

# Personalized Item Ranking from Implicit User Feedback: A Heterogeneous Information Network Approach

**Mukul Gupta**

Information Systems Area  
Indian Institute of Management Indore, India  
[mukulg@iimdr.ac.in](mailto:mukulg@iimdr.ac.in)

**Pradeep Kumar**

Information Technology & Systems Area  
Indian Institute of Management Lucknow, India  
[pradeepkumar@iiml.ac.in](mailto:pradeepkumar@iiml.ac.in)

**Bharat Bhasker**

Decision Science & Systems Area  
Indian Institute of Management Raipur, India  
[bhasker@imrapur.ac.in](mailto:bhasker@imrapur.ac.in)

## Abstract

*In today's era of the digital world with information overload, generating personalized recommendations for the e-commerce users is a challenging and interesting problem. Recommendation of top-N items of interest to a user of e-commerce is highly challenging using binary implicit feedback. The training data is usually very sparse and have binary values capturing a user's action or inaction. Due to the sparseness of data and lack of explicit user preferences, the recommendations generated by model-based and neighborhood-based approaches are not effective. Of late, network-based item recommendation methods, which utilize item related meta-information, are beginning to attract increasing attention for binary implicit feedback data. In this work, we propose a heterogeneous information network based recommendation model for personalized top-N recommendations using binary implicit feedback data. To utilize the potential of meta-information related to items, we utilize the concept of meta-path. To improve the effectiveness of the recommendations, the popularity of items and interest of users are leveraged simultaneously. Personalized weight learning of various meta-paths in the network is performed to determine the intrinsic interests of users from the binary implicit feedback data. To show the effectiveness, the proposed model is experimentally evaluated using the real-world dataset.*

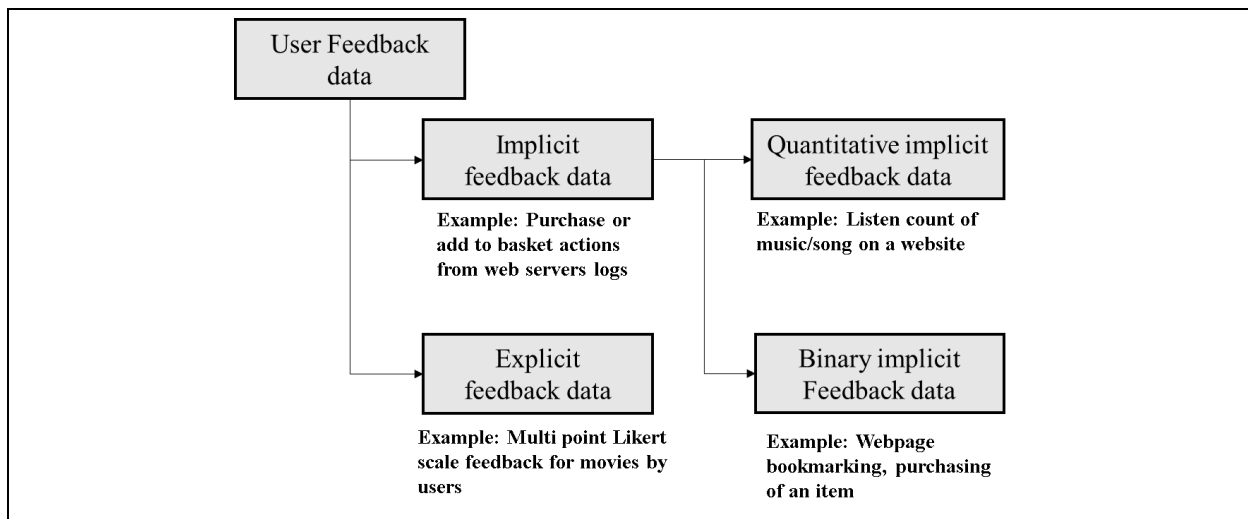
**Keywords:** recommendation system, binary implicit feedback, heterogeneous information network

Citation: Gupta, M., Kumar, P. and Bhasker, B. (2017). "Personalized Item Ranking from Implicit User Feedback: A Heterogeneous Information Network Approach," *Pacific Asia Journal of the Association for Information Systems*, 9(2), pp. 23-42.

## Introduction

Information filtering has become very crucial to the success of E-commerce for providing accurate recommendations to the users. In the last decade, research on personalized recommendation approaches has gained momentum in the industry as well as in academia (Berkovsky and Freyne, 2015). For recommendation generation, we can utilize explicit or binary implicit user feedback

(Aggarwal, 2016). A Model-based collaborative filtering technique like matrix factorization has effectively utilized explicit feedback data for recommendation generation (Koren et al., 2009). However, for binary implicit feedback data, the performance of matrix factorization based technique is poor due to the lack of explicit signals of interest of users (McFee et al., 2012). Figure 1 shows the broad classification of user feedback data.



**Figure 1 - Classification of user feedback data**

In the explicit feedback data, user preferences are usually mentioned in terms of ratings/stars specified on a scale to express the specific level of like or dislike for an item by a user (Aggarwal, 2016). A model-based filtering technique like matrix factorization can effectively utilize explicit feedback data for recommendation generation (Koren et al., 2009). However, in implicit feedback data, only the interactions between users and items are mentioned. The implicit feedback data can be further classified into two types namely, quantitative implicit feedback and binary implicit feedback. In quantitative implicit feedback, the interaction between a user and item is mentioned in terms of frequency of interaction e.g. frequency of listening to a song by a user on a website (Aggarwal, 2016; Bauer and Nanopoulos, 2014; Hu et al.,

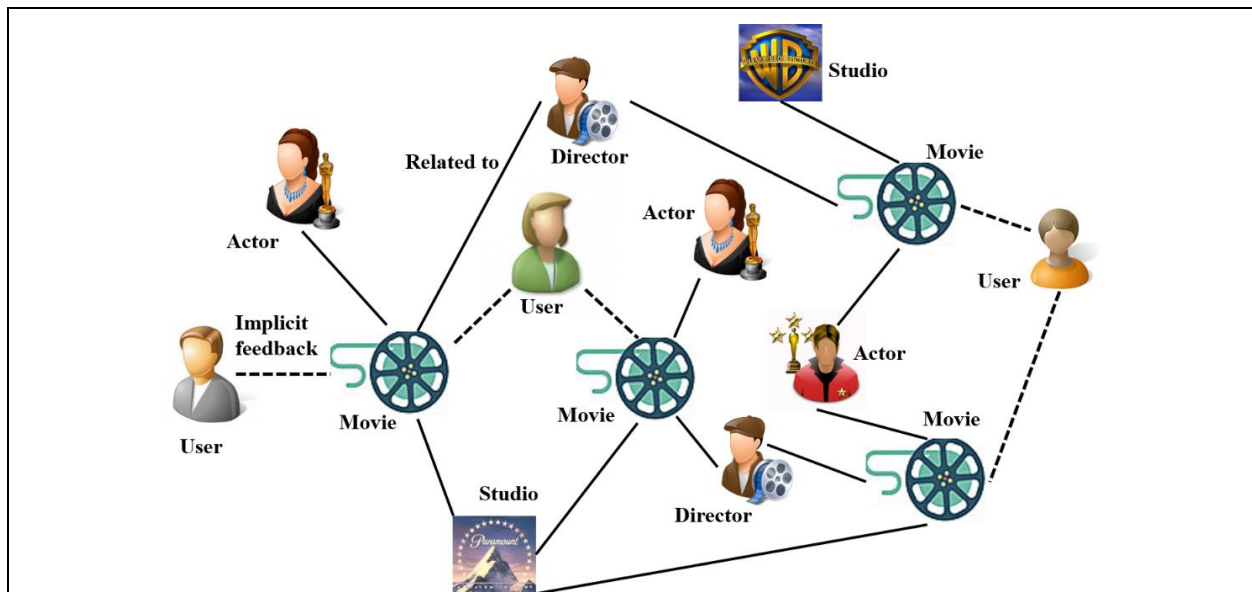
2008). If the frequency value of interaction between a user and item is zero then it represents the lack of interaction for that item. However, in the case of binary implicit feedback data, values are from the set  $\{1, 0\}$  where a 1 represents a user's interaction with the item and a 0 represents a lack of interaction (Yu et al., 2014; Rendle et al., 2009; Pan et al., 2008). Binary implicit feedback data is considered highly noisy as a 0 in the data is due to either negative feedback (user is not interested in the item) or unobserved interaction (user is not aware of the item for now) (Yu et al., 2014; Rendle et al., 2009; Pan et al., 2008).

Neighborhood-based filtering techniques have shown good performance for binary implicit feedback data as compared to matrix factorization based technique (McFee et al., 2012). User-based collaborative filtering

(UBCF) is one of the accurate neighborhood-based technique for generating recommendations. However, it has been experimentally proven that it is not scalable due to the time requirement that increases with the number of users as well as the number of items (Linden et al., 2003; Sarwar et al., 2001). The number of users for an E-commerce website is usually high and typically increases rapidly that makes UBCF impractical when recommendations are to be generated in real time (Linden et al., 2003; Sarwar et al., 2001). Item-based collaborative filtering (IBCF) can solve this problem of scalability, and it generates accurate recommendations that make IBCF preferred over UBCF (Deshpande and Karypis, 2004; Linden et al., 2003). The reason of IBCF to be scalable is that it can be used as a model based approach. The item-item similarity matrix is computed offline and utilized as a pre-computed model for generating recommendations in real time (Linden et al., 2003; Deshpande and Karypis, 2004). Since the number of items is less as compared to the number of users for a typical E-commerce website; therefore, model based IBCF is suitable in practical situations. Though, the model based IBCF is faster and scalable but the quality of recommendations is significantly hampered if the data used for

training of the model is sparse (Bobadilla et al., 2013).

In this work, to deal with the data sparsity problem for binary implicit feedback data, we form a heterogeneous item information network to incorporate the meta-information related to items to be recommended using meta-path based framework. Using the heterogeneous item information network, we can measure the similarity/relatedness between items using meta-information rather than the sparse user-item binary implicit feedback matrix. Figure 2 shows an example snippet of a heterogeneous movie information network. The meta-information related to movies like directors, actors and studios have been incorporated to form a heterogeneous information network. Users are related to movies using binary implicit feedback that has been shown as dashed lines. Solid lines in Figure 2 shows the various relationships between movies and meta-information related to movies. Using this heterogeneous movie information network, we can measure the similarity between movies following different paths. Each path has a semantic meaning and can be useful for measuring the relatedness between items.



**Figure 2 - Example of heterogeneous movie information network**

By forming a heterogeneous item information network, we can deal with the problem of data sparseness. However, to improve the quality of recommendations, we need to determine the intrinsic interest of users from the binary implicit feedback data. For determining the intrinsic interest of users, we perform the personalized weight learning of different meta-paths to find those paths that represent the interest of users. It has been shown that users would experience/purchase those items first which are as per their interest (Bobadilla et al., 2013). Also, in a study by Barman and Dabeer (2012), it has been shown that items that are highly popular have good chance to be experienced/purchased. In this work, we leveraged simultaneously the popularity of items and intrinsic interest of users to generate accurate recommendations for users.

In this work, the focus is on utilizing the binary implicit feedback data for recommendation generation. In the case of binary implicit feedback data, due to the lack of explicit signals of user preferences, the recommendation task is more challenging. The majority of the earlier research works utilized explicit feedback data due to the convenience of getting explicit signals of user preferences that makes recommendation task easier as compared to the case of binary implicit feedback data (Yu et al., 2014; Bobadilla et al., 2013; McFee et al., 2012; Hu et al., 2008). However, in most of the practical situations, users are unwilling to devote time to giving explicit ratings (Demiriz, 2004). Also, the explicit ratings, given by the users, are not consistent. Another reason for considering binary implicit feedback data in this work is its easy availability and accessibility from user logs (Demiriz, 2004). The contributions of this paper may be summarized as follows. For binary implicit feedback data, a heterogeneous item information network based approach is proposed, which is scalable and generates accurate personalized recommendations as compared to the model based IBCF. The

popularity of items and interest of the user are simultaneously leveraged in the proposed model to improve the quality of personalized recommendations for users. This work proposes a technique to perform the personalized weight learning to determine the intrinsic interest of users and integrate the semantics of various meta-paths for generating accurate personalized recommendations.

The rest of the paper is organized as follows. The next section presents the related work on the generation of recommendations using binary implicit feedback data, followed by background for the heterogeneous item information network. Then, the proposed recommendation model for personalized item ranking for users is presented. After that, the experimental setup and results are discussed. Finally, the paper concludes with a discussion.

## **Related Work**

Matrix factorization, for recommendation generation using explicit feedback data, is a popular model based technique and has been studied extensively (Koren et al., 2009). Hu et al. (2008) and Pan et al. (2008) have proposed matrix factorization based filtering techniques for implicit feedback data. However, their approaches for implicit feedback data work on the pointwise preference assumption and optimized for predicting whether an item will be selected by a user. For the task of generating top-*N* recommendations, recently, Rendle et al. (2009) have proposed Bayesian Personalized Ranking based Matrix Factorization (BPRMF) methodology for binary implicit feedback data. Their methodology is based on the pairwise-preference assumption and optimized for ranking of items. BPRMF has shown better performance for ranking of items as compared to the methodology based on pointwise-preference assumption (Rendle et al., 2009). Of late, it has been shown by McFee et al. (2012) that a neighborhood-based filtering technique like item-based

collaborative filtering (IBCF) is more effective than BPRMF for binary implicit feedback data. However, for highly sparse binary implicit feedback data, the effectiveness of recommendations generated by IBCF is affected adversely (Bobadilla et al., 2013).

For improving the effectiveness of recommendations and dealing the problem of the sparseness of binary implicit feedback data, researchers have incorporated the different types of information related to items to be recommended (Li et al., 2013; Basilico and Hofmann, 2004). However, recently, researchers are more interested in extracting the knowledge from the heterogeneous linked information related to items rather than the content information (Gan, 2016; Shi et al., 2015; Yu et al., 2014; Bellogin et al., 2013). Recently, the effective and information rich mining results using the heterogeneous information networks (Gupta et al., 2017; Gupta et al., 2015; Sun and Han, 2013) have attracted the attention of researchers for exploring their applicability in the domain of recommendation generation (Yu et al., 2014; Shi et al., 2015).

In this paper, the focus is on the applicability of the heterogeneous information network for information filtering to generate the effective top- $N$  personalized recommendations. Also, using the heterogeneous information network, the problem of data sparsity is handled for binary implicit feedback data. Recently, Yu et al. (2014) and Shi et al. (2015) formed the heterogeneous information network using the meta-information related to items and performed the information filtering for recommendation generation. Yu et al. (2014), in their work, utilized binary implicit feedback data for recommendation generation; however, Shi et al. (2015) utilized explicit feedback data in their work for recommendation generation.

For recommendation generation, user preference diffusion based methodology was utilized by Yu et al. (2014) for different meta-paths in the heterogeneous information network for items. For the training of their proposed model, they utilized Bayesian

ranking-based optimization methodology (Rendle et al., 2009). For personalization of the recommendations, in their model, they performed clustering of users. However, their work has some limitations. First, for preference diffusion along various paths in the network, they utilize PathSim which can utilize only symmetric meta-paths in a network (Sun et al., 2011). Therefore, their model cannot utilize the full potential of the linked information present in the heterogeneous information network as there would be various asymmetric meta-paths in the network. Further, for personalization of the recommendations, their model requires clustering of users. However, finding the appropriate number of clusters for a dataset is a challenging and difficult problem and the effectiveness of the recommendations from their model significantly depends on the quality of the clusters created. This makes their recommendation model difficult to apply for real-world applications.

Considering the limitations of the aforementioned approaches, in this paper, we form a heterogeneous information network for items and fully utilize the potential of the linked information present in the network by considering symmetric and asymmetric meta-paths. Using this network, personalized recommendations are generated for binary implicit feedback. In this paper, we have proposed a recommendation model with a personalized weight learning methodology and demonstrated its generalizability by comparing it with different techniques using a real-world dataset MovieLens-100K. Also, in this work, we have incorporated the meta-path generation methodology (Gupta et al., 2017) which alleviates the need for manual meta-path enumeration (Gupta et al., 2016).

To determine the intrinsic interest of users, personalized weight learning of various meta-paths in the network is to be performed. In this work, a regularization based methodology for personalized weight learning is proposed which is effective to determine the intrinsic interest of users. By personalized weight learning, we can

understand the importance of various meta-paths and accordingly personalized recommendations can be generated for users. For example, to generate personalized recommendations for movies, the interest of users should be considered as one user may be the fan of a particular actor while another user may like movies of a particular genre. Therefore, the semantics of various meta-paths should be considered and personalized weight assignment should be performed while generating recommendations.

## Background

In this section, we present the background for this work. We also present the formal definition of personalized item ranking using the heterogeneous item information network. Some frequently used notations have been listed in Table 1.

In this work, we utilized binary implicit feedback matrix as defined in Definition 1.

Table 1 - Important Notations	
Notation	Description
$U = \{u_1, u_2, \dots, u_m\}$	Set of all $m$ users
$I = \{e_1, e_2, \dots, e_n\}$	Set of all $n$ items
$Q_L$	Binary implicit feedback matrix for training
$G, T_G$	Heterogeneous information network (HIN) for items and its schema
$A, R$	Set of object types and set of relations in HIN
$P = \{P_1, P_2, \dots, P_K\}$	Set of $K$ meta-paths from HIN
$Sim(a_i, b_j   P_q)$	Relatedness between object $a_i$ and $b_j$ in HIN following meta-path $P_q$
$Pop(e_j)$	Popularity of item $e_j$ in the training data
$W^{ U  \times  P }$	Personalized weight matrix for meta-paths
$B_{u_i} = [\beta_1, \dots, \beta_K]$	Importance of the set of meta-paths $P = \{P_1, P_2, \dots, P_K\}$ for user $u_i$
$RS^{ U  \times  I }$	Personalized recommendation score matrix

**Definition 1 (Binary implicit feedback matrix).** Binary implicit feedback matrix  $Q \in \mathbb{R}^{m \times n}$  for  $m$  users  $U = \{u_1, u_2, \dots, u_m\}$  and  $n$  items  $I = \{e_1, e_2, \dots, e_n\}$  is defined as

$$Q_{ij} = \begin{cases} 1, & \text{if } (u_i, e_j) \text{ interaction is observed} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

In the binary implicit feedback matrix, the value 1 represents that the user has interacted with the item. For example, in case of movies,  $Q_{ij} = 1$  implies that the user  $u_i$  has watched the movie  $e_j$ . However,  $Q_{ij} = 0$  implies no interaction. The value 1 represents that the user has experienced that item but it does not necessarily mean that the feedback is positive. The feedback may be positive or negative. It just shows the interaction between a user and item. The value 0 does not imply a negative feedback

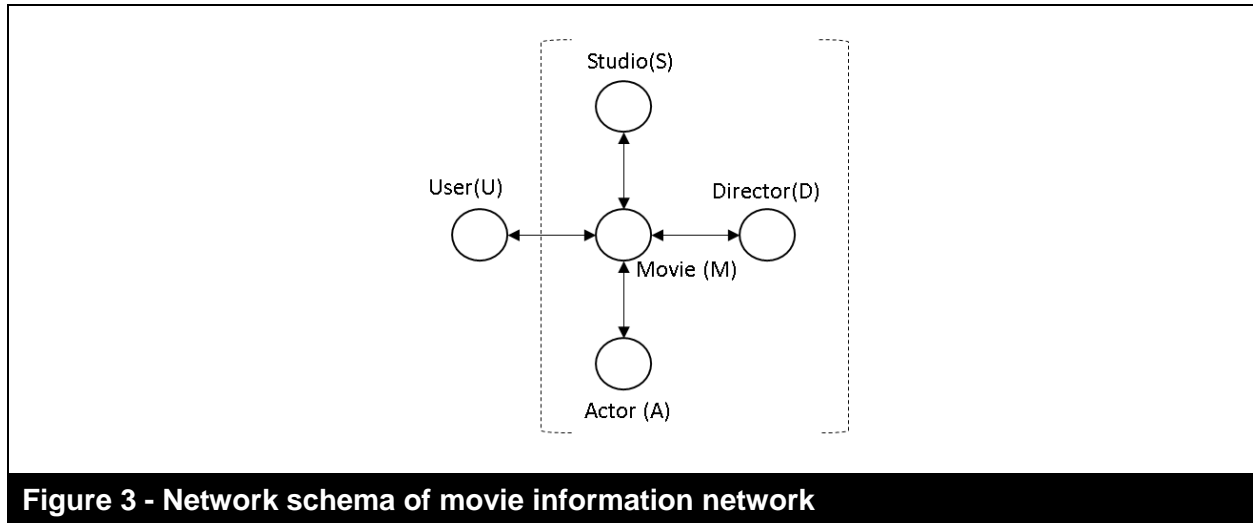
but it shows the lack of interaction between a user and item.

We define information network and network schema as follows which is similar to the definition by Sun et al. (2011).

**Definition 2 (Information Network and Network Schema).** An information network is defined as a directed graph  $G = (V, E)$  with an object type mapping function  $\phi: V \rightarrow A$  and a link type mapping function  $\psi: E \rightarrow R$ , where each object  $v \in V$  belongs to a particular object type  $\phi(v) \in A$ , and each link  $e \in E$  belongs to a particular relation  $\psi(e) \in R$ . However, the network schema, denoted as  $T_G = (A, R)$ , is the meta-level representation of  $G = (V, E)$  and it is a directed graph over object types  $A$  and relations  $R$ .

When the types of objects  $|A| > 1$  or the number of relations  $|R| > 1$ , the network is heterogeneous information network. In this work, we form a heterogeneous information network for items to be recommended. We utilize meta-information related to items to form the network and it is termed as

heterogeneous item information network. Figure 2 shows the snippet of a heterogeneous information network for movies. It includes meta-information related to movies like actors, director, studios. The network schema of this information network is shown in Figure 3.



**Figure 3 - Network schema of movie information network**

This network schema has four types of nodes i.e. movie, actor, director, and studio. Users have also been included in this network as they have watched the movies. Thus the whole network consist of five different types of nodes and four different types of bidirectional relationships between nodes. A path on a network schema is termed as meta-path. Now, we formally define the problem of recommendation generation using meta-path based framework in the heterogeneous network as follows.

**Definition 3 (Problem Definition).** For the given binary implicit feedback data  $Q$  with heterogeneous information network  $G$  for items, the objective is to online generate a personalized ranked list of top- $N$  items for an active user  $u$  after offline forming a personalized recommendation model for the user.

So, in this work, we design a recommendation model which utilizes the meta-information related to items and generates online personalized

recommendations for users after training of the model.

## Meta-path Based Framework for Personalized Item Ranking

For recommendation generation, first, a heterogeneous item information network is formed that incorporates meta-information related to items. Then, following different meta-paths, relatedness is measured for users and items. The relatedness value for those items will be high which are as per the interest of an active user. For relatedness calculation, we first define the meta-path similar to Sun et al. (2011).

**Definition 4 (Meta-path).** A meta-path  $P$  is a path defined on network schema  $T_G = (A, R)$  and denoted in the form of  $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \dots \xrightarrow{R_l} A_{l+1}$  that defines a composite relation  $R = R_1 \circ R_2 \circ \dots \circ R_l$  of length  $l$  between source object type  $A_1$  and target object type  $A_{l+1}$  using composition operator  $\circ$  on relations.

When there is no ambiguity regarding the relationship between two object types then we can represent the meta-path  $P$  using only object types as  $P = (A_1 A_2 \dots A_{l+1})$ . For example, in Figure 3, several meta-paths exist like *MSM* (*movie – studio – movie*), *MDM* (*movie – director – movie*), and *UMAM* (*user – movie – actor – movie*). Each meta-path has a semantic meaning and can be utilized for the relatedness measurement between objects. For example, meta-path *MDM* links those movies that have been

directed by same director. Using this meta-path we can calculate the relatedness between movies on the basis of director using DPRel (Gupta et al., 2015). Likewise, we can explain the semantic meaning of other meta-paths and utilize them for relevance measurement.

### Model Description

The proposed recommendation model consist of two modules: offline and online module as shown in Figure 4.

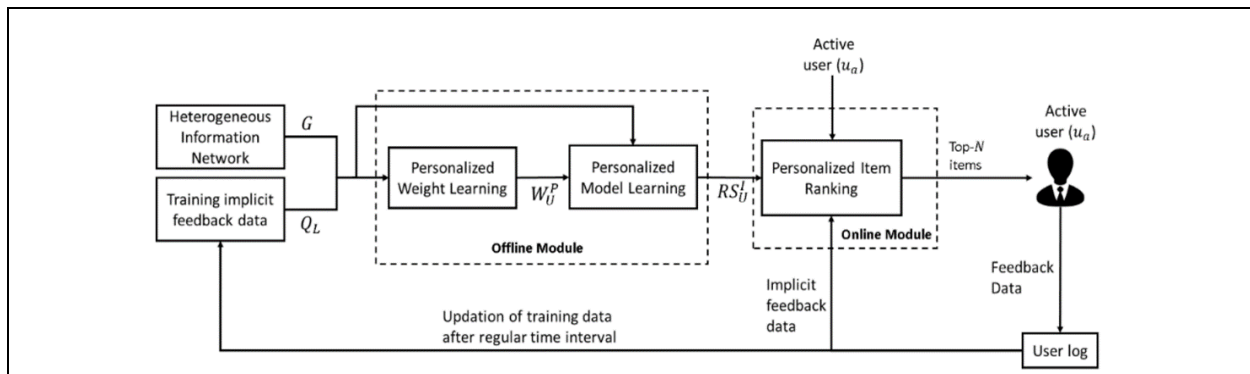


Figure 4 - Proposed Recommendation Model

Offline module performs two tasks, namely, personalized weight learning for different meta-paths and personalized model learning for inexperienced items for each user. After learning the personalized weights of different meta-paths, the output of the personalized weight learning is supplied as input to the personalized model learning. The online module generates personalized recommendations of top- $N$  items using the trained personalized model for an active user using the feedback data of that user from the user log. Feedback data from the user log is used to update the offline module.

However, immediate update of the offline module is not done because it is time intensive to retrain the whole model. The delayed update would save the time with a little reduction in quality of recommendations as suggested by Linden et al. (2003). We can retrain the model (offline module) after regular time intervals.

### Personalized Weight Learning (PWL)

PWL determines the personalized weights of different meta-paths for each user. For a user, the importance of a meta-path may be higher as compared to other meta-paths. For example, a user may be the fan of a particular actor and views most of those movies in which that actor has performed. In that case, the importance of the meta-path *MAM* (*movie – actor – movie*) would be high for that user. Therefore, it is important to determine the weight of each meta-path for a user.

After determining the personalized weights of meta-paths for a user, we can effectively recommend items to that user. For personalized weight learning, we consider the items that have already been experienced by a user (training data). From that set of items, we can perform supervised learning using Algorithm-1 shown in Figure 5 to determine the weights of various meta-paths for that user.



Let  $B_u = [\beta_1, \beta_2, \dots, \beta_K]$  be the vector of importance for different meta-paths such that  $\beta_k$  is the importance of the meta-path  $P_k$  for the user  $u$  and  $\beta_k \geq 0, \forall k$ . For a user, we can find  $B_u = [\beta_1, \beta_2, \dots, \beta_K]$  by solving the following optimization problem:

$$B_u^* = \arg \min_{B_u=[\beta_1, \beta_2, \dots, \beta_K]} L(B_u) \quad (2)$$

Here,  $L(B_u)$  is the regularization based cost function defined in Equation (3). The first term in Equation (3) ensures that the importance of each meta-path follows the set of items already experienced by the user.

The second term is the smoothness constraint in Equation (3).

$$\begin{aligned} L(B_u) &= \frac{1}{2} \sum_{e_i, e_j \in T_u} \left\| R(e_i, e_j) \right. \\ &\quad \left. - \sum_{k=1}^K \beta_k R(e_i, e_j) \text{Sim}(e_i, e_j | P_k) \right\|_2^2 \\ &\quad + \frac{\lambda}{2} \|B_u\|_2^2 \\ \text{s. t. } &\beta_k \geq 0, \forall k = 1, \dots, K \end{aligned} \quad (3)$$

### Input:

$P = \{P_1, P_2, \dots, P_K\}$ : meta-paths from  $G$  in format of *item* – \* – *item*

$Q_L$ : training data

$\lambda$ : regularization parameter

$\text{Sim}(a, b | P_k)$ : relatedness between objects  $a$  and  $b$  following meta-path  $P_k, \forall k = 1, \dots, K$

$\epsilon$ : convergence tolerance

### Output:

$W_U^P$ : personalized weights of meta-paths

### Begin

1. **foreach** user  $u_i$  **do**
2.     initialize  $B = [\beta_1, \beta_2, \dots, \beta_K] > 0$
3.     **repeat**
4.          $B_{old} = B$ ;
5.         Calculate  $\beta_k, \forall k = 1, \dots, K$  using Equation (6)
6.          $\beta'_k = \max(0, \beta_k), \forall k = 1, \dots, K$ ;
7.          $B = [\beta'_1, \beta'_2, \dots, \beta'_K]$  using  $\beta'_k, \forall k = 1, \dots, K$
8.     **until**  $|B - B_{old}| \leq \epsilon$
9.     **for**  $q \leftarrow 1$  to  $K$  **do**
10.          $W_{u_i}^{P_k} = \frac{\beta_k}{\sum_k \beta_k}$
11.     **end for**
12. **end for**

### End

Figure 5 - Algorithm-1 for Personalized Weight Learning

Here,  $T_u$  is the set of items already experienced by the user  $u$  (training data for user  $u$ ),  $\text{Sim}(e_i, e_j | P_k)$  is the relatedness

between items  $e_i$  and  $e_j$  following meta-path  $P_k$ , and  $\lambda$  is the regularization parameter. For measuring the relatedness between items, we use relevance measure

DPRel (Gupta et al., 2015). The function  $R(e_i, e_j)$  is defined as follows:

$$R(e_i, e_j) = \begin{cases} 1, & \text{if } e_i, e_j \in T_u \text{ and } i \neq j \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

The pseudo code of the algorithm for PWL is shown in Figure 5. This algorithm first determines the importance of each meta-path for a user as per the interest and then normalizes the value to generate the weights for each meta-path. To solve the Equation (2), we take the partial derivative of  $L(B_u)$  w.r.t each  $\beta_k$  and set those derivatives to 0 to obtain  $\beta_k, \forall k$  i.e.

$$\frac{\partial L(B_u)}{\partial \beta_k} = 0 \quad (5)$$

After solving the Equation (5), we find the value of  $\beta_k, \forall k$  as given in Equation (6),

$$\beta_k = \frac{\sum_{e_i, e_j \in T_u} R^2(e_i, e_j) \text{Sim}(e_i, e_j | P_k) f(e_i, e_j)}{\lambda + \sum_{e_i, e_j \in T_u} R^2(e_i, e_j) \text{Sim}^2(e_i, e_j | P_k)} \quad (6)$$

Where  $f(e_i, e_j) = (1 - \sum_{r \neq k} \beta_r \text{Sim}(e_i, e_j | P_r))$

Then we apply Algorithm-1 to learn  $B_u = [\beta_1, \beta_2, \dots, \beta_K]$ . After finding  $B_u^*$  for the user, we find the weights for different meta-paths for the user by normalizing the value of importance of meta-paths. Let  $W \in \mathbb{R}^{|U| \times |P|}$  be the weight matrix such that an entry, denoted as  $W_{u_i}^{P_k}$ , is the weight of the meta-path  $P_k$  for the user  $u_i$  and  $\sum_k W_{u_i}^{P_k} = 1$ . Then, using Equation (7), we can find the weights of meta-paths for a user.

$$W_{u_i}^{P_k} = \frac{\beta_k}{\sum_k \beta_k} \quad (7)$$

After learning the personalized weights of meta-paths, we supply the output of PWL i.e.  $W_{ij}^P$  to personalized model learning to learn the personalized recommendation scores for items.

### Personalized Model Learning (PML)

PML is used for generating personalized recommendation scores for items. The trained model, which is the output of this

algorithm, is the sorted recommendation scores (in descending order) for inexperienced items for each user. The PML takes as input the personalized weights for different meta-paths from PWL. The pseudo code of the Algorithm-2 for PML is shown in Figure 6. This algorithm takes into consideration the interest of the active user and popularity of items. The popularity of item  $e_j$  is defined in Equation (8).

$$\text{Pop}(e_j) = \sum_{i=1}^m Q_L[u_i, e_j] \quad (8)$$

Where  $Q_L[u_i, e_j] = 1$  if user  $u_i$  has experienced the item  $e_j$ , otherwise it is zero.

The interest of the active user is captured by the personalized weights for meta-paths. If the weight of a meta-path is high, then that meta-path explains the intrinsic interest of a user in a better way than other meta-paths. To find the items that are related to the previously experienced items by the user, we measure the relatedness scores between all inexperienced items and the active user following different meta-paths taken into consideration. The higher the relatedness value for an item, the more closely related that item is to the user's previously experienced items. For an active user and item pair, the recommendation score is generated by taking the sum of weighted products of item popularity and relatedness score following all meta-paths taken into consideration. Finally, for each user, all inexperienced items are sorted in descending order of their recommendation scores and stored as a pre-computed model. The output of the PML is used as an input to the online module for generating personalized item ranking.

To find the items that are as per the interest of a targeted user, we need to calculate the relatedness between the targeted user and items following a specified meta-path. For that, we utilize the relevance measure DPRel (Gupta et al., 2015).

**Input:**

$P = \{P_1, P_2, \dots, P_K\}$ : meta-paths from  $G$  in format of *user – item – \* – item*

$Sim(a, b|P_k)$ : relatedness between objects  $a$  and  $b$  following meta-path  $P_k, \forall k = 1, \dots, K$

$Q_L$ : training data

$W_U^P$ : personalized weights of meta-paths for users

**Output:**

$RS_U^I$ : recommendation score matrix

**Begin**

1. **for all**  $u_i$  and  $e_j$  **do**

2.      $RS_{u_i}^{e_j} = 0$  // initialization

3.     **for**  $q \leftarrow 1$  to  $K$  **do**

4.          $RS_{u_i}^{e_j} = RS_{u_i}^{e_j} + W_{u_i}^{P_q} \times Pop(e_j) \times Sim(u_i, e_j|P_q)$

5.     **end for**

6. **end for**

7. **for all**  $u_i$  **do**

8.      $RS_{u_i}^I = (1 - Q_L[u_i, *]) \times RS_{u_i}^I$

9.      $RS_{u_i}^I = sort(RS_{u_i}^I)$

10. **end for**

**End**

**Figure 6 - Algorithm-2 for Personalized Model Learning**

### Personalized Item Ranking (PIR)

PIR is used in the online module to generate personalized recommendations of top- $N$  items in constant time for an active user. It takes as input the learned model from PML. For an active user, it produces a list of top- $N$  items and recommends those items to that user. PIR takes into consideration the binary implicit feedback data from user log to generate the list of recommended items for a targeted user. Typically, the number of items

recommended are 10 to 15 for a targeted user. Thus, the online module of the proposed system produces recommendations in constant time. The pseudo code of Algorithm-3 for PIR is shown in Figure 7. In the algorithm for PIR, if a recommended item is experienced by the active user, then the recommendation score of that item is updated to zero and accordingly the recommend list of items is updated instantly.

**Input:**

Q: implicit feedback data

 $u_a$ : active user $RS_{u_a}^I$ : recommendation score of active user for items**Output:** $R_{u_a}^N$ : top-K recommended items for active user**Begin**

```

1. for user  $u_a$  do
2.    $RS_{u_a}^{IN} = RS_{u_a}^I[1, \dots, N]$ 
3.    $tQ[u_a, I_N] = 1 - Q[u_a, I_N]$ 
4.    $RS_{u_a}^{IN} = RS_{u_a}^{IN} \times tQ[u_a, I_N]$ 
5.    $c \leftarrow 0, p \leftarrow NULL$ 
6.   for  $r \leftarrow 1$  to  $N$  do
7.     if ( $RS_{u_a}^{e_r} = 0$ )
8.        $c = c + 1$ 
9.        $p = p \cup r$ 
10.    end if
11.  end for
12.   $R_{u_a}^N = RS_{u_a}^I[(1, \dots, N + c) - p]$ 
13. end for
End

```

**Figure 7 - Algorithm-3 for Personalized Item Ranking**

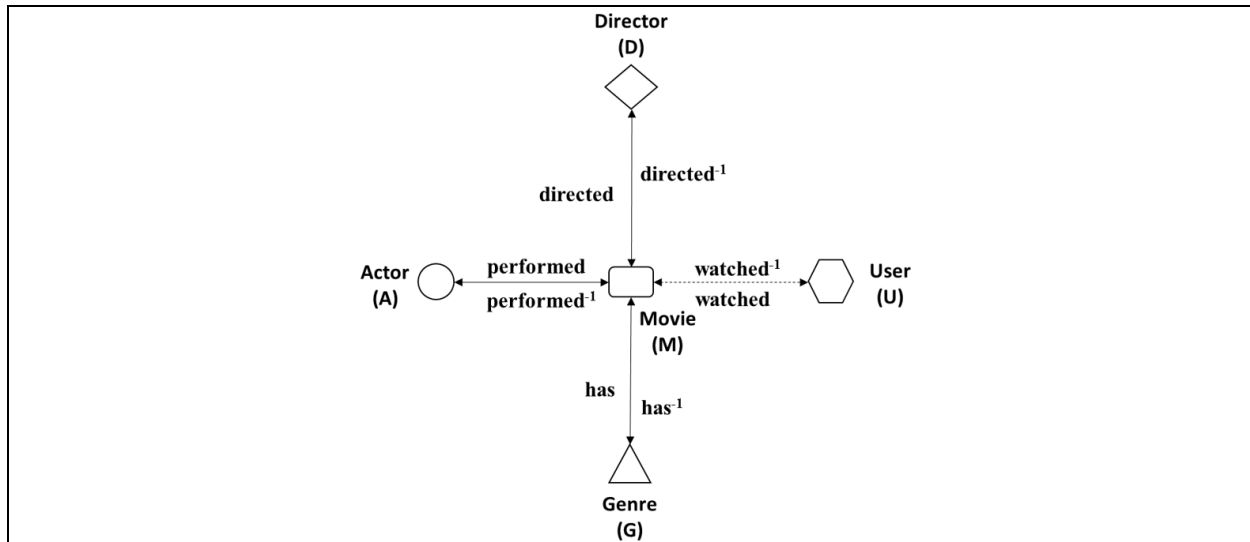
## Experimental Setup and Results

In this section, we discuss the experimental setup and results of experiments. All experiments were performed on a system with Intel Core i5 processor and 4 GB RAM using R version 3.3.1.

### Dataset

For experiments, we utilized MovieLens-100k dataset (<http://grouplens.org/datasets/movielens/>) to form the heterogeneous item information network and to show the effectiveness of the proposed approach (Yu et al., 2014). The information related to movies like actors, directors and genres are incorporated to form the heterogeneous item information network for

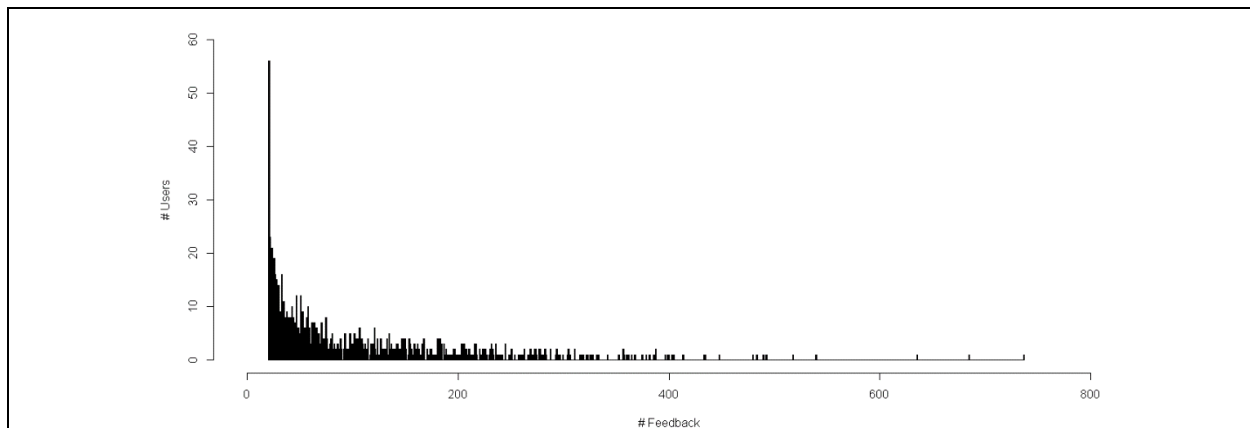
this dataset. This information is collected from IMDB site (<http://www.imdb.com/>). For each movie in the dataset, we collected top-5 actors according to their credits in the movie. The final heterogeneous item information network contains total 8,526 objects (represented as nodes in the network) and 112,866 links. In this heterogeneous network, there are 5 types of objects i.e. 943 users, 1,682 movies, 4,818 actors, 1,064 directors and 19 genres. The number of relations in this network is four i.e. between movies and actors (*movie – actor*), movies and directors (*movie – director*), movies and genres (*movie – genre*), and between movies and users (*movie – user*). All these relations are bidirectional. The network schema of this heterogeneous network is shown below in Figure 8.



**Figure 8 - Network Schema for movie dataset**

This dataset has  $10^5$  ratings. For empirical study in this work, these real ratings are considered to form a corresponding binary implicit feedback matrix  $Q$  as defined in Equation (1). If a user has watched a movie,

whether liked it or not, that user-movie pair has 1 in the binary implicit feedback matrix. Otherwise, for that user-movie pair, the binary implicit feedback matrix has 0. The distribution of users' feedback in this dataset is shown in Figure 9.



**Figure 9 - Distribution of feedback for movie dataset**

For evaluation purpose, we split the binary implicit feedback matrix  $Q$  into training and test matrices. Following 80% - 20% splits of  $Q$  into training and test data, five train ( $train_i, i = 1, \dots, 5$ ) and test matrices ( $test_i, i = 1, \dots, 5$ ) are created where all  $test_i$  are disjoint. The system is, first, trained using a train matrix ( $train_i$ ), then, testing is performed using the corresponding test

matrix ( $test_i$ ). This process is performed for all train and test matrix pairs.

### Meta-paths

For experiments, instead of manual enumeration of meta-paths, we utilized the methodology proposed in Gupta et al. (2017) for generation of meta-paths. The meta-path generation algorithm (Gupta et al., 2017) takes into consideration the network schema

and generates all the meta-paths starting and ending with the specified nodes up to the specified length. In our experiments, for the network schema shown in Figure 8, the algorithm produces nine different meta-paths in the format of  $item - * - item$  to compute the relatedness between items for PWL (considering only the movie and its meta-information part). Meta-paths  $MDM$ ,  $MAM$ ,

and  $MGM$  are symmetric. However, meta-paths  $MDMAM$ ,  $MGMAM$ ,  $MDMGM$ ,  $MAMDM$ ,  $MAMGM$ , and  $MGMMDM$  are asymmetric. For PML, the meta-paths are taken in the format of  $user - item - * - item$  i.e. we compute the relatedness between users and items. All these meta-paths are listed in Table 2.

**Table 2 - Meta-paths for movie dataset**

Meta-path ( $P$ )	Notation
$user - (movie - director - movie)$	$U(MDM)$
$user - (movie - actor - movie)$	$U(MAM)$
$user - (movie - genre - movie)$	$U(MGM)$
$user - (movie - director - movie - actor - movie)$	$U(MDMAM)$
$user - (movie - genre - movie - actor - movie)$	$U(MGMAM)$
$user - (movie - director - movie - genre - movie)$	$U(MDMGM)$
$user - (movie - actor - movie - director - movie)$	$U(MAMDM)$
$user - (movie - actor - movie - genre - movie)$	$U(MAMGM)$
$user - (movie - genre - movie - director - movie)$	$U(MGMMDM)$

In this work, we have taken director, actor, and genre as the meta-information for movies as in typical situations these are the factors that determine the user's selection for a movie (Yu et al., 2014; Shi et al., 2015). However, one can incorporate other meta-information related to movies and form the network schema to generate a large number of meta-paths. For experiments, we considered only those meta-paths that are of length not more than 4 (considering only the movie and its meta-information part in the network schema as shown in Figure 8). For example, meta-path  $MDMAM$  has length 4. The reason for not taking longer paths is that as we increase the length of a meta-path, the number of noisy relations would increase between objects following that meta-path and that meta-path would not be semantically correct (Yu et al., 2014). We also do not consider those meta-paths that have identical sub-paths as these meta-paths would be noisy (Kong et al., 2012). For example, meta-path  $MAMAM$  ( $movie - actor - movie - actor - movie$ ) is a noisy path as it has two identical sub-paths i.e.  $MAM$  ( $movie - actor - movie$ ). Therefore, we have not

considered this path for recommendation generation.

### Algorithms for Comparison and Performance Evaluation

The performance of the proposed method is compared with the following state-of-the-art techniques that are applicable to binary implicit feedback data to show the effectiveness of the proposed approach.

**Popularity:** In Popularity based recommendation, the most popular items are recommended to the users.

**Neighborhood Model:** We utilize IBCF for recommendations as this technique is scalable and has experimentally shown better accuracy than UBCF in many cases (Deshpande and Karypis, 2004).

**Association Rule Model:** Association rule (AR) based recommendation model develops, first, a set of association rules, and then apply those rules while generating recommendations (Demiriz, 2004).

**Bayesian Personalized Ranking based Matrix Factorization (BPRMF):** This model is optimized for personalized item ranking task

for binary implicit feedback data (Rendle et al., 2009).

*HeteRec*: It is a user preference diffusion score based technique. We utilize both the global and personalized recommendation technique (Yu et al., 2014) for comparison.

For the first three recommendation techniques mentioned above, we utilized the “*recommenderlab*” package in R (Hahsler, 2011) and for BPRMF we utilized the “*rrecsys*” package in R (Ludovik, 2016). To compare the proposed method with aforementioned techniques, we utilized all the meta-paths listed in Table 2. However, we compare the performance of the proposed method separately with HeteRec by utilizing only symmetric meta-paths as HeteRec cannot utilize asymmetric meta-paths.

The confidence of users on a recommender system depends directly on the accuracy of the list of recommended items (Bobadilla et al., 2013; Bellogín et al., 2013). Since in this work, the binary implicit feedback dataset is utilized, and the objective is to generate a list of top- $N$  recommended items; therefore, a relevance based metric would be suitable for evaluation purpose. We consider, for an active user, relevant items are those items that belong to his test set (items that have 1 in the test set for the active user) (Bellogín et al., 2013). Precision is the widely used relevance based measures to evaluate the recommendation quality of a recommendation system. Since, in this work, only top- $N$  recommended items are taken into consideration for a user; therefore we utilized  $Precision@N$  (Bobadilla et al., 2013) that is defined in Equation (9).

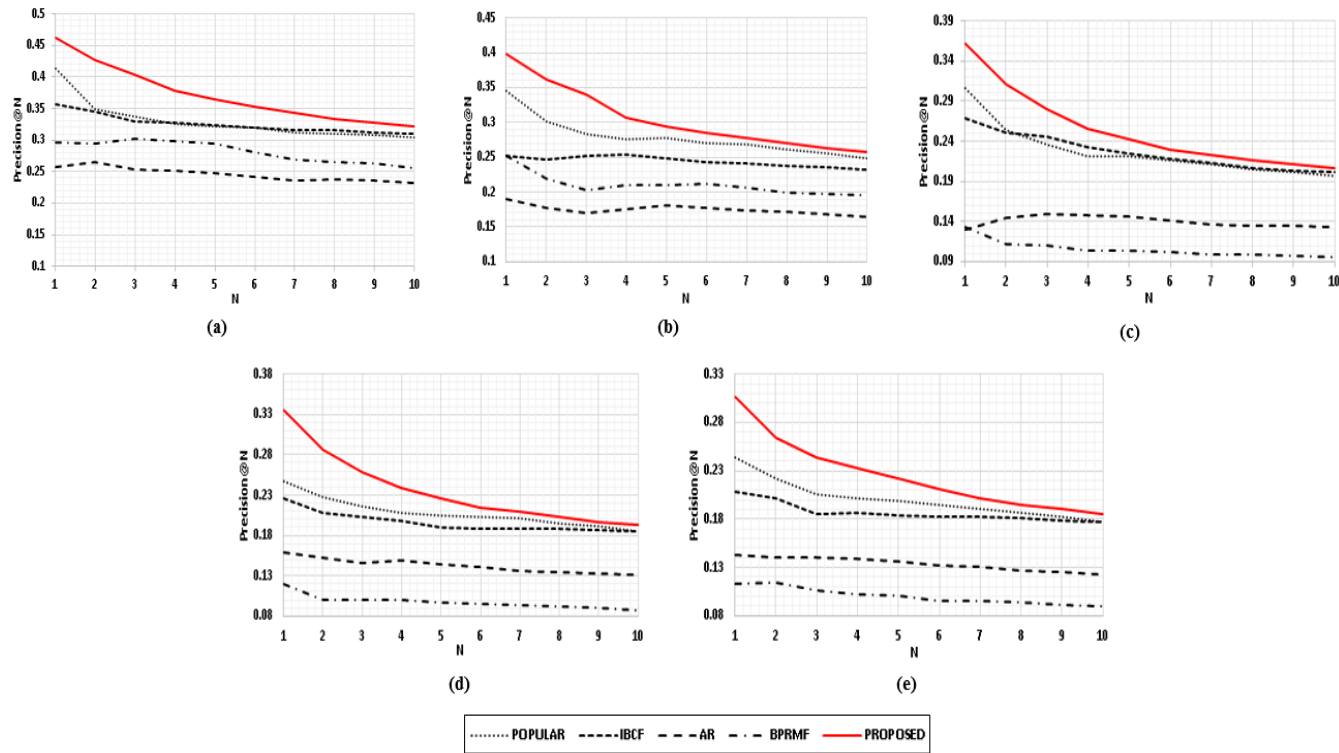
$$Precision@N = \frac{|relevant \cap recommended|}{N} \quad (9)$$

Where *relevant* are the items which have 1 in the test matrix for the active user and *recommended* are the items which are *recommended* by the system to that user.

## Results and Discussion

In Figure 10,  $Precision@N$  graphs for all test sets are shown for different values of  $N$ . Figure 10 shows the comparison of the proposed method with aforementioned methods i.e. IBCF, BPRMF, Popularity, and AR that are applicable to binary implicit feedback data. In the figure, we can see that the performance of IBCF is not good due to the sparsity of the training data. As we can see from the Figure 10, the performance of IBCF and BPRMF is not good even as compared to the Popularity based method. The reason for low performance of IBCF and BPRMF may be because of the sparseness of the dataset. Nguyen et al. (2014) in their work has also shown that the performance of IBCF and BPRF for sparse dataset is low as compared to the Popularity based method. For all test sets, we can see that in Figure 10, the curves for the proposed method is better as compared to other methods. It shows that the proposed method generates accurate recommendations for users. From the figures, we can see that the performance of the proposed method is stable, and it consistently outperforms other methods for all test sets. From the results, it is evident that the combination of the popularity of items and the intrinsic interest of users gives improved performance.

Now, in Figure 11, we show the results of the comparison of the proposed method with HeteRec. In these results, for the proposed method, we have taken the average precision value for all five test sets at  $N = 1$ ,  $N = 5$  and  $N = 10$  and compared the results with HeteRec as reported in Yu et al. (2014). We compared the performance of the global (HeteRec-g) as well as the personalized (HeteRec-p) method of recommendation using HeteRec with the proposed method. From the results, we can see that the proposed method outperforms both the global and personalized HeteRec in recommendation generation for users.



**Figure 10 - Results for test sets ( $test_i, i = 1, \dots, 5$ ) after performing the training of the proposed model using corresponding train sets ( $train_i, i = 1, \dots, 5$ ).**

Notes: In these experiments, for PWL algorithm, the value of regularization parameter is 0.5 i.e.  $\lambda = 0.5$  and the convergence tolerance is set to 0.001 i.e.  $\epsilon = 0.001$  for all experiments. We take the value of N from 1 to 10. For IBCF, we tuned the neighborhood size from {30, 40, 50, 60, 70, and 80}. For BPRMF, the threshold for convergence is  $1e-05$  and the number of factors are tuned from {10, 20, 30, 40, ..., 200}.



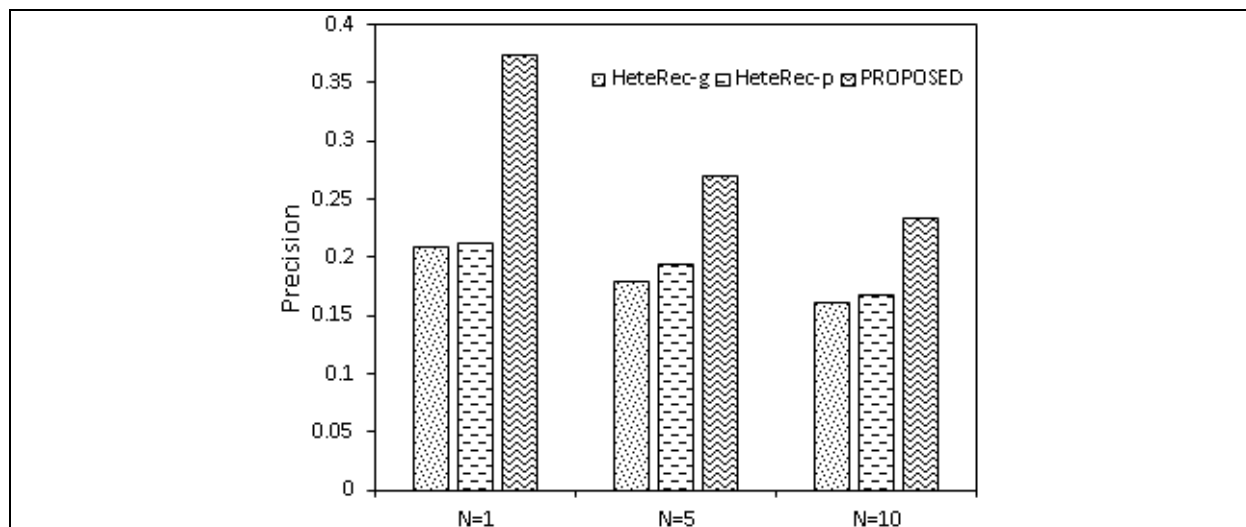


Figure 11 - Performance comparison of proposed method with HeteRec

## Conclusion and Future Research Directions

In this work, we study recommender systems for scalability and quality of recommendations for binary implicit feedback data. We found that item-based collaborative filtering is one of the most popular recommendation technique that is used by many E-commerce websites like Amazon.com for large scale recommendations. For binary implicit feedback data, to improve the quality of IBCF recommendations without affecting the scalability, we incorporate the meta-information about the items to be recommended. For that, we utilized the meta-path based framework for personalized recommendation generations. To improve the quality of recommendations, we combine the intrinsic interest of users and popularity of items. To determine the intrinsic interest of users, we propose a regularization based optimization that determines the personalized weights for meta-paths for users. The meta-path that represents the intrinsic interest of a user is assigned the higher weight as compared to other meta-paths. The online module of the proposed system generates recommendations, for an active user, in constant time. For evaluation,

the proposed method is compared with state-of-the-art recommender systems applicable to binary implicit feedback data namely IBCF, BPRMF, Popularity based and Association rule-based to show the effectiveness of the proposed approach. We also compared the proposed method with HeteRec.

Experimental results show that the proposed method outperforms the aforementioned methods and also performs the personalized weight learning effectively. Interesting future research directions include the incorporation of the demographic information related to users for improving the quality of recommendations. Also, we plan to apply the proposed model for large datasets in other domains like music and book recommendation using meta-information related to items.

## References

- Aggarwal, C. C. (2016). *Recommender Systems: The Textbook*. Springer.
- Barman, K., and Dabeer, O. (2012). Analysis of a collaborative filter based on popularity amongst neighbors. *Information Theory, IEEE Transactions on*, 58(12), pp.7110-7134.

- Basilico, J., and Hofmann, T. (2004). Unifying collaborative and content-based filtering. *Proceedings of the twenty-first international conference on Machine learning, ACM*.
- Bauer, J., and Nanopoulos, A. (2014). Recommender systems based on quantitative binary implicit customer feedback. *Decision Support Systems*, 68, pp. 77-88.
- Berkovsky, S., and Freyne, J. (2015). Web Personalization and Recommender Systems. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM*.
- Bellogín, A., Cantador, I., and Castells, P. (2013). A comparative study of heterogeneous item recommendations in social systems. *Information Sciences*, 221, pp. 142-169.
- Bobadilla, J., Ortega, F., Hernando, A., and Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46, pp. 109-132.
- Demiriz, A. (2004). Enhancing product recommender systems on sparse binary data. *Data Mining and Knowledge Discovery*, 9(2), pp. 147-170.
- Deshpande, M., and Karypis, G. (2004). Item-based top-n recommendation algorithms. *ACM Transactions on Information Systems*, 22(1), pp. 143-177.
- Gan, M. (2016). COUSIN: A network-based regression model for personalized recommendations. *Decision Support Systems*, 82, pp. 58-68.
- Gupta, M., Kumar, P., and Bhasker, B. (2017). HeteClass: A Meta-path based framework for transductive classification of objects in heterogeneous information networks. *Expert Systems with Applications*, 68, pp. 106-122.
- Gupta, M., Kumar, P., and Bhasker, B. (2016). Personalized item ranking from binary implicit user feedback using heterogeneous information network. *Proceedings of the Pacific Asia Conference on Information Systems*.
- Gupta, M., Kumar, P., and Bhasker, B. (2015). A New Relevance Measure for Heterogeneous Networks. *Proceedings of the Big Data Analytics and Knowledge Discovery, Springer International Publishing*.
- Hahsler, M. (2011). recommenderlab: A Framework for Developing and Testing Recommendation Algorithms.
- Hu, Y., Koren, Y., and Volinsky, C. (2008). Collaborative filtering for binary implicit feedback datasets. *Proceedings of the 8th IEEE International Conference on Data Mining, IEEE*.
- Kong, X., Yu, P.S., Ding, Y., and Wild, D.J. (2012). Meta path-based collective classification in heterogeneous information networks. *Proceedings of the 21st ACM international conference on Information and knowledge management, ACM*.
- Koren, Y., Bell, R. and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), pp. 30-37.
- Linden, G., Brent S., and Jeremy Y. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *Internet Computing, IEEE*, 7(1), pp. 76-80.
- Li, Y. M., Wu, C. T., and Lai, C. Y. (2013). A social recommender mechanism for e-commerce: Combining similarity, trust, and relationship. *Decision Support Systems*, 55(3), pp. 740-752.
- Ludovik Çoba (2016). recsys: Environment for Assessing Recommender Systems. R package version 0.9.5.4.

- McFee, B., Bertin-Mahieux, T., Ellis, D.P. and Lanckriet, G.R. (2012). The million song dataset challenge. *Proceedings of the 21st international conference companion on World Wide Web, ACM*.
- Nguyen, H. T., Almenningen, T., Havig, M., Schistad, H., Kofod-Petersen, A., Langseth, H., and Ramampiaro, H. (2014). Learning to Rank for Personalised Fashion Recommender Systems via Binary implicit Feedback. *Proceedings of Mining Intelligence and Knowledge Exploration, Springer International Publishing*.
- Pan, R., Zhou, Y., Cao, B., Liu, N.N., Lukose, R., Scholz, M. and Yang, Q. (2008). One-class collaborative filtering. *Proceedings of the Eighth IEEE International Conference on Data Mining, IEEE*.
- Rendle, S., Freudenthaler, C., Gantner, Z. and Schmidt-Thieme, L. (2009). BPR: Bayesian personalized ranking from binary implicit feedback. *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence, AUAI Press*.
- Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. *Proceedings of the 10th international conference on World Wide Web, ACM*.
- Shi, C., Zhang, Z., Luo, P., Yu, P.S., Yue, Y., and Wu, B. (2015). Semantic path based personalized recommendation on weighted heterogeneous information networks. *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, ACM*.
- Sun, Y., Han, J., Yan, X., Yu, P.S., and Wu, T. (2011). PathSim: Meta path-based top-k similarity search in heterogeneous information networks. *Proceedings of the international conference on VLDB*.
- Sun, Y., and Han, J. (2013). Mining heterogeneous information networks: a structural analysis approach. *ACM SIGKDD Explorations Newsletter, 14(2)*, pp. 20-28.
- Yu, X., Ren, X., Sun, Y., Gu, Q., Sturt, B., Khandelwal, U., Norick, B., and Han, J. (2014). Personalized entity recommendation: A heterogeneous information network approach. *Proceedings of the 7th ACM international conference on Web search and data mining, ACM*.

## About the Authors

**Mukul Gupta** is currently an Assistant Professor in Information Systems area at Indian Institute of Management Indore, India. He completed his FPM in IT & Systems area from Indian Institute of Management Lucknow, India. He did his M.Tech from Dayalbagh Educational Institute, Agra (India) in Computer Science and B.Tech from Uttar Pradesh Technical University, Lucknow (India) in Computer Science and Engineering. His current research interest includes e-Commerce, Recommendation Systems, Information Networks, Machine Learning, Social Media Analytics, Web and Data Mining.

**Dr. Pradeep Kumar** is currently an Associate Professor in IT and Systems area at Indian Institute of Management Lucknow, India. Prior to joining IIM, he was associated with SET Labs, Infosys Technologies Ltd. as a researcher. He served Institute for development and research in Banking Technology (IDRBT), established by Reserve Bank of India (RBI), as a Research Fellow. He received his Ph.D. from Department of Computer and Information Sciences, Hyderabad University, India. He holds M.Tech. and B.Sc.(Engg.) in Computer Science. His area of interest includes Data Warehousing, Data Mining, Web Mining, Text Mining and Big Data analytics. In his credit he has more than 30 authored research papers in international journals and conferences of repute.

**Dr. Bharat Bhasker** is a Professor in the area of Decision Science and Systems, at Indian Institute of Management Raipur, India. He holds a Bachelor's degree in Electronics and Communications Engineering from University of Roorkee, India; Master's degree and Doctorate in Computer Science from Virginia Polytechnic Institute and State University, USA. Prior to joining IIM Raipur, he was with MDL Information Systems and Sybase Inc., California, USA and was the

architect of the massively parallel DBMS, Sybase MPP. He also served as a Visiting Professor of Information Systems, Business Management School, University of Maryland, University of California, and University of Texas, USA. His research interests include Distributed Database Management, Data Mining, Personal Recommendation Systems, and Agent based Electronic Shopping. He has also authored two books on Electronic Commerce.