# Supervised Learning in Matrix Completion Framework for Recommender System Design

Anupriya Gogna, Angshul Majumdar

IIIT-Delhi, Delhi
INDIA
anupriyag@iiit.ac.in, angshul@iiitd.ac.in

## Abstract

Recommender systems primarily utilize the, highly sparse, explicit rating information to make relevant predictions. This data scarcity places a limit on the accuracy of prediction. In this work we attempt to alleviate the problem of data sparsity by using secondary information. Most existing works incorporate auxiliary information in a (bi-linear) matrix factorization setup; whilst our model is based on a (convex) matrix completion framework. In this work, we use auxiliary information about users and items to impose additional constraints on the recovered rating values; adopting ideas from supervised learning. Alongside, we also propose a method to utilize the information map extracted from supervised learning approach to handle the cold start problem. Most works that address the cold start problem are focused on users with very few ratings - this is not the pure cold-start problem. However, in this work we target new users and items which have no ratings available for them; and only has the associated metadata. We propose an algorithm using split Bregman technique for solving our formulations. Comparison of our design with existing state of the art methods for RS design on the movie recommender systems clearly indicate the superiority of our formulation over existing methods.


COMAD, March 11-13, 2016, Pune.


## 1. Introduction

Today recommender systems (RS) [1,2] are the workhorse behind all Business-to-Client eCommerce portals. To facilitate the user, a recommender system predicts the user's choices and suggests a handful of items; if the prediction is good the user buys it. The importance of accurate recommendation and hence the focus on building efficient RS is very clear - better the prediction, more is the revenue for the portal.

RS largely rely on some form of feedback provided by users on a subset of items, such as purchase information, like/dislike options or explicit rating data, to predict the ratings on yet unrated items. Gathering this information involves a user's active participation, either by means of purchase or some form of interviewing process (like seeking user's rating on a selected set of items), which is not always a plausible scenario. Lack of this preference information, especially in case of new users registering on the system can be major bottleneck in improving customer satisfaction. It is essential for RS to provide satisfactory suggestions to such (new) users as well, failing in which can cause potential loss of customers and revenue.

In absence of any explicit predilection information, the rating prediction for new users (user cold start problem) can be based on available secondary data like user's demographics. Consider for example distribution based on age grouping; children in age group of 1-10 will most likely have affinity for animation movies; similarly, young adults (say 20-30 years) can have affinity for action/thriller. Similarly, women may have in general affinity for rom-com or family genres whereas males might be more inclined towards action. On similar lines, metadata for new items (such as their category information) can be used to gauge user's interest in them; thereby solving the item cold start problem. For example, a user who liked comedies in past will most likely enjoy comic recommendations. Thus, auxiliary data can prove to be a valuable source of information in RS; idea being exploited in several works [3,4]. Despite the difficulty in garnering collaborative information for new users or items, most existing works [5,6] handling the cold start problem work with users and/or items which have small number of ratings available for them i.e. solve the partial cold start problem. Several works rely on building interviewing process [7,8] to collect (new) user's ratings on few selected items, which might not be convenient in all scenarios outside the academia. For example, e-retailers such as amazon or alibaba does not gather such information from new users and sites garnering such information are becoming increasingly rare.

The metadata used for prediction to solve the cold start problem can also be used to augment the rating dataset for warm start users. The explicit rating data (from users) is much more reliable than implicitly gathered information but suffer from extreme data sparsity. To alleviate this data sparsity, user/item auxiliary information can be exploited.

Traditionally, collaborative filtering (CF) [9,10] techniques have been used as de-facto approach to harness the explicit rating data for rating prediction. In recent past, researchers have proposed models based on CF schemes to assimilate user/ item metadata as well. This additional data has been used to augment the explicit rating data in either a memory based setup [4,11] or in latent factor framework [12,13].

Neighborhood based models [14] although easy to implement, do not always yield the best of results [15]; latent factor models [16] being more powerful. These models assume that user's choices on items are determined by very few factors. The user has an affinity towards these factors whereas the items possesses these factors to a lesser or greater extent. Thus both the users and items can be characterized as vectors of latent factors; user's rating on an item expressed as an inner product between the user and item latent factor vectors.

In light of the arguments presented above, in this work, we aim to use auxiliary data in a latent factor framework to improve prediction accuracy for both warm start and pure cold start scenarios. The highlight of our approach is that the proposed method to solve the cold start problem for new users/items is a direct increment of the model proposed for improving prediction accuracy for existing users; thereby handling both the major problems without increased resource requirement or complexity. Such a comprehensive model has not been proposed.

Another highlight of our model is use of matrix completion formulation, instead of more commonly employed matrix factorization (MF) [17], which attempts to recover the rating matrix as a product of two matrices – user's latent factor and item's latent factor matrix. Matrix factorization is computationally fast, but unfortunately it is bilinear and hence non-convex. Recently, researchers in signal processing showed that, instead of formulating the latent factor model as a matrix factorization problem, it can be recast as a low-rank matrix completion problem (LRMC) - a convex formulation [18,19]. As discussed above, latent factor model assumes that an item's rating is a function of a handful of features (latent factors). As, the entire rating matrix is a result of interaction amongst the latent factor vectors of users and items, its structure is governed by the small number of factors only. This results in the low rank nature of the rating matrix enabling use of LRMC techniques for rating prediction. We formulate our proposition as an augmented matrix completion problem (with additional regularization terms) – which enjoys the benefit of a convex formulation, deriving ideas from supervised learning.

In addition, unlike most recent works, which use user's social profile or trust network as additional data source [13], we use user's demography and item category information to supplement the rating database. It is difficult for RS to acquire social relation data for user; limiting the applicability of models using the same. Most RS maintain a database of item genre/category (for example an online book store will always have books categorized as per genre). Also, usually users are required to fill up some basic information (like age, gender etc.) while registering on an online portal. Hence, this information is readily available and at no extra cost, enabling a wider applicability of our model.

The novelty of our approach lies in the use of easily and widely available data (user demography and item genres) in a supervised learning environment to generate effective recommendations. We group together (label) users based on their demographic information – age, gender and occupation. The rating prediction is done under the additional constraint of maintaining label consistency. Similar strategy is adopted for items as well by using the genres as classification labels. Use of additional information (as constraints) to augment the matrix completion model reduces the problem search (solution) space, making the problem less underdetermined. There are few works [4,12] that incorporate demographic data of users, however none of them follow the principles of supervised learning followed in our work. Also, as indicated in results section; ours is a far superior formulation.

We extend our supervised learning based model to mitigate the cold start problem as well. The label consistency model is used to derive a relation between a user's and/or item's class labeling and their rating pattern. This information is used to predict the ratings and make effective recommendations for new users (or new items) for which secondary information is available. When a new user enters a system, their record is updated and so is the case with a new item (say movie) made available at an online portal. Thus, in absence of any rating data for such cold start conditions, this information makes effective recommendations plausible. We also design an algorithm based on split Bregman technique for our formulations.

## 2. Related Work

### 2.1 Matrix Factorization Framework for Latent Factor Model

The explicit rating provided by a user ( $R_{i,j}$ , user *i* on item *j*) can be viewed as a combination of two factors – baseline estimate and interaction component. The baseline constitutes user and item biases. There are some users who are overtly critical and tend to rate everything on the lower side of the scale – they have negative bias; similarly, there are some movies which are always rated on the higher side – they have positive bias. The 'interaction' part models a user's affinity for an item.

Usually baseline is computed offline by solving (2) via stochastic gradient descent algorithm [17].

$$\min_{b_i,b_j} \sum_{i,j\in avail\_rat} \left\|R_{i,j} - b_j - b_i - \mu\right\|_2^2 + \delta\left(\left\|b_j\right\|_2^2 + \left\|b_i\right\|_2^2\right) \quad (2)$$

where, $\mu$ is the global mean; $b_i$ is $i^{th}$ the user bias and $b_j$ is the item bias of $j^{th}$ item; $\left(\mu + b_i + b_j\right)$ is the baseline component; $\delta$ is the regularization parameter.

The interaction (*Y*) between the user and the item $\left(Y_{i,j} = R_{i,j} - b_j - b_i - \mu\right)$ is modelled in terms of latent factors. Consider the case of movie ratings; choice of a movie is determined by very few factors - genre, director, cast, music etc. Each movie possesses these factors to a certain extent, and each user has affinity towards these factors. Based on this model one can represent a user (*i*) by a vector $U_i$ and an item (*j*) by a vector $V_j$ corresponding to latent factors. The 'interaction' can hence be expressed as inner product of two $\left\langle U_i, V_j \right\rangle$.

The problem in CF is that all the user ratings are not available; a typical user will only rate a small percentage of all the items. Thus, if we consider the interaction matrix (Y), it is incomplete. The problem in CF is to predict all the missing ratings - i.e. fill in the rating matrix. This can be expressed as an inverse problem [14]; $Y = M \odot (UV)$, where $M$ is a binary mask having 1's in place of available ratings and 0 elsewhere.

This problem is solved via the following optimization:

$$\min_{U,V} \left\|Y - M \odot (UV)\right\|_F^2 + \lambda\left(\left\|U\right\|_F^2 + \left\|V\right\|_F^2\right) \quad (3)$$

This problem is non-convex in *U* and *V*, owing to the bi-linearity. Thus there is no convergence guarantee.

### 2.2 Matrix Completion

If, we consider all the users and the items, the interaction matrix will be represented as $Z = UV$ ; Z $\left(Z \in \mathbb{R}^{K\times N}\right)$ is complete interaction matrix with *K* users and *N* items.

Traditionally latent factor models formulated the interaction component as a matrix factorization problem. However, if we concentrate on rating prediction (only the interaction *Z*), we do not need to solve for the user (*U*) and the item (*V*) factor matrices separately, as long as we can estimate Z. Recent studies proposed estimating Z directly, by solving the inverse problem $Y = M \odot Z$.

This is an under-determined inverse problem with infinitely many solutions. In order to find a reasonable solution, one needs some prior assumption regarding *Z*. Even though *Z* is a very large matrix (hundreds of thousands of users and items), it has a very low-rank; the rank being the same as the number of latent factors. Thus predicting the missing interactions turns out to be a Matrix Completion problem (4)

$$\min_{Z} \left\|Y - M \odot Z\right\|_F^2 + \lambda \left\|Z\right\|_* \quad (4)$$

The nuclear norm penalty promotes a low-rank solution [20]. In this section, we review few LRMC algorithms briefly.

Toh, & Yun [21] proposed Accelerated Proximal Gradient (APG) algorithm for LRMC. It employs Proximal Gradient (PG) [22] method with an appropriate step size and an extra interpolation step to achieve faster convergence. The iterative algorithm can be summarized as follows

$$\begin{aligned} W^k &= X^k + \frac{t^{k-1}-1}{t^k}\left(X^k - X^{k-1}\right) \\ G^k &= W^k - \left(\tau^k\right)^{-1} A^T\left(A\left(W^k\right) - b\right) \\ X^{k+1} &= S_{\tau^k}\left(G^k\right); t^{k+1} = \frac{1+\sqrt{1+4\left(\tau^k\right)^2}}{2} \\ s.t.\ b &= vec\left(Y\right); A : block\ diag\ \ form\, of M \end{aligned} \quad (5)$$

Authors in [23] proposed a method for low-rank matrix recovery using the Iterative Least Square (IRLS) technique. It aims at minimizing the weighted Frobenius norm, $\left\|W_p^{(1/2)} X\right\|_F^2$ of matrix, *X*. A low rank matrix (*X*) results if weighting matrix $W_p$ is chosen appropriately. IRLS algorithm for nuclear norm minimization consists of following iterates

$$\begin{aligned} X^k &= \arg\min\left\{Tr\left(W_p^{\,k-1} X^T X\right) : A\left(X\right) = b\right\} \\ W_p^{\,k} &= \left(\left(X^k\right)^T\left(X^k\right) + \gamma^k I\right)^{p/2 - 1} \end{aligned} \quad (6)$$

Most of the existing methods for LRMC require large number of iterations for convergence on large datasets. We propose an algorithm for our augmented matrix completion formulation based on split Bregman technique [24]. Use of split Bregman helps achieve faster convergence and improved recovery accuracy.

### 2.3 Use of Auxiliary Information

To augment the (sparse) explicit rating dataset several researchers have utilized available secondary data. In this section we review some of the techniques for the same.

Authors in [25] proposed a similarity measure $\left(sim_{mod}\right)$ to determine nearest neighbours based on both rating data and demographic information (7).

$$sim_{mod} = sim_{dem} \times sim_{rat} + sim_{rat} \tag{7}$$

where, $sim_{dem}$ is similarity computed using demographics and $sim_{rat}$ is computed using explicit rating data.

Rating data is augmented with geo-spatial information, in a neighbourhood based model, for photograph recommendation in [26]. They used geographical tag data to group photographs into clusters and propagate ratings amongst the members of the same cluster. Thus, a dense rating matrix is obtained which is used as input to neighbourhood based CF algorithm.

Authors in [12] used graph regularization to augment the matrix factorization model. User and item graphs were constructed by utilizing user's demographic and social profile data and item's genre classification.

$$\begin{aligned} &\min_{U,V} \left\| Y - M \odot (UV) \right\|_F^2 + \lambda \left( Tr\left( U^T G_{lu} U \right) \right) \\ &+ \gamma \left( Tr\left( V^T G_{lv} V \right) \right) \end{aligned} \tag{8}$$

where, $G_{lu}$ and $G_{lv}$ are the graph Laplacians for user and item graphs respectively.

In [27] social network information and ratings are used in a PMF (Probabilistic Matrix Factorization) framework. Standard PMF models latent factor vectors as independent Gaussian priors. In [12] PMF is modified to allow for correlation between these Gaussian priors, incorporating similarity amongst items/users.

Most Existing works, as discussed above, augment the conventional matrix factorization framework with secondary data. Also, they mainly rely on grouping of users and/or items and promoting similarity amongst latent factor vector of similar (grouped) users and/or items. Though, authors in [28] augmented matrix completion model, their model is also based on grouping together similar users. They minimized the rating variation amongst similar users (9). They do not exploit item metadata in their framework

$$\min_{Z} \left\| Y - A(Z) \right\|_F^2 + \lambda \left\| Z \right\|_* + \sum_{G \in Groups} \left( \mu_G \sum_{g \in G} \mathrm{var}_g \left( R \right) \right) \tag{9}$$

In this work, we build up on the (convex) matrix completion model incorporating user/item metadata (demographic information and item categorization) in a label consistent (supervised learning) framework. Also, our formulation can exploit both item and user metadata. Use of label consistency model helps us derive linear maps from rating space to item/user label domain. This assists in solving the rating prediction problem for new users and new items (cold start). None of prior art targets both warm and cold start users together.

### 2.4 Cold Start Poblem

The problem of providing effective suggestions to new users or recommending new items to existing users – the cold start problem, is a big challenge in RS design. We review some of prior art in the area.

In [29] used a trust based measure to determine similar users instead of rating based similarity for cases where very few ratings are available. They argued that because trust propagates, there can be many more similar users than if (very few) ratings are considered, making predictions better. Authors in [6] used social tags as a means of relating users to items. The predictions are based on the frequency of tags and the semantic relationships between tags and items.

Works like [30] use small amount of rating information alone to target partial cold start problem. They based their predictions on a new similarity measure that also consider the frequency and count of co-rated items to remove disparity between users with highly varied rating patterns.

Authors in [31] used user's demographics to model an alpha-community space model. Once a new user's communities are defined, one recommendation list per community is generated based on adhoc level of agreement recommendation process.

Most works, as highlighted above, solve the cold start problem for cases where some rating information is available. We, in this work attempt to solve the pure cold start problem. Also, unlike existing methods which attempt to separately solve the cold start problem, our framework is a cohesive model aiming for improvements in accuracy for existing users and mitigating the cold start problem.

## 3. Proposed Formulation

In this section, we describe our proposed formulation for design of a RS incorporating user-item metadata to improve prediction accuracy. The design is also extended to solve the pure cold start problem. The novelty of our work lies in formulating a matrix completion based model for exploiting user (demographic profile) metadata and item categories along with the ratings. We augment the LRMC model with label consistent constraints, derived from user/item metadata, imposed on the rating matrix. Also, the highlight of our design is that we put forth a comprehensive model to handle two major problems afflicting the RS – improving quality of prediction and the cold start problem.

### 3.1 Problem Formulation

#### *3.1.1 Low Rank nature of rating matrix*

As discussed above, we perform offline baseline estimation and work with interaction component alone. Once the complete interaction matrix ($Z$) is recovered (using proposed formulation) the baseline estimates are added back.

Latent factor model states that the interaction between users and items is governed by a small number of factors – the latent factors; say, for books the latent factors may be author and genre; for movies director, genre, cast etc. As the interaction matrix is a function of very few variables (~40-50) as compared to matrix dimensions (hundreds of thousands of users and items), the matrix is fairly low rank. The low-rank property of $Z$ can be used to predict the missing ratings using LRMC framework. Thus predicting the missing interactions turns out to be a Matrix Completion problem (10).

$$\min_{Z} \|Y - A(Z)\|_F^2 + \lambda \|Z\|_* \tag{10}$$

where, $A$ is a binary mask, which is 1's in place of available rating values and *0* otherwise; $Z$ is the completely filled matrix of interaction component; Y interaction component of available ratings.

### 3.1.2 Incorporating Metadata

Nuclear norm minimization (10) requires that for a rank $r$ matrix of size $n \times n$, at least $(6n-5r)r$ samples be available [20]. For the case of RS design, size of matrix is at least $1000 \times 1000$, thereby requiring around 23% of the ratings to be available for reasonable reconstruction accuracy (assuming rank to be 40). However, in real world datasets, the available information is less than 10%, in some cases even as low as 1%.

Hence, there is considerable need for additional information, which can alleviate data sparsity to improve prediction accuracy. In this paper, we make use of user's demographic data and item genre information to augment the rating data for a movie recommender system. Often, during the process of sign up users are required to enter their basic demographic data. Also, all portals maintain a database of their item categories. Thus, collecting this information invites no additional cost. Even for new users and new items, this metadata is readily available; even if collaborative information is missing.

Our model utilizes a label information data (matrix) defining relations between users and/or items and the class they belong to. For users, classes are defined on the basis of age, gender and their occupation; for items, multiple genres form the distinct classes. Our framework can make use of any additional available information as well for classification purpose. We incorporate label data into the matrix completion framework by modifying (10) to include additional label consistent regularization terms.

Considering user metadata, we define multiple classes based on gender, age brackets and different occupational profiles; user can simultaneously belong to multiple classes. Using this label information a user-class label matrix $(L_u)$ is defined, such that $L_u(i,c)=1$ if user $i$ belongs to class $c$ else 0. Let us consider an example wherein we form 2 distinct gender (M/F) groups, P distinct non-overlapping age groups (say 1-17, 18-24 and so on) and Q distinct occupational categories. The label matrix $(L_u)$ will have a row corresponding to each user and columns corresponding to $(2+P+Q)$ classes as shown in fig. 1. Let us consider a user (User 1), who is a male in age group of 18-24 and a lawyer by profession. The classification information of this user can be used to fill up first row of $L_u$. Similarly, for a female in age group of 60+ and an artist by profession, corresponding row will be as shown in row 2 and so on.

This class label matrix provides additional data to help predict the missing values in the rating matrix. The ratings are predicted under the add-on constraint of maintaining label consistency (appended as a regularization term) as

$$\min_{Z,W_u} \|Y - A(Z)\|_F^2 + \lambda \|Z\|_* + \lambda_u \|L_u - ZW_u\|_F^2 \tag{11}$$

where, $W_u$ is the linear map from user-item rating space to user-class space. It defines relation between a user's class and their ratings; $\lambda_u$ is the regularization parameter governing the relative importance given to rating data and the demographic information.

Similar model is built for items as well; establishing a relation between the item genre and the ratings given to them by users. Each item (movie, in this case) may belong to several classes (genres). A class-item label matrix $(L_v)$ is constructed such that $L_v(c,j)=1$ if item $j$ belongs to class $c$ else 0. Similar to formulation discussed in (12) we propose item metadata based framework (12)

$$\min_{Z,W_v} \|Y - A(Z)\|_F^2 + \lambda \|Z\|_* + \lambda_v \|L_v - W_v Z\|_F^2 \tag{12}$$

where, $W_v$ is the linear map (to be estimated) from user-item rating space to class-item space and $\lambda_v$ is the

| | Gender 1 (M) | Gender 2 (F) | Age 1 (1-17) | Age 2 (18-24) | … | Age P (60+) | Occ. 1 (Tech.) | Occ. 2 (Artist) | … | Occ. Q (Lawyer) |
|---|---|---|---|---|---|---|---|---|---|---|
| User 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| User 2 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| .. | .. | .. | .. | .. | .. | .. | .. | .. | .. | .. |
| User \|U\| | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |

**Figure 1. Construction of label matrix**

regularization parameter.

We also club together both formulations to exploit both item and user metadata simultaneously as shown in (13).

$$\min_{Z,W_v,W_u} \left\|Y-A(Z)\right\|_F^2+\lambda\left\|Z\right\|_* + \lambda_v\left\|L_v-W_vZ\right\|_F^2+\lambda_u\left\|L_u-ZW_u\right\|_F^2 \tag{13}$$

Equation (13) illustrates our final formulation for supplementing the matrix completion model with item and user metadata. Use of additional information helps improve the robustness and accuracy of our recommender system by making the problem less underdetermined.

### 3.1.3 Alleviating Cold Start Problem

For new users or items there are no ratings; making rating prediction a challenge. We propose to use the information map ($W_v$ and $W_u$) extracted from solving (13), almost as a by-product, to solve the pure cold start problem.

First let us consider, the information map $W_u$ i.e. one generated using user metadata. It is a map from rating information to user label (classification) space. The map primarily correlates the ratings or user's choice with the demographic profile of a user. Consider a new user $U_{new}$ entering a system. As he/she signs up on the portal, their demographic information is captured. Thus, a vector $(U_{coldstart})$ defining class labelling of the said user $\left(\begin{bmatrix}U_{new-c_1} & U_{new-c_2} & . & U_{new-c_{cu}}\end{bmatrix}\right)$ can be constructed, where *cu* is the number of classes considered for users. From solution top (13), we have the deciphered map $W_u$. The new user's demographic information (label vector) and the deciphered map can be related as

$$\begin{bmatrix}U_{new-c_1} & . & . & U_{new-c_{cu}}\end{bmatrix}=\begin{bmatrix}Z_{new-i_1} & . & . & Z_{new-i_N}\end{bmatrix} \times\begin{bmatrix}W_{u(11)} & W_{u(12)} & W_{u(13)} & . \\ W_{u(21)} & W_{u(22)} & . & . \\ . & . & . & . \\ . & . & . & W_{u(i_N c_{cu})}\end{bmatrix} \tag{14}$$

where, $\left(Z_{coldstart}=\begin{bmatrix}Z_{new-i_1} & . & . & Z_{new-i_N}\end{bmatrix}\right)$ is the vector defining the new user's rating (interaction components) for each item in the database (total number of items, *N*). Equation (14) can be written as set of linear equation (15)

$$U_{coldstart}=Z_{coldstart}W_u \tag{15}$$

Predicted interaction part for new user, $Z_{coldstart}$, can be obtained by solving (15) using any conjugate gradient type algorithm.

Similar approach can be followed for item cold start problem as well by utilizing the genre information of new item $(V_{new})$ and the information map $W_v$. As a new item (say movie in our case) is added to the system, its genre information is easily available. The information map, $W_v$, establishes a relation between the rating data and the genre of items i.e. it captures information relating user's choice of an item to its genre content. This information map is used to determine user's preference for a new item.

Similar to equation constructed above for users, we can formulate item cold start problem as

$$\begin{bmatrix}V_{new-c_1} \\ V_{new-c_2} \\ . \\ V_{new-c_{cv}}\end{bmatrix}=\begin{bmatrix}W_{v(11)} & W_{v(12)} & . & . \\ W_{v(21)} & . & . & . \\ . & . & . & . \\ . & . & . & W_{v(c_{cv}u_K)}\end{bmatrix}\begin{bmatrix}Z_{new-u_1} \\ Z_{new-u_2} \\ . \\ Z_{new-u_K}\end{bmatrix} \tag{16}$$

where, $V_{coldstart}=\begin{bmatrix}V_{new-c_1} & . & . & V_{new-c_{cv}}\end{bmatrix}^T$ is the class label vector for the new item (*cv*: number of distinct classes); $Z_{coldstart}=\begin{bmatrix}Z_{new-u_1} & . & . & Z_{new-u_K}\end{bmatrix}^T$ defines interaction component of ratings by all existing users for new item.

Equation (16) can be compactly written as in (17) and solved using a conjugate gradient solver.

$$V_{coldstart}=W_vZ_{coldstart} \tag{17}$$

Hence, our model can be used to mitigate both user and item end (pure) cold start problem, as an extension of our label consistent model, without significant computational burden.

## 3.2 Algorithm Design

In this section, we present the algorithm for our proposed formulation (13) using split Bregman technique.

Use of split Bregman technique [24] aids in faster convergence and lower recovery errors, as no cooling of regularization parameter is required and thus optimal values of regularization parameters for each of the sub problem can be set.

Firstly, in order to enable splitting of multiple norm terms, we introduce proxy variables (*P* and *Q*) in our formulation (13) as in (18).

$$\min_{Z,W_v,W_v,P,Q}\left\|Y-A(Z)\right\|_F^2+\lambda\left\|Z\right\|_*+\lambda_v\left\|L_v-W_vP\right\|_F^2+ \lambda_u\left\|L_u-QW_u\right\|_F^2+\mu_v\left\|P-Z-B1\right\|_F^2+\mu_u\left\|Q-Z-B2\right\|_F^2 \tag{18}$$

where, $B1$ and $B2$ are the Bregman variables.

Use of Bregman variables ensures that the equality between original and proxy variables need not be strictly enforced from the start. Updation of Bregman variables helps add back the error thus making the algorithm self-correcting and also helps in faster convergence.

We split our formulation into simpler sub problems using Alternating Direction method of Multipliers.

Sub Problem 1

$$\min_{Z}\left\|Y-A(Z)\right\|_F^2+\lambda\left\|Z\right\|_* +\mu_v\left\|P-Z-B1\right\|_F^2+\mu_u\left\|Q-Z-B2\right\|_F^2 \tag{19}$$

Sub Problem 2

$$\min_{P}\lambda_v\left\|L_v-W_vP\right\|_F^2+\mu_v\left\|P-X-B1\right\|_F^2 \tag{20}$$

Sub Problem 3

$$\min_{Q} \lambda_u \|L_u - QW_u\|_F^2 + \mu_u \|Q - X - B2\|_F^2 \quad (21)$$

Sub Problem 4

$$\min_{W_v} \|L_v - W_v P\|_F^2 \quad (22)$$

Sub Problem 5

$$\min_{W_u} \|L_u - QW_u\|_F^2 \quad (23)$$

Now, focusing on sub problem 1, it can be recast as

$$\min_{Z} \left\| \begin{pmatrix} Y \\ \sqrt{\mu_v}(P - B1) \\ \sqrt{\mu_u}(Q - B2) \end{pmatrix} - \begin{pmatrix} A \\ \sqrt{\mu_v} I \\ \sqrt{\mu_u} I \end{pmatrix} Z \right\|_F^2 + \lambda \|Z\|_* \quad (24)$$

Equation (24) can be solved by soft thresholding of singular values [32] as follows

$$Z \leftarrow Soft\left( Singular\ value(T), \frac{\lambda}{2\alpha} \right)$$

$$T = Z + \frac{1}{\alpha} \left( \begin{pmatrix} A \\ \sqrt{\mu_v} I \\ \sqrt{\mu_u} I \end{pmatrix}^T \left( \begin{pmatrix} Y \\ \sqrt{\mu_v}(P - B1) \\ \sqrt{\mu_u}(Q - B2) \end{pmatrix} - \begin{pmatrix} A \\ \sqrt{\mu_v} I \\ \sqrt{\mu_u} I \end{pmatrix} Z \right) \right) \quad (25$$

where, $Soft(t,u) = sign(t)\ \max(0, |t| - u)$ and

$$\alpha \geq \max \left( eig \begin{pmatrix} A \\ \sqrt{\mu_v} I \\ \sqrt{\mu_u} I \end{pmatrix}^T \begin{pmatrix} A \\ \sqrt{\mu_v} I \\ \sqrt{\mu_u} I \end{pmatrix} \right).$$

2nd subproblem can be cast as a least square expression as

$$\min_{P} \left\| \begin{pmatrix} \sqrt{\lambda_v} L_v \\ \sqrt{\mu_v}(X + B1) \end{pmatrix} - \begin{pmatrix} \sqrt{\lambda_v} W_v \\ \sqrt{\mu_v} I \end{pmatrix} P \right\|_F^2 \quad (26)$$

Similarly, sub problem 3 can be recast as follows

$$\min_{Q} \left\| \begin{pmatrix} \sqrt{\lambda_u} L_u \\ \sqrt{\mu_u}(X + B2) \end{pmatrix} - Q \begin{pmatrix} \sqrt{\lambda_v} W_u \\ \sqrt{\mu_u} I \end{pmatrix} \right\|_F^2 \quad (27)$$

Equation (22), (23), (26) and (27), are simple least square expressions which can be efficiently solved using any conjugate gradient type solver. In each iteration Bregman variables are updated as follows

$$B2 = B2 + Z - Q \quad (28)$$

$$B1 = B1 + Z - P \quad (29)$$

The iterations continue till convergence. The complete algorithm (LCMC-Label consistent matrix Completion) is given in fig 2.

## 4. Experiment and Results

We demonstrate the performance of our algorithm for a movie recommender system. We conducted experiments on 100K and 1M Movielens datasets (http://grouplens.org/datasets/movielens/). To the best of our knowledge, these are the only public datasets which provide relevant user and item metadata with ratings.

*Initialize variables,*
*Set regularization parameters; max_iter*
*while not convergence*
$// \ Solve\ for\ Z;\ Z \leftarrow Soft\left( Singular\ value(T), \frac{\lambda}{2\alpha} \right)$
$// \ Solve\ for\ P;\ Solve \begin{pmatrix} \sqrt{\lambda_v} L_v \\ \sqrt{\mu_v}(X + B1) \end{pmatrix} = \begin{pmatrix} \sqrt{\lambda_v} W_v \\ \sqrt{\mu_v} I \end{pmatrix} P$
$// \ Solve\ for\ Q;\ Solve \begin{pmatrix} \sqrt{\lambda_u} L_u \\ \sqrt{\mu_u}(X + B2) \end{pmatrix} = Q \begin{pmatrix} \sqrt{\lambda_v} W_u \\ \sqrt{\mu_u} I \end{pmatrix}$
$// \ Solve\ for\ W_v;\ Solve \quad \min_{W_v} \|L_v - W_v P\|_F^2$
$// \ Solve\ for\ W_u;\ Solve \quad \min_{W_u} \|L_u - QW_u\|_F^2$
*// Update Bregman Variable;*
$B2 = B2 + Z - Q; B1 = B1 + Z - P$
*end while*

**Figure 2. Algorithm - LCMC**

### 4.1 Description of Datasets

Both the datasets contain ratings on a scale of 1-5. 100K dataset contains 100K ratings given by 943 users on 1682 movies and 1M dataset has 1M ratings on around 3952 movies given by 6040 users. Both datasets have less than 5% of the ratings available and hence the improvement achieved by using metadata can be adequately gauged.

For users 30 groups are constructed – 2 for gender (M/F), 7 for multiple age-brackets (1-17, 18-24, 25-34, 35-44, 45-49, 50-55 and 56+) and 21 for various occupations. For items, 19 groups are formed, each representing a different genre. This information is used to construct label matrices $(L_u, L_v)$ as discussed in section 3.

### 4.2 Experimental Setup and Evaluation Criteria

We conducted 5-fold cross validation on both the datasets; 80% of the ratings forming the train set and remaining 20% used for testing. The simulations are carried out on system with i7-3770S CPU @3.10GHz with 8GB RAM. For cold start testing, 80% of users (items) were kept as part of training data and test done on remaining 20% users (items).

For offline baseline estimation, value of $\delta$ in (2) is set as $1e-3$. The value of regularization parameters for our formulation (18) is selected using greedy L-curve technique [33]. The values for both 100K and 1M dataset are $\lambda = 1e+1$, $\lambda_u = 1e-1$, $\lambda_v = 1e-1$, $\mu_u = 1$, $\mu_v = 1$. The overall accuracy of our model is evaluated using MAE (Mean absolute error) (30) and RMSE (root mean square error) (31).

$$MAE = \frac{\sum_{i,j} R_{i,j} - \hat{R}_{i,j}}{|R|} \quad (30)$$

$$RMSE = \sqrt{\frac{\sum_{i,j}\left(R_{i,j} - \hat{R}_{i,j}\right)^2}{|R|}} \quad (31)$$

where, $R$ and $\hat{R}$ are the actual and predicted ratings and $|R|$ is the cardinality of the rating matrix $R$.

The relevance of recommendations for each user is measured in terms of precision (32) and recall (33) [34] for top-N recommendations. The values depicted in the results are the average of values computed for each user. Precision and recall curves are plotted for varying number of recommendations.

$$Precision = \frac{\#t_p}{\#t_p + \# f_p} \quad (32)$$

$$Recall = \frac{\#t_p}{\#t_p + \# f_n} \quad (33)$$

Here, $t_p$ denotes true positive (item relevant and recommended), $f_p$ denotes false positive (item irrelevant and recommended) and $f_n$ denotes false negative (item relevant and not recommended). An item is marked relevant if it's rated as above 3 else irrelevant.

### 4.3 Analyzing impact of metadata

In this section we present the results of our proposed formulations – Label consistent matrix completion with user metadata (LCMC-U) (11), Label consistent matrix completion with item metadata (LCMC-I) (12), Label consistent matrix completion with user and item metadata (LCMC-UI) (13).

We compare the result of our work with state of the art matrix completion and matrix factorization algorithms – Accelerated Proximal gradient (APG) [21], Block Co-ordinate descent based Non negative matrix factorization (BCD-NMF) [35], Factored item similarity model (FISM) [36] and Probabilistic matrix factorization (PMF) [37].

To further highlight the contribution of user/item metadata in improving recommendation accuracy, we also show the results for following two (sub) formulations:

1. MC: Formulation exploiting just the rating data in a nuclear norm minimization framework i.e. user/item metadata is not utilized (34).

$$\min_{Z} \|Y - A(Z)\|_F^2 + \lambda \|Z\|_* \quad (34)$$

For solving (34) we adopt split Bregman technique, similar to one used for our formulation, to maintain consistency of algorithm efficiency and highlight the contribution of our model (13).

2. LC: Formulation exploiting only the label consistency constraints i.e. without the low rank nature of rating matrix being taken into consideration (35).

$$\min_{Z,W_v,W_v} \|Y - A(Z)\|_F^2 + \lambda_v \|L_v - W_v Z\|_F^2 + \lambda_u \|L_u - ZW_u\|_F^2 \quad (35)$$

Equation (35) is a least squares formulation which can be easily solved.

TABLE 1. ERROR MEASURES

| | 100K Dataset | | 1M Dataset | |
|---|---|---|---|---|
| **Algorithm** | **MAE** | **RMSE** | **MAE** | **RMSE** |
| **LCMC-U** | **0.7230** | **0.9207** | **0.6767** | **0.8634** |
| **LCMC-I** | **0.7224** | **0.9216** | **0.6766** | **0.8612** |
| **LCMC-UI** | **0.7193** | **0.9145** | **0.6731** | **0.8559** |
| MC | 0.7351 | 0.9319 | 0.6813 | 0.8711 |
| LC | 0.7481 | 0.9473 | 0.7186 | 0.9094 |
| APG | 0.8847 | 3.7076 | 0.9782 | 3.8109 |
| PMF | 0.7564 | 0.9639 | 0.7241 | 0.9127 |
| BCD-NMF | 0.7582 | 0.9816 | 0.6863 | 0.8790 |
| FISM | 0.7432 | 0.9439 | 0.7196 | 0.9102 |

Table 1 illustrates the MAE and RMSE values for the 100K and 1M datasets for various algorithms. The results obtained for nuclear norm minimization algorithm using split Bregman technique (MC) indicate that it gives around 3% lower MAE and 3.5% lower RMSE value than the next best latent factor model based MF algorithm i.e. PMF. Also, MC is superior than the neighborhood inspired factor model (FISM) and achieves a 1.5% lower MAE than the latter. This demonstrates the efficiency of our algorithm using split Bregman technique over other methods.

Also, our formulation using only the metadata (label consistency) constraints also yields fairly good results. We are able to outperform the existing matrix factorization algorithm as well (i.e. PMF and BCD-NMF) by around 1.7%. It gives results quite close (MAE 0.7481) to those obtained using FISM (MAE 0.7432). Thus, both our individual formulations, one including rating information and other involving metadata give good results. Then the obvious next step is to combine both information sources to get improved prediction accuracy, as in our combined formulation LCMC.

Comparison of our formulations incorporating user/item metadata (LCMC) with one using just the rating data (MC) corroborate our claim that use of secondary information indeed improves recovery accuracy. Our proposed formulations are able to better the MAE and RMSE values by around 2% over the MC algorithm. Using both user and item metadata yields slightly better result than each of them individually.

For 1M dataset also MC formulation outperforms existing MC/MF algorithms. Use of secondary data is able to achieve a reduction of around 1.5% in error measures over formulations just exploiting rating data.

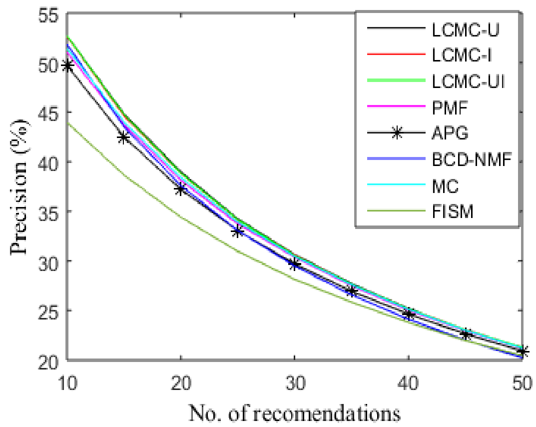


**Figure 3. Precision Curve (100K dataset)**

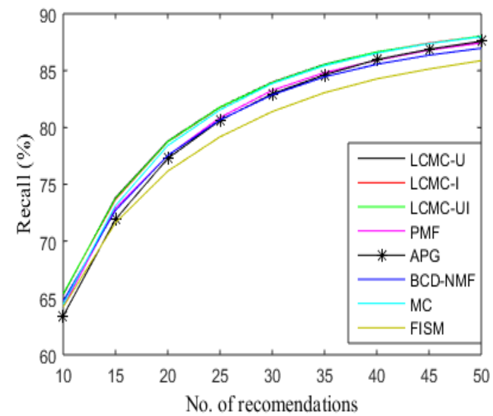


**Figure 4. Recall Curve (100K dataset)**

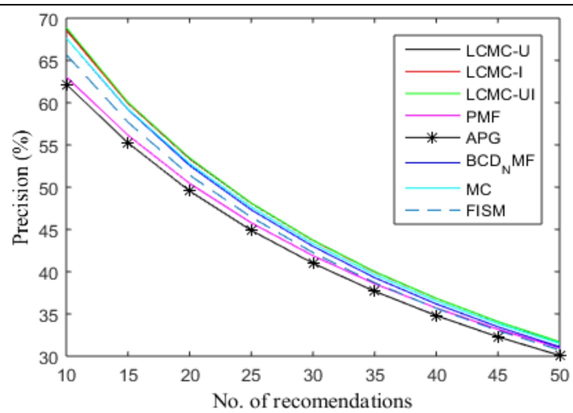


**Figure 5. Precision Curve (1M dataset)**

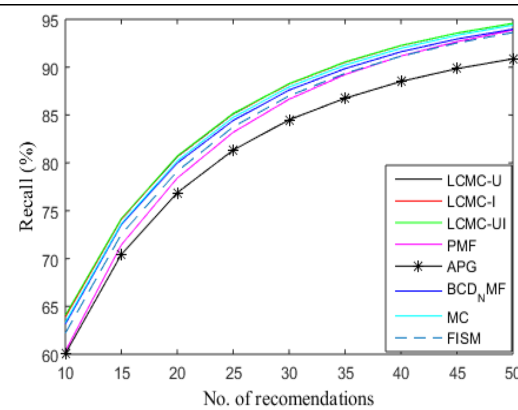


**Figure 6. Recall Curve (1M dataset)**

The precision and recall curves for all the algorithms for 100K and 1M dataset are given in figures 3-6. Here also, our formulations (LCMC) show better performance than the algorithms compared against. However, there isn't much difference between the precision and recall values for LCMC formulation using either individual user or item metadata or a combination of both. Also, the improvement using our algorithm is more pronounced for the 1M dataset.

### 4.4 Comparison with existing techniques

In this section we showcase the superiority of our supervised learning based approach for assimilating user/item metadata over other methods utilizing similar information. We compare the performance of our formulation against a neighbourhood based method (KNN) proposed in [25] and against a latent factor MF based formulation (Graph Reg) using graph regularization [12]. We also compared our work against two other works - a semi supervised learning based non negative matrix factorization (SSNMF) technique proposed in [38] and matrix completion framework with user metadata (MCAI) proposed in [28].

TABLE 2. ERROR MEASURES

| | 100K Dataset | | 1M Dataset | |
|---|---|---|---|---|
| **Algorithm** | **MAE** | **RMSE** | **MAE** | **RMSE** |
| **LCMC-UI** | **0.7193** | **0.9145** | **0.6739** | **0.8559** |
| KNN | 0.8302 | 1.0467 | 0.8198 | 0.9989 |
| SSNMF | 0.7723 | 1.0112 | 0.7285 | 0.9401 |
| Graph Reg | 0.7577 | 0.9616 | 0.7233 | 0.9139 |
| MCAI | 0.7206 | 0.9187 | 0.6749 | 0.8622 |

Table 2 shows the comparison of error measures for 100K and 1M datasets. Amongst all the algorithms for both the datasets KNN gives the poorest results. This is owing to the fact that neighbourhood based methods are simple heuristic measures which perform worse than latent factor models. On comparison to latent factor formulation – Graph Reg – our method yields more than

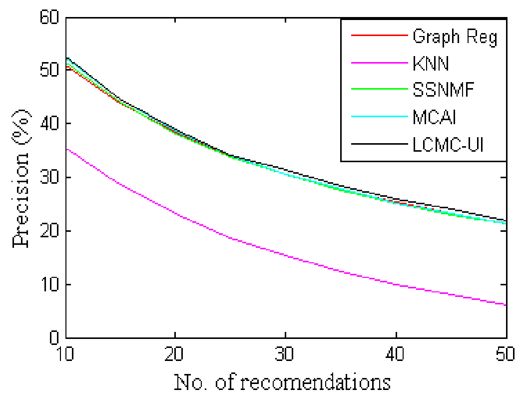


**Figure 7. Precision Curve (100K dataset)**

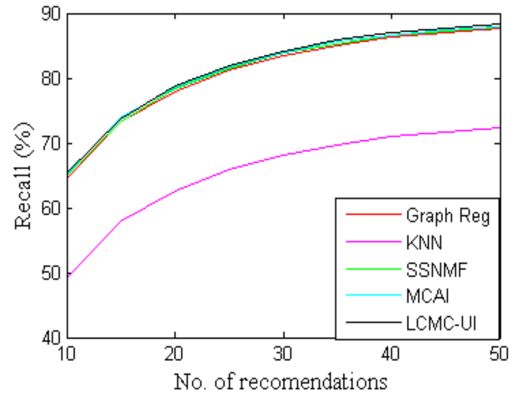


**Figure 8. Recall Curve (100K dataset)**

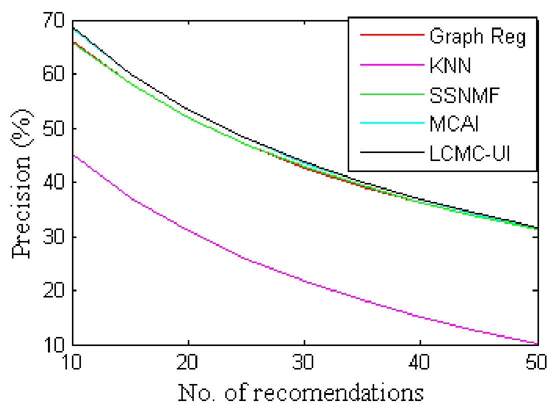


**Figure 9. Precision Curve (1M dataset)**

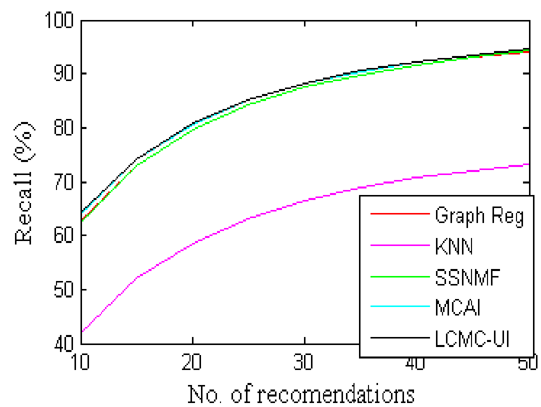


**Figure 10. Recall Curve (1M dataset)**

5% lower MAE and RMSE values. Also, as compared to semi-supervised learning approach adopted in [38] our label consistent formulation is much better at capturing the metadata information. We are able to get ~8% reduction in MAE and RMSE values. It is also partly contributed by use of our algorithm designed using split Bregman approach. Comparison to another matrix completion based approach (MCAI) also indicates that our formulation is able to achieve a reduction in both MAE and RMSE. It can be contributed to use of our novel label consistent formulation that enables use of both user and item metadata. Thus, it validates our claim that our label consistent formulation is able to better capture the correlation amongst users and items based on their associated metadata.

The precision and recall curves for these methods are given in figure 7-10. On this measure also, it's clear that our method performs better or at least comparable than the other two compared against. The improvement is more significant for 1M dataset, owing to higher sparsity of the rating dataset.

### 4.5 Cold Start Problem

In this section we present our results for both the user and item (pure) cold start problem (U-CS and I-CS). For evaluation of our algorithm, we compute MAE and RMSE values. None of the existing works report results on both (user and item) cold start problems and hence we compare against different works. For comparison, we report the results indicated in the recent works. Table 3 gives results for our algorithm for item and user cold start condition for 100K and 1M datasets.

TABLE 3. ERROR MEASURES FOR COLD START

| Algorithm | MAE | RMSE |
|---|---|---|
| User Cold Start - 100K | 0.7275 | 0.9217 |
| Item Cold Start - 100K | 0.7271 | 0.9214 |
| User Cold Start - 100K | 0.7100 | 0.8984 |
| Item Cold Start - 100K | 0.7099 | 0.8983 |

From the above data it can be observed that our design methodology for solving the cold start problem gives fairly good results. The MAE and RMSE values for cold start (users or items) is sufficiently close to those obtained for existing (warm start) users and items; as shown in results discussed in section 4.4.

Results shown in previous works are very limited with most of the works solving the user end cold start problem. In [39] authors solved new user cold start problem by proposing a hybrid system based on SCOAL. They segregate users into groups based on available information and design separate prediction model for each group. The new user, based on this demographic profile, is assigned to closest group and his ratings predicted accordingly. They reported a MAE of 0.93 for the 100K dataset, 29% higher than our MAE (0.73).

In [11] authors used known classification algorithms in combination with similarity techniques (similarity computed based on demographic information) and prediction mechanisms to retrieve recommendations. They conducted experiment on Movielens 1M dataset and reported an MAE of 0.75 and RMSE of 0.95. Our corresponding values for 1M dataset are 0.71 and 0.89.

Thus our algorithm significantly outperforms existing, state of the art, works for mitigating the cold start problem.

## 5. Conclusion

In this work, we propose a formulation to incorporate user-item metadata in a supervised learning augmented matrix completion framework. Our design targets accuracy improvement for new users and rating prediction for new users and items. Most existing works incorporate secondary information in a matrix factorization framework. However, MF being bi-linear and hence non-convex formulation does not provide convergence guarantees. We augment the convex matrix completion framework to include available metadata.

We defined multiple classes for both users and items based on available secondary information. Using this information, label matrices were constructed and used as additional information source. The rating values were predicted under the additional constraint of maintaining this label consistency. Use of add-on constraint helps reduce solution search space; in effect reducing the underdetermined nature of the problem. We also propose an algorithm using split Bregman technique for our proposed formulation.

Our design for cold start problem also uses information generated using the proposed label consistent model and hence proves efficient in terms of computational load. Most existing works focus on cases where a few ratings are available, whereas in this paper we solve a more challenging, pure cold start problem.

We illustrated the efficiency of our algorithm structure by comparing a basic matrix completion framework using split Bregman with existing MF/MC methods. We find that we are able to achieve better results than state of the art techniques in low-rank matrix completion. Secondly, we demonstrate the improvement obtained using the base MC formulation by augmenting it with label consistent information. Comparison with existing methods using metadata also shows the superiority of our design. In case of cold start problem, our framework is able to generate far superior results than the existing state of the art methods for both new user and new items. In the future, we would like to extend our design for other recommender system as well as for simultaneous new user-new item cold start problem.

## References


[1] Bobadilla, Jesús, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. "Recommender systems survey." Knowledge-Based Systems 46 (2013): 109-132.

[2] Park, Deuk Hee, Hyea Kyeong Kim, Il Young Choi, and Jae Kyeong Kim. "A literature review and classification of recommender systems research." Expert Systems with Applications 39, no. 11 (2012): 10059-10072

[3] Melville, P., Mooney, R. J., & Nagarajan, R. (2002, July). Content-boosted collaborative filtering for improved recommendations. In AAAI/IAAI (pp. 187-192).

[4] Vozalis, M. G., & Margaritis, K. G. (2007). Using SVD and demographic data for the enhancement of generalized collaborative filtering. Information Sciences, 177(15), 3017-3037.

[5] Victor, Patricia, Martine De Cock, Chris Cornelis, and A. Teredesai. "Getting cold start users connected in a recommender system's trust network." Computational Intelligence in Decision and Control 1 (2008): 877-882.

[6] Zhang, Zi-Ke, Chuang Liu, Yi-Cheng Zhang, and Tao Zhou. "Solving the cold-start problem in recommender systems with social tags." EPL (Europhysics Letters) 92, no. 2 (2010): 28002

[7] Zhou, K., Yang, S. H., & Zha, H. (2011, July). Functional matrix factorizations for cold-start recommendation. In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval (pp. 315-324). ACM.

[8] Houlsby, N., Hernandez-lobato, J. M., & Ghahramani, Z. (2014). Cold-start active learning with robust ordinal matrix factorization. In Proceedings of the 31st International Conference on Machine Learning (ICML-14) (pp. 766-774).

[9] Schafer, J. Ben, Dan Frankowski, Jon Herlocker, and Shilad Sen. "Collaborative filtering recommender systems." In The adaptive web, pp. 291-324. Springer Berlin Heidelberg, 2007

[10] Cacheda, Fidel, Víctor Carneiro, Diego Fernández, and Vreixo Formoso. "Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems." ACM Transactions on the Web (TWEB) 5, no. 1 (2011): 2

[11] Lika, Blerina, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. "Facing the cold start problem in recommender systems." Expert Systems with Applications 41, no. 4 (2014): 2065-2073.

[12] Gu, Quanquan, Jie Zhou, and Chris HQ Ding. "Collaborative Filtering: Weighted Nonnegative Matrix Factorization Incorporating User and Item Graphs." In SDM, pp. 199-210. 2010.
[13] Zhang, Y., Chen, W., & Yin, Z. (2013). Collaborative filtering with social regularization for TV program recommendation. Knowledge-Based Systems,54, 310-317
[14] Wang, Jun, Arjen P. De Vries, and Marcel JT Reinders. "Unifying user-based and item-based collaborative filtering approaches by similarity fusion." In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 501-508. ACM, 2006.
[15] Adomavicius, Gediminas, and Alexander Tuzhilin. "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions." Knowledge and Data Engineering, IEEE Transactions on 17, no. 6 (2005): 734-749
[16] Hofmann, Thomas. "Latent semantic models for collaborative filtering." ACM Transactions on Information Systems (TOIS) 22, no. 1 (2004): 89-115
[17] Koren, Yehuda, Robert Bell, and Chris Volinsky. "Matrix factorization techniques for recommender systems." Computer 8 (2009): 30-37
[18] Jaggi, Martin, and Marek Sulovsk. "A simple algorithm for nuclear norm regularized problems." In Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 471-478. 2010
[19] Shamir, Ohad, and Shai Shalev-Shwartz. "Collaborative filtering with the trace norm: Learning, bounding, and transducing." (2011
[20] Candès, Emmanuel J., and Benjamin Recht. "Exact matrix completion via convex optimization." Foundations of Computational mathematics 9, no. 6 (2009): 717-772.
[21] Toh, K. C., & Yun, S. (2010). An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. Pacific Journal of Optimization, 6(615-640), 15.
[22] Nesterov, Y., & Nesterov, I. E. (2004). Introductory lectures on convex optimization: A basic course (Vol. 87). Springer.
[23] Mohan, K., & Fazel, M. (2012). Iterative reweighted algorithms for matrix rank minimization. The Journal of Machine Learning Research, 13(1), 3441-3473.
[24] Goldstein, T., & Osher, S. (2009). The split Bregman method for L1-regularized problems. SIAM Journal on Imaging Sciences, 2(2), 323-343.
[25] Vozalis, M., & Margaritis, K. G. (2004, August). Collaborative filtering enhanced by demographic correlation. In AIAI Symposium on Professional Practice in AI, of the 18th World Computer Congress
[26] Ostrikov, Alexander, Lior Rokach, and Bracha Shapira. "Using geospatial metadata to boost collaborative filtering." In Proceedings of the 7th ACM conference on Recommender systems, pp. 423-426. ACM, 2013
[27] Zhou, Tinghui, Hanhuai Shan, Arindam Banerjee, and Guillermo Sapiro. "Kernelized Probabilistic Matrix Factorization: Exploiting Graphs and Side Information." In SDM, vol. 12, pp. 403-414. 2012
[28] Gogna, Anupriya, and Angshul Majumdar. "Matrix completion incorporating auxiliary information for recommender system design." Expert Systems with Applications 42, no. 14 (2015): 5789-5799.
[29] Massa, Paolo, and Paolo Avesani. "Trust-aware recommender systems." In Proceedings of the 2007 ACM conference on Recommender systems, pp. 17-24. ACM, 2007.
[30] Bobadilla, JesúS, Fernando Ortega, Antonio Hernando, and JesúS Bernal. "A collaborative filtering approach to mitigate the new user cold start problem." Knowledge-Based Systems 26 (2012): 225-238
[31] Nguyen, An-Te, Nathalie Denos, and Catherine Berrut. "Improving new user recommendations with rule-based induction on cold user data." In Proceedings of the 2007 ACM conference on Recommender systems, pp. 121-128. ACM, 2007
[32] Mohan, K., & Fazel, M. (2010, June). Reweighted nuclear norm minimization with application to system identification. In American Control Conference (ACC), 2010 (pp. 2953-2959). IEEE.
[33] Lawson, C. L., & Hanson, R. J. (1974). Solving least squares problems (Vol. 161). Englewood Cliffs, NJ: Prentice-hall
[34] Shani, G., & Gunawardana, A. (2011). Evaluating recommendation systems. In Recommender systems handbook (pp. 257-297). Springer US.
[35] Xu, Y., & Yin, W. (2013). A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. SIAM Journal on Imaging Sciences, 6(3), 1758-1789
[36] Kabbur, S., Ning, X., & Karypis, G. (2013, August). Fism: factored item similarity models for top-n recommender systems. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 659-667). ACM
[37] Mnih, A., & Salakhutdinov, R. (2007). Probabilistic matrix factorization. InAdvances in neural information processing systems (pp. 1257-1264).
[38] Lee, Hyekyoung, Jiho Yoo, and Seungjin Choi. "Semi-supervised nonnegative matrix factorization." Signal Processing Letters, IEEE 17, no. 1 (2010): 4-7.
[39] Pereira, Andre Luiz Vizine, and Eduardo Raul Hruschka. "Simultaneous co-clustering and learning to address the cold start problem in recommender systems." Knowledge-Based Systems 82 (2015): 11-19