Data Science in e-commerce

Donghui Yan

University of Massachusetts Dartmouth

June 15, 2018

Outline

- Introduction
- Movie rating and recommendation
- Shopping item recommendation
- News recmmendation

Data science

- Phenomenal growth in last two decades
  - Availability of big data
  - Advance in computing technology
  - Development of statistics and machine learning methods
- Wide range of applications
  - E-commerce, healthcare, finance, sports statistics, credit scoring, entertainment etc
  - ♠ Relevant in any domain as long as there is data.

Data science applications in e-commerce

- Computational advertising
- Logistics and supply chain management
- Customer service (value estimation, potential customers, promotions etc)
- Fraud detection
- Item recommendation (focus of this talk)
  - Movie rating and recommendation
  - Shopping items recommendation
  - News recommendation.

Recommendation systems

- Content-based
  - Recommended items similar to those liked by user in the past
  - Focus on properties of items
    - Similarity of items with what the user liked in the past
    - e.g., similarity of movies by actors, directors, genres, subject etc
- Collaborative filtering based
  - Recommended items that people with similar preferences liked
  - Focus on the relationship between users and items
    - e.g., user who bought item A may also buy item B
- Hybrid approach
  - Combination of the two.

Illustration by examples

- Movie rating and recommendation
- Shopping items recommendation
- News content recommendation.

Movie rating and recommendation

Movie rating

Movie rating and recommendation

- Given some existing user ratings
  - Predict user's rating to a movie he has not rated before
  - Recommend a movie to a user
- Example
  - T_1: Twilight, T_2: Titantic, C: Casablanca, G: Gettysburg, S: Star wars, H: Harry potter
  - To fill up those missing values.

	T_1	T_2	C	G	S	H
Alice	4			5	1	
Bob	5	5	4			
Candice			2	4	5	
David		3				3
Eve	5	2	4			

Movie profile

- A record representing important characteristics of item
  - Actors of the movie
  - The director
  - Genre
  - The author
  - Subject
  - Production: year, screen size/color, artists, composer etc
  - Ratings
- Similarity of items
  - Cosine distance.

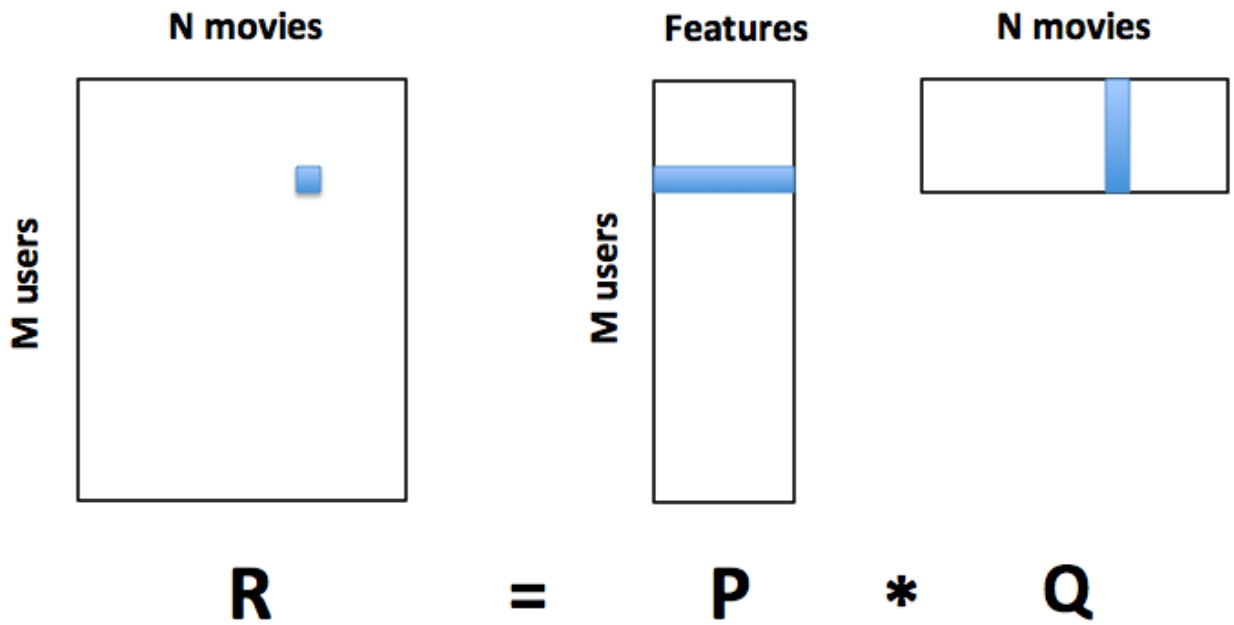
Approaches

- Treat as a missing value problem
  - Expectation-maximization (EM) algorithm
- Treat as low-rank matrix completion problem
  - Convex relaxation (Candes and Recht 2009)
  - Gradient descent (Keshaven, Montanari and Oh 2010)
- Non-negative matrix factorization
  - Idea: there exists some latent features that characterizes how a user rates a movie
    - E.g., two users would rate a certain movie high if they both like the actors, or its genre.

Non-negative matrix factorization

- Given user-rating matrix R, find matrices P and Q s.t.

$$R_{m\times n} \approx P_{m\times|f|} \times Q^T_{n\times|f|}$$



Non-negative matrix factorization

- Estimated ratings $\hat{r}_{ij}$ given by

$$\hat{r}_{ij} = \sum_{k=1}^{|f|} p_{ik} q_{kj}$$

- Minimize discrepancy between $\hat{r}_{ij}$ and given ratings r_{ij}

$$\arg\min_{P,Q} \sum_i \sum_j (r_{ij} - \hat{r}_{ij})^2$$

- Regularization

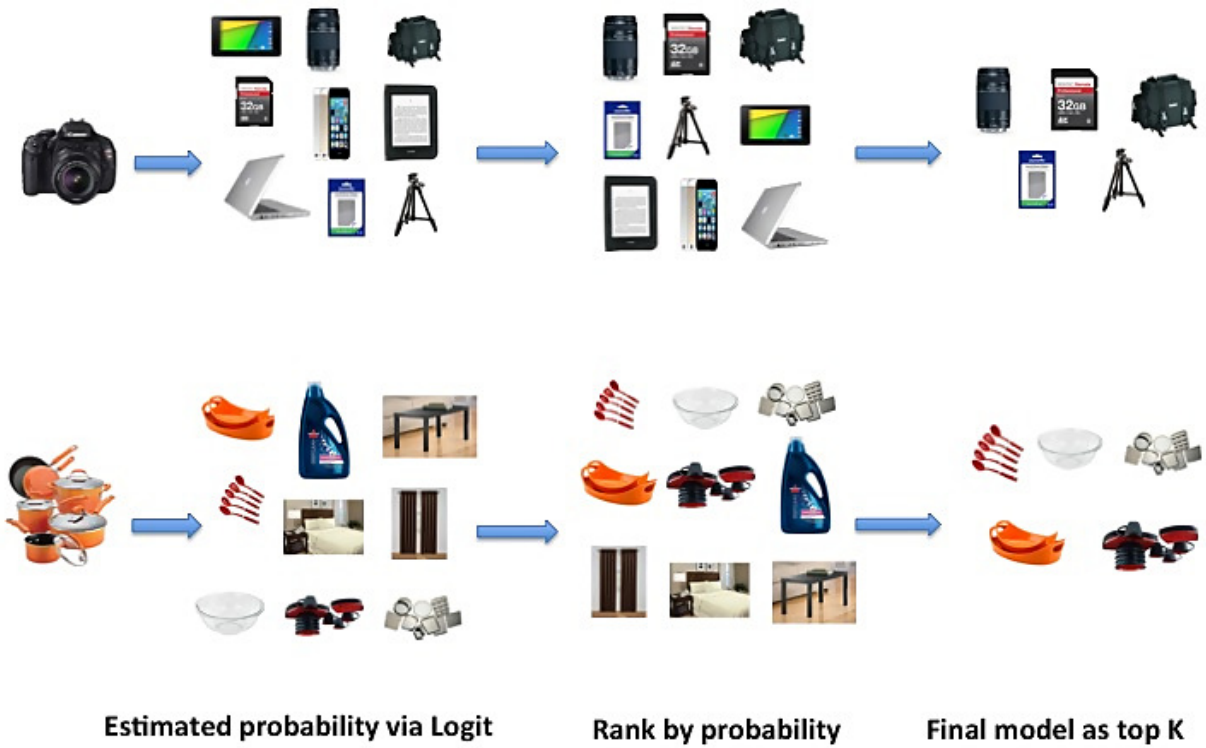
$$\arg\min_{P,Q} \sum_i \sum_j (r_{ij} - \hat{r}_{ij})^2 + \lambda(||P|| + ||Q||).$$

Shopping items recommendation

Item recommendation

Shopping items recommendation

- An effective way to promote sales
  - e.g., *Amazon* boosts its sales by 20-30%, *Walmart* by 5-10%



What items to recommend when a user buys A?

- Only a few items would be recommended
  - Given the space limit of a web page
- Strategy: show items a user would *most likely* buy
  - Or items that leads to the most profit (expected)
- Existing approaches
  - Counts/frequency based
    - Counts of times buying item B_i when A is bought
    - Recommendation by counts
  - Mixed effects model
  - Posterior probability $P(B \mid A)$ based
    - Or Profit$(B) \cdot P(B \mid A)$ for the most profit
    - Then *rank* those B's according to likelihood to buy or profit
    - Pick a few *top* ones (e.g., 5) to recommend.

Estimating posterior probability $P(B \mid A)$

- Data: historical sales record
  - Assume item set $\mathcal{B} = \{B_1, ..., B_T\}$
  - e.g., a user bought $A, B_1, ..., B_t$ in a transaction
  - Any pair of items among $A, B_1, ..., B_t$ called co-bought

- Additionally, item profile
  - e.g., item price, categorical information etc.

Estimating posterior probability $P(B \mid A)$

- Aggregate historical transactions

```
A, B_1, cntAB1, profileB_1, ...
A, B_2, cntAB2, profileB_2, ...
          ............
A, B_T, cntABT, profileB_T, ...
```

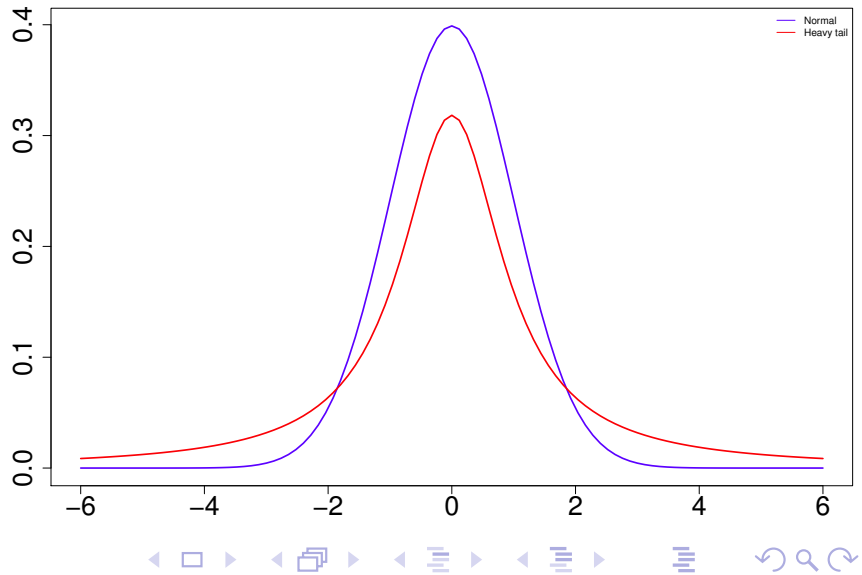
  - $cntAB_i$ = # times A and B_i bought together in a transaction
- *Logistic regression* produces $P(B \mid A)$ for all possible B

$$\log \frac{P(B \mid A)}{1 - P(B \mid A)} = \beta_0 + \beta_1 \cdot cntAB + \beta_2 \cdot profileB + ...$$

Challenges: data sparsity

- Typically with count or data resulted from aggregation
  - *Small or no record* associated for fairly large proportion of data
- A problem commonly encountered in practice
  - e.g., # times two items are bought together
  - e.g., # clicks of some web links or advertised items
  - e.g., total amount of sales of items

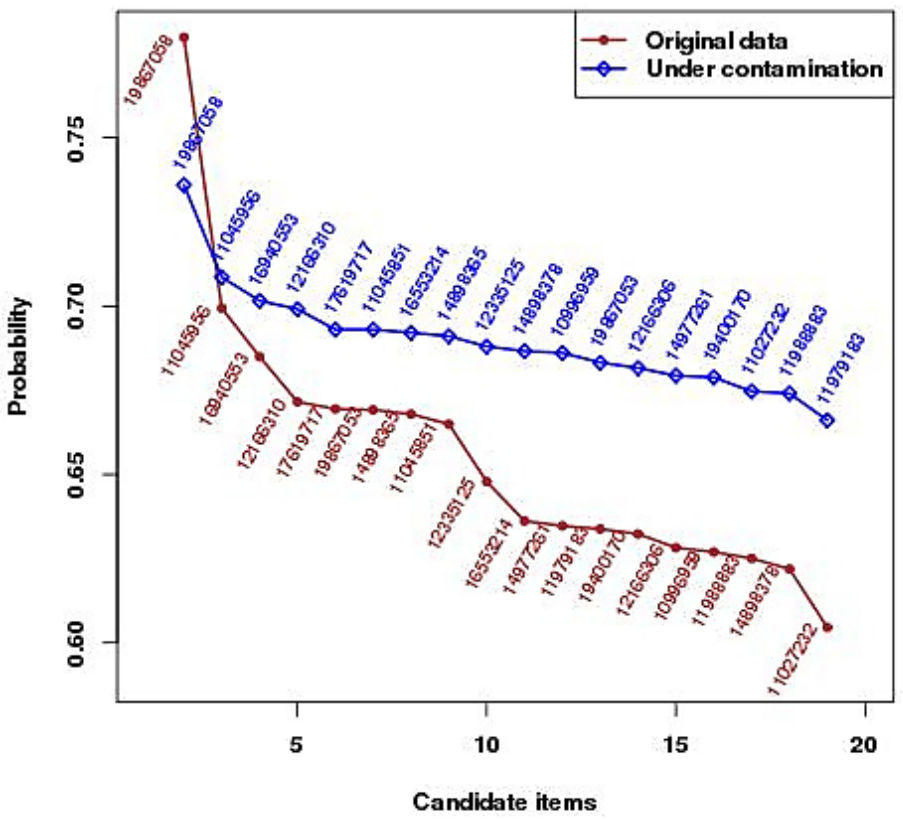
- Why does this happen so often?
  - Many quantities of interest have a *heavy-tailed distribution*
    - e.g., power law distribution
    - As a result of *aggregation* or *preferential attachment* (H. Simon 1955).



Potential solutions

- Aggregate quantities for similar items
  - e.g., use co-bought counts of the respective categories
  - e.g., sum up # clicks of similar items
  - Here 'similarity' may be obtained by other features or sementics
- Use a highly correlated surrogate variable
  - e.g., use # times two items put into the *same shopping cart* instead of co-bought counts
  - Much larger data available and may better reflect customer's intention
- Will those change the underlying model a lot?
  - *No* for predictive models if above properly implemented
  - *Data contamination results* by Yan, Chen et al (2011).

Influence of data contamination to Logit-based ranking

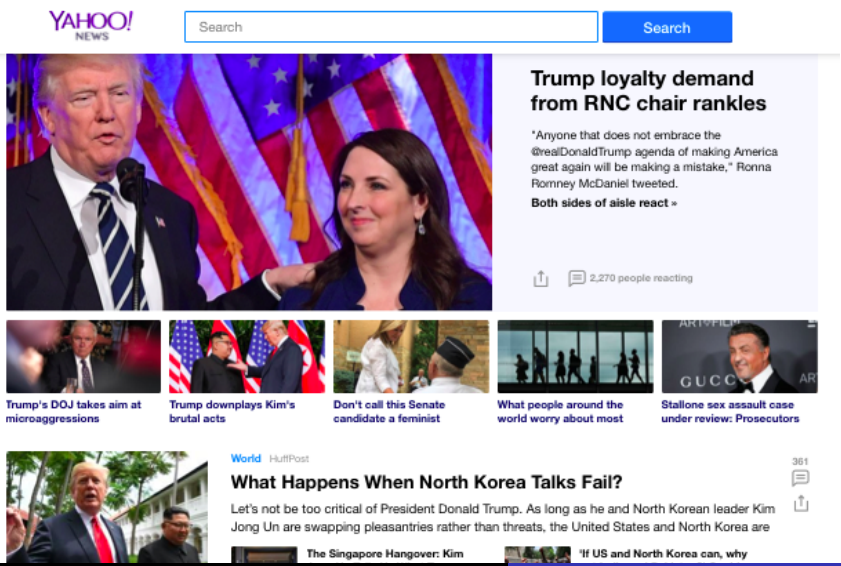


News recommendation

News recommendation

News recommendation

- Explosive growth of online contents
  - Too much for user to consume
  - Need to selectively show news to users
- To improve user experience
  - News contents match user's interests
  - Potentially attracts users to read more (user engagement).
- Relevant to all online news portals.

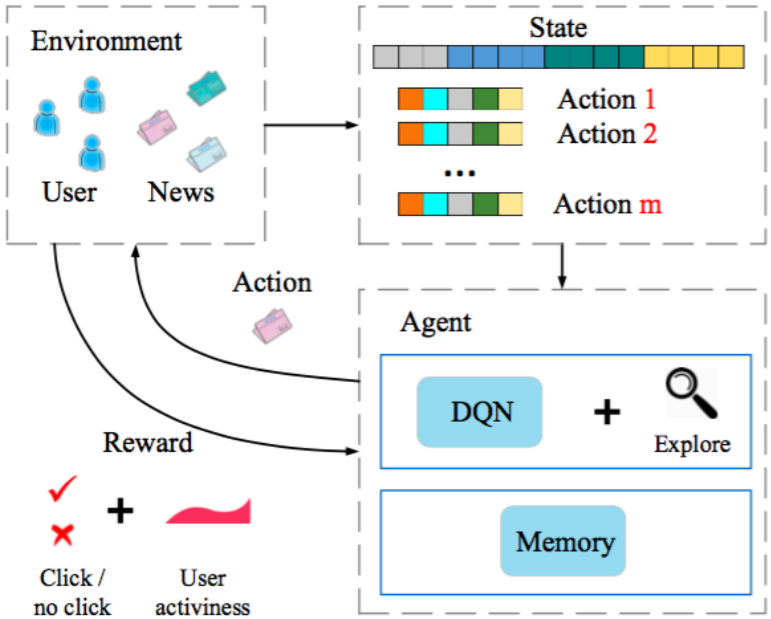


News recommendation

- Many existing approaches
  - News content-based
  - Collaborative filtering based
- Reinforcement learning based
  - An emerging trend
  - Allows to better capture dynamic nature of news features and user preference
  - Better models of user-news interaction
    - Explicitly incorporates user feedback

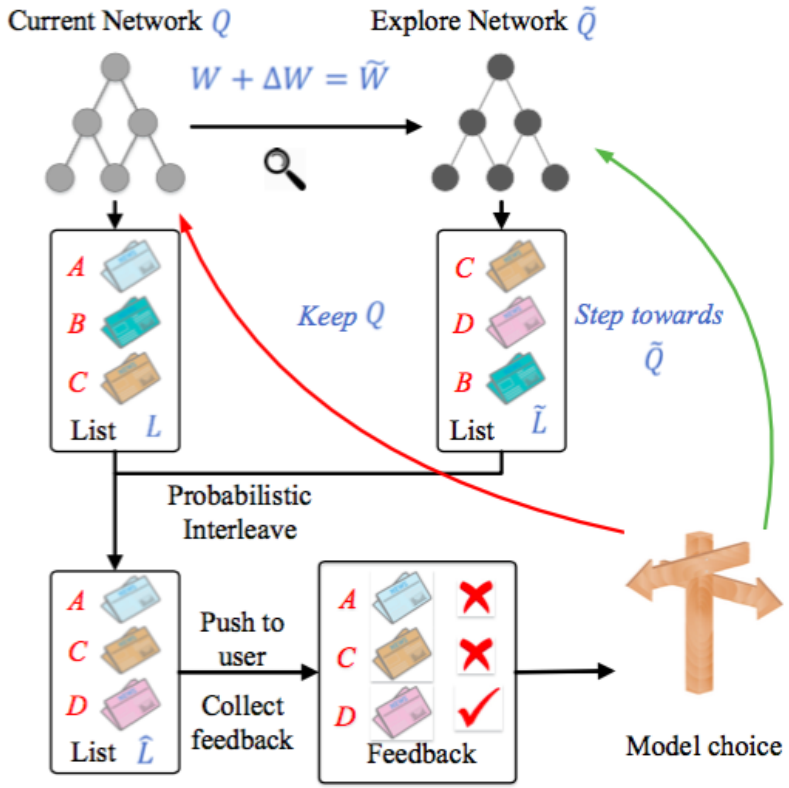
Reinforcement learning for news recommendation

- Zhang et al @WWW 2018
  - Capture the temporal changes of user preference
  - News clicks + user activeness as reward.



Zhang et al @WWW 2018

- Learning Q network



Summary

- Many applications of data science in e-commerce
- Enabled by advances in statistics and machine learning.

The end

Thank you!