

Lessons on Innovation and Collaboration from the Netflix Competition

James P. White <jpwhite@uci.edu>

Abstract

The Netflix Prize Challenge has received a great deal of attention since its launch in 2006. With its conclusion (and the launch of a second contest) there is an even greater flurry of coverage and interest[7] [2] [9] [12]. Besides the technical advances in collaborative filtering and recommender systems, the process of the contest and how the competitors managed their development efforts (both successful and unsuccessful) are worth studying for those attacking other difficult technical challenges.

Competition vs. Collaboration

The Netflix Competition starkly highlights the conflicts between competition and collaboration, and how a sufficient incentive and motivating challenge are needed to overcome them.

When the contest started, a large number of people began working on the problem (there were eventually 4000 entrants[11]). For those that were able to create a solution that was an improvement over Cinematch, they also discovered that they were not going to achieve the 10% improvement required to win without combining their efforts with others (or at least not in a timely manner). So they're driven by an unavoidable game theoretic value proposition to deal with the process of joining with others. But of course those need to be others that they're able and willing to collaborate with, and, most importantly, have or will have solutions that will improve their results.

The way teams came together, especially late in the contest, and the improved results that were achieved suggest that this kind of Internet-enabled approach, known as crowdsourcing, can be applied to complex scientific and business challenges.

That certainly seemed to be a principal lesson for the winners. The blending of different statistical and machine-learning techniques “only works well if you combine models that approach the problem differently,” said Chris Volinsky, a scientist at AT&T Research and a leader of the BellKor team. “That’s why collaboration has been so effective, because different people approach problems differently.”

- Steve Lohr, New York Times[8]

The winning team – BellKor Pragmatic Chaos – has seven members, and the runners-up (who had the same result but submitted their entry 20 minutes after BPC) has thirty members. Perhaps thirty is too many considering that twenty minutes cost them one million dollars and communication costs is one of the big problems with teams[4].

In the second edition of the Netflix Contest, this imperative to collaborate is reinforced. By just using progress payments to the current leader at specific times (six months and eighteen months)[5], the incumbent has a lot of power to attract and select additional members, while individuals are at a big disadvantage. The only real option for individuals and small teams being to decide which collaboration to join. Although it is possible to imagine that a particularly desirable competitor could be sought (and even bid) after by the collaborations – research as a competitive team sport (or market for talent if you

prefer).

Issues vs. Scale of Association

	Individuals	Teams	Alliances	Community	Public
Openness					
Decision-making					
Motivation					
Compensation					

Closed vs. Open

One of the difficult issues in moving up the scale of association is information disclosure. When dealing with proprietary information, negotiating each move is greatly complicated by the requirements of non-disclosure. Naturally that is a dilemma since disclosure of some degree is required in order for the move to take place. A great many potentially beneficial collaborations never happen because of this problem.

By requiring that the winner's solution to be made public, Netflix lessened this conflict significantly. Because the winner would have their solution publicized, competitor's were already committed to that outcome, and losers had nothing to lose, unless they believed they could benefit from their solution in some other way as a "Plan B".

The conflict remains though for many during the contest, and thus is still a deterrent to collaboration at that time (perhaps when collaboration would be most beneficial overall).

Cartesian Thinking At Work

The intensive work by mathematical and informatics specialists required to solve the contest's challenge (and the challenge itself) is a classic application of the Cartesian epistemology of computers[3] and of scientific culture¹. It is worthwhile to note though, that success in the challenge required some clearly outside-of-the-box thinking. BPC's addition of considering not just users' ratings, but also just which ones they rated regardless of rating, was key to winning the first progress prize[1]. The final method added to the winning entry to get it over the 10% threshold was to recognize and account for the fact that how users rate movies changes over time[6].

"Recommender Systems" vs. "Collaborative Filtering"

The shift away from the user's perspective and into the domain of scientific speciality (per the process of Cartesian Thinking) is visible in the promotion of the term "Recommender Systems" in favor of "Collaborative Filtering". It is noteworthy that the promoter of the change in terminology is the specialists' organization, the ACM, which has a conference called the "ACM Recommender Systems Conference"². The winners themselves though seem to be ambivalent, using the term "collaborative

1 Bowers uses distinctly Cartesian thinking though in the identification of particular personality types ("mindsets") with entire cultures. While it may well be said that cultures have personalities too (and they do, however politically incorrect that opinion may be), Bowers makes his cases with the thinking/writing of particular individuals (as in choosing the term "Cartesian" rather than "Age of Reason" or "Industrial Revolution").

2 <http://recsys.acm.org/>

filter” fairly interchangeably with “recommender system”.

A Repeatable Model?

At one level the Netflix competition is already a repeat because it is an application of the Grand Challenge model. But there are unique aspects that would be important for those who want to get the same kind of results.

Perhaps the most unique aspect of this competition is the role that publicizing proprietary data (Netflix’s users’ ratings database) played. By enabling academics and amateurs to freely engage in a real-world problem at a real-world scale, an entire field of research has been advanced and made more useful.

Publicizing the winning solution also plays a big role. Not only does it attract academics, who publish their work anyhow, but it also facilitates (or even compels) collaboration by parties who would otherwise remain competitors. An aid to collaboration could be to create an escrow agency as a trusted third party to process disclosures (including, perhaps, verifying results and other such sticky wickets) and handle contracts during the competition.

There have been calls for Google and Amazon to conduct their own competitions[10]. Certainly they have the proprietary data and applications of the same kind as Netflix uses. An interesting point is that Google has publicized some of their data, such as the trillion word N-gram database³. An interesting analysis would be to see what the results of that release (which was also made in 2006) has been vs. the Netflix competition. Certainly it hasn’t gained nearly as much publicity for Google as Netflix has received. It is easy to imagine that if Google had attached a million dollars and a challenge question to that release it certainly would have gotten a lot ink and more research effort. But one can also imagine that Google may be selective in how much more publicity they want to get.

Another source of these Grand Challenges is DARPA and other government agencies[13]. Public interest groups could play too by conducting a challenge around public data (perhaps after liberating it via FOIA)⁴. For companies and groups that don’t have the technorati cachet of a Netflix or Amazon (Google being by itself in another class above), it is less obvious that this approach would work as well. But with the right problem, data set, prize, and willingness to publicize the results, it is certainly possible.

A Sustainable Model?

Some aspects of the Netflix Competition are not sustainable. The level of effort expended in pursuit of the one million dollar prize was extremely high. If these competition become more common, not only will the novelty factor be gone (and fatigue set in), but there would be competition amongst the contests for the competitors’ time and attention. But the dollar level can go quite low by American standards given the global nature of the competition, as seen with TopCoders and InnoCentive.

On the other hand, the Netflix Contest had a distinctly academic flavor from the start, and many entrants were not (or claimed not to be) much influenced by the prize money.

“Having an academic flavor to the competition has really helped it to sustain energy for two-and-a-half years.”

3 <http://googleresearch.blogspot.com/2006/08/all-our-n-gram-are-belong-to-you.html>

4 Humanitarian Open Source efforts are something of a role model here too. “Trinity’s Humanitarian Open Source Software Project Snags Major Grant”, <http://www.tmcnet.com/usubmit/2009/10/05/4407179.htm>

- Chris Volinsky, executive director of statistics research at AT&T Research and a member of BellKor's Pragmatic Chaos[11]

“We've already had a \$10 million payoff internally from what we've learned... So for us, the \$1 million prize was secondary, almost trivial.”

- Arnab Gupta, chief executive of Opera Solutions (employer of key members of The Ensemble)[8]

A further question on compensation is whether it is possible to have a prize model that is more equitable in terms of rewarding good efforts but gets results that are as good overall (or even better?). For example, what if the prize wasn't fixed but instead was larger for larger teams? Obviously there is some point of diminishing returns, but it seems quite likely to me that the winner-takes-all model is not optimal. The Ensemble (runner-up) showed⁵ that a simple merge of their solution with the winner's is 0.13% better than either of them alone. That is a whopping improvement and makes it pretty clear that if those two teams had allied then they probably could have won the contest a year ago (my quick guess without checking the contest progress numbers – but they were stuck looking for less than 1% for a long while).

Bibliography

1. Bell, R. and Koren, Y. Lessons from the Netflix prize challenge. *ACM SIGKDD Explorations Newsletter*, (2007), 75-79.
2. Bell, R.M., Bennett, J., Koren, Y., and Volinsky, C. The Million Dollar Programming Prize. *IEEE Spectrum*, 28-33. <http://www.spectrum.ieee.org/computing/software/the-million-dollar-programming-prize>.
3. Bowers, C.A. How Computers Contribute to the Ecological Crisis. *The Trumpeteer* 8, 1 (1991), 13-15.
4. Brooks, F. *The mythical man-month : essays on software engineering*. Addison-Wesley Pub. Co., Reading Mass., 1995.
5. Hunt, N. Netflix Prize: Forum / Netflix Prize 2 (Yes, a sequel!). *Netflix Prize Forum*, 2009. <http://www.netflixprize.com/community/viewtopic.php?id=1520>.
6. Koren, Y. Collaborative filtering with temporal dynamics. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM (2009), 447-456.
7. Linden, G. The Netflix prize, computer science outreach, and Japanese mobile phones. *Communications of the ACM* 52, 10 (2009), 8.
8. Lohr, S. A \$1 Million Research Bargain for Netflix, and Maybe a Model for Others. *New York Times*. <http://www.nytimes.com/2009/09/22/technology/internet/22netflix.html?pagewanted=all>.
9. Lohr, S. Netflix Awards \$1 Million Prize and Starts a New Contest. *Bits Blog*, 2009. <http://bits.blogs.nytimes.com/2009/09/21/netflix-awards-1-million-prize-and-starts-a-new-contest/?hp>.
10. Manjoo, F. The Netflix Prize was brilliant. Google and Microsoft should steal the idea. *Slate Magazine*, 2009. <http://www.slate.com/id/2229225/pagenum/all/>.
11. Monroe, D. Just for you. *Communications of the ACM* 52, 8 (2009), 15-17.
12. Naone, E. What's Next for the Netflix Algorithms? *Technology Review*, 2009. <http://www.technologyreview.com/computing/23635/?a=f>.
13. Rampell, C. Should the government start handing out prizes for science breakthroughs? *Slate Magazine*, 2008. <http://www.slate.com/id/2182663/pagenum/all/#p2>.

5 <http://www.the-ensemble.com/content/netflix-prize-conclusion>