

2017 Machine Learning Summit

人工智能在Etsy电商平台中的技术实践

洪亮劼

美国电子商务平台 Etsy 数据科学主管、前雅虎研究院高级研发经理

洪亮劫

- **Etsy数据科学主管** (2016年 – 至今)
 - 搜索与发掘
 - 个性化与推荐系统
 - 计算广告
 - **雅虎研究院高级研发经理** (2013年 – 2016年)
 - 搜索与个性化科学
-
- 在国际顶级机器学习和数据挖掘会议**SIGIR, WWW, KDD, CIKM, AAAI, WSDM, RecSys** 和**ICML**等上发表多篇论文。
 - **3**篇国际会议最佳论文、**2000**多论文引用量，以及**H-Index**达到**17**。
 - 在国际顶级机器学习和数据挖掘会议**KDD, WWW, SIGIR, WSDM, AAAI, EMNLP, ICWSM, ACL, CIKM, IJCAI** 任程序委员会成员，担任多个国际顶级期刊审稿人。

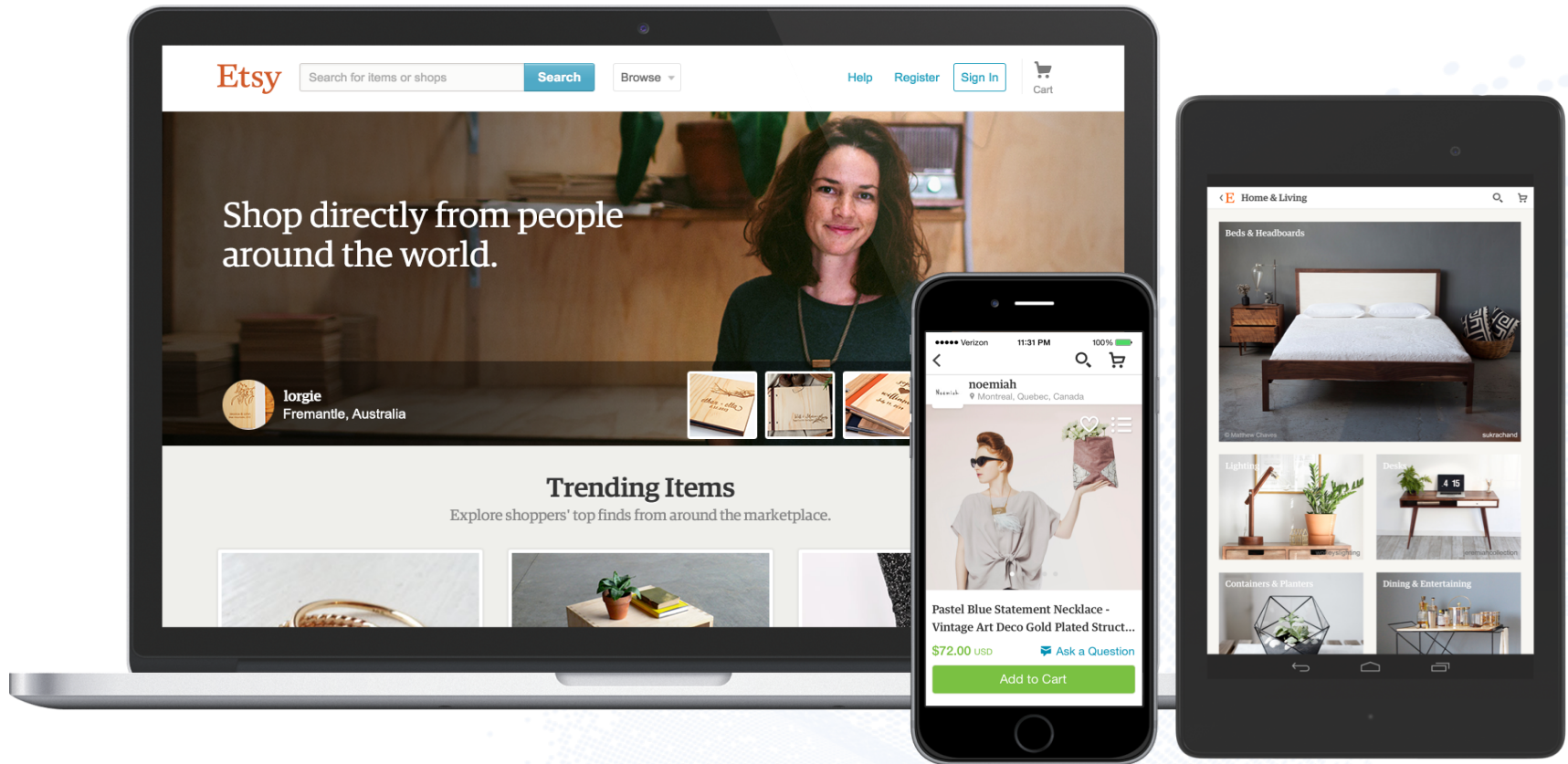
人工智能在Etsy电商平台的技术实践

- Etsy，一个服务全球中小手工业者的电商平台
- Etsy的人工智能挑战
 - 基于深度学习的搜索方案（KDD 2016）
 - 基于大型概率图模型的个性化推荐方案（KDD 2014）
 - 基于大型集成学习的计算广告平台方案（AdKDD 2017/KDD 2017）



Etsy – 全球化的电商平台

Etsy是一个联结全球创意手工者的电商平台



Etsy – 全球化的电商平台



Artifact Bags
Omaha, NE

Photo by: Dana Damewood and Jackie Sterba



Clap Clap
Los Angeles, CA

Photo by: Bert Youn and Mimi Kim



redravenstudios
Pittsburgh, PA

Photo by: Janelle Bedycki



Little Hero Capes
Somerset, MA

Photo by: Rich Vintage Photography



Cattails Woodwork
Hermitage, PE, Canada

Photo by: Cattails Woodwork



Room for Emptiness
Berlin, Germany

Photo by: Room for Emptiness



sukrachand
Brooklyn, NY

Photo by: sukrachand



Nicole Porter Design
Saint Paul, MN

Photo by: Nicole Porter Design



noemiah
Montreal, QC, Canada

Photo by: noemiah



Lorgie
Fremantle, WA, Australia

Photo by: Lorgie



Jeremiah Collection
San Francisco, CA

Photo by: Matthew Reamer



Docksmith
Brunswick, ME

Photo by: Docksmith



purlBKnit
Brooklyn, NY

Photo by: purlBKnit



Julia Astreou
Nicosia, Cyprus

Photo by: Panagiotis Mina



Moira K. Lime
Omaha, NE

Photo by: Moira K. Lime



Nested Yellow
Portland, OR

Photo by: Jessica Dremov and Nested Yellow



Habitables
Madrid, Spain

Photo by: Habitables



Woodstoring
Kaunas, Lithuania

Photo by: Iлона & Martynas from Instudija



karoArt
Dublin, Ireland

Photo by: Christine Burns



ADIKILAV
Jerusalem, Israel

Photo by: Shlomit Koslowe



My A La Mode Boutique
Ecuador

Photo by: My A La Mode Boutique

| Etsy – 全球化的电商平台

200万
活跃卖家

3000万
活跃买家

\$28.4亿美元
年销售额

4500万
件独特的商品



| Etsy – 全球化的电商平台

200万
活跃卖家

86%
卖家是女性

95%
卖家从家办公

76%
的卖家认为他们的网店是一门生意



| Etsy – 全球化的电商平台

30%

的卖家把其创新中小手工业当做唯一职业

65%

的卖家开始把Etsy的收入当做重要收入来源

79%

的卖家把Etsy的网店当做他们的门户展示平台



| Etsy – 全球化的电商平台

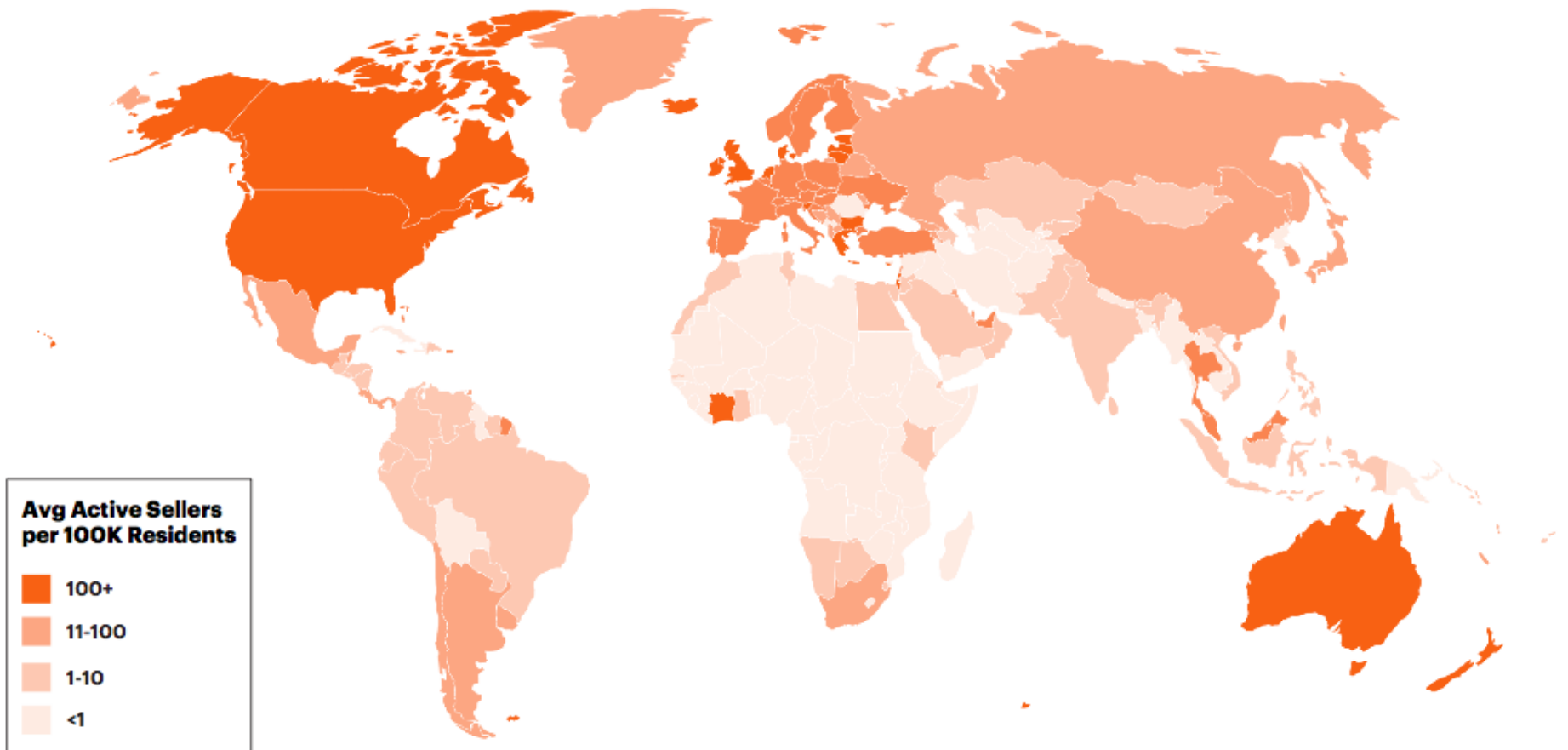
2500万
活跃买家

87%
的买家是女性

92%
的买家认为Etsy能够提供独特的在其他地方找不到的商品



Etsy – 全球化的电商平台



Etsy的人工智能挑战与机遇

对买家而言

- 如何让买家在上百万的商品中找到满意的、独特的商品
- 如何引导并且启发买家去挖掘他们原本不知道的商品
- 如何根据不同的场景（比如婚礼、乔迁、毕业等）给买家推荐合适的商品

对卖家而言

- 如何能够接触到更加广泛的潜在买家
- 如何有效地运营自己的广告策略
- 如何通过不同的渠道和潜在买家进行交流和沟通

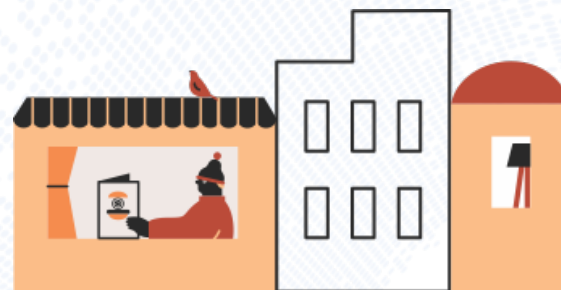
对平台而言

- 如何能够构建一个健康平衡的环境
- 如何能够加快卖家买家的交互



Etsy的人工智能挑战与机遇

- **搜索与发掘**
 - 查询建模 (Query Modeling)
 - 用户意图建模 (User Intent Modeling)
 - 排序 (Learning to Rank)
- **个性化推荐**
 - 用户建模 (User Profiling)
 - 物品建模 (Item Modeling)
 - 推荐排序 (Recommender Ranking)
- **计算广告**
 - 点击率建模 (Click-Through Rate Modeling)
 - 转化率建模 (Conversion Rate Modeling)
 - 出价策略优化 (Bid Optimization)



Etsy - 搜索与发掘

Etsy [Sell on Etsy](#) [Register](#)

July 4th [Clothing & Accessories](#) [Jewelry](#) [Craft Supplies & Tools](#) [Weddings](#) [Entertainment](#) [Home & Living](#) [Kids & Baby](#) [Vintage](#)

All categories






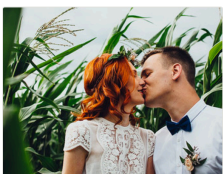


- Clothing
- Weddings
- Craft Supplies & Tools
- Accessories
- Jewelry
- Paper & Party Supplies
- Bath & Beauty
- Home & Living
- Shoes
- Art & Collectibles
- Toys & Games
- Pet Supplies
- Bags & Purses
- Books, Movies & Music
- Electronics & Accessories

Shop location

- Anywhere
- China
- Custom

All categories > "wedding dress" (232,752 Results)







Sort by: Relevancy

 <p>Audrey Hepburn Retro Tea Length Wedding Dress ieie \$299.95</p>	 <p>Sexy Open Back Wedding Dress, Lace Wedding Dress LaceMarry \$314.00</p>	 <p>Wedding Suspender, Brown Leather Suspender Heysir \$23.00</p>	 <p>Rush Order ROSE GOLD Wedding Hanger LynnClaire \$26.00</p>
 <p>Detachable lace skirt for wedding dress, ... mimetik \$119.00</p>	 <p>Dress FW14-15 Wedding dress Boho ... BabyDollShopRu \$280.00</p>	 <p>Boho wedding dress "Moonlight" BridalGardenStudio \$330.00</p>	 <p>Exquisite Lace Wedding Dress V Shape ... LaceBridal \$259.00</p>

中国联通 LTE 10:01 99%

< Filter

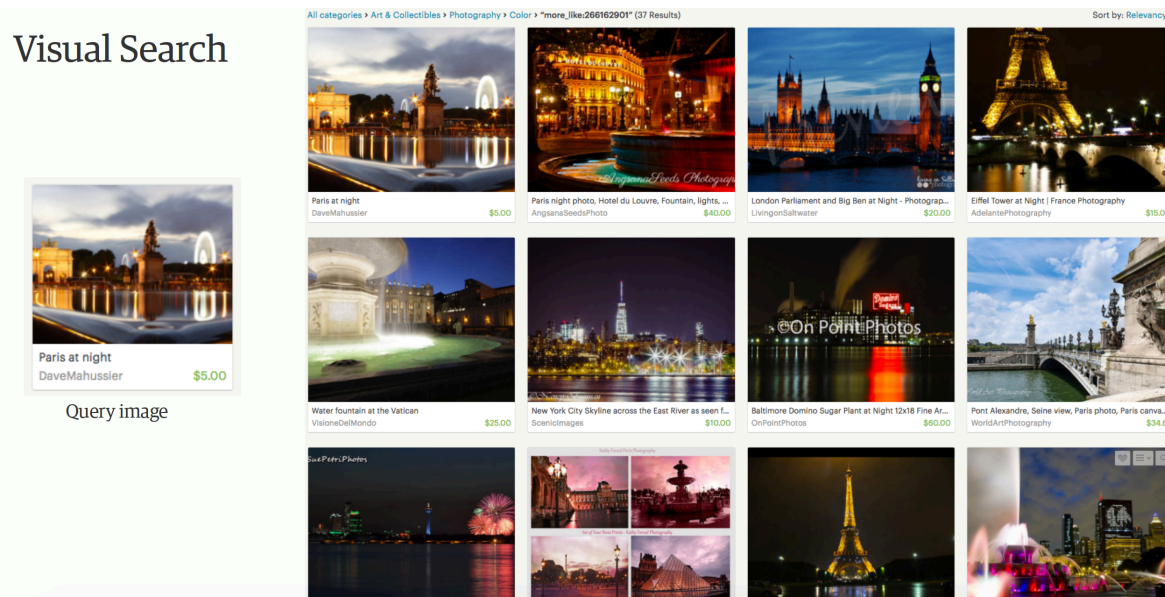
1,719,287 items

 <p>Ad Groomsmen Gift Box ... ScissorMill \$35.00</p>	 <p>Ad Navy and Gold Weddi ... starboardpr... \$70.00</p>
 <p>BEST SELLER Unique wedding gift - ... HelloAm \$45.00</p>	 <p>Wedding Gift, Gift for ... HappyYouH... \$43.50</p>
 <p>Heart-shaped map wedding gift Jennifer & Mathew 10.08.2014</p>	 <p>Family tree wedding gift</p>

Home Favorites Search You Cart

基于深度图像特征的大规模多模搜索排序算法

- Corey Lynch, Kamelia Aryafar, and Josh Attenberg. **Images Don't Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank.** In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD '16**). ACM, New York, NY, USA, 541-548.



Etsy – 基于深度图像特征的搜索排序

- > **1.42亿**张图片
- 对文字（标题、描述）信息的**重要补充**
- 当**没有文字信息或者文字信息不准确**的时候，图像信息就成为了**重要来源**



“Red Short dress, Prom Dress, **wedding dress**, dress, ...”



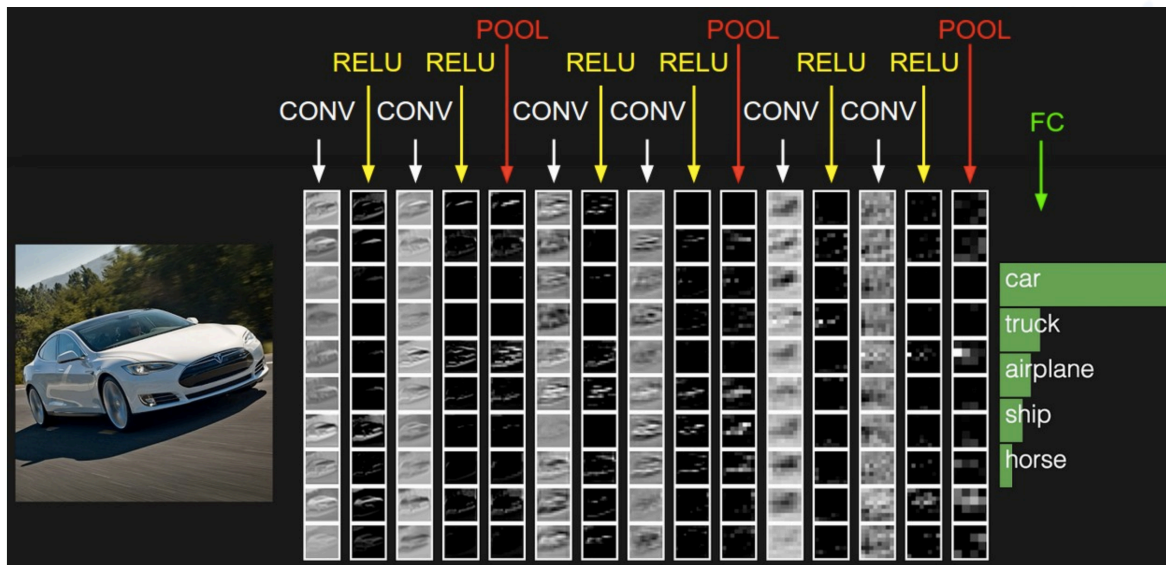
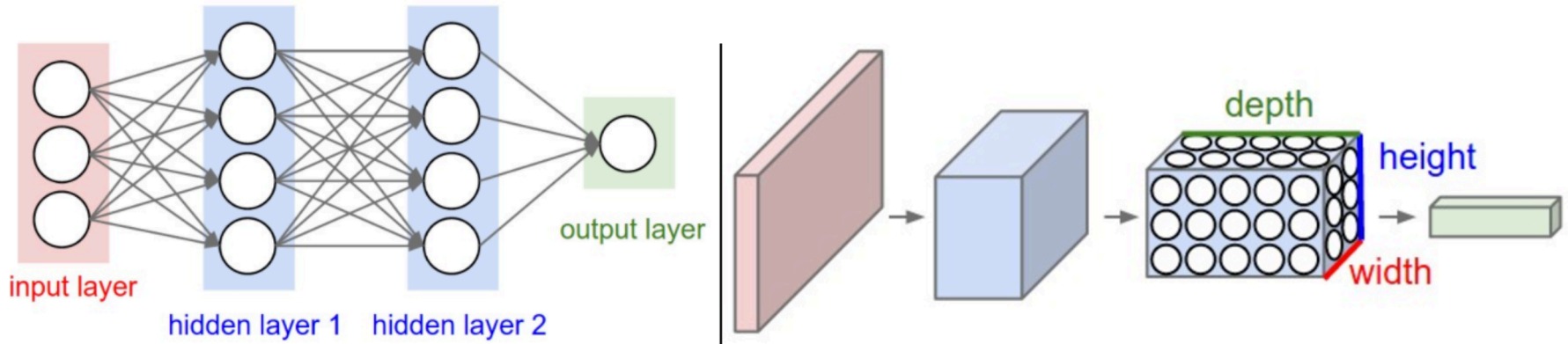
“Pocket Knife wedding shower ideas **wedding dresses**, beach ...”



“Yellow dress. Retro dress **Wedding dress**. Flared skirt...”

Etsy – 基于深度图像特征的搜索排序

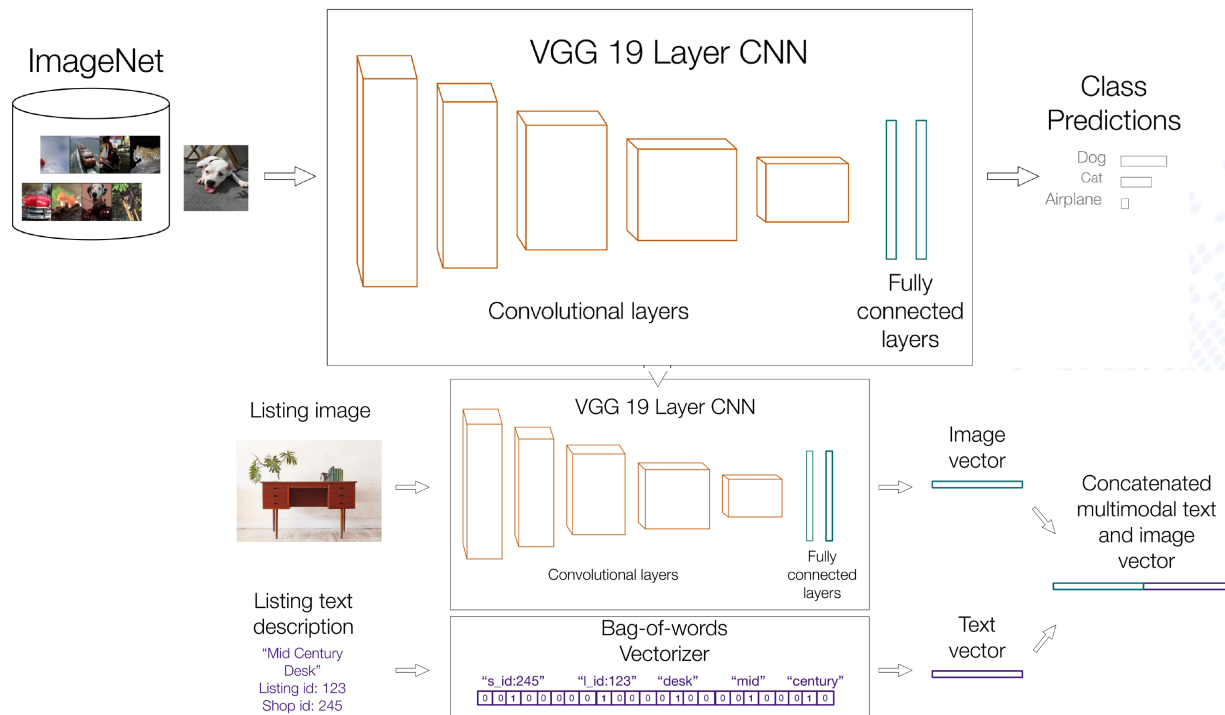
深度图像特征提取



Etsy – 基于深度图像特征搜索排序

深度图像特征提取 + 迁移学习

- 从ImageNet上学习深度CNN模型
- 提取出训练得到的CNN模型的参数
- 在Etsy数据上使用学出的模型参数，以及结合文本信息来训练排序算法



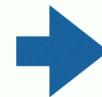
Etsy – 基于深度图像特征的搜索排序

多模排序算法训练

Etsy mid century desk Search

Item	Seller	Price
Mid Century Desk / Entryway Table	OrWaDesigns	\$475.00 USD
Mid Century Desk 48" or 60" Minim...	FlintAlleyFurniture	\$565.00 USD
French vintage mid century desk t...	lestrictmaximum	\$55.09 USD
mid century modern secretary des...	dsartereno	\$799.00 USD
Red Gooseneck Mid Century Desk ...	OldOakVintage	\$28.11 USD
Mid Century Desk from the Lane A...	owlsongvintage	\$1,475.00 USD

The image shows a search results page for 'mid century desk' on Etsy. A blue arrow points from the search results to a zoomed-in view of two items. The first item, a red gooseneck desk lamp, is enclosed in a red border and labeled 'ignored'. The second item, a wooden desk, is enclosed in a green border and labeled 'clicked'.

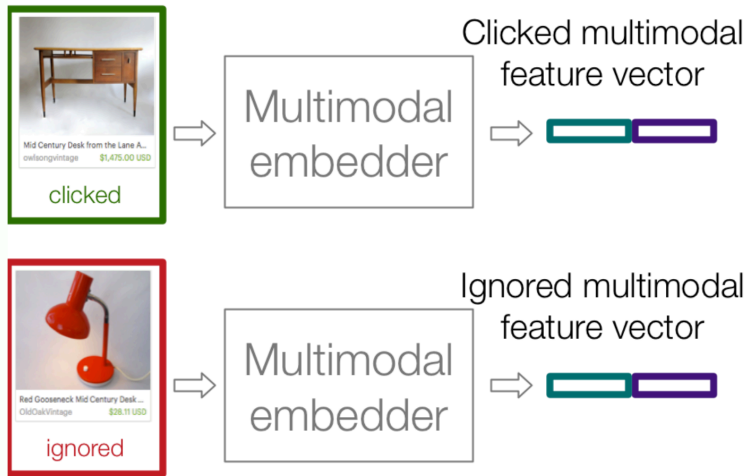


Red Gooseneck Mid Century Desk ...
OldOakVintage \$28.11 USD
ignored

Mid Century Desk from the Lane A...
owlsongvintage \$1,475.00 USD
clicked

Etsy – 基于深度图像特征搜索排序

多模排序算法训练



Algorithm 1 Multimodal Embedding of Listings

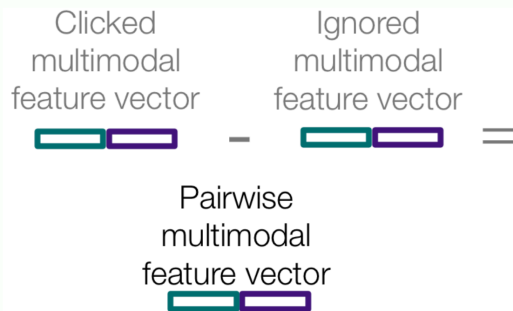
```

1: procedure EMBEDMULTIMODAL( $d_i$ )
2:    $d_{T_i} \leftarrow \text{BoW}(\text{text})$ 
3:    $d_{I_i} \leftarrow \text{VGG}(\text{image})$ 
4:    $d_{MM_i} \leftarrow [d_{T_i}, d_{I_i}]$ 
5:   return  $d_{MM_i}$ 
  
```

Algorithm 2 Generate Pairwise Classification Instances

```

1: procedure GETPAIRWISEINSTANCES( $\{\{d_i^+, d_i^-\}\}$ )
2:    $L \leftarrow \{\}$ 
3:   for  $i = 1 \dots |P|$  do  $\triangleright |P|$  labeled tuples
4:      $d_{MM_i}^+ \leftarrow \text{EmbedMultimodal}(d_i^+)$ 
5:      $d_{MM_i}^- \leftarrow \text{EmbedMultimodal}(d_i^-)$ 
6:     Draw  $r$  uniformly at random from  $[0, 1)$ 
7:     if  $r > 0.5$  then
8:        $x_i \leftarrow d_{MM_i}^+ - d_{MM_i}^-$ 
9:        $y_i \leftarrow +1$ 
10:    else
11:       $x_i \leftarrow d_{MM_i}^- - d_{MM_i}^+$ 
12:       $y_i \leftarrow -1$ 
13:     $L = L.append(\langle x_i, y_i \rangle)$ 
14:    return  $L$   $\triangleright$  The list of classification instances.
  
```



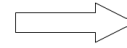
Etsy – 基于深度图像特征的搜索排序

多模排序算法训练

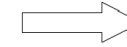
Text-only ranking for
“bar necklace”

Multimodal ranking for
“bar necklace”

90th
percentile



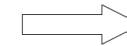
80th
percentile



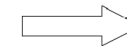
70th
percentile



60th
percentile

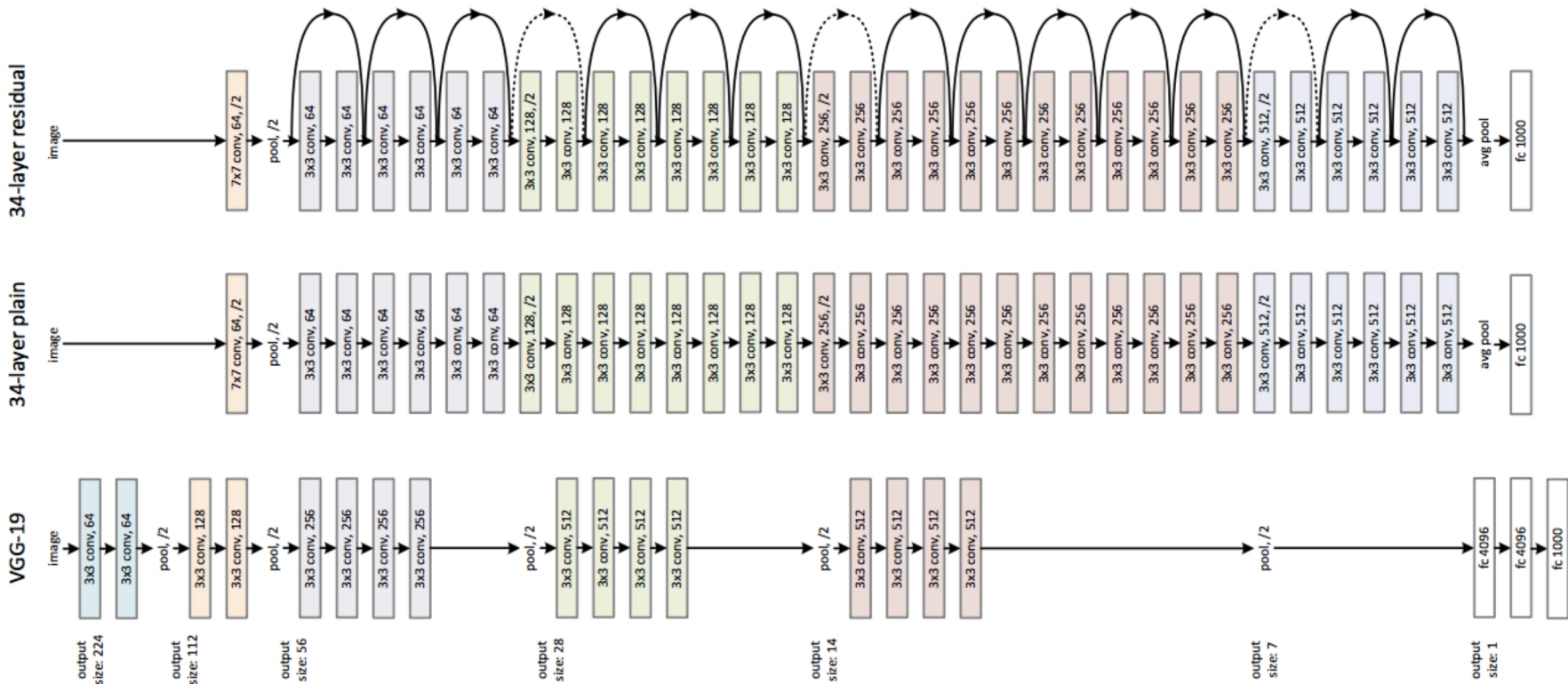


50th
percentile



Etsy – 基于深度图像特征的搜索排序

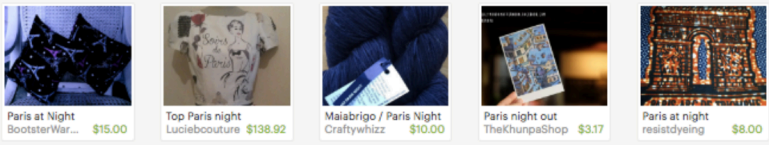
多模排序算法训练



Etsy – 基于深度图像特征的搜索排序

基于视觉信息的相似推荐

Similar Items [Shop More](#)



Paris at Night
BoosterWar... \$15.00


Top Paris night
Luciebouture \$138.92

Malabrigo / Paris Night
Craftywhizz \$10.00

Paris night out
TheKhunpaShop \$3.17

Paris at night
resistdyeing \$8.00


DaveMahussier
Favorite shop



5 items

Favorite Like this item?
Add it to your favorites to revisit it later.

Instant download



Q zoom

Item details Shop policies

Taken outside of the Louvre in Paris, France.

Paris at night
\$5.00 [Ask a question](#)

Overview

- Handmade item
- Instant Digital Download: 1 JPG included
- Made to order
- Favorited by: 1 person

This shop accepts Etsy gift cards

Add to cart

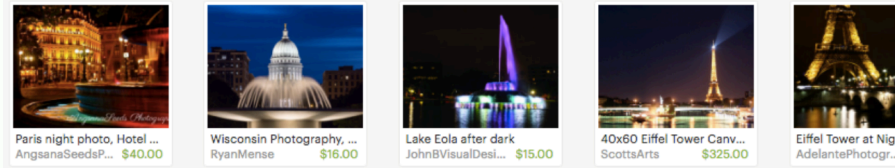
Favorite Add to -

Tweet Pin it tumblr. + Like 0

DaveMahussier
in United States

Control

Similar Items [Shop More](#)



Paris night photo, Hotel ...
AngsanaSeedsP... \$40.00

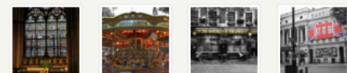
Wisconsin Photography, ...
RyanMense \$16.00

Lake Eola after dark
JohnBVisualDesi... \$15.00

40x60 Eiffel Tower Carv...
ScottsArts \$325.00

Eiffel Tower at Night
AdelantePhotogr...


DaveMahussier
Favorite shop



5 items

Favorite Like this item?
Add it to your favorites to revisit it later.

Instant download



Q zoom

Item details Shop policies

Paris at night
\$5.00 [Ask a question](#)

Overview

- Handmade item
- Instant Digital Download: 1 JPG included
- Made to order
- Favorited by: 1 person

This shop accepts Etsy gift cards

Add to cart

Favorite Add to -

Tweet Pin it tumblr. + Like 0

Variant

Etsy – 个性化推荐



Curated by Etsy, Based on **Your History**.

Our Picks for You

Search available items



Portrait Art of French Composer Ge...
ArtyMargit **\$26.34**



Map of Far E...
HunnapPrint



SALE: Violin Fine Art Photography M...
KEnczPhotography **\$20.00**



Original 189...
BooksMapsa



zoom

Item details

★★★★★ (53)

Shipping & Policies

Hand-drawn print of French composer Georges Bizet. Signed and dated.

Click on the photo to see others in this series: Brahms, Pachelbel and Vivaldi....

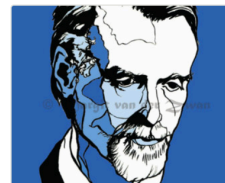
This limited edition print is printed on fine art 280gsm paper. Makes the perfect present for young or old. I take every care to package my artwork securely, to ensure it arrives in perfect condition.

Title: "Bizet" #0776
Size: A4 (21 x 30cm)
Comes unframed
limited edition: 100 prints

The complete number of limited edition French composer artwork will be...



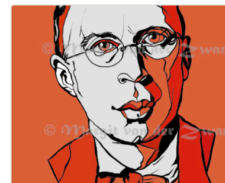
ArtyMargit
in Manchester, United Kingdom



Limited edition Print of Zolta...
\$26.34



Khachaturian Composer port...
\$26.34



Prokofiev Portrait Print Limite...
\$26.34

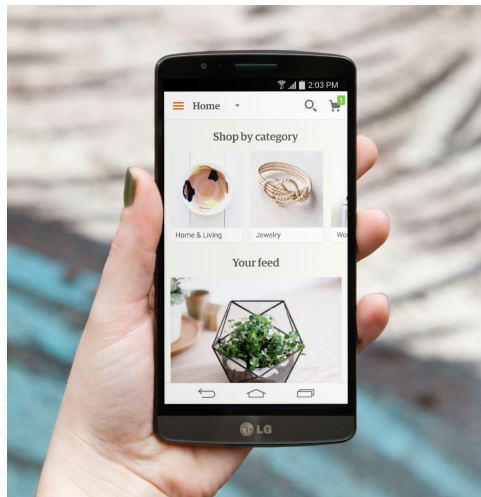


Music Art Print of Composer ...
\$26.34

Etsy – 个性化推荐

基于概率图模型的大规模个性化推荐系统

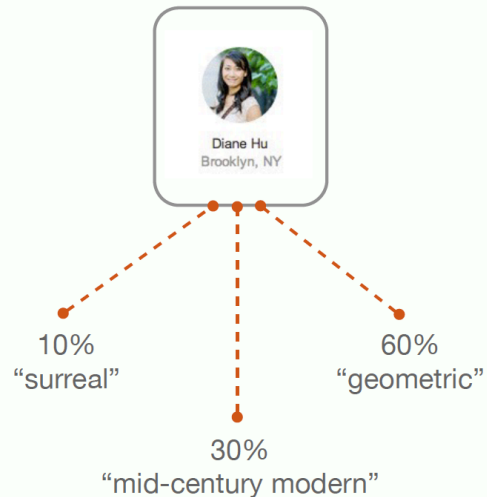
- Diane J. Hu, Rob Hall, and Josh Attenberg. **Style in the long tail: Discovering Unique Interests with Latent Variable Models in Large Scale Social E-Commerce.** In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (**KDD '14**). ACM, New York, NY, USA, 1640-1649.
- **KDD 2014工业界最佳论文**



Etsy – 个性化推荐

Latent Dirichlet Allocation

Learn **style profiles** for each user using LDA



1

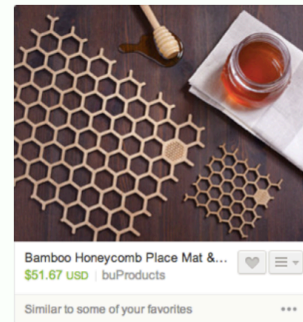
Define what each style looks like



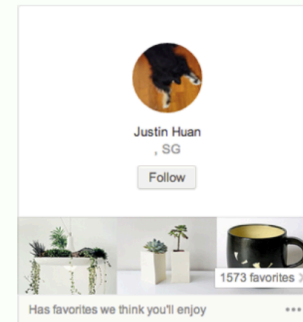
= "mid-century modern"

2

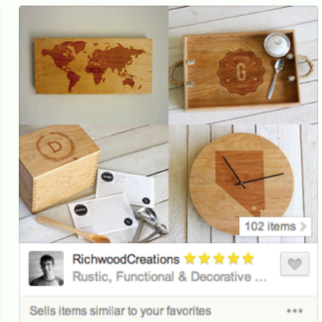
Use style profiles to generate personalized content



ITEM RECS



USER REC



SHOP REC

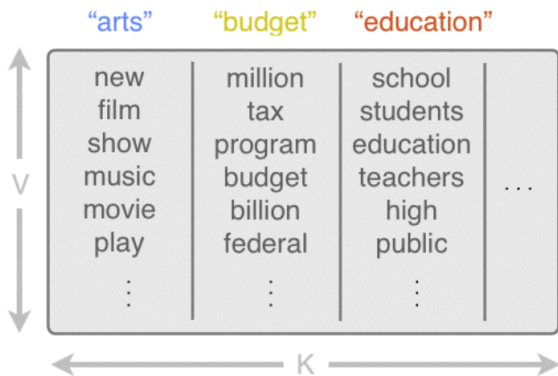
Etsy – 个性化推荐

Latent Dirichlet Allocation

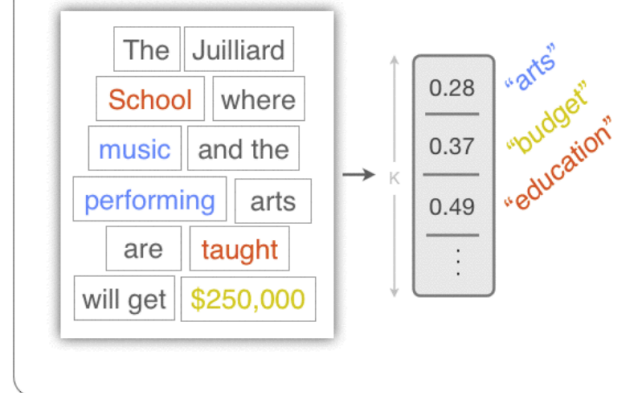
Large collection of text documents



Topics as distribution over words



Documents as distribution over topics



Etsy – 个性化推荐

Latent Dirichlet Allocation

Article
about Juilliard

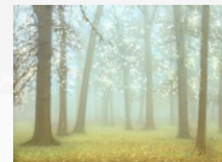
The
Juilliard School
where music and
the performing
arts are taught
will get \$250,000



Diane's
favorited items



#101975185



#63876344



#100109163

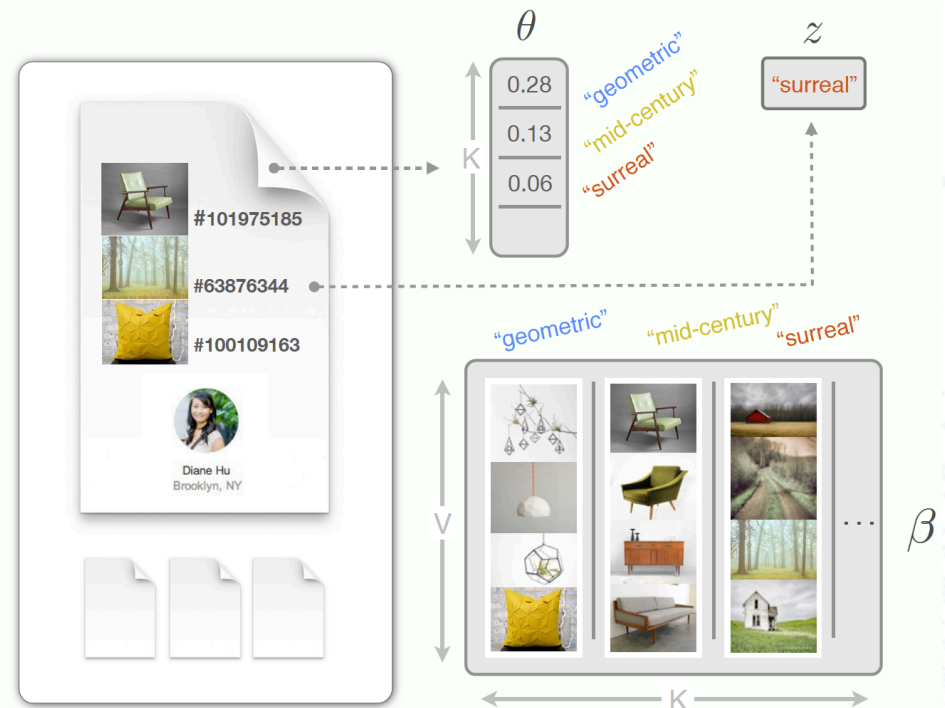
Etsy – 个性化推荐

Latent Dirichlet Allocation

Assume: Each user's favorited items are generated by this process:

For each user u ,

1. Draw a style profile:
 $\theta \sim \text{Dirichlet}(\alpha)$
2. For each item, x_n that user u has favorited,
 - (a) Draw a style:
 $z_n \sim \text{Multinomial}(\theta)$
 - (b) Draw an item:
 $x_n \sim \text{Multinomial}(\beta_{z_n})$



Etsy – 个性化推荐

Latent Dirichlet Allocation

1) Inference:

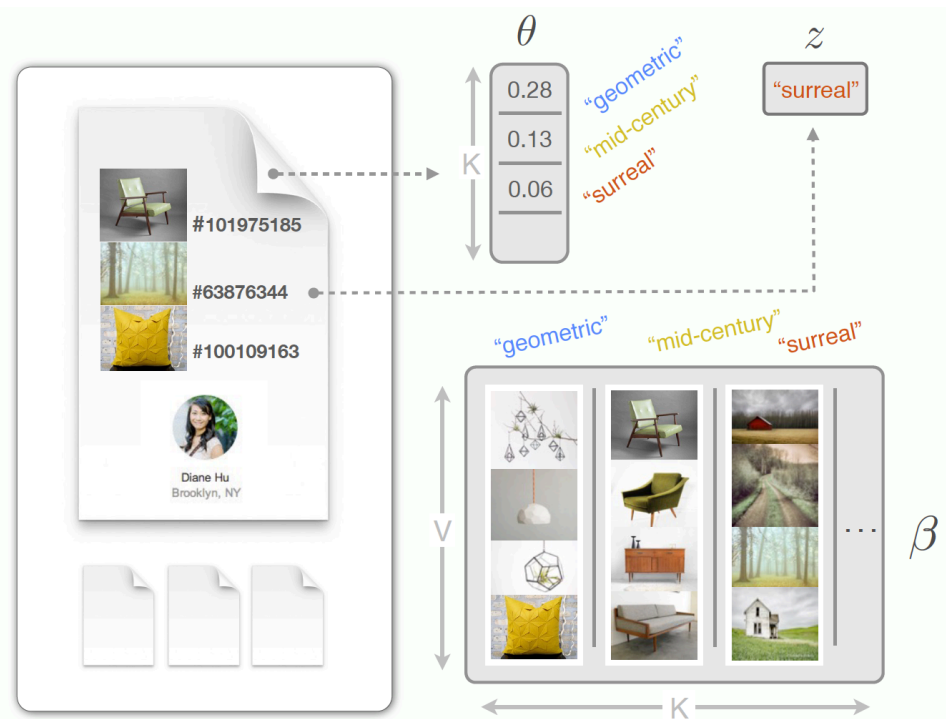
Determine posterior distribution:

$$p(\theta, \mathbf{z} | \mathbf{w}, \alpha, \beta) = \frac{p(\theta, \mathbf{z}, \mathbf{w} | \alpha, \beta)}{p(\mathbf{w} | \alpha, \beta)}$$

2) Estimation:

Choose α and β that maximize the log-likelihood of all user's data:

$$\mathcal{L}(\alpha, \beta) = \sum_{m=1}^M \log p(\mathbf{w} | \alpha, \beta)$$

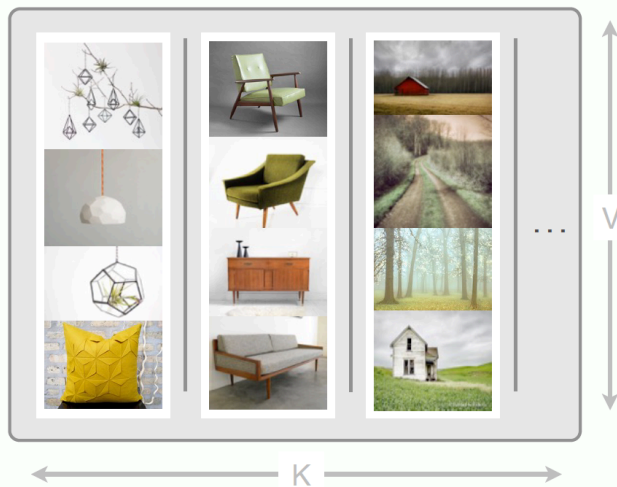


Etsy – 个性化推荐

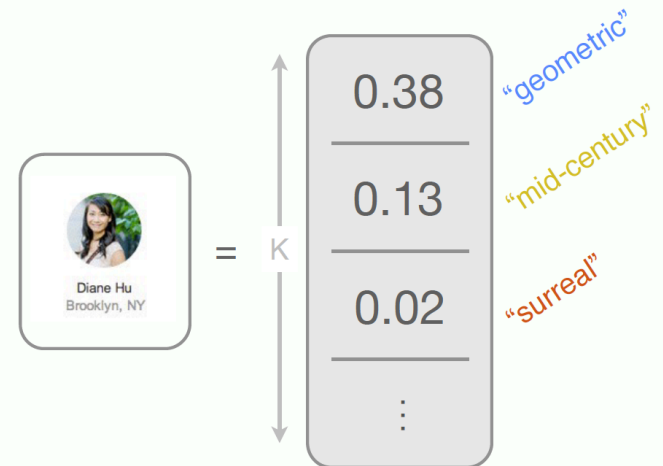
Latent Dirichlet Allocation

Discover popular styles on Etsy as a distribution over items

“geometric” “mid-century” “surreal”

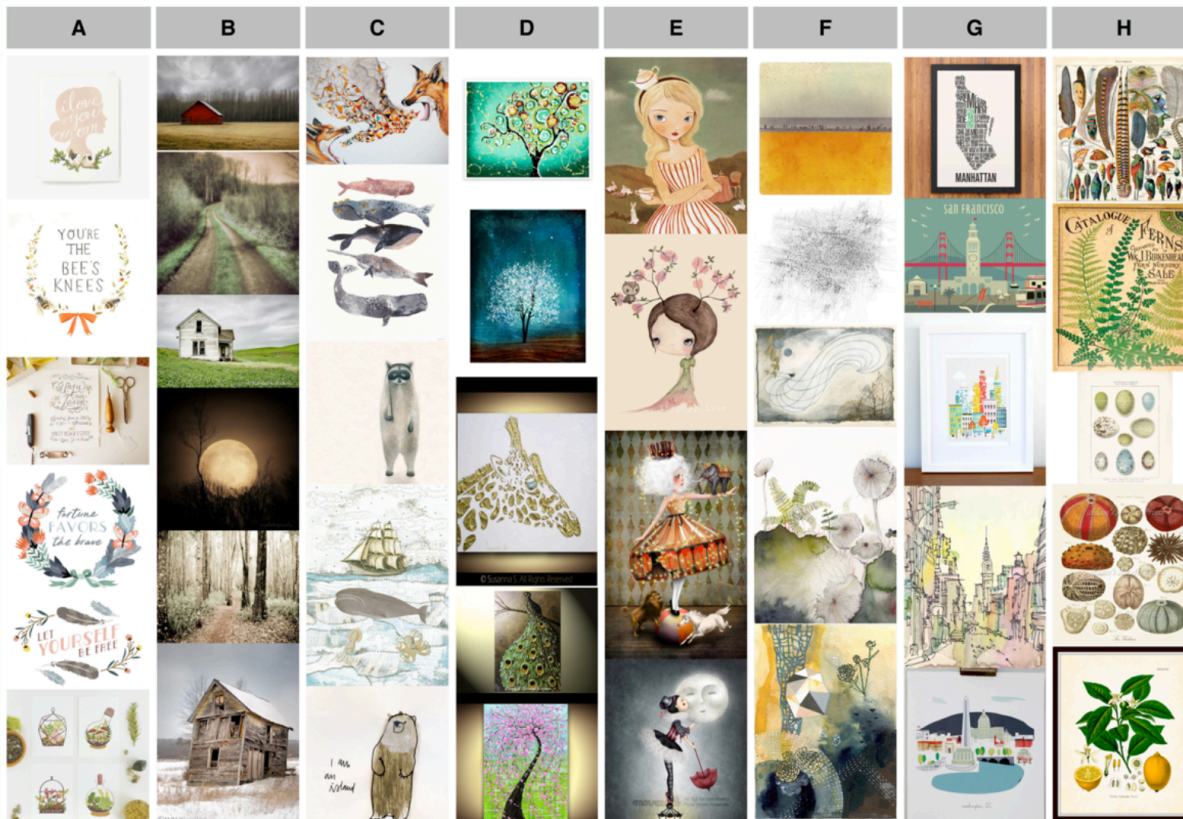


Represent each user as a distribution over popular styles, i.e. “style profile”



Etsy – 个性化推荐

LDA学习到的各种“类型样式”画像



Example of learned styles that contain art prints:

A = Botanical

B = Surreal landscapes

C = Whimsical

D = Acrylic/Abstract

E = French Dolls

F = Whimsical/Abstract

G = Cities

H = Vintage

Etsy – 个性化推荐

从每个用户的“类型样式”画像中推荐用户喜爱的物品

MY FAVORITES



STYLE #428



STYLE #54



STYLE #655



STYLE #87



Etsy – 个性化推荐

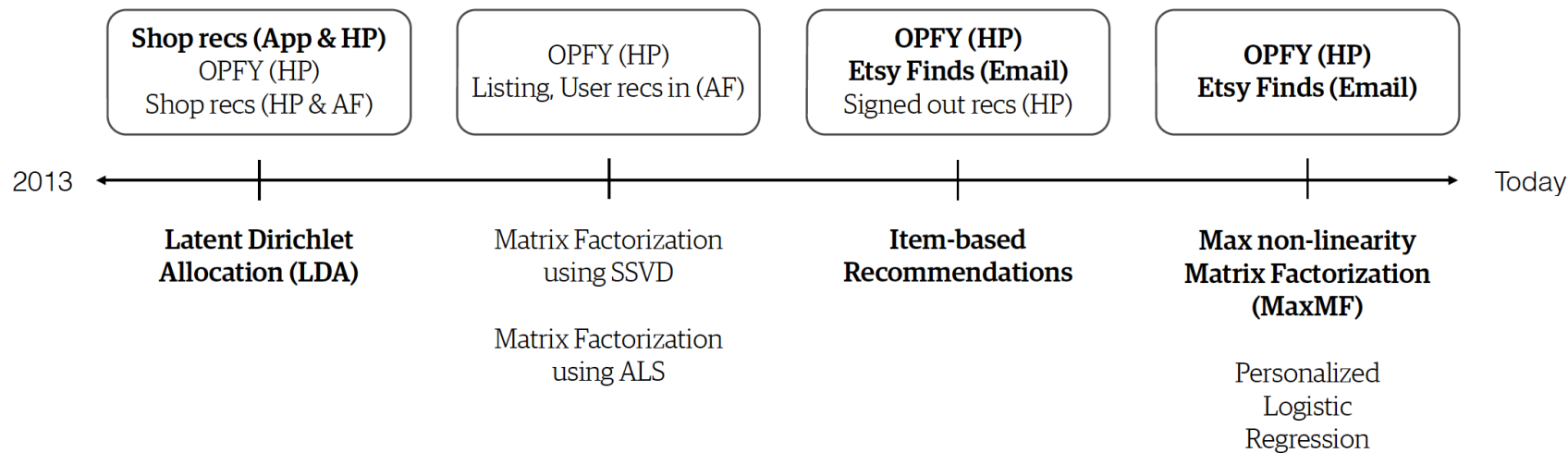
利用每个用户的“类型样式”画像：

- **用户推荐**
 - 利用LSH做快速的近似KNN，然后得到Top-N最相近的用户
- **产品推荐**
 - 利用用户画像和主题-物品 (Topic-Item) 矩阵乘积
 - 返回Top-N结果给用户
- **商铺推荐**
 - 用商铺 (Shop) 替代物品重建模型
 - 利用产品推荐来返回商铺推荐



Etsy – 个性化推荐

一个简短的历史



Etsy - 计算广告

Etsy [Sell on Etsy](#) [Register](#)

[Clothing & Accessories](#) [Jewelry](#) [Craft Supplies & Tools](#) [Weddings](#) [Entertainment](#) [Home & Living](#) [Kids & Baby](#) [Vintage](#)

Show results for: All categories > "stud earrings" (363,150 Results) Sort by: Relevancy ▾

All categories

- Jewelry
- Craft Supplies & Tools
- Weddings
- Accessories
- Home & Living
- Paper & Party Supplies
- Bath & Beauty
- Art & Collectibles
- Toys & Games
- Bags & Purses
- Clothing
- Pet Supplies
- Electronics & Accessories
- Shoes
- Books, Movies & Music

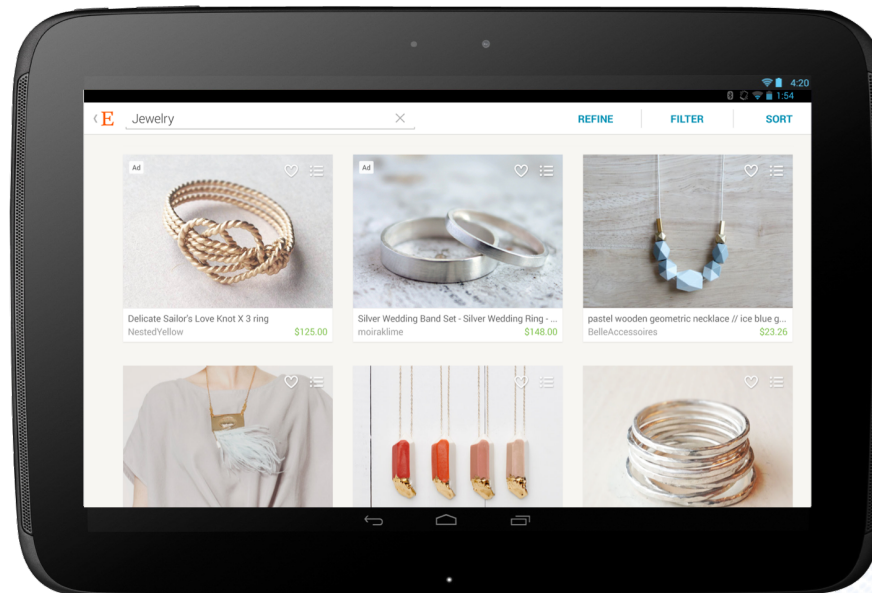
Refine your search

Shop location

<small>Ad</small>	<small>Ad</small>	<small>Ad</small>	<small>Ad</small>
14K Amethyst Earrings Genuine AAA Amet... PristinityJewelry \$99.00	SINGLE 2mm 14k SOLID Rose Gold Stud E... tinysparklestudio \$37.50	AAA Morganite Earring, Fancy Scroll Earrin... GNGJewel \$185.00	Rainbow glitter studs mom wife wood stu... starlightwoods \$27.75
Small Earring Set of Three Mix and Match...	Mint Succulent Earrings. Gift. Polymer cla...	Hook Stud Earrings, Sterling Silver, Gold P...	Rose gold Druzy earrings, Faux Druzy stud...

大规模广告点击率预估系统

- Kamelia Aryafar, Devin Guillory and Liangjie Hong. **An Ensemble-based Approach to Click-Through Rate Prediction for Promoted Listings at Etsy**. To appear in the proceedings of **AdKDD & TargetAd 2017** workshop, held in conjunction with the 23rd SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD 2017**), Halifax, Nova Scotia, August, 2017.



Etsy – 计算广告

大规模广告点击率预估系统

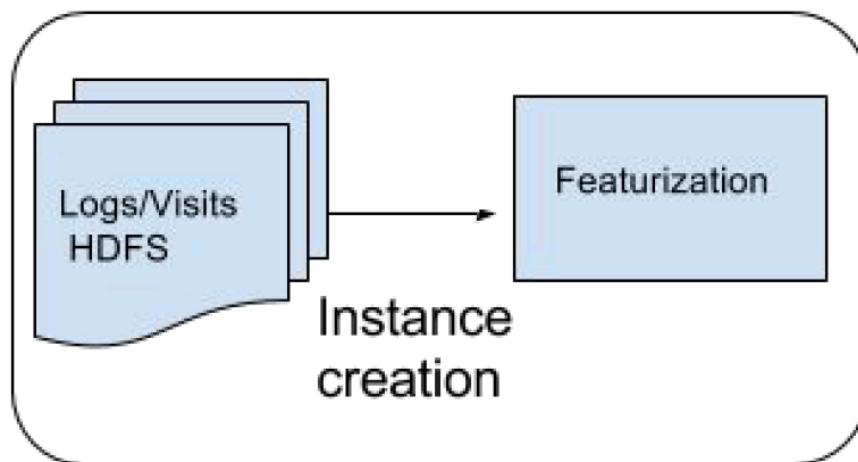
- 数据收集
- 特征工程 (Feature Engineering)
- 学习算法
- 集成学习 (Ensemble Learning)



Etsy – 计算广告

数据收集

- 训练数据：30天的广告用户交互历史记录
- 平衡采样
- 评估数据：前一天的广告用户交互历史



Etsy – 计算广告

特征工程

- 简单基于一元 (Unigram) 的文本特征
- 卖家、广告、用户的历史信息
- 广告本身的一些信息 (比如, 收藏数量, 单价等)
- 基于SVD的特征
- 利用特征哈希, 把大量的稀疏的特征映射到固定维度的空间

$$A_k = U_k \Sigma_k V_k^T$$

$m \times n$ $m \times r$ $r \times r$ $r \times n$

学习训练

- 使用Logistic Regression模型
- 利用FTRL-Proximal算法进行在线更新模型参数

Given a sequence of gradients $\mathbf{g}_t \in \mathbb{R}^d$, OGD performs the update

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta_t \mathbf{g}_t,$$

where η_t is a non-increasing learning-rate schedule, e.g., $\eta_t = \frac{1}{\sqrt{t}}$. The FTRL-Proximal algorithm instead uses the update

$$\mathbf{w}_{t+1} = \arg \min_{\mathbf{w}} \left(\mathbf{g}_{1:t} \cdot \mathbf{w} + \frac{1}{2} \sum_{s=1}^t \sigma_s \|\mathbf{w} - \mathbf{w}_s\|_2^2 + \lambda_1 \|\mathbf{w}\|_1 \right),$$

Etsy – 计算广告

集成学习

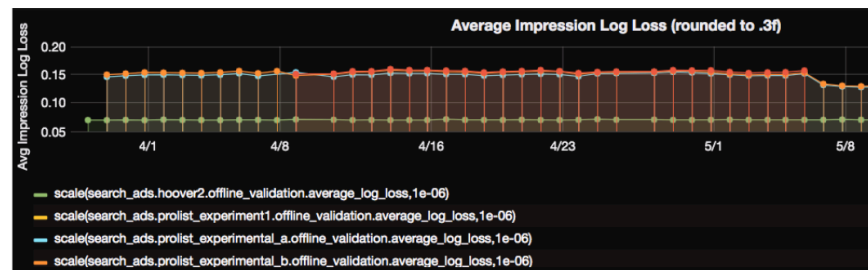
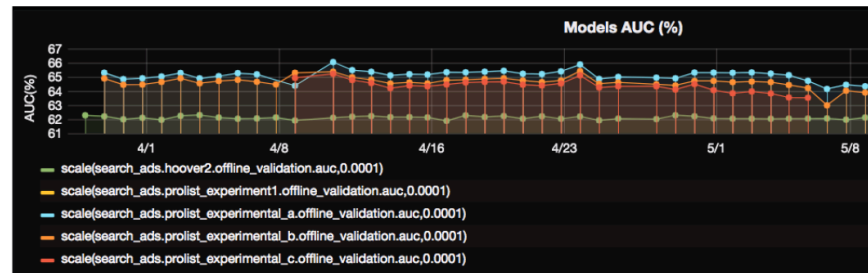
- 对数据不足的一批广告商采用一个模型
- 对数据充足的广告商采用另外一个模型
- 然后再把这两个模型结合起来



Etsy – 计算广告

实验结果

	Historical			Content-based			Ensemble		
	mixed	cold	warm	mixed	cold	warm	mixed	cold	warm
AUC (%)	+1.56	+1.89	+1.55	-1.57	+6.39	-1.74	+1.87	+8.09	+1.79
Log Loss ($\times 10^3$)	-0.016	-0.048	-0.018	+0.311	-0.194	+0.335	-0.200	-0.330	-0.193
Normalized Entropy ($\times 10^3$)	-0.29	-1.23	-0.31	+5.67	-5.00	+6.01	-3.64	-8.51	-3.46



总结

- Etsy , 一个服务全球中小手工业者的电商平台
- Etsy的人工智能挑战
 - 基于深度学习的搜索方案 (KDD 2016)
 - 基于大型概率图模型的个性化推荐方案 (KDD 2014)
 - 基于大型集成学习的计算广告平台方案 (AdKDD 2017/KDD 2017)

2017 Machine Learning Summit

人工智能在Etsy电商平台中的技术实践

谢谢大家，欢迎提问