Amogh Param, Brandon Wu

brandonwu@cs.ucla.edu

Abstract

[Draw your reader in with an engaging abstract. It is typically a short summary of the document.   
When you’re ready to add your content, just click here and start typing.]

EE239 Project 1

Collaborative Filtering

# Introduction

In this report we study a matrix representation of a Recommendation System that utilizes Collaborative Filtering. There is much industry interest in such systems; for example the Netflix suggestion system for its users may use a scheme similar to that presented here. The problem we will study in detail is the following: based on feedback data from users indicating a rating of items, we construct a matrix,where the element corresponds to the rating given by user on item*.*

We would like to extrapolate this matrix to infer which unrated items that a user will likely enjoy. Therefore, we adopt the following construction: where the rows of are vectors that characterize a particular user, the columns of characterize a particular item, and represents the error, parameterized by a factor.

The factorization algorithm used throughout this report is the *Alternating Least Squares Method* which seeks to minimize the following cost function:

(Equation 1)

We will build the recommendation system based on a dataset consisting of 100K movie ratings collected by GroupLens[[1]](#footnote-1).

# Part 1: A Simple Factorization Using ALS

Our Recommendation System uses the NMF MATLAB toolbox[[2]](#footnote-2) to perform the matrix factorization that minimizes the cost function in Equation 1. In this section, we used the NMF tools to factorize. The metric we used to determine the closeness of the factorization was the squared error (Equation 1). We only considered entries that of the prediction matrix that corresponded to actual prediction the user actually made, so Equation 1 was modified by an additional weight matrix where if user rated movie and 0 otherwise. We also varied the value of parameter to determine the optimal size of matrices and. Small values of encode less information which makes the reconstruction of less accurate but large values of could result in over-fitting the dataset. The results for squared error versus parameter are shown in Table 1 below:

Table 1: Squared Error vs k

|  |  |  |  |
| --- | --- | --- | --- |
| k | 10 | 50 | 100 |
| sq. error | 5.81E+04 | 4.39E+05 | 3.71E+05 |

Although the squared error decreased with increasing, the difference is not significant.

# Part 2: A Modified Construction Using Non-Trivial Weights

In the second part, we apply the same cost function but this time, we reverse the roles of the and (weight) matrices in the factorization step. We again applied the ALS factorization algorithm for different values of parameter and measured the squared error. The results are summarized below in Table 2.

Table 2: Squared error vs k

|  |  |  |  |
| --- | --- | --- | --- |
| k | 10 | 50 | 100 |
| sq. error | 5.71E+04 | 2.50E+04 | 1.18E+04 |

# Part 3: 10-Fold Cross Validation

In this section, we show results from the same system built using 10-Fold Cross validation. The original dataset is transformed via random permutation, then divided into a testing and training set. The training set is used to construct the matrix which is factorized into and. The values of the prediction matrix are validated against the ratings in the testing set. We define a new metric based on this prediction error shown below in Equation 2.

Equation 2

Where is the testing set and is the prediction matrix defined by and. The results of the 10-Fold Cross Validation is shown below in Table 3.

Table 3: Ten-Fold Cross Validation Prediction Error

|  |  |
| --- | --- |
| fold | error |
| 1 | 1.5592 |
| 2 | 1.5476 |
| 3 | 1.5413 |
| 4 | 1.5685 |
| 5 | 1.537 |
| 6 | 1.5629 |
| 7 | 1.5686 |
| 8 | 1.538 |
| 9 | 1.5632 |
| 10 | 1.5688 |

The minimum error occurred in the 5th fold and the maximum error occurred in the 10th fold. The average error is.

# Precision And Recall

1. http://grouplens.org/datasets/movielens/ [↑](#footnote-ref-1)
2. https://sites.google.com/site/nmftool/ [↑](#footnote-ref-2)