Aaron Drexler

2/27/22

DSC 630

**Netflix Case Study**

“We’re quite curious, really. To the tune of one million dollars.” This is how Netflix introduced the [initiative of a contest](https://www.netflixprize.com/rules.html) aimed at improving its movie recommendations back in 2006. Netflix was already a household name at the time. But instead of video streaming, which didn’t exist yet, the company was in the business of [DVD rental by mail](https://en.wikipedia.org/wiki/Netflix), with the website serving as its online library. Cinematch, the site’s recommendation system, was “doing pretty well”, Netflix reckoned, but could surely be made even better. To this end, Netflix offered a Grand Prize of one million dollars to anyone who could come up with an algorithm that beat the accuracy of Cinematchrecommendations by a margin of ten percent. Accuracy referred to how closely predicted ratings of movies would match subsequent actual ratings.

The movie rental company Netflix launched this competition to improve the movie recommendations it provides to customers. The company challenged the world by requiring that the winner improve upon Netflix’s own established recommendation capabilities by 10 percent.

PA contests such as the Netflix Prize leverage competitive spirit to garner scientific advancement. Like a horse race, a competition levels the playing field and unambiguously singles out the best entrant. With few limitations, almost anyone in the world—old, young, tall, or short—can participate by downloading the data, forming a predictive model, and submitting. In 2008, Martin Chabbert and Martin Piotte launched a mission to win the $1 million Netflix Prize, the most high-profile analytical competition of its time.

**Business Understanding: Defining the Problem**

Netflix is a prime example of PA in action, as a reported 70 percent of Netflix movie choices arise from its online recommendations. Product recommendations are increasingly important for the retail industry in general. More than a sales ploy, these tailored recommendations provide relevancy and personalization that customers actively seek.

The $1 million Netflix Prize attracted a white-hot spotlight and built a new appreciation for the influence crowdsourcing has to rally an international wealth of bright minds. In total, 5,169 teams formed to compete in this contest, submitting 44,014 entries by the end of the event.

**Data Understanding**

To ensure submissions are objectively compared, prediction competitions employ a clever trick: The competitor must submit not a predictive model, but its predictive scores, as generated for an evaluation data set within which the correct answers—the target values that the model is meant to infer—are withheld. Netflix Prize models predict how a customer would rate a movie (based on how he or she has rated other movies). The true ratings are suppressed in the publicly posted evaluation data, so submitters can’t know exactly which examples they’re getting right and which they’re getting wrong at the time of submission. All said, to launch the competition, Netflix released to the public over 100 million ratings from some 480,189 customers (anonymized for privacy considerations, with names suppressed).

PA APPLICATION: MOVIE RECOMMENDATIONS  
1. What’s predicted: What rating a customer would give to a movie.  
2. What’s done about it: Customers are recommended movies that they are predicted to rate highly.

Here is the top portion of the final leaderboard:

**Table

Description automatically generated**

**Data Preparation**

Here’s where the power to advance PA begins. As the contest wore on, various teams started merging and combining their algorithms in a concerted effort to reach the elusive ten-percent threshold. Combining two or more sophisticated predictive models is simple: Just apply predictive modeling to learn how to combine them together. Since each model comes about from machine learning, this is an act of “learning on top of learning”—meta-learning.

A picture containing indoor

Description automatically generated

**Modeling/ Deployment**

Two teams, BellKor (from AT&T Research) and BigChaos (a strikingly young-looking team from a small start-up in Austria), formed an alliance. They joined forces and blended predictive models to form an über-team.

Competitors-turned-collaborators, with two distinct, intricate models that have been developed in very different ways, combine each of their models. This merging also combined with Martin Chabbert and Martin Piotte’s model to create a new super team. The Martin’s team was named Pragmatic Theory, and when merged with the uber team combination of BellKor and BigChaos, the layout of their model merge was as follows:

Diagram

Description automatically generated

This was an example of something called the ensemble effect. When joined in an ensemble, predictive models compensate for one another’s limitations, so the ensemble as a whole is more likely to predict correctly than its component models are.

**Summary and Conclusions**

Ensembles remain robust even as they become increasingly complex. They seem to be immune to this limitation, as if soaked in a magic potion against overlearning. John Elder, who humorously calls ensemble models a “secret weapon,” identified this dilemma in a research paper and dubbed it “the generalization paradox of ensembles.” Enter The Ensemble Effect. By simply joining models together, we enjoy the benefit of cranking up our model’s structural complexity while retaining a critical ingredient: robustness against overlearning.

**Citations**

Siegel, Eric. (2016). *Predictive analytics: the power to predict who will click, buy, lie, or die*. New Jersey: John Wiley & Sons.

**https://www.yusp.com/blog-posts/netflix-prize-story-personalization/**