Name(s)	Project Number
Andrew B. Liu	
	30389

Project Title

Collaborative Filtering in Large-Scale Recommender Systems

Abstract

Objectives/Goals

Recommender systems attempt to suggest specific products to the customers most likely to buy them, and play a large role in e-commerce and business revenues. Most recommender systems employ collaborative filtering (CF), which makes personal recommendations based on the ratings of other customers with similar preferences. While CF has achieved some success, it is limited by real-world scalability and sparsity. The Singular Value Decomposition (SVD) predictor is widely used in CF to overcome these barriers, because it saves memory through approximation and numerically defines the relationship between item and user attributes. This research investigates several aspects of SVD, mainly the number of features, bias incorporation, and sparsity.

Methods/Materials

I ran experiments on the large and genuine Netflix Prize data, and measured the accuracy of each run with root mean squared error (RMSE) between predicted and actual ratings for a test set of item-user pairs. I conducted experiments on my Dell Inspiron, with 320GB of hard disk drive and 4GB of RAM. Program files were written in Java and adapted from oknetflix's implementation.

Results

Experimental results first showed a convergent relationship between number of features and RMSE. After 100 features, the successive addition of 10 more features never lowered the RMSE by more than 0.0003. The second experiment showed that bias had limited impact on SVD, with an RMSE difference of 0.0002 between SVD's with and without bias. The sparsity experiment illustrated a positive linear correlation between sparsity (percent of data removed) and RMSE ($r^2 = 0.9992$), with an LSRL of $y = (0.000426)x + 0.9077$.

Conclusions/Discussion

I conclude that tackling sparsity, not features or bias, has the largest impact on and greatest potential for improving SVD. Sparsity should be a focus in future SVD and CF research and in businesses' data gathering considerations.

Summary Statement

My project identifies and quantifies sparsity as a larger influence than features or bias on the Singular Value Decomposition predictor for collaborative filtering.

Help Received

Dad looked over research paper and advised me on experimentation