Collaborative Filtering via Ensembles of Matrix Factorizations

Mingrui Wu

Max Planck Institute for Biological Cybernetics, Tübingen, Germany

August 12th, 2007

Introduction

- Collaborative Filtering
  - Rating matrix $\mathbf{Y} = [y_{ij}] \in \mathbb{R}^{m \times n}$, with missing values
  - Predict y_{ij} for given user-item pairs, matrix completion task
- Netflix prize competition, about 480,000 users and 18,000 movie items, 100 million ratings
- Our team: HAT, “Have a Try”.

CF via Matrix Factorization

Low dimensional linear factor model [Hofmann, 2004], it assumes that only a small number of factors can influence the ratings. In a k-factor model

$$\mathbf{Y} \approx \mathbf{U}\mathbf{V}$$

where $\mathbf{U} \in \mathbb{R}^{m \times k}$ and $\mathbf{V} \in \mathbb{R}^{k \times n}$.
For each user-item pair, we have,

$$y_{ij} = \sum_{l=1}^{k} u_{il} v_{jl} = \mathbf{u}_i^\top \mathbf{v}_j$$

User i is modeled by $\mathbf{u}_i \in \mathbb{R}^k$, while item j is modeled by $\mathbf{v}_j \in \mathbb{R}^k$.

Three MF Methods

- RMF, Regularized Matrix Factorization.
- MMMF, Maximum Margin Matrix Factorization.
- NMF, Non-negative Matrix Factorization.

RMF

- Goal: find $\mathbf{U} \in \mathbb{R}^{m \times k}$, $\mathbf{V} \in \mathbb{R}^{k \times n}$, such that $\mathbf{u}_i^\top \mathbf{v}_j \approx y_{ij}$, $ij \in \mathcal{S} = \{ij \mid y_{ij} > 0\}$.
- Objective function:

$$\min_{\mathbf{U} \in \mathbb{R}^{m \times k}, \mathbf{V} \in \mathbb{R}^{n \times k}} \frac{\lambda}{2}(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2) + \sum_{ij \in \mathcal{S}} (y_{ij} - \mathbf{u}_i^\top \mathbf{v}_j)^2$$

where $\lambda > 0$ and $\|\mathbf{U}\|_F^2 = \sum_{pq} u_{pq}^2$.

MMMF

Basic idea:

- factor matrices $\mathbf{U} \in \mathbb{R}^{m \times k}$, $\mathbf{V} \in \mathbb{R}^{k \times n}$
- $r - 1$ thresholds $\theta_{i1}, \ldots, \theta_{i,r-1}$ for each user i

$$\begin{array}{ll} \theta_{ia} > \mathbf{u}_i^\top \mathbf{v}_j & \text{if } a >= y_{ij} \\ \theta_{ia} < \mathbf{u}_i^\top \mathbf{v}_j & \text{if } a < \; y_{ij} \end{array}$$

- $r = 5$, then $y_{ij} = 3 \Leftrightarrow \theta_{i1}, \theta_{i2} < \mathbf{u}_i^\top \mathbf{v}_j < \theta_{i3}, \theta_{i4}$

MMMF

- Objective function [Rennie & Srebro, 2005, DeCoste, 2006]:

$$J(\mathbf{U},\mathbf{V},\theta) = \frac{\lambda}{2}(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2) + \sum_{a=1}^{r-1}\sum_{ij\in\mathcal{S}} h(T_{ij}^a[\theta_{ia} - \mathbf{u}_i^\top \mathbf{v}_j])$$

where

$$T_{ij}^a = \begin{cases} +1 & \text{if } a \geq y_{ij} \\ -1 & \text{if } a < y_{ij} \end{cases}$$

$h(\cdot)$ is a smoothed hinge loss function:

$$h(z) = \begin{cases} \frac{1}{2} - z & \text{if } z < 0 \\ 0 & \text{if } z > 1 \\ \frac{1}{2}(1-z)^2 & \text{otherwise} \end{cases}$$

Hinge Loss and Smoothed Hinge Loss

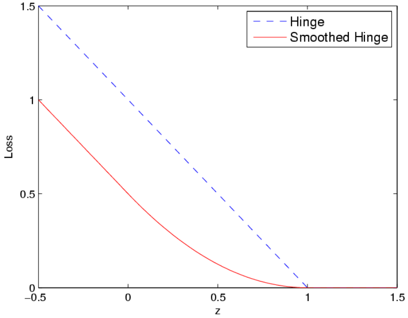

Figure: Hinge loss and smoothed hinge loss

NMF

- Goal: find *non-negative* matrices $\mathbf{U} \in \mathbb{R}^{m \times k}$ and $\mathbf{V} \in \mathbb{R}^{k \times n}$, such that $\mathbf{Y} \approx \mathbf{UV}$.
- Measure of divergence [Lee & Seung, 2000]:

$$D(A||B) = \sum_{ij}(a_{ij} \log \frac{a_{ij}}{b_{ij}} - a_{ij} + b_{ij})$$

  similar to Kullback-Leible divergence.
- To perform NMF, we need to minimize $D(\mathbf{Y}||\mathbf{UV})$ with respect to $\mathbf{U}$ and $\mathbf{V}$, subject to the constraints $u_{il}, v_{jl} \geq 0$

MF Ensembles

Two possible approaches:

- Picking up the ”best” one:
  - one particular MF algorithm
  - using one single set of optimal parameters
- Ensemble learning approach, combine the results obtained with
  - different MF algorithms: RMF, MMMF, NMF
  - different parameters for each MF algorithm: dimensionality, learning rate, regularization parameter, etc.

Ensemble approach has been used for CF in [DeCoste, 2006]. Detailed analysis of ensemble learning methods can be found in [Dietterich, 2002].

Experimental Results

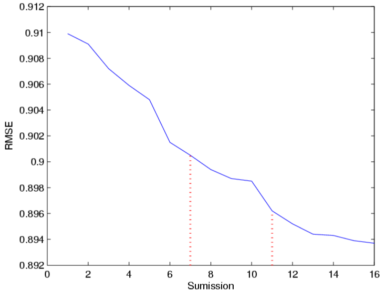

Figure: RMSE of our 16 submissions on the quiz dataset.

Summary

- MF is an effective approach for CF.
- Three current MF approaches: RMF, MMMF and NMF.
- Using ensemble approach to improve the performance.

Bennett, J., & Lanning, S. (2007).
The netflix prize.
Proc. KDD Cup and Workshop.

DeCoste, D. (2006).
Collaborative prediction using ensembles of maximum margin matrix factorizations.
Proc. 23rd International Conference on Machine Learning.

Dietterich, T. G. (2002).
Ensemble learning.
In *The handbook of brain theory and neural networks*. The MIT Press.

Hofmann, T. (2004).
Latent semantic models for collaborative filtering.
ACM Transactions on Information Systems, *22*, 89–115.

Lee, D. D., & Seung, H. S. (2000).
Algorithms for non-negative matrix factorization.
In *Advances in neural information processing systems 12*.

Rennie, J. M. M., & Srebro, N. (2005).

Fast maximum margin factorization for collaborative prediction.
Proc. 22nd International Conference on Machine Learning.