All Together Now: A Perspective on the NETFLIX PRIZE

Robert M. Bell, Yehuda Koren, and Chris Volinsky

When the Netflix Prize was announced in October of 2006, we initially approached it as a fun diversion from our 'day jobs' at AT&T. Our group had worked for many years on building profiles of customer patterns for fraud detection, and we were comfortable with large data sets, so this seemed right up our alley. Plus, it was about movies, and who doesn't love movies? We thought it would be a fun project for a few weeks.

Boy, were we wrong (not about the fun part, though). Almost three years later, we were part of a multinational team named as the winner of the $1 million prize for having the greatest improvement in root mean squared error (RMSE) over Netflix's internal algorithm, Cinematch.

The predominant discipline of participants in the Netflix Prize appears to have been computer science, more specifically machine learning. While something of a stereotype, machine learning methods tend to center on algorithms (black boxes), where the focus is on the quality of predictions—rather than 'understanding' what drives particular predictions.

In contrast, statisticians tend to think more in terms of models with parameters that carry inherent interest for explaining the world. Leo Breiman's article, "Statistical Modeling: The Two Cultures," which was published in *Statistical Science*, provides various views on this contrast. Our original team consisted of two statisticians and a computer scientist, and the diversity of expertise and perspective across these two disciplines was an important factor in our success.

Fundamental Analysis Challenge

The Netflix Prize challenge concerns recommender systems for movies. Netflix released a training set consisting of data from almost 500,000 customers and their ratings on 18,000 movies. This amounted to more than 100 million ratings. The task was to use these data to build a model to predict ratings for a hold-out set of 3 million ratings. These models, known as collaborative filtering, use the collective information of the whole group to make individualized predictions.

Movies are complex beasts. Besides the most obvious characterization into genres, movies differ on countless dimensions describing setting, plot, characters, cast, and many more subtle features such as tone or style of the dialogue. The Movie Genome Project (*www.jinni.com/movie-genome.html*) reports using "thousands of possible genes." Consequently, any finite model is likely to miss some of the signal, or explanation, associated with people's ratings of movies.

On the other hand, complex models are prone to overfitting, or matching small details rather than the big picture—especially where data are scarce.