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.

Customers Who Bought This Item Also Bought

A Beautiful Mind (Widescreen Awards Edition) DVD ~ Paul Bettany
★★★★☆ (667) $8.99

Finding Forrester DVD ~ Sean Connery
★★★★☆ (217) $6.99

Rounders (Collector's Edition) DVD ~ Matt Damon
★★★★½ (165) $11.49

Rain Man (Special Edition) DVD ~ Tom Cruise
★★★★½ (128) $9.99

- Much online consumption driven by CF. [Anderson 06]

YouTube Google News ebay NETFLIX amazon.com

- Nearest Neighbor (NN) methods common.

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 $\in \overline{S}$, ratings $\in S = \overline{S} \cup \{?\}$.
- 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 $\overline{S}$:

$$p_{\nu_k, x^{k-1}, W}.$$

- Scalar prediction $\tilde{x}^W_{\nu_k} = \mathrm{E}[x_{\nu_k}]$ w.r.t. $p_{\nu_k, x^{k-1}, W}$.
- RMS distortion:

$$d_n(p, \nu, Y, Z) = \sqrt{\frac{1}{n} \sum_{k=1}^{n} \mathrm{E}\left[\left(\tilde{x}^Y_{\nu_k} - \tilde{x}^{(Y,Z)}_{\nu_k}\right)^2\right]}.$$

Linear CF Algorithms

- Probabilistic CF algorithm p:

  for each W, exists PMF $\hat{\psi}^{p,W}$ over $\overline{S}^N \times S^N$ for each n, x, s.t.

  $$p_{n,x,W} = \hat{\psi}^{p,W}(\overline{x}_n|x).$$

- Linear CF algorithm p:

  for any $W_1 \in S^{N\times M_1}$ and $W_2 \in S^{N\times M_2}$,

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

Results

Theorem

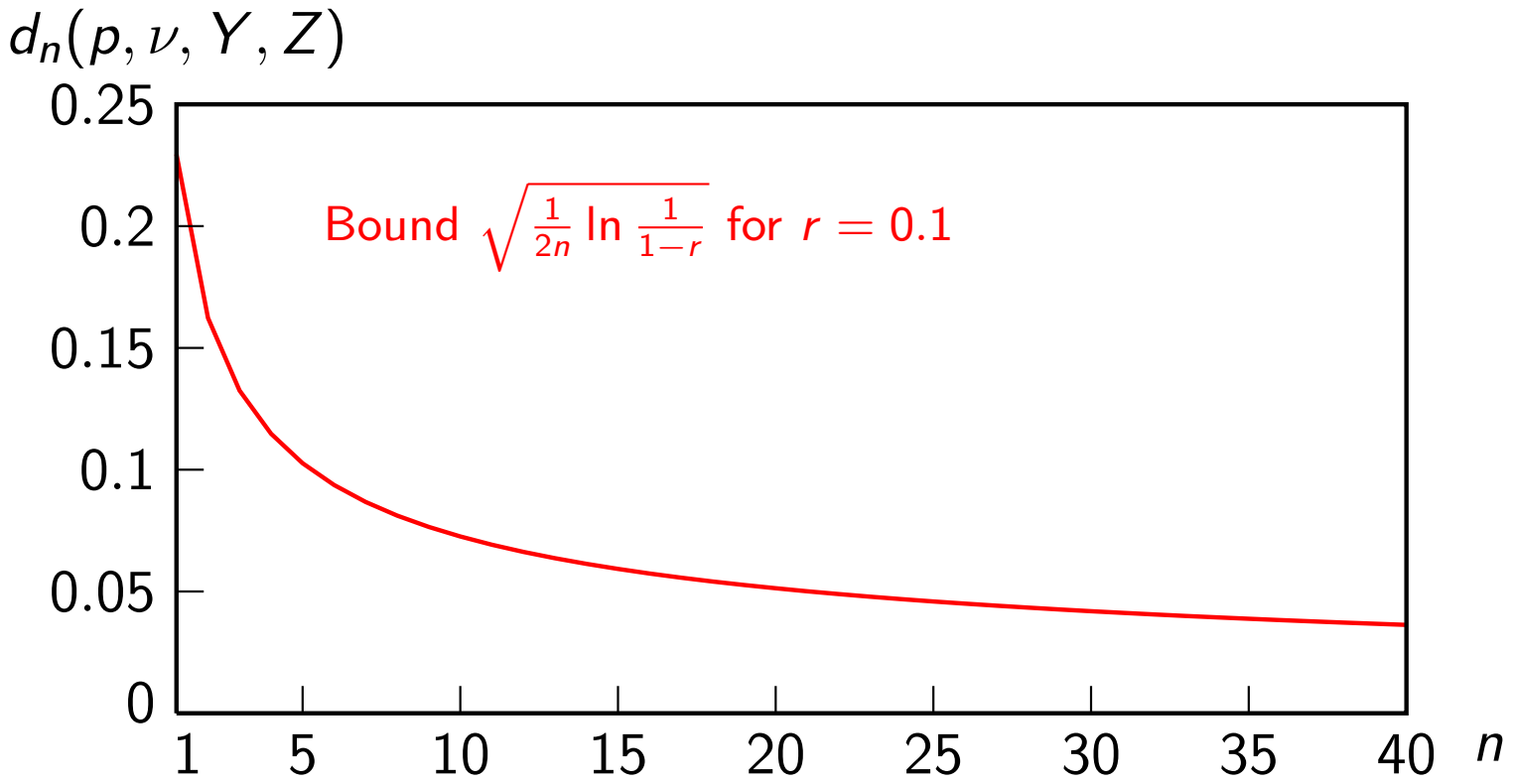
p is a linear CF algorithm. For all M, r, Y, Z, and ν,

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

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

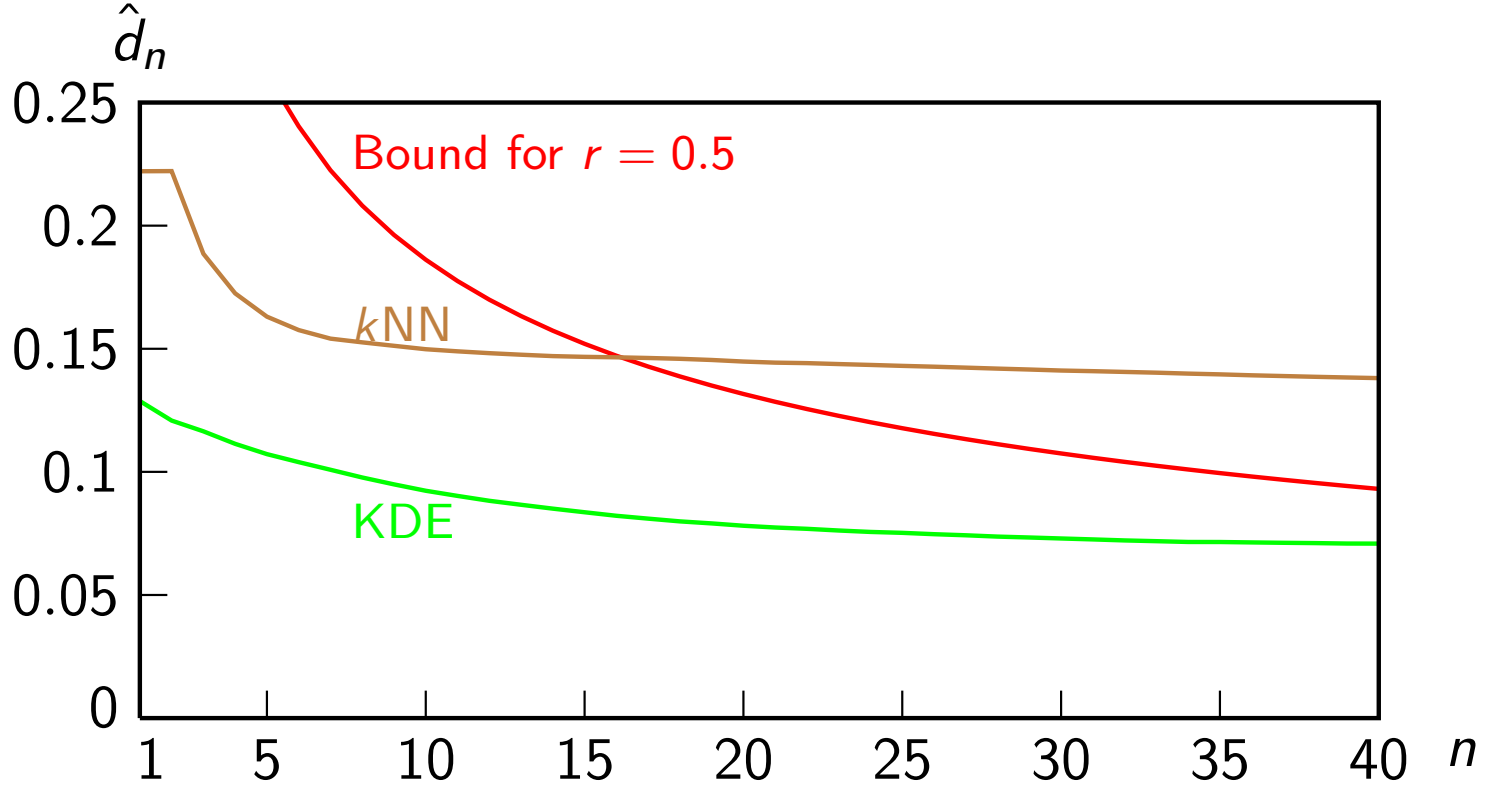


Binary setting: $r \leq 0.1$, $n \geq 22 \Rightarrow$ % correct predictions $\searrow \leq 0.05$.

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: kNN.

Results


$\hat{d}_n$
0.25
0.2
0.15
0.1
0.05
0
Bound for $r = 0.5$
kNN
KDE
1
5
10
15
20
25
30
35
40
n

Conclusion

- 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.