

Recommendation System Construction of Expedia Data via Collaborative Filtering

Menglan Jiang, Yijun Jiang, Muyao Sun, Yizheng Wang, Yizhe Zhang

Duke University

Statistical Science Department

April 2, 2017

Introduction

Motivation:

- Based on the given user-hotel data, our group attempts to build up a recommender system, which is used to predict the "preference" that a user would give to a hotel.

	Item 1	Item 2	Item 3	...	Item n
User 1	2	3	?	...	5
User 2	?	4	3	...	?
User 3	3	2	?	...	3
...
User m	1	?	5	...	4

User-based and Item-based Collaborative Filtering

Adjusted Cosine Similarity

$$u_i \in R^m, i \in (1, \dots, K); v_j \in R^n, j \in (1, \dots, L)$$

$$\text{sim}(u_i, u_{i'}) = \frac{(u_i - \bar{u})^T W_1 (u_{i'} - \bar{u})}{\sqrt{(u_i - \bar{u})^T W_1 (u_i - \bar{u})} \sqrt{(u_{i'} - \bar{u})^T W_1 (u_{i'} - \bar{u})}}, \text{ same for } v_j$$

$$\text{weight}(i, i') = (1 + \text{sim}(i, i'))/2$$

Estimating Score

$$\hat{S}_{ij} = \beta_0 + \beta_1 \frac{\sum_{i'} S_{i'j} \sigma_{ii'}}{\sum_{i'} I_{i'j} \sigma_{ii'}} + \beta_2 \frac{\sum_{i'} S_{i'j} \lambda_{ii'}}{\sum_{i'} I_{i'j} \lambda_{ii'}}$$

Loss Function

$$L(\hat{S}_{ij}, S_{ij}) = \sum_{ij} I_{ij} (\hat{S}_{ij} - S_{ij})^2$$

Result

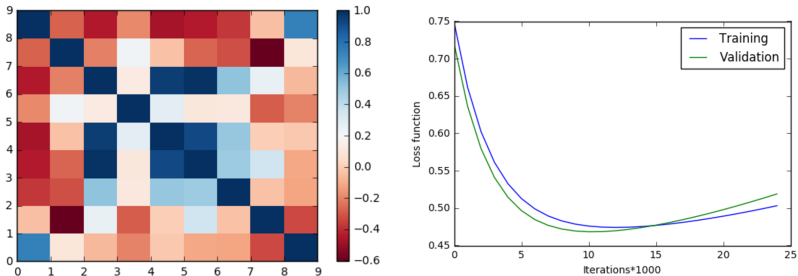


Figure: Left: similarity matrix. Right: learning curve

```
In [423]: iter, loss_train,
          loss_test, pred = loss(beta, train) # output the prediction error
          beta

Out[423]: array([ 0.73371751,  1.23797033, -0.36479982])
```

Figure: Predicted coefficient