香港中文大學
The Chinese University of Hong Kong

The ACM Conference Series on Recommender Systems

# Modeling Assimilation-Contrast Effects in Online Product Rating System: Debiasing and Recommendation

Xiaoying Zhang, Junzhou Zhao, John C.S. Lui
Dept. of Computer Science & Engineering
The Chinese University of Hong Kong

# Online rating systems

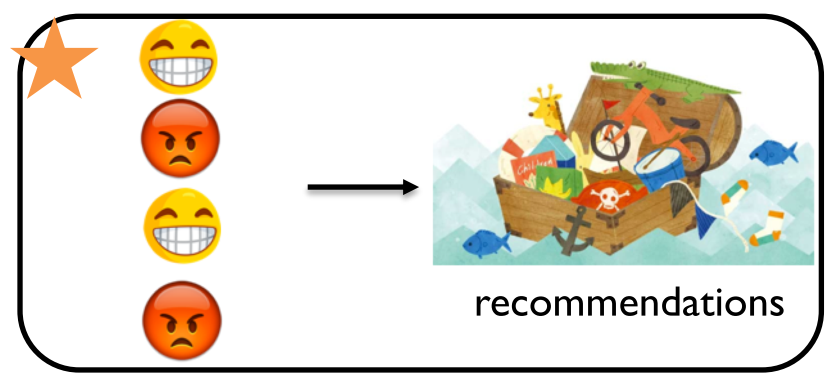
- Online ratings systems play an important role in our life

# Online rating systems

- Online rating systems allow users to rate/review the products they consumed.


Function 1: Aiding wise purchase decisions


Function 2: Aiding future recommendations



# Online rating systems

- Online rating systems allow users to rate/review the products they consumed.

Function 1: Aiding wise purchase decisions

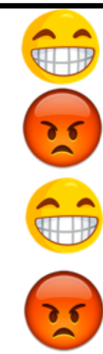

Product 1 Product 2

**The unbiasedness of ratings:**

- users' ratings indeed reflect their true evaluations.

**Crucial**

Function 2: Aiding future recommendations



recommendations

# The ruin of the unbiasedness of ratings

Empirical studies

- Social news aggregation web : Muchnik et al. [Science'13], Weninger et al. [HT'15]
- Rating jokes: Adomavicius et al. [IntRS@Recsys'16]
- Music lab: Salganik et al. [Science'06]

…..

Disclosing historical opinions distorts subsequent users' decision making

The unbiasedness of ratings is ruined!

# Research problem

- For each single rating in **real rating system**:
  - How to discover the distortions from historical ratings?
  - How to restore its unbiasedness?
- Main challenge:


Historical ratings


The next single rating

# Contributions


Empirical Observations

- The Assimilation-Contrast effects
- A psychological explanation

Modelling

- Historical Influence Aware Latent Factor Model (HIALF)

Experiments

- Evaluation from three different aspects on 42 million real ratings.


Application

- Debiased recommender system
- Wiser purchase decisions

# Empirical Measurements

- 42 million real ratings from Amazon and Tripadvisor
- Our problem:

Prior expectation $e_{p,i}$ $\xrightarrow{?}$ The next ratings $r_{p,i}$

- $r_{p,i}$: the $i$-th rating of product $p$
- $\mathcal{H}_{p,i} \triangleq (r_{p,1}, \dots, r_{p,i-1})$: historical ratings of $r_{p,i}$
- prior expectation: $e_{p,i} = \frac{1}{i-1} \sum_{k=1}^{i-1} r_{p,k}$

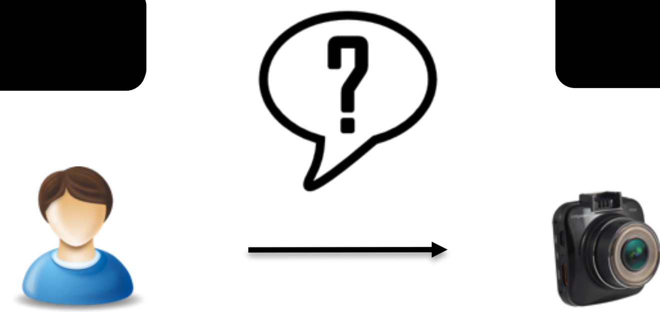
# Empirical Measurements

- Factors that affect users' given ratings

Historical ratings

[redacted]

[redacted]

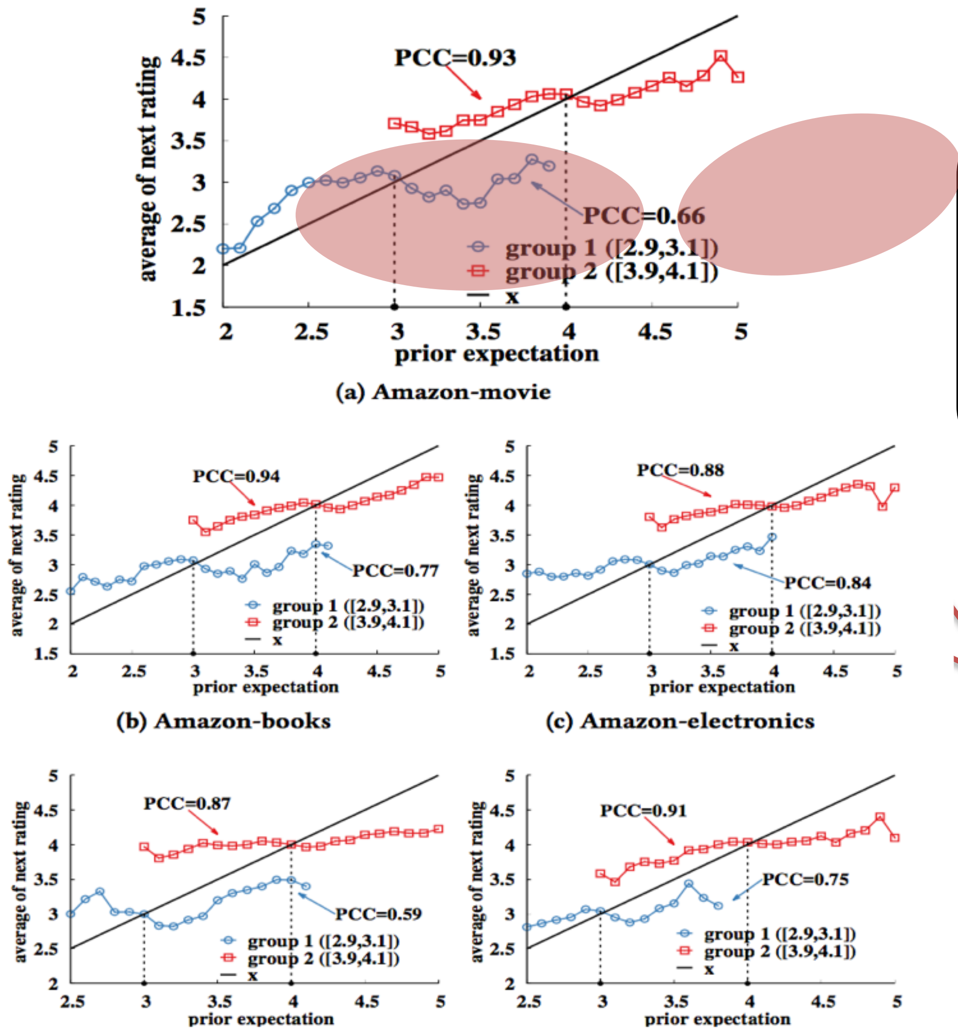


Group products with similar average ratings.

- Two selected groups: average rating in [2.9, 3.1] ( [3.9, 4.1])
- approximately true quality : 3 (4)

How prior expectation influence the next rating, on average ?

# Empirical Observations




(a) Amazon-movie

(b) Amazon-books

(c) Amazon-electronics

(d) Amazon-clothes

(e) Tripadvisor

## Observation 1

- Products' historical ratings do affect the next rating
  - PCC ∈ [0.59,0,94] → a positive correlation

## Observation 2

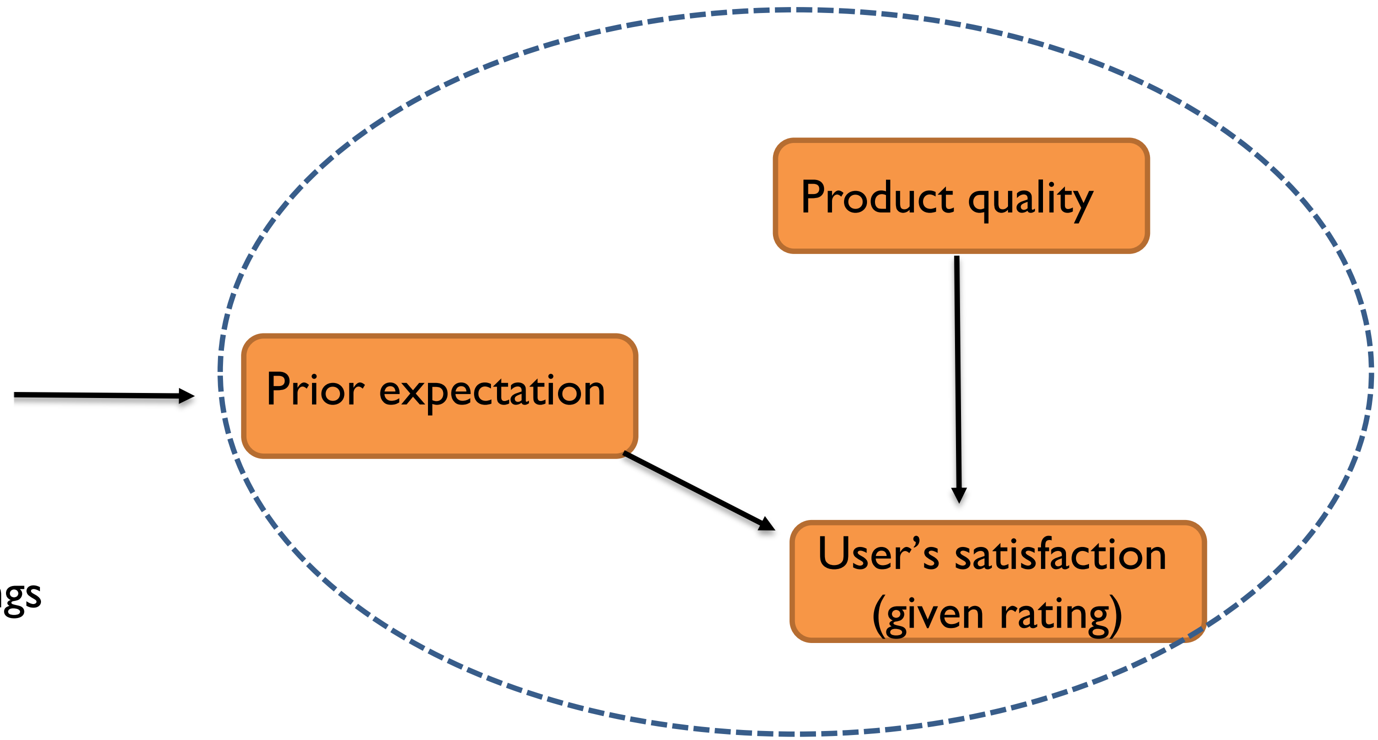
- Each curve is divided into two parts by the group's *approximately true quality*.

# Explanation of Observations

- Why do historical ratings influence its next rating?

**The Users' Satisfaction Theory**
Anderson (1973)

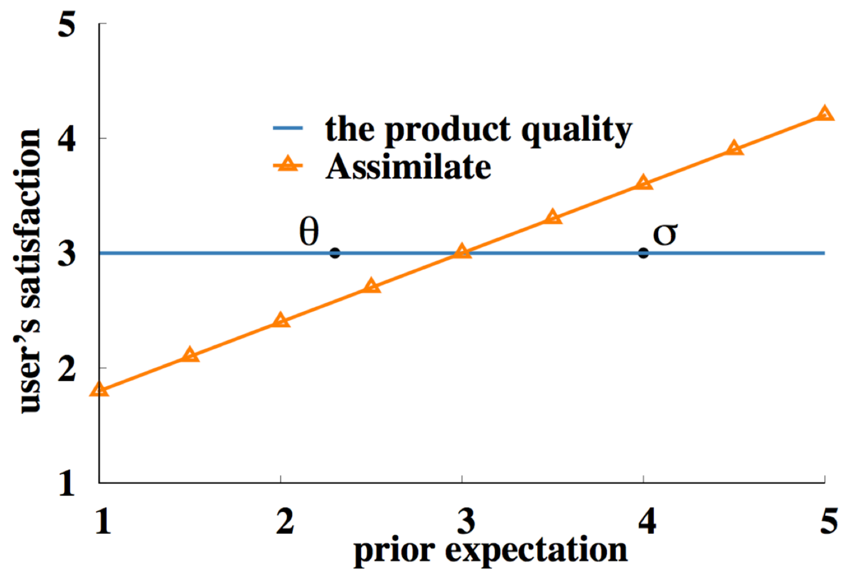

Historical ratings



# Explanation of Observations

- Why does the influence of historical ratings behave consistently as before?

- **Users' satisfaction theory (Anderson (1973)) in psychology**
  - "Assimilate" theory
  - "Contrast" theory
  - "Assimilate-Contrast" theory

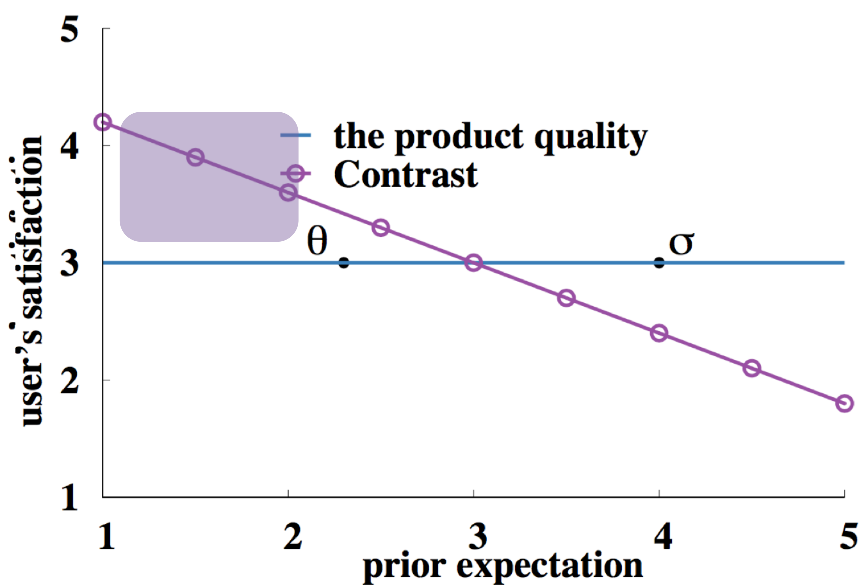



Sample representation
(quality (blue line): 3)

# Explanation of Observations

- Why does the influence of historical ratings behave consistently as before?

- **Users' satisfaction theory (Anderson (1973)) in psychology**
  - "Assimilate" theory
  - "Contrast" theory
  - "Assimilate-Contrast" theory

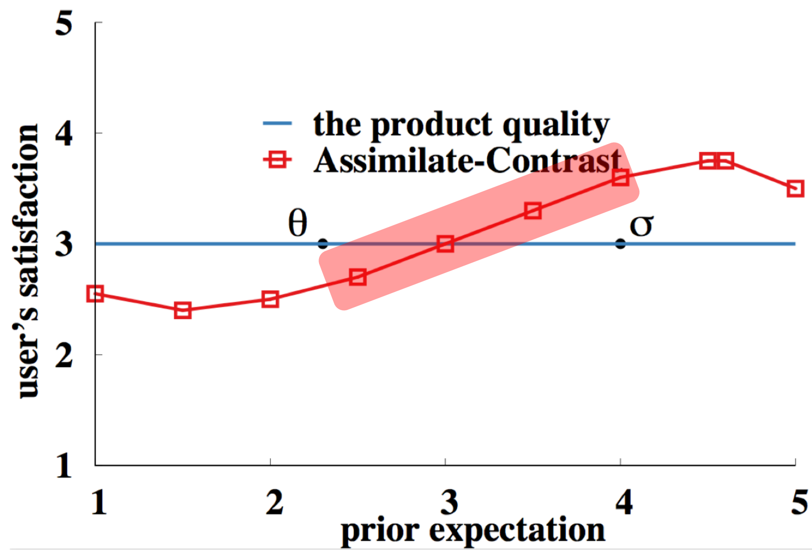



Sample representation (quality (blue line): 3)

# Explanation of Observations

- Why does the influence of historical ratings behave consistently as before?

- **Users' satisfaction theory (Anderson (1973)) in psychology**
  - "Assimilate" theory
  - "Contrast" theory
  - "Assimilate-Contrast" theory

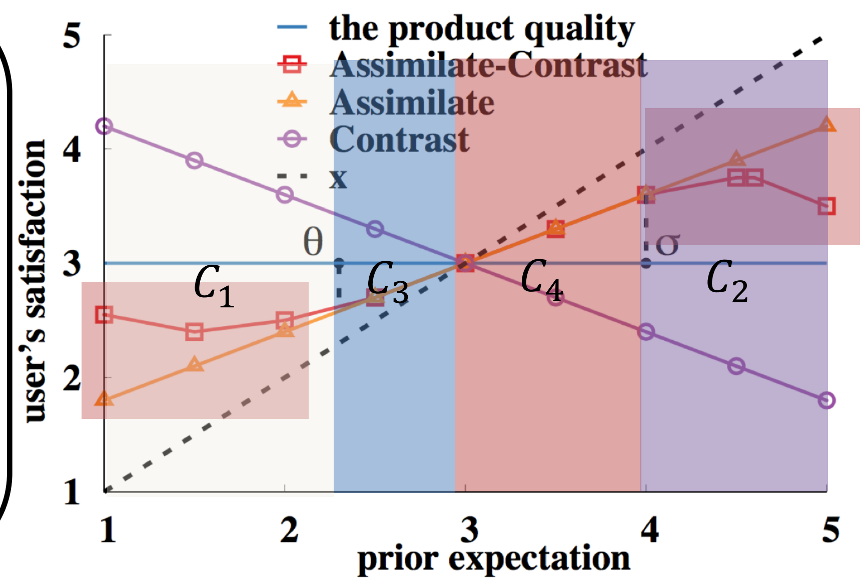



Sample representation
(quality (blue line): 3)

# Explanation of Observations

- Why does the influence of historical ratings behave consistently as before?

- **Users' satisfaction theory (Anderson (1973)) in psychology**
  - "Assimilate" theory
  - ~~"Contrast" theory~~
  - "Assimilate-Contrast" theory




Hypothesis Test on slope difference in $C_1$ and $C_3$ ($C_2$ and $C_4$ ) → "Assimilate-Contrast" theory !

# Deficiencies of existing works

- HEARD (HEARD[KDD'14]): how historical ratings $\mathcal{H}_{p,i}$ influence the general rating distribution after next M ratings ?
- Can HEARD explain our observations?

**Conclusion**:

- HEARD **fails to explain** our empirical observations
- We need **a better model**.

# Contributions


Empirical Observations

- The Assimilation-Contrast effects
- A psychological explanation

Modelling

- Historical Influence Aware Latent Factor Model (HIALF)

Experiments

- Evaluation from three different aspects on 42 million real ratings.


Application

- Debiased recommender system
- Wiser purchase decisions

# Our Proposed Model

- Historical Influence Aware Latent Factor Model (HIALF)
  - HIALF predicts the $i$-th rating of product $p$ given by user $u$:

User bias

How easily will user $u$ be influenced ?

$$\hat{r}_{p,i,u} = b_u + q_{u,p} + \alpha_u h_{p,i}$$

Distortion from historical ratings

product $p$'s quality in $u$'s view

# Our Proposed Model

- Historical Influence Aware Latent Factor Model (HIALF)
  - HIALF predicts the $i$-th rating of product $p$ given by user $u$:

$$\hat{r}_{p,i,u} = b_u + q_{u,p} + \alpha_u h_{p,i}$$

product $p$'s quality in $u$'s view

$$q_{u,p} = g + \boxed{b_p} + \boxed{\boldsymbol{x}_u^T \boldsymbol{y}_p}$$

product p's intrinsic quality

user u's overall interest in product p

# Our Proposed Model

- Historical Influence Aware Latent Factor Model (HIALF)
  - HIALF predicts the $i$-th rating of product $p$ given by user $u$:

$$\hat{r}_{p,i,u} = b_u + q_{u,p} + \alpha_u \boxed{h_{p,i}}$$

Distortion from historical ratings

$$h_{p,i} = f\left(\left|\mathcal{H}_{p,i}\right|\right)\boldsymbol{\beta}(e_{p,i} - q_{u,p})$$

$f(x)$: scaling effect when $\left|\mathcal{H}_{p,i}\right| = x$

- $\left|\mathcal{H}_{p,i}\right|$ boosts $h_{p,i}$ (Latané(1981) )

$\boldsymbol{\beta}(x)$ : induced bias when $x$=$e_{p,i} - q_{u,p}$

- The "Assimilate-Contrast" theory

# Our Proposed Model

- $\boldsymbol{\beta}(x)$: induced bias when $x = e_{p,i} - q_{u,p}$.
  - Kernel regression ! **(a non-parametric way)**:
    - A good way to validate model: the assimilation-contrast effects.

- If we know a set of samples $\{(x_l, r_l)\}_{l=1}^{n}$ from $\boldsymbol{\beta}(x)$, we can approximate $\boldsymbol{\beta}(x)$ by:

$$\widehat{\boldsymbol{\beta}}(x) = \frac{\sum_{l=1}^{n} w(x, x_l) r_l}{\sum_{i=1}^{n} w(x, x_i)}$$

  - $w(x, x_l) = \exp(-\kappa(x - x_l)^2)$ : greater weight to $x$ closer to $x_l$
- How we get $\{(x_l, r_l)\}_{l=1}^{n}$?
  - $x_l$ : in a limited range ($e_{p,i} \in [1,5], q_{u,p} \in [1,5]$)
    - our dataset: $\{x_1, \ldots, x_l \ldots, x_n\} = \{-4, -3.5, \ldots, 3.5, 4\}$
  - $r_l$: parameters, learned from data.

# Our Proposed Model

- $f(x)$: a scaling effect when $\left|\mathcal{H}_{p,i}\right| = x$
  - $f(x) \propto |\mathcal{H}_{p,i}|$
  - slope of f (x ) (Latané(1981)):
    - decrease as x increases
    - positive

$$f(x) = \boxed{\frac{a}{1+\exp(-bx)}} - \boxed{\frac{a}{2}}$$

A sigmod function

$f(0) = 0$

# Our Proposed Model

- Prior expectation $e_{p,i}$
  - users focus more on recent ratings
  - a general formula

$$e_{p,i} = \frac{\sum_{k=1}^{i-1} \xi(i-k) r_{p,k}}{\sum_{k=1}^{i-1} \boxed{\xi(i-k)}}$$

$\xi(d) = \exp(-\boxed{\gamma} d)$: larger weight for smaller $d$ (recent ratings)

How much users prefer recent ratings

- $\gamma = 0$: $e_{p,i}$ = average of $\mathcal{H}_{p,i}$
- larger $\gamma$ → users focus more on recent rating
- set by cross-validation.

# Model Inference

Real rating

Predicted rating

$$\min_{\Theta} \sum_{(p,i,u)\in K} \left(\boxed{r_{p,i,u}} - \boxed{\hat{r}_{p,i,u}}\right)^2 + \lambda_{rec}\left(b_u^2 + b_p^2 + \left|\left|\boldsymbol{x_u}\right|\right|^2 + \left|\left|\boldsymbol{y_p}\right|\right|^2\right)$$

$$+\lambda_f(a^2 + b^2) + \lambda_\beta\left(\sum_l v_l^2\right) + \lambda_\alpha(\alpha_u^2)$$

$\Theta = \{g, \{b_u\}, \{b_p\}, \{\boldsymbol{x_u}\}, \{\boldsymbol{y_p}\}, \{\alpha_u\}, a, b, \{v_l\}\}$, we use stochastic gradient descent (SGD) to find the parameters.

# Contributions


**Empirical Observations**

- The Assimilation-Contrast effects
- A psychological explanation

**Modelling**

- Historical Influence Aware Latent Factor Model (HIALF)

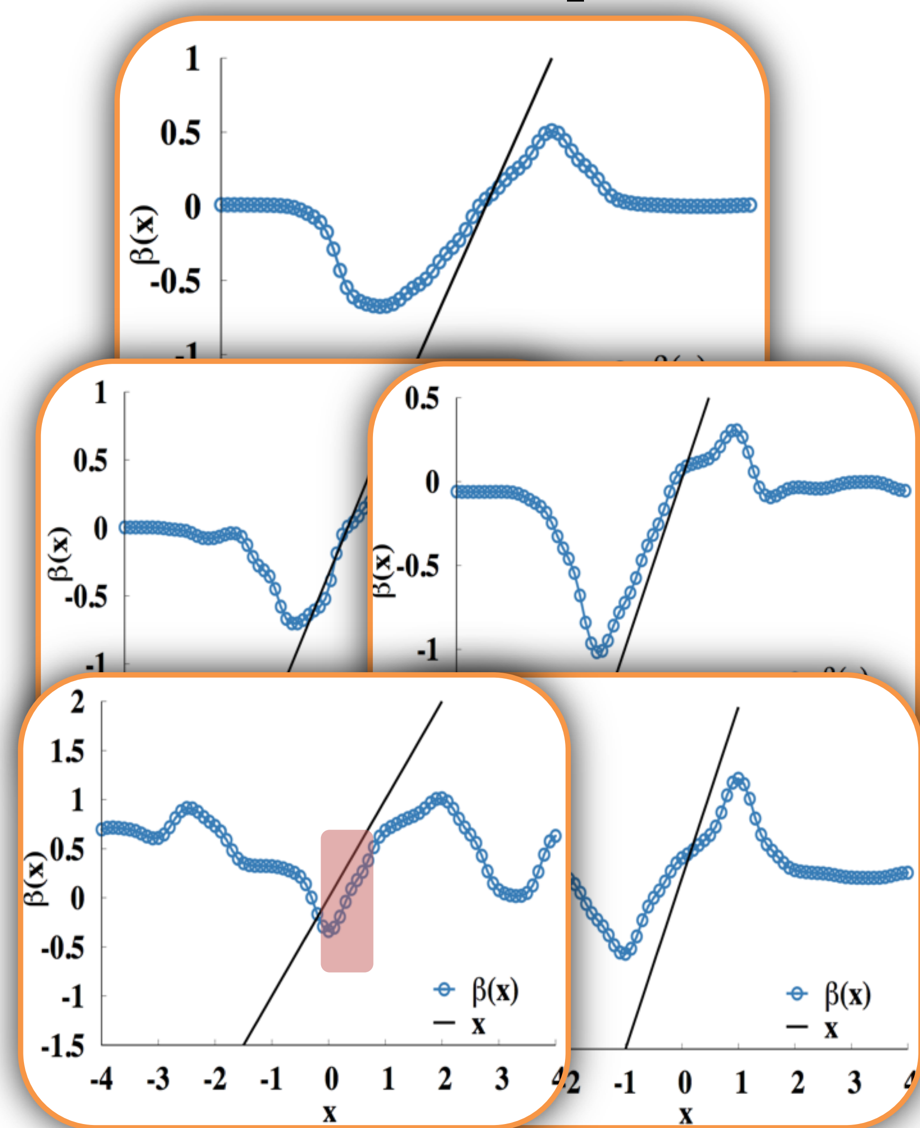
**Experiments**

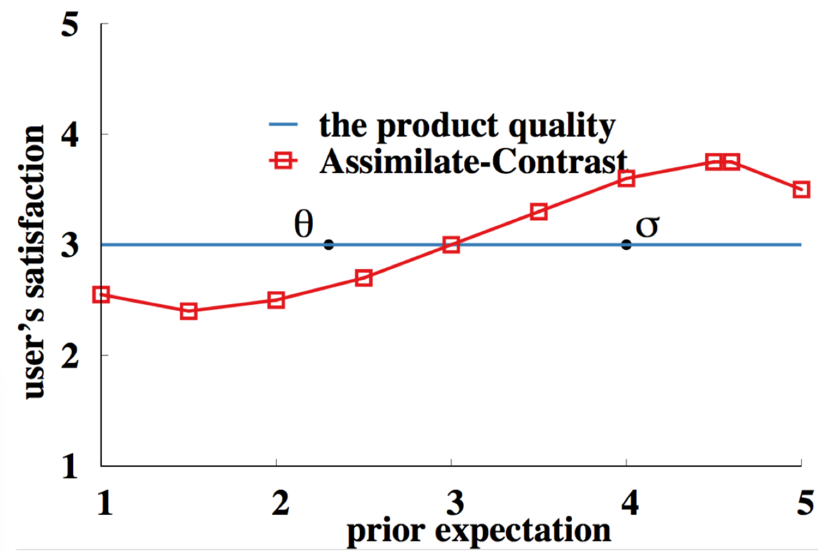
- Evaluation from three different aspects on 42 million real ratings.


**Application**

- Debiased recommender system
- Wiser purchase decisions

# Experiments: $\beta(x)$




Sample representation

All $\beta(x)$ perfectly match the assimilation-contrast effects.

# Experiments : Prediction

Table 3: *MSE* on five datasets

| | Amazon-movie | Amazon-books | Amazon-electronics | Amazon-clothes | Tripadvisor |
|---|---|---|---|---|---|
| HEARD | 1.5826 | 1.5548 | 3.1170 | 2.1550 | 1.3135 |
| LF | 1.2794 | 1.0777 | 1.9634 | 1.4123 | 1.0074 |
| HIALF-AVG | 1.2054 | 1.0619 | 1.9357 | 1.3985 | 0.9805 |
| HIALF | **1.1194** | **1.0318** | **1.8764** | **1.3759** | **0.9405** |
| benefits of HIALF over HEARD | 29.27% | 32.83% | 39.80% | 35.17% | 28.40% |
| benefits of HIALF over LF | 12.51% | 4.26% | 4.43% | 2.58% | 6.64 % |

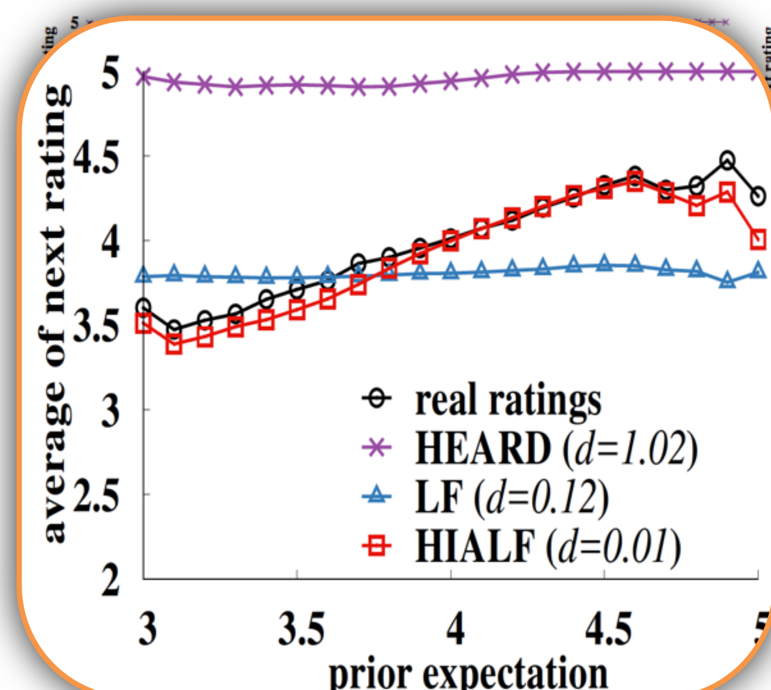
our model significantly outperforms alternatives on all datasets.

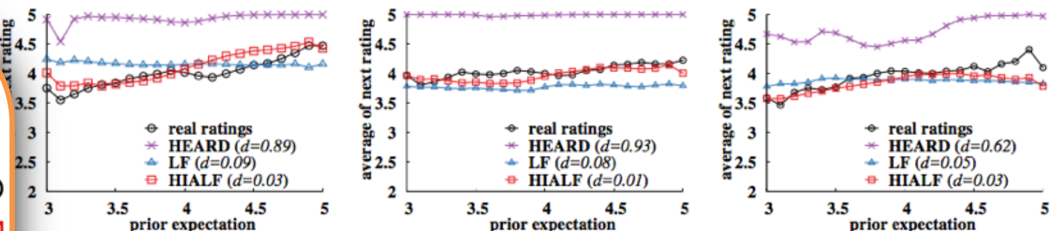
- Over HEARD: 33% reduction, on average
- Over LF : 6% reduction, on average

- Suffered distortions → gained benefits
  - Amazon-movie suffers the largest influence→ it gains the largest benefits vs LF
  - Amazon-clothes suffers the smallest influence→ it gains the smallest benefits vs LF

# Experiments :Fitting

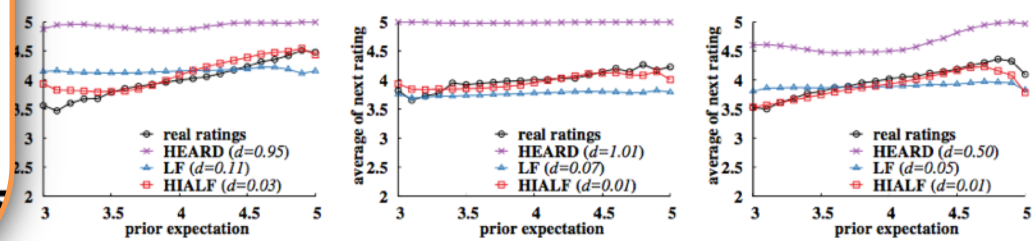
- Fitting empirical observations
  - prior expectation: average of historical ratings




(c) Amazon-books (d) Amazon-clothes (e) Tripadvisor

n: the new general formula




(c) Amazon-books (d) Amazon-clothes (e) Tripadvisor

HIALF perfectly fits previous observations in real ratings!

# Contributions


Empirical Observations

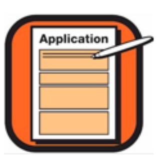
- The Assimilation-Contrast effects
- A psychological explanation

Modelling

- Historical Influence Aware Latent Factor Model (HIALF)

Experiments

- Evaluation from three different aspects on 42 million real ratings.




Application

- Debiased recommender system
- Wiser purchase decisions

# Application(1)

- Debiased Recommender System
  - We can generate the debiased recommendation score of product $p$ for user $u$ as :

$$rec(p,u) = g + b_u + b_p + \boldsymbol{x}_{\boldsymbol{u}}^{\boldsymbol{T}} \boldsymbol{y}_{\boldsymbol{p}}$$

  - Evaluation:
    - Test set : ratings without historical ratings
    - Metric : Root Mean Square Error (RMSE)

| category | LF | debiased recsys |
|---|---|---|
| Amazon-movie | 1.0639 | 1.0465 |
| Amazon-books | 0.9125 | 0.8922 |
| Amazon-electronics | 1.2273 | 1.2083 |
| Amazon-clothes | 1.1239 | 1.1034 |
| Tripadvisor | 1.1919 | 1.1776 |

Better recommendations !

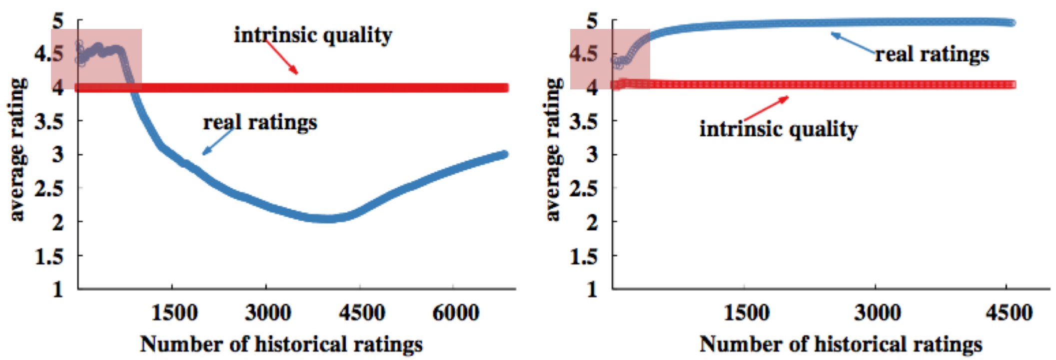
# Application(2)

- Exposing the intrinsic product quality for wiser purchase decisions

The $i$-th rating of product $p$ without historical distortions

$$q_p^* = \boxed{\frac{1}{n}\sum_{i=0}^{n}} \boxed{(g + b_p + \boldsymbol{x}_{\tilde{u}(p,i)}^{\boldsymbol{T}} \boldsymbol{y_p})}$$

  - Why we need to reveal the intrinsic quality ?



(a) sample product 1 (b) sample product 2

- similar intrinsic quality (4)
- similar initial ratings (4.5)
- After a sequence of ratings with different trends, the average rating are 3.2 and 4.9 respectively (differ at about 1.7).

# Contributions


Empirical Observations

- The Assimilation-Contrast effects
- A psychological explanation

Modelling

- Historical Influence Aware Latent Factor Model (HIALF)

Experiments

- Evaluation from three different aspects on 42 million real ratings.


Application

- Debiased recommender system
- Wiser purchase decisions

# Questions?

# Reference


- Muchnik et al. [Science'13] Muchnik L, Aral S, Taylor S J. Social influence bias: A randomized experiment. Science, 2013, 341(6146): 647-651.
- Adomavicius et al. [IntRS@Recsys'16] Gediminas Adomavicius, Jesse Bockstedt, Shawn P. Curley, and Jingjing Zhang. Understanding Effects of Personalized vs. Aggregate Ratings on User Preference . In IntRS Workshop@RecSys 2016. 14–21.
- Salganik et al. [Science'06] Matthew J Salganik, Peter Sheridan Dodds, and Duncan J Watts. Experimental study of inequality and unpredictability in an artificial cultural market. Science 311, 5762 (2006), 854–856.
- Weninger et al. [HT'15] Tim Weninger, Thomas James Johnston, and Maria Glenski. 2015.Random voting effects in social-digital spaces: A case study of reddit post submissions. In Proceedings of the 26th ACM conference on hypertext & social media. ACM,293–297
- HEARD[KDD'14] Wang, Ting, Dashun Wang, and Fei Wang. "Quantifying herding effects in crowd wisdom." KDD,2014
- Adomavicius et al. [IntRS@Recsys'14] Gediminas Adomavicius, Jesse Bockstedt, Shawn Curley, and Jingjing Zhang.2014. De-biasing user preference ratings in recommender systems. In IntRSWorkshop@RecSys 2014. 2–9.
- Krishnan et al.[Recsys'14] Sanjay Krishnan, Jay Patel, Michael Franklin, and Ken Goldberg. 2014. Socialinfluence bias in recommender systems: a methodology for learning, analyzing, and mitigating bias in ratings. In Proceedings of the 8th ACM Conference on Recommender systems. 137–144.
- Richard(2014) Richard L Oliver. . Satisfaction: A behavioral perspective on the consumer. Routledge.
- Anderson (1973) Rolph E Anderson. Consumer dissatisfaction: The effect of disconfirmed expectancy on perceived product performance. Journal of marketing research (1973)
- Latané(1981) Bibb Latané. 1981. The psychology of social impact. American psychologist. 1981
- Zhu et.al (2014) Haiyi Zhu and Bernardo A Huberman. 2014. To switch or not to switch: understanding social influence in online choices. American Behavioral Scientist 58, 10 (2014), 1329–1344.