WIKIPEDIA

Netflix Prize

The **Netflix Prize** was an open competition for the best collaborative filtering algorithm to predict user ratings for films, based on previous ratings without any other information about the users or films, i.e. without the users or the films being identified except by numbers assigned for the contest.

The competition was held by Netflix, an online DVD-rental and video streaming service, and was open to anyone who is neither connected with Netflix (current and former employees, agents, close relatives of Netflix employees, etc.) nor a resident of certain blocked countries (such as Cuba or North Korea).[1] On September 21, 2009, the grand prize of US$1,000,000 was given to the BellKor's Pragmatic Chaos team which bested Netflix's own algorithm for predicting ratings by 10.06%.[2]

Contents

Problem and data sets

Netflix provided a *training* data set of 100,480,507 ratings that 480,189 users gave to 17,770 movies. Each training rating is a quadruplet of the form `<user, movie, date of grade, grade>`. The user and movie fields are integer IDs, while grades are from 1 to 5 (integral) stars.[3]

The *qualifying* data set contains over 2,817,131 triplets of the form `<user, movie, date of grade>`, with grades known only to the jury. A participating team's algorithm must predict grades on the entire qualifying set, but they are only informed of the score for half of the data, the *quiz* set of 1,408,342 ratings. The other half is the *test* set of 1,408,789, and performance on this is used by the jury to determine potential prize winners. Only the judges know which ratings are in the quiz set, and which are in the test set —this arrangement is intended to make it difficult to hill climb on the test set. Submitted predictions are scored against the true grades in terms of root mean squared error (RMSE), and the goal is to reduce this error as much as possible. Note that while the actual grades are integers in the range 1 to 5, submitted predictions need not be. Netflix also identified a *probe* subset of 1,408,395 ratings within the *training* data set. The *probe*, *quiz*, and *test* data sets were chosen to have similar statistical properties.

In summary, the data used in the Netflix Prize looks as follows:

- Training set (99,072,112 ratings not including the probe set, 100,480,507 including the probe set)

  - Probe set (1,408,395 ratings)
- Qualifying set (2,817,131 ratings) consisting of:

  - Test set (1,408,789 ratings), used to determine winners

- Quiz set (1,408,342 ratings), used to calculate leaderboard scores

For each movie, title and year of release are provided in a separate dataset. No information at all is provided about users. In order to protect the privacy of customers, "some of the rating data for some customers in the training and qualifying sets have been deliberately perturbed in one or more of the following ways: deleting ratings; inserting alternative ratings and dates; and modifying rating dates".[2]

The training set is such that the average user rated over 200 movies, and the average movie was rated by over 5000 users. But there is wide variance in the data—some movies in the training set have as few as 3 ratings,[4] while one user rated over 17,000 movies.[5]

There was some controversy as to the choice of RMSE as the defining metric. Would a reduction of the RMSE by 10% really benefit the users? It has been claimed that even as small an improvement as 1% RMSE results in a significant difference in the ranking of the "top-10" most recommended movies for a user.[6]

Prizes

Prizes were based on improvement over Netflix's own algorithm, called *Cinematch*, or the previous year's score if a team has made improvement beyond a certain threshold. A trivial algorithm that predicts for each movie in the quiz set its average grade from the training data produces an RMSE of 1.0540. Cinematch uses "straightforward statistical linear models with a lot of data conditioning".[7]

Using only the training data, Cinematch scores an RMSE of 0.9514 on the quiz data, roughly a 10% improvement over the trivial algorithm. Cinematch has a similar performance on the test set, 0.9525. In order to win the grand prize of $1,000,000, a participating team had to improve this by another 10%, to achieve 0.8572 on the test set.[2] Such an improvement on the quiz set corresponds to an RMSE of 0.8563.

As long as no team won the grand prize, a *progress* prize of $50,000 was awarded every year for the best result thus far. However, in order to win this prize, an algorithm had to improve the RMSE on the quiz set by at least 1% over the previous progress prize winner (or over Cinematch, the first year). If no submission succeeded, the progress prize was not to be awarded for that year.

To win a progress or grand prize a participant had to provide source code and a description of the algorithm to the jury within one week after being contacted by them. Following verification the winner also had to provide a non-exclusive license to Netflix. Netflix would publish only the description, not the source code, of the system. (To keep their algorithm and source code secret, a team could choose not to claim a prize.) The jury also kept their predictions secret from other participants. A team could send as many attempts to predict grades as they wish. Originally submissions were limited to once a week, but the interval was quickly modified to once a day. A team's best submission so far counted as their current submission.

Once one of the teams succeeded to improve the RMSE by 10% or more, the jury would issue a *last call*, giving all teams 30 days to send their submissions. Only then, the team with best submission was asked for the algorithm description, source code, and non-exclusive license, and, after successful verification; declared a grand prize winner.

The contest would last until the grand prize winner was declared. Had no one received the grand prize, it would have lasted for at least five years (until October 2, 2011). After that date, the contest could have been terminated at any time at Netflix's sole discretion.

Progress over the years

The competition began on October 2, 2006. By October 8, a team called WXYZConsulting had already beaten Cinematch's results.[8]

By October 15, there were three teams who had beaten Cinematch, one of them by 1.06%, enough to qualify for the annual progress prize.[9] By June 2007 over 20,000 teams had registered for the competition from over 150 countries. 2,000 teams had submitted over 13,000 prediction sets.[3]

Over the first year of the competition, a handful of front-runners traded first place. The more prominent ones were:[10]

- WXYZConsulting, a team of Wei Xu and Yi Zhang. (A front runner during November–December 2006.)
- ML@UToronto A, a team from the University of Toronto led by Prof. Geoffrey Hinton. (A front runner during parts of October–December 2006.)
- Gravity, a team of four scientists from the Budapest University of Technology (A front runner during January–May 2007.)
- BellKor, a group of scientists from AT&T Labs. (A front runner since May 2007.)

On August 12, 2007, many contestants gathered at the KDD Cup and Workshop 2007, held at San Jose, California.[11] During the workshop all four of the top teams on the leaderboard at that time presented their techniques. The team from IBM Research — Yan Liu, Saharon Rosset, Claudia Perlich, and Zhenzhen Kou — won the third place in Task 1 and first place in Task 2.

Over the second year of the competition, only three teams reached the leading position:

- BellKor, a group of scientists from AT&T Labs. (front runner during May 2007 - September 2008.)
- BigChaos, a team of Austrian scientists from commendo research & consulting (single team front runner since October 2008)
- BellKor in BigChaos, a joint team of the two leading single teams (A front runner since September 2008)

2007 Progress Prize

On September 2, 2007, the competition entered the "last call" period for the 2007 Progress Prize. Over 40,000 teams from 186 different countries had entered the contest. They had thirty days to tender submissions for consideration. At the beginning of this period the leading team was BellKor, with an RMSE of 0.8728 (8.26% improvement). followed by Dinosaur Planet (RMSE = 0.8769; 7.83% improvement), and Gravity (RMSE = 0.8785; 7.66% improvement). In the last hour of the last call period, an entry by "KorBell" took first place. This turned out to be an alternate name for Team BellKor.

On November 13, 2007, team KorBell (formerly BellKor) was declared the winner of the $50,000 Progress Prize with an RMSE of 0.8712 (8.43% improvement).[12] The team consisted of three researchers from AT&T Labs, Yehuda Koren, Robert Bell, and Chris Volinsky.[13] As required, they published a description of their algorithm.[14]

2008 Progress Prize

The 2008 Progress Prize was awarded to the team BellKor. Their submission combined with a different team, BigChaos achieved an RMSE of 0.8616 with 207 predictor sets.[15] The joint-team consisted of two researchers from commendo research & consulting GmbH, Andreas Töscher and Michael Jahrer (originally team BigChaos) and three researchers from AT&T Labs, Yehuda Koren, Robert Bell, and Chris Volinsky (originally team BellKor).[16] As required, they published a description of their algorithm.[17][18]

This was the final Progress Prize because obtaining the required 1% improvement over the 2008 Progress Prize would be sufficient to qualify for the Grand Prize. The prize money was donated to the charities chosen by the winners

2009

On June 26, 2009 the team "BellKor's Pragmatic Chaos", a merger of teams "Bellkor in BigChaos" and "Pragmatic Theory", achieved a 10.05% improvement over Cinematch (a Quiz RMSE of 0.8558). The Netflix Prize competition then entered the "last call" period for the Grand Prize. In accord with the Rules, teams had thirty (30) days, until July 26, 2009 18:42:37 UTC, to make submissions that will be considered for this Prize.[19]

On July 25, 2009 the team "The Ensemble", a merger of the teams "Grand Prize Team" and "Opera Solutions and Vandelay United", achieved a 10.09% improvement over Cinematch (a Quiz RMSE of 0.8554).[20][21]

On July 26, 2009, Netflix stopped gathering submissions for the Netflix Prize contest.[22]

The final standing of the Leaderboard at that time showed that two teams met the minimum requirements for the Grand Prize. "The Ensemble" with a 10.10% improvement over Cinematch on the Qualifying set (a Quiz RMSE of 0.8553), and "BellKor's Pragmatic Chaos" with a 10.09% improvement over Cinematch on the Qualifying set (a Quiz RMSE of 0.8554).[23] The Grand Prize winner was to be the one with the better performance on the Test set.

On September 18, 2009, Netflix announced team "BellKor's Pragmatic Chaos" as the prize winner (a Test RMSE of 0.8567), and the prize was awarded to the team in a ceremony on September 21, 2009.[24] "The Ensemble" team had matched BellKor's result, but since BellKor submitted their results 20 minutes earlier, the rules award the prize to BellKor.[21][25]

The joint-team "BellKor's Pragmatic Chaos" consisted of two Austrian researchers from Commendo Research & Consulting GmbH, Andreas Töscher and Michael Jahrer (originally team BigChaos), two researchers from AT&T Labs, Robert Bell, and Chris Volinsky, Yehuda Koren from Yahoo! (originally team BellKor) and two researchers from Pragmatic Theory, Martin Piotte and Martin Chabbert.[26] As required, they published a description of their algorithm.[27]

The team reported to have achieved the "dubious honors" (*sic* Netflix) of the worst RMSEs on the *Quiz* and *Test* data sets from among the 44,014 submissions made by 5,169 teams was "Lanterne Rouge", led by J.M. Linacre, who was also a member of "The Ensemble" team.

Cancelled sequel

On March 12, 2010, Netflix announced that it would not pursue a second Prize competition that it had announced the previous August. The decision was in response to a lawsuit and Federal Trade Commission privacy concerns.[28]

Privacy concerns

Although the data sets were constructed to preserve customer privacy, the Prize has been criticized by privacy advocates. In 2007 two researchers from The University of Texas at Austin were able to identify individual users by matching the data sets with film ratings on the Internet Movie Database.[29][30]

On December 17, 2009, four Netflix users filed a class action lawsuit against Netflix, alleging that Netflix had violated U.S. fair trade laws and the Video Privacy Protection Act by releasing the datasets.[31] There was public debate about privacy for research participants. On March 19, 2010, Netflix reached a settlement with the plaintiffs, after which they voluntarily dismissed the lawsuit.

See also

- Crowdsourcing
- Open innovation
- Innovation competition
- Inducement prize contest

References

1. "The Netflix Prize Rules" (http://www.netflixprize.com//rules). Retrieved 2014-04-01.
2. "The Netflix Prize" (https://web.archive.org/web/20090924184639/http://www.netflixprize.com/community/viewtopic.php?id=1537). Archived from the original (http://www.netflixprize.com//community/viewtopic.php?id=1537) on 2009-09-24. Retrieved 2012-07-09.
3. James Bennett; Stan Lanning (August 12, 2007). "The Netflix Prize" (https://web.archive.org/web/20070927051207/http://www.netflixprize.com/assets/NetflixPrizeKDD_to_appear.pdf) (PDF). *Proceedings of KDD Cup and Workshop 2007*. Archived from the original (http://www.netflixprize.com/assets/NetflixPrizeKDD_to_appear.pdf) (PDF) on September 27, 2007. Retrieved 2007-08-25.

4. Sigmoid Curve (2006-10-08). "Miss Congeniality" (https://www.webcitation.org/65tSoXKuf?url=http://www.netflixprize.com/community/viewtopic.php?id=164). *Netflix Prize Forum*. Archived from the original (http://www.netflixprize.com/community/viewtopic.php?id=164) on 2012-03-03. Retrieved 2007-08-25.

5. "prodigious" (2006-10-06). "A single customer that rated 17,000 movies" (https://www.webcitation.org/65tSp5RPI?url=http://www.netflixprize.com/community/viewtopic.php?id=141). *Netflix Prize Forum*. Archived from the original (http://www.netflixprize.com/community/viewtopic.php?id=141) on 2012-03-03. Retrieved 2007-08-25.

6. Yehuda Koren (2007-12-18). "How useful is a lower RMSE?" (https://www.webcitation.org/65tSpdL9E?url=http://www.netflixprize.com/community/viewtopic.php?id=828). *Netflix Prize Forum*. Archived from the original (http://www.netflixprize.com/community/viewtopic.php?id=828) on 2012-03-03.

7. "Netflix Prize Frequently Asked Questions" (https://web.archive.org/web/20070821194257/http://www.netflixprize.com/faq). Archived from the original (http://www.netflixprize.com/faq) on 2007-08-21. Retrieved 2007-08-21.

8. "Netflix Prize Rankings" (http://www.hackingnetflix.com/2006/10/netflix_prize_r.html). *Hacking NetFlix*. October 9, 2006. Retrieved 2007-08-21.

9. "Netflix Prize (I tried to resist, but...)" (http://jsnell.iki.fi/blog/archive/2006-10-15-netflix-prize.html). *Juho Snellman's Weblog*. October 15, 2006. Retrieved 2007-08-21.

10. "Top contenders for Progress Prize 2007 chat" (http://www.research.att.com/~volinsky/netflix/leaders.gif).

11. "The KDD Cup and Workshop 2007" (http://www.cs.uic.edu/~liub/Netflix-KDD-Cup-2007.html).

12. Prizemaster (2007-11-13). "Netflix Progress Prize 2007 awarded to team KorBell" (https://www.webcitation.org/65tSs5fyw?url=http://www.netflixprize.com/community/viewtopic.php?id=799). *Netflix Prize Forum*. Archived from the original (http://www.netflixprize.com/community/viewtopic.php?id=799) on 2012-03-03.

13. "$50,000 Progress Prize is Awarded on First Anniversary of $1 Million Netflix Prize" (http://www.netflix.com/MediaCenter?id=6422).

14. R. Bell; Y. Koren; C. Volinsky (2007). "The BellKor solution to the Netflix Prize" (http://www.netflixprize.com/assets/ProgressPrize2007_KorBell.pdf) (PDF).

15. Robert Bell; Yehuda Koren; Chris Volinsky (2008-12-10). "The BellKor 2008 Solution to the Netflix Prize" (http://www2.research.att.com/~volinsky/netflix/Bellkor2008.pdf) (PDF). *Netflix Prize Forum*.

16. "Netflix Awards $50,000 Progress Prize in Year Two of Multi-Year, Multi-National Netflix Prize Competition" (http://netflix.mediaroom.com/index.php?s=43&item=295).

17. A. Töscher; M. Jahrer (2008). "The BigChaos solution to the Netflix Prize 2008" (http://www.netflixprize.com/assets/ProgressPrize2008_BigChaos.pdf) (PDF).

18. R. Bell; Y. Koren; C. Volinsky (2008). "The BellKor solution to the Netflix Prize 2008" (http://www.netflixprize.com/assets/ProgressPrize2008_BellKor.pdf) (PDF).

19. "BellKor's Pragmatic Chaos" (http://www.research.att.com/~volinsky/netflix/bpc.html). 2009-06-26.

20. "The Ensemble" (http://www.the-ensemble.com/). 2009-07-25.

21. "Netflix Prize Leaderboard" (http://www.netflixprize.com//leaderboard?showtest=t&limit=20). 2009-07-26.

22. "Contest Closed" (http://www.netflixprize.com/closed). 2009-07-26.

23. "The Netflix Prize Comes To A Buzzer-Beater, Nailbiting Finish" (https://techcrunch.com/2009/07/26/the-netflix-prize-comes-to-a-buzzer-beater-nailbiting-finish/). 2009-07-26.

24. "Grand Prize awarded to team BellKor's Pragmatic Chaos" (https://web.archive.org/web/20120507045732/http://www.netflixprize.com/community/viewtopic.php?id=1537). Netflix Prize Forum. 2009-09-21. Archived from the original (http://www.netflixprize.com/community/viewtopic.php?id=1537) on 2012-05-07.

25. Steve Lohr (2009-09-21). "A $1 Million Research Bargain for Netflix, and Maybe a Model for Others" (https://www.nytimes.com/2009/09/22/technology/internet/22netflix.html). New York Times.

26. "Netflix Awards $1 Million Netflix Prize and Announces Second $1 Million Challenge" (http://netflix.mediaroom.com/index.php?s=43&item=327).

27. Andreas Töscher & Michael Jahrer (2009-09-21). "The BigChaos Solution to the Netflix Grand Prize" (http://www.commendo.at/en-publications-48.html). commendo.

28. "Netflix Prize Update" (http://blog.netflix.com/2010/03/this-is-neil-hunt-chief-product-officer.html). Netflix Prize Forum. 2010-03-12.

29. Narayanan, Arvind; Shmatikov, Vitaly. "How To Break Anonymity of the Netflix Prize Dataset". arXiv:cs/0610105 (https://arxiv.org/abs/cs/0610105)

30. Demerjian, Dave (15 March 2007). "Rise of the Netflix Hackers" (http://archive.wired.com/science/discoveries/news/2007/03/72963). *wired.com*. Wired. Retrieved 13 December 2014.

31. Singel, Ryan. "Netflix Spilled Your Brokeback Mountain Secret, Lawsuit Claims" (https://www.wired.com/threatlevel/2009/12/netflix-privacy-lawsuit/). Retrieved 11 August 2017.

External links

- Official website
- Netflix Prize on RecSysWiki
- Kate Greene (2006-10-06). "The $1 million Netflix challenge". *Technology Review*.
- Robert M. Bell; Jim Bennett; Yehuda Koren & Chris Volinsky (May 2009). "The Million Dollar Programming Prize". *IEEE Spectrum*.
- Robust De-anonymization of Large Sparse Datasets by Arvind Narayanan and Vitaly Shmatikov
- Robert M. Bell, Yehuda Koren and Chris Volinsky (2010), "All together now: A perspective on the NETFLIX PRIZE", *Chance*, **23** (1): 24, doi:10.1007/s00144-010-0005-2
- Andrey Feuerverger; Yu He & Shashi Khatri (2012), "Statistical Significance of the Netflix Challenge", *Statistical Science*, **27** (2): 202–231, arXiv:1207.5649, doi:10.1214/11-STS368
- The Netflix $1 Million Prize - Netflix never used its $1 million algorithm due to engineering costs (2009) - Saint