Journal of Computer Science

Original Research Paper

Community Aware Recommendation System with Explicit and Implicit Link Prediction

[1]Muhammed E Abd Alkhalec Tharwat, [1]Mohd Farhan Md Fudzee, [1]Shahreen Kasim, [1]Azizul Azhar Ramli, [2]Mohanad Sameer Jabbar and [3]Farazdaq Nahedh Alsamawi

[1]Faculty of Computer Science and Information Technology, University Tun Hussein Onn Malaysia, 86400, Malaysia
[2]Department of Medical Instruments Techniques Engineering, Technical College of Engineering, Albayan University, Iraq
[3]Department of Computer Science and E-Learning, Curricula Directorate MOE, Iraq

Article history
Received: 12-02-2023
Revised: 16-06-2023
Accepted: 19-06-2023

Corresponding Author:
Muhammed E Abd Alkhalec Tharwat
Faculty of Computer Science and Information Technology,
University Tun Hussein Onn Malaysia, 86400, Malaysia
Email: gi170023@student.uthm.edu.my

Abstract: Recommendation systems are essential tools that help users discover content they may be interested in, amidst the vast amount of information available online. However, current methods, such as using historical user-item interactions and collaborative filtering, have limitations in accurately predicting user preferences. Our research aims to address these challenges and improve the performance of recommendation systems. In this article, we propose a new approach to recommendation systems using a method called Probabilistic Matrix Factorization (PMF). We transform the standard PMF method into a community-based PMF that takes into account implicit relationships between users and items. To achieve this, we use a machine learning technique called Reduced Kernel Extreme Learning Machine (RKELM). Our proposed framework is designed to integrate these implicit relationships and identify communities of users with similar preferences based on PMF. We conducted a comparative analysis of our newly developed model against existing methods, using two well-known datasets. Various performance metrics, such as prediction errors, were employed to evaluate the effectiveness of our proposed community-based PMF approach with RKELM. Our model demonstrates improved performance, achieving a 7% improvement for the Douban dataset and a 4% improvement for the Last.fm dataset. Despite the improvements demonstrated by our model, potential limitations and challenges may still exist, such as scalability to larger datasets or adaptability to different domains. Future work could explore these aspects and investigate further enhancements to our approach.

Keywords: Recommendation Systems, Collaborative Filtering, Kernel Extreme Learning Machine, Root Mean Square Error

Introduction

The rapid development of Artificial Intelligence (AI) in general and Machine Learning (ML) in particular has opened the door to a wide range of smart applications (Eirinaki *et al*., 2018). Most of them aim at exploiting the data availability and analyzing them to obtain useful information that can help at automating various types of services and operations (Wang and Siau, 2019). One of the important operations is recommendation systems RSs that are essential for marketing on one side and for predicting and matching user-items preferences on the other side (Beheshti *et al*., 2020). In the data era and with the advancement of ML as a new field, RSs have emerged significantly by contributing a significant part to the revenue of firms and companies. It has been found by Krishnappa *et al*. (2015) that a high percentage of YouTube video views is impacted by RSs algorithms.

RSs are assigned the role of identifying individual interests and recommending certain products according to their preference (Tharwat *et al*., 2020). Identifying individual interests is done through a complicated process of analyzing the data and inferring the preference based on various aspects such as the explicit rating of products and linking that with the background of the product, the social circle, and linking that to the friend's background, extracting products association from their friend's preference. Data linking and association is feasible at this time because human is closer to the online platform and the user decisions are reflected in online activities that generate data for analysis. Furthermore, dealing with data does not have individual perspectives only, rather, it combines both individual and group perspectives. Consequently, developing an algorithm for RS that takes into account individual preference as well as his/her community and social network impact is an emerging trend (Tharwat *et al*., 2021). In this context, many works have exploited the existence of social networks in the prediction of users' preferences by matching users with his/her friends in terms of linking or not liking a certain item. The friendship relation has two types: Explicit and implicit. Both of them were used as input for RS algorithms. However, non-of the previous works in the literature have integrated friendships relations with community identification for RS. Such a combination is useful for many purposes: Firstly, it focuses the RS process on the most important part of the social network rather than using the whole social network. Secondly, it extends the friendship from the narrow region in the social network to a more contextual region named a community. Thirdly, it mitigates the issue of time and memory complexity that results from executing the algorithm on the entire social network instead of executing them on a small part of the social network, i.e., the communities. The contribution of this research can summarize in three points: First It reformulates the model of PMF from general PMF to community-based PMF, second It designs a new framework to integrate the implicit relations and identify communities based on probability matrix factorization, last evaluates the proposed work based on two datasets for recommendation systems and compares it with classical approaches. Typically, RS uses collaborative filtering CF. We define CF as an RS algorithm that uses the assumption of "sharing preference of items is associated with similar users". This assumption is simple and effective at the same time. Hence, it was used by many commercial systems such as Amazon. However, it still suffers from various weaknesses; firstly, sparsity that exists in the user-item rating matrix which occurs because many users do not provide ratings for items or they are considered passive users. Secondly, the issue of cold start which occur at the beginning stage of the operation of RS where no adequate users have yet provided their preference. Thirdly, the dependence of CF on similarity calculation methods which are in most cases simple and based on non-valid assumptions, e.g., users have at least rated some items in common. An implementation approach that uses the concept CF is matrix factorization MF (Koren *et al*., 2009) and probabilistic matrix factorization PMF (Ma *et al*., 2018).

Literature Review and Problem Statement

Social networks as a separate topic have attracted researchers to solve various matters and to conduct deep analyses of them. One of the important topics is link prediction and the analysis of the relation (Sun *et al*., 2016). This part has also found its way to recommendation systems with social network awareness which has increased only recently (Yang *et al*., 2018). Most of the approaches use the framework of matrix factorization (Xue *et al*., 2017) and probabilistic matrix factorization (Yin *et al*., 2018) as a baseline for developing RS with social network awareness. The real reason for incorporating social networks into RS is to mitigate the cold start problem by leveraging the friendship relations in the items predictions (Li *et al*., 2018). One of the early works that have used matrix factorization with social network information is the work of (Yang *et al*., 2016) which used MF to extract low dimensional latent features by using their trust relationship. The work was named TrustMF. This approach was criticized for ignoring two facts, namely, the trust relations are not necessarily similar to preference relations and the consideration of user's preference should be area oriented. Segmentation-based Matrix Factorization (SPMF) was proposed based on that by Peng and Xin (2019). Another development of trust information provided by social networks is extracting topical information from attention-based recurrent neural networks RNN. This was done in the work of (Zhang *et al*., 2019) which used an attention model with RNN for weighting different words according to the topic in the review document; however, it has not considered the implicit link prediction in the social network for improving the recommendation network. Other researchers have attempted to develop a recommendation system for supporting industries' preferences in terms of supply chain management. In the work of Zare *et al*. (2020), a hybrid model that combines similarity detection, an artificial neural network with fuzzy logic in order to analyze data from LinkedIn based on the industries of the user. The role of using the neural network Artificial Neural Network (ANN) is to capture the knowledge of the user preference and to represent it in a reduced structure that supports scalability. The article has presented a detailed framework with evaluation metrics of recommendation systems in terms of RMSE, precision, recall, and other metrics. The authors have not stated details about the neural network training part and the supportability of online training. The exploitation of implicit link prediction of the social network has been proposed by various researchers. In the work of Reafee *et al*. (2016), four similarity metrics were used to predict implicit links and they were used as additional information with modified probabilistic matrix factorization with two terms: The first one includes explicit links and the second one included implicit links. This study was criticized for the limitation of the similarity metrics such as resource allocation (Al-Sabaawi *et al*., 2020) which has extended the model to multi-step resource allocation MSRA which uses the information of friends of friends under multi-step. A significant limitation of such a model is the non-capability of learning any pattern from the social network regarding the implicit links prediction using the similarity metrics. Another limitation is the lacking of community awareness which can be useful for concentrating the solution of PMF on the identified communities. In the work of Moradabadi and Meybodi, (2018), learning automata LA was used to predict links. In their work, there is one automaton for each test link with the goal of finding the true weight of the link. The whole set of LA is responsible for calculating the weights of all links. However, this approach has been used verified in the goal of a recommendation system using the power of link prediction in the social network.

The Aim and Objectives of the Study

This article provides several contributions in the area of recommendation systems with social networking awareness. They can be presented as follows.

It reformulates the model of PMF from general PMF to community-based PMF with implicit link awareness.

It designs a new framework to integrate the implicit relations and identify communities based on probability matrix factorization. It is designated as Integrated-Probabilistic Matrix Factorization (IN-PMF) via incorporating the IR-RKELM (Tharwat *et al*., 2020) and NDS-CD-DA (Tharwat *et al*., 2021).

It evaluates IN-PMF based on two datasets for recommendation systems and compares it with classical approaches of stand-alone RS using PMF, explicit relation-aware PMF, and implicit-explicit aware PMF (Reafee *et al*., 2016).

Materials and Methods

Before we introduce our developed algorithm, we provide an overview of the Probabilistic Matrix Factorization (PMF). Given user and item data, the goal of PMF is to detect the features behind the interaction between elements (items) and users. Mathematically, given a set of *m* users, a set of *n* items, and rating matrix *R* of size $m \times n$ that is analyzed to a multiplication of *k*-rank factors, we define the latent users and item features as $U \in R^{k\times n}$ and $V \in R^{k\times m}$ with the dimension of *k* where $k \ll m, k \ll n$. The conditional probability over the observed rating is defined in Eq. (1):

$$p\left(R\middle|UV,\sigma_R^2\right)=\prod_{i=1}^{m}\prod_{i=1}^{n}\left[N\left(R_{ij}\middle|g\left(U_i^TV_i\right),\sigma_R^2\right)\right] \tag{1}$$

where, $N\left(x\middle|\mu,\sigma^2\right)$ Denotes the probability density function of the Gaussian distribution with mean μ and standard deviation σ^2. I_{ij} Denotes the indicator function which equals 1 if the user I rated the item *j* and 0 otherwise. *g* Denotes the logistic function $(x)=\frac{1}{1+e^{-x}}$. The prior of the user feature vector in Eqs. (2-3):

$$p\left(U\middle|\sigma_U^2\right)=\prod_{i=1}^{m}N\left(U_i\middle|0,\sigma_U^2 I\right) \tag{2}$$

$$p\left(V\middle|\sigma_V^2\right)=\prod_{j=1}^{n}N\left(Vj\middle\|0,\sigma_V^2 I\right) \tag{3}$$

Using Bayesian inference, the prediction of the rating is by minimizing the sum of squares of factorization error with quadratic regularization terms (Mnih and Salakhutdinov, 2007) in Eq. (4):

$$E=\frac{1}{2}\sum_{i=1}^{m}\sum_{i=1}^{n}I_{ij}\left(R_{ij}-g\left(U_i^T\right)\right)^2+\frac{\lambda U}{2}\sum_{i=1}^{m}U_i^TU_i\ \frac{\lambda v}{2}\sum_{j=1}^{n}V_j^TV_j \tag{4}$$

where, $\lambda_U=\sigma_R^2/\sigma_U^2$ and $\lambda_V=\sigma_R^2/\sigma_V^2$

Methodology

This section presents the methodology of providing community-aware RS with Explicit and Implicit link prediction. It consists of a framework and an algorithm for the community-aware RS with explicit and implicit relations. Evaluation metrics and the datasets used for evaluation are also presented.

Integrated Probabilistic Matrix Factorization Framework

This section presents the developed framework for the recommendation system with awareness of social networks in Fig. 1. The awareness of the social network is from two aspects, the first one is the explicit and implicit link aspect and the second one is from the community partitioning aspect. Hence, the framework uses two types of data: The first one is a social network which is a non-direct input and the second one is the user-item rating matrix which is a direct input. Basically, two algorithms are adopted to operate on the first input, namely, the implicit link prediction algorithm and the second the community identification algorithm. The result of the first one is the implicit links prediction and the result of the second one is the communities which represent subsets of the entire social network.

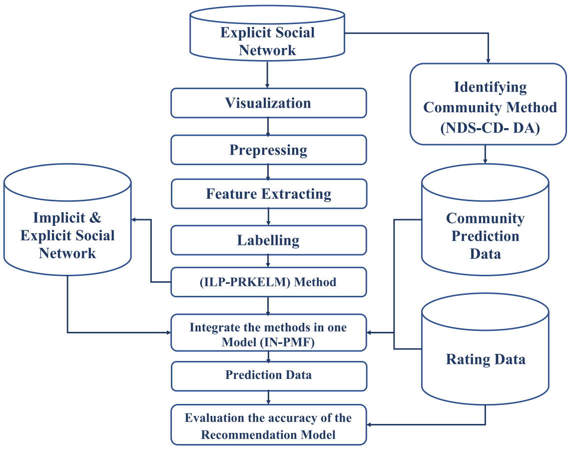


Fig. 1: Integrated probabilistic matrix factorization for recommendation IN-PMF

Algorithm of Integrated Probabilistic Matrix Factorization IN-PMF

This section presents our developed integrated probabilistic matrix factorization IN-PMF. As it is shown in the pseudocode in Algorithm 1, the algorithm accepts as input the social network, the rating data, the maximum number of communities, maximum iterations, and the population size. The output of the algorithm is the predicted rating. The algorithm is combined with three main steps: The first one is dividing the social network into set communities as given in line 4. Each one is regarded as a sub-social network for the following steps. The next step is to operate the implicit link prediction using the reduced kernel extreme learning machine RKELM as it is shown in line 6. The result of this step is a social network with both its explicit and implicit relations which is used in the probabilistic matrix factorization as it is given in line 7. The latter has the goal of generating the rating predictions which will be compared with the ground truth in order to generate the root mean square error in line 8.

Algorithm 1: Pseudocode of integrated probabilistic matrix factorization IN-PMF

Input
1. SocialNetwork, RatingDataTraining, RatingDataTesting, MaxNumberofCommunities
2. MaxIteration, PopulationSize

Output
1. PredictedRating, RMSE

Start
1: Communities=Divide (MaxNumberofCommunities, MaxIteration, PopulationSize, SocialNetwork)
2: For each of the communities
3: Apply RKELM for extracting implicit relationship
4: For each of the communities
5: PredictedRating=Apply PMF with weighting for implicit and explicit
6: Compare RatingDataTesting ground truth with PredictedRating and generate RMSE
7: End

Implicit *Link Prediction*

The other component of the proposed framework is the implicit link prediction-probabilistic reduced kernel extreme learning machine ILP-PRKELM (Tharwat *et al*., 2022) which is presented in the pseudocode of Algorithm 2. The input of the algorithm is the social network SN, the minimum considered a number of links for prediction NL and the size of the subset used for building the model size of the subset. The output is implicit social network SNI and the testing accuracy.

The data is defined over a feature space as H, the kernel is $\Phi : x \rightarrow H$ notated as $k(x_i, x_i)$ where, $x_i, x_i \in x$ The kernel variant of ELM is given as Eq. (5):

$$\begin{aligned} y(x) &= \left[k\begin{pmatrix} x, x_1 \\ k(x, x_N) \end{pmatrix} \right] \left(\frac{1}{\lambda} + \Omega_{ELM\,j} \right)^{-1} T \\ \Omega ELM_i &= HH^T = h(x, x_i) \\ j &= 1, \ldots N \end{aligned} \tag{5}$$

There are three famous types of kernels: Linear kernel $k(x_i, x_i) = x_i^T x_i$, radial basis function $k(x_i, x_j) = e^{(-\gamma x_i - x_j^2)}$ $\gamma > 0$, and polynomial kernel $k(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$ the kernel will be calculated based on parts of the data with the condition of having at least a set of samples from each of the two classes. The percentage that is used is named as the sampling factor. The output of the classifier is $y_i = (p_i^1, p_i^2)$ where p_i^1 denotes the probability of existing of an implicit relationship and p_i^2 denotes the probability of the non-existing of an implicit relationship. The equation $p_i^1 + p_i^2 = 1$ must be applied regardless of the value of *i*. ELM was adapted to output the probabilities p_i^1, p_i^2 using Eq. (6):

$$p_i = \frac{y - \min(y)}{\max(y) - \min(y)} \tag{6}$$

Equation (7) enables us to know how much the classifier is confident about a certain record to represent an implicit relationship. Our model of prediction will use a confident threshold t_{conf} to predict the implicit relation based on the inequality:

$$C(y_i) = \begin{cases} 1 \; if \; p_i^1 > t_{con\,f} \\ 0 \; otherwise \end{cases} \tag{7}$$

DOI: 10.3844/jcssp.2023.953.963

The system evaluation is based on the recognition rate of the implicit relation. The accuracy of the recognition is calculated as follows Eq. (8):

$$Accuaracy = 1 - \frac{Number\,of\,\,miss\,classification}{Total\,number\,of\,\,record} \tag{8}$$

The implicit relationship of a friend in the social network can be achieved based on the algorithm see Algorithm 2. The algorithm starts with the input of the parameters. Next, building the adjacency matrix which determines if there is a connection between two nodes or not. Next, the algorithm goes through the nodes and selects the nodes that have at least a number of links equal to NL. The selected nodes will be used for inferring the implicit relationship based on RKELM trained on the explicit relations considering extracted features from the network. The extracted features are calculated for any two nodes A and B in a graph:

1. $\prod(A,B)$ The size of the intersection between $\Gamma(A)$ $\Gamma(B)$ and or $|\Gamma(A)\cap\Gamma(B)|$. Where $\Gamma(x)$ denotes the neighbors of *x*
2. $\Omega(A,B)$ The size of the union of $\Gamma(A)$ and $\Gamma(B)$ or $|\Gamma(A) U \Gamma(B)|$
3. The multiplication of the degree of each of the two nodes that their link is predicted
4. The resource allocation index: Considering a pair of nodes *A* and *B* which are not directly connected. Node *A* can send some resources to node *B*, with their common neighbors as transmitters. The amount of resources that *B* receives from *A* is calculated as Eq. (9):

$$S_{AB}^{RA} = \sum_{z\in\Gamma(A)\cap\Gamma(B)} \frac{1}{|\Gamma(Z)|} \tag{9}$$

Algorithm 2: Proposed ILP-PRKELM algorithm

```
Input
  1. SN
  2. NL
  3. SizeOfSubset
Output
  1. SNI
  2. TestingAccuracy
Start
  AdjcencyList=BuildAdj(SN)
  for each node i in the SN
      FriendNo=checkFriends(i,AdjcencyList)
      if(FriendNo)<NL
          remove node i from SN
  end
  EN=SelectRandom(AdjcencyList, sizeOfSubset);
  End
  Data=[];
  for each record in EN
      features=ExtractFeatures(record)
      add features to Data
      add a label to Data
  End
  [trainingData,testingData]= DataPartitioning(Data)
  trainedELM=ELMtraining(trainingData)
  predictedResutls= ELMPrediction( trainedELM,
     testingData.Records,threshold)
  testingAccuracy = Evaluate(predictedResutls,
     testingData.Labels)
  SNI= buildNetwork (testingData.Records,
     predictedResutls)
  End
```

Community Identification

Assuming we have an evolving social network SN_t with respect to time t = 1, 2…T and assumed a maximum number of communities in the social network MaxCom. Considering that a solution $S_t^j = (n_i), n_i \in [1,..MaxCom], i \in 1,..m$ where m denotes the number of nodes in the social network and each solution $S_t^j \in PF_t$. Each of the solutions S_t^j is part of the Pareto front (set of non-dominated solutions) with respect to two objectives, namely, modularity and normalized mutual information. Our goal is to operate a multi-objective optimization algorithm for finding PF_t which will result in decomposing the social network $SN_t = \{SN_1, SN_2, ...SN_T\}$. The selected algorithm is the one presented in Tharwat *et al*. (2021). The pseudocode of the selected algorithm is given in Algorithm 3 and the output of the algorithm is the Pareto fronts of each of the solutions of community detection of the social networks. As we see in the pseudocode, the algorithm accepts a snapshot of the social network at each moment and provides the corresponding Pareto front. It runs based on three parameters: The maximum number of community MaxCom, the maximum number of iterations max iterations, the number of solutions a number of solutions, and objective functions definition for modularity and normalized mutual information. The first objective is the modularity at moment *t* which is *Qt* and is given by Eq. (10):

$$Q_t = \sum_{S=1}^{CN_t} \left[\frac{l_s}{m} - \left(\frac{d_s}{2m} \right)^2 \right] \tag{10}$$

where, l_s Denotes the number of edges in the community *s*; d_s denotes the sum of the degree of nodes inside the community. The second objective is the normalized mutual information *NMI* between two partitions *A* and *B* measure that is given by Eq. (11):

$$NMI = \frac{-2\sum_{i=1}^{kA} \sum_{j=1}^{KB} c_{ij} \log\left(c_{ij} \cdot |v| / c_i . c_j\right)}{\sum_{i=1}^{KA} c_i \log\left(c_i . / |v|\right) + \sum_{j=1}^{KB} c_j \log\left(c_j . |v|\right)} \quad (11)$$

Algorithm 3: Pseudocode for community identification in an evolving social network

Input

1. $G - Mean = TP / (TP + FN) * TN / (TN + FP)$ //evolving social network
2.MaxCom, // the maximum number of communities in the network
3.maxIterations
4. number of solutions
5. objectiveFuntions=[@Modularity, @NMI]

Output

1. $PF = \{PF_1, PF_2, ...PF_T\}$ // evolving Pareto front

Start

1: for t=1 until T
2: N[1]=numberOfNodes(SN(t))
3. solutionSpaceBoundary=[1 MaxCom]
4. solutionSpaceDimension=N[1]
5. objectiveFuntions=[@Modularity,@NMI]
6. initateNSGAII(solutionSpaceBoundary, solutionSpaceDimension);
7. PF[1]=callNSGAIICommunities(maxIterations, numberofSolutions ,objectiveFuntions);
8. End

End

Evaluation Metrics

In order to evaluate the proposed recommendation system, we used five metrics *RMSE*, Recall, Precision, G-mean, and F-measure. The first metrics Root Mean Square Error (RMSE) which is presented in Eq. (12):

$$RMSE = \sum_{i=1}^{N} \frac{\sqrt{yi - \tilde{y}_i}}{N} \quad (12)$$

where, yi denotes the ground truth of the training, $\tilde{y}_i$ denotes the predicted value of the rating N Denotes the number of samples. The recall is the next metric used. It is a measure that determines the number of correct positive predictions made from all the positive predictions that could have been made. Recall metrics presented in Eq. (13):

$$Recall = TP / (TP + FN) \quad (13)$$

where, *TP* represents the True Positive; *FN* represent the False Negative

The next metric is Precision. It is a metric that quantifies the number of correct positive predictions made. Precision, therefore, calculates the accuracy for the minority class. It is calculated as the ratio of correctly predicted positive examples divided by the total number of positive examples that were predicted (Mobasher *et al*., 2001). Precision metrics presented in Eq. (14):

$$Precision = TP / (TP + FP) \quad (14)$$

where, *TP* represents the true positive; *FP* represents False Positives

Other measures are the Geometric average/Mean (GM). In mathematics, the mean G is the average value or mean that indicates the central direction of a set of numbers by finding the output of its values. Basically, we multiply the numbers completely and take the nth root of the multiplied numbers, where n is the total number of data values. G-mean metrics presented in Eq. (15):

$$G - Mean = TP / (TP + FN) * TN / (TN + FP) \quad (15)$$

The last metric we've used for our work is the F-measure. F1-Score or F-Measure is an evaluation measure of a rating defined as the harmonic average of accuracy and recall. It is a statistical measure of the accuracy of a test or model. Mathematically, it is expressed as follows Eq. (16):

$$F - Measure = 2 * Precision * Recall / Precision * Recall \quad (16)$$

Dataset

We used two datasets. The first one is the Last.fm. it was released in the framework of the 2nd international workshop on information heterogeneity and fusion in recommender systems (Cantador *et al*., 2011). It's consisted of more than 40 million active clients distributed in more than 190 countries. Last.fm dataset allows the clients to list online music and allow the clients to extend friendly relations with other clients. This dataset only records the listening of each client for particular artists. The latter one is the Douban dataset crawled by Ma *et al*. (2011). Douban is a social website designed on 6/3/2005 in China, it provides books, music, and movies recommendation based on user rating and evaluation. It's like Facebook where users can have friendship relations through email communication. The rating is secret ranges from 1/5 which is useless to 5/5 which is most useful. The basic statistics are in Table 1. The data were partitioned into two sets: Training and testing. The training contains balanced friendship and non-friendship relations while the testing is random 500 samples with also balance friendship and non-friendship.

Table 1: Statistics of Last. fm and Douban datasets

Datasets	No. of users	No. of explicit relations
Last. fm	1123	22126
Douban	1384	50738

DOI: 10.3844/jcssp.2023.953.963

Results and Discussion

In this research, we performed the analysis and developed our model using the MATLAB environment, a widely used platform for scientific computing and data analysis. This information can serve as a foundation for future research seeking to build upon or replicate our approach.

For comparing our developed IN-PMF with the benchmark, we apply it to 2 datasets Last.fm and Douban respectively using the parameters depicted in Table 2. The results of recommendation predictions were generated for each of the similarity metrics, namely, Jaccard Coefficient (JC), Common Neighbors (CN), Preferential Attachment (PA), and Resource Allocation index (RA) as it is observed in Figs. 2-3, comparison between developed IN-PMF and the benchmark, and error compression between the communities version and the full social version for the Last.fm dataset respectively. The training curve for each of the two approaches within 100 iterations, we note that our developed IN-PMF has accomplished lower RMSE for all similarity coefficients compared with the benchmark. It is observed that RMSE for IN-PMF was in the range of 0.71 while it has increased to the 0.74 range for the benchmark. This is equivalent to an improvement percentage of 4% in the error of predictions. Similarly, as it is presented in Figs. 4-5 Comparison between developed IN-PMF and the benchmark and error compression between the communities version and the full social version that provides the convergence curve for Douban, IN-PMF which is based on partitioning the social network into communities and extending it with implicit links has accomplished lower RMSE compared with the benchmark that deals with the entire network. The least reached RMSE for IN-PMF was less than 0.946 compared to RMSE for the benchmark higher than 1.026. This means an improvement percentage of 7%.

For more evaluation, we present the comparison of the classification metrics between the two approaches of recommendation system, community-based and full society based. As it is shown in Figs. 6-7 for Last.fm and Douban respectively, we find that the community based has outperformed the society based in terms of recall, precision, G-mean, and F-measure for the Douban dataset which is shown in Fig. 7. This applies on all methods of similarities used for the recommendation system. However, for the Last.fm in Fig. 6, the classification metrics of the community-based were inferior which is interpreted by the nature of the data. In other words, performing the recommendation using community partitioning is not always useful for improving performance.

Overall, the proposed algorithm, Integrated-Probabilistic Matrix Factorization (IN-PMF), offers several unique features that differentiate it from existing techniques in the area of recommendation systems with social networking awareness. These features contribute to its superiority over existing techniques.

Community-based PMF with implicit link awareness: IN-PMF reformulates the standard PMF model to focus on community-based PMF. This enables the algorithm to better account for implicit relationships between users and items, which is often neglected in traditional PMF models.

Novel framework integration: The new framework, IN-PMF, integrates two cutting-edge techniques, Implicit Relations-Reduced Kernel Extreme Learning Machine (IR-RKELM) and Network-Driven Selection - Community Detection-Data Aggregation (NDS-CD-DA). This combination allows the algorithm to effectively identify communities based on probability matrix factorization and to integrate implicit relationships more efficiently.

Comprehensive evaluation and comparison: IN-PMF is evaluated using two datasets for recommendation systems and compared against classical approaches of stand-alone RS using PMF, explicit relation aware PMF, and implicit-explicit aware PMF. This thorough evaluation demonstrates the superior performance of IN-PMF in terms of prediction accuracy and overall effectiveness compared to existing techniques.

Table 2: The parameters used for the evaluation with their corresponding values

Parameter name	Value	
Number of iterations	200	
Number of solutions	100	
Maximum number of communities	50	
Number of hidden neurons	5-20	
Confidence value	0-0.9	
Activation function	sin, sig, hardlim, tribas, radbas	
Sampling factor	0.01	
Kernel Type	a = 0.1	
Kernel 1	Sigmoid	a = 0.1, d = 2
Kernel 2	linear	a = 0.1
Kernel 3	Polynomial $(ax_i^T.x_j + r)^d$	c = 0
Kernel 4	RBF $e^{(-ax_i - x_j^2)}$	1 linear, 2 polynomial, 14 RBF
Kernel 5	Inverse multiquadric $\frac{1}{\sqrt{\lVert x-y \rVert^2 + c^2}}$	1 linear, 16 polynomials
Kernel Set	Set 1	1 linear, 2 polynomials, 14 RBF
	Set 2	1 linear, 16 polynomials
	Set 3	1 linear, 16 polynomials

DOI: 10.3844/jcssp.2023.953.963

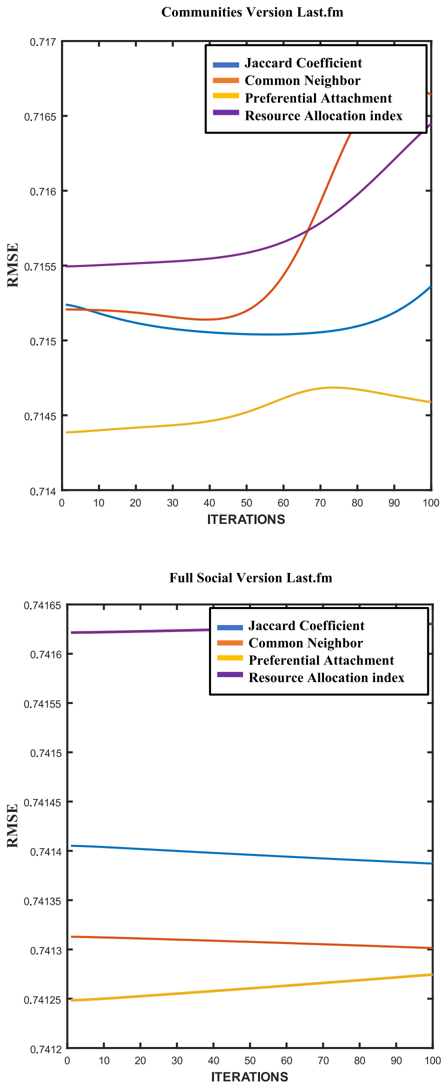


Fig. 2: Comparison between developed IN-PMF and the benchmark (Last.fm)

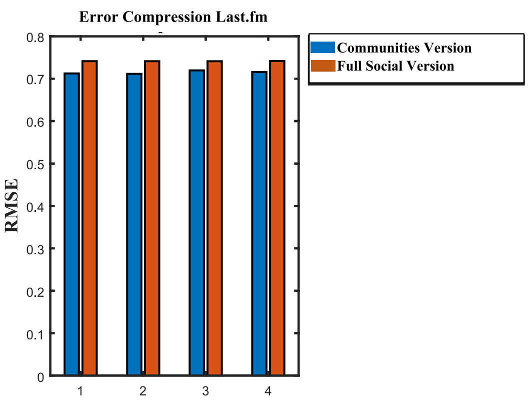


Fig. 3: Error compression between the community's version and the full social version (Last.fm)

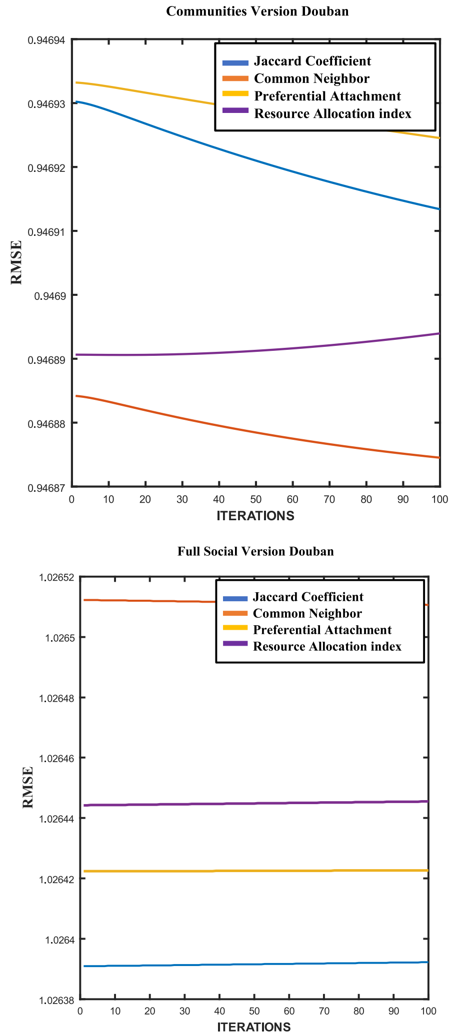


Fig. 4: Comparison between developed IN-PMF and the benchmark (Douban)

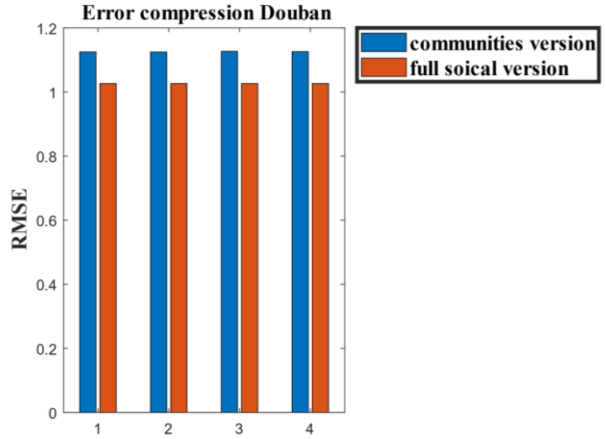


Fig. 5: Error Compression betwaeen the communities' version and the full social version (Douban)

DOI: 10.3844/jcssp.2023.953.963

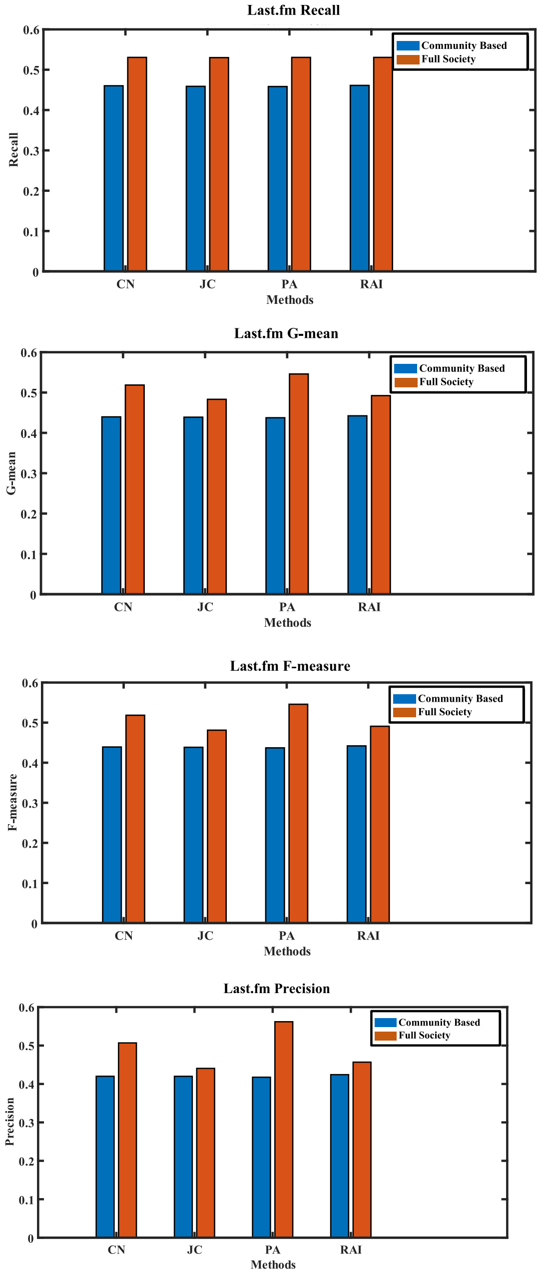
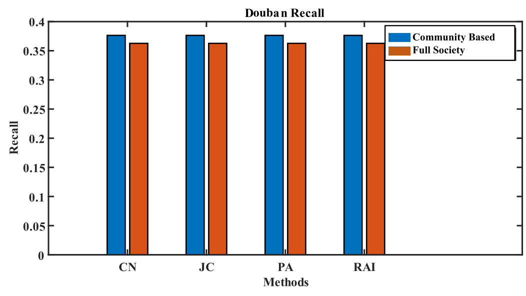


Fig. 6: Classification metrics of the recommendation system using the Last.fm dataset



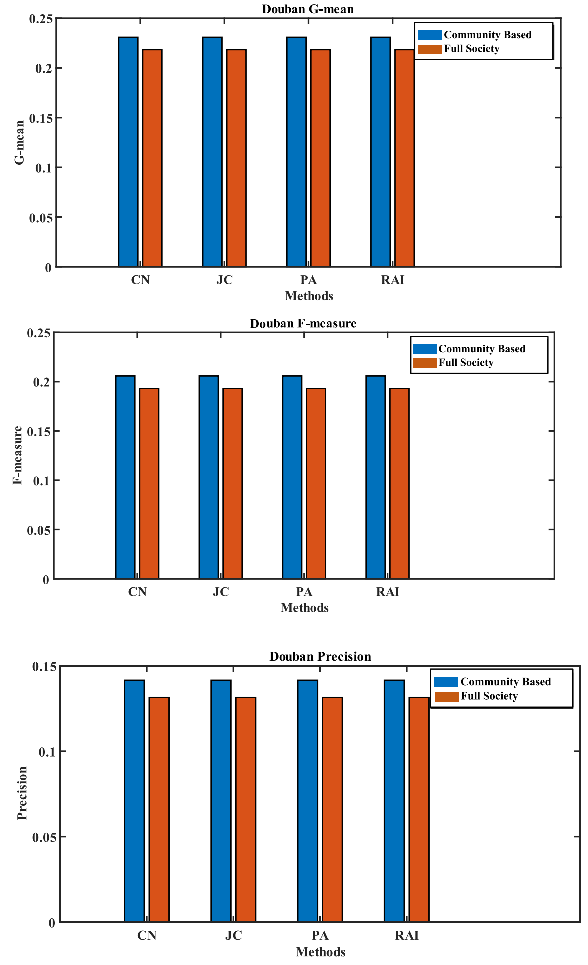


Fig. 7: Classification metrics of the recommendation system using the Douban dataset

Conclusions

This article has formulated the problem of recommendation prediction using probabilistic matrix factorization based on community awareness. It proposed a new framework that assumes both implicit link predictions and community partitioning of the social network for improving the prediction of ratings. The concept is to avoid using non-relevant data from other communities in the training of the user's preference model in the subject community. In addition, this approach assists in increasing the computation efficiency of the algorithm. The framework is designated as Integrated IN-PMF. Its comparison with the benchmark based on two datasets Douban and Last.fm shows its superiority over the benchmark with an improvement percentage of the testing error with a percentage of 0.04 for Last.fm and 0.07 for Douban. Future work is to investigate using other types of similarity measures for the solution of PMF and using other types of community partitioning.

Acknowledgment

This study is supported by the Ministry of Higher Education (MOHE). The authors are grateful for this support.

Funding Information

This study is supported by the Ministry of Higher Education (MOHE) under Fundamental Research Grant Scheme (FRGS) reference code FRGS/1/2018/ICT04/UTHM/02/3.

Author's Contributions

Muhammed E Abd Alkhalec Tharwat: The author of the main findings and the implementation.

Mohd Farhan Md Fudzee: Supervisor guided to think and success the methodology.

Shahreen Kasim: Co-Supervisor guided to do the literature reviews.

Azizul Azhar Ramli: Co-Supervisor guided gap analysis.

Mohanad Sameer Jabbar: Technical part support to run the code.

Farazdaq Nahedh Alsamawi: Formatted and written process.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and that no ethical issues are involved.

Reference

Al-Sabaawi, A. M. A., Karacan, H., & Yenice, Y. E. (2020). Exploiting implicit social relationships via dimension reduction to improve recommendation system performance. *PloS One*, *15*(4), e0231457. https://doi.org/10.1371/journal.pone.0231457

Beheshti, A., Yakhchi, S., Mousaeirad, S., Ghafari, S. M., Goluguri, S. R., & Edrisi, M. A. (2020). Towards cognitive recommender systems. *Algorithms*, *13*(8), 176. https://doi.org/10.3390/a13080176

Cantador, I., Brusilovsky, P., & Kuflik, T. (2011, October). Second workshop on information heterogeneity and fusion in recommender systems (HetRec2011). In *Proceedings of the 5th ACM Conference on Recommender Systems* (pp. 387-388). https://dl.acm.org/doi/pdf/10.1145/2043932.2044016

Eirinaki, M., Gao, J., Varlamis, I., & Tserpes, K. (2018). Recommender systems for large-scale social networks: A review of challenges and solutions. *Future Generation Computer Systems*, *78*, 413-418. https://doi.org/10.1016/j.future.2017.09.015

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, *42*(8), 30-37. https://doi.org/10.1109/MC.2009.263

Krishnappa, D. K., Zink, M., Griwodz, C., & Halvorsen, P. (2015). Cache-centric video recommendation: An approach to improve the efficiency of youtube caches. *ACM Transactions on Multimedia Computing, Communications and Applications (TOMM)*, *11*(4), 1-20. https://doi.org/10.1145/2716310

Li, G., Cai, Z., Yin, G., He, Z., & Siddula, M. (2018). Differentially Private Recommendation System Based on Community Detection in Social Network Applications. Security & Communication Networks. https://doi.org/10.1155/2018/3530123

Ma, H., Yang, H., Lyu, M. R., & King, I. (2018, October). Sorec: Social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management* (pp. 931-940). https://doi.org/10.1145/1458082.1458205

Ma, H., Zhou, D., Liu, C., Lyu, M. R., & King, I. (2011, February). Recommender systems with social regularization. In Proceedings of the Fourth ACM International Conference on Web Search and Data Mining (pp. 287-296). https://doi.org/10.1145/1935826.1935877

Mnih, A., & Salakhutdinov, R. R. (2007). Probabilistic matrix factorization. Advances in neural information processing systems, 20. https://proceedings.neurips.cc/paper_files/paper/2007/hash/d7322ed717dedf1eb4e6e52a37ea7bcd-Abstract.html

Mobasher, B., Dai, H., Luo, T., & Nakagawa, M. (2001, November). Effective personalization based on association rule discovery from web usage data. In *Proceedings of the 3rd International Workshop on Web Information and Data Management* (pp. 9-15). https://doi.org/10.1145/502932.502935

Moradabadi, B., & Meybodi, M. R. (2018). Link prediction in weighted social networks using learning automata. *Engineering Applications of Artificial Intelligence*, *70*, 16-24. https://doi.org/10.1016/j.engappai.2017.12. 006

Peng, W., & Xin, B. (2019). A social trust and preference segmentation-based matrix factorization recommendation algorithm. *EURASIP Journal on Wireless Communications and Networking*, *2019*, 1-12. https://doi.org/10.1186/s13638-019-1600-4

Reafee, W., Salim, N., & Khan, A. (2016). The power of implicit social relation in rating prediction of social recommender systems. *PloS One*, *11*(5), e0154848. https://doi.org/10.1371/journal.pone.0154848

Sun, Y., Xu, H., Bertino, E., & Li, D. (2016, June). User preference based link inference for social network. In *2016 IEEE International Conference on Web Services (ICWS)* (pp. 188-195). IEEE. https://doi.org/10.1109/ICWS.2016.32

Tharwat, M. E. A. A., Fudzee, M. F. M., Kasim, S., Ramli, A. A., & Ali, M. K. (2021). Multi-objective NSGA-II based community detection using dynamical evolution social network. *International Journal of Electrical and Computer Engineering*, *11*(5), 4502. https://doi.org/10.11591/ijece.v11i5.pp4502-4512

Tharwat, M. E. A. A., Fudzee, M. F. M., Kasim, S., Ramli, A. A., & Madni, S. H. H. (2022, May). Friendship Prediction in Social Networks Using Developed Extreme Learning Machine with Kernel Reduction and Probabilistic Calculation. In *Recent Advances in Soft Computing and Data Mining: Proceedings of the Fifth International Conference on Soft Computing and Data Mining (SCDM 2022), May 30-31, 2022* (pp. 56-68). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-00828-3_6

Tharwat, M. E. A. A., Jacob, D. W., Fudzee, M. F. M., Kasim, S., Ramli, A. A., & Lubis, M. (2020). The role of trust to enhance the recommendation system based on social network. *Int. J. Adv. Sci. Eng. Inf. Technol*, *10*, 1387-1395.

Wang, W., & Siau, K. (2019). Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda. *Journal of Database Management (JDM)*, *30*(1), 61-79. https://doi.org/10.4018/JDM.2019010104

Xue, H. J., Dai, X., Zhang, J., Huang, S., & Chen, J. (2017, August). Deep matrix factorization models for recommender systems. In *IJCAI* (Vol. *17*, pp. 3203-3209). https://www.ijcai.org/Proceedings/2017/0447.pdf

Yang, B., Lei, Y., Liu, J., & Li, W. (2016). Social collaborative filtering by trust. *IEEE transactions on pattern analysis and machine intelligence*, *39*(8), 1633-1647. https://doi.org/10.1109/TPAMI.2016.2605085

Yang, D., Chen, L., Liang, J., Xiao, Y., & Wang, W. (2018, August). Social tag embedding for the recommendation with sparse user-item interactions. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, (pp. 127-134). IEEE. https://doi.org/10.1109/ASONAM.2018.8508802

Yin, Y., Chen, L., & Wan, J. (2018). Location-aware service recommendation with enhanced probabilistic matrix factorization. *IEEE Access*, *6*, 62815-62825. https://doi.org/10.1109/ACCESS.2018.2877137

Zare, A., Motadel, M. R., & Jalali, A. (2020). Presenting a hybrid model in social networks recommendation system architecture development. *AI & Society*, *35*, 469-483. https://doi.org/10.1007/s00146-019-00893-z

Zhang, W., Liu, F., Xu, D., & Jiang, L. (2019). Recommendation system in social networks with topical attention and probabilistic matrix factorization. *PloS One*, *14*(10), e0223967. https://doi.org/10.1371/journal.pone.0223967