Manipulation Robustness of Collaborative Filtering Systems

Xiang “Robbie” Yan

Department of Electrical Engineering
Stanford University
(Joint Work with Benjamin Van Roy)

April 15, 2009
Abstract

- CF heuristics deployed in industry highly manipulable.

- *Linear* CF algorithms much more robust.

- We provide theoretical guarantee and empirical support.
Preliminaries

- Introduction of CF systems.
- Much online consumption driven by CF. [Anderson 06]
- Nearest Neighbor (NN) methods common.

Customers Who Bought This Item Also Bought

- *A Beautiful Mind* (Widmark Awards Edition) DVD ~ Paul Bettany
  • 5 stars • (867) • $8.99
- *Finding Forrester* DVD ~ Sean Connery
  • 5 stars • (272) • $6.99
- *Rounders* (Collector’s Edition) DVD ~ Matt Damon
  • 4.5 stars • (881) • $11.49
- *Rain Man* (Special Edition) DVD ~ Tom Cruise
  • 4.5 stars • (128) • $9.99

---

Introduction | Framework | Algorithms | Experiments | Conclusion

Xiang “Robbie” Yan | Manipulation Robustness of Collaborative Filtering Systems
Manipulation

Manipulation likely frequent.

- Incentivized by revenue impact.

- Rampant manipulation in other recommendation systems.

- Example: Christianity and homosexuality on Amazon. [Olsen 2002]

- Method: inject ratings “typical” except on targets. [Burke 2005]
Model

- \( N \) products.
- Types \( \in \bar{S} \), ratings \( \in S = \bar{S} \cup \{?\} \).
- \( M \) ratings vectors, each \( \in S^N \).
- Fraction of manipulated data \( r \).
- Honest ratings vectors \( Y \in S^{N \times (1-r)M} \).
- Manipulated ratings vectors \( Z \in S^{N \times rM} \).
- Training set \((Y, Z)\).
CF Algorithms

- Active user inspects and rates products \( \nu_1, \ldots, \nu_n \).
- CF algorithm maps \( \{1, \ldots, N\} \times S^n \times S^{N \times M} \) to PMF over \( \mathcal{S} \):
  \[
p_{\nu_k, x^{k-1}, \mathcal{W}}.
  \]
- Scalar prediction \( \hat{x}_{\nu_k}^{\mathcal{W}} = \mathbb{E}[x_{\nu_k}] \) w.r.t. \( p_{\nu_k, x^{k-1}, \mathcal{W}} \).
- RMS distortion:
  \[
d_n(p, \nu, Y, Z) = \sqrt{\frac{1}{n} \sum_{k=1}^{n} \mathbb{E} \left[ \left( \hat{x}_{\nu_k}^{Y} - \hat{x}_{\nu_k}^{(Y,Z)} \right)^2 \right]}.
  \]
Linear CF Algorithms

- Probabilistic CF algorithm $p$:
  
  for each $W$, exists PMF $\hat{\psi}^{p,W}$ over $\bar{S}^N \times S^N$ for each $n, x$, s.t.
  
  $$p_{n,x,W} = \hat{\psi}^{p,W}(\bar{x}_n | x).$$

- Linear CF algorithm $p$:
  
  for any $W_1 \in S^{N \times M_1}$ and $W_2 \in S^{N \times M_2}$,

  $$\hat{\psi}_{W_1,W_2}^{p} = \frac{M_1}{M_1 + M_2} \hat{\psi}^{p,W_1}_{W_1} + \frac{M_2}{M_1 + M_2} \hat{\psi}^{p,W_2}_{W_2}.$$

Xiang “Robbie” Yan

Manipulation Robustness of Collaborative Filtering Systems
Theorem

$p$ is a linear CF algorithm. For all $M$, $r$, $Y$, $Z$, and $\nu$,

$$d_n(p, \nu, Y, Z) \leq \sqrt{\frac{1}{2n} \ln \frac{1}{1 - r}}.$$ 

Intuition:

- $\hat{\psi}^{p,W}$ is convex combination of $\hat{\psi}^{p,Y}$ and $\hat{\psi}^{p,Z}$.
- Weights on the latter decays as $n$ increases.
Practical Implications

Bound $\frac{1}{2n} \ln \frac{1}{1-r}$ for $r = 0.1$

Binary setting: $r \leq 0.1$, $n \geq 22 \Rightarrow$ % correct predictions $\downarrow \leq 0.05$. 
Empirical Evaluation

- Netflix data: (user, movie, rating) tuples.

- 5K users, 500 movies, 200K ratings.

- 20% test data.

- Injected 50% manipulated ratings vectors. In each,
  - Half sampled from empirical distribution.
  - Half are 1’s.

- Applied
  - Linear CF algorithm: kernel density estimation.
  - NN algorithm: $k$NN.
Results

Bound for $r = 0.5$.
Theoretical and empirical analyses demonstrate that

- NN algorithms highly manipulable.
- Linear CF algorithms counter manipulation.
- Users should rate a minimum number of products.

Our framework and insights can

- Help make existing CF systems more robust.
- Help assess additional countermeasures.