

# Relational learning using bilinear models and its application in E-commerce

Shenghuo Zhu  
shenghuo.zhu@alibaba-inc.com

Alibaba Group

May 2, 2015

Collaborators:

Rong Jin, Qi Qian, Lijun Zhang, Mehrdad Mahdavi, Tianbao Yang

# Recommendation in e-commerce

今日团 | 女装 男装 鞋包 内衣 饰品 运动 美妆 童装 零食 母婴 百货 汽车 家电 数码 家装 家纺 [明日预告]

10点新团开抢

默认 销量↓ 价格↑

全部 今日新团开团 昨日开团

预告

搜索新团商品

Q

下一页 1/7

天弘基金  
天弘安康养老  
2014年度开放式混合金牛基金  
好的不止余额宝  
同风险产品中收益最高  
同收益产品中风险最低  
2014年收益率 15.2%

金牛基金 VPPi保本策略 2014年化15%

申购费率仅0.3% [多城团购] [天弘基金]天弘安康养老混...

¥1000

24217 件已售

PHILIPS  
新品首发  
抢鲜PHONE7  
免费试用60天  
首发价 1999  
好评返现500元

全网首发 三秒换芯 金牌服务

好评返500 抱团抢鲜六 [新品首发] 飞利浦家用高端直...

¥1999

1003 件已售

Microsoft Lumia640  
新品首发  
WIN10抢先体验 指日可待  
双卡双待双通双4G

天猫独家首发 首批支持WP10 赠价值399礼

抢优惠券 再返10元 [Microsoft/微软]Lumia 640 诺基亚 ...

¥1299

8 件已售

Proscenic  
浦桑尼克扫地机器人  
新品  
专业级仿人工  
拖地机器人  
GPS导航盒

专业级拖地机 室内GPS导航 清扫地图规划

悬赏挑战 晒单再返100 Proscenic 30JO专业级拖地机器...

¥1599

631 件已售

Midea  
1:1超低废水比  
废水减少66%  
已爆售687台  
免单疯抢中

废水减少66% 滤芯到期提醒 出水健康直饮

预定享免单 [Midea/美的]美的净水器家用直饮高端厨房...

¥1598

763 件已售

Microsoft Lumia640XL  
新品首发  
WIN10抢先体验 指日可待  
双卡双待双通双4G

天猫独家首发 首款支持WP10 赠价值399礼

领优惠券+再返10元 [Microsoft/微软] Lumia 640 XL 4G...

¥1899

4 件已售

# Personalized search in e-commerce

淘宝网  
tw.taobao.com

搜“新品春装”试试，新品春装



三 所有商品类目

所有分类 > 数码相机/单反相机/摄像机 > 该条件下查找

共 7,58 万件相关商品

筛选区

综合 销量 信用 价格 旺旺在线 天猫 折扣 发货地

< 1/100 >

相关分类

数码相机  
单反相机  
单电/微单  
摄像机  
镜头  
数码相机

更多



¥6199.00 销量 0

Casio/卡西欧 EX-TR550 美颜自拍神器 新款



¥198.00 销量 93

FUGU富富DC-F18数码相机 1800万像素 相机



¥298.00 销量 107

家用 拍照就“它”了 数码相机 1500万像素



¥298.00 销量 83

数码相机 2000万像素 8倍光学变焦 美颜神器



掌櫃熱賣



¥1088.00 ¥1208.80

效果超贊 正品30倍长焦高清数码相机SX510

销量 129



¥60.00 销量 2062



¥428.00 销量 1745



¥428.00 销量 1317



¥5999.01 销量 805



¥198.00 ¥247.60

# Recommendation to sellers

商品名称  商品ID   如果您商品数量多，可以直接

共找到符合聚划算商品机审条件的商品11个，不符合商品12个

商品	品牌	类目
 <p>90%白鸭绒初棉高端定制2014冬装新 女装羽绒服女中长款修身加厚</p>	初棉	羽绒服
 <p>初棉2015春装纯色套头高领打底毛衣 底衫女冬秋长袖针织衫女韩版</p>	初棉	毛衣
 <p>初棉牛仔衣2015新款春装女士百搭外 牛仔衬衫女长袖韩版牛仔衬衣</p>	初棉	衬衫
 <p>初棉2015春装新款文艺纯棉方领衬衫 长袖修身白色衬衣韩版女显瘦</p>	初棉	衬衫

From Taobao's seller interface.

# Relation and ranking

In the above tasks, we consider the relationship between

- buyer and item
- buyer/query and item
- seller and item
- seller and buyer
- ...

and rank items (or buyers) conditioned on a buyer (or a seller, a query-buyer pair).

# Relation as function

- Buyer and item:  $u$  buyer,  $v$  item,
  - ▶ scoring function:  $y(v; u)$
- Ranking by scores:
  - ▶  $y(v_i; u) > y(v_j; u) \implies u$  prefers  $v_i$  over  $v_j$ .

# Ranking by segmentation

- Assume that given  $K$  underlying user segment, users,  $u$ , belonging to segment  $k$  share a same scoring function:

$$y(v; u) = g_k(v)$$

- User  $u$  bought item  $w$ . Let all users that bought item  $w$  be segment  $k$ .  $g_k(v)$  be the preference scores (purchase history) of item  $v$  in segment  $k$ . It is item-based collaborative filtering.

# Ranking using mixture

- User  $u$  bought more than one item. The above strict segmentation assumption is relaxed. It is usually considered to use similarity between users.
- In a general term, scoring function of  $u$  is a linear combination of  $g_k$ :

$$y(v; u) = \sum_{k=1}^K \beta_k g_k(v),$$

where segmentation is latent.



# Ranking using mixture

- User  $u$  bought more than one item. The above strict segmentation assumption is relaxed. It is usually considered to use similarity between users.
- In a general term, scoring function of  $u$  is a linear combination of  $g_k$ :

$$y(v; u) = \sum_{k=1}^K \beta_k g_k(v),$$

where segmentation is latent.

# Ranking as matrix factorization

- For each user  $u$ ,

$$y(v; u) = \sum_k \beta_k g_k(v)$$

- As each user has its own  $\beta_k = f_k(u)$

$$y(v; u) = \sum_k f_k(u) g_k(v)$$

- Put  $y(u, v)$ ,  $f_k(u)$  and  $g_k(v)$  as matrices

$$\mathbf{Y} = \mathbf{GF}^T \approx \mathbf{T}$$

- Convex: use low rank constraint of  $Y$ , [YLZG09].

# Ranking as matrix factorization

- For each user  $u$ ,

$$y(v; u) = \sum_k \beta_k g_k(v)$$

- As each user has its own  $\beta_k = f_k(u)$

$$y(v; u) = \sum_k f_k(u) g_k(v)$$

- Put  $y(u, v)$ ,  $f_k(u)$  and  $g_k(v)$  as matrices

$$Y = GF^T \approx T$$

- Convex: use low rank constraint of  $Y$ , [YLZG09].

# Ranking as matrix factorization

- For each user  $u$ ,

$$y(v; u) = \sum_k \beta_k g_k(v)$$

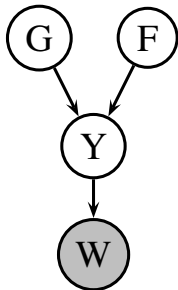
- As each user has its own  $\beta_k = f_k(u)$

$$y(v; u) = \sum_k f_k(u) g_k(v)$$

- Put  $y(u, v)$ ,  $f_k(u)$  and  $g_k(v)$  as matrices

$$\mathbf{Y} = \mathbf{GF}^T \approx \mathbf{T}$$

- Convex: use low rank constraint of  $Y$ , [YLZG09].



# Ranking as matrix factorization

- For each user  $u$ ,

$$y(v; u) = \sum_k \beta_k g_k(v)$$

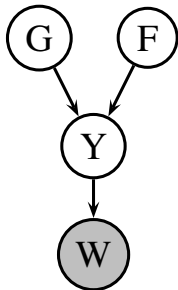
- As each user has its own  $\beta_k = f_k(u)$

$$y(v; u) = \sum_k f_k(u) g_k(v)$$

- Put  $y(u, v)$ ,  $f_k(u)$  and  $g_k(v)$  as matrices

$$\mathbf{Y} = \mathbf{GF}^T \approx \mathbf{T}$$

- Convex: use low rank constraint of  $Y$ , [YLZG09].



# Ranking in bilinear model

- User feature  $x_u$ , and item feature  $z_v$ .

$$f_k(u) = \langle a_k, x_u \rangle, \quad g_k(v) = \langle b_k, z_v \rangle$$

$$y(v; u) = \sum_k f_k(u)g_k(v) = \langle x_u, Wz_v \rangle$$

where  $W = \sum_k a_k b_k^\top$ .

- Put  $y(v; u)$ ,  $f_k(u)$  and  $g_k(v)$  as matrices

$$Y = X^\top W Z$$

# Ranking in bilinear model

- User feature  $x_u$ , and item feature  $z_v$ .

$$f_k(u) = \langle a_k, x_u \rangle, \quad g_k(v) = \langle b_k, z_v \rangle$$

$$y(v; u) = \sum_k f_k(u)g_k(v) = \langle x_u, Wz_v \rangle$$

where  $W = \sum_k a_k b_k^\top$ .

- Put  $y(v; u)$ ,  $f_k(u)$  and  $g_k(v)$  as matrices

$$\mathbf{Y} = \mathbf{X}^\top \mathbf{W} \mathbf{Z}$$

- How to control the complexity of learning space?
  - ▶ Rank of  $W$ , or nuclear norm  $\|W\|_*$ .
- When features have high dimensions, can we take the advantage of low complexity of  $W$  to reduce the computational complexity?
  - ▶ The model is essentially a linear model:

$$y \equiv \text{vec}(\mathbf{Y}) = (Z \otimes X)^\top \text{vec}(W) \equiv x^\top w.$$

- ▶ A projection approach of linear model is presented in this talk.



- How to control the complexity of learning space?
  - ▶ Rank of  $W$ , or nuclear norm  $\|W\|_*$ .
- When features have high dimensions, can we take the advantage of low complexity of  $W$  to reduce the computational complexity?
  - ▶ The model is essentially a linear model:

$$y \equiv \text{vec}(\mathbf{Y}) = (Z \otimes X)^\top \text{vec}(W) \equiv x^\top w.$$

- ▶ A projection approach of linear model is presented in this talk.

# Solve via random projection

- To solve:

$$w = \arg \min_w \frac{\lambda}{2} \|w\|^2 + \sum_i \ell(x_i^\top w, y_i).$$

- Approach

- ▶ Generate a random projection  $R$  of rank  $m$ , and let  $\tilde{x}_i = Rx_i$ .
- ▶ Solve:

$$v = \arg \min_v \frac{\lambda}{2} \|v\|^2 + \sum_i \ell(\tilde{x}_i^\top v, y_i).$$

- ▶ Recover:  $\hat{w} = R^\top v$ .

# Solve via random projection

- To solve:

$$w = \arg \min_w \frac{\lambda}{2} \|w\|^2 + \sum_i \ell(x_i^\top w, y_i).$$

- Approach

- ▶ Generate a random projection  $R$  of rank  $m$ , and let  $\tilde{x}_i = Rx_i$ .
- ▶ Solve:

$$v = \arg \min_v \frac{\lambda}{2} \|v\|^2 + \sum_i \ell(\tilde{x}_i^\top v, y_i).$$

- ▶ Recover:  $\hat{w} = R^\top v$ .

- $\hat{w}$  is limited in the subspace spanned by  $R$ , as  $\hat{w} = R^T v$ .

## Theorem 3 of [ZMJ<sup>+</sup>13]

For any  $0 < \varepsilon \leq 1/2$ , with a probability  $1 - \exp(-(d-r)/32) - \exp(-m/32) - \delta$ , we have

$$\|\hat{w} - w_*\|_2 \geq \frac{1}{2} \sqrt{\frac{d-r}{m}} \left( 1 - \frac{\varepsilon \sqrt{2(1+\varepsilon)}}{1-\varepsilon} \right) \|w_*\|_2.$$

# Dual space

- Dual variable and function

$$\ell_*(\alpha, y_i) = \sup_{\xi} \{\alpha_i \xi - \ell(\xi, y_i)\}$$

- Dual problem

$$\alpha = \arg \min_{\alpha} \frac{1}{2\lambda} \alpha^\top X^\top X \alpha + \sum_i \ell_*(\alpha_i, y_i).$$

- Dual problem after random projection

$$\hat{\alpha} = \arg \min_{\alpha} \frac{1}{2\lambda} \alpha^\top X^\top R^\top R X \alpha + \sum_i \ell_*(\alpha_i, y_i).$$

## Proposition

For any  $0 < \varepsilon \leq 1/2$ , with a probability at least  $1 - \delta$ , we have

$$\|\tilde{\alpha} - \alpha_*\|_K \leq \frac{\varepsilon}{1 - \varepsilon} \|\alpha_*\|_K,$$

provided  $m = \Omega(\varepsilon^{-2} \log \delta^{-1})$ .

- Generate a random projection  $R$  of rank  $m$ , and let  $\tilde{x}_i = Rx_i$ .
- Solve:

$$\arg \min_v \frac{\lambda}{2} \|v\|^2 + \sum_i \ell(\tilde{x}_i^\top v, y_i).$$

- Obtain dual variables:  $\tilde{\alpha}_i = \ell'(\tilde{x}_i^\top v, y_i)$
- Recover primal solution:  $\tilde{w} = -\frac{1}{\lambda} \sum_i \tilde{\alpha}_i x_i$ .

- Generate a random projection  $R$  of rank  $m$ , and let  $\tilde{x}_i = Rx_i$ .
- Solve:

$$\arg \min_v \frac{\lambda}{2} \|v\|^2 + \sum_i \ell(\tilde{x}_i^\top v, y_i).$$

- Obtain dual variables:  $\tilde{\alpha}_i = \ell'(\tilde{x}_i^\top v, y_i)$
- Recover primal solution:  $\tilde{w} = -\frac{1}{\lambda} \sum_i \tilde{\alpha}_i x_i$ .



# Approximation error of DuRP

## Theorem 2 of [ZMJ<sup>+</sup>13]

For any  $0 < \varepsilon \leq 1/2$ , with a probability at least  $1 - \delta$ , we have

$$\|\tilde{w} - w_*\|_2 \leq \frac{\varepsilon}{1 - \varepsilon} \|w_*\|_2,$$

provided  $m \geq \frac{(r+1) \log(2r/\delta)}{c\varepsilon^2}$ .

## Dual variables in bilinear model

- Dual variables in  $\alpha$  can be reshaped to a matrix  $A$ , where the nonzero entries correspond to the user-item pairs having interaction.
- Then the recovered matrix is written as

$$W = XAZ^T.$$

- Very high dimension in its linear representation, i.e.  $Z \otimes X$ 
  - ▶ Random projection:  $R = R_2 \otimes R_1$ . [QJZL13].
- Recovered matrix  $W = XAZ^T$  is of high dimension and usually dense, thus is difficult to apply to online service.
  - ▶ Approximated by multiplication of two low rank matrices, using approximate SVD [HMT11].

- Many applications using relational data
- A bilinear model is a straightforward approach
- To learn from massive data, dual recovery random projection.



N. Halko, P. Martinsson, and J. Tropp.

Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011.



Qi Qian, Rong Jin, Shenghuo Zhu, and Yuanqing Lin.

An integrated framework for high dimensional distance metric learning and its application to fine-grained visual categorization.

Technical Report 2013-TR102, NEC Laboratories America, 2013.  
arXiv:1402.0453.



Kai Yu, John Lafferty, Shenghuo Zhu, and Yihong Gong.

Large-scale collaborative prediction using a nonparametric random effects model.

In *ICML'09: Proceedings of the 26th Annual International Conference on Machine Learning*, pages 1185–1192, New York, NY, USA, 2009. ACM.



Lijun Zhang, Mehrdad Mahdavi, Rong Jin, Tianbao Yang, and Shenghuo Zhu.

Recovering optimal solution by dual random projection.

In *COLT'13: The 26th Annual Conference on Learning Theory*, 2013.