智能推荐技术--案例分析: Netflix

潘微科

Netflix

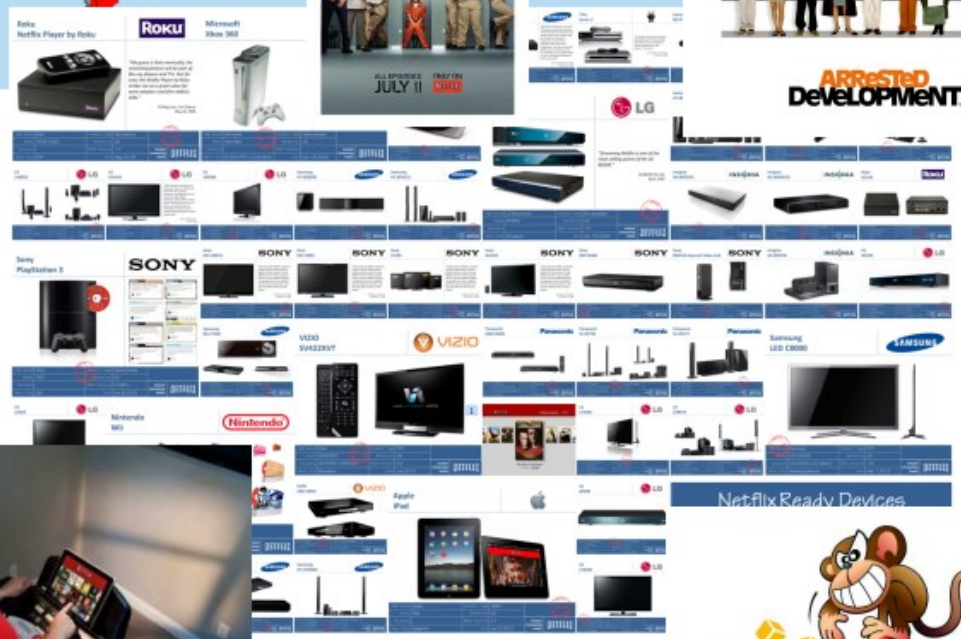
百万美元大赛

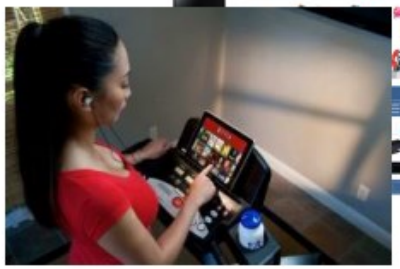
- 2 October **2006** ~ 18 September **2009**

- 10% improvement of root mean square error (RMSE) performance

2006

Netflix





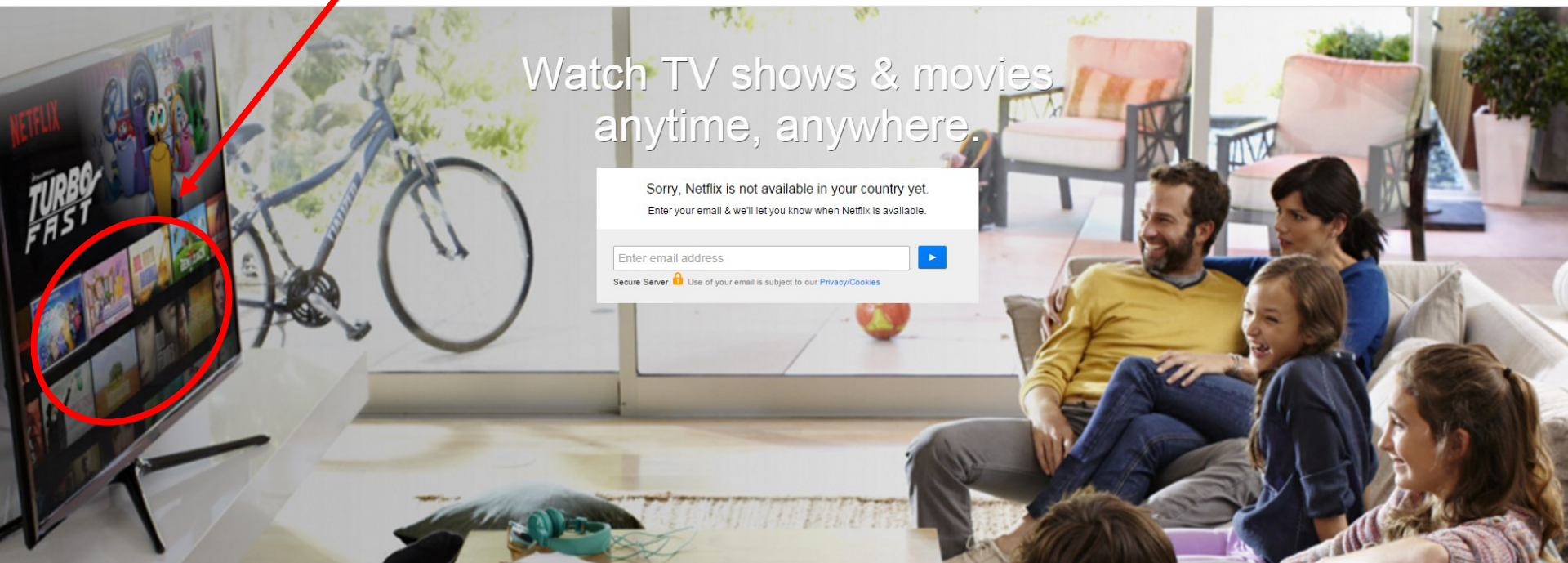
amazon web services™

2014

Netflix

价值

- Help members find content that they'll enjoy to maximize satisfaction and retention
- Netflix's New 'My List' Feature Knows You Better Than You Know Yourself (Because of Algorithms)
- Every impression is a recommendation



Netflix

数据

- Member behavior (play, search, rating)
- Impression
- Time
- Social
- Metadata
- Geo-information
- Device information

Netflix

数据规模

- > 50M members
- > 40 countries
- > 1000 device types
- > 7B hours in Q2 2014
- Plays: > 70M/day
- Searches: > 4M/day
- Ratings: > 6M/day
- Log 100B events/day
- 31.62% of peak US downstream traffic

Netflix

算法(1/2)

- SVD and other matrix factorizations (MF), restricted Boltzmann machines (RBM), factorization machines (FM)
- Learning to ranking (pointwise)/classification/regression: linear/ordinal regression, Logistic regression (LR), Elastic nets, support vector machine (SVM), Bayesian networks, decision tree (DT), gradient boosted decision trees (GBDT), random forests (RF), naïve Bayes
- Learning to ranking (pairwise): RankSVM, RankBoost, RankNet, FRank, …
- Learning to ranking (listwise): RankCosine, ListNet; genetic programming or simulated annealing, LambdaMart, SVM-MAP, AdaRank, …

Netflix

算法(2/2)

- Markov models and graph algorithms (e.g., topic sensitive PageRank)
- Clustering: k-means, affinity propagation (AP), spectral clustering, latent Dirichlet allocation (LDA), Chinese restaurant processes (CRP), hierarchical Dirichlet process (HDP), locality-sensitive hashing (LSH)
- Association rule mining (ARM)
- Gaussian processes (GP)
- Deep artificial neural networks (Deep ANN)

Netflix

评估

- Consumption
  - Accuracy
  - Novelty
  - Diversity
  - Freshness
  - Scalability
  - ...

Netflix

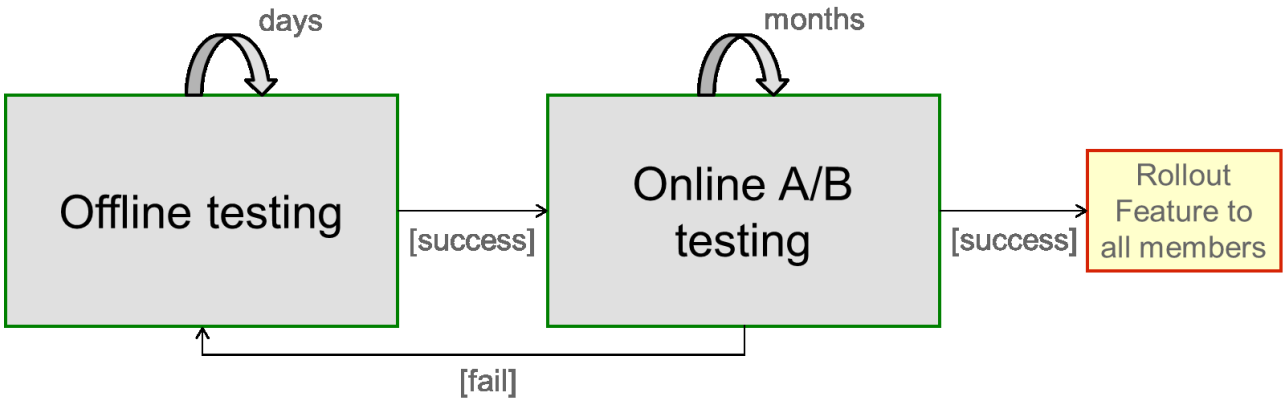
技术(1/3)

- Ranking = Scoring + Sorting + Filtering

Netflix

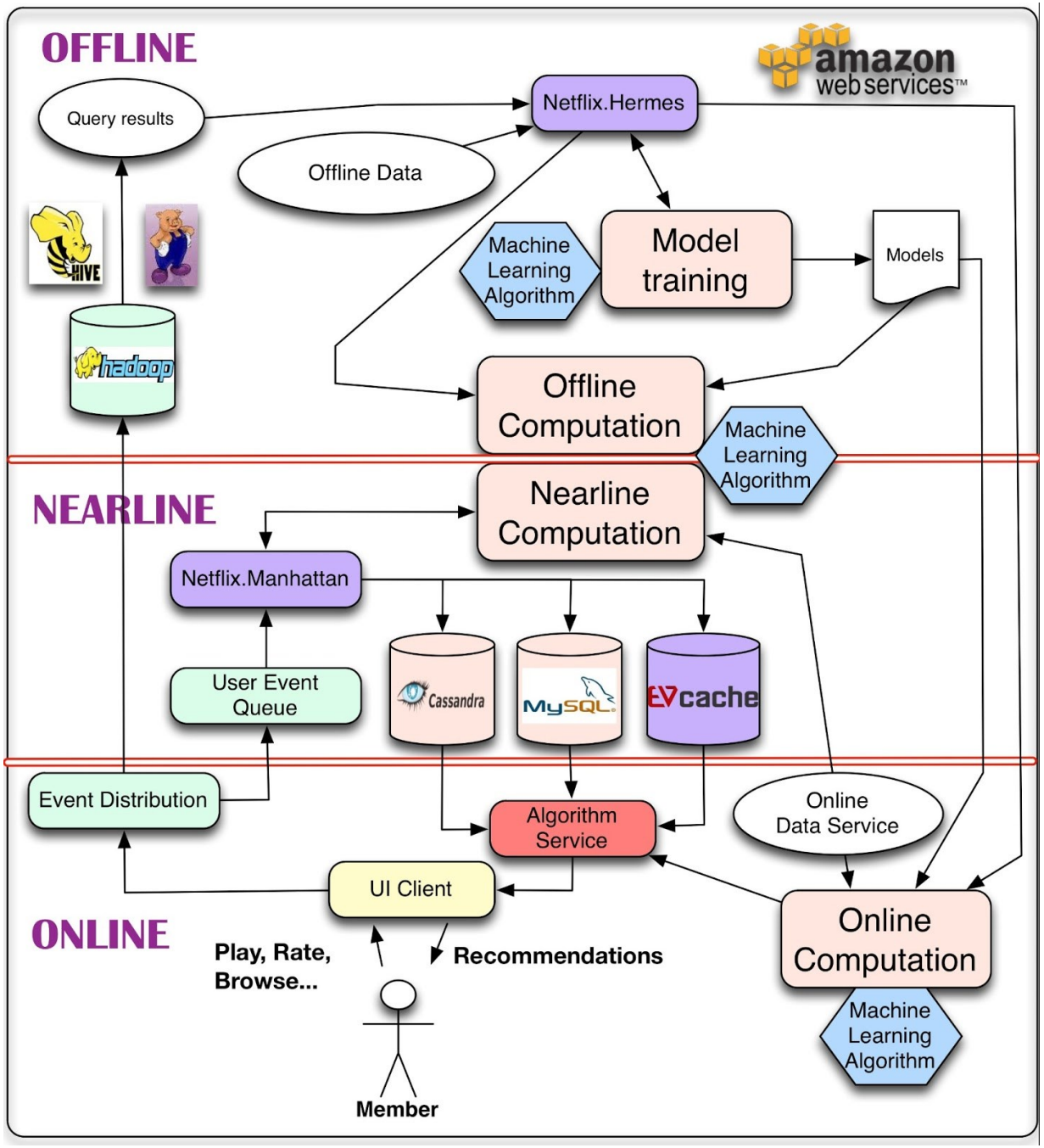
技术(2/3)

- Online testing and online A/B testing



技术(3/3)

Netflix



Netflix

数据平台/系统架构

- Hadoop
- Spark
- Scala/Breeze
- GraphX
- Multi-core machine learning
- …

Netflix

10 Lessons (1/3)

- More data and better models
- You might not need all your Big Data
- The fact that a more complex model does not improve things does not mean you don't need one
  - More complex features may require a more complex model
- Be thoughtful about your training data
  - Time traveling: usage of features that originated after the event you are trying to predict

Xavier Amatriain. 10 Lessons Learned from Building ML Systems. November 2014.

Netflix

10 Lessons (2/3)

- Learn to deal with (the curse of) Presentation Bias
  - Impression bias
- The UI is the algorithm's only communication channel with that which matters most: the users
  - UI->Algorithm->UI
- Data and Models are great. You know what's even better? The right evaluation approach
  - Offline/Online test, A/B test, long-/short- term metrics

Xavier Amatriain. 10 Lessons Learned from Building ML Systems. November 2014.

Netflix

10 Lessons (3/3)

- Distributing algorithms is important, but knowing at what level to do it is even more important
- It pays off to be smart about choosing your hyperparameters
- There are things you can do offline and there are things you can't… and there is nearline for everything in between

Xavier Amatriain. 10 Lessons Learned from Building ML Systems. November 2014.

参考文献

- Ehtsham Elahi and Yves Raimond. **Spark and GraphX in the Netflix Recommender System**. May 19, 2015.
- Xavier Amatriain. **10 Lessons Learned from Building ML Systems**. November 2014.
- Xavier Amatriain. **The Recommender Problem Revisited**. RecSys tutorial, October 2014.
- Xavier Amatriain. **Distributing ML Algorithms: from GPUs to the Cloud**. MMDS, June 2014.
- Xavier Amatriain and Justin Basilico. **System Architectures for Personalization and Recommendation**. The Netflix Tech Blog. March 27, 2013. http://techblog.netflix.com/2013/03/system-architectures-for.html
- Xavier Amatriain. **Netflix Recommendations - Beyond the 5 Stars**. October 22, 2012.