UROC Research Probabilistic Machine Learning

Jacob Schopmeyer

Indiana University

Abstract

For my research, my goal was to replicate research previously done on probabilistic machine learning which utilized a Variational Bayes in an attempt to improve upon Netflix's Cinematch movie recommendation system. Netflix provides movie recommendations for each of its users based upon what a user scores similar or related movies that other users have watched. It utilizes its Cinematch algorithm to analyze what a user has watched compared to what others have watched to provide recommendations of movies that the user would more likely enjoy. [1]

Models

For this problem, Netflix possesses a partially populated matrix that contains all of the scores users have rated movies; however, no user has seen every movie and no movie has been seen by every person; therefore, we need a way to analyze the overlaps and averages to determine what movies would best fit for each user. We found that a variational inference model would best suit this partially observable space, since we cannot predict everything exactly. In our problem, we utilize the equation below for the variational free energy:

$$\mathbb{E}_{Q(U,V)}[\log p(M,U,V) - \log Q(U,V)] \tag{1}$$

Within the equation, the values of U represent all the ratings each user gives and the V represents all the scores given for each movie. While M is a matrix which is IxJ size, where U is the row and V is the columns, where each element in the matrix can be referenced by $M_{i,j}$. We break down the $P(M,U,V)$ into $P(M|U,V)*P(U,V)$ to allow us to condition for the M. $p(U)$ and $p(V)$ are prior values that are able to be calculated by using their respective variances and rating values. These priors are then conditioned upon to estimate values of M for $P(M|U,V)$. This is calculated by multiplying $m_{i,j}$ by the inner product of U_i and V_j, while accounting for Gaussian noise with τ^2. The second part of (1), $\log Q(U,V)$, represents the variational distribution which we find and use in training the model. We then utilize the Kullback-Leibler divergence which allows us to optimize the algorithm by finding the difference between these two distributions P and Q from (1).

Methods

For this data, we look to compare two different models a baseline and a probabilistic matrix factorization bayes algorithm. **The Baseline algorithm** utilizes the average scores a user gave for movies multiplied by the average scores that the movie received divided by 2 to get a predicted score for that movie for the specific user. **Probabilistic Bayes Factorization Matrix** utilizes the algorithm and equation described in the Models section to train itself and predict values for each of the scores with a prediction matrix.

Results

The goal of my research was to learn how to implement the Variational Bayes approach and compare it to a baseline approach for the Netflix recommendation problem. To measure the accuracy of each of these two models, I utilized RSME (root-mean-square-deviation) to assess the difference between the actual score and the predicted score (where a low rmse is better).

$$RMSE = \sqrt{\frac{\Sigma(Predicted - Actual)^2}{N}} \tag{2}$$

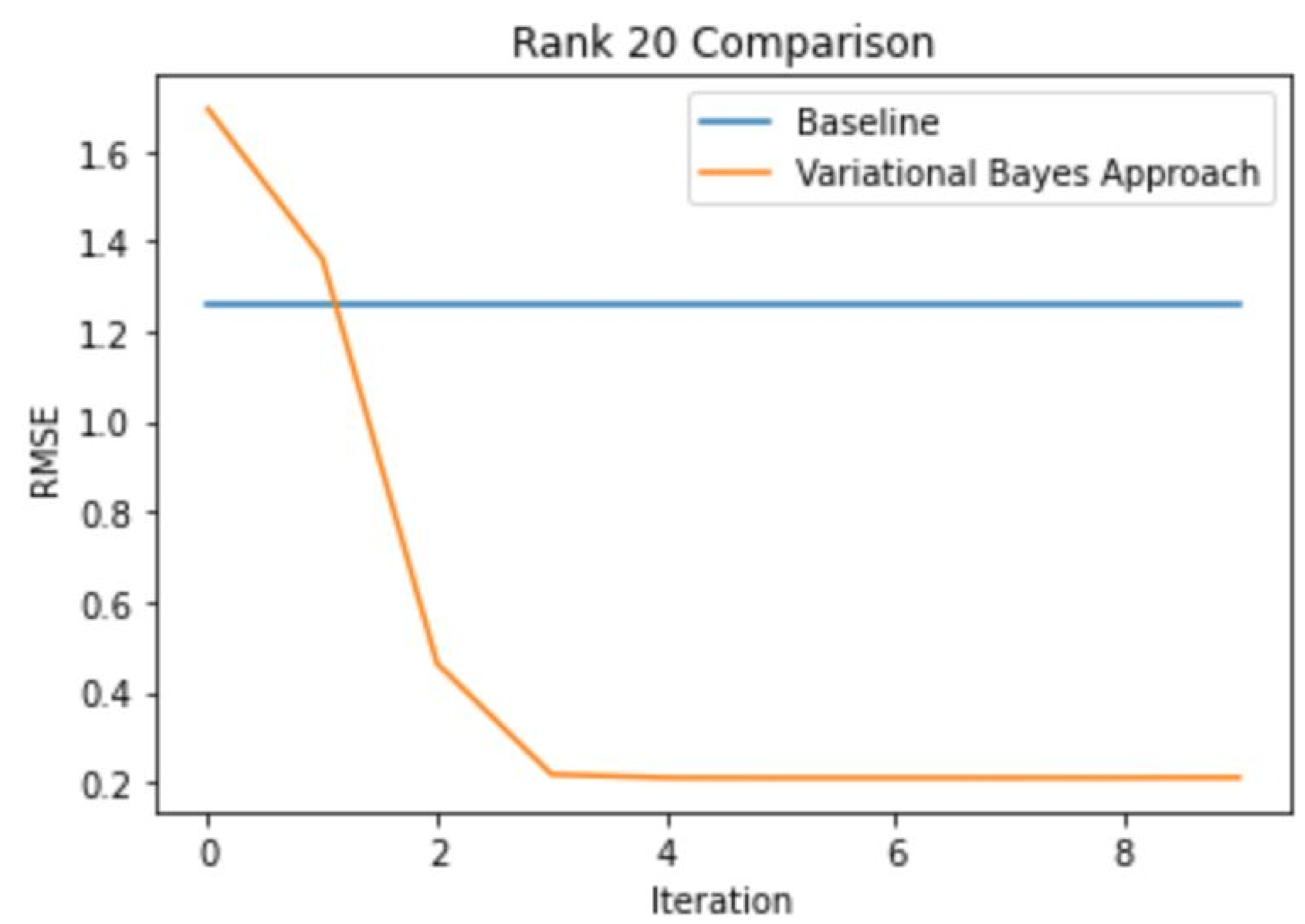

Fig. 1: Plot of Rank 20 Comparison

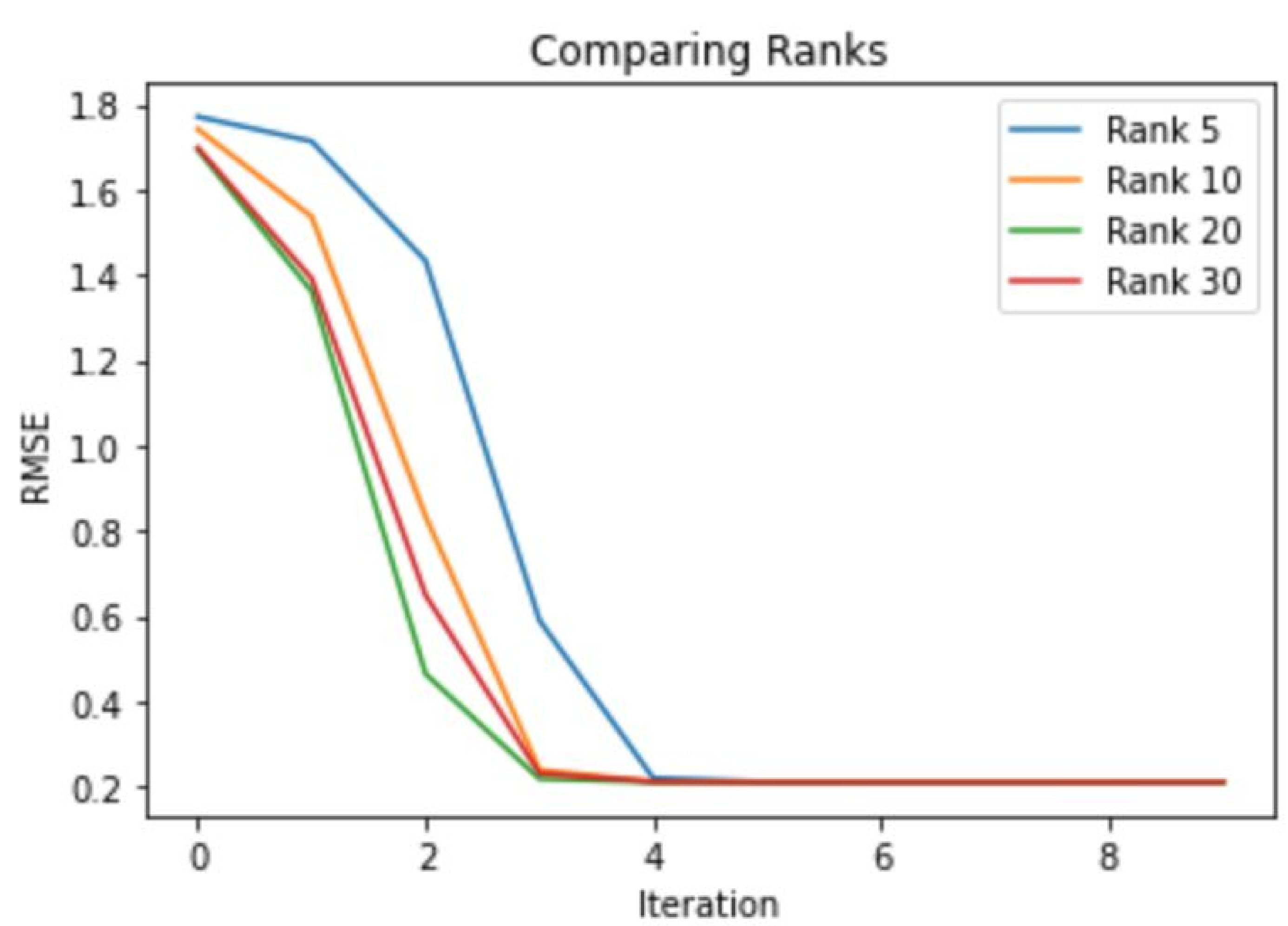

Fig. 2: Plot comparing ranks

Analysis

Looking at the Figure 1, we can see the the Variational Bayes approach does vastly better than the RMSE approach in predicting which movies a user would like after a couple iterations. However, this can be expected since the baseline algorithm relies purely on the means of scores and does not alter it's approach nor does it 'learn'. While the variational bayes approach, initially sets default values and scores for the predictions, but after each iteration it reassesses and redefines these default values to better predict a person's rating of a movie. As we can see in the graphic above, around the 3rd iteration for the rank 20 variational bayes converges to around 0.21.

Figure 2 is comparing how the probabilistic bayes matrix factorization model varies in convergence for different ranks. The rank determines the size that the values used for calculating the RMSE will be decomposed to. The higher the rank the more values it stores, but the longer it takes to run.

Future Work

Implementing and utilizing direct loss minimization to improve upon the predictions.

References

[1] Yew Lim and Yee Teh. "Variational Bayesian Approach to Movie Rating Prediction". In: (2007).