The Thirteenth International Symposium on Artificial Life and Robotics 2008(AROB 13th '08),
B-Con Plaza, Beppu, Oita, Japan, January 31-February 2, 2008

A Unified Approach to Collaborative and Feature-based Recommendation based on Probabilistic Latent Semantic Models

Ami Ishihara[1], Jun-ichiro Hirayama[1], Takashi Takenouchi[1], Shin Ishii[2]

[1]Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara, 630-0192
[2]Graduate School of Informatics, Kyoto University
Gokasho, Uji, Kyoto, 611-0011

Abstract

Recommendation systems have recently been put into practical uses in e-commerce web sites. In these systems, collaborative filtering (CF) is often used. Recommendation by CF is based on only user's ratings, and does not use information of attributes with respect to items. On the other hand, content-based filtering uses similarity of content information of items, for recommendation, rather than the user's ratings. Recently several "hybrid" methods have been proposed; those methods make recommendations by utilizing both the rating and content information about items. For published datasets, we can often use additional demographic information which typically represents status of users such as ages and occupations. In this study, we combine two types of probabilistic models, the aspect model and the naive Bayes model to deal with the rating, content and demographic information in a unified manner. In our model, content and demographic information given the rating is represented by the naive Bayes model and a distribution of the rating is represented by the aspect model. When applying the proposed method to the "MovieLens" dataset, the proposed method improves the prediction performance compared with the aspect model.

1 Introduction

Recommendation systems have recently been put into practical uses in various application domains. Such applications include, for example, recommendation of books or movies on some e-commerce web sites [1]. In these systems, collaborative filtering (CF) is used as a key technique to realize effective recommendation of items for users. In general, CF recommends an item for a target user by utilizing observed ratings given by other users whose rating pattern is similar to that of the target user. Such an approach to CF is contrastive to another popular approach for recommendation, content-based filtering, which does not employ other users' ratings but use additional content information about the items. While the original CF method utilizes only ratings for items, recently several "hybrid" methods of collaborative filtering and content-based filtering, have also been proposed; the hybrid methods incorporate additional content information into the framework of CF.

While the CF methods cannot make appropriate recommendation for a new item until the new item is rated by a substantial number of users, the "hybrid" methods can predict the rating by utilizing the content-related information and hence make a recommendation. In many practical situations, however, we can additionally use users' demographic information, which represents user's age, occupations and contextual situation such as seasons, day or night, and locations. These additional demographic information, if available, can also be useful for further improvement of performance of CF. Currently, however, little has been reported on such an attempt to incorporate such information within a unified manner.

In this study, we investigate a way of combining such additional information which is available in realistic contexts of recommendation applications, and develop a unified framework. For this purpose, we combine two types of probabilistic generative models: the aspect model and the naive Bayes model. The aspect model is a probabilistic mixture model for CF, which has latent classes and associates co-occurrences among users, items and ratings with a set of latent variables. The naive Bayes is one of the basic techniques for feature-based classification, assuming a simple generative model in which each feature depends only on a target variable to be estimated, i.e., the rating variables in our context. By applying the aspect model as a prior distribution of the rating variable in the naive Bayes model, we then propose a novel model which can incorporate additional information into the framework of CF in a natural manner. In section 2, we briefly introduce the related model: the aspect model [5] in which the collaborative filtering is extended with a probabilistic model, and the naive Bayes model. In section 3, our novel model is described, and then, an experiment with the MovieLens

The Thirteenth International Symposium on Artificial Life and Robotics 2008(AROB 13th '08),
B-Con Plaza, Beppu, Oita, Japan, January 31-February 2, 2008

dataset is done in section 4.

2 Traditional Methods

First, we describe basic settings and notations. Let I be a number of items and U be a number of users. We assume that there is a rating matrix $\boldsymbol{R}$ which is an $I \times U$ matrix and whose component each represents a rating value $r_{i\,u} \in \{1, \cdots, K\}$ for an item $i \in \{1, \cdots, I\}$ voted by a user $u \in \{1, \cdots, U\}$. In practical situations, a large part of $\boldsymbol{R}$ is not observed and then a part of $\boldsymbol{R}$ is available as a dataset. Let $\mathcal{V} = \{(i,u) | r_{i\,u} \text{ is observed}\}$ be a set of indexes of observed ratings in $\boldsymbol{R}$. We denote $\boldsymbol{V} = \{r_{i\,u} | (i,u) \in \mathcal{V}\}$ and $\tilde{\boldsymbol{V}} = \{r_{i\,u} | (i,u) \notin \mathcal{V}\}$ as sets of observed and unobserved ratings, respectively. In addition to the rating information $\boldsymbol{V}$, we assume that content information $\boldsymbol{C} = \{\boldsymbol{c}_i\}_{i=1}^{I}$ for items and demographic information $\boldsymbol{D} = \{\boldsymbol{d}_u\}_{u=1}^{U}$ for users are available. Here, let $\boldsymbol{c}_i$, a vector of length L, correspond to the content information associated with the item i and $c_{i\,l} \in \{1, \cdots, S_l\}$ be the l-th component of $\boldsymbol{c}_i$, and let $\boldsymbol{d}_u$, a vector of length H, correspond to the demographics information associated with the user u. The h-th component $d_{u\,h}$ of $\boldsymbol{d}_u$ is in $\{1, \cdots, T_h\}$.

2.1 Aspect Model

Hofmann *et al*. [5] proposed a probabilistic generative model for collaborative filtering, based on the probabilistic latent semantic analysis, or the aspect model [4]. The model basically assumes a mixture model for rating variables $r_{i\,u}$, and the mixing rate varies over users. While being different from original formulation, the generative model of rating variables can be given as follows. Let z_u be an indicator of the latent class associated with user u. Then, the joint distribution of $r_{i\,u}$ and z_u is given as

$$p(r_{i\,u}, z_u | \boldsymbol{\theta}_i, \quad _u) = p(z_u | \quad _u) p(r_{i\,u} | z_u, \boldsymbol{\theta}_i),$$

$$p(z_u | \quad _u) = \prod_{m=1}^{M} [\phi_{u\,m}]^{\ (z_u\ m)},$$

$$p(r_{i\,u} | z_u, \boldsymbol{\theta}_i) = \prod_{m=1}^{M} \prod_{k=1}^{K} [\theta_{i\,k}^{(m)}]^{\ (r_{i,u}\ k)\ (z_u\ m)},$$

where $\boldsymbol{\theta}_i = \{\theta_{i\,k}^{(m)}; k = 1, \cdots, K, m = 1, \ldots, M\}$ and $\ _u = \{\phi_{u\,m}; m = 1, \cdots, M\}$ are sets of model parameters which define distributions. Note that $\phi_{u\,m}$ represents a probability that a user u belongs to a latent class m, and $\theta_{i\,k}^{(m)}$ is the probability of rating $r_{i\,u}$ being equal to k when the item i under the latent class m. We thus assume $\sum_{m=1}^{M} \phi_{u\,m} = 1$ and $\sum_{k=1}^{K} \theta_{i\,k}^{(m)} = 1$ are satisfied. $\delta(a,b)$ is the Kronecker's delta function, which takes one for $a = b$, or zero otherwise.

Given a dataset $\boldsymbol{V}$, this model can be trained by the Expectation-Maximization (EM) algorithm [3].

2.2 Feature-based Rating Estimation

In the naive Bayes model for content-based filtering [2], each component of $\boldsymbol{c}_i$ is assumed to be independently generated given the rating $r_{i\,u}$. Given a prior distribution of rating, $p(r_{i\,u})$, the joint distribution of $r_{i\,u}$ and $\boldsymbol{c}_i$ is written as

$$p(r_{i\,u}, \boldsymbol{c}_i) = p(r_{i\,u}) \prod_{l=1}^{L} p(c_{i\,l} | r_{i\,u}). \tag{1}$$

Provided that no additional information is available, the rating value is estimated simply by maximizing its posterior distribution:

$$p(r_{i\,u} | \boldsymbol{c}_i) \propto p(r_{i\,u}) \prod_{l=1}^{L} p(c_{i\,l} | r_{i\,u}),$$

which can be easily calculated.

3 Unified approach

In addition to the items' content information, the users' demographic information is sometimes available. We also use the additional information to estimate the rating in the same way as in the content-based filtering. Assuming a naive Bayes model again, the joint distribution of $\boldsymbol{d}_u$ and the rating $r_{i\,u}$ can be written as

$$p(r_{i\,u}, \boldsymbol{d}_u) = p(r_{i\,u}) \prod_{h=1}^{H} p(d_{u\,h} | r_{i\,u}), \tag{2}$$

with a specification of prior distribution, $p(r_{i\,u})$. Furthermore, we can integrate the two naive Bayes models to estimate the rating based on both of the content and demographic information. Unifying equations (1) and (2), we can formulate the joint distribution over these variables as

$$p(\boldsymbol{c}_i, \boldsymbol{d}_u, r_{i\,u}) = p(r_{i\,u}) \prod_{l=1}^{L} p(c_{i\,l} | r_{i\,u}) \prod_{h=1}^{H} p(d_{u\,h} | r_{i\,u}). \tag{3}$$

The posterior distribution of $r_{i\,u}$ can again be readily calculated to estimate the rating based on the two kinds of information. Equation (3) contains a prior distribution of $p(r_{i\,u})$, which is usually assumed as non-informative in the context of content-based (or demographic) filtering, by reflecting the lack of knowledge about the distribution of rating. When one consideres the use of collaborative filtering, however, the rating distribution of each user is actually modeled by utilizing the relationship to other users. Based on this fact, we present our main idea of this study: in order to integrate the collaborative and feature-based approaches for rating estimation, we use the aspect model as the prior distribution of rating variables in the naive Bayes model. The joint distribution of all the variables of interest, then, is given as

$$p(\boldsymbol{c}_i, \boldsymbol{d}_u, r_{i\,u}, z_u|\boldsymbol{\theta}_i, \quad_u, \quad, \boldsymbol{\omega})$$
$$= p(z_u| \quad_u)p(r_{i\,u}|z_u, \boldsymbol{\theta}_i)\prod_{l=1}^{L} p(c_{i\,l}|r_{i\,u}, \quad_l)\prod_{h=1}^{H} p(d_{u\,h}|r_{i\,u}, \boldsymbol{\omega}_h). \tag{4}$$

where $\quad = \{\quad_l\}_{l=1}^{L}$ and $\boldsymbol{\omega} = \{\boldsymbol{\omega}_h\}_{h=1}^{H}$. Here $\quad_l = \{\quad_{l\,s}^{(k)}; s = 1, \cdots, S_l, k = 1, \ldots, K\}$ and $\boldsymbol{\omega}_h = \{\omega_{h\,t}^{(k)}; t = 1, \cdots T_h, k = 1, \cdots, K\}$ are sets of parameters describing each component of (4) as

$$p(c_{i\,l}|r_{i\,u}, \quad_l) = \prod_{k=1}^{K}\prod_{s=1}^{S_l} \quad_{l\,s}^{(k)} \quad ^{(c_{i,l}\ s)\ (r_{i,u}\ k)},$$
$$p(d_{u\,h}|r_{i\,u}, \boldsymbol{\omega}_h) = \prod_{k=1}^{K}\prod_{t=1}^{T_h} \omega_{h\,t}^{(k)} \quad ^{(d_{u,h}\ t)\ (r_{i,u}\ k)}.$$

Note that $\quad_{l\,s}^{(k)}$ represents a probability that the content $c_{i\,l}$ becomes s when the item i is rated as k, and $\omega_{h\,t}^{(k)}$ is a probability that the demographics $d_{u\,h}$ becomes t when the user u votes the rating k. Then, we have $\sum_{s=1}^{S_l} \quad_{l\,s}^{(k)} = 1$ and $\sum_{t=1}^{T_h} \omega_{h\,t}^{(k)} = 1$. According to the EM algorithm, we can estimate these parameters by maximizing the following free energy.

$$F[q(\tilde{\boldsymbol{V}}, \boldsymbol{z}), \boldsymbol{\theta}, \quad, \quad, \boldsymbol{\omega}] = \left\langle \log \frac{p(\boldsymbol{R}, \boldsymbol{z}, \boldsymbol{C}, \boldsymbol{D}|\boldsymbol{\theta}, \quad, \quad, \boldsymbol{\omega})}{q(\boldsymbol{V}, \boldsymbol{z})} \right\rangle_{q(\tilde{\boldsymbol{V}}, \boldsymbol{z})} \tag{5}$$

where $p(\boldsymbol{R}, \boldsymbol{z}, \boldsymbol{C}, \boldsymbol{D}|\boldsymbol{\theta}, \quad, \quad, \boldsymbol{\omega})$ is the likelihood for complete data. We assume $q(\tilde{\boldsymbol{V}}, \boldsymbol{z}) = \prod_{u=1}^{U} q(\tilde{\boldsymbol{V}}_u, z_u)$ where $\tilde{\boldsymbol{V}}_u$ is a set of unobserved ratings of the user u. The E-step and M-step are written as follows.

[E-step]: Let $\hat{\boldsymbol{\theta}}$, $\hat{\quad}$, $\hat{\quad}$ and $\hat{\boldsymbol{\omega}}$ be the estimated parameters in the previous step. We maximize the F with respect to $q(\tilde{\boldsymbol{V}}, \boldsymbol{z})$ with the fixed parameters $\hat{\boldsymbol{\theta}}$, $\hat{\quad}$, $\hat{\quad}$ and $\hat{\boldsymbol{\omega}}$. Consequently, for $r_{i\,u} \in \tilde{\boldsymbol{V}}_u$, we obtain $\hat{q}(r_{i\,u}, z_u) = \hat{q}(r_{i\,u}|z_u)\hat{q}(z_u)$ where

$$\hat{q}(r_{i\,u} = k|z_u) \propto \exp(\quad(k, m, \boldsymbol{c}_i, \boldsymbol{d}_u)),$$
$$\hat{q}(z_u = m) \propto \exp(\quad(m, \boldsymbol{c}_i, \boldsymbol{d}_u)),$$
$$\quad(k, m, \boldsymbol{c}_i, \boldsymbol{d}_u) = \log\hat{\theta}_{i\,k}^{(m)}$$
$$+ \sum_{l=1}^{L}\sum_{s=1}^{S_l}\delta(c_{i\,l}, s)\log\hat{\quad}_{l\,s}^{(k)} + \sum_{h=1}^{H}\sum_{t=1}^{Th}\delta(d_{u\,h}, t)\log\hat{\omega}_{h\,t}^{(k)},$$
$$\quad(m, \boldsymbol{c}_i, \boldsymbol{d}_u) = \log\hat{\phi}_{u\,m}$$
$$+ \sum_{i\in\mathcal{V}_u}\sum_{k=1}^{K}\delta(r_{i\,u}, k)\log\hat{\theta}_{i\,k}^{(m)} + \sum_{i\in\mathcal{V}}\quad(m, \boldsymbol{c}_i, \boldsymbol{d}_u)),$$
$$\quad(m, \boldsymbol{c}_i, \boldsymbol{d}_u) = \log\sum_{k=1}^{K}\exp(\quad(k, m, \boldsymbol{c}_i, \boldsymbol{d}_u)).$$

[M-step]: Update parameters $\quad$, $\boldsymbol{\theta}$, $\quad$ and $\boldsymbol{\omega}$ as to maximize the free energy F for the fixed $\hat{q}(\tilde{\boldsymbol{V}}, \boldsymbol{z})$ as

$$\hat{\phi}_{u\,m} = \hat{q}(z_u),$$
$$\hat{\theta}_{i\,k}^{(m)} \propto \sum_{u=1}^{U}\big\langle\delta(r_{i\,u}, k)\delta(z_u, m)\big\rangle,$$
$$\hat{\quad}_{l\,s}^{(k)} \propto \sum_{i=1}^{I}\delta(c_{i\,l}, s)\sum_{u=1}^{U}\big\langle\delta(r_{i\,u}, k)\big\rangle,$$
$$\hat{\omega}_{h\,t}^{(k)} \propto \sum_{u=1}^{U}\delta(d_{u\,h}, t)\sum_{i=1}^{I}\big\langle\delta(r_{i\,u}, k)\big\rangle,$$

where

$$\big\langle\delta(r_{i\,u}, k)\delta(z_u, m)\big\rangle = \begin{cases}\hat{q}(z_u)\delta(r_{i\,u}, k) & \text{if } (i, u) \in \mathcal{V}\\ \hat{q}(z_u)\hat{q}(r_{i\,u}|z_u) & \text{if } (i, u) \notin \mathcal{V},\end{cases}$$
$$\big\langle\delta(r_{i\,u}, k)\big\rangle = \begin{cases}\delta(r_{i\,u}, k) & \text{if } (i, u) \in \mathcal{V}\\ \hat{q}(r_{i\,u}) & \text{if } (i, u) \notin \mathcal{V}.\end{cases}$$

By repeating above two steps until (5) converges, we obtain the estimated parameters $\hat{\boldsymbol{\theta}}$, $\hat{\quad}$, $\hat{\quad}$ and $\hat{\boldsymbol{\omega}}$. By plugging those parameters into (4), we can calculate the posterior distribution of rating $r_{i\,u}$ for the item i given content and demographic information for the user u. The distribution of the rating given additional information, $\boldsymbol{c}_i$ and $\boldsymbol{d}_u$, is written as

$$p(r_{i\,u}|\boldsymbol{c}_i, \boldsymbol{d}_u) \propto \sum_{m=1}^{M} p(z_u|\hat{\quad}_u)p(r_{i\,u}|z_u, \hat{\boldsymbol{\theta}}_i)$$
$$\times\prod_{l=1}^{L} p(c_{i\,l}|r_{i\,u}, \hat{\quad}_l)\prod_{h=1}^{H} p(d_{u\,h}|r_{i\,u}, \hat{\boldsymbol{\omega}}_h),$$

and we predict the rating $\tilde{r}_{i\,u}$ as

$$\tilde{r}_{i\,u} = \arg\max_{r_{i,u}} p(r_{i\,u}|\boldsymbol{c}_i, \boldsymbol{d}_u).$$

4 Experiments

4.1 Datasets

We examined prediction performance of the rating of the proposed method by comparing with the traditional aspect model for collaborative filtering. In experiments, we used a part of the MovieLens dataset which was collected by the GroupLens project [6]. We selected a part of the dataset for the rating matrix $\boldsymbol{R}$ which contains information of 50 users and 100 items, in which 83% of ratings is missing. We randomly selected 70% of rating data from the rating matrix $\boldsymbol{R}$ and used it as a training dataset $\boldsymbol{V}$, and remained 30% of rating data were used as a test dataset $\tilde{\boldsymbol{V}}$. Test performances were measured by the prediction accuracy for the test dataset $\tilde{\boldsymbol{V}}$. We set $K = 5$, $I = 100$, $U = 10$, $L = 19$ and $H = 3$.

4.2 Evaluation Criteria

For performance evaluation, we calculated the mean absolute error (MAE) of the predicted rating $\tilde{r}_{i\,u}$ from the actual rating $r_{i\,u}$:

$$\frac{1}{n}\sum_{(i\ u)\in\mathcal{W}}\left|\tilde{r}_{i\,u} - r_{i\,u}\right|,$$

where $\mathcal{W}$ is a set of arbitary indexes and n is the number of components included in $\mathcal{W}$. The smaller MAE value implies better performance.

4.3 result

Figure 1 shows the average of 10 trials of the MAE scores for the training dataset $\boldsymbol{V}$, against various values of M, the number of latent classes. We observed that better performance was attained at a larger M value by both of the methods. Figure 2 shows the average of MAE for the test dataset, and we observe that the proposed method outperformed the aspect model. Note that the averaged test MAE increases for a large value of M, while the training MAE decreases; this implies over-fitting of the proposed method to the training dataset.

5 Conclusions

We proposed a probabilistic method for collaborative filtering based on the aspect model and the naive Bayes model. In our model, information of item-content and user-demographic is integrated in a unified manner. In an experiment with a synthetic dataset which was obtained from a realistic benchmark one, we observed that such integration of information effectively worked, and then the proposed method outperformed the traditional method.

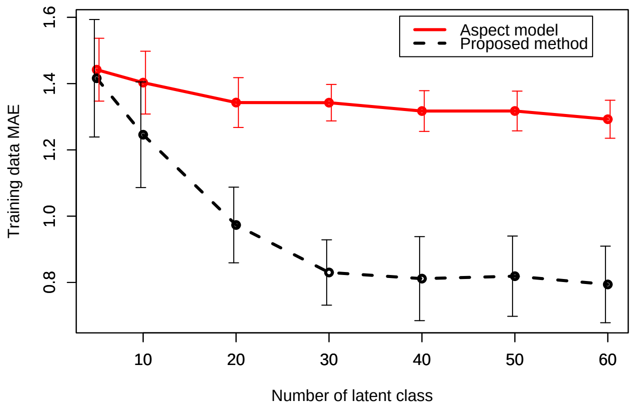

Figure 1: MAE for the training data, The number of users and items are 50 and 100, respectively

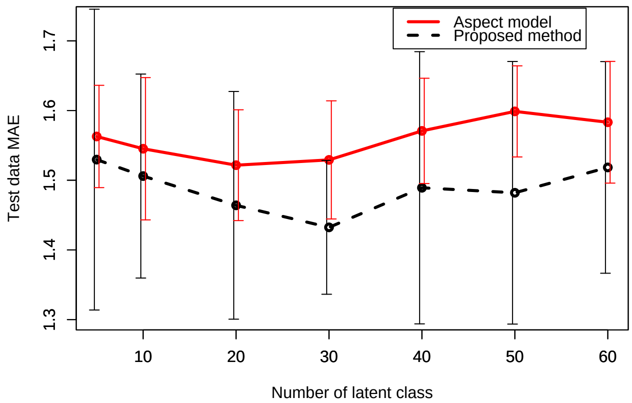

Figure 2: MAE for the test data, The number of users and items are 50 and 100, respectively

References

[1] Amazon.com : "http://www.amazon.com/"

[2] Breese, J. S., and Heckerman, D. , and Kadie, C., Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In *Proceeding of the 14th Conference on Uncertainty in Artificial Intelligence*, pp.43-52, 1998.

[3] Dempster, A. P., Laird, N. M. , and Rubin, D. B., Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B*, 39, pp.1-38, 1977

[4] Hofmann, T., Probabilistic Latent Semantic Analysis. In *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, pp.264-273, 1999.

[5] Hofmann, T., and, Pusicha, J., Latent class models for collaborative filtering. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence*, pp.688-693, 1999.

[6] Movielens : "http://www.movielens.org/"