Collaborative Filtering

based on user similarity

Zhe Xie, Kugelmann Stephan,
Massé Benoit, Freitag Francois

13/12/2012

Task

Several users already rated an item

One user haven't rated it yet
→ Will he like it ?

Task

Several users already rated an item

One user haven't rated it yet
→ Will he like it ?

Goal : Predict his rating for this item
How : Similarity between users

Main issue

How to measure the similarity?

Efficiency of the algorithm ?

Algorithm

```
message  #     Ken  Lee  Meg  Nan
        1       1    4    2    2
        2       5    2    4    4
        3                 3
        4       2    5         5
        5       4    1         1
        6       ?    2    5    ?
```

Figure 5: a sample matrix of ratings.

Step 1: Sample matrix

Algorithm

message #	Ken	Lee	Meg	Nan
1	1	4	2	2
2	5	2	4	4
3			3	
4	2	5		5
5	4	1		1
6	?	2	5	?

Figure 5: a sample matrix of ratings.

Step 1: Sample matrix

$$r_{KL} = \frac{Cov(K, L)}{\sigma_K \sigma_L}$$

$$= \frac{\sum_i \left(K_i - \bar{K}\right)\left(L_i - \bar{L}\right)}{\sqrt{\sum_i \left(K_i - \bar{K}\right)^2} \sqrt{\sum_i \left(L_i - \bar{L}\right)^2}}$$

$$= \frac{-2 - 2 - 2 - 2}{\sqrt{10} \sqrt{10}} = -0.8$$

Step 2: Define the covariance matrix

Algorithm

message #	Ken	Lee	Meg	Nan
1	1	4	2	2
2	5	2	4	4
3			3	
4	2	5		5
5	4	1		1
6	?	2	5	?

Figure 5: a sample matrix of ratings.

Step 1: Sample matrix

1	-0.8	0.6	0
-0.8	1	-0.5	0.6
0.6	-0.5	1	0.3
0	0.6	0.3	1

$$r_{KL} = \frac{Cov(K,L)}{\sigma_K \sigma_L}$$

$$= \frac{\sum_i (K_i - \bar{K})(L_i - \bar{L})}{\sqrt{\sum_i (K_i - \bar{K})^2} \sqrt{\sum_i (L_i - \bar{L})^2}}$$

$$= \frac{-2 - 2 - 2 - 2}{\sqrt{10}\sqrt{10}} = -0.8$$

Step 2: Define the covariance matrix

$$K_{6_{Pred}} = \bar{K} + \frac{\sum_{J \in raters}(J_6 - \bar{J})r_{KJ}}{\sum_J |r_{KJ}|} =$$

$$3 + \frac{2r_{KM} - r_{KL}}{|r_{KM}| + |r_{KL}|} = 3 + \frac{2 - (-.8)}{|1| + |-.8|} = 4.56$$

Step 3: Define the rating based on the others users weighted rating

Introduction to the dataset

Dataset: MovieLens data set

collected by the GroupLens Research Project at the University of Minnesota

http://ict.ewi.tudelft.nl/~jun/CollaborativeFiltering.html

User	Movie	Grade
1	1	4
2	1	1
3	2	5

Extraction from the dataset

943 users
1682 movies
Grades from 1 to 5
100 000 ratings

Histogram