Master Thesis
Recommender Systems Comparison

Vasileios Symeonidis

27-05-2017

Contents

Part I
Master Thesis

1 Intro

This is the intro suction for this master thesis. Why we need recommendation systems? Retailers can propose the right product to the right target group. User get advertisements the may be interested in.[2]

History, what has been tried so far?

2 Collaborative filtering

What is collaborative filtering. [2]

2.1 Content based

$$w = R^{-1}M^T \tag{1}$$

Normalized

$$w = (\lambda I + R^T R)^{-1} R^T M \tag{2}$$

2.2 Latent Factors

test sadsad

sad

asd

ds

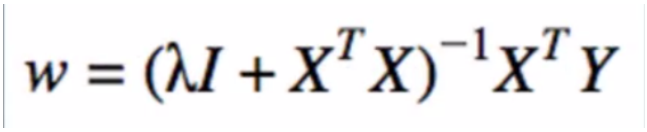

Figure 1: **Equation**

3 Our Experiment

3.1 Infrastructure

3.1.1 Apache Spark

What is map reduce How spark differentiates from its predecessors, hadoop yarn

Resilient Distributed Datasets (RDDs)
mllibs
add spark jira note broadcasting rdd //cite the mastering apache spark book [1]
new trends on spark https://github.com/apache-spark-on-k8s/spark

3.2 Dataset

[3]

3.3 Metrics

3.3.1 Mean Absolute Error

$$\sum_{0}^{i} x_i^2 \tag{3}$$

3.3.2 Execution Time

$$\sum_{0}^{i} x_i^2 \tag{4}$$

4 Results

Table 1: **Content Based Algorithm Results**

Training Dataset	Testing Dataset	Mean Absolute Error	Execution time (ms)
ml-100k/u1.base	ml-100k/u1.test	1.6467431428213226	30514
ml-100k/u2.base	ml-100k/u2.test	1.6055222166704628	27714
ml-100k/u3.base	ml-100k/u3.test	1.608925907479106	27164
ml-100k/u4.base	ml-100k/u4.test	1.6259192043203685	26687
ml-100k/u5.base	ml-100k/u5.test	1.6284658627202895	27124
ml-100k/ua.base	ml-100k/ua.test	1.6425364580036836	26640
ml-100k/ub.base	ml-100k/ub.test	1.6357196576385744	26861

5 Conclusion

As a conclusion we can see that als is better on both metrics from the content based.

Table 2: **Latent Factors Algorithm Results**

Training Dataset	Testing Dataset	Mean Absolute Error	Execution time (ms)
ml-100k/u1.base	ml-100k/u1.test	1.1818684937209607	10195
ml-100k/u2.base	ml-100k/u2.test	1.1800652808093945	6517
ml-100k/u3.base	ml-100k/u3.test	1.1783366748334452	5377
ml-100k/u4.base	ml-100k/u4.test	1.1730543877181654	5433
ml-100k/u5.base	ml-100k/u5.test	1.1686585291940668	5217
ml-100k/ua.base	ml-100k/ua.test	1.2008035300836668	5214
ml-100k/ub.base	ml-100k/ub.test	1.2134460078406009	5083

6 References

References

[1] "Apache Spark lightning-fast cluster computing." `https://spark.apache.org/`. Accessed: 2017-05-21.

[2] P. Melville and V. Sindhwani, "Recommender systems," *Encyclopedia of Machine Learning and Data Mining*, pp. 1056–1066, 2017.

[3] "MovieLens grouplens." `https://grouplens.org/datasets/movielens/`. Accessed: 2017-05-22.

Part II
Appendices

A Code used

A.1 User Based Collaborative Filtering

A.2 Product Based Collaborative Filtering

A.3 Latent Factors

A.4 infra code

B Metrics

B.1 What is the mean absolute error

B.2 Time

List of Tables

List of Figures