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

Objective function:

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

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

RMF
 MMMMF  

Basic idea:

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

\[
\begin{align*}
\theta_{ia} &> \mathbf{u}_i^\top \mathbf{v}_j & \text{if } a \geq y_{ij} \\
\theta_{ia} &< \mathbf{u}_i^\top \mathbf{v}_j & \text{if } a < y_{ij}
\end{align*}
\]

- $r = 5$, then $y_{ij} = 3 \iff \theta_{i1}, \theta_{i2} < \mathbf{u}_i^\top \mathbf{v}_j < \theta_{i3}, \theta_{i4}$
Objective function [Rennie & Srebro, 2005, DeCoste, 2006]:

$$J(U, V, \theta) = \frac{\lambda}{2}(\|U\|_F^2 + \|V\|_F^2) + \sum_{a=1}^{r-1} \sum_{ij \in S} h(T_{ij}^a[\theta_{ia} - u_i^T v_j])$$

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
Goal: find non-negative matrices $U \in \mathbb{R}^{m \times k}$ and $V \in \mathbb{R}^{k \times n}$, such that $Y \approx UV$.

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-Leibler divergence.

To perform NMF, we need to minimize $D(Y||UV)$ with respect to $U$ and $V$, 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.


Fast maximum margin factorization for collaborative prediction.

*Proc. 22nd International Conference on Machine Learning.*