

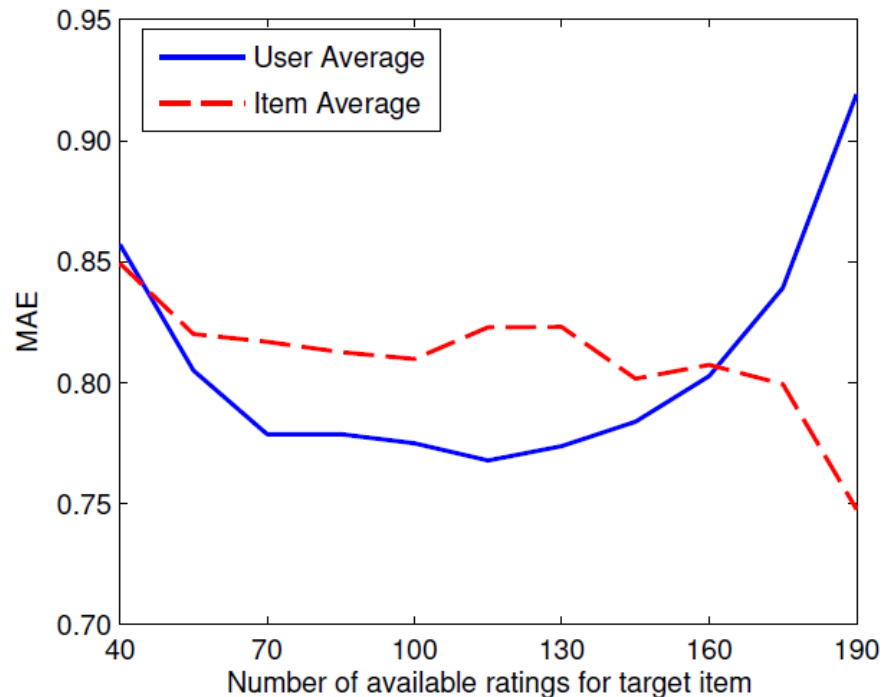
Automatic Feature Induction for Stagewise Collaborative Filtering

Joonseok Lee
College of Computing
Georgia Institute of Technology



Empirical Observation

- ▶ Different recommendation systems perform better than others for some users or items, but not for others.
- ▶ Example:



Ensemble Model

- ▶ Weights in the combination should be **functions of inputs**, rather than constants.

$$F^{(K)}(u, i) = \sum_{k=1}^K \alpha_k(u, i) f_k(u, i)$$

- ▶ Example of features:
 - Rating count for a user
 - Standard deviation of an item

Inducing Local Features

- ▶ **Assumption**: Similar users/items share well-performing CF method on them as well.
- ▶ **Step 1**: Randomly select an anchor user and an anchor item.
- ▶ **Step 2**: Apply kernel smoothing, e.g,

$$K_{h,(u^*,i^*)}^{(1)}(u,i) \propto \left(1 - \frac{d(u^*,u)}{h}\right) I(d(u^*,u) \leq h)$$
$$K_{h,(u^*,i^*)}^{(2)}(u,i) \propto \left(1 - \frac{d(i^*,i)}{h}\right) I(d(i^*,i) \leq h)$$

Stagewise Learning

- ▶ **Greedy** stagewise learning with least squares.

$$(\beta_k, h_k, f_k) = \arg \min_{\beta_k \in \mathbb{R}, h_k \in \mathcal{H}, f_k \in \mathcal{F}} \sum_{(u,i) \in R} \left(F^{(k-1)}(u,i) + \beta_k h_k(u,i) f_k(u,i) - R_{u,i} \right)^2$$

- ▶ **Sampling** is needed as $|\mathcal{H}|$ is very big.
- ▶ Surprisingly, randomly-chosen **small number** of features actually works quite well!

Experiment

