Industrial Engineering
& Management Systems
Vol 13, No 4, December 2014, pp.421-431
ISSN 1598-7248 │ EISSN 2234-6473 │
http://dx.doi.org/10.7232/iems.2014.13.4.421

Supervised Learning-Based Collaborative Filtering Using Market Basket Data for the Cold-Start Problem

Wook-Yeon Hwang*
Data Analytics Department, Institute for Infocomm Research, A*STAR, Singapore 138632

Chi-Hyuck Jun
Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang, Korea

(Received: July 12, 2014 / Revised: November 7, 2014 / Accepted: November 7, 2014)

ABSTRACT
The market basket data in the form of a binary user-item matrix or a binary item-user matrix can be modelled as a binary classification problem. The binary logistic regression approach tackles the binary classification problem, where principal components are predictor variables. If users or items are sparse in the training data, the binary classification problem can be considered as a cold-start problem. The binary logistic regression approach may not function appropriately if the principal components are inefficient for the cold-start problem. Assuming that the market basket data can also be considered as a special regression problem whose response is either 0 or 1, we propose three supervised learning approaches: random forest regression, random forest classification, and elastic net to tackle the cold-start problem, comparing the performance in a variety of experimental settings. The experimental results show that the proposed supervised learning approaches outperform the conventional approaches.

Keywords: Market Basket Data, Cold-Start Problem, Supervised Learning-Based Collaborative Filtering, Random Forest, Elastic Net

* Corresponding Author, E-mail: hwangwy@i2r.a-star.edu.sg

1. INTRODUCTION

Since Goldberg *et al*. (1992) first introduced collaborative filtering (CF) which assumes that a good way to find the content a certain customer is interested in is to find other people who have interests similar to him or her, the CF has been widely used for recommender systems in a variety of industries, such as Web pages, movies, articles and products (Hill *et al*., 1995; Lee *et al*., 2001; Resnick *et al*., 1994; Shardanand and Maes, 1995). The objective of the CF is to suggest new items for an active user based on the user's previous likings and the opinions of other like-minded users (Sarwar *et al*., 2001). The CF can be roughly classified into two classes: memory-based CF approach and model-based CF approach (Breese *et al*., 1998). Some model-based CF methodologies have been developed using such approaches as Bayesian network, clustering, regression, association rule, eigentaste, and so on. (Breese *et al*., 1998; Goldberg *et al*., 2001; Leung *et al*., 2008; Park *et al*., 2012).

The model-based CF approaches are originally conducted on the basis of voting scores. However, some of the voting scores are usually missing. To tackle such a problem, some researchers have proposed a new CF scheme using market basket data (Mild and Reutterer, 2001, 2003), which can be transformed into a so-called binary user-item matrix or a binary item-user matrix having customers (users) and products (items) consisting of ones (purchases) and zeros (non-purchases). This scheme can lead to a poor recommendation accuracy due to the information loss. The market basket data in the form of a binary user-item matrix or a binary item-user matrix can be modelled as a binary classification problem (Lee *et al*., 2005; Lee and Olafsson, 2009). Lee *et al*. (2005) have proposed a binary logistic regression approach for market basket data in the form of a binary user-item matrix whose principal components are predictor variables. Since our research is based on their method, we will explain this method in more detail in Section 2. Lee and Olafsson (2009) have extended the binary logistic regression approach further to two-way cooperative prediction for CF recommendations, where the user-based CF and item-based CF approaches are combined (Breese *et al*., 1998; Sarwar *et al*., 2001). For making a prediction, the user-based CF approach using a binary user-item matrix utilizes user similarities, whereas the item-based CF approach using a binary item-user matrix uses item similarities.

Moreover, new users or items can limit the CF approaches. The phrase cold-start describes the situation when almost nothing is known about customer preferences (e.g., a start-up company has no little or no purchase history) (Schein *et al*., 2002). The problem of making recommendations for new users or new items can be also considered as a cold-starting problem. Such a cold-start problem is prevalent in recommendation systems (Schein *et al*., 2002; Ahn, 2008; Park and Chu, 2009; Lika *et al*., 2014). If users or items are sparse in the training data, the binary classification problem can be also considered as a cold-start problem. In this paper, we call the cold-start problem a new user problem if users are sparse in the training data, whereas it is called a new item problem if items are sparse in the training data. The binary logistic regression approach may not function appropriately if principal components are inefficient for the cold-start problem. We assume that the market basket data can be also modelled as a special regression problem, whose response is either 0 or 1. Therefore, we propose three supervised learning approaches: random forest regression, random forest classification, and elastic net to tackle the cold-start problem.

The remainder of this paper is organized as follows. In Section 2, we review in detail existing approaches. Section 3 describes the proposed supervised learning approaches. In Section 4, a numerical study is conducted to compare the proposed approaches with existing methods. Concluding comments are presented in Section 5.

2. EXISTING APPROACHES

2.1 The Correlation Approach

Breese *et al*. (1998) first proposed a user-based CF approach based on Pearson correlation coefficients. Mild and Reutterer (2003) provided a modified version of the approach. Suppose that there are n users and m items in the form of a binary user-item matrix. Let v_{ij} be an indicator variable of the event that item j is purchased by user u_i (i = 1, 2, ⋯, n; j = 1, 2, ⋯, m) as follows.

$$v_{ij} = \begin{cases} 1, & \text{when } u_i \text{ purchased item } j \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $\boldsymbol{V} = (v_{ij})$ is called a binary user-item matrix.

The mean score for user u_i is defined as

$$v_i = \frac{1}{|I_i|}\sum_{j \in I_i} v_{ij}, \tag{2}$$

where I_i is the set of items which user u_i purchases. Then the predicted score by the active user u_a for the item j, P_{aj} is defined as

$$P_{aj} = k_a \sum_{i=1}^{n} w(a, i) v_{ij'} \tag{3}$$

where $w(a, i) = \dfrac{\sum_j (v_{aj} - \bar{v}_a)(v_{ij} - \bar{v}_i)}{\sqrt{\sum_j (v_{aj} - \bar{v}_a)^2 \sum_j (v_{ij} - \bar{v}_i)^2}}$ and

$$k_a = \frac{1}{\sum_{i=1}^{n} |w(a, i)|}.$$

2.2 The Binary Logistic Regression Approach

In order to predict the preference of the active user u_a for the item j that has not been purchased yet, a predictive model for predicting the class of item j can be expressed based on Eq. (1) as

$$\boldsymbol{v_j} = \boldsymbol{f}(\boldsymbol{v_1}, \cdots, \boldsymbol{v_{j-1}}, \boldsymbol{v_{j+1}}, \cdots, \boldsymbol{v_m}) + \varepsilon,\ j = 1, \cdots, m, \tag{4}$$

where v_j is the jth column of the binary user-item matrix, f is a predictive function for the item j to be predicted and ε is a random error vector. Lee *et al*. (2005) consider a binary logistic regression for the predictive function, where principal components are predictor variables due to high correlation among items. Let $\sum_j$ be the (m-1)×(m-1) covariance matrix of the binary user-item matrix excluding the jth column. The covariance matrix can be decomposed into a product of two matrices as follows (Hastie *et al*., 2001).

$$\textstyle\sum_j = \boldsymbol{PAP}', \tag{5}$$

where Λ is a diagonal matrix whose elements are the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_{m-1}$ of the covariance matrix and

$\mathbf{P} = \left(p_{ij}\right)$ is (m-1)×(m-1) orthogonal matrix whose jth column is the eigenvector corresponding to the jth eigenvalue. Then the principal components are represented as follows.

$$\begin{aligned} \boldsymbol{PC_1} &= p_{11}\boldsymbol{v}_1 + \cdots p_{(j-1)1}\boldsymbol{v}_{\boldsymbol{j}-1} + p_{(j+1)1}\boldsymbol{v}_{(\boldsymbol{j}+1)} + \cdots + p_{m1}\boldsymbol{v_m} \\ \boldsymbol{PC_2} &= p_{21}\boldsymbol{v}_1 + \cdots p_{(j-1)2}\boldsymbol{v}_{\boldsymbol{j}-1} + p_{(j+1)2}\boldsymbol{v}_{(\boldsymbol{j}+1)} + \cdots + p_{m2}\boldsymbol{v_m} \\ &\vdots \\ \boldsymbol{PC_{m-1}} &= p_{(m-1)}\boldsymbol{v}_1 + \cdots p_{(j-1)(m-1)}\boldsymbol{v}_{j-1} + p_{(j+1)(m-1)}\boldsymbol{v}_{(\boldsymbol{j}+1)} \\ &\quad + \cdots + p_{mm}\boldsymbol{v_m}. \end{aligned} \tag{6}$$

The first k principal components can be selected for predictors considering the eigenvalues or prediction performance. The logistic regression model can be represented based on the principal components as follows (Hastie *et al*., 2001).

$$p\left\{\boldsymbol{v_j} = 1\right\} = \frac{\exp(\boldsymbol{\beta pc})}{1 + \exp(\boldsymbol{\beta pc})}, \tag{7}$$

where $\boldsymbol{\beta} = \left(\beta_1, \beta_2, \cdots, \beta_k\right)$ and $\boldsymbol{PC} = \left(\boldsymbol{PC_1}, \boldsymbol{PC_2}, \cdots, \boldsymbol{PC_k}\right)$. Since $p\left\{v_j = 1\right\}$ is considered as purchase probabilities, we can recommend the N items corresponding to the N highest probabilities (Lee *et al*., 2005).

3. PROPOSED SUPERVISED LEARNING APPROACHES

3.1 Random Forest

A random forest is a collection of decision trees based on a re-sampling technique (Breiman, 2001; Hwang and Lee, 2013). Let v_j be the jth column of the binary user-item matrix and $\boldsymbol{V^j}$ be the binary user-item matrix excluding the jth column, where $j = 1, 2, \cdots, m$. A bootstrap sample $\boldsymbol{V_i^j}$ ($i = 1, 2, \cdots, B$) for each tree is drawn from the training data $\boldsymbol{V^j}$. Since each tree $\boldsymbol{v_j} = f_i(\boldsymbol{V_i^j})$ casts equally one vote for predicting the final response, ideal trees for a random forest should be independently and identically distributed. To remove correlations between trees, the bootstrapped sample only has o variables randomly chosen from the m variables ($o \ll m$). Growing a tree can be achieved by a traditional tree algorithm, such as the classification and regression tree (CART). In the CART algorithm, the best split for regression is determined to reduce an impurity measure such as mean squared error, whereas the best split for classification is determined to reduce an impurity measure such as the Gini index and information entropy (Breiman *et al*, 1999). After we repeat the above procedure B times, purchase probabilities are determined by Eq. (8).

$$P\left\{v_j = 1\right\} = \frac{1}{B}\sum_{i=1}^{B}\hat{f}_i(\boldsymbol{V_i^j}). \tag{8}$$

3.2 Elastic Net

We consider the following linear regression model.

$$\boldsymbol{v_j} = \boldsymbol{V^j\beta} + \boldsymbol{\varepsilon}, \tag{9}$$

where $\boldsymbol{v_j} = (v_{1j}, v_{2j}, \cdots, v_{nj})^T \in \mathbb{R}^n$ is a vector of responses, $\boldsymbol{V^j}$ is an $n \times (m-1)$ matrix for predictors, $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_{m-1})^T \in \mathbb{R}^{m-1}$ is a vector for parameters, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)^T \in \mathbb{R}^n$ is an n-dimensional random error, whose components are uncorrelated random variables having mean zero and common variance σ^2. When the predictors are highly correlated, Zou and Hastie (2005) suggest imposing a quadratic penalty in addition to LASSO's one-norm penalty and call the resulting procedure an elastic net (ENET). Also, Friedman *et al*. (2009) propose regularization paths for generalized linear models via coordinate descent. Then the elastic net estimator with smoothing parameter $\alpha > 0$ and $\lambda > 0$ can be defined as

$$\begin{aligned} \hat{\boldsymbol{\beta}} = \arg\min_{\beta}\Bigg\{ & \frac{1}{2n}\left\|\boldsymbol{v_j} - \boldsymbol{V^j\beta}\right\|^2 \\ & + \lambda\left(\alpha\sum_{j=1}^{m-1}\left|\beta_j\right| + \frac{(1-\alpha)}{2}\sum_{j=1}^{m-1}\beta_1^2\right)\Bigg\}. \end{aligned} \tag{10}$$

If α is zero, the ENET becomes the ridge regression (Hoerl and Kennard, 1970), while α is one, the EN becomes the LASSO (Tibshirani, 1996).

4. NUMERICAL EXPERIMENTS

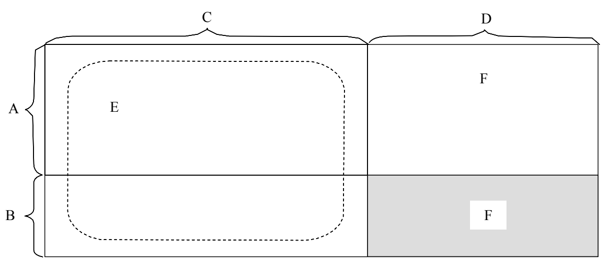
4.1 Experimental Settings

We consider the experimental settings used by Lee *et al*. (2005). To evaluate the proposed supervised learning approaches, the EachMovie data set, available on the website (http://www.research.digital.com/SRC/eachmovie/), is considered. This historical data set consists of 2,811,983 voting scores from 72,916 users on 1628 movies and videos. Although the voting scores are originally on a numeric six-point scale with [0.0, 0.2, 0.4, 0.6, 0.8, 1.0], our proposed scheme requires a binary user-item matrix or a binary item-user matrix and so we change the values of voted cells into ones and the null values of non-voted cells into zeros. For this experiment, 725 users and 257 movies are randomly selected.

The data set prepared for the user-based CF approach is divided into a training set (A) and a test set (B) as illustrated in Figure 1. The training set (A) is utilized to build predictive models. We then consider every active user in the test set for making recommendations. Also, we randomly divide the items into a set of items that treated as predictor variables (C) and a set of items regarded as response variables (D). The information of the four sections in Figure 1 is described in Table 1. The area by dotted line (E) in Figure 1 will be used to calculate similarities between training users and test users in the correlation approach by Eq. (3). The grey shaded area of the data (F) will be first blinded and then used to measure performance of the proposed approaches. The division of the data set is illustrated in Figure 1. On the other hand, the transposed data set for the item-based CF approach is divided into a training set (C) and a test set (D) as illustrated in Figure 2. Also, we divide the customers into a set of customers that treated as predictor variables (A) and a set of customers regarded as response variables (B). The information of the four sections in Figure 2 is described in Table 2. The area by dotted line (E) in Figure 1 will be used to calculate similarities between training items and test items in the correlation approach by Eq. (3). The grey shaded area of the data (F) will be first blinded and then used to measure performance of the proposed approaches. The division of the data set is illustrated in Figure 2. Our performance measure is based on the precision, which is generally used in information retrieval research and defined by

$$\text{Precision} = \frac{\text{hitting number}}{\text{Top} - N} \tag{11}$$

where 'Top-N' is the number of first N items that are recommended by a CF scheme and 'hitting number' is the

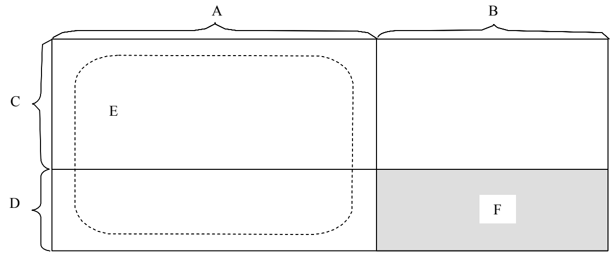


A: Training users
B: Active users
C: Items referred to as independent variables
D: Items referred to as dependent variables
E: Data for calculating the similarities between A and B in the correlation approach
F: Test data set

Figure 1. Division of the experimental data set for the user-based collaborative filtering approach.

Table 1. Summary of each section of data set for the user-based collaborative filtering approach

Section	No. of users	No. of items	Non-zeros	Proportion of ones (%)
A×C	604	207	33972	27.17
A×D	604	50	6519	21.59
B×C	121	207	9234	36.87
B×D	121	50	2046	33.82



A: Users referred to as independent variables
B: Users referred to as dependent variables
C: Training items
D: Active items
E: Data for calculating the similarities between C and D in the correlation approach
F: Test data set

Figure 2. Division of the experimental data set for the item-based collaborative filtering approach.

Table 2. Summary of each section of data set for the item-based collaborative filtering approach

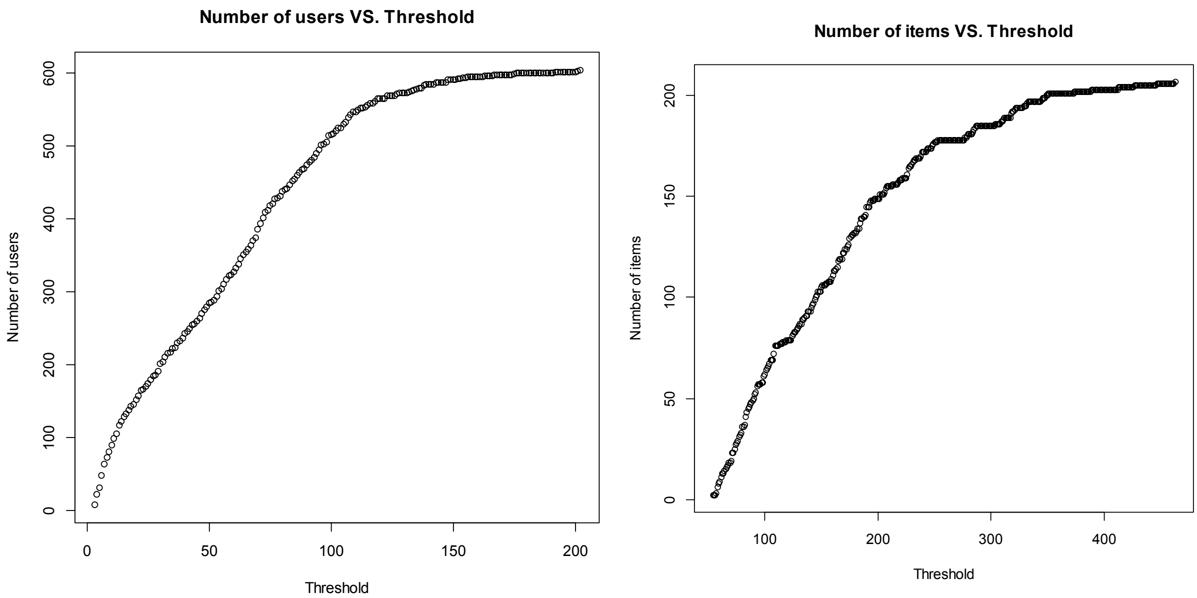
Section	No. of items	No. of users	Non-zeros	Proportion of ones (%)
C×A	207	604	33972	27.17
C×B	207	121	9234	36.87
D×A	50	604	6519	21.59
D×B	50	121	2046	33.82

actual Top-N obtained from the section F. For our performance measure, we will consider the average of the precisions calculated for values of N ranging from 1 to 5.

In the division of the experimental data set for the user-based CF approach, section A×C (training data) is considered for a cold-start problem. For the new user problem, we decrease the number of users by removing users, who have the number of ones greater than or equal to a threshold which ranges from 202 to 3. For the new item problem, we decrease the number of items by removing items, which have the number of ones greater than or equal to a threshold which ranges from 463 to 54. On the other hand, in the division of the experimental data set for the item-based CF approach, section C×A (training data) is likewise considered for a cold-start problem. In accordance with the threshold values, the numbers of users and items are reduced respectively as shown in Figure 3. We consider the user-based CF and item-based CF approaches for the new user and new item problems. Therefore, the four cases: new user and user-based CF, new user and item-based CF, new item and user-based CF, new item and item-based CF are considered for the cold-start problem.

4.2 Results

Table 3 summarizes average precisions for the smallest threshold values, which are 3 for the new user problem and 54 for the new item problem. The smallest



(a) (b)

Figure 3. Numbers of users and items corresponding to threshold values. (a) Number of users vs. threshold. (b) Number of items vs. threshold.

threshold values reflect the most extreme cold-start problems. The correlation approach is not available for the new user and item-based CF and new item and user-based CF cases because the training observations have all the zeros, while random forest classification is not available for the new user and user-based CF and new item and item-based CF cases because the training responses have all the zeros. In order to ensure the best average precisions, the binary logistic regression approach selects the number of principal components equal to 12 for the new item and item-based CF case and equal to 10 for the rest cases. For elastic net, $\alpha = 0.1$ and $\lambda = 0.2$ are selected for the new user and user-based CF case, $\alpha = 0.1$ and $\lambda = 0.25$ for the new user and item-based CF case, $\alpha = 0.05$ and $\lambda = 0.2$ for the new item and user-based CF case, and $\alpha = 0.15$ and $\lambda = 0.2$ for the new item and item-based CF case. Random forest uses the default values of the parameters in R 3.0.1. Despite that, random forest regression performs well for the four cases. It is also interesting that even the correlation approach outperforms the binary logistic regression approach for the new user and user-based CF and new item and item-based CF cases.

4.2.1 New user and user-based CF case

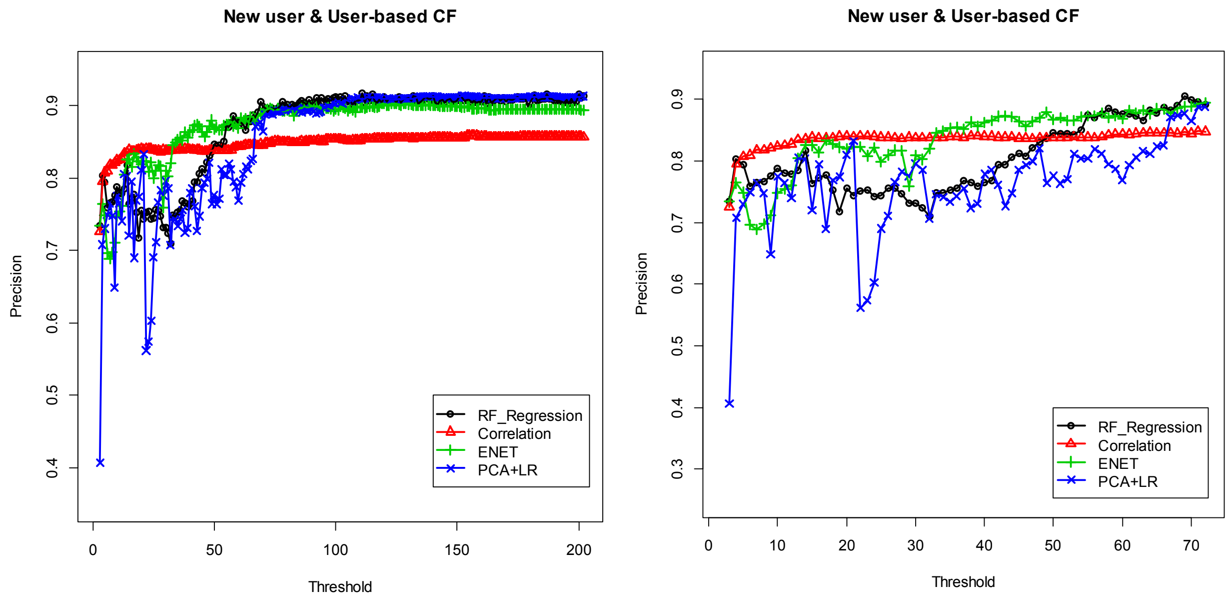
We observe average precisions for all the threshold values in Figure 3(a). For the threshold values below 30, the correlation approach outperforms other approaches as shown in Figure 4(b). When users are sparse, other approaches may not function appropriately because the training observations (users) are not enough for the supervised learning tasks. For the threshold values between 30 and 70, elastic net performs the best. For the threshold values above 70, the correlation approach performs the worst, while other approaches show similar performance as observed in Figure 4(a).

4.2.2 New user and item-based CF case

We observe average precisions for all the threshold values in Figure 3(a). For the threshold values below 70, random forest classification outperforms other approaches as shown in Figure 5(b). Although users are sparse, it may function appropriately because the number of training observations (items) is still fixed at 207. For the threshold values below 50, the binary logistic regression approach is clearly outperformed by the proposed supervised learning approaches. For the threshold values above 70, the approaches show similar performance as

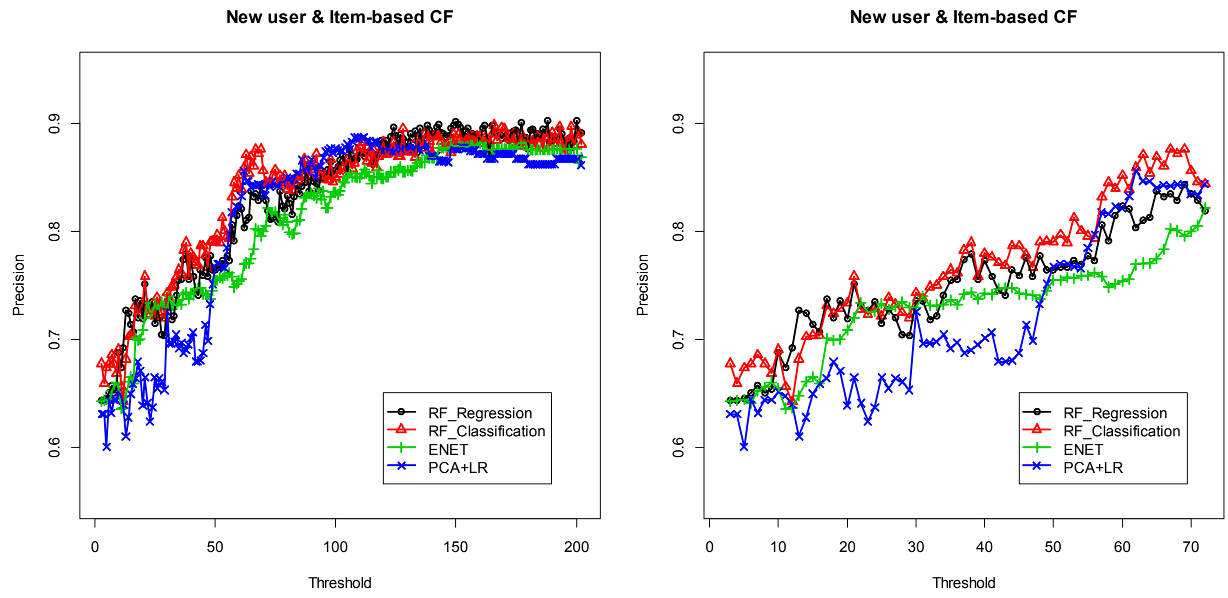
Table 3. Average precisions for the smallest thresholds

	Correlation	PCA+LR	RF_R	RF_C	ENET
New user and user-based CF	0.725	0.407 (10)	0.740	NA	0.740
New user and item-based CF	NA	0.630 (10)	0.643	0.657	0.643
New item and user-based CF	NA	0.830 (10)	0.836	0.795	0.838
New item and item-based CF	0.504	0.399 (12)	0.612	NA	0.498



(a) (b)

Figure 4. New user and user-based CF case. (a) Maximum threshold 202. (b) Maximum threshold 72.



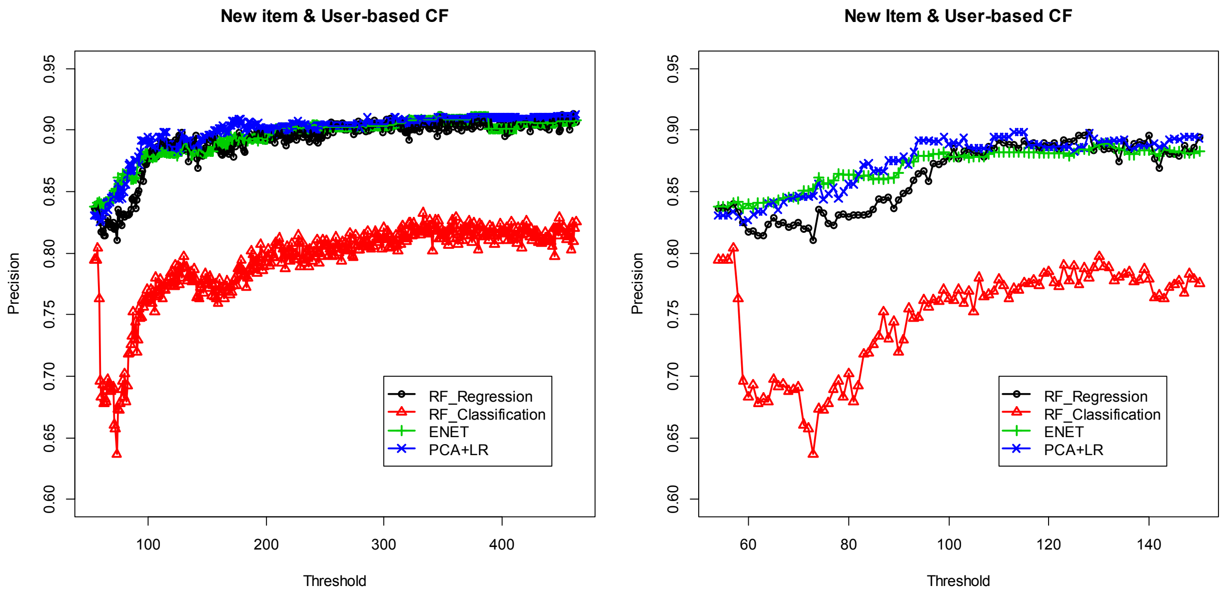
(a) (b)

Figure 5. New user and item-based CF case. (a) Maximum threshold 202. (b) Maximum threshold 72.

shown in Figure 5(a).

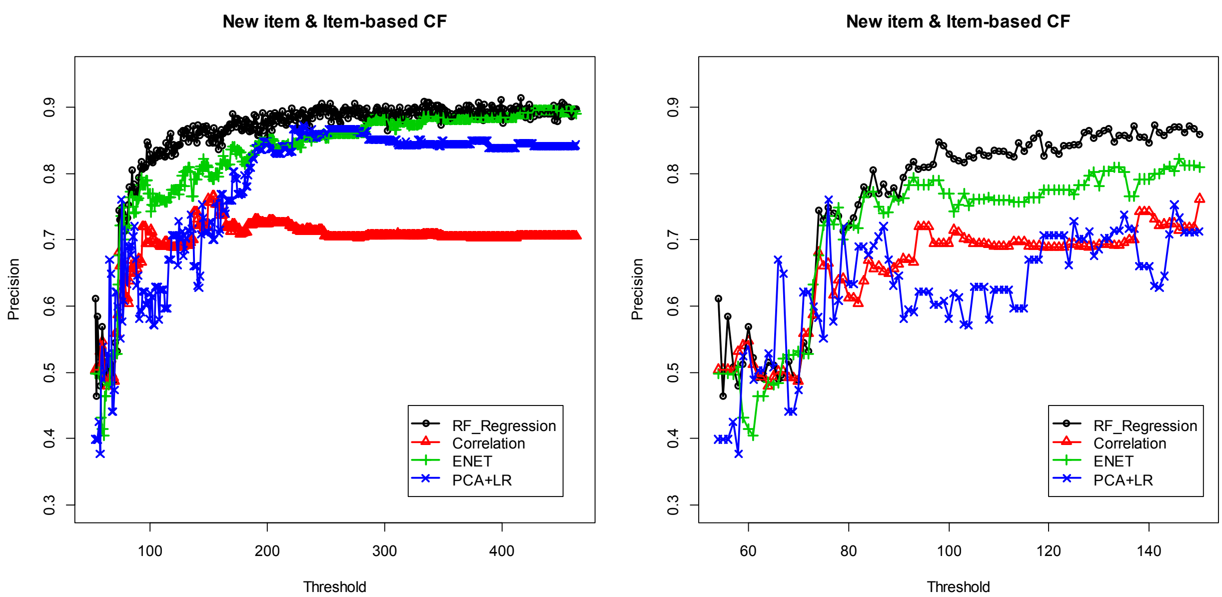
4.2.3 New item and user-based CF case

We observe average precisions for all the threshold values in Figure 3(b). Throughout the threshold values, random forest classification is clearly outperformed by other approaches as shown in Figure 6(a). Random forest classification may not function appropriately because its performance depends on the explanatory variables (items). For the threshold values below 100, the binary logistic regression approach as well as elastic net performs better than random forest regression as shown in Figure 6(b). As a result, we can deduce that random forest is affected more by explanatory variables (items) than observations (users). For the threshold values above 100, the three approaches show similar performance. Most significantly, the binary logistic regression approach is not outperformed by other approaches.



(a) (b)

Figure 6. New item and user-based CF case. (a) Maximum threshold 463. (b) Maximum threshold 150.



(a) (b)

Figure 7. New item and item-based CF case. (a) Maximum threshold 463. (b) Maximum threshold 150.

4.2.4 New item and item-based CF case

We observe average precisions for all the threshold values in Figure 3(b). For the threshold values below 80, the difference between the approaches is not clear where the correlation approach emulates other approaches. In the case, the proposed supervised learning approaches may not function appropriately because the training observations (items) are not enough for the supervised learning tasks. For the threshold values between 80 and 150, random forest regression clearly outperforms other approaches as shown in Figure 7(b). Also, for the threshold values between 90 and 110, even the correlation approach performs better than the binary logistic regression approach.

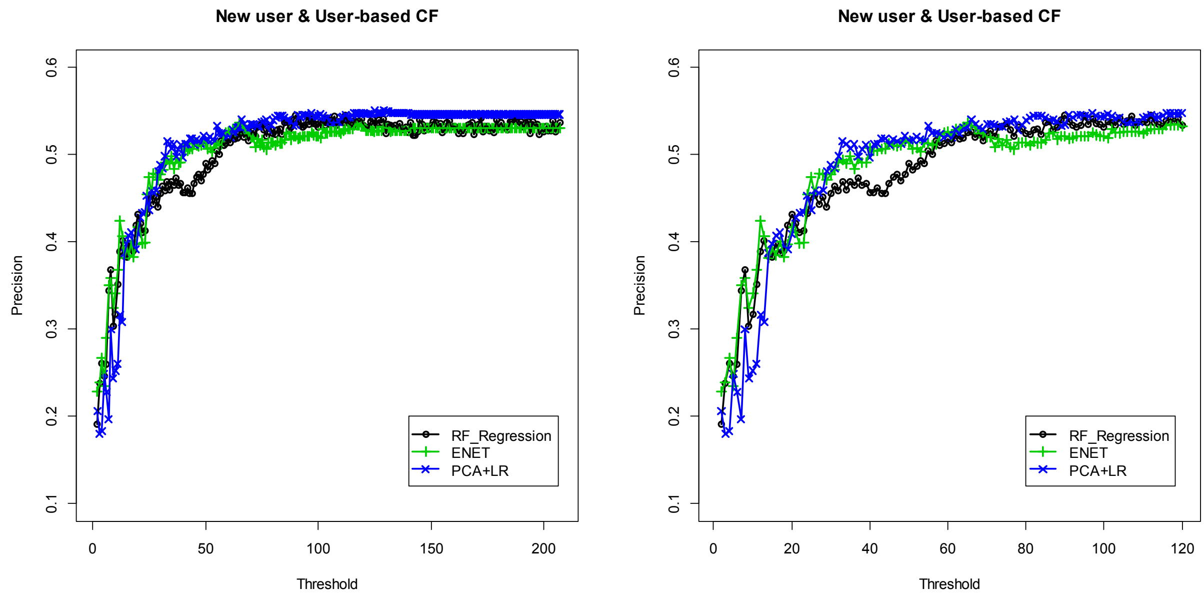
4.3 Analysis for the MovieLenz Data Set

Similar to the EachMovie data set, we analyze the MovieLenz data set, available on the R package, recommenderlab contains about 100,000 ratings (1-5) from

943 users on 1,664 movies (Hahsler, 2014). For this experiment, the first 725 users and 257 movies are selected.

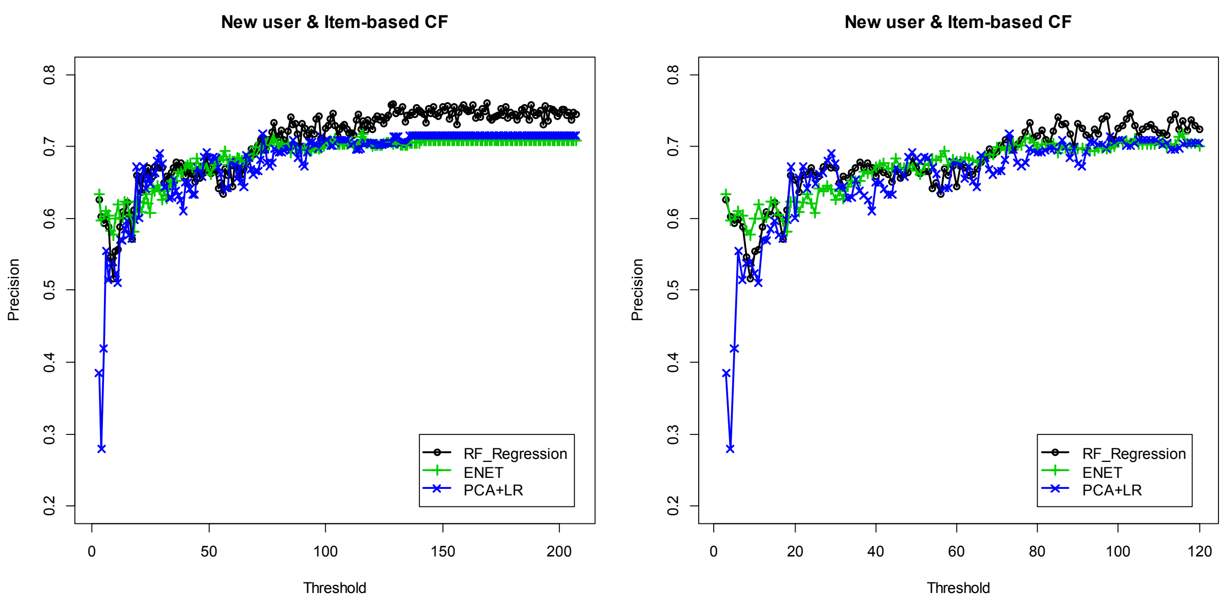
4.3.1 New user and user-based CF case

For the threshold values below 20, the logistic regression approach is outperformed by the proposed approaches as shown in Figure 8(b). When users are extremely sparse, the logistic regression approach may not function appropriately because the training observations (users) are not enough for the supervised learning tasks. For the threshold values above 60, the logistic regression approach slightly outperforms the proposed approaches, which show similar performance as observed in Figure 8(a).



(a) (b)

Figure 8. New user and user-based CF case. (a) Maximum threshold 207. (b) Maximum threshold 120.



(a) (b)

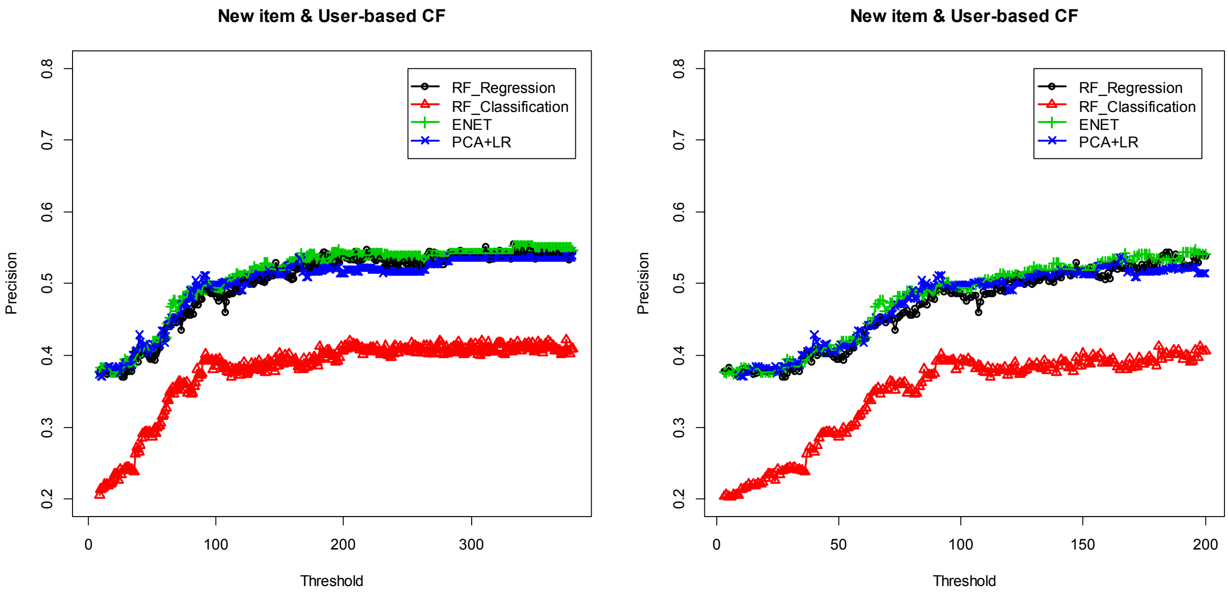
Figure 9. New user and item-based CF case. (a) Maximum threshold 207. (b) Maximum threshold 120.

4.3.2 New user and item-based CF case

For the threshold values above 80, random forest regression outperforms other approaches as shown in Figure 9(a). Although users are not extremely sparse, it may function appropriately. For the threshold values below 20, the binary logistic regression approach is clearly outperformed by the proposed supervised learning approaches.

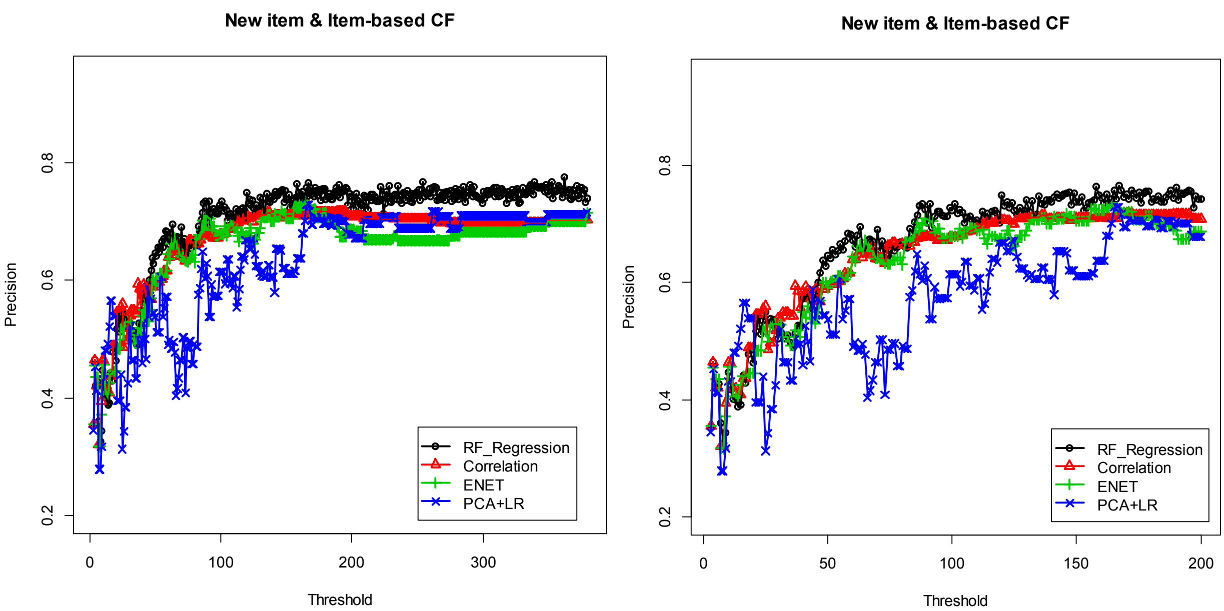
4.3.3 New item and user-based CF case

Throughout the threshold values, random forest classification is clearly outperformed by other approaches as shown in Figure 10(a). Random forest classification may not function appropriately because its performance depends on the explanatory variables (items). For the threshold values above 100, the binary logistic regression approach is slightly outperformed by elastic net.



(a) (b)

Figure 10. New item and user-based CF case. (a) Maximum threshold 379. (b) Maximum threshold 200.



(a) (b)

Figure 11. New item and item-based CF case. (a) Maximum threshold 379. (b) Maximum threshold 200.

4.3.4 New item and item-based CF case

For the threshold values below 150, the logistic regression approach performs the worst, whereas the correlation approach and elastic net shows similar performance as shown in Figure. 11(b). For the threshold values above 50, random forest regression clearly outperforms other approaches as shown in Figure 11(a). For the threshold values above 200, elastic net is outperformed by other approaches.

5. CONCLUSIONS

The market basket data in the form of binary user-item matrix and item-user matrix can be modelled as a binary classification problem or a regression problem. If users or items are sparse, the problem can be considered as a cold-start problem. In order to tackle the cold-start problem, we proposed the three supervised learning approaches: random forest regression, random forest classification, and elastic net.

For the new user and user-based CF case, the correlation approach and elastic net perform the best. For the new user and item-based CF case, random forest classification shows the best performance. For the new item and user-based CF case, the binary logistic regression approach and elastic net outperform other approaches. For the new item and item-based CF case, random forest regression works the best. The correlation approach is not available for the new user and item-based CF and new item and user-based CF cases, while random forest classification is not available for the new user and user-based CF and new item and item-based CF cases. Our research results led us to conclude that the performance and availability of the approaches mainly depend on the characteristics of the training data. Therefore, we need to select a CF approach to understanding the training data and statistical methods. In future research, we can apply the three supervised learning approaches to the market basket data in the form of a binary user-item matrix or a binary item-user matrix across a variety of recommender systems.

ACKNOWLEDGMENTS

The work by Chi-Hyuck Jun was supported by the National Research Foundation of Korea (Grant No. 2013R1A2A2A03068323).

REFERENCES

Ahn, H. J. (2008), A new similarity measure for collaborative filtering to alleviate the new user cold-starting problems, *Information Sciences*, **178**(1), 37-51.

Breese, J. S., Heckerman, D., and Kadie, C. (1998), Empirical analysis of predictive algorithms for collaborative filtering, *Technical Report MSR-TR-98-12*, Microsoft Research, Redmond, WA.

Breiman, L. (2001), Random forests, *Machine Learning*, **45**(1), 5-32.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1999), *Classification and Regression Trees*, CRC Press, New York, NY.

Friedman, J., Hastie, T., and Tibshirani, R. (2009), Regularization paths for generalized linear models via coordinate descent, Department of Statistics, Stanford University, Stanford, CA.

Goldberg, D., Nichols, D., Oki, B., and Terry, D. (1992), Using collaborative filtering to weave an information tapestry, *Communications of the ACM*, **35**(12), 61-70.

Goldberg, K., Roeder, T., Gupta, D., and Perkins, C. (2001), Eigentaste: a constant time collaborative filtering algorithm, *Information Retrieval Journal*, **4**(2), 133-151.

Hahsler, M. (2014), recommenderlab: a framework for developing and testing recommendation algorithms, http://cran.r-project.org/web/packages/recommenderlab/vignettes/recommenderlab.pdf.

Hastie, T., Tibsharani, R., and Friedman, J. (2001), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York, NY.

Hill, W., Stead, L., Rosenstein, M., and Furnas, G. (1995), Recommending and evaluating choices in a virtual community of use, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Denver, CO, 194-201.

Hoerl, A. E. and Kennard, R. W. (1970), Ridge regression: biased estimation for nonorthogonal problems, *Technomerics*, **12**(1), 55-67.

Hwang, W. Y. and Lee, J. S. (2013), Shifting artificial data to detect system failures, *International Transactions in Operational Research*, Advanced online publication, doi: 10.1111/itor.12047.

Lee, C. H., Kim, Y. H., and Rhee, P. K. (2001), Web personalization expert with combining collaborative fil-tering and association rule mining technique, *Expert Systems with Applications*, **21**(3), 131-137.

Lee, J. S. and Olafsson, S. (2009), Two-way cooperative prediction for collaborative filtering recommendations, *Expert Systems with Applications*, **36**(3), 5353-5361.

Lee, J. S., Jun, C. H., Lee, J. W., and Kim, S. Y. (2005), Classification-based collaborative filtering using market basket data, *Expert Systems with Applications*, **29**(3), 700-704.

Leung, C. W., Chan, S. C., and Chung, F. (2008), An empirical study of a cross-level association rule mining approach to cold-start recommendations, *Knowledge-Based Systems*, **21**(7), 515-529.

Lika, B., Kholomvatsos, K., and Hadjiefthymiades, S. (2014), Facing the cold start problem in recommender systems, *Expert Systems with Applications*, **41**(4), 2065-2073.

Mild, A. and Reutterer, T. (2001), Collaborative filtering methods for binary market basket data analysis, *Active Media Technology, Lecture Notes in Computer Science*, **2252**, 302-313.

Mild, A. and Reutterer, T. (2003), An improved collaborative filtering approach for predicting cross-category purchase based on binary market basket data, *Journal of Retailing and Consumer Services*, **10**(3), 123-133.

Park, D. H., Kim, H. K., Choi, I. Y., and Kim, J. K. (2012), A literature review and classification of recommender systems research, *Expert Systems with Applications*, **39**(11), 10059-10072.

Park, S. T. and Chu, W. (2009), Pairwise preference regression for cold-start recommendation, *Proceedings of the third ACM Conference on Recommender Systems (RecSys2009)*, New York, NY, 21-28.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994), GroupLens: an open architecture for collaborative filtering of netnews, *Proceedings of the ACM Conference on Computer Supported Cooperative (CSCW1994)*, Chapel Hill, NC, 175-186.

Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2001), Item-based collaborative filtering recommendation algorithms, *Proceedings of the 10th international World Wide Web Conference (WWW10)*, Hong Kong, 285-295.

Schein, A., Popescul A., and Ungar, L. H. (2002), Methods and metrics for cold-start recommendations, *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Tampere, Finland, 253-260.

Shardanand, U. and Maes, P. (1995), Social information filtering: algorithms for automating word of mouth, *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI1995)*, Vancouver, Canada, 210-217.

Tibshirani, R. (1996), Regression shrinkage and selection via the lasso, *Journal of Royal Statistical Society Series B: Methodological*, 58(**1**), 267-288.

Zou, H. and Hastie, T. (2005), Regularization and variable selection via the elastic net, *Journal of Royal Statistical Society Series B: Statistical Methodology*, **67**(2), 301-320.