Netflix Prize Analysis

The Netflix prize was announced on October 2nd 2006 and soon became quite the phenomenon in the tech world, attracting over 20,000 teams within the first year. To qualify for the Grand Prize of a million dollars, the root mean squared error (RMSE) of a Participant's submitted predictions on the test subset must be less than or equal to 90% of 0.9525, or 0.8572 (the "qualifying RMSE").

At the beginning of the netflix competition the most common method used to provide solutions was the nearest neighbour algorithm, which presents the most similar objects to the input where similarity can be measured by a variety of metrics, with the euclidian metric being one of them. Soon more sophisticated systems started becoming the norm, and a breakthrough was achieved when a certain "Simon Funk" suggested using Singular Value Decomposition to aid with the recommendations.

Ensemble methods were used by nearly all the leading teams, and with reason, as they largely improve upon results by individual models. As long as the models are not highly correlated, we can expect to see an increase in the performance. The basic idea is that the more diverse the constituent models, the better the performance of the aggregate; theoretically, most linear and non-linear aggregate models always outperform the constituent models.