Recommendation with Non-Random Missing Data

Kuhwan Jeong [1]

[1]Department of Statistics, Seoul National University, South Korea

July, 2017

Outline

1. Introduction
2. Missing Data Theory
3. Models for Non-Random Missing Data
   - MM/CPT-v
   - MM/Logit-vd
   - RAPMF-r
   - RAPMF-c
   - Bayesian-BM/OR
4. Proposed Model

Introduction

	item1	item2	item3	item4	item5
user1	5	4			
user2		5		4	
user3	4			4	
user4	5		5		
user5		4			5

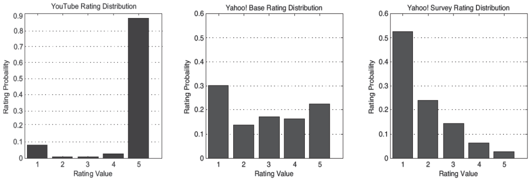
- Generally, previous works on recommender system can be classified into *neighborhood-based methods* and *model-based methods*.
- Neighborhood-based methods are likely to predict rating values in the range of 4 to 5.
- Model-based methods usually predict rating values based on (1) all rating values with missing values being replaced to 0, or (2) observed rating values only (*missing at random* assumption).

Introduction

Rating Frequency	Preference Level				
	Hate	Don't like	Neutral	Like	Love
Never	6.76%	4.69%	2.33%	0.11%	0.07%
Very Infrequently	1.59%	4.17%	9.46%	0.54%	0.35%
Infrequently	1.63%	4.44%	24.87%	1.48%	0.20%
Often	12.46%	22.50%	26.83%	25.30%	5.46%
Very Often	77.56%	64.20%	36.50%	72.57%	93.91%

- Yahoo! LaunchCast users were asked to report the frequency with which they choose to rate a song given their preference for that song [5].
- The choice to rate a song depends on the user's opinion of that song.
- Users were also directly asked if they thought their preferences for a song do not affect whether they choose to rate it.
  - 64.85% of users responded that their preferences do affect their choice to rate a song [5].

Introduction



(d) YouTube (e) Yahoo! User (f) Yahoo! Random

- About 90% of the ratings in the YouTube data take the maximum rating value [7].
- The marginal distributions of
  (e) existing rating values, and
  (f) new rating values of randomly selected songs [5]
  are significantly different.

Missing Data Theory

- Let N be the number of users, D be the number of items, and V be the number of rating values.
- Let $X = [x_{nd}]_{N \times D}$ where x_{nd} denotes the rating of user n for item d.
- Let $R = [r_{nd}]_{N \times D}$ be a binary matrix where $r_{nd} = 1$ denotes the observation of x_{nd}.
- X is divided into X^{obs} and X^{mis} where

$$X^{obs} = \{x_{nd} : r_{nd} = 1\}, \;\; X^{mis} = \{x_{nd} : r_{nd} = 0\}.$$

Missing Data Theory

- We model the data as a two-step procedure:
  1. A *data model* $P(X|\theta)$ generates the full data.
  2. A *missing data model* $P(R|X,\mu)$ determines which elements in X are observed.

$$P(R,X|\mu,\theta) = P(R|X,\mu)P(X|\theta)$$

- According to [4], there are three kinds of missing data assumptions:
  1. Missing Completely At Random (MCAR):

$$P(R|X,\mu) = P(R|\mu).$$

  2. Missing At Random (MAR):

$$P(R|X,\mu) = P(R|X^{obs},\mu).$$

  3. Not Missing At Random (NMAR)

Missing Data Theory

- Under the MAR assumption, the missing data model can be ignored

$$\begin{aligned}\int_{X^{mis}} P(R, X|\mu, \theta) dX^{mis} &= \int_{X^{mis}} P(R|X^{obs}, \mu) P(X|\theta) dX^{mis} \\ &= P(R|X^{obs}, \mu) \int_{X^{mis}} P(X|\theta) dX^{mis} \\ &= P(R|X^{obs}, \mu) P(X^{obs}|\theta)\end{aligned}$$

- Since the above derivation fails to hold under the NMAR assumption, we should consider the missing data model.

Models for Non-Random Missing Data

1. MM/CPT-v ([5])

- Multinomial Mixture Data Model $P(X|\theta, \phi)$

$$\begin{aligned} \theta = (\theta_1, \dots, \theta_K) &\sim Dir(\alpha, \dots, \alpha), \\ P(z_n = k|\theta) &= \theta_k, \\ \phi_{kd} = (\phi_{kd1}, \dots, \phi_{kdV}) &\sim Dir(\beta, \dots, \beta), \\ P(x_{nd} = v|z_n = k, \phi) &= \phi_{kdv}. \end{aligned}$$

  – The ordinal nature of rating values is ignored.

- CPT-v Missing Data Model $P(R|X, \mu)$

$$\begin{aligned} \mu = (\mu_1, \dots, \mu_V) &\sim Beta(a, b)^V, \\ P(r_{nd} = 1|x_{nd} = v, \mu) &= \mu_v. \end{aligned}$$

  – The probability of missing depends only on the rating value.

Models for Non-Random Missing Data

2. MM/Logit-vd ([6, 7])

- Multinomial Mixture Data Model $P(X|\theta, \phi)$

- Logit-vd Missing Data Model $P(R|X, \mu, \omega)$

$$\begin{aligned}
\mu &= (\mu_1, \ldots, \mu_V) \sim N(0, \sigma_1^2)^V, \\
\omega &= (\omega_1, \ldots, \omega_D) \sim N(0, \sigma_2^2)^D, \\
P(r_{nd} = 1 | x_{nd} = v, \mu, \omega) &= \frac{1}{1 + \exp(-\mu_v - \omega_d)}.
\end{aligned}$$

  - The probability of missing depends on the rating value and item.
  - Only marginal effects are included in the model.

Models for Non-Random Missing Data

3. RAPMF-r ([3])

- Probabilistic Matrix Factorization $P(X|U, W, \sigma^2)$

$$
\begin{aligned}
U = (U_1, \ldots, U_N) &\sim N(0, \sigma_U^2)^{K \times N}, \\
W = (W_1, \ldots, W_D) &\sim N(0, \sigma_W^2)^{K \times D}, \\
x_{nd}|U_n, W_d, \sigma^2 &\sim N(U_n^T W_d, \sigma^2).
\end{aligned}
$$

- Rating Dominant Response Model $P(R|U, W, \mu, \sigma^2)$

$$
\begin{aligned}
\mu = (\mu_1, \ldots, \mu_V) &\sim N(0, \sigma_\mu^2)^V, \\
P(r_{nd}|U_n, W_d, \mu, \sigma^2) &= \sum_{v=1}^{V} \alpha_v^{r_{nd}} (1 - \alpha_v)^{1 - r_{nd}} N(v|U_n^T W_d, \sigma^2),
\end{aligned}
$$

where $\alpha_v = \frac{e^{\mu_v}}{1 + e^{\mu_v}}$.

Models for Non-Random Missing Data

4. RAPMF-c ([3])

- Probabilistic Matrix Factorization $P(X|U, W, \sigma^2)$
- Context-aware Response Model $P(R|U, W, \mu, \sigma^2)$

$$\begin{aligned} \theta_U &\sim N(0, \lambda_U^2)^K, \\ \theta_W &\sim N(0, \lambda_W^2)^K, \\ \mu = (\mu_1, \ldots, \mu_V) &\sim N(0, \sigma_\mu^2)^V, \\ P(r_{nd}|U_n, W_d, \mu, \sigma^2) &= \sum_{v=1}^{V} \alpha_{ndv}^{r_{nd}} (1 - \alpha_{ndv})^{1-r_{nd}} N(v|U_n^T W_d, \sigma^2), \end{aligned}$$

where $\alpha_{ndv} = \frac{e^{\mu_v + U_n^T \theta_U + W_d^T \theta_W}}{1+e^{\mu_v + U_n^T \theta_U + W_d^T \theta_W}}$.

Models for Non-Random Missing Data

5. Bayesian-BM/OR ([1])

- Binomial Mixture Model $P(X|\theta, \phi)$

$$\begin{aligned} \theta = (\theta_1, \ldots, \theta_K) &\sim Dir(\alpha, \ldots, \alpha), \\ P(z_n = k|\theta) &= \theta_k, \\ \phi_k = (\phi_{k1}, \ldots, \phi_{kD}) &\sim Beta(a, b)^D, \\ x_{nd}|z_n = k, \phi &\sim 1 + Bin(V-1, \phi_{kd}). \end{aligned}$$

- Bernoulli OR Model $P(R|X, \mu, \nu, \gamma)$

$$\begin{aligned} \mu &= (\mu_1, \ldots, \mu_N) \sim Beta(c, d)^N, \;\; U_{nd}|\mu_n \sim Bern(\mu_n), \\ \nu &= (\nu_1, \ldots, \nu_D) \sim Beta(e, f)^D, \;\; M_{nd}|\nu_d \sim Bern(\nu_d), \\ \gamma &= (\gamma_1, \ldots, \gamma_V) \sim Beta(g, h)^V, \;\; T_{nd}|x_{nd} = v, \gamma \sim Bern(\gamma_v), \\ r_{nd} &= 1 - (1 - U_{nd})(1 - M_{nd})(1 - T_{nd}). \end{aligned}$$

Proposed Model

Motivation

	명량	국제시장	베테랑	아바타	도둑들
용찬	2	2	5	5	4
일상	1	2	5	4	5
세민	2	5	4	2	2
규승	2	4	5	2	1

- In previous works, complex relationships among users, items and rating values are not modeled in the missing data model.
- There are different patterns in rating and missing.

Proposed Model

- Data Model $P(X|\theta)$

$$\begin{aligned} \theta_n = (\theta_{n1}, \ldots, \theta_{nD}) &\sim DP\big(\alpha_1, F^D\big), \\ x_{nd}|\theta_n &\sim G(\theta_{nd}). \end{aligned}$$

  – The distribution pair (F, G) can be beta-binomial or gamma-truncated Poisson distribution.

- Missing Data Model $P(R|X, \mu)$

$$\begin{aligned} \mu_n = (\mu_{n1}, \ldots, \mu_{nV}) &\sim DP\big(\alpha_2, Beta(a, b)^V\big) \\ P(r_{nd} = 1|x_{nd} = v, \mu_n) &= \mu_{nv}. \end{aligned}$$

References

[1] Yong-Deok Kim and Seungjin Choi. Bayesian binomial mixture model for collaborative prediction with non-random missing data. In *Proceedings of the 8th ACM Conference on Recommender systems*, pages 201–208. ACM, 2014.

[2] Daryl Lim, Julian McAuley, and Gert Lanckriet. Top-n recommendation with missing implicit feedback. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 309–312. ACM, 2015.

[3] Guang Ling, Haiqin Yang, Michael R Lyu, and Irwin King. Response aware model-based collaborative filtering. *arXiv preprint arXiv:1210.4869*, 2012.

[4] Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*. John Wiley & Sons, 2014.

References

[5] Benjamin Marlin, Richard S Zemel, Sam Roweis, and Malcolm Slaney. Collaborative filtering and the missing at random assumption. *arXiv preprint arXiv:1206.5267*, 2012.

[6] Benjamin M Marlin and Richard S Zemel. Collaborative prediction and ranking with non-random missing data. In *Proceedings of the third ACM conference on Recommender systems*, pages 5–12. ACM, 2009.

[7] Benjamin M Marlin, Richard S Zemel, Sam T Roweis, and Malcolm Slaney. Recommender systems, missing data and statistical model estimation. In *IJCAI proceedings-international joint conference on artificial intelligence*, volume 22, page 2686, 2011.

[8] Harald Steck. Training and testing of recommender systems on data missing not at random. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 713–722. ACM, 2010.