

Chan Wai Ming Student: Supervisor: Dr. SO, H C

EE4382 Final Year Project 2013-2014

E-Commerce Collaborative Filtering System based on Cloud Computing

1. Introduction

E-Commerce industry is growing big since more people are buying and selling things in the Internet nowadays. The customer transaction histories are valuable the company for make to recommendation or advertisement, which can be done by Collaborative Filtering. The industry is collecting TBs or even PBs of data every day in which traditional data analysis may not be capable to handle vast volume of data. Therefore, distributive Collaborative Filtering algorithm is needed in Big Data Analytics.

2. Objective

Our project uses Netflix Prize dataset which is a real world dataset containing 100 millions of users' movie ratings. The dataset forms a 2D sparse matrix. Our goal is to predict the missing elements in the sparse matrix. As a result, user preference of interests are obtained.



4. Cloud Computing

Our project has developed a distributed stochastic gradient descent algorithm to solve the big matrix factorization problem.

Programme: BENG3-INFE



With the use of Amazon Web Services (AWS), a computer cluster is setup easily. Our distributive program is run on the computer cluster which brings several advantages:

- 1) execution time decreases with more computers;
- 2) run in low cost machine;
- 3) computer cluster hardware is configurable (e.g. CPU, memory and storage size);
- 4) data storage is unlimited scalable.





The low-rank matrix can be factorized into two factor matrices. By using stochastic gradient descent updating rule, the factor matrices are estimated from the training dataset in an iterative manner. After the training process, the prediction matrix is formed by multiplying the factor matrices together.



Based on the Netflix dataset, it is shown that our algorithm gives more accurate prediction in terms of root mean squared error (see yellow line).