



Netflix Technology Blog

Follow

Learn more about how Netflix designs, builds, and operates our systems and engineering organizations

Apr 6, 2012 · 9 min read

Netflix Recommendations: Beyond the 5 stars (Part 1)

by Xavier Amatriain and Justin Basilico (*Personalization Science and Engineering*)

In this two-part blog post, we will open the doors of one of the most valued Netflix assets: our recommendation system. In Part 1, we will relate the Netflix Prize to the broader recommendation challenge, outline the external components of our personalized service, and highlight how our task has evolved with the business. In Part 2, we will describe some of the data and models that we use and discuss our approach to algorithmic innovation that combines offline machine learning experimentation with online AB testing. Enjoy... and remember that we are always looking for more star talent to add to our great team, so please take a look at [our jobs page](#).

The Netflix Prize and the Recommendation Problem

In 2006 we announced the Netflix Prize, a machine learning and data mining competition for movie rating prediction. We offered \$1 million to whoever improved the accuracy of our existing system called *Cinematch* by 10%. We conducted this competition to find new ways to improve the recommendations we provide to our members, which is a key part of our business. However, we had to come up with a proxy question that was easier to evaluate and quantify: the *root mean squared error* (RMSE) of the predicted rating. The race was on to beat our RMSE of 0.9525 with the finish line of reducing it to 0.8572 or less.

A year into the competition, the Korbell team won the first Progress Prize with an 8.43% improvement. They reported more than 2000 hours of work in order to come up with the final combination of 107 algorithms that gave them this prize. And, they gave us the source code. We looked at the two underlying algorithms with the best performance in the ensemble: *Matrix Factorization* (which the community generally called SVD, *Singular Value Decomposition*) and *Restricted Boltzmann Machines* (RBM). SVD by itself provided a 0.8914

RMSE, while RBM alone provided a competitive but slightly worse 0.8990 RMSE. A linear blend of these two reduced the error to 0.88. To put these algorithms to use, we had to work to overcome some limitations, for instance that they were built to handle 100 million ratings, instead of the more than 5 billion that we have, and that they were not built to adapt as members added more ratings. But once we overcame those challenges, we put the two algorithms into production, where they are still used as part of our recommendation engine.

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40
4	Opera Solutions and Vandelay United	0.8588	9.84	2009-07-10 01:12:31
5	Vandelay Industries I	0.8591	9.81	2009-07-10 00:32:20
6	PragmaticTheory	0.8594	9.77	2009-06-24 12:06:56
7	BellKor in BigChaos	0.8601	9.70	2009-05-13 08:14:09
8	Dace	0.8612	9.59	2009-07-24 17:18:43

If you followed the Prize competition, you might be wondering what happened with the final Grand Prize ensemble that won the \$1M two years later. This is a truly impressive compilation and culmination of years of work, blending hundreds of predictive models to finally cross the finish line. We evaluated some of the new methods offline but the additional accuracy gains that we measured did not seem to justify the engineering effort needed to bring them into a production environment. Also, our focus on improving Netflix personalization had shifted to the next level by then. In the remainder of this post we will explain how and why it has shifted.

From US DVDs to Global Streaming

One of the reasons our focus in the recommendation algorithms has changed is because Netflix as a whole has changed dramatically in the last few years. Netflix launched an instant streaming service in 2007, one year after the Netflix Prize began. Streaming has not only changed the way our members interact with the service, but also the type of data available to use in our algorithms. For DVDs our goal is to help people fill their queue with titles to receive in the mail over the coming days and weeks; selection is distant in time from viewing, people select

carefully because exchanging a DVD for another takes more than a day, and we get no feedback during viewing. For streaming members are looking for something great to watch right now; they can sample a few videos before settling on one, they can consume several in one session, and we can observe viewing statistics such as whether a video was watched fully or only partially.

Another big change was the move from a single website into hundreds of devices. The integration with the [Roku player](#) and the [Xbox](#) were announced in 2008, two years into the Netflix competition. Just a year later, Netflix streaming made it into the [iPhone](#). Now it is available on a multitude of devices that go from a myriad of [Android devices](#) to the latest [AppleTV](#).

Two years ago, we went international with the [launch in Canada](#). In 2011, we added [43 Latin-American countries](#) and territories to the list. And just recently, we launched in [UK and Ireland](#). Today, Netflix has more than 23 million subscribers in 47 countries. Those subscribers streamed 2 billion hours from hundreds of different devices in the last quarter of 2011. Every day they add 2 million movies and TV shows to the queue and generate 4 million ratings.

We have adapted our personalization algorithms to this new scenario in such a way that now 75% of what people watch is from some sort of recommendation. We reached this point by continuously optimizing the member experience and have measured significant gains in member satisfaction whenever we improved the personalization for our members. Let us now walk you through some of the techniques and approaches that we use to produce these recommendations.

Everything is a Recommendation

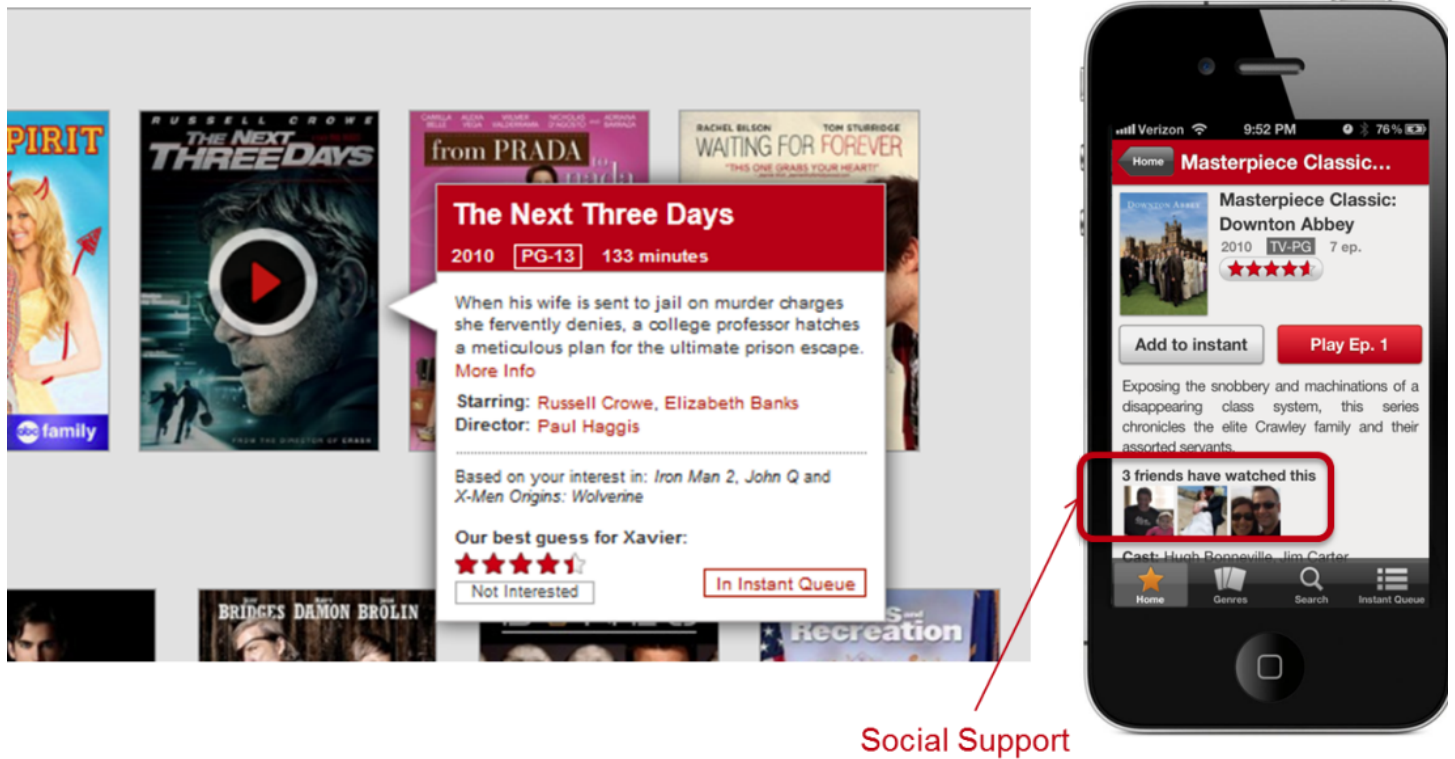
We have discovered through the years that there is tremendous value to our subscribers in incorporating recommendations to personalize as much of Netflix as possible. Personalization starts on our homepage, which consists of groups of videos arranged in horizontal rows. Each row has a title that conveys the intended meaningful connection between the videos in that group. Most of our personalization is based on the way we select rows, how we determine what items to include in them, and in what order to place those items.

Take as a first example the Top 10 row: this is our best guess at the ten titles you are most likely to enjoy. Of course, when we say “you”, we really mean everyone in your **household**. It is important to keep in

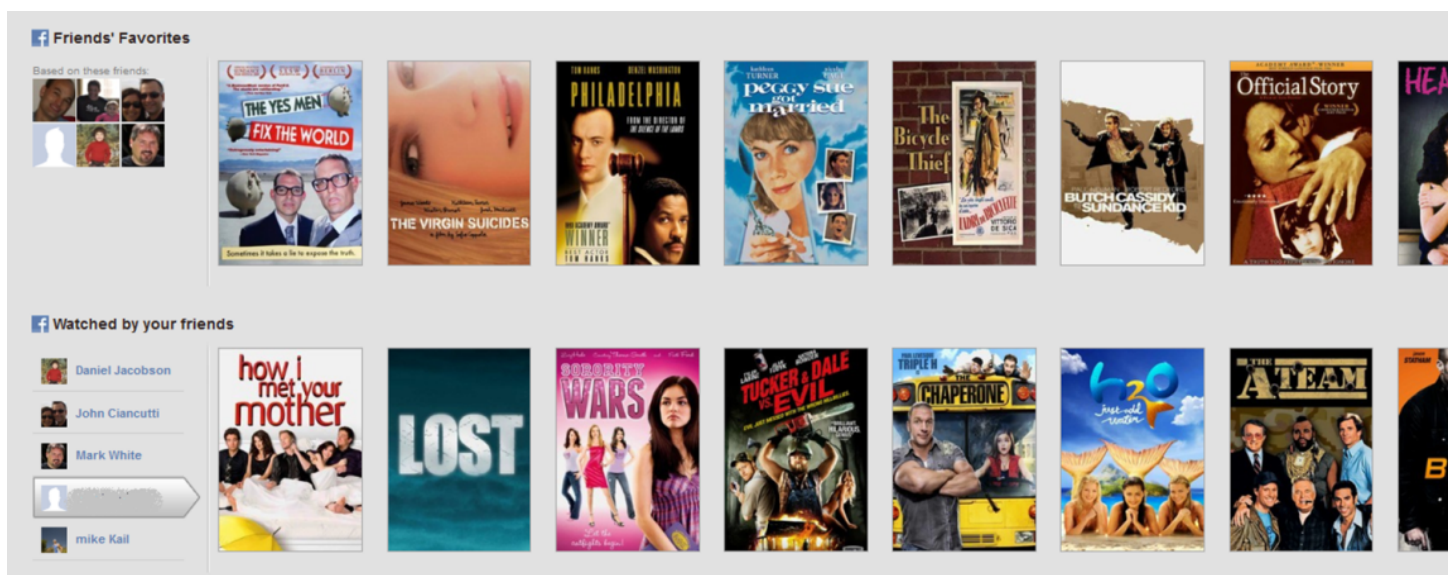
mind that Netflix' personalization is intended to handle a household that is likely to have different people with different tastes. That is why when you see your Top10, you are likely to discover items for dad, mom, the kids, or the whole family. Even for a single person household we want to appeal to your range of interests and moods. To achieve this, in many parts of our system we are not only optimizing for accuracy, but also for **diversity**.



Another important element in Netflix' personalization is **awareness**. We want members to be aware of how we are adapting to their tastes. This not only promotes trust in the system, but encourages members to give feedback that will result in better recommendations. A different way of promoting trust with the personalization component is to provide **explanations** as to why we decide to recommend a given movie or show. We are not recommending it because it suits our business needs, but because it matches the information we have from you: your explicit taste preferences and ratings, your viewing history, or even your friends' recommendations.



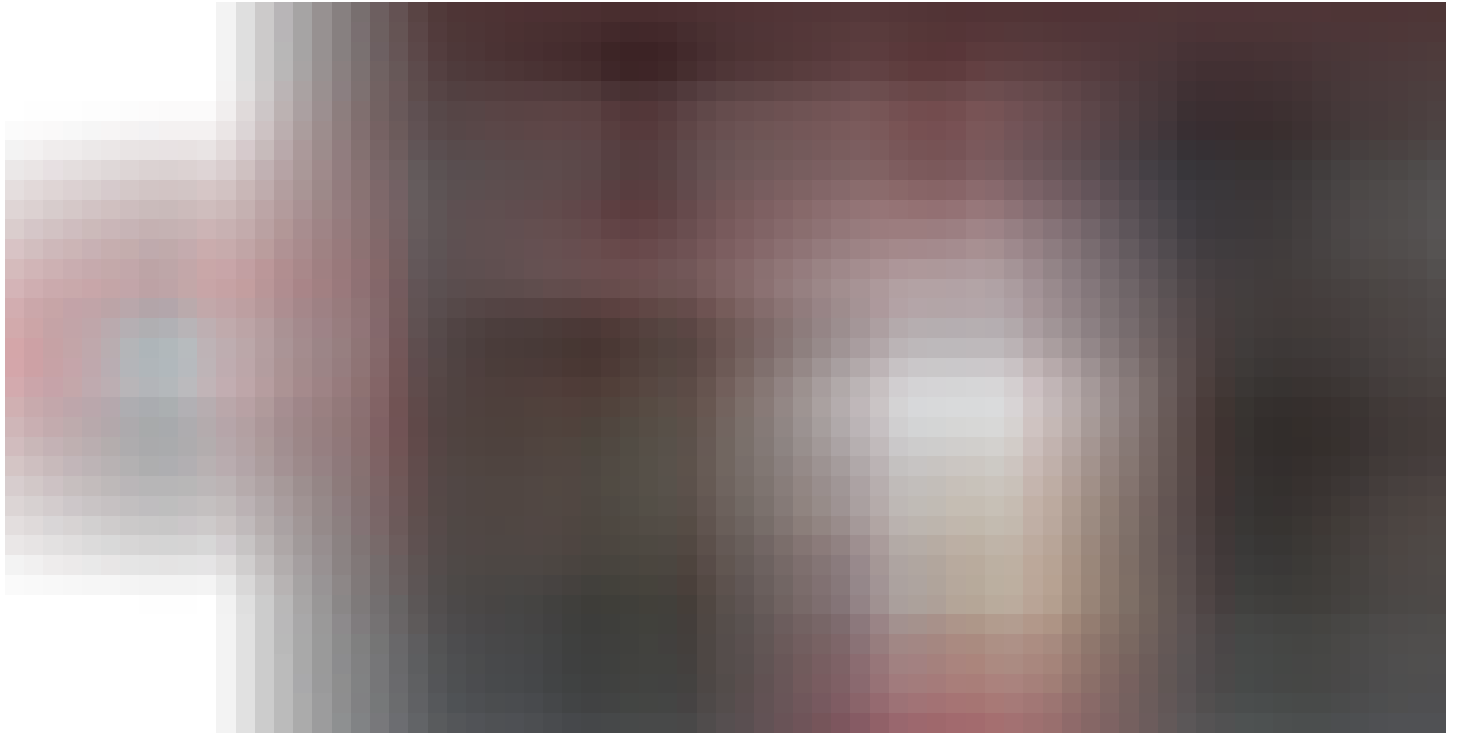
On the topic of friends, we recently released our Facebook connect feature in 46 out of the 47 countries we operate—all but the US because of concerns with the VPPA law. Knowing about your friends not only gives us another signal to use in our personalization algorithms, but it also allows for different rows that rely mostly on your **social** circle to generate recommendations.



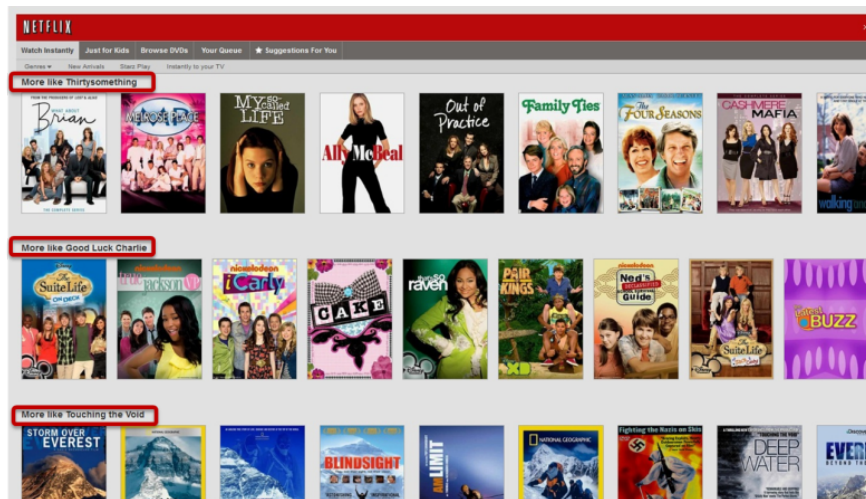
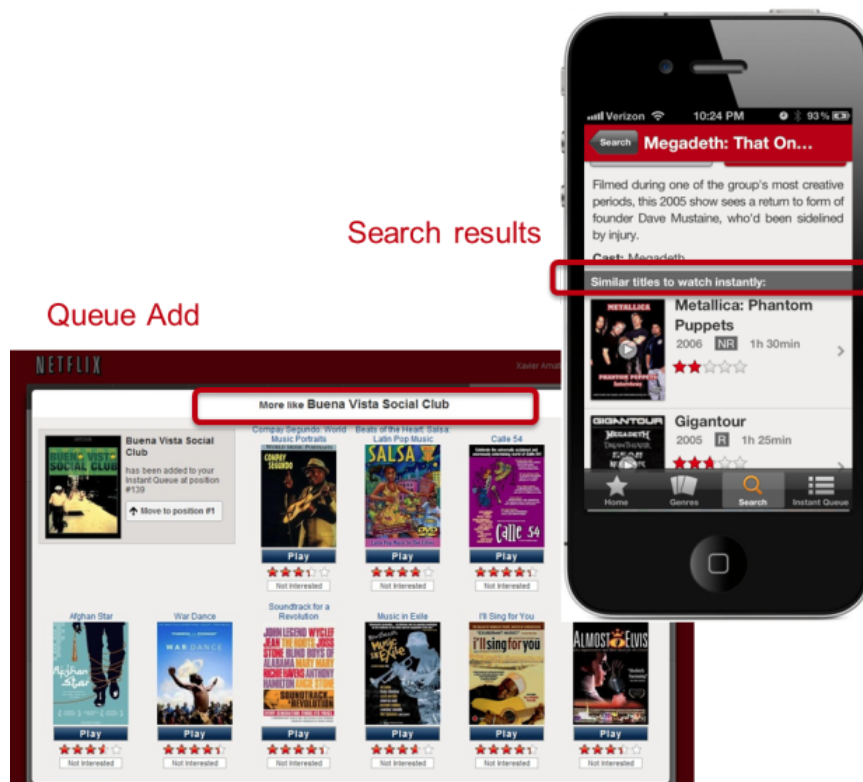
Some of the most recognizable personalization in our service is the collection of “**genre**” rows. These range from familiar high-level categories like “Comedies” and “Dramas” to highly tailored slices such as “Imaginative Time Travel Movies from the 1980s”. Each row represents 3 layers of personalization: the choice of genre itself, the subset of titles selected within that genre, and the ranking of those titles. Members connect with these rows so well that we measure an increase in member retention by placing the most tailored rows higher on the page instead of lower. As with other personalization elements, **freshness** and diversity is taken into account when deciding what genres to show from the thousands possible.



We present an explanation for the choice of rows using a member’s implicit genre preferences—recent plays, ratings, and other interactions—, or explicit feedback provided through our taste preferences survey. We will also invite members to focus a row with additional explicit preference feedback when this is lacking.



Similarity is also an important source of personalization in our service. We think of similarity in a very broad sense; it can be between movies or between members, and can be in multiple dimensions such as metadata, ratings, or viewing data. Furthermore, these similarities can be blended and used as features in other models. Similarity is used in multiple contexts, for example in response to a member's action such as searching or adding a title to the queue. It is also used to generate rows of "adhoc genres" based on similarity to titles that a member has interacted with recently. If you are interested in a more in-depth description of the architecture of the similarity system, you can read about it in [this past post](#) on the blog.



In most of the previous contexts—be it in the Top10 row, the genres, or the similars—**ranking**, the choice of what order to place the items in a row, is critical in providing an effective personalized experience. The goal of our ranking system is to find the best possible ordering of a set of items for a member, within a specific context, in real-time. We decompose ranking into scoring, sorting, and filtering sets of movies for presentation to a member. Our business objective is to maximize member satisfaction and month-to-month subscription retention, which correlates well with maximizing consumption of video content. We therefore optimize our algorithms to give the highest scores to titles that a member is most likely to play and enjoy.

Now it is clear that the Netflix Prize objective, accurate prediction of a movie’s rating, is just one of the many components of an effective recommendation system that optimizes our members enjoyment. We also need to take into account factors such as context, title popularity, interest, evidence, novelty, diversity, and freshness. Supporting all the different contexts in which we want to make recommendations requires a range of algorithms that are tuned to the needs of those contexts. In the next part of this post, we will talk in more detail about the ranking problem. We will also dive into the data and models that make all the above possible and discuss our approach to innovating in this space.

On to part 2:

Netflix Recommendations: Beyond the 5 stars (Part 2)

More insight into our broader personalization technology.

medium.com

See Also:

Netflix Prize: Home

The Netflix Prize sought to substantially improve the accuracy of predictions about how much...

www.netflixprize.com

Testing Netflix on Android

When Netflix decided to enter the Android ecosystem, we faced a daunting set of challenges.

medium.com

"More Like This..." — Building a network of similarity

Ever wondered what makes "Inception" similar to "12 Monkeys"?



Originally published at techblog.netflix.com on April 6, 2012.

