Изготвили:

Сияна славова, 24963

Иван Капукаранов, 24958

Проект за курса „Препоръчващи системи“

Задача: recSys 2013 challenge

Съдържание

[Проблем 1](#_Toc454038202)

[Съществуващи решения 1](#_Toc454038203)

[Избрано решение 1](#_Toc454038204)

[Експерименти 1](#_Toc454038205)

[Резултати 1](#_Toc454038206)

[Използвана литература 1](#_Toc454038207)

# Проблем

Данните за задачата са предоставени от [Yelp](http://www.yelp.com/), като съдържат над:

* 10,000 фирми (business)
* 8,000 check – in sites
* 40,000 потребителя (users)
* 200,000 ревюта (reviews)

Трябва да се предскаже рейтинга, който потребител ***u*** би дал на продукт ***i***. За целта са дадени обучаващо и тестово множество, като тестовото множество съдържа само колони потребител и фирма.  
  
Оценяването на алгоритъма става чрез използването на средно квадратична грешка.

# Съществуващи решения

# Избрано решение

За реализиране на задачата използваме [Apache Mahout](http://mahout.apache.org/). За предварителна обработка на данните и оценяване на алгоритмите използваме [R](https://www.r-project.org/).

# 

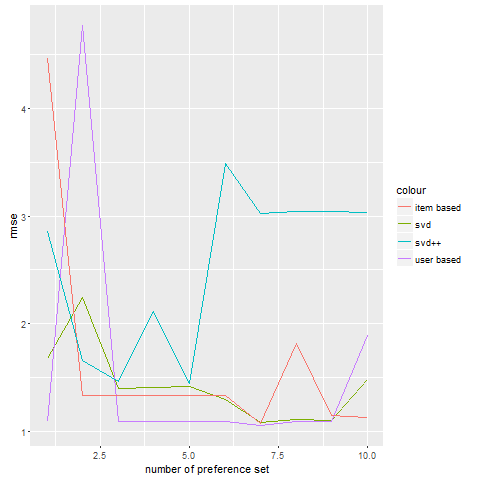
# Експерименти

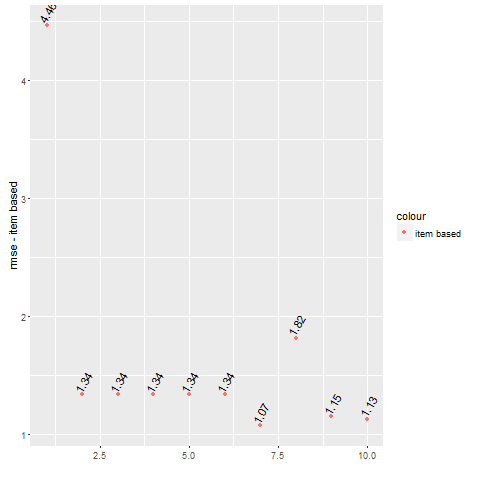
## Параметри

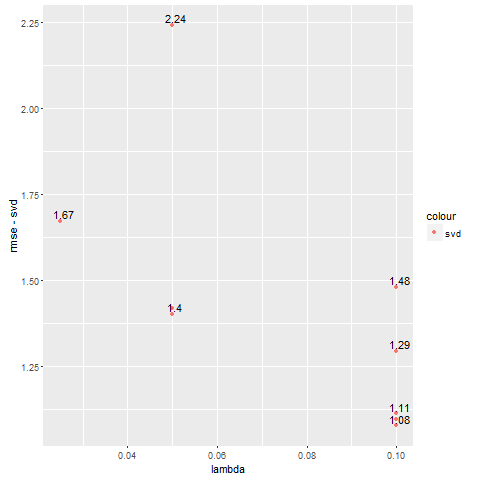
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| File name | userBased | svd | svd++ | cold start | hybrid |
| preferences\_1 | threshold: 0.05 | numFeatures: 2 lambda: 0.05 numIterations: 100 | numFeatures: 2 numIterations: 20 |  |  |
| preferences\_2 | threshold: 0.05 | numFeatures: 2 lambda: 0.05 numIterations: 100 | numFeatures: 2 numIterations: 20 | if item based is used, get item avg, else get user avg |  |
| preferences\_3 | threshold: 0.1 | numFeatures: 2 lambda: 0.1 numIterations: 100 | numFeatures: 2 numIterations: 10 | if item based is used, get item avg, else get user avg |  |
| preferences\_3\_hybrid | threshold: 0.1 | numFeatures: 2 lambda: 0.1 numIterations: 100 | numFeatures: 2 numIterations: 10 | if item based is used, get item avg, else get user avg | 0.5\*userBased + 0.5 \* other |
| preferences\_3\_hybrid\_weighted | threshold: 0.1 | numFeatures: 2 lambda: 0.1 numIterations: 100 | numFeatures: 2 numIterations: 10 | if item based is used, get item avg, else get user avg | weighted: 0.6\*userBased + 0.4 \* other |
| preferences\_4 | threshold: 0.025 | numFeatures: 2 lambda: 0.025 numIterations: 100 | numFeatures: 2 numIterations: 30 | if item based is used, get item avg, else get user avg |  |
| preferences\_5 | threshold: 0.2 | numFeatures: 2 lambda: 0.05 numIterations: 50 | numFeatures: 3 numIterations: 20 | if item based is used, get item avg, else get user avg |  |
| preferences\_6 | threshold: 0.25 | numFeatures: 2 lambda: 0.05 numIterations: 150 | numFeatures: 3 numIterations: 30 | if item based is used, get item avg, else get user avg |  |
| preferences\_7 | threshold: 0.5 | numFeatures: 2 lambda: 0.1 numIterations: 100 | numFeatures: 2 numIterations: 10 | get avg of item avg and user avg |  |
| preferences\_8 | threshold: 0.1 | numFeatures: 5 lambda: 0.1 numIterations: 100 | numFeatures: 5 numIterations: 10 | get avg of item avg and user avg |  |

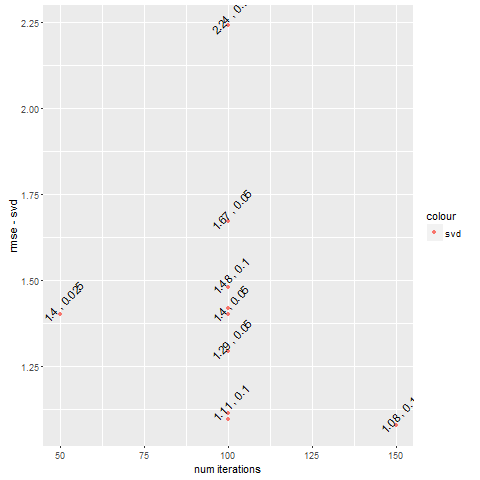
# Резултати

## Графики

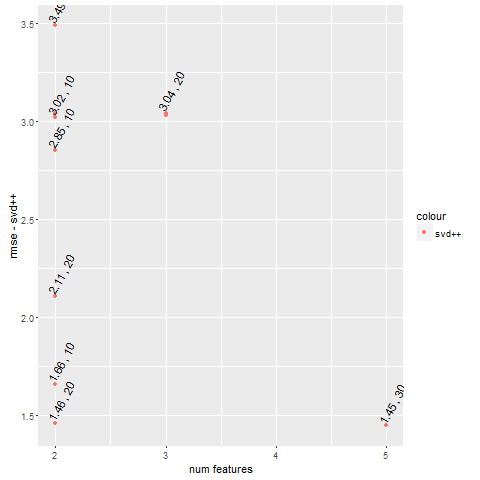




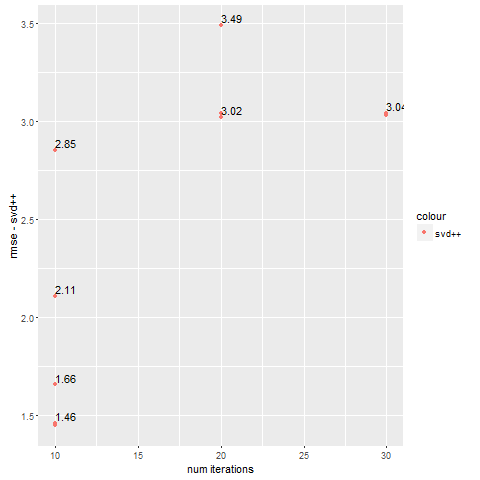


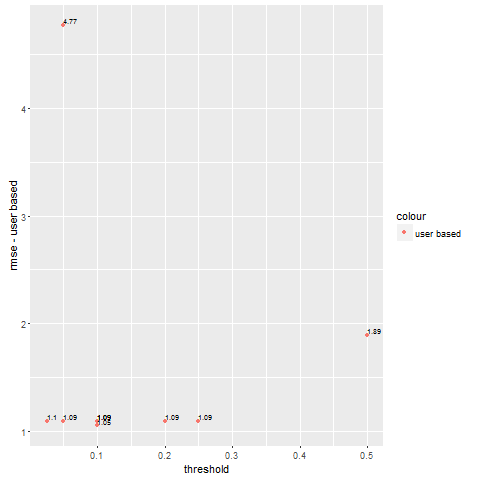


Фигура 1: Грешка, lambda



Фигура 2: Грешка, Брой итерации





## RMSE

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| File name: | user based: | item based: | svd: | svd++: |
| preferences\_1 | train: 1.0410  test: 4.7702 | train: 1.4466  test: 4.4641 | train: 2.3610  test: 2.2418 | train: 2.3002  test: 3.4902 |
| preferences\_2 | train: 1.1281  test: 1.0949 | train: 1.5795  test: 1.3360 | train: 2.3296  test: 1.4017 | train: 2.2153  test: 3.0218 |
| preferences\_3 | train: 0.9950  test: 1.0948 | train: 1.4686  test: 1.3360 | train: 1.7947  test: 1.2947 | train: 1.6787  test: 2.8548 |
| preferences\_3\_hybrid | train: 1.0230  test: 1.0948 | train: 1.5111  test: 1.1489 | train: 1.9077  test: 1.1135 | train: 1.7298  test: 1.6571 |
| preferences\_3\_hybrid\_weighted | train: 1.0872  test: 1.0948 | train: 1.5787  test: 1.1281 | train: 1.8786  test: 1.0980 | train: 1.4379  test: 1.4604 |
| preferences\_4 | train: 1.1274  test: 1.0952 | train: 1.4988  test: 1.3360 | train: 3.4880  test: 1.6725 | train: 2.5378  test: 3.0412 |
| preferences\_5 | train: 1.1339  test: 1.0941 | train: 1.5411  test: 1.3360 | train: 2.0570  test: 1.4030 | train: 2.1565  test: 3.0403 |
| preferences\_6 | train: 1.1074  test: 1.0944 | train: 1.5931  test: 1.3360 | train: 2.5270  test: 1.4187 | train: 2.4845  test: 3.0311 |
| preferences\_7 | train: 1.1615  test: 1.8926 | train: 1.5152  test: 1.8163 | train: 1.8921  test: 1.4814 | train: 1.6597  test: 2.1100 |
| preferences\_8 | train: 1.0518  test: 1.0496 | train: 1.5910  test: 1.0726 | train: 1.9671  test: 1.0781 | train: 1.6125  test: 1.4467 |

# Използвана литература

* Идея: <https://www.kaggle.com/c/yelp-recsys-2013>
* Apache Mahout: <http://mahout.apache.org/>