The $1 Million Netflix Challenge
VP Jim Bennett discusses how recommendation systems suggest your next movie and the challenges of building a better one.

By Kate Greene

Earlier this week, Netflix, the online movie rental service, announced it will award $1 million to anyone who can come up with an algorithm that improves the accuracy of its movie recommendation service.

In doing so, the company is putting out a call to researchers who specialize in machine learning--the type of artificial intelligence used to build systems that recommend music, books, and movies. The entrant who can increase the accuracy of the Netflix recommendation system, which is called Cinematch, by 10 percent by 2011 will win the prize.

Recommendation systems such as those used by Netflix, Amazon, and other Web retailers are based on the principle that if two people enjoy the same product, they're likely to have other favorites in common too.

But behind this simple premise is a complex algorithm that incorporates millions of user ratings, tens of thousands of items, and ever-changing relationships between user preferences.

To deal with this complexity, algorithms for recommendation systems are "trained" on huge datasets. One dataset used in Netflix's system contains the star ratings--one to five--that Netflix customers assign to movies. Using this initial information, good algorithms are able to predict future ratings, and therefore can suggest other films that an individual might like.

Because access to such a dataset is critical to improving the quality of its recommendation systems, the company also released 100 million recommendations--stripped of any personal identifying information--according to Jim Bennett, vice president of recommendations systems at Netflix.

We spoke with Bennett this week about how recommendation systems work--and the challenges of building a better one.
**Technology Review:** Before building a better recommendation system, it would be useful to understand your current approach. How does Cinematch work?

**Jim Bennett:** First, you collect 100 million user ratings for about 18,000 movies. Take any two movies and find the people who have rated both of them. Then look to see if the people who rate one of the movies highly rate the other one highly, if they liked one and not the other, or if they didn't like either movie. Based on their ratings, Cinematch sees whether there's a correlation between those people. Now, do this for all possible pairs of 65,000 movies.

**TR:** So Cinematch would recommend movies to me based on the evaluations of people who rated movies to the way I did. Does that method work for all movies at Netflix?

**JB:** A lot of the really obscure discs, for instance, the "How to Mow a Lawn" DVDs, don't have very many ratings and this method doesn't work as well. For movies with a large number of ratings, you do substantially well. But to make it work, there needs to be a lot of data-tuning because people can sometimes have interesting rating patterns.

**TR:** Like what?

**JB:** For example, there are many people who rate a movie with only one star or five stars. And there are some people who just rate everything with three stars. What you're looking for is an interesting spread of opinions because you're trying to capture correlations. That's the core of the engine.

**TR:** How do you quantitatively measure the accuracy of your system?

**JB:** We trained Cinematch on 100 million ratings and asked it to predict what the other 3 million would be. We compared ours with the actual answers. We do that every day. We get about 2 million ratings per day and we track the daily fluctuations of the system. We expect to measure submissions to the contest [the same way]. The actual prize dataset is 103 million ratings, but we only released 100 million of them.

**TR:** In order to win the $1 million prize, a new algorithm needs to improve the accuracy of recommendations by 10 percent over Cinematch. You're also rewarding a $50,000 "progress" prize each year for the algorithm that shows the most improvement over the previous year's best algorithm, by at least 1 percent. What will these percentage improvements mean to a Netflix customer?

**JB:** If you go to the website and rate 100 movies for us, the red stars shown under each movie are personalized for you. We use these ratings to adjust the prediction away
from the average recommendation, according to your taste. A three-percent difference, for instance, might make a difference of one-quarter star. We have millions of people rating millions of DVDs, and that quarter-star difference helps us sort the list. The individual movie recommendation might not get so much better, but, overall, the set of recommended movies is very different. Move a battleship a little bit, and it makes a huge difference.

TR: Why are recommendation systems so hard to improve?

JB: One of the reasons is there are no datasets. Many of the machine-learning applications require fairly substantial datasets that easily have millions of data points. There are lots of different approaches to solving the problem, but they all need large datasets. And as with many datasets, once we've applied the techniques to those datasets, there's no place to go.

TR: So you're looking for an algorithm that tackles the problem in a completely different way than Cinematch?

JB: Correct. As far as we know, there are many good ideas out in the field. We just can't test them all. We know that there are people who are really on top of the literature who know the ins and outs of [recommendation systems] and we'd really like to know which ones would be better.

TR: What are some approaches, discussed in the literature, which could work, but haven't been tested with movie recommendations yet?

JB: It's hard to say. There was an article in Science a few months ago [July 28, 2006] that used an interesting combination of two types of neural networks [a computational method that sorts data similar to the human brain]. One neural network supervises the machine learning and the other steers that learning. At Netflix, we look at correlations between ratings, and that's a linear model. Not all knowledge can be represented by a linear combination of features. This particular model in Science uses a nonlinear approach. I think that technique could be quite good.

TR: Are there any other pressing technical challenges at Netflix that might be solved by offering a prize?

JB: I wouldn't want to speculate on more contests. Are there other technical challenges? Absolutely. Beyond the systems challenge of keeping the recommendation engines up and running with an increasing customer base, we also have a huge number of challenges within the company--like trying to ship two millions discs a day to
people. And there are interesting challenges ahead as we get ready for the download world [where people can download movies via the Internet]. The company's filled with tremendous challenges.

Copyright Technology Review 2006.