Matrix Co-factorization for Recommendation with Rich Side Information and Implicit Feedback

Yi Fang and Luo Si

Department of Computer Science
Purdue University
West Lafayette, IN 47906, USA
fangy@cs.purdue.edu

HetRec 2011
Outline

- Background
- Challenges
- Our approach
- Experiments
- Conclusion
Recommender systems have been ubiquitously adopted in many applications such as e-commerce, social bookmarking, and subscription based services.

Most of them focus on the areas of leisure activities such as art (e.g., movies and books), fashion (e.g., music and gaming), and food (e.g., restaurants).

This work investigates the task of recommendation in online scientific communities.
Nanohub

Nanohub (http://www.nanohub.org) is an online scientific community for research, education and collaboration in nanotechnology.

It comprises numerous resources with an active user base. These resources include lectures, seminars, tutorials, publications, events and so on.

The task is to recommend relevant resources to the users.
Matrix Co-factorization for Recommendation with Rich Side Information and Implicit Feedback

HetRec 2011
ECE 495N: Fundamentals of Nanoelectronics Lecture Notes (Fall 2009)

By Mehdi Saimani Jelodar, Supriyo Datta (editor)

Electrical and Computer Engineering, Purdue University, West Lafayette, IN

Lecture notes for the Fall 2009 teaching of ECE 495: Fundamentals of Nanoelectronics.

Abstract

Lecture notes for the Fall 2009 teaching of ECE 495: Fundamentals of Nanoelectronics.

These notes closely parallel the online lectures from the Fall 2008 teaching and may be useful to students who are using the online lectures.

Cite this work

Researchers should cite this work as follows:


BibTex  EndNote

Time  Fall Semester, 2009

Location  Electrical Engineering, Purdue University, West Lafayette, IN

Tags  course lectures  nanoelectronics  NEGF  quantum transport  transistors
Vladimir M. Shalaev

Organization: Purdue University

Employment Status: University / College Faculty

Web Site: http://cobweb.ecn.purdue.edu/~shalaev/

Biography: Vladimir M. Shalaev, the Robert and Anne Burnett Professor of Electrical and Computer Engineering at Purdue University, specializes in nanophotonics and plasmonic nanomaterials. He is a Fellow of the American Physical Society and a Fellow of the Optical Society of America. He earned a doctoral degree in physics and mathematics in 1983 and a master's degree in physics, with highest distinction, in 1979, both from the Krasnoyarsk State University in Russia. Shalaev, who came to Purdue in 2001, was the George W. Gardiner Professor of Physics at New Mexico State University. He previously taught and conducted research at the Krasnoyarsk State University and University of Toronto. Shalaev also was a Humboldt Foundation Fellow at the University of Heidelberg in Germany. He authored and edited 4 books, published 14 invited book chapters, and over 200 research papers. Dr. Shalaev is a co-editor of the Elsevier Book series "Advances in Nano-Optic Nano-Photonics," co-editor of Applied Physics B - Lasers and Optics, editor of J. of Optical Society of America, and a chair of a topical group "Photonic Metamaterials" of the OSA.
Challenges

▶ There exists very rich information about resources and users.
▶ The users in the scientific communities tend not to give explicit ratings to the resources, even though they have clear preference in their minds.
This work proposes matrix co-factorization techniques to incorporate rich user and resource information into recommendation with implicit feedback.

Our main contribution is to factorize implicit feedback, user, and item content matrices into shared subspaces so that the rich side information can be exploited for recommendation with implicit feedback.

The experiments on Nanohub show that the proposed method can effectively improve the recommendation performance.
Related Work

- Hybrid methods of Content-based Filtering (CBF) and Collaborative Filtering (CF)
- One-Class Collaborative Filtering (OCCF) for implicit feedback
- To the best of our knowledge, there is no prior work on incorporating both user and item information for implicit feedback
With explicit feedback users tell us both what they like and what they dislike, but with implicit user feedback, there is no negative examples. This setting is referred to as one-class collaborative filtering.

A naive approach is to treat all missing values as negative examples (i.e., AMAN) and then directly apply matrix factorization techniques.

A better method proposed is to treat all missing values as negative, but with weights controlling their relative contribution to the loss function.

\[ J(P, Q) = \sum_{i=1}^{n} \sum_{j=1}^{m} W_{i,j} (R_{ij} - P_i Q_j)^2 \]
The above OCCF models do not consider the rich side information that are available in many real-world systems.

For user $i$ and word $w$, $U_{iw}$ is the TFIDF weight calculated from user profiles. Similarly, the matrix $T_{jw}$ encodes the item information.

Our method is motivated by the assumption that the latent features that determine whether a user likes a given item, and the latent features that determine the content of that item, can be mapped into a shared space in which they are likely to be similar.

We constrain our factorizations to use a common matrix to model the features of each item.
The feedback matrix $R \approx PQ$, and the item content matrix $T \approx YQ$, with the latent feature matrix $Q$ contributing to both matrices. Similarly, the user content matrix can also be approximated by $U \approx PX$ with the coupled factor $P$:

$$J(P, Q, X, Y) = \| W \otimes (R - PQ) \|_F^2 + \lambda_1(\| U - PX \|_F^2 + \| T - YQ \|_F^2)$$

where $\otimes$ denotes element-wise product and $W$ is used for weighting implicit feedback. To prevent overfitting, a regularization term can be appended to the objective function $J$.

We then aim to find a solution by minimizing the following loss function:

$$J(P, Q, X, Y) = \| W \otimes (R - PQ) \|_F^2 + \lambda_1(\| U - PX \|_F^2 + \| T - YQ \|_F^2) + \lambda_2(\| P \|_F^2 + \| Q \|_F^2 + \| X \|_F^2 + \| Y \|_F^2)$$
Weighting Scheme for Implicit Feedback

- For positive example:

\[ W_{ij} = 1 + \beta f_{ij} \]

where \( \beta \) controls the increase rate of confidence. In the experiment, we set it to be 0.1.

- For negative example:

\[ W_{ij} = 1 - \text{sim}(i, j) \]

where \( \text{sim}(i, j) \) is the cosine similarity between \( U_i \) and \( T_j \).
The low-rank matrices $P$, $Q$, $X$ and $Y$ can be solved by weighted Alternative Least Square.

$$\frac{\partial J}{2\partial P} = (W \otimes (PQ - R))Q^T + \lambda_1(PX - U)X^T + \lambda_2||P||_F$$

Let the partial derivative $\frac{\partial J}{\partial P_i} = 0$, we get

$$P_i = (R_i\tilde{W}_iQ^T + \lambda_1U_iX^T)(QW_iQ^T + \lambda_1XX^T)$$

$$+ \lambda_2(\sum_j W_{ij})I^{-1}$$

(2)

where $\tilde{W}_i$ is a diagonal matrix with entries of $i^{th}$ row in $W$ on the diagonal.
Similarly, let $\frac{\partial J}{\partial Q_j} = 0$, $\frac{\partial J}{\partial X_j} = 0$ and $\frac{\partial J}{\partial Y_i} = 0$, we get

$$Q_j = (R_j^T \tilde{W}_j P + \lambda_1 T_j^T Y)(P^T W_j P + \lambda_1 Y^T Y$$

$$+ \lambda_2 (\sum_i W_{ij}) I)^{-1}$$

$$X_j = U_j^T P (P^T P + \lambda_2/\lambda_1 I)^{-1}$$

$$Y_i = T_i Q^T (QQ^T + \lambda_2/\lambda_1 I)^{-1}$$

where $\tilde{W}_j$ is a diagonal matrix with entries of $j^{th}$ column in $W$ on the diagonal. The update is then repeated until convergence. The computational complexity of the algorithm is $O(Nk^2 mn)$ where $N$ is the number of iterations.
Table: The Matrix Co-factorization for Rich side information and Implicit feedback algorithm

**Algorithm** MCRI

**Input:** $R, U, T, W, k$

**Output:** $P, Q, X, Y$

1: Initialize $P, Q, X, Y$

2: Initialize $W_{ij}$

3: if $R_{ij} = 0$ then

4: $W_{ij} = 1 - \text{sim}(i, j)$

5: else

6: $W_{ij} = 1 + \beta f_{ij}$

7: end if

8: repeat

9: Eqn. (2), Eqn. (3), Eqn. (4) and Eqn. (5)

10: until $P, Q, X, Y$ converge

11: return $P, Q, X, Y$
Experimental Setup

- Nanohub testbed
- We split the data into three parts: the data from year 2001 to 2008 is used for training, the data in 2009 is a validation set, and the 2010 data is for testing
- The training and validation set includes 10,013 users and 4,430 resources, and the test set contains 6,029 users and 3,673 resources
- Evaluation metric: Mean Percentage Ranking (MPR)
Effect of Number of Latent Factors

Figure: Impact of varying the number of latent factors in wAMAN and MCRI
**Effect of Side Information**

**Table:** Comparison of MCRI with different configurations. MCRI₀ incorporates no side information. MCRIₚ incorporates only user information. MCRIₜ incorporates only resource information. MCRIₚt incorporates both user and resource information. MCRI₀ is the baseline in Gain.

<table>
<thead>
<tr>
<th></th>
<th>MCRI₀</th>
<th>MCRIₚ</th>
<th>MCRIₜ</th>
<th>MCRIₚt</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPR</td>
<td>0.245</td>
<td>0.219</td>
<td>0.187</td>
<td>0.164</td>
</tr>
<tr>
<td>Gain%</td>
<td>-</td>
<td>2.6</td>
<td>5.8</td>
<td>8.1</td>
</tr>
</tbody>
</table>
Effect of Weighting Schemes

Table: Comparison of different weighting schemes in MCRI. UNI denotes uniform weighting, UO denotes user-oriented weighting, and IO is item-oriented [?]. UNI, UO and IO only weight on negative instances. PO denotes only weighting on positive instances. “Both” denotes the weighting that combines UO and PO. UNI is the baseline in Gain.

<table>
<thead>
<tr>
<th></th>
<th>UNI</th>
<th>UO</th>
<th>IO</th>
<th>PO</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPR</td>
<td>0.201</td>
<td>0.187</td>
<td>0.194</td>
<td>0.213</td>
<td>0.164</td>
</tr>
<tr>
<td>Gain%</td>
<td>-</td>
<td>1.4</td>
<td>0.7</td>
<td>-1.2</td>
<td>3.7</td>
</tr>
</tbody>
</table>

Yi Fang and Luo Si (Purdue University)
Table: Comparison of MCRI$_{UT}$ with other methods. AMAN is the baseline in Gain.

<table>
<thead>
<tr>
<th>AMAN</th>
<th>wAMAN</th>
<th>wAMAN+CB</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPR</td>
<td>0.320</td>
<td>0.267</td>
</tr>
<tr>
<td>Gain%</td>
<td>-</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>CB</td>
<td>MCRI$_{UT}$</td>
</tr>
<tr>
<td>MPR</td>
<td>0.268</td>
<td>0.164</td>
</tr>
<tr>
<td>Gain%</td>
<td>5.2</td>
<td>15.6</td>
</tr>
</tbody>
</table>
This paper presents a principled approach to exploiting rich user and item information with implicit feedback.

The experiments are conducted on an online scientific community dataset, which has been rarely investigated in the prior work.

The experimental results have shown the proposed model can effectively incorporate the side information and improve the recommendation performance.

In the future work, we will conduct more comprehensive experiments on large-scale recommender systems.