Received November 16, 2015, accepted December 15, 2015, date of publication December 21, 2015, date of current version December 31, 2015.

*Digital Object Identifier 10.1109/ACCESS.2015.2510659*

# A Comprehensive Recommender System Model: Improving Accuracy for Both Warm and Cold Start Users

**ANUPRIYA GOGNA, (Student Member, IEEE), AND ANGSHUL MAJUMDAR, (Member, IEEE)**
IIIT-Delhi, New Delhi, 110020, India

Corresponding author: A. Gogna (anupriyag@iiitd.ac.in)

**ABSTRACT** Sparsity of the ratings available in the recommender system database makes the task of rating prediction a highly underdetermined problem. This poses a limit on the accuracy and the quality of prediction. In this paper, we utilize secondary information pertaining to user's demography and item categories to enhance prediction accuracy. Within the matrix factorization framework, we introduce additional supervised label consistency terms that match the user and item factor matrices to the available secondary information (metadata). Matrix factorization model—conventionally employed in collaborative filtering techniques—yields dense user and dense item factor matrices—the assumption is that users have an affinity toward all latent factors and items possess all latent factors. Our formulation, based on a recent work, aims to recover a dense user and a sparse item factor matrix—this is a more reasonable model. Human beings show a natural interest toward all the factors, but every item cannot possess all the factors; this leads to a sparse item factor matrix. A natural outcome of our proposal is a solution to the pure cold start problem. We utilize the label consistency map generated from the proposed model to make reasonable recommendations for new users and new items which have not (been) rated yet. We demonstrate the performance of our model for a movie recommendation system. We also design an efficient algorithm for our formulation.

**INDEX TERMS** Auxiliary information, blind compressive sensing, cold start, latent factor model, matrix factorization.

## I. INTRODUCTION

Recommender systems (RS) [13], [25] help mitigate the problem of information overload by saving users from the discouraging job of filtering through huge pile of information/ items, and finding the relevant. The contribution of RS towards improving customer satisfaction fuels the interest of both academia and industry in design of efficient techniques for the same.

The most popular technique for RS design is collaborative filtering (CF) [18], [34]. It uses the feedback given by users to a subset of items to make relevant recommendations on the remaining. Feedback can either be explicit – users clearly providing ratings (say on scale of 1-5) or implicit – inferred from user's behavior like browsing history and buying pattern. Explicit ratings, though more reliable are scanty. On the other hand, implicit ratings are easier to obtain but are not very informative; especially for negative feedbacks, e.g. if a person does not buy a product one does not know if he/she does not like it or does not know about it. In this work we will work with explicit ratings for evident reasons.

The rating database can be exploited in two ways – via memory based methods or latent factor models (LFM). Memory based methods [3], [5], [38] are simple heuristic strategies, which though more intuitive, perform poorer than their (latent factor) model based counterpart in terms of accuracy of prediction and speed of computation [1].

Latent factor models [12], [40] build a lower dimensional representation of data under the assumption that a user's choice is guided by very few (latent) factors. Users show affinity towards these factors and items possess them to varying degree. Thus, both items and users can be defined as a vector of (latent) features and the interaction between the two (modeled as inner product between corresponding latent factor vectors) indicates the degree of user's penchant for the item.

Commonly used formulation for rating prediction, based on LFM, is matrix factorization (MF) [17]. It aims to recover the rating matrix as a product of two matrices; one corresponding to users' latent factor vectors and other representing the items. Conventionally employed MF model promotes recovery of a dense structure for both user and item latent factor vectors.

A recent work [9] has proposed to modify standard MF model to factor the rating matrix into a dense user and a sparse item latent factor component. A human being is expected to be well rounded in taste; thus he/she is expected to have curiosity towards all possible features, e.g. a user can like all movie genres. Hence, for a user the latent factor vector will be dense, having non-zero values corresponding to all the factors. However, no item (say a movie) will possess all these factors; a movie will never belong to all genres. In effect, item's latent factor vector should be highly sparse. The said formulation has been cast into blind compressive sensing (BCS) [8] framework. It has been shown to yield higher accuracy than the conventional MF framework wherein both user and item matrices are dense. Motivated by the improvements in [9], we follow the dense (user) – sparse (item) factor model as the basis for this work.

An improved latent factor model [9] alone does not suffice in improving the prediction accuracy significantly. This is because of the highly sparse nature of the rating dataset which puts a limit on achievable accuracy. The problem becomes even more pronounced in cold start condition i.e. for new users and new items which have no rating data associated with them. Thus, cold start problem cannot be solved by direct implementation of MF models discussed above. Insufficiency in providing relevant suggestions to new users can cause potential loss of customers for online portals, and hence cold start problem assumes great relevance for RS design.

Fortunately, along with the explicit ratings certain other pertinent information is also available in RS database; same can be utilized to compensate for insufficiency of collaborative (rating) data. Such information includes users' demographic profile, their social circle data, item descriptors or tagging information. This secondary information can be used to reduce the under-determinacy of (rating prediction) problem for warm start (existing) users as well as to provide basis for rating prediction in cold start condition.

In this work, our objective is to utilize auxiliary information about users and items in a latent factor framework using concepts from supervised learning. The novelty of our approach lies in presenting a comprehensive model, based on BCS framework augmented with supervised learning, for exploiting collaborative as well as auxiliary data suited to both warm and cold start users.

We make use of user demographic data and item category information to augment explicit rating database. Most online portals maintain classification data of their products or services, for example an online movie rental website maintains a database of movie genres. Also, demographic information about users is usually collected during the process of signing up. The ease of availability of this information, as compared to social tagging/network data used in most existing models [26], [27], [35], widens the applicability of our design. In addition, this information is available even for new users or new items which have no collaborative representation in the database.

We propose to devise a framework that uses available information (ratings and/or metadata) to generate high quality recommendations. Our model involves classifying users and items into several overlapping clusters, based on label information derived from available metadata. For example, users may be classified based on their demographic profile; similarly, items can be classified in accordance with the category (say genre for movies) they belong to. The class labelling is not restrictive and a user/item can simultaneously belong to multiple classes. Our primary model recovers the latent factor matrices for (warm start) users and items such that in addition to satisfying the (rating) data consistency constraints they are also consistent with the class label information. By introducing supervised learning into the BCS based matrix factorization model, we are able to improve recovery accuracy by effect of reducing data sparsity. Our model works well even for partial cold start condition (very few available ratings). In such cases, the limited rating information is suitably supported by use of available metadata to generate more accurate predictions as compared to using rating information alone.

A simple extension of our principal formulation presents a solution to the pure (item and user) cold start problem as well. To this effect, we learn a label consistency map – relating user or item label data to their latent factor vectors – almost on the side lines, while solving our primary formulation. This label consistency map along with auxiliary data of new users/items can be used to make predictions for them. Our proposition is based on the belief that users sharing similar classification profiles (say similar demographics) will tend to display similar preferences. Similarly, items belonging to common genre will possess similar (characterizing) latent factor vectors. Thus, label consistency map learned for existing users or items is equally applicable to new users and items as well.

Current works target the two problems - accuracy enhancement for warm start users [11], [37] and rating prediction for cold start scenario [30], [41] independently. There is no existing model, to the best of our knowledge, which attempts to solve both the relevant problems in a common framework. Also, most of existing literature solves partial cold start problem (using some amount of rating data with auxiliary information), whereas we require only auxiliary data for new users or items. Thus, our design enables high prediction accuracy for existing as well as new users within the same efficient framework. We also derive an efficient algorithm for our suggested formulation.

The main contribution of our work can be summarized as follows

- We present a framework to incorporate user/item metadata into the LFM, modeled as BCS formulation, using ideas from supervised learning.
- We extend our design to handle the pure cold start problem as well; thereby proposing a comprehensive model for both warm and cold start situations.

- We derive an efficient algorithm using Majorization-Minimization technique [6], for solving our formulation.

## II. RELATED WORK

In this section, first we briefly review the existing matrix factorization frameworks for latent factor model followed by a discussion of works incorporating auxiliary information into the basic framework. We also discuss the current literature on cold start problem.

### A. LATENT FACTOR MODELS - MATRIX FACTORIZATION APPROACH

Latent factor models [12] are constructed under the belief that relatively a very small number of features - the latent factors - govern a user's preference for any item. For instance, considering the case of movie recommendations, a user's choice for a movie can be influenced by characteristics such as genre, cast, language etc. To model this belief, a user ($u$) can be described as a vector ($U_u$) capturing the degree of his/her liking for each of the relevant features or latent factors. On similar lines, any movie ($m$) can be described as a vector ($V_m$) outlining the extent to which it possesses these features. The affinity of a user for an item can be modeled as an inner product between the individual feature vectors of user and item (1).

$$interaction(u, m) = \langle U_u, V_m \rangle \tag{1}$$

However, actual (explicit) ratings are not just a result of pure interaction but also have certain biases embedded in them. Users who usually rate items higher than the (global) average have a positive user bias. Likewise, items which are very popular like award winning movies tend to be rated high by almost all users giving them positive item bias. The net rating $R_{u,m}$ (by user $u$ on item $m$) can be modeled as follows [17].

$$R_{u,m} = \langle U_u, V_m \rangle + b_u + b_m + g \tag{2}$$

where, $b_u$, $b_m$ are the $u^{th}$ user's bias and $m^{th}$ item's bias; $g$ is the global mean of the rating matrix.

The term $b_u + b_m + g$ forms the baseline estimate which can be easily computed by solving the following

$$\min_{b_u, b_m} \sum_{u,m \in \Theta} \left( \left( R_{u,m} - b_u - b_m - g \right)^2 + \delta \left( b_u^2 + b_m^2 \right) \right) \tag{3}$$

where, $\delta$ is the regularization parameter. Equation (3) is solved using stochastic gradient descent algorithm [45].

Latent factor models use the matrix factorization framework [17] to recover the interaction component. It attempts to recover the same as a product of two matrices – users' ($U$) and items' ($V$) latent factor matrices.

If $Y$ is the interaction component of the available ratings such that $Y_{u,m} = R_{u,m} - b_u - b_m - g$ and $A$ is the binary mask (1's in place of available ratings and 0 for missing values), MF model to determine latent factors can be written as

$$\min_{U,V} \|Y - A(UV)\|_F^2 + \lambda \left( \|U\|_F^2 + \|V\|_F^2 \right) \tag{4}$$

where, $\lambda$ is the regularization parameter. The formulation in (4) recovers a dense user factor matrix and a dense item factor matrix promoted by the use of Frobenius norm penalty terms.

Some of the existing algorithm for matrix factorization include Probabilistic matrix factorization (PMF) [28], non-negative matrix factorization (NMF) [20] and block co-ordinate descent NMF (BCD-NMF) [39]. All of them solve the problem outlined in (4).

### B. BLIND COMPRESSIVE SENSING FRAMEWORK FOR LATENT FACTOR MODELS

Recently, authors in [9] suggested a modification to the conventional factorization scheme (4). Since this a new and relatively unknown model, we will discuss it briefly. A dense user vector highlights the condition that a user displays some degree of affinity towards all features. Considering the scenario of book recommendations, it translates into the proposition that any user will display some like/dislike towards all genres or all authors; which is a valid belief. However, considering items (books) a dense vector indicates that (all) book(s) will also belong to all genres or be linked to (written by) all authors; which is not conceivable. Hence, the latent factor vector for any item (a book in this case) should have a sparse structure having zeros corresponding to features not at all possessed (present) by the item. In light of the above argument, [9] presented a modification to the structure of matrix factorization problem by replacing the Frobenius norm constraint on item factor matrix by $l_1$ norm constraint promoting sparse recovery for item feature vectors (5).

$$\min_{U,V} \|Y - A(UV)\|_F^2 + \lambda_u \|U\|_F^2 + \lambda_v \|vec(V)\|_1 \tag{5}$$

where, $Y \in R^{M \times N}$, $U \in R^{M \times F}$, $V \in R^{F \times N}$ and $M$, $N$, $F$ are the number of users, items and latent factors considered respectively; $vec$ implies vectorized form; $\lambda_u$ and $\lambda_v$ are the regularization parameters.

The above formulation fits into the framework of blind compressive sensing (BCS) (6) [8].

$$\min_{c, \hat{\Psi}} \left[ \sum_{i=1}^{l} \left\| \phi_i \left( \hat{\Psi} c \right) - y_i \right\|_2^2 \right] + \lambda \|c\|_1 \quad s.t. \ \left\| \hat{\Psi} \right\|_F^2 \leq const \tag{6}$$

As, the work [9] is not well-known we briefly justify the choice of BCS for solving the problem in (5). According to the CS convention, in (6), $\phi$ is the sensing matrix and $\hat{\Psi}$ is the sparsifying basis/dictionary. For problem in (5), $A$ is a Dirac/sampling operator and $U$ replaces the dictionary $\hat{\Psi}$, both ($U$ and $\hat{\Psi}$) having small valued coefficients throughout (dense structure); thereby justifying the equivalence.

### C. COMBINING RATING DATA WITH AUXILIARY INFORMATION

Highly sparse nature of the explicit rating data proves to be a major factor limiting the accuracy of estimating the latent factor vectors. Several works have been proposed which augment the MF model (4), by including supplementary data, in order to improve the accuracy of estimation. We review some of them in this section.

Authors in [16] employed a combination of neighborhood and latent factor model based strategies to achieve higher accuracy. They combined the global outlook of latent factor models and similarity measure of neighborhood approach into a global optimization framework.

In [26] authors proposed a modified MF model (7) which includes an additional regularization term penalizing the deviation of a user's latent factor vector from other users in his/her trust network.

$$\min_{U,V} \|Y - A(UV)\|_F^2 + \lambda\left(\|U\|_F^2 + \|V\|_F^2\right) + \beta \sum_i \sum_{f \in F_i^+} sim(i,f)\left\|U_i - U_f\right\|_F^2 \tag{7}$$

where, $sim(i,f)$– similarity amongst users (based on rating pattern) – is used to weigh individual members of trust network differently. The model ensures that users in common trust network share similar latent factor vectors. Gradient based method was used to solve the formulation.

In [11], authors used graph regularization to include auxiliary data pertaining to user's demography and social profile with item categorization in a non-negative matrix factorization framework (8).

$$\min_{U,V} \|Y - A(UV)\|_F^2 + \lambda\left(Tr\left(U^T L_u U\right)\right) + \gamma\left(Tr\left(V^T L_v V\right)\right) \tag{8}$$

where, $L_u$ and $L_v$ are the graph Laplacians. The edges in the user/item graphs have similarity amongst users/items as the weight factors. The resulting optimization problem is formulated as low-rank semi-definite program. A similar graph based strategy was adopted in [36], where graphical representation was used to populate the sparse rating database using domain information.

Authors in [42] and [43] proposed to incorporate social network information into the MF framework. In [43] prior distribution of rows and columns of the latent factor matrix is described as a Gaussian process, whose correlation function captures the correlation between the vectors describing various users or items. Thus, it ensures that any interdependency or relation between latent factor vectors for individual items or users is accounted for. They employed stochastic gradient descent for solving their formulation.

All the existing latent factor model based works (utilizing secondary data) employ standard matrix factorization framework, yielding dense user-item latent factors. In this work, we incorporate supplementary information in a modified matrix factorization framework; following a blind compressive sensing formulation. We attempt to include auxiliary information in the form of class labels while allowing users and/or item to simultaneously belong to several classes. We also design an efficient algorithm based on Majorization minimization technique for our model.

### D. COLD START PROBLEM

For new users, registering on online portal, preference data (ratings) is unavailable; same is the case for new items being added. In such cold start conditions, auxiliary data proves to be a useful resource for making reasonable predictions. Several works have been undertaken in the past which propose a solution to the cold start problem.

Most works concentrate on users or items which have some rating information available for them and auxiliary data is used to supplement this information; thus they do not solve the pure cold start problem. We briefly review some of these approaches.

Authors in [29] used the available user metadata (demographic details) to model an alpha-community space model. Once a new user's community is defined, one recommendation list per community is generated based on adhoc level of agreement recommendation process.

Several works like [2] proposed to modify the similarity measure (using rating data alone) to make it more suitable for cold start users. The new similarity measure not only considers actual ratings but also the frequency and count of co-rated items to remove disparity between users with highly varied rating patterns.

Authors in [32] utilized the available rating data alone and relied on imputation to reduce data sparsity. They used auto-adaptive imputation method to fill in the missing ratings before applying neighborhood based scheme for rating prediction.

Most of existing literature [27], [44] that augment rating information with auxiliary data for cold start users utilize social profile or trust network of users to find similar users (amongst existing ones). Massa and Avesani [27] proposed a trust based measure to compute similarity and determine neighbors of a new user. Their model is based on the reasoning that in case of cold start users if rating data is used, the available set of neighbors will be very few. On the other hand, as trust propagates, it provides a measure to include a wider number of users to select the neighbors from. Authors in [44] 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.

Unlike most of existing literature, we in this work tackle the pure cold start problem (no rating information available) for new users as well as new items. Also, our proposed scheme has a cohesive structure targeting effective recommendations for both warm and cold start users.

## III. PROBLEM FORMULATION AND ALGORITHM DESIGN

In this section, we describe our proposed formulation for incorporating secondary data into a modified matrix factorization framework. The novelty of our work lies in formulating a new label consistent model for exploiting user and item metadata along with the explicit rating information. Each user and item can be assigned to one or more classes based on the available metadata and their latent factor vectors recovered such that they are consistent with the assigned class labels. We use the information generated from this supervised learning strategy to generate recommendations for cold start users and items as well.

## A. PROBLEM FORMULATION

### 1) INCORPORATING AUXILIARY DATA IN BLIND COMPRESSIVE SENSING FRAMEWORK

In this section, we discuss our design for a latent factor model based efficient RS which uses explicit ratings, user demography and item genre information.

As discussed above, ratings are composed of interaction and baseline components. In this work, baseline estimation is done offline using (3) and we only work with the interaction component in our proposed framework. Offline computation of baseline reduces the run times considerably without impacting the recovery accuracy. Once, the 'interaction' component is recovered, we adjust the biases based on offline estimation.

Our design is built on the latent factor model i.e. under the assumption that both users and items can be characterized in terms of their latent factor vectors. For greater accuracy in recovering the latent factor vectors, we require a large number of available ratings. In case of RS design, we typically have less than 10% of ratings available; data sparsity placing a limit on the achievable accuracy. In order to mitigate the effect of rating data sparsity i.e. reduce the underdetermined nature of the problem, we utilize user and item metadata.

We borrow ideas from supervised learning [14] to incorporate this metadata into our base formulation (5). Our model is constructed on the premise that users can be clubbed or categorized into several classes (of similar users) based on the available secondary information. In our work, we form classes as per users' age group, gender and occupation. A user can simultaneously belong to several classes. This class label information is used to learn the latent factor vectors of users in a supervised environment, such that they are consistent with the class label information. This secondary information (class label) provides additional constraints (thereby restricting the search space) which in effect reduces the under determinacy of the problem.

We model this idea as an additional regularization term appended to the base BCS formulation (5) which penalizes any deviation from class label consistency (9).

$$\min_{U,V,C} \|Y - A(UV)\|_F^2 + \lambda_u \|U\|_F^2 + \lambda_v \|vec(V)\|_1 + \mu_u \|W - UC\|_F^2 \tag{9}$$

where, $W \in R^{M \times C_u}$, $C \in R^{F \times C_u}$, $C_u$ is the number of classes constructed based on user metadata, $\mu_u$ is the regularization parameter. Class information matrix $(W)$ is constructed such that $W_{i,j} = 1$ if user $i$ belongs to class $j$ else 0.

The matrix $C$, which is learned as part of the optimization process, is basically a linear map from latent factor space to the classification domain. User latent factor matrix $(U)$ is learned so as to be consistent with the class label information via the mapping defined by $C$.

The idea is extended to items as well. All items are categorized based on their genre; each item belonging to one or more genres. Similar to (9), item information can be included as given in (10)

$$\min_{U,V,D} \|Y - A(UV)\|_F^2 + \lambda_u \|U\|_F^2 + \lambda_v \|vec(V)\|_1 + \mu_v \|Q - DV\|_F^2 \tag{10}$$

where, $Q \in R^{C_v \times N}$, $D \in R^{C_v \times F}$, $C_v$ is the number of classes constructed based on item genres, $\mu_v$ is the regularization parameter. Class information matrix $(Q)$ is constructed s.t. $Q_{i,j} = 1$ if movie (item) $i$ belongs to class (genre) $j$ else 0.

The matrix $D$ forms a linear map from the item latent factor domain to classification domain. Item latent factor matrix $(V)$ is learned so as to be consistent with the available class label (genre) information.

We also utilize item and user metadata together; yielding the combined formulation (11).

$$\min_{U,V,C,D} \|Y - A(UV)\|_F^2 + \lambda_u \|U\|_F^2 + \lambda_v \|V\|_1 + \mu_u \|W - UC\|_F^2 + \mu_v \|Q - DV\|_F^2 \tag{11}$$

In (11), explicit ratings are utilized to learn latent factor vectors by means of data consistency term $\left(\|Y - A(UV)\|_F^2\right)$ and labeling information used by means of label consistency terms $\left(\|W - UC\|_F^2, \|Q - DV\|_F^2\right)$.

Introducing supervised learning into the latent factor model helps improve prediction accuracy by alleviating the problem of rating matrix sparsity. The value of regularization parameters is found using $l$-curve technique [19]. Using (9) for rating prediction improves prediction accuracy for warm start users.

In our formulation we used user demographics and item category; however, the model can easily accommodate any other classification criteria (such as social network information or item descriptors like movie cast/director) as per the available information.

### 2) MITIGATING COLD START PROBLEM

Our proposed model can effectively target both the partial as well as pure cold start problem. In case of partial cold start problem i.e. for users or items with limited collaborative information, model presented in (11) can be applied. The lack of adequate amount of rating data can be compensated by use of metadata and help improve rating prediction accuracy.

More importantly, a simple extension to our label consistent formulation (11) generates quality recommendations in pure cold start conditions as well. The category information of new items added to the repository is invariably available in a RS database. Also, when a new user registers on the system, he/she is more often than not required to enter his/her age/gender related information. Thus for a new user or new item, no rating data is available, but associated metadata is readily available. Our model exploits the same for predicting their ratings.

Our formulation solves the pure cold start problem for both users and items, unlike most existing works capable of solving only partial cold start problem. To this effect we use the linear map ($C$ and $D$), generated almost on the side-lines while solving (11) along with the available metadata.

First, let us consider the new user cold start problem. Our design is based on the idea that uses' demographics have an impact on their preference. For example, it can be safely assumed that children will like animation movies or older generation will enjoy family/drama genre. Similarly, men may favour action movies more than women. Thus, users having similar demographic profiles can be defined by similar latent factor vectors. This theory can be exploited by using the mapping ($C$) between users' classification domain and their latent factor vectors to generate latent factors for new users if classification data is available.

When a new user registers on the system, his/her relevant demographic information is captured. Thus, for the said user the class label information (similar to that used for constructing label consistency matrix $W$) is available. Let $L_{new\_user} \in R^{1\times C_u}$ be the class label vector for the new user.

We use this information along with the linear (information) map $C$ to estimate the latent factor vector for the new user. As discussed above, matrix $C$ is a map from latent factor space to user class domain. It primarily establishes a generic relation between a user's demographic profile and its latent factor vector. Hence, it gives information about the user's preference for features given his/her age, gender and occupational profile. This information can be exploited for cold start users as well.

The class label vector $L_{new\_user}$ of new user and the linear map $C$ can be related as follows

$$L_{new\_user} = U_{new\_user} \times C \tag{12}$$

where, $U_{new\_user} \in R^{1\times F}$($F$ being the number of latent factors considered) is the latent factor vector of the new user.

Equation (12) can be solved efficiently using any conjugate gradient based solver to estimate new user's latent factor vector. Once $U_{new\_user}$ is recovered, (interaction component) rating by the user can be computed as $Z_{new\_user} = U_{new\_user} \times V$, where $V$ is the latent factor matrix for existing items.

Similar model can be constructed for new items as well. Here also, we argue that items belonging to same genre will share similarities in their latent factor vectors. Thus, the mapping from item classification to item latent factor space, derived using existing items' auxiliary data, can be used for new items as well.

A new item can be characterized in terms of its genre data as a vector $L_{new\_item} \in R^{C_v\times 1}$. For item cold start case we use the information matrix $D$ which relates latent factor space with the item label space. Matrix $D$ establishes a relation between the latent factor vector of an item and its category (genre). For a new items we have the genre information and hence, similar to the case for new user, its latent factor vector $\left(V_{new\_item}\right)$ can be estimated using (13)

$$L_{new\_item} = D \times V_{new\_item} \tag{13}$$

Using previously estimated user latent factor matrix ($U$), users' rating (preference) for the new item can be computed as $Z_{new\_item} = U \times V_{new\_item}$.

Our design procedure can thus be used to solve both new user and new item pure cold start problem, without significant add-on computations. Our model requires solving a simple linear system of equation to enable rating prediction for new users or items.

### *B. ALGORITHM DESIGN*

In this section we design an efficient algorithm using Majorization Minimization (MM) technique [6] for our combined formulation (11). The use of MM approach allows us to break a complex optimization problem into simpler and more efficiently solvable steps.

Firstly, we employ Method of Alternating Direction [7] to split our formulation (11) into separate simpler sub problems, each minimizing over a single variable.

Sub Problem 1

$$\min_U \|Y - A(UV)\|_F^2 + \lambda_u \|U\|_F^2 + \mu_u \|W - UC\|_F^2 \tag{14}$$

Sub Problem 2

$$\min_V \|Y - A(UV)\|_F^2 + \lambda_v \|vec(V)\|_1 + \mu_v \|Q - DV\|_F^2 \tag{15}$$

Sub Problem 3

$$\min_C \|W - UC\|_F^2 \tag{16}$$

Sub Problem 4

$$\min_D \|Q - DV\|_F^2 \tag{17}$$

Considering sub problem 1; it can be cast as a least square problem and thus has a closed form solution. However, given the large size of the rating matrix, computing the pseudo inverse to obtain closed form solution, becomes a computational challenge. To eliminate prohibitively large computations, we use MM technique.

MM technique [6] involves replacing a complex optimization problem by its majorizer which is simpler to solve. A function $g(x)$ is a majorizer of $f(x)$ iff $g(x) \geq f(x) \forall x$ and $g(x) = f(x)$ at $x = x_k$.

We replace the sub problem in (14) by the following expression

$$\begin{aligned} \min_U \|Y - A(UV)\|_F^2 + \lambda_u \|U\|_F^2 + \mu_u \|W - UC\|_F^2 \\ + (z - z_k)^T \left(\beta I - A^T A\right)(z - z_k); \\ \beta \geq \max\left(eig\left(A^T A\right)\right) \end{aligned} \tag{18}$$

where, $z = vec(UV)$ ($vec$: vectorised form).

The add-on term being non-negative ensures (18) is a majorizer of (14). After some mathematical manipulations it can be shown that (18) reduces to the form given in (19).

$$\min_{U} \ \|S - UV\|_F^2 + \lambda_u \|U\|_F^2 + \mu_u \|W - UC\|_F^2$$
$$s.t.\ S = U_k V_k + \frac{1}{\beta} A^T \left(Y - A\left(U_k V_k\right)\right) \tag{19}$$

Equation (19) can be written as $(S + \mu_u W) V^T = U\left(VV^T + \lambda_u I + \mu_u CC^T\right)$ which being a simple least square can be solved easily.

Similar application of MM technique helps cast (15) as

$$\min_{V} \|S - UV\|_F^2 + \lambda_v \|vec(V)\|_1 + \mu_v \|Q - DV\|_F^2 \tag{20}$$

which can be rewritten as

$$\min_{V} \left\| \begin{pmatrix} S \\ \sqrt{\mu_v} Q \end{pmatrix} - \begin{pmatrix} U \\ \sqrt{\mu_v} D \end{pmatrix} V \right\|_F^2 + \lambda_v \|vec(V)\|_1 \tag{21}$$

Equation (21) can be updated via soft thresholding [5], wherein the solution is given by

$$V \leftarrow Soft\left(B, \frac{\lambda_v}{2\alpha}\right);\ \alpha \geq \max\left(eig\left(\begin{pmatrix} U \\ \sqrt{\mu_v} D \end{pmatrix}^T \begin{pmatrix} U \\ \sqrt{\mu_v} D \end{pmatrix}\right)\right)$$
$$B = V + {}^1\!/_\alpha \left(\begin{pmatrix} U \\ \sqrt{\mu_v} D \end{pmatrix}^T \left(\begin{pmatrix} S \\ \sqrt{\mu_v} Q \end{pmatrix} - \begin{pmatrix} U \\ \sqrt{\mu_v} D \end{pmatrix} V\right)\right) \tag{22}$$

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

Equation (16) and (17) are simple least squares; efficiently solvable by any conjugate gradient method. For cold start problem, we can again solve two least square expression, (12) and (13) using conjugate gradient. The complete algorithm (LC_BCS) is given in figure 1.

```
%%% CORE ALGORITHM
Set regularization parametrs, initalize variables
While not convergence
// Solve for V :    V <- Soft
  ((V + 1/α ((U; √μv D)^T ((S; √μv Q) - (U; √μv D) V))), λv/2α)
// Solve for U : (S + μu W) V^T = U (V V^T + λu I + μu C C^T)
// Solve for C :  min_C ||W - UC||_F^2
// Solve for D:   min_D ||Q - DV||_F^2
end while
%%% FOR USER COLD START PROBLEM
L_new_user = U_new_user × C
R_new_user = (U_new_user × V) + Baseline
%%% FOR ITEM COLD START PROBLEM
L_new_item = D × V_new_item
R_new_item = (U × V_new_item) + Baseline
```

**FIGURE 1.** Algorithm for LC_BCS.

## IV. EXPERIMENT AND RESULTS

We conducted experiment on two Movielens datasets [46] - 100K and 1M. These datasets from Grouplens are the most widely used publically available datasets for evaluating the performance of recommender systems. Both datasets have user and item metadata available. We are not aware of any other public dataset that provides the required metadata. Our results are compared against existing standard matrix factorization models and other existing approaches incorporating metadata for improving accuracy.

### A. DESCRIPTION OF DATASET

The 100K dataset has 100,000 ratings provided by 943 users for 1682 movies. The 1M dataset has around 3900 movie ids, 6040 users and 1 million ratings. Both the datasets, having ratings on a scale of 1-5, are extremely sparse and hence an optimum candidate for benefiting from use of metadata.

For classification of users (in both datasets) we consider 30 classes/categories - 7 age groups (1-17, 18-24, 25-34, 35-44, 45-49, 50-55, 56+), 2 gender groups (M/F) and 21 occupational categories. Items are assigned to one or more of 19 classes, each representing a single genre.

We briefly review construction of class matrices. Consider the user class matrix ($W$), having dimension 943 × 30. Each row corresponds to classification profile of one user; first seven columns representing age groups, next two gender and the remaining 21 occupations. Let user 1 be a 20 year old male working as an educator (occupation number 2), then $W_{1,2} = 1$, $W_{1,8} = 1$ and $W_{1,11} = 1$; rest of elements in row 1 being zero. Similarly for user 2, a 30 year old female working as an artist (occupation number 3), $W_{2,3} = 1$, $W_{2,9} = 1$ and $W_{2,12} = 1$; rest of elements in row 2 being zero. Similar strategy is adopted for constructing item class matrix as well.

### B. EXPERIMENTAL SETUP AND EVALUATION

For evaluating performance for warm start users we carried out 5-fold cross validation; in each instance 80% ratings forming the training data and remaining 20% acting as test set. For evaluating cold start model, 80% users (or items) were kept as part of training sample and test conducted on the remaining 20%. The simulations were carried out on system with i7-3770S CPU @3.10GHz processor with 8GB RAM.

The algorithms are evaluated on the basis of Mean absolute error (MAE) (23) and Root mean square error (RMSE) (24) to determine rating prediction accuracy.

$$MAE = \frac{\sum_{u,m} R_{u,m} - \hat{R}_{u,m}}{|R|} \tag{23}$$

$$RMSE = \sqrt{\frac{\sum_{u,m} \left(R_{u,m} - \hat{R}_{u,m}\right)^2}{|R|}} \tag{24}$$

 

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

The ranking efficiency of various formulations for top-N recommendations is evaluated via precision (25) and recall (26) [33].

$$Precision = \frac{\#t_p}{\#t_p + \#f_p} \tag{25}$$

$$Recall = \frac{\#t_p}{\#t_p + \#f_n} \tag{26}$$

where, $t_p$ denotes true positive (item relevant and recommended), $f_p$ is false positive (item irrelevant and recommended) and $f_n$ is false negative (item relevant and not recommended). To differentiate the relevant and irrelevant items we mark the items rated as 4 or 5 as irrelevant and those rated below (1-3) as irrelevant to the user.

### C. IMPACT OF INCORPORATING METADATA

In this section we demonstrate the improvement in accuracy obtained for warm start users by employing our models - LC_BCS_U (9), LC_BCS_I (10) and LC_BCS_UI (11) - compared to formulations exploiting only the rating.

We compare our proposed approach against the base formulation – BCS framework [9]. We also demonstrate comparison with other MF techniques - Matrix factorization (PMF) [28], SGD [17], and Block co-ordinate descent non-negative matrix factorization (BCD-NMF) [39] and Factored item similarity model (FISM) [15].

Table 1 lists the evaluation metrics for various algorithms on 100K dataset. BCS formulation yields improvement (~1% in MAE) over standard MF models, hence validating its use as our base formulation. It is also evident that use of additional information indeed improves the recovery accuracy. If used alone, item metadata yields better results than user metadata.

Combined model (LC_BCS_UI) helps improve the accuracy even further. It is able to obtain a reduction of over 2% in MAE and RMSE compared to BCS base model.

Our model incorporating user/item metadata also yields better results in terms of top-N recommendations. The precision and recall values (given for top-N recommendation, @N) for our model are consistently higher than those obtained using other methods.

Table 2 gives the results for 1M dataset. Here also similar pattern can be observed. Moreover, because of greater sparsity of 1M dataset, the improvement via use of metadata is more pronounced in this case; 3% drop in MAE and RMSE with respect to BCS model. The precision obtained via use of metadata is also much higher than for algorithms using explicit ratings alone.

### D. COMPARISON WITH EXISTING METHODS

This section contains comparison between our combined formulations (LC_BCS_UI) and existing works using user/item metadata targeted at improving accuracy for warm start users. We compare our work against the neighborhood (KNN)

**TABLE 1.** Evaluation metrics (100K dataset) – warm start.

| Algorithm | MAE | RMSE | Precision (%) | | | | | Recall (%) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | @10 | @20 | @30 | @40 | @50 | @10 | @20 | @30 | @40 | @50 |
| SGD | 0.7432 | 0.9421 | 50.72 | 37.95 | 30.28 | 24.97 | 21.17 | 63.76 | 76.95 | 82.63 | 85.25 | 86.80 |
| PMF | 0.7564 | 0.9639 | 50.52 | 37.74 | 30.06 | 24.71 | 21.09 | 63.56 | 76.76 | 82.43 | 85.06 | 86.56 |
| FISM | 0.7431 | 0.9439 | 43.95 | 34.41 | 38.14 | 23.79 | 20.46 | 63.20 | 75.15 | 80.39 | 83.27 | 84.07 |
| BCD-NMF | 0.7582 | 0.9816 | 51.33 | 37.22 | 29.2 | 23.88 | 20.04 | 64.13 | 76.77 | 82.04 | 84.7 | 86.10 |
| BCS | 0.7356 | 0.9409 | 51.33 | 38.05 | 30.14 | 24.82 | 21.31 | 64.16 | 77.57 | 82.89 | 85.47 | 86.86 |
| LC_BCS_U | 0.7253 | 0.9257 | 52.28 | 38.57 | 30.49 | 25.13 | 21.28 | 65.08 | 78.47 | 83.81 | 86.54 | 88.04 |
| LC_BCS_I | 0.7206 | 0.9201 | 52.36 | 38.66 | 30.67 | 25.17 | 21.26 | 65.67 | 78.55 | 83.98 | 86.57 | 87.97 |
| LC_BCS_UI | 0.7199 | 0.9196 | 52.48 | 38.91 | 30.82 | 25.38 | 21.47 | 65.14 | 78.75 | 84.15 | 86.77 | 88.18 |

**TABLE 2.** Evaluation metrics (1M dataset) – warm start.

| Algorithm | MAE | RMSE | Precision (%) | | | | | Recall (%) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | @10 | @20 | @30 | @40 | @50 | @10 | @20 | @30 | @40 | @50 |
| SGD | 0.6936 | 0.8763 | 62.46 | 49.85 | 41.3 | 35.1 | 30.39 | 60.35 | 77.18 | 84.8 | 88.81 | 91.17 |
| PMF | 0.7241 | 0.9127 | 63.06 | 50.45 | 41.9 | 35.7 | 30.79 | 60.32 | 78.41 | 86.65 | 91.17 | 93.82 |
| FISM | 0.7196 | 0.9102 | 65.69 | 51.42 | 42.26 | 35.67 | 30.75 | 62.2 | 79.12 | 87 | 91.15 | 93.63 |
| BCD-NMF | 0.6863 | 0.8790 | 66.95 | 52.64 | 42.57 | 35.82 | 30.79 | 62.51 | 79.18 | 86.72 | 90.69 | 93.03 |
| BCS | 0.6917 | 0.8789 | 67.19 | 52.36 | 42.94 | 36.16 | 31.17 | 62.53 | 79.91 | 87.57 | 91.47 | 93.82 |
| LC_BCS_U | 0.6796 | 0.8735 | 68.2 | 53.14 | 43.58 | 36.71 | 31.64 | 63.62 | 80.44 | 88.14 | 92.12 | 94.51 |
| LC_BCS_I | 0.6721 | 0.8577 | 68.2 | 53.15 | 43.58 | 36.71 | 31.64 | 63.64 | 80.45 | 88.14 | 92.13 | 94.51 |
| LC_BCS_UI | 0.6709 | 0.8567 | 68.22 | 53.17 | 43.59 | 36.72 | 31.64 | 63.67 | 80.47 | 88.15 | 92.15 | 94.51 |

IEEE Access The journal for rapid open access publishing

**TABLE 3.** Evaluation metrics (100K dataset) – warm start.

| Algorithm | MAE | RMSE | Precision (%) | | | | | Recall (%) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | @10 | @20 | @30 | @40 | @50 | @10 | @20 | @30 | @40 | @50 |
| KNN | 0.8302 | 1.0146 | 35.65 | 23.09 | 15.22 | 9.97 | 6.14 | 49.21 | 62.52 | 68.11 | 70.92 | 72.37 |
| Graph-Reg | 0.7577 | 0.9616 | 50.96 | 38.39 | 30.52 | 25.27 | 21.45 | 64.51 | 77.82 | 83.41 | 86.32 | 87.67 |
| SSNMF | 0.7723 | 1.0112 | 51.6 | 38.24 | 30.46 | 25.06 | 21.21 | 64.79 | 79.23 | 83.8 | 86.49 | 87.96 |
| BCS_M | 0.7202 | 0.9200 | 52.49 | 38.88 | 30.5 | 25.30 | 21.45 | 65.11 | 78.73 | 84.16 | 86.76 | 88.15 |
| LC_BCS_UI | 0.7199 | 0.9196 | 52.48 | 38.91 | 30.82 | 25.38 | 21.47 | 65.14 | 78.75 | 84.15 | 86.77 | 88.18 |

**TABLE 4.** Evaluation metrics (1M dataset) – warm start.

| Algorithm | MAE | RMSE | Precision (%) | | | | | Recall (%) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | @10 | @20 | @30 | @40 | @50 | @10 | @20 | @30 | @40 | @50 |
| KNN | 0.8198 | 0.9989 | 45.23 | 30.91 | 21.74 | 15.15 | 10.25 | 41.73 | 58.63 | 66.48 | 70.63 | 73.12 |
| Graph-Reg | 0.7233 | 0.9139 | 65.99 | 52.03 | 42.82 | 36.26 | 31.38 | 62.43 | 76.69 | 87.54 | 91.69 | 94.23 |
| SSNMF | 0.7285 | 0.9401 | 66.23 | 51.91 | 42.74 | 36.15 | 31.25 | 62.73 | 79.63 | 87.48 | 91.63 | 94.12 |
| BCS_M | 0.6738 | 0.8623 | 67.99 | 53.01 | 43.11 | 36.2 | 35.98 | 63.54 | 80.23 | 87.76 | 91.98 | 94.31 |
| LC_BCS_UI | 0.6709 | 0.8567 | 68.22 | 53.17 | 43.59 | 36.72 | 31.64 | 63.67 | 80.47 | 88.15 | 92.15 | 94.51 |

based method proposed in [37], graph regularized matrix factorization (Graph-Reg) formulation proposed in [11] and semi-supervised approach (SSNMF) proposed in [21]. We also compared our design against another BCS based model (BCS_M) presented in [10].

Authors is [37] put forth a neighborhood based model which uses a modified similarity measure (utilizing user's demographic information) to determine similar users. In [11] an augmented MF model with graph based regularization is proposed. Additional regularization terms capture the similarity amongst users/items based on available metadata. A semi-supervised learning based factorization model was proposed in [21] for general matrix factorization problem. They augmented the basic non-negative matrix factorization framework to exploit information from multiple sources for generating the factor matrices. However, their framework is highly restrictive imposing the condition of a common factor (matrix) while factorizing information from various sources. A BCS based framework is suggested in [10] for utilizing both user and item metadata. However, unlike our model none of above discussed strategies can generate relevant predictions for cold start users or items.

Table 3 and 4 gives the evaluation measures for 100K and 1M dataset respectively. It can be seen that our formulation (LC_BCS_UI) achieves an improvement of $\sim$15% in recovery accuracy as compared to KNN based approach; highlighting the ability of latent factor models in better handling the rating prediction problem. In comparison to standard MF model (dense user-item latent factor matrices) based designs i.e. Graph-Reg and SSNMF, our model achieves $\sim$5% reduction in error values. The BCS based model, BCS_M gives better results than the other works compared against. Our model (LC_BCS_UI) gives results slightly better than the BCS_M approach. This establishes the efficiency BCS framework in better handling the rating information and our more efficient algorithm design. Albeit the two BCS based models show nearly similar accuracy for warm start users, our model's ability to perform in cold start conditions as well, is an important advantage over BCS_M. The precision recall values also substantiate the claim that our design is a better model than other approaches.

**TABLE 5.** Run time comparison (100K dataset).

| Algorithm | Run Times (seconds) |
|---|---|
| KNN | 350.23 |
| Graph-Reg | 3.15 |
| SSNMF | 20.14 |
| BCS_M | 61.21 |
| LC_BCS_UI | 2.39 |

To highlight the efficiency of our algorithm design we also study the computation cost of each algorithm in terms of run time values. Table 5 shows the run times of various methods utilizing rating as well as secondary data for the 100K dataset.

As can be observed from the values in table 5, our method has the smallest run time in comparison to all other approaches. We are almost 30 times faster than BCS_M, the algorithm closest to our formulation in terms of prediction accuracy. Thus, our proposition offers a highly efficient framework targeting substantial gain in recovery accuracy with a simultaneous reduction in computational complexity (run times) as well. In addition, as shown in the next section, it offers an effective solution to cold start problem as well.

### *E. COLD START PROBLEM*

In this section we illustrate the efficiency of our formulation in solving pure user/item cold start problem.

**TABLE 6.** Error measures (100K dataset) – cold start.

| Algorithm | MAE | RMSE |
|---|---|---|
| User Cold Start | 0.7275 | 0.9224 |
| Item cold Start | 0.7273 | 0.9214 |

**TABLE 7.** Error measures (1M dataset) – cold start.

| Algorithm | MAE | RMSE |
|---|---|---|
| User Cold Start | 0.7082 | 0.8984 |
| Item cold Start | 0.7176 | 0.9148 |

Table 6 and 7 gives the MAE and RMSE values for both user and item cold start problem on 100K and 1M datasets respectively. The values clearly indicate that even for new users and items, our design generates highly relevant suggestions. The results are poorer than those reported for warm start scenario using LC_BCS_UI model, which is understandable given lack of any collaborative data. However, the linear mapping from rating domain to classification domain, used for cold start situation, is learned using both rating as well as demographic information and hence is robust even in case of cold start scenario. It is validated by the observations in table 3-7 that our design yields better results for cold start users than other works for even warm start users. This can be credited to our efficient use of available data compared to existing frameworks.

There are very limited works that target the pure cold start problem and almost all concentrate on new users and not on new items. We compare our model against two recent works [31], [22]. Both the works solve only user cold start problem.

In [31] a hybrid scheme based on SCOAL algorithm was proposed to alleviate the user cold start problem. SCAOL is used to cluster together users and builds a separate prediction model for each cluster. For each new user the closest cluster is identified (based on demographic details) and its preference predicted based on the applicable model. They reported a MAE of 0.93 for the 100K dataset, 28% higher than our MAE (∼0.73).

In [22] authors used combination of existing classification (computed based on demographic information) techniques and similarity based prediction mechanisms to retrieve recommendations. They conducted experiment on 1M Movielens dataset and reported an MAE of 0.75 and RMSE of 0.95. Our corresponding values for 1M dataset are 0.7082 and 0.8984 (an improvement of 6%).

## V. CONCLUSION

In this work, we propose to improve the accuracy of rating prediction in recommender systems by including user and item metadata in the latent factor model for both warm and cold start conditions. Instead of the conventional matrix factorization formulation, we use a recently proposed modified matrix factorization model as our baseline. The modified model is based on recovering a dense user and a sparse item latent factor matrix following a blind compressive sensing formulation. It is shown to outperform conventional MF algorithms.

We augment the basic BCS model, utilizing only the explicit user feedback, with information from secondary sources to reduce the under determined nature of the problem. Our new formulation expands on the BCS framework via supervised learning based design. We categorize users and items based on auxiliary information and their latent factor vectors are recovered such that (additionally) label consistency is maintained. Most existing works focus on using social information which is not very readily information. We utilize user demographic data and item categories which are more widely and easily accessible.

Natural progression of our model helps solve both the user and item cold start problem. Our design proves to be computationally efficient as both warm and cold start users are handled within the same comprehensive framework. Existing works mostly focus on solving either of the two problems. Our proposed formulation is solved using an efficient algorithm based on Majorization-minimization technique.

## REFERENCES

[1] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, Jun. 2005.
[2] H. J. Ahn, "A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem," *Inf. Sci.*, vol. 178, no. 1, pp. 37–51, Jan. 2008.
[3] R. M. Bell and Y. Koren, "Improved neighborhood-based collaborative filtering," in *Proc. KDD-Cup Workshop*, 2007, pp. 7–14.
[4] I. Daubechies, M. Defrise, and C. De Mol. (2003). "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint." [Online]. Available: http://arxiv.org/abs/math/0307152
[5] C. Desrosiers, and G. Karypis, "A comprehensive survey of neighborhood-based recommendation methods," in *Recommender Systems Handbook*. New York, NY, USA: Springer, 2011, pp. 107–144.
[6] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak, "Majorization–minimization algorithms for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2980–2991, Dec. 2007.
[7] M. Fukushima, "Application of the alternating direction method of multipliers to separable convex programming problems," *Comput. Optim. Appl.*, vol. 1, no. 1, pp. 93–111, Oct. 1992.
[8] S. Gleichman and Y. C. Eldar, "Blind compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6958–6975, Oct. 2011.
[9] A. Gogna and A. Majumdar. (2015). "Blind compressive sensing framework for collaborative filtering." [Online]. Available: http://arxiv.org/abs/1505.01621
[10] A. Gogna and A. Majumdar, "Blind compressive sensing formulation incorporating metadata for recommender system design," *APSIPA Trans. Signal Inf. Process.*, vol. 4, p. e2, Jul. 2015.
[11] Q. Gu, J. Zhou, and C. Ding, "Collaborative filtering: Weighted nonnegative matrix factorization incorporating user and item graphs," in *Proc. SDM*, 2010, pp. 199–210.
[12] T. Hofmann, "Latent semantic models for collaborative filtering," *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 89–115, Jan. 2004.
[13] M. F. Hornick and P. Tamayo, "Extending recommender systems for disjoint user/item sets: The conference recommendation problem," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 8, pp. 1478–1490, Aug. 2012.
[14] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, Nov. 2013.
[15] S. Kabbur, X. Ning, and G. Karypis, "FISM: Factored item similarity models for top-N recommender systems," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2013, pp. 659–667.

[16] Y. Koren, "Factorization meets the neighborhood: A multifaceted collaborative filtering model," in *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2008, pp. 426–434.
[17] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 8, pp. 30–37, Aug. 2009.
[18] Y. Koren and R. Bell, "Advances in collaborative filtering," in *Recommender Systems Handbook*. New York, NY, USA: Springer, 2011, pp. 145–186.
[19] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, vol. 161. Englewood Cliffs, NJ, USA: Prentice-Hall, 1974.
[20] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, pp. 556–562.
[21] H. Lee, J. Yoo, and S. Choi, "Semi-supervised nonnegative matrix factorization," *IEEE Signal Process. Lett.*, vol. 17, no. 1, pp. 4–7, Jan. 2010.
[22] B. Lika, K. Kolomvatsos, and S. Hadjiefthymiades, "Facing the cold start problem in recommender systems," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 2065–2073, Mar. 2014.
[23] S. G. Lingala and M. Jacob, "Blind compressive sensing dynamic MRI," *IEEE Trans. Med. Imag.*, vol. 32, no. 6, pp. 1132–1145, Jun. 2013.
[24] Q. Liu, E. Chen, H. Xiong, C. H. Q. Ding, and J. Chen, "Enhancing collaborative filtering by user interest expansion via personalized ranking," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 1, pp. 218–233, Feb. 2012.
[25] Q. Liu, E. Chen, H. Xiong, Y. Ge, Z. Li, and X. Wu, "A cocktail approach for travel package recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 2, pp. 278–293, Feb. 2014.
[26] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, "Recommender systems with social regularization," in *Proc. 4th ACM Int. Conf. Web Search Data Mining*, 2011, pp. 287–296.
[27] P. Massa and P. Avesani, "Trust-aware recommender systems," in *Proc. ACM Conf. Recommender Syst.*, 2007, pp. 17–24.
[28] A. Mnih and R. Salakhutdinov, "Probabilistic matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 1257–1264.
[29] A.-T. Nguyen, N. Denos, and C. Berrut, "Improving new user recommendations with rule-based induction on cold user data," in *Proc. ACM Conf. Recommender Syst.*, 2007, pp. 121–128.
[30] S.-T. Park, D. Pennock, O. Madani, N. Good, and D. DeCoste, "Naïve filterbots for robust cold-start recommendations," in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2006, pp. 699–705.
[31] A. L. V. Pereira and E. R. Hruschka, "Simultaneous co-clustering and learning to address the cold start problem in recommender systems," *Knowl.-Based Syst.*, vol. 82, pp. 11–19, Jul. 2015.
[32] Y. Ren, G. Li, J. Zhang, and W. Zhou, "Lazy collaborative filtering for data sets with missing values," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1822–1834, Dec. 2013.
[33] A. Gunawardana and G. Shani, "Evaluating recommendation systems," in *Recommender Systems Handbook*. New York, NY, USA: Springer, 2011, pp. 257–297.
[34] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Adv. Artif. Intell.*, vol. 2009, Jan. 2009, Art. ID 4.
[35] X. Tang, Y. Xu, and S. Geva, "Learning higher-order interactions for user and item profiling based on tensor factorization," in *Proc. 20th Int. Conf. Intell. User Interfaces*, 2015, pp. 213–224.
[36] A. Tiroshi, S. Berkovsky, M. A. Kaafar, D. Vallet, T. Chen, and T. Kuflik, "Improving business rating predictions using graph based features," in *Proc. 19th Int. Conf. Intell. User Interfaces*, 2014, pp. 17–26.
[37] M. Vozalis and K. G. Margaritis, "Collaborative filtering enhanced by demographic correlation," in *Proc. 18th World Comput. Congr. AIAI Symp. Prof. Pract. AI*, 2004, pp. 1–10.
[38] J. Wang, A. P. de Vries, and M. J. T. Reinders, "Unifying user-based and item-based collaborative filtering approaches by similarity fusion," in *Proc. 29th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2006, pp. 501–508.
[39] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1758–1789, 2013.
[40] P. Yu, L. Lin, R. Wang, J. Wang, and F. Wang, "A unified latent factor correction scheme for collaborative filtering," in *Proc. 11th Int. Conf. Fuzzy Syst. Knowl. Discovery (FSKD)*, Aug. 2014, pp. 581–586.
[41] Z.-K. Zhang, C. Liu, Y.-C. Zhang, and T. Zhou, "Solving the cold-start problem in recommender systems with social tags," *EPL (Europhys. Lett.)*, vol. 92, no. 2, p. 28002, Nov. 2010.
[42] Y. Zhang, W. Chen, and Z. Yin, "Collaborative filtering with social regularization for TV program recommendation," *Knowl.-Based Syst.*, vol. 54, pp. 310–317, Dec. 2013.
[43] T. Zhou, H. Shan, A. Banerjee, and G. Sapiro, "Kernelized probabilistic matrix factorization: Exploiting graphs and side information," in *Proc. SDM*, vol. 12. 2012, pp. 403–414.
[44] H. Zou, Z. Gong, N. Zhang, W. Zhao, and J. Guo, "TrustRank: A cold-start tolerant recommender system," *Enterprise Inf. Syst.*, vol. 9, no. 2, pp. 117–138, 2015.
[45] [Online]. Available: http://sifter.org/~simon/journal/20061211.html, accessed Jul. 2015.
[46] [Online]. Available: http://grouplens.org/datasets/movielens, accessed Nov. 2014.

**ANUPRIYA GOGNA** received the M.Tech. (ECE) degree from the Indira Gandhi institute of Technology Delhi, in 2012. She has been a Research Scholar with the Indraprastha Institute of Information Technology Delhi since 2013. She has authored several research papers in reputed journals and conferences. Her research interests include signal processing, metaheuristics, compressive sensing, and sparse signal recovery with a focus on matrix factorization and collaborative filtering algorithms.

**ANGSHUL MAJUMDAR** received the Ph.D. degree from the University of British Columbia, in 2012. He has been an Assistant Professor with the Indraprastha Institute of Information Technology Delhi since 2012. His works on a wide range of topics include biomedical signal and image processing, remote sensing, data analytics, and machine learning. His research focuses on algorithms and applications of sparse signal recovery and low rank matrix recovery. He has authored over 100 journal and conference papers since the inception of his research career in 2007. He has recently published a book entitled *Compressed Sensing-Based MRI Reconstruction* (Cambridge University Press).