axis shows the (minimum threshold/100) used for all the methods. The *F-score* when the top 5 articles are recommended is shown in the left plot. The *perplexity* of the HMM and aspect model are compared in the right plot.

Figure C1. Sensitivity Analysis