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

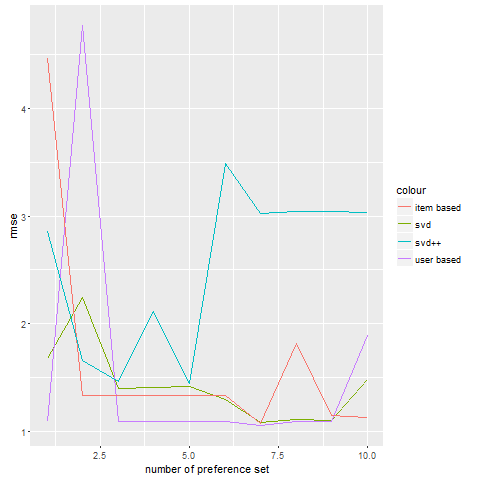
За реализиране на задачата използваме [Apache Mahout](http://mahout.apache.org/). За предварителна обработка на данните и оценяване на алгоритмите използваме [R](https://www.r-project.org/).

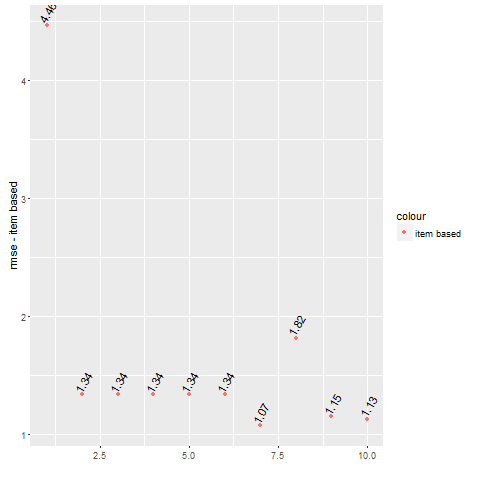
# 

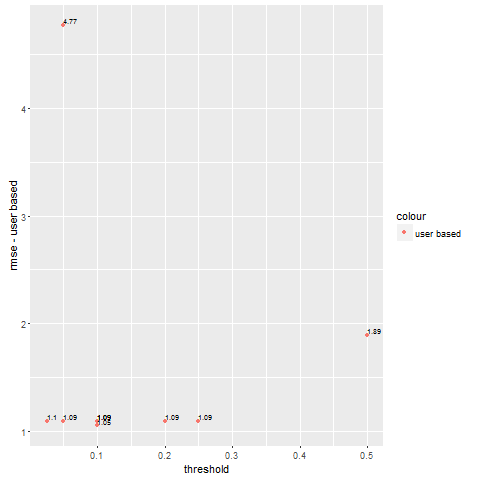
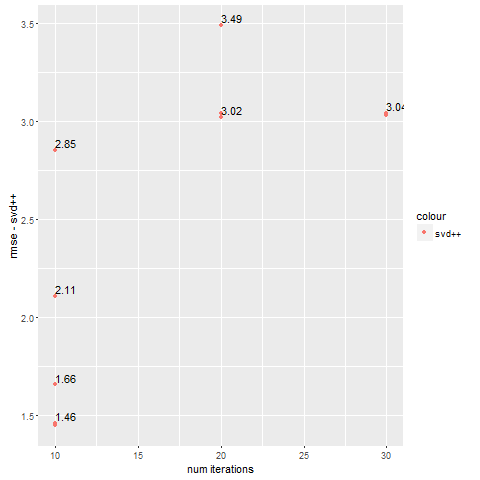
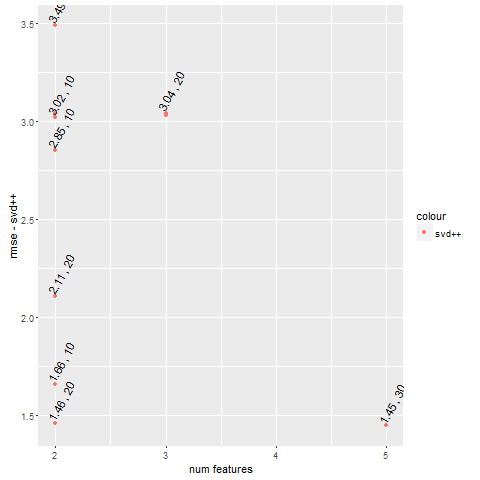
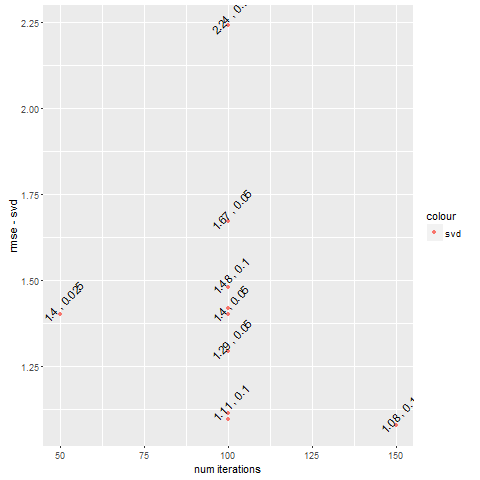
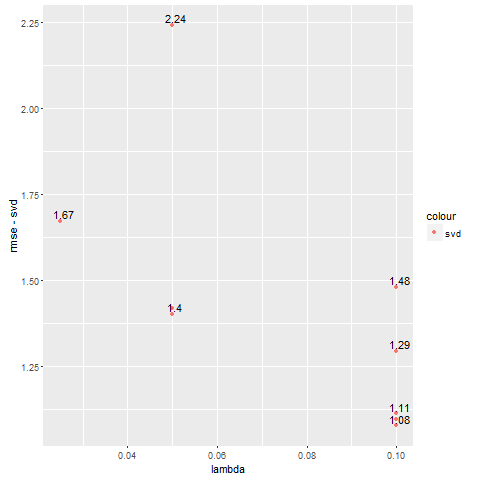
# Експерименти

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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 |  |  |
| 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 |  |  |
| preferences\_5 | threshold: 0.2 | numFeatures: 2 lambda: 0.05 numIterations: 50 | numFeatures: 3 numIterations: 20 |  |  |
| preferences\_6 | threshold: 0.25 | numFeatures: 2 lambda: 0.05 numIterations: 150 | numFeatures: 3 numIterations: 30 |  |  |
| 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, най –добро при preference 8 user based ~1.05

# Използвана литература

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