Machine Learning and Data Science at Etsy

Liangjie Hong
Head of Data Science

Liangjie Hong

- Head of Data Science at Etsy.
- Senior Manager of Research at Yahoo Research in Sunnyvale, CA
  Leading science efforts for personalization and search sciences.
- Published papers in SIGIR, WWW, KDD, CIKM, AAAI, WSDM, RecSys and ICML (2000+ citations)
- WWW 2011 Best Poster Paper Award
  WSDM 2013 Best Paper Nominated
  RecSys 2014 Best Paper Award
- Program committee members in KDD, WWW, SIGIR, WSDM, AAAI, EMNLP, ICWSM, ACL, CIKM, IJCAI and various journal reviewers
- PhD in Machine Learning from Lehigh University

Etsy

Etsy – A Global Marketplace


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 Bendycki


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


Woodstorming
Kaunas, Lithuania
Photo by: Ilona & 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

By The Numbers

1.6M
active sellers
AS OF MARCH 31, 2016

25M
active buyers
AS OF MARCH 31, 2016

$2.39B
annual GMS
IN 2015

35+M
items for sale
AS OF MARCH 31, 2016

Photo by Kirsty-Lyn Jameson

Work and Culture

852

employees around the world

AS OF MARCH 31, 2016

9

offices in 7 countries

AS OF MARCH 31, 2016

54%

female employees

46%

male employees

AS OF DECEMBER 31, 2015

Photo by Emily Andrews

Work and Culture

1.6M
active sellers
AS OF MARCH 31, 2016

86%
of sellers are women
2014 ETSY SELLER SURVEY

95%
of sellers run their Etsy shop from home
2014 ETSY SELLER SURVEY

76%
consider their shop a business
2014 ETSY SELLER SURVEY

Photo by Moira K. Lime

Passionate and Loyal Business Owners

30%

focus on their creative businesses as their sole occupation

2014 ETSY SELLER SURVEY

65%

started their Etsy shop as a way to supplement income

2014 ETSY SELLER SURVEY

79%

started their Easy shop as an outlet for creativity

2014 ETSY SELLER SURVEY

Photo by Panagiotis Mina

Engaged and Thoughtful Buyer Base

25M

active buyers

AS OF MARCH 31, 2016

87%

of Etsy buyers are women

2014 ETSY BUYER SURVEY

92%

of buyers agree Etsy offers products they can't find elsewhere

2014 ETSY BUYER SURVEY

Photo by Jean-Michael Seminaro

AI in E-commerce

AI Challenges

For Buyers

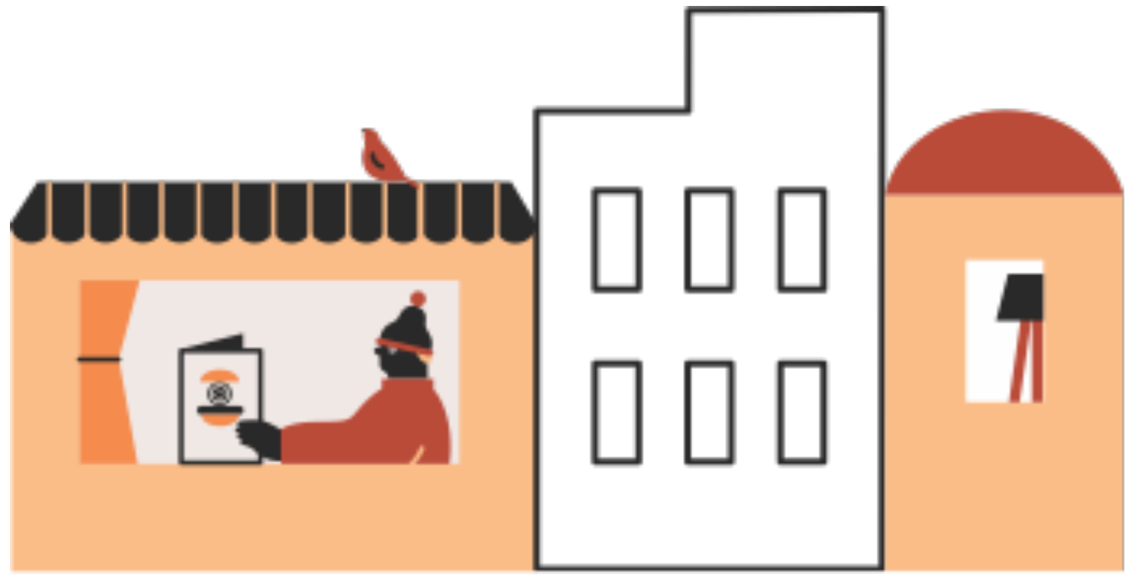
- How to choose unique and satisfied products among millions?
  How to lead and guide buyers to discover products that they wouldn't buy at the first place?
  How to recommend appropriate products for different occasions?

For Sellers

- How to reach larger audience and potential buyers?
  How to run advertising campaign more effectively?
  How to communicate with buyers through different channels?

For Platform

- How to build a healthy platform?
  How to speed-up buyer and seller communication?



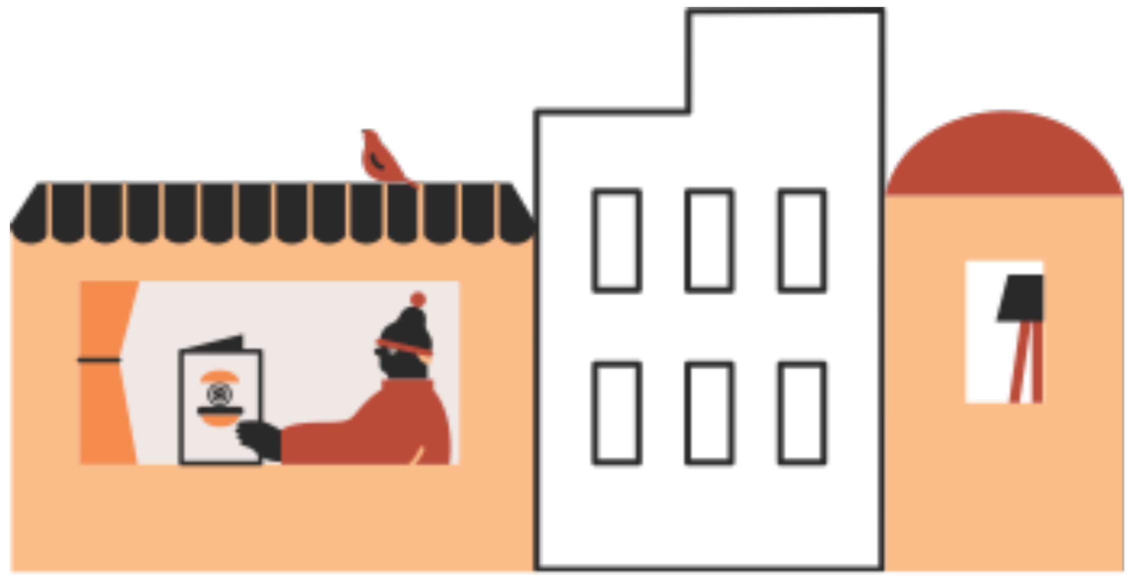
AI in E-commerce

AI Challenges

- **Search and Discovery**
  Query Modeling
  User Intent Modeling
  Learning to Rank

- **Personalization and Recommendation**
  User Profiling
  Item Modeling
  Recommender Ranking

- **Computational Advertising**
  Click-Through Rate Modeling
  Conversion Rate Modeling
  Bid Optimization

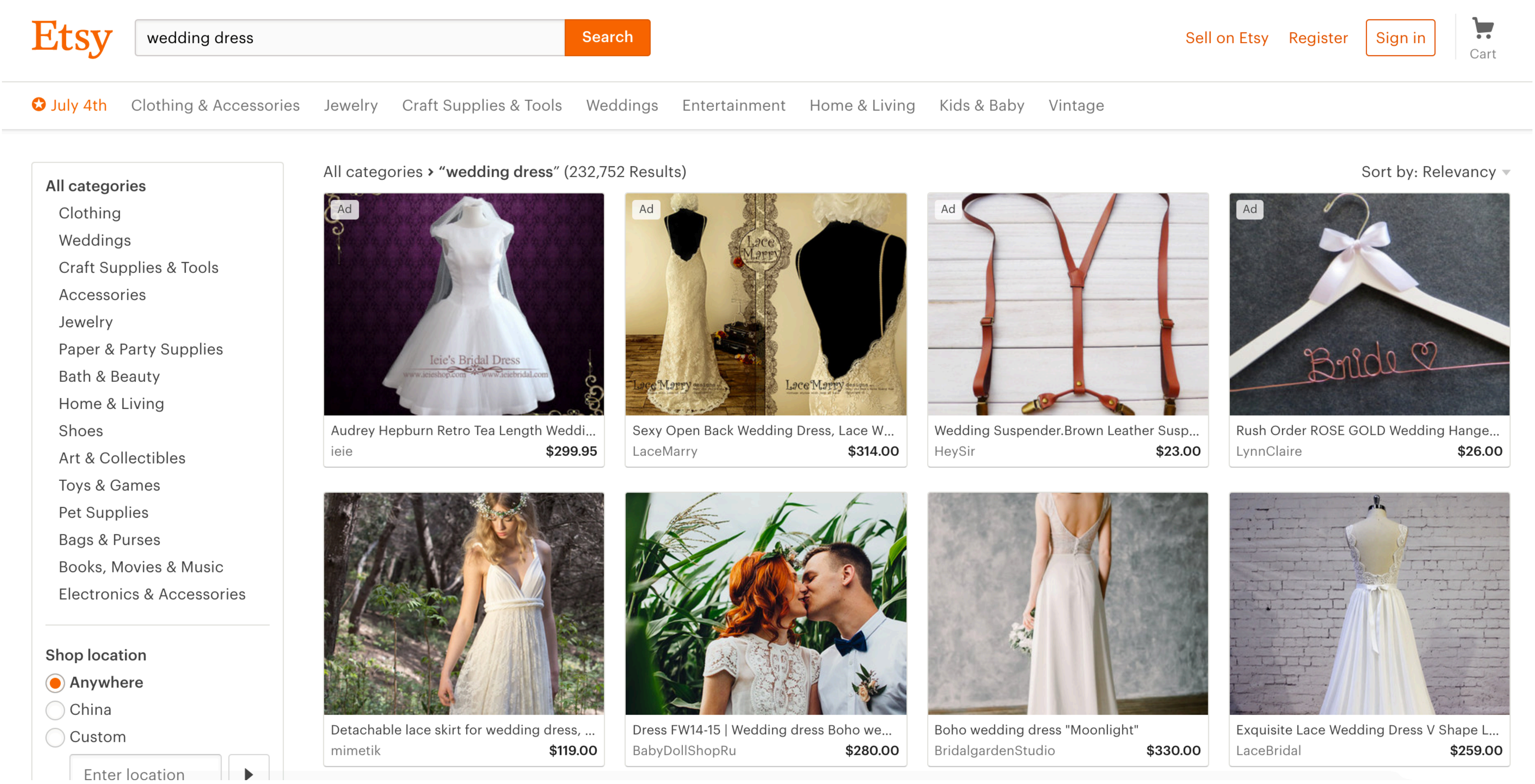


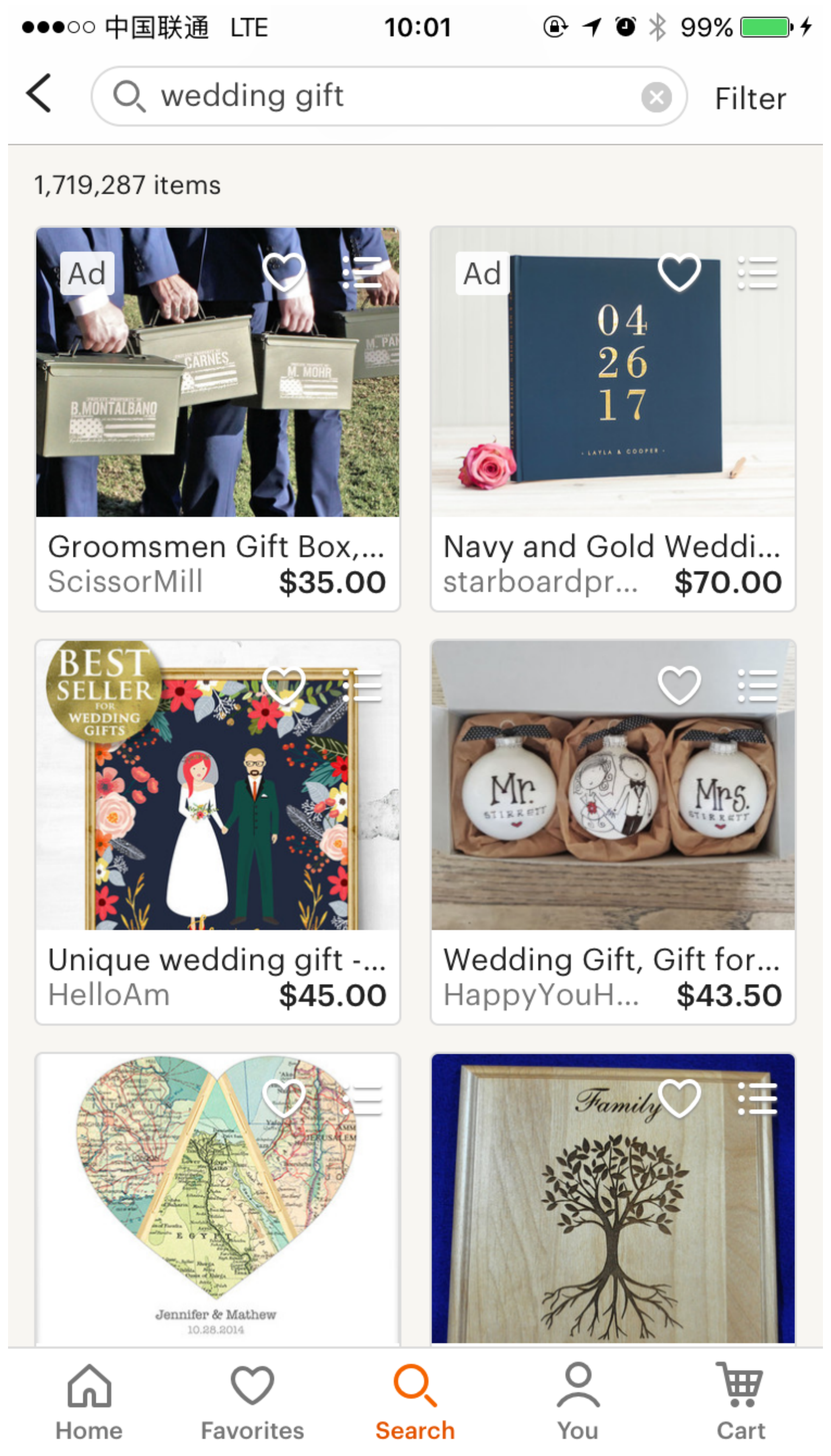
AI in E-commerce

AI in E-commerce at Etsy

- Multi-modal Deep-learning based Search Solution (KDD 2016)
- Probabilistic Graphical Model based Personalization Recommendation (KDD 2014)
- Ensemble Learning based CTR Prediction Solution (AdKDD 2017/KDD 2017)

Etsy – Search and Discovery

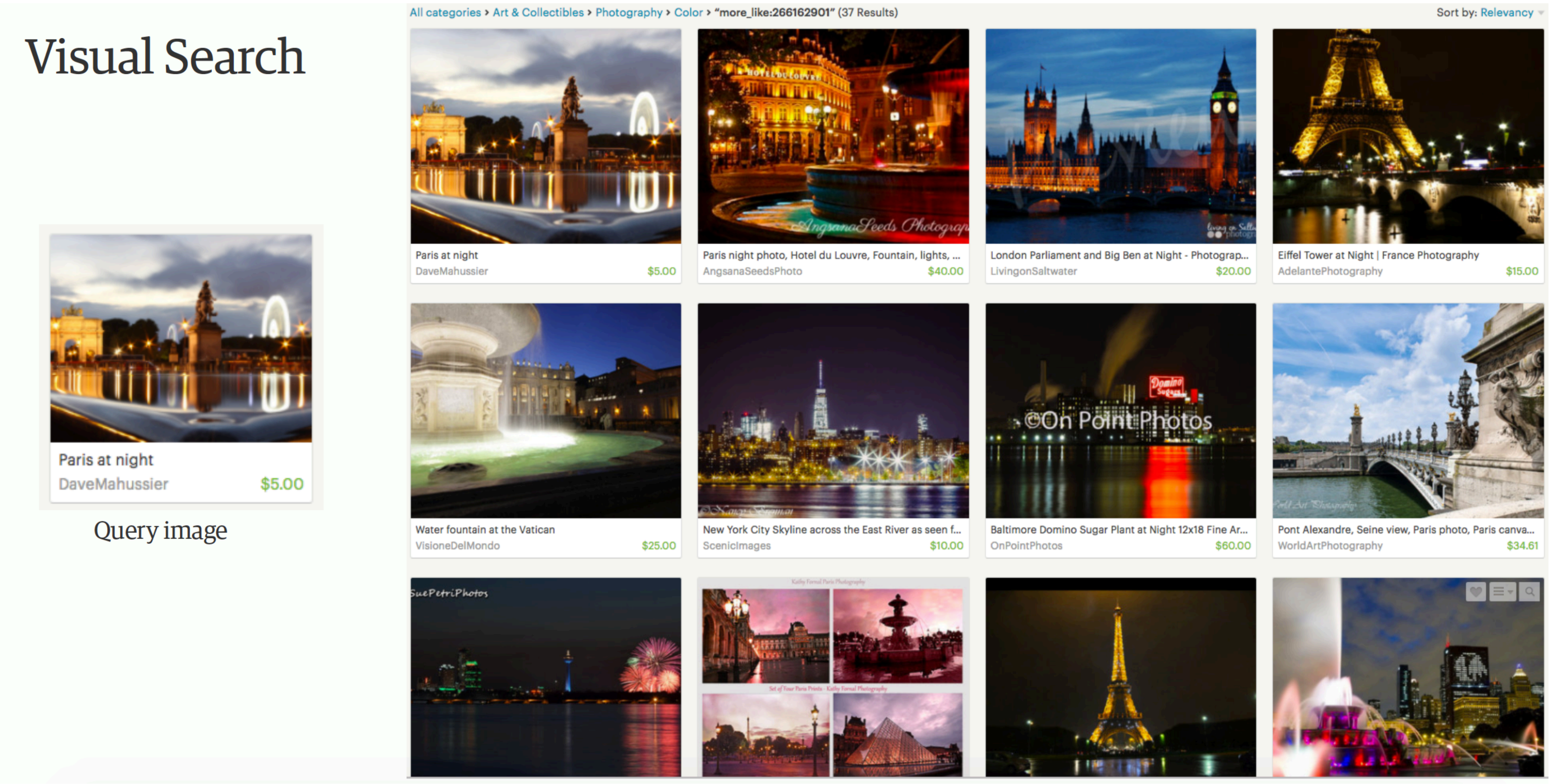




Etsy – Search and Discovery

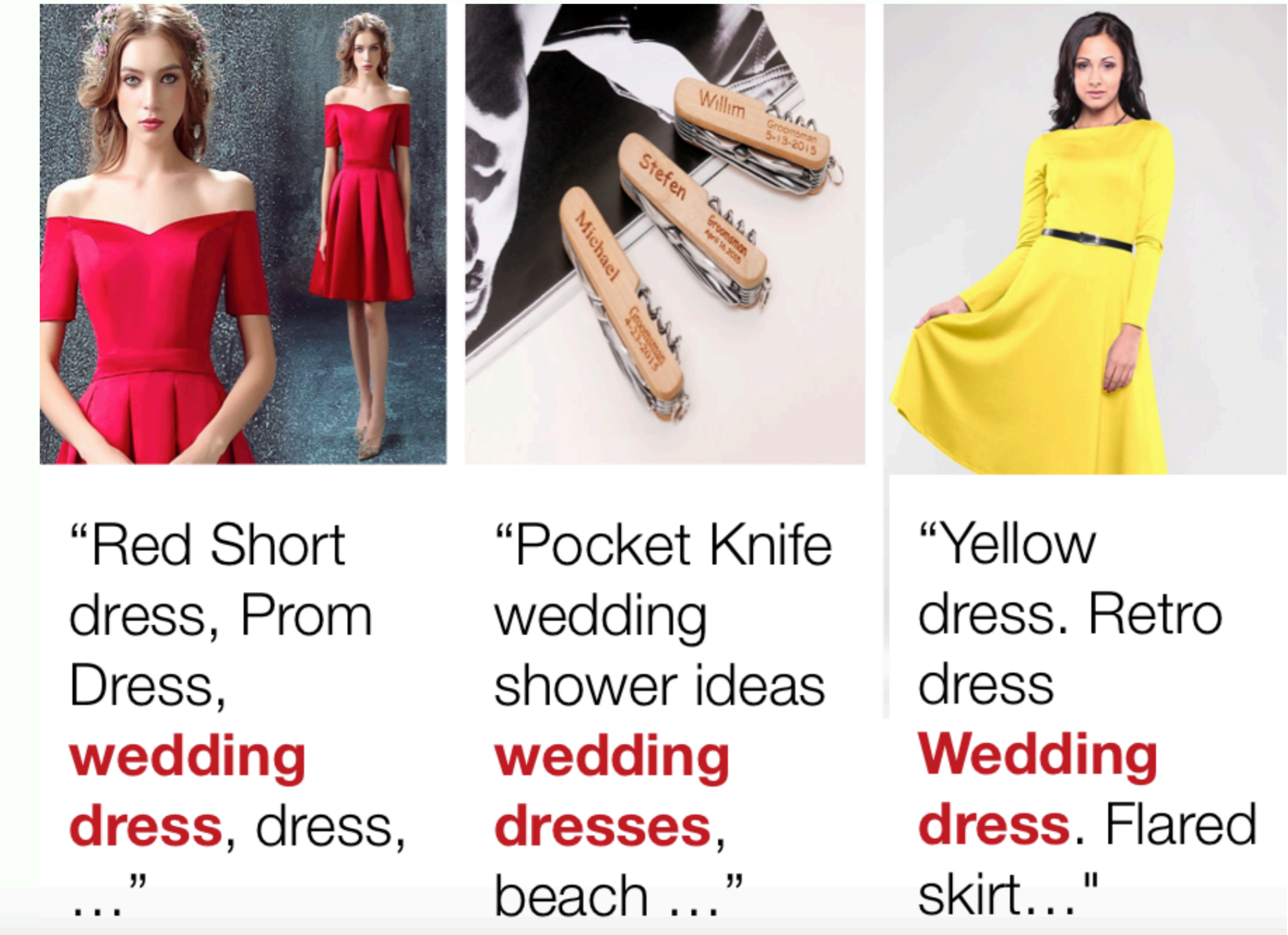
Deep Learning based Multi-modal Learning to Rank

- 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.

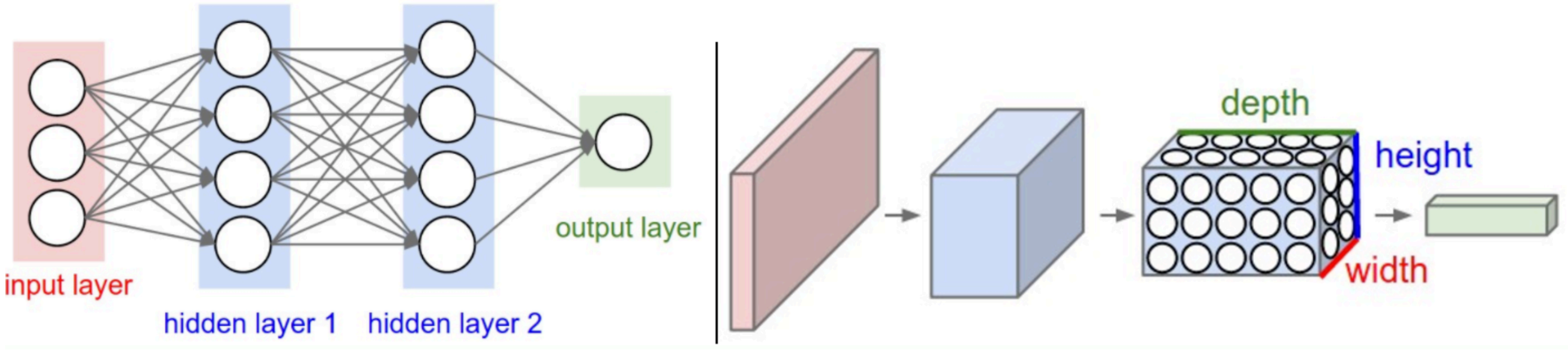


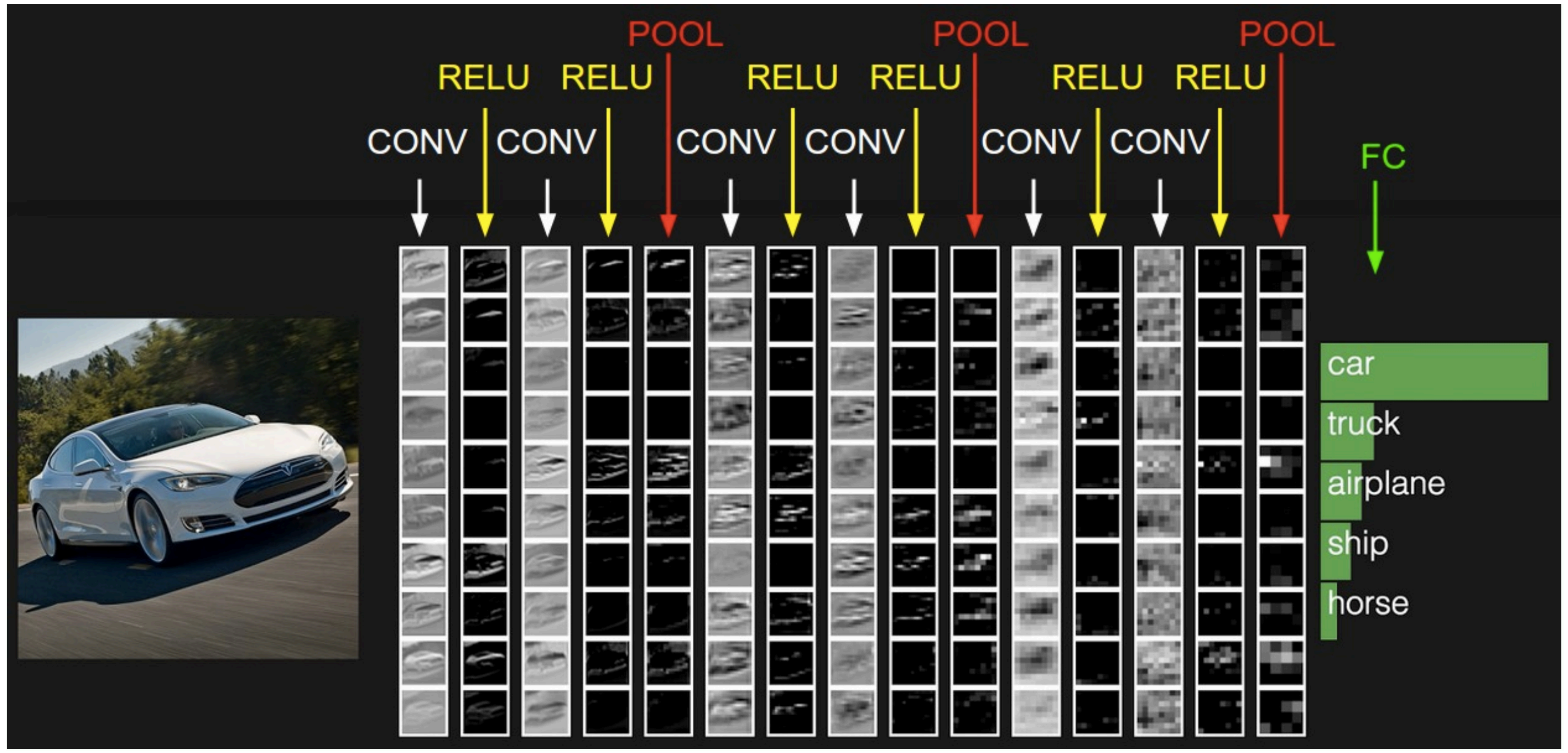
Etsy – Search and Discovery

- **>142M images**
- Important complementary information to text
- Potential only information



Etsy – Search and Discovery

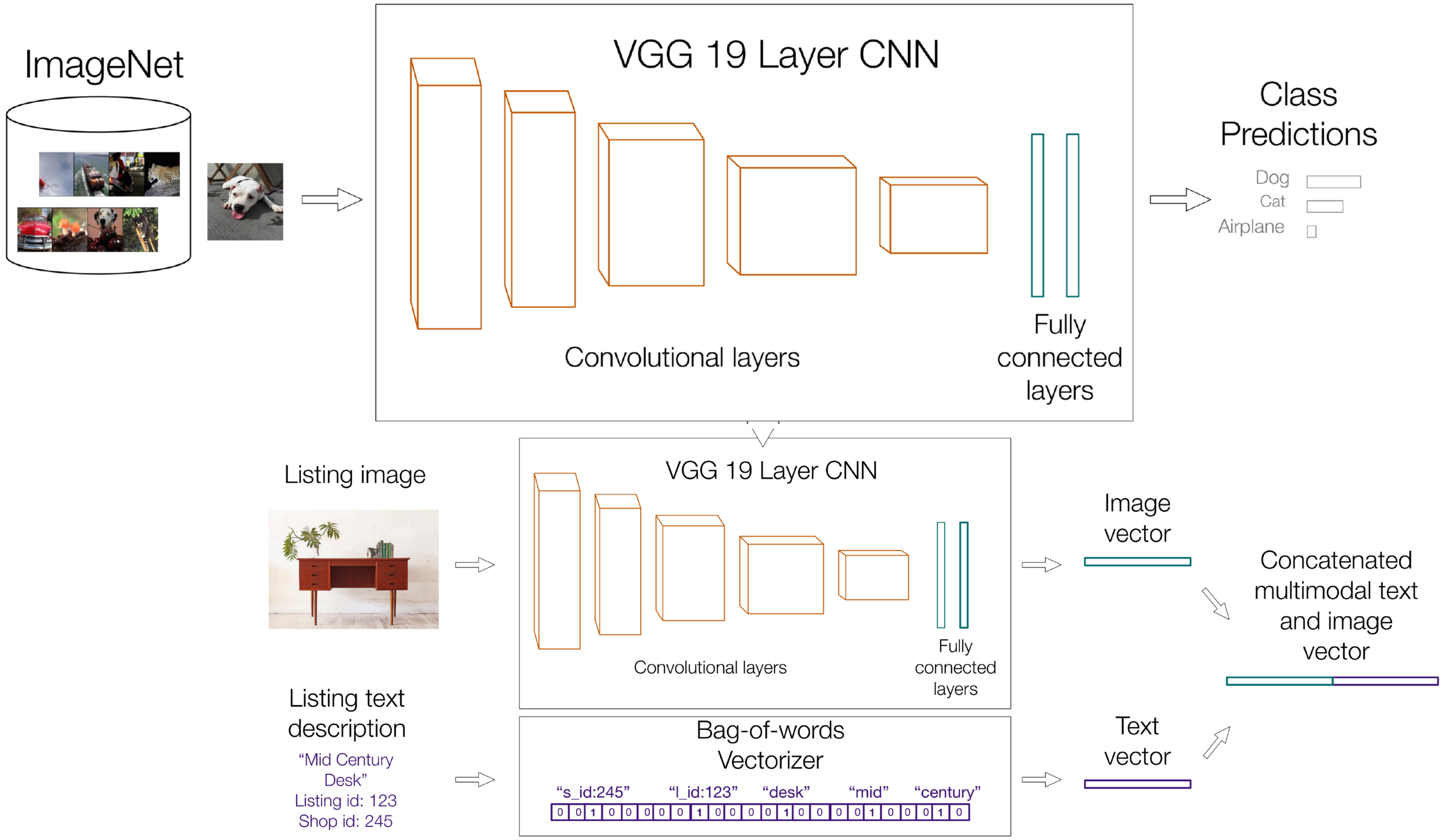




Etsy – Search and Discovery

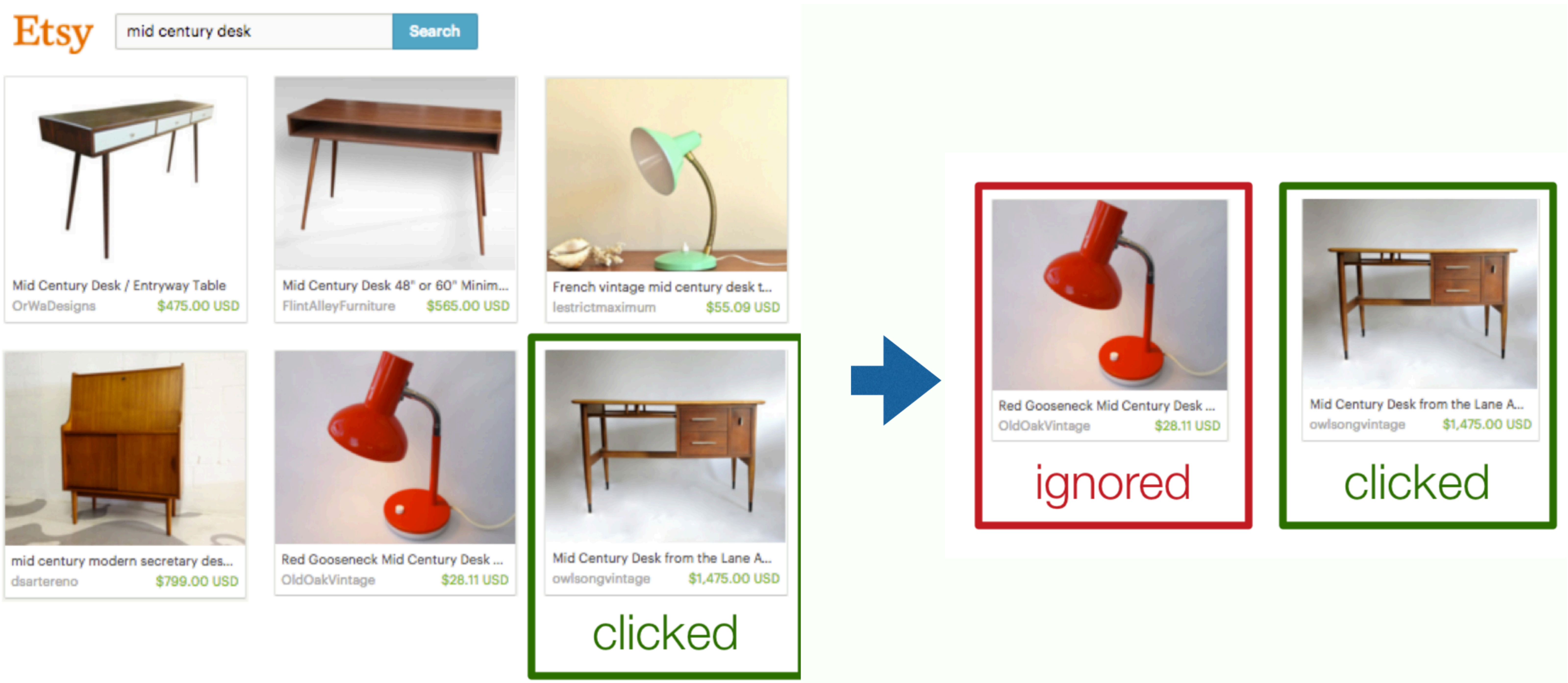
Deep Learning-based Feature Extraction + Transfer Learning

- Learn CNN model from ImageNet
- Extract CNN's parameters
- Apply Model on Etsy's data and combine with text info

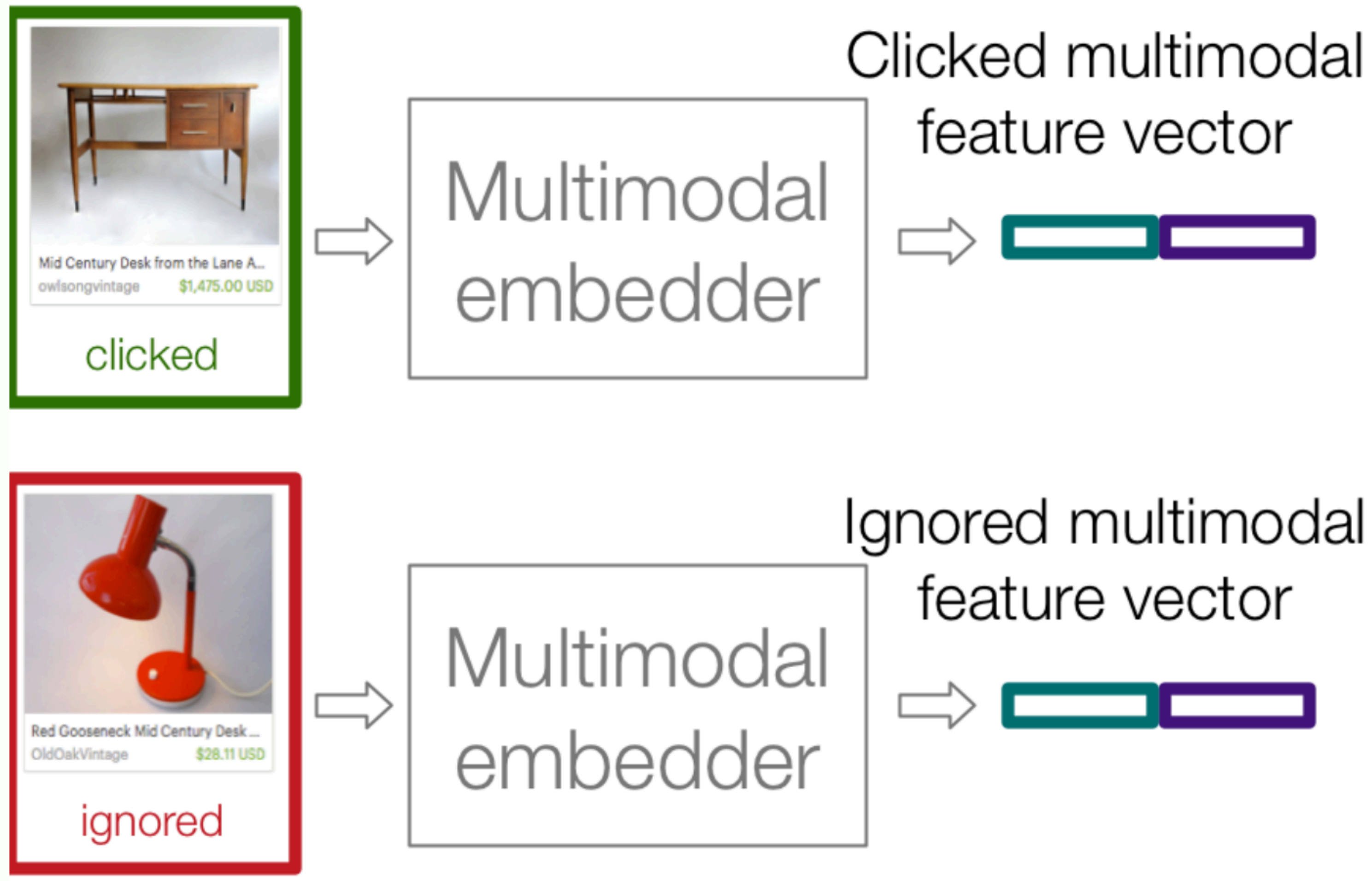


Etsy – Search and Discovery

Multi-modal Learning to rank

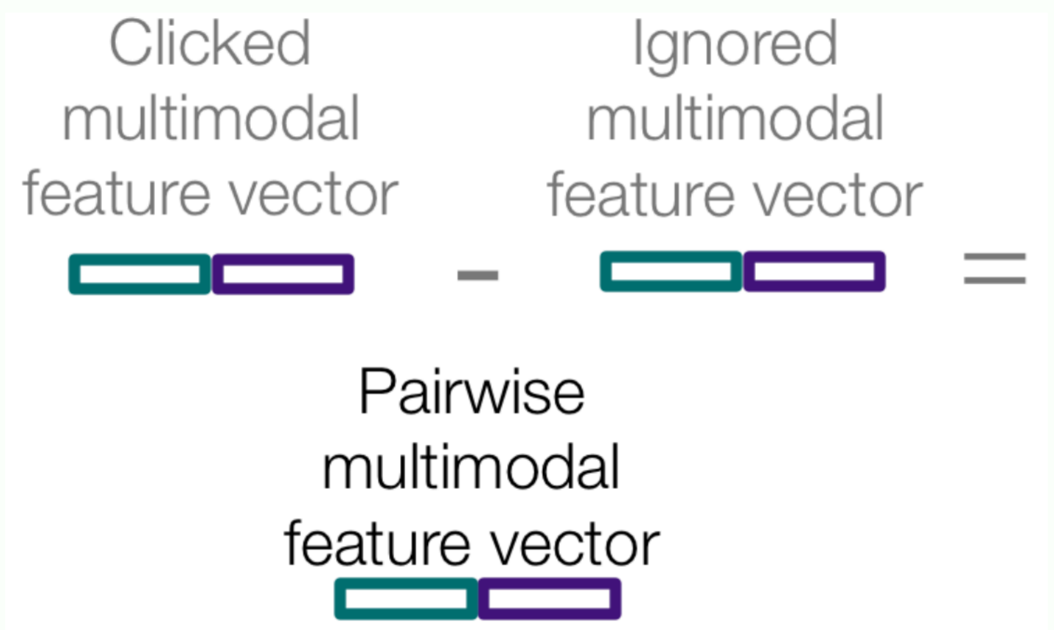


Etsy – Search and Discovery



Algorithm 1 Multimodal Embedding of Listings

procedure EMBEDMULTIMODAL($\mathbf{d}_i$)
  $\mathbf{d}_{T_i} \leftarrow \mathbf{BoW}(text)$
  $\mathbf{d}_{I_i} \leftarrow \mathbf{VGG}(image)$
  $\mathbf{d}_{MM_i} \leftarrow [\mathbf{d}_{T_i}, \mathbf{d}_{T_i}]$
  return $\mathbf{d}_{MM_i}$



Algorithm 2 Generate Pairwise Classification Instances

procedure GETPAIRWISEINSTANCES($\{\langle \mathbf{d}_i^+, \mathbf{d}_i^- \rangle\}$)
  $L \leftarrow \{\}$
  for $i = 1 \dots |P|$ **do** ▷ $|P|$ labeled tuples
    $\mathbf{d}_{MM_i}^+ \leftarrow$ **EmbedMultimodal**($\mathbf{d}_i^+$)
    $\mathbf{d}_{MM_i}^- \leftarrow$ **EmbedMultimodal**($\mathbf{d}_i^-$)
    Draw r uniformly at random from $[0, 1)$
    if $r > 0.5$ **then**
      $x_i \leftarrow \mathbf{d}_{MM_i}^+ - \mathbf{d}_{MM_i}^-$
      $y_i \leftarrow +1$
    else
      $x_i \leftarrow \mathbf{d}_{MM_i}^- - \mathbf{d}_{MM_i}^+$
      $y_i \leftarrow -1$
  $L = L.append(\langle x_i, y_i \rangle)$
  return L ▷ The list of classification instances.

Etsy – Search and Discovery

Text-only ranking for "bar necklace"	Multimodal ranking for "bar necklace"

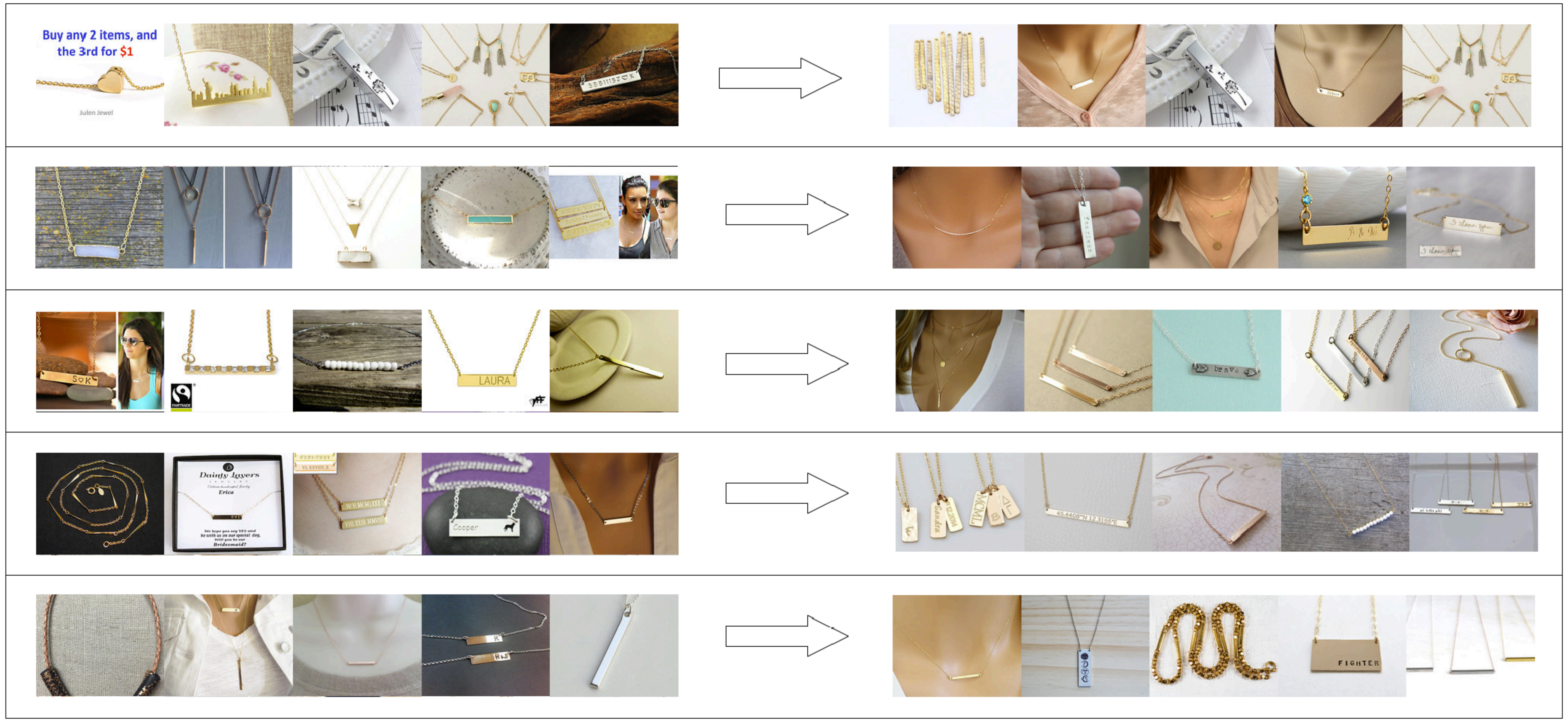
90th percentile

80th percentile

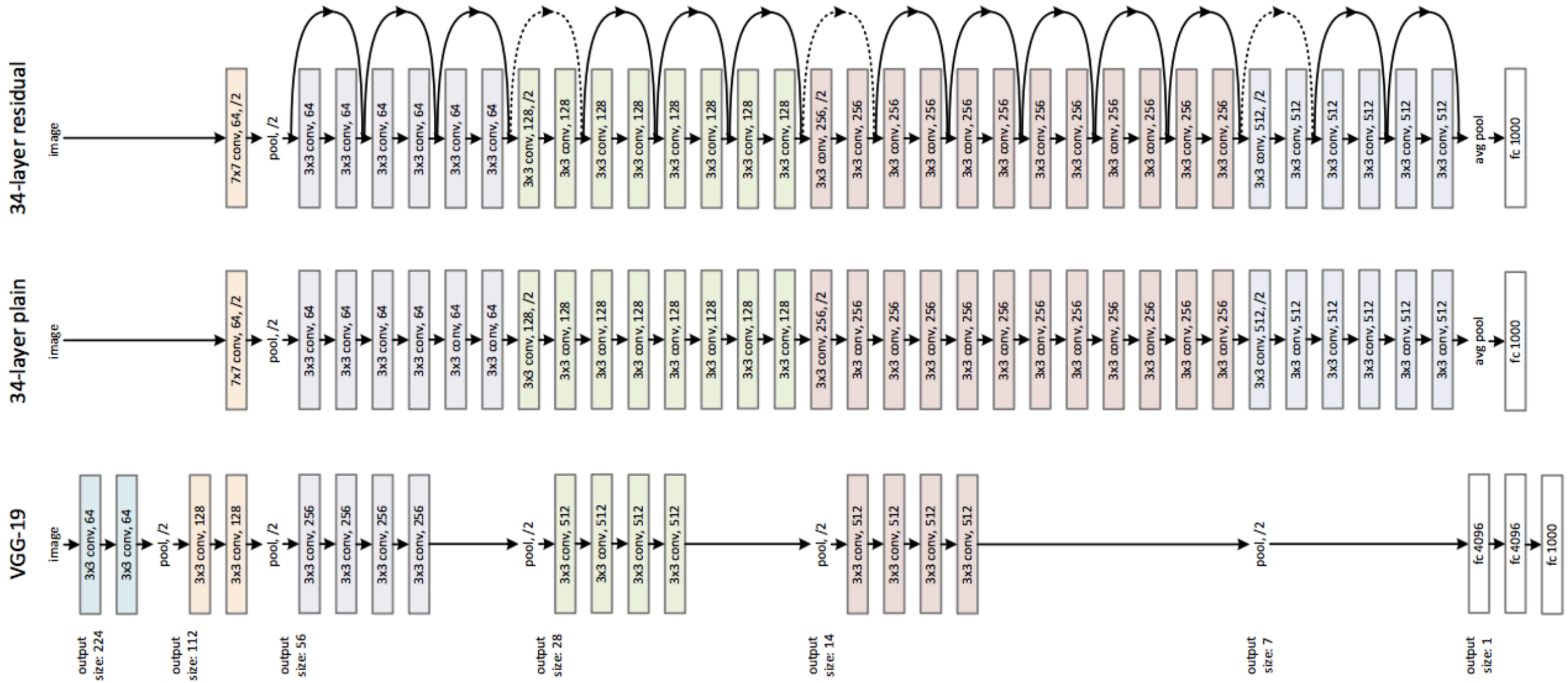
70th percentile

60th percentile

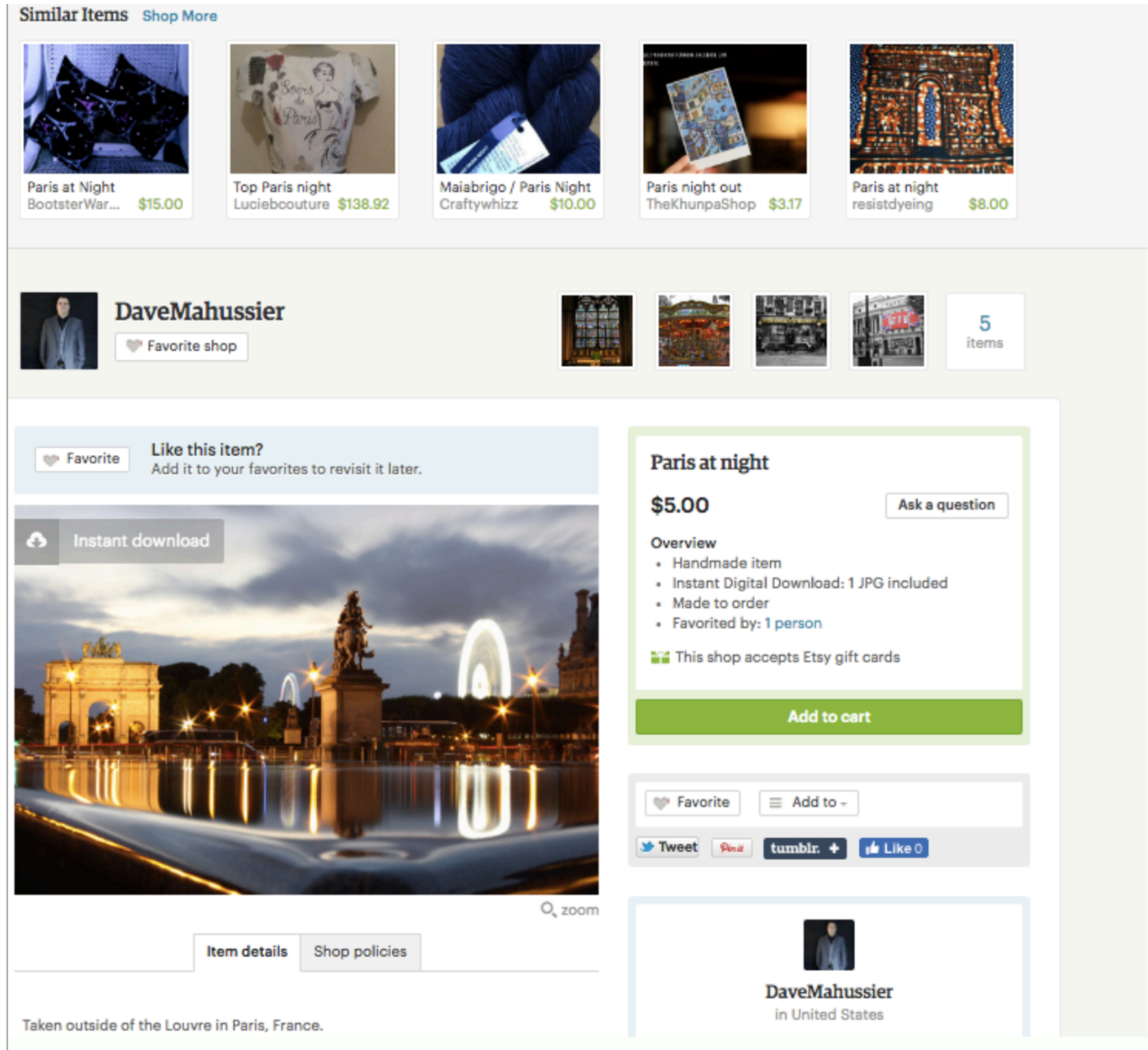
50th percentile



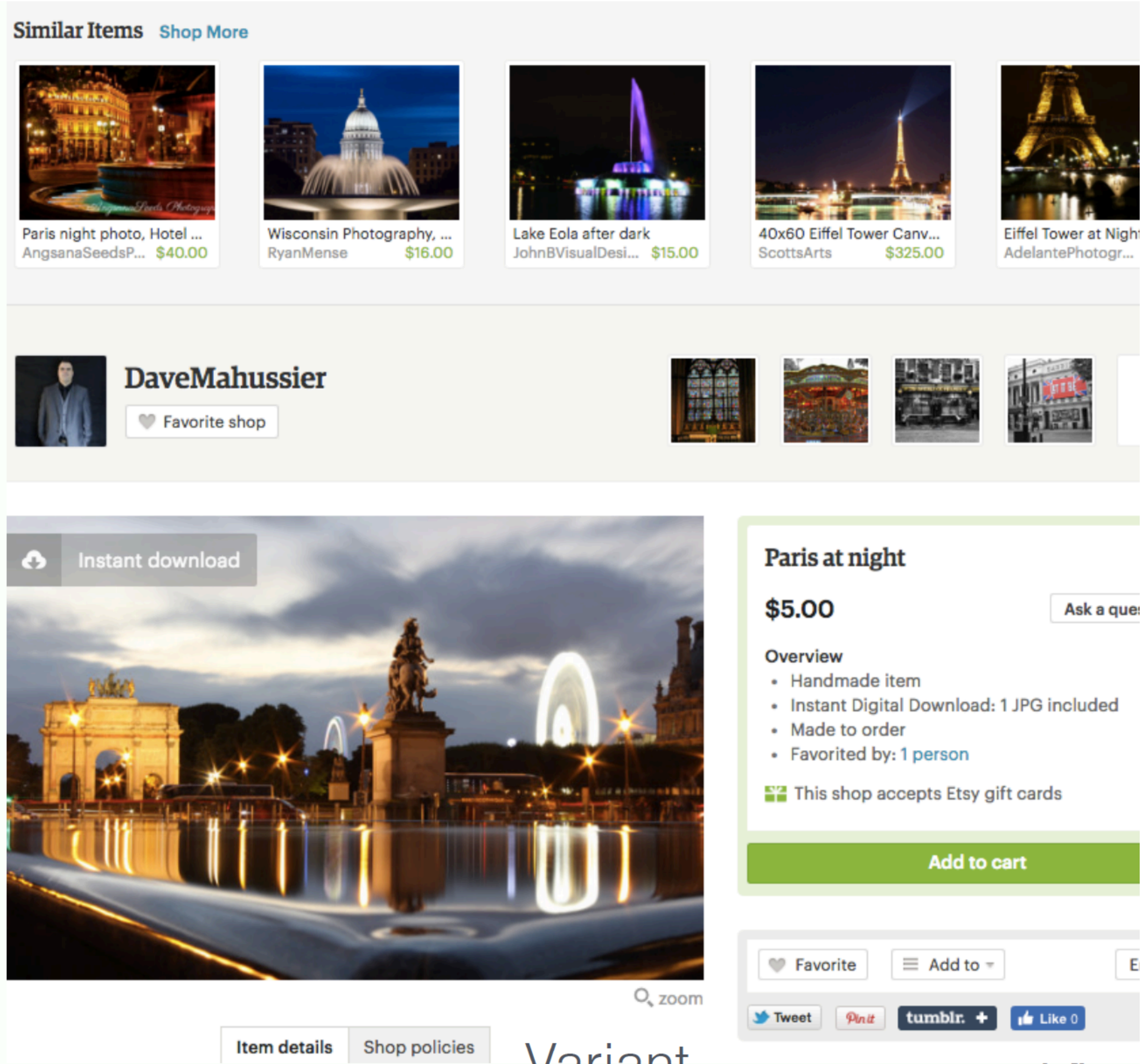
Etsy – Search and Discovery



Etsy – Search and Discovery

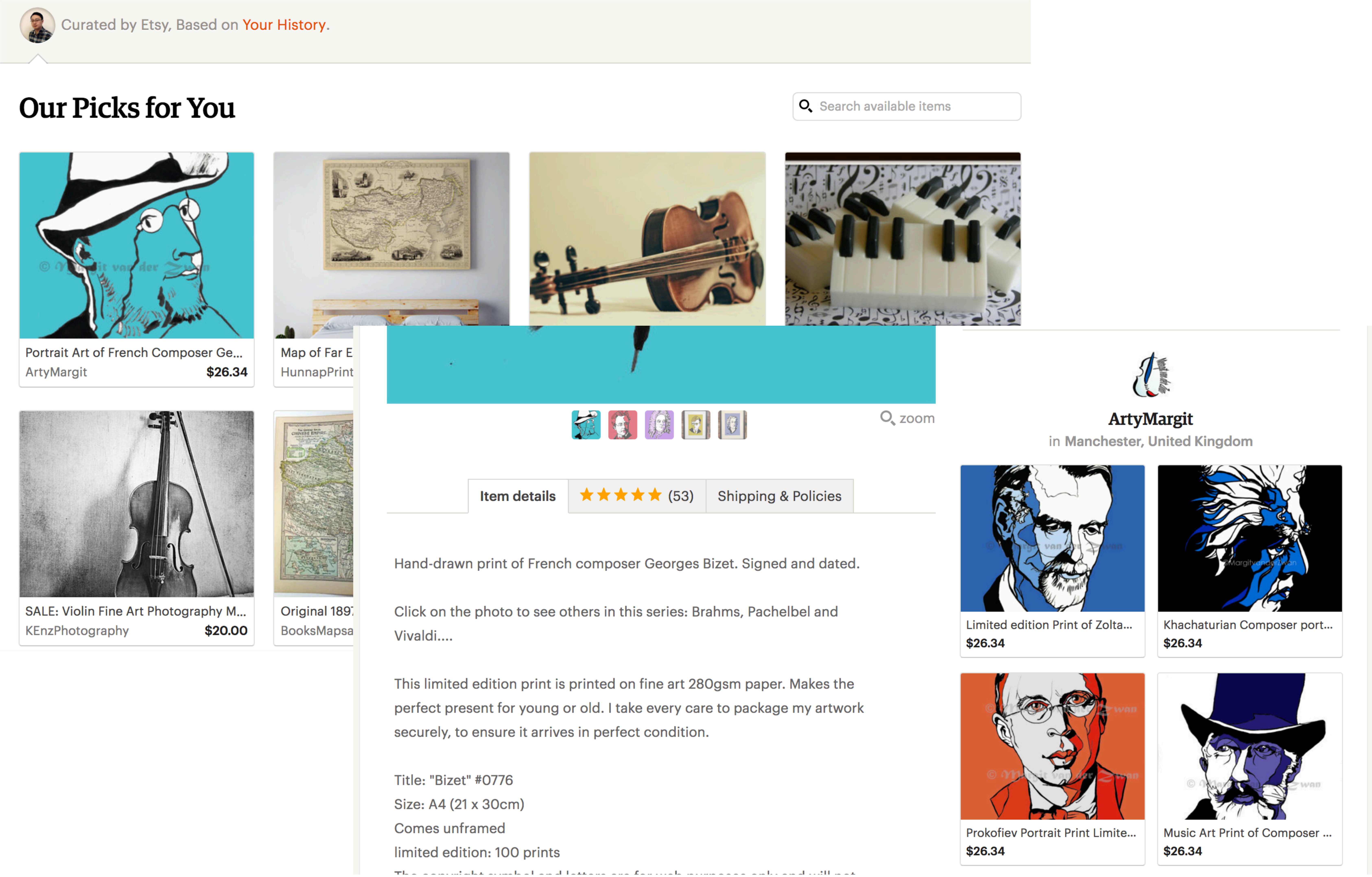


Control



Variant

Etsy – Personalization and Recommendation

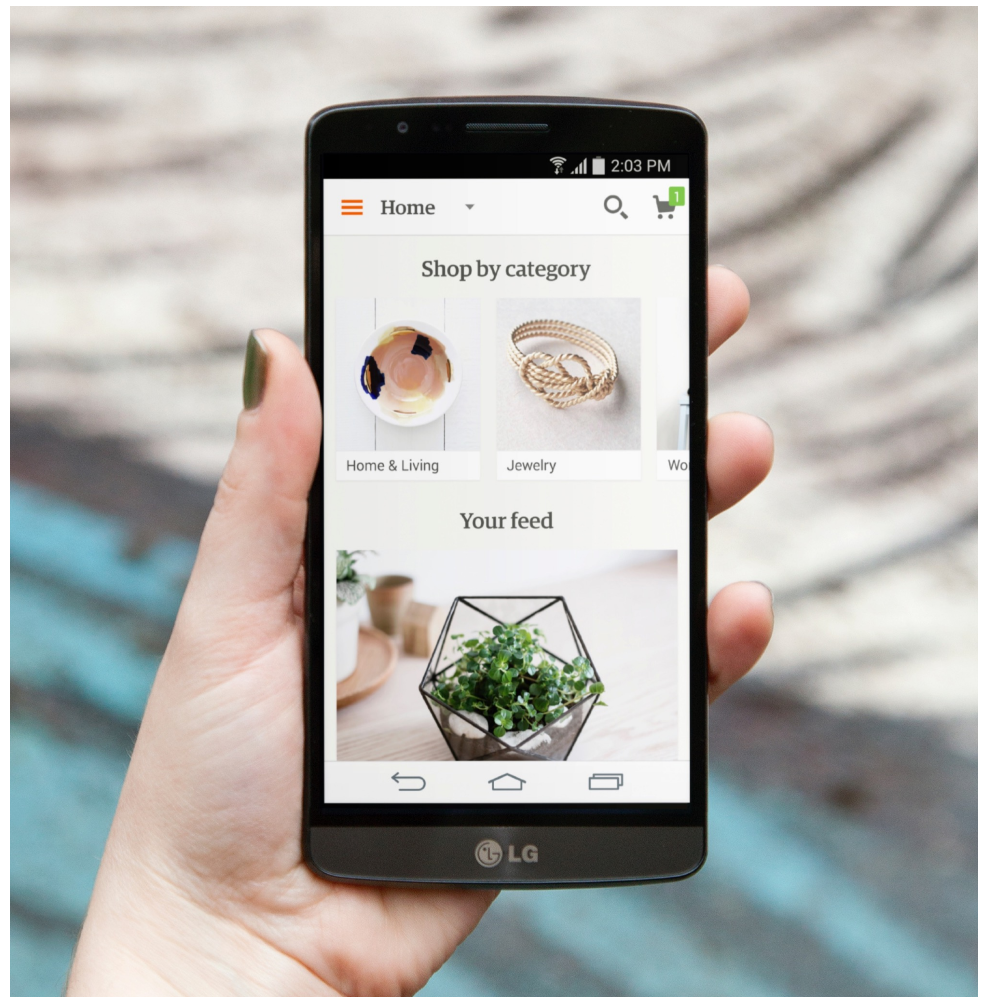


Etsy – Personalization and Recommendation

Probabilistic Graphical Model based Personalization Recommendation

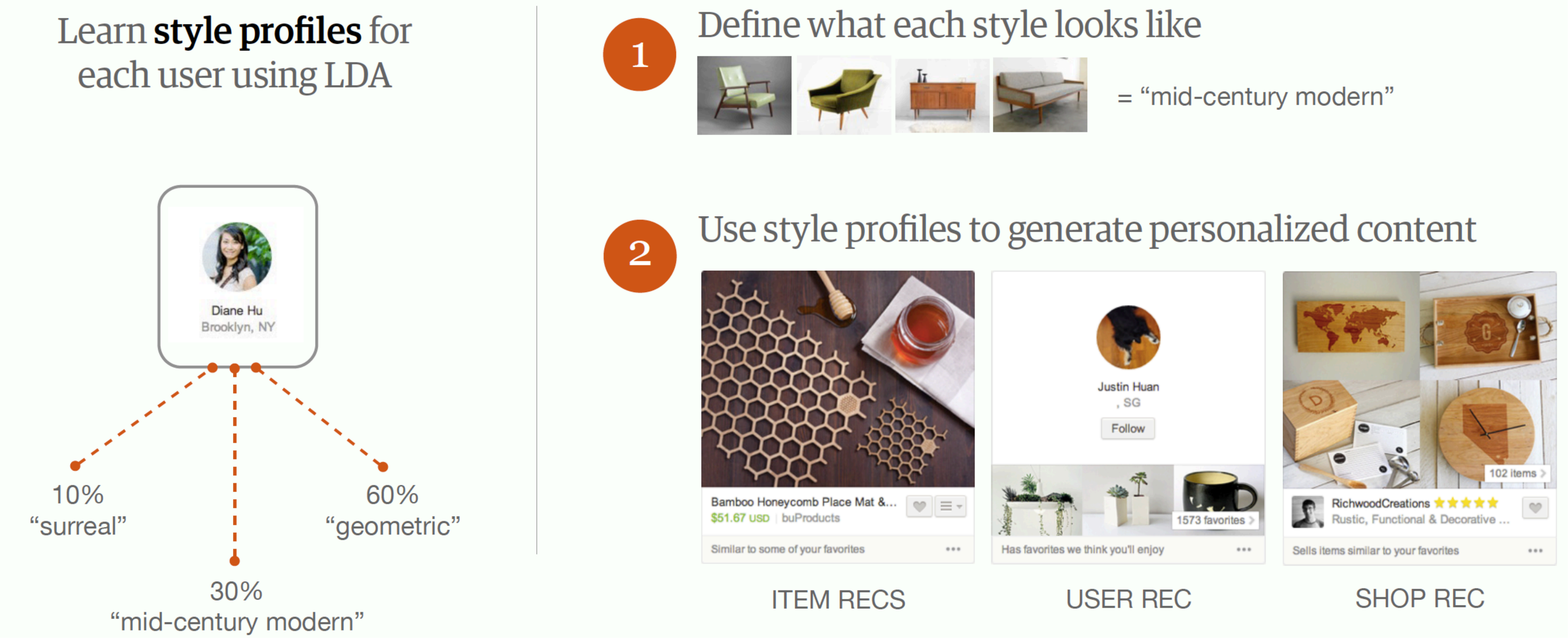
- Diane J. Hu, Rob Hall, and Josh Attenberg. **Style in the long tail: Discovering Unique Interests with Latent Variable Models in Large Scale Social E-Commerce**. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (**KDD '14**). ACM, New York, NY, USA,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

- **KDD 2014 Industrial Best Paper**



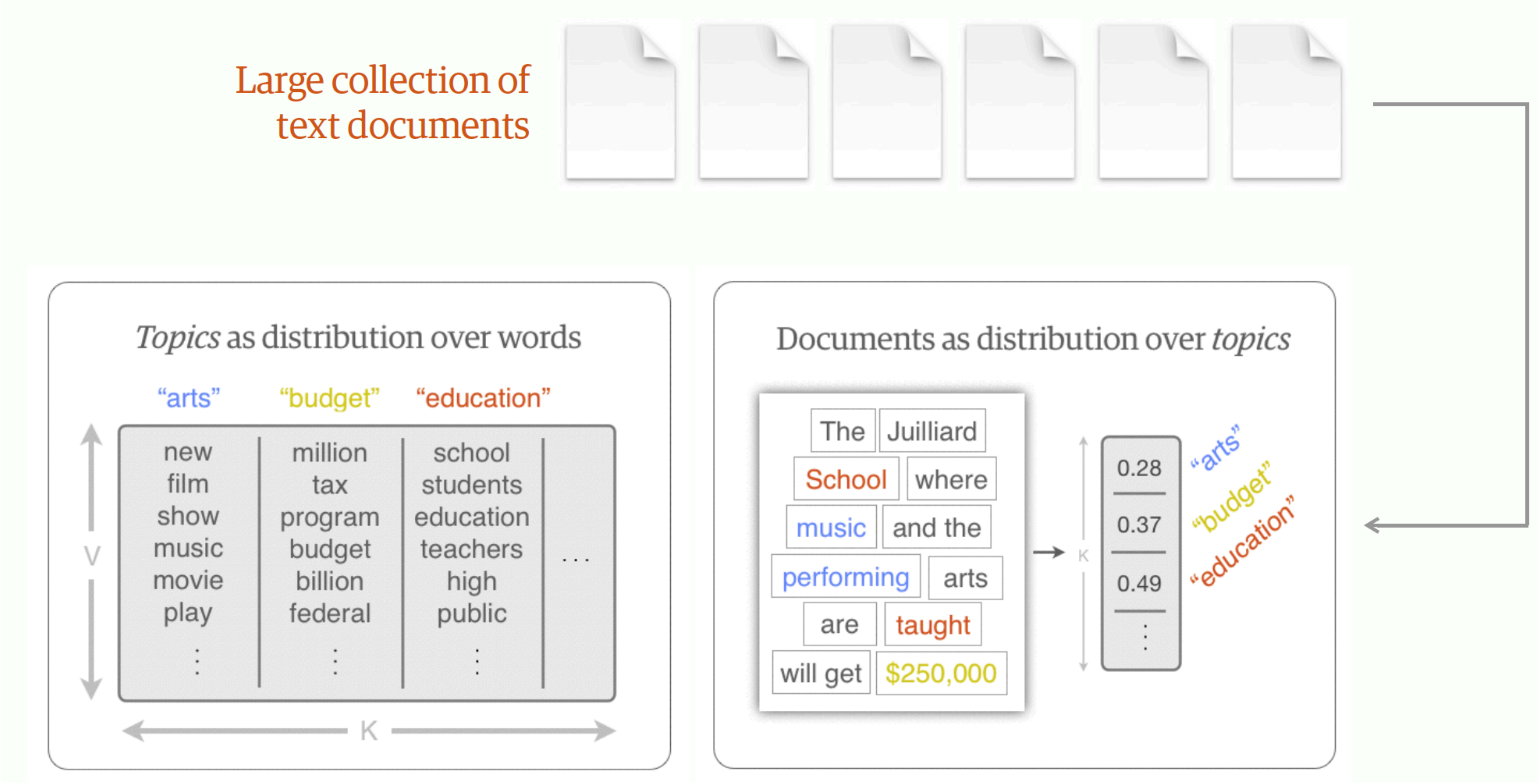
Etsy – Personalization and Recommendation

Latent Dirichlet Allocation



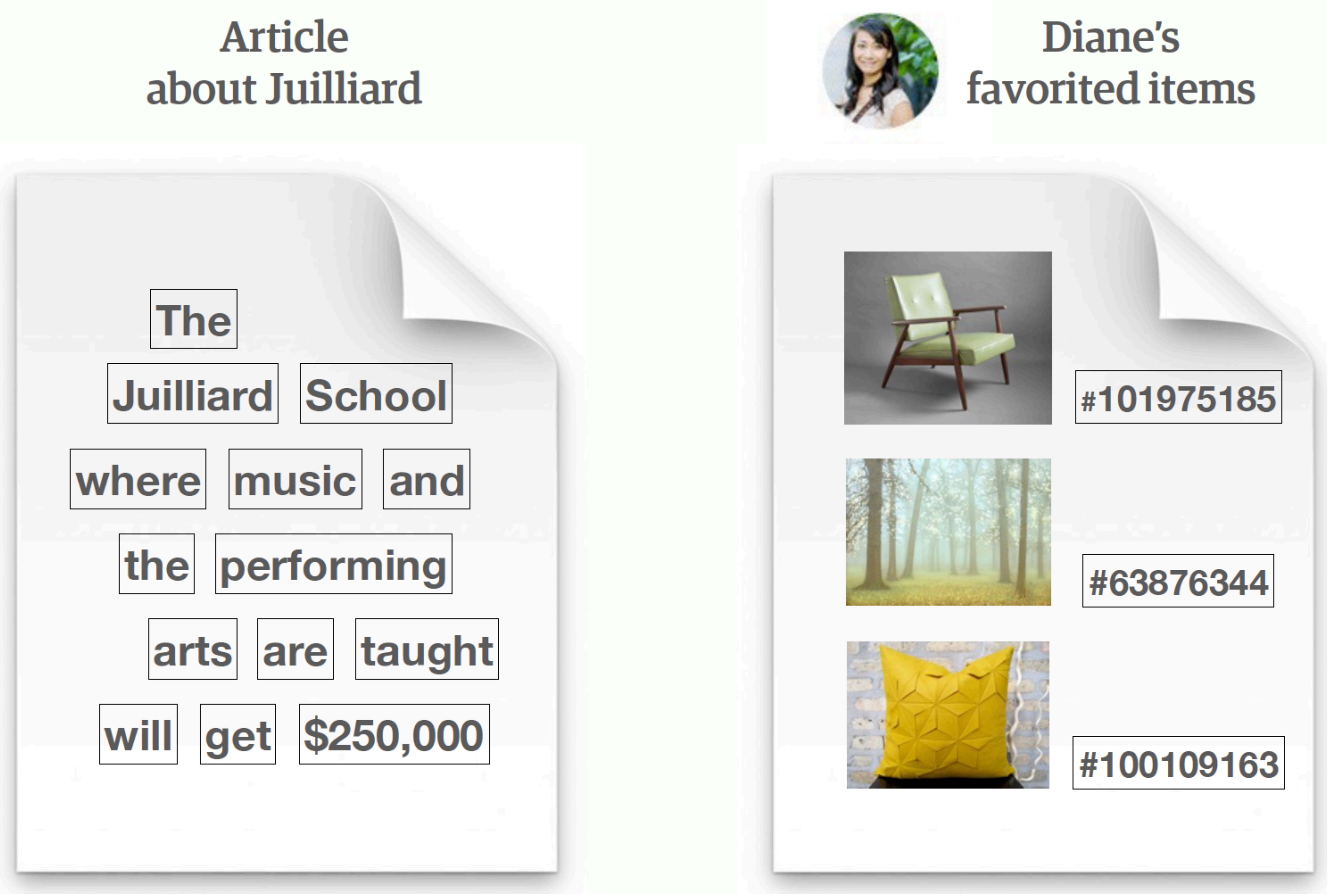
Etsy – Personalization and Recommendation

Latent Dirichlet Allocation



Etsy – Personalization and Recommendation

Latent Dirichlet Allocation



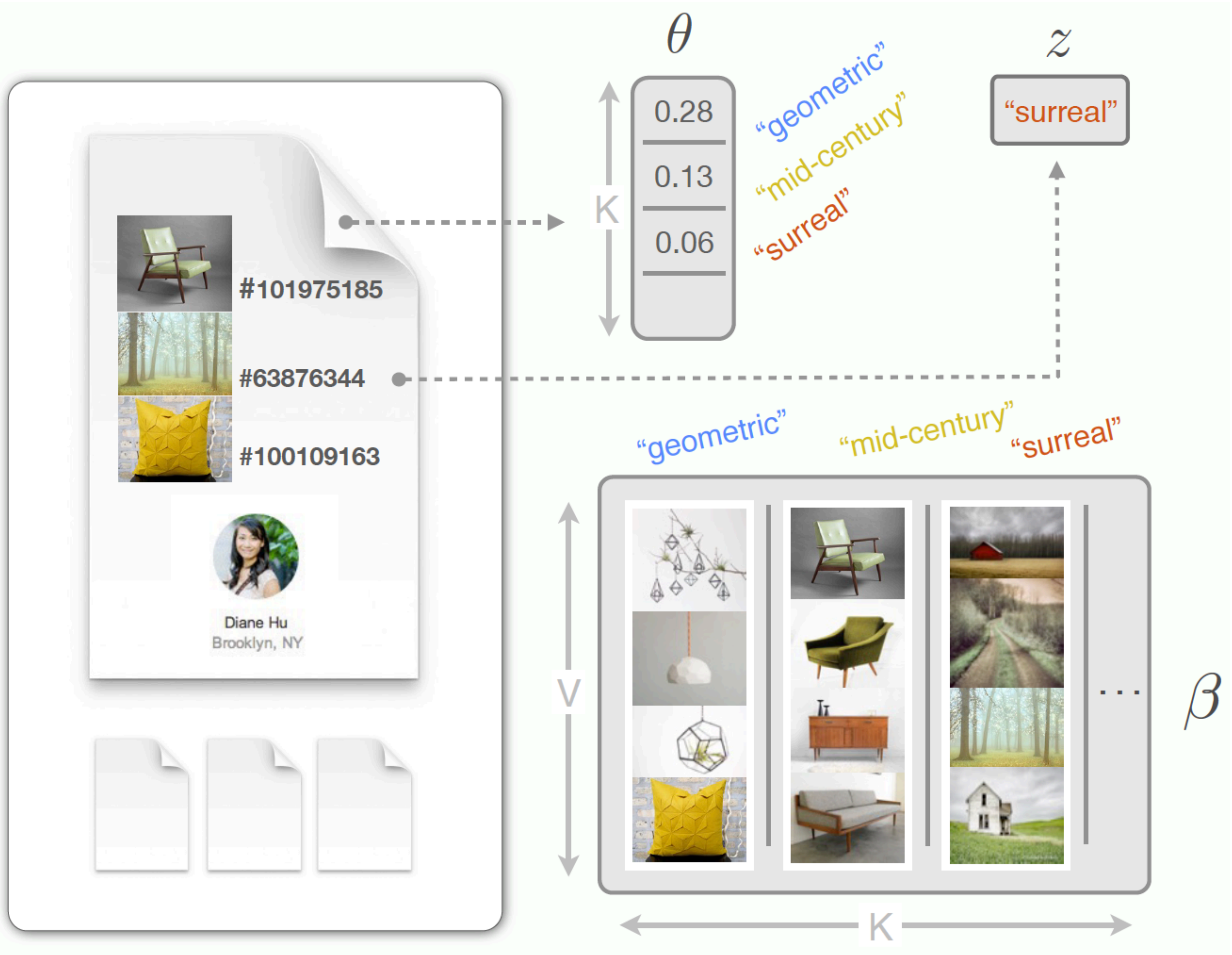
Etsy – Personalization and Recommendation

Latent Dirichlet Allocation

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

For each user u,

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



Etsy – Personalization and Recommendation

Latent Dirichlet Allocation

1) Inference:

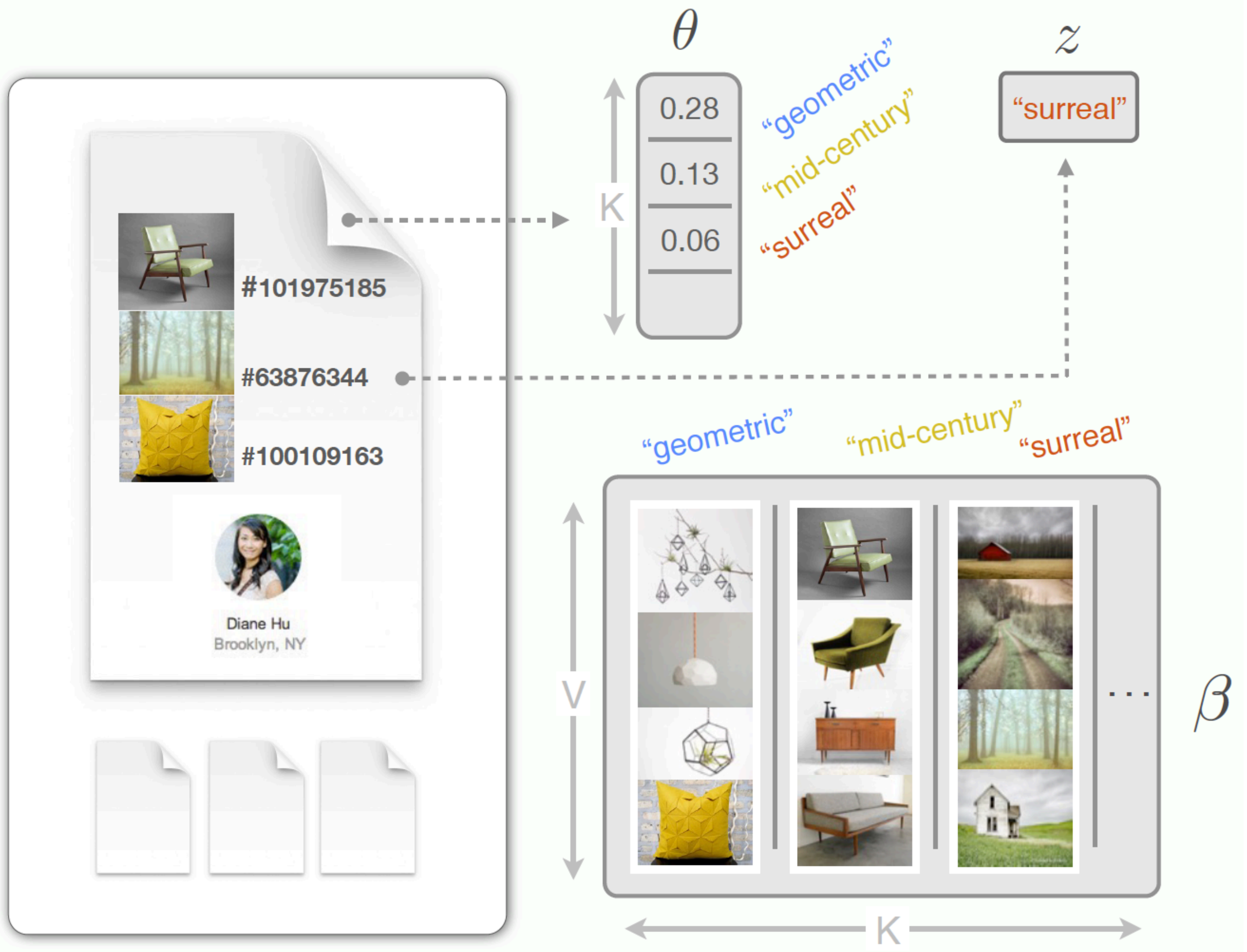
Determine posterior distribution:

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

2) Estimation:

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

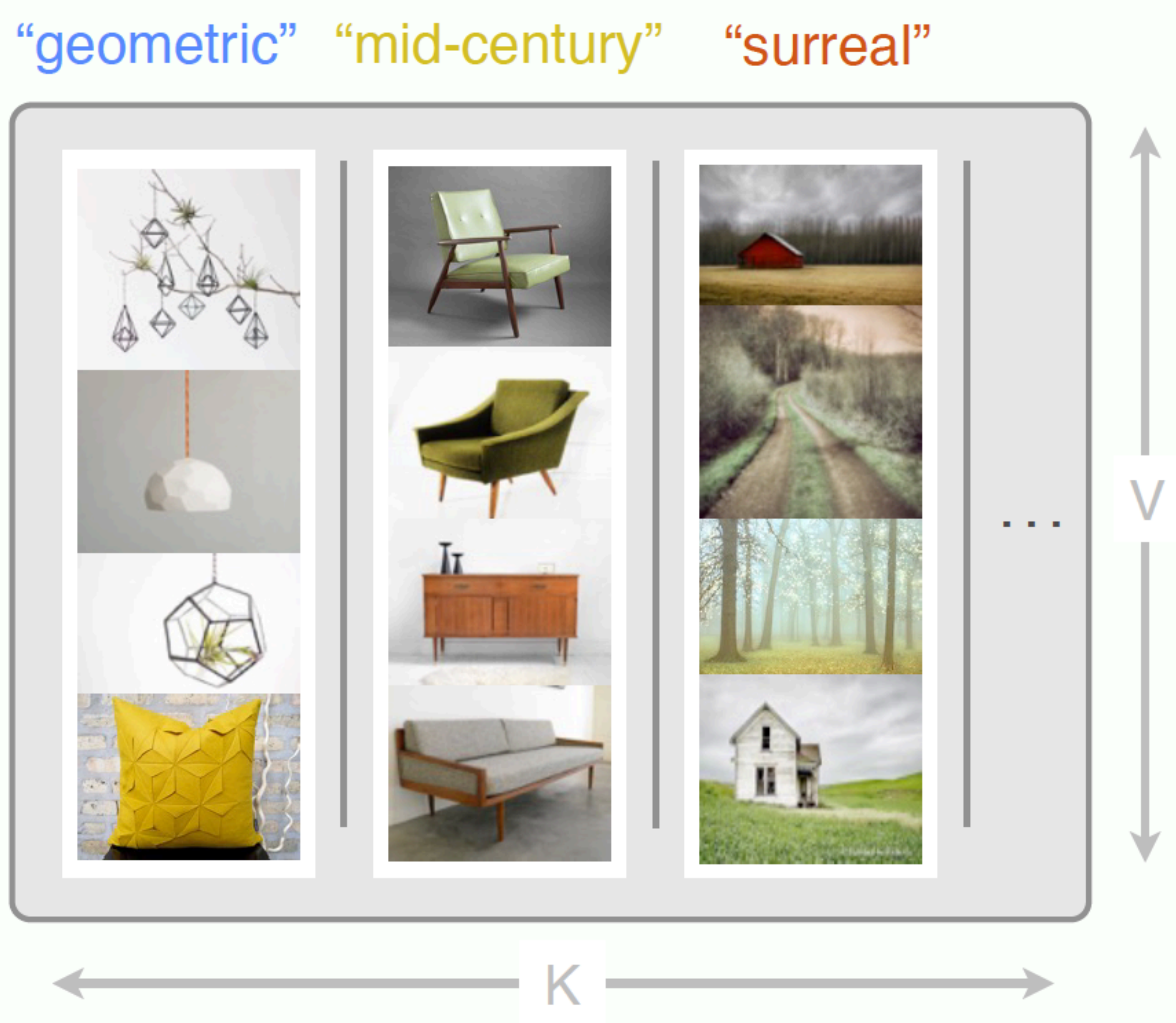
$$\mathcal{L}(\alpha, \beta) = \sum_{m=1}^{M} \log p(\mathbf{w} | \alpha, \beta)$$



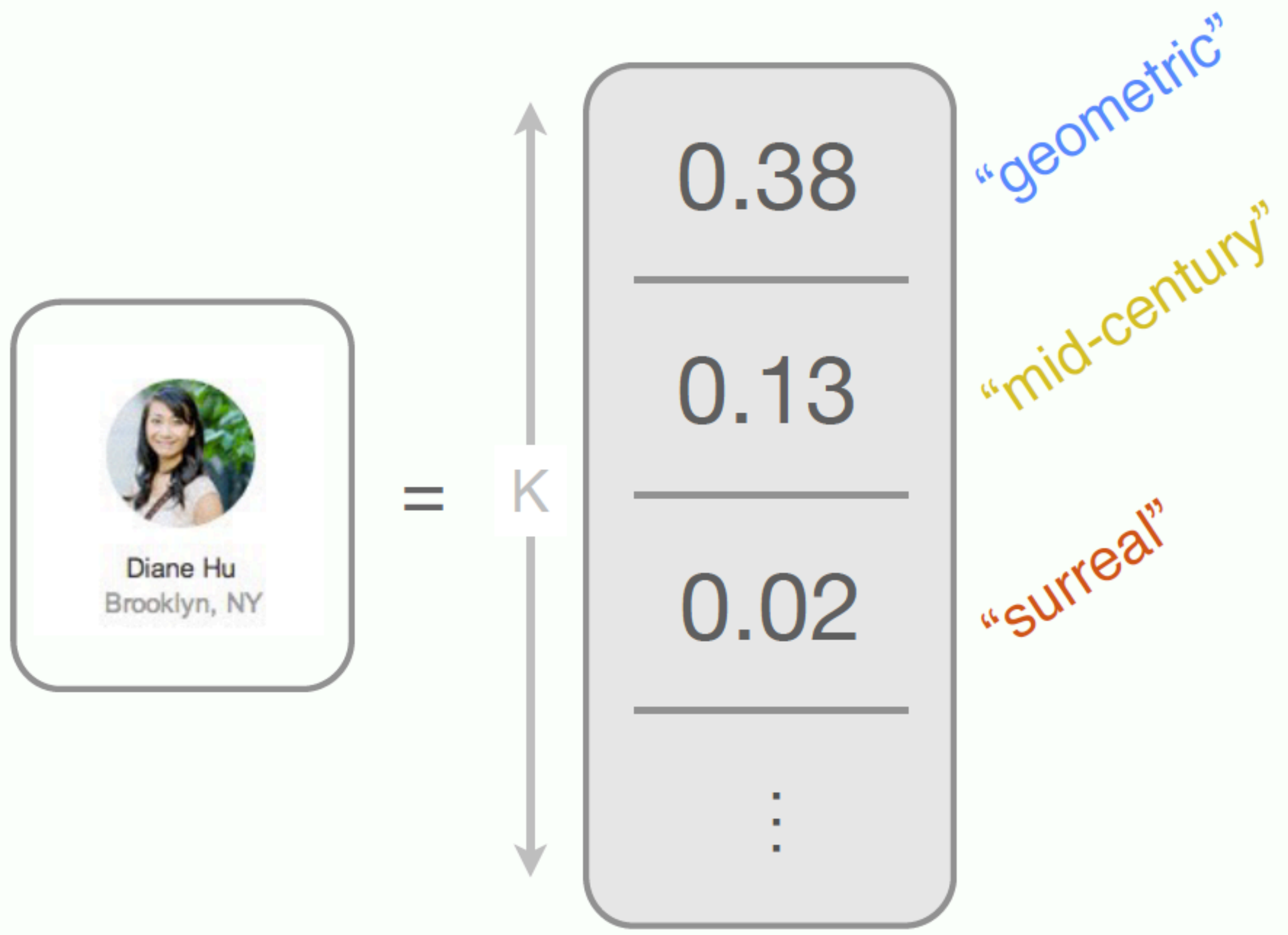
Etsy – Personalization and Recommendation

Latent Dirichlet Allocation

Discover popular styles on Etsy as a distribution over items

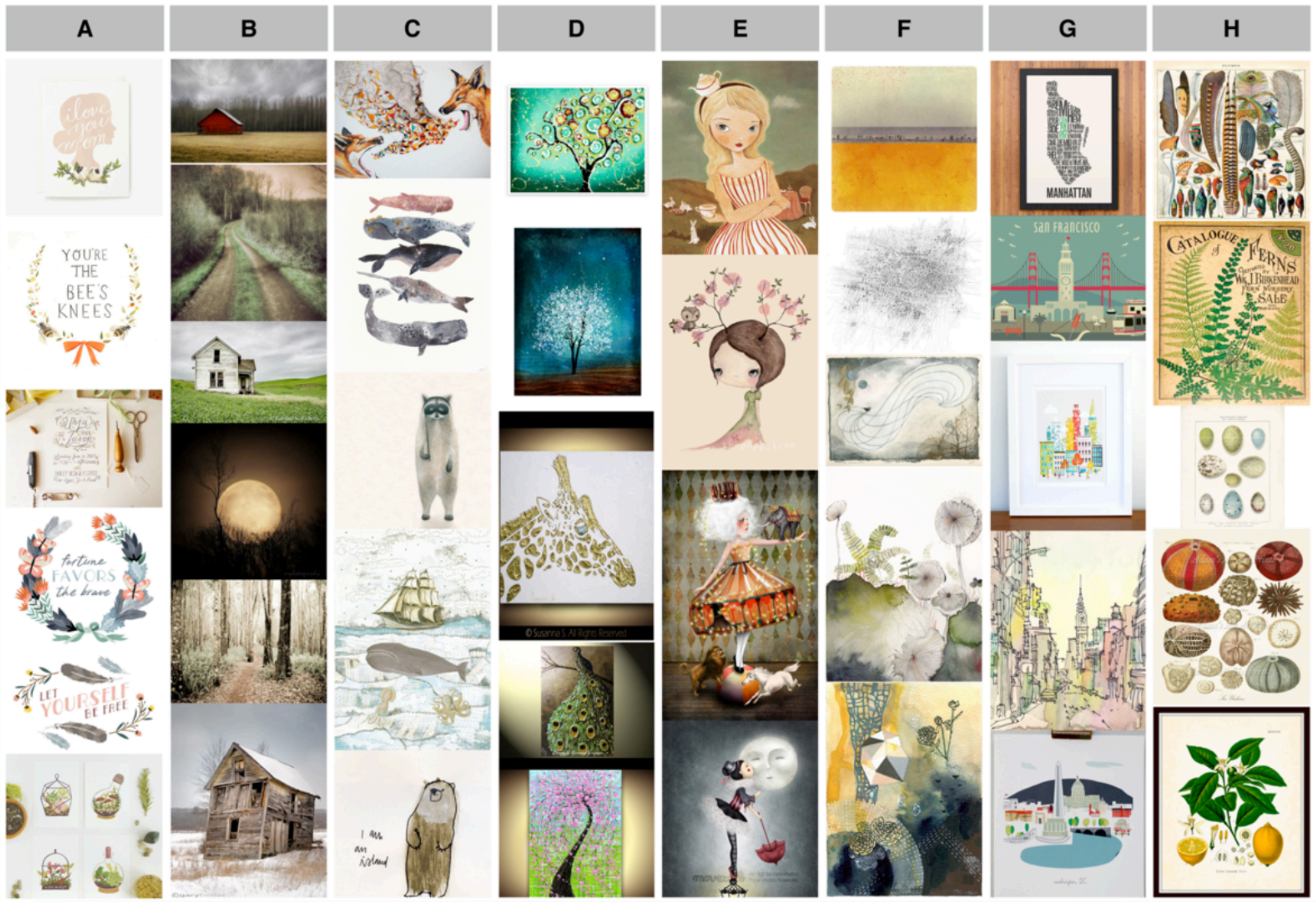


Represent each user as a distribution over popular styles, i.e. "**style profile**"



Etsy – Personalization and Recommendation

Learned Styles from LDA



Example of learned styles that contain art prints:

A = Botanical

B = Surreal landscapes

C = Whimsical

D = Acrylic/Abstract

E = French Dolls

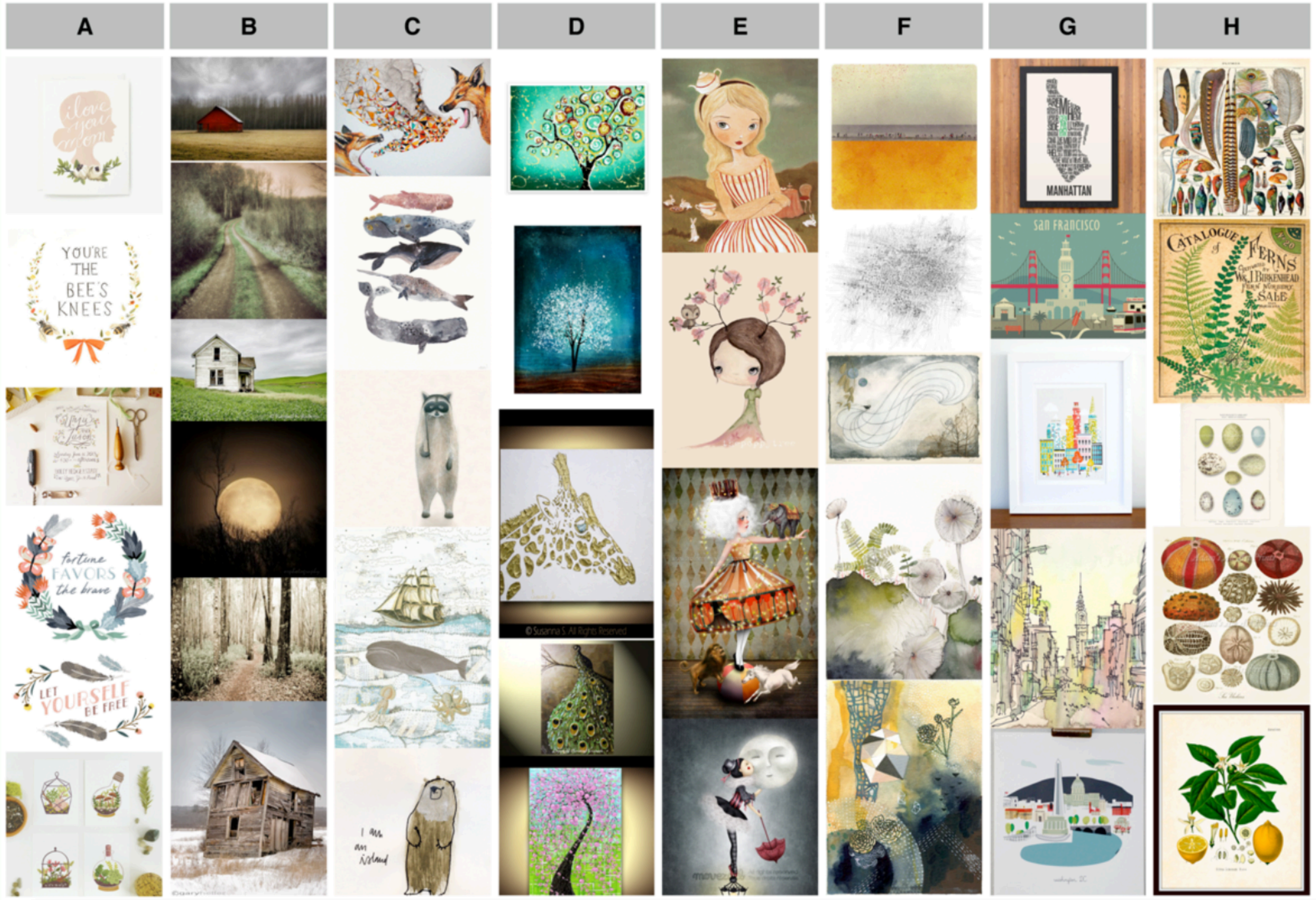
F = Whimsical/Abstract

G = Cities

H = Vintage

Etsy – Personalization and Recommendation

Learned Styles from LDA



Example of learned styles that contain art prints:

A = Botanical

B = Surreal landscapes

C = Whimsical

D = Acrylic/Abstract

E = French Dolls

F = Whimsical/Abstract

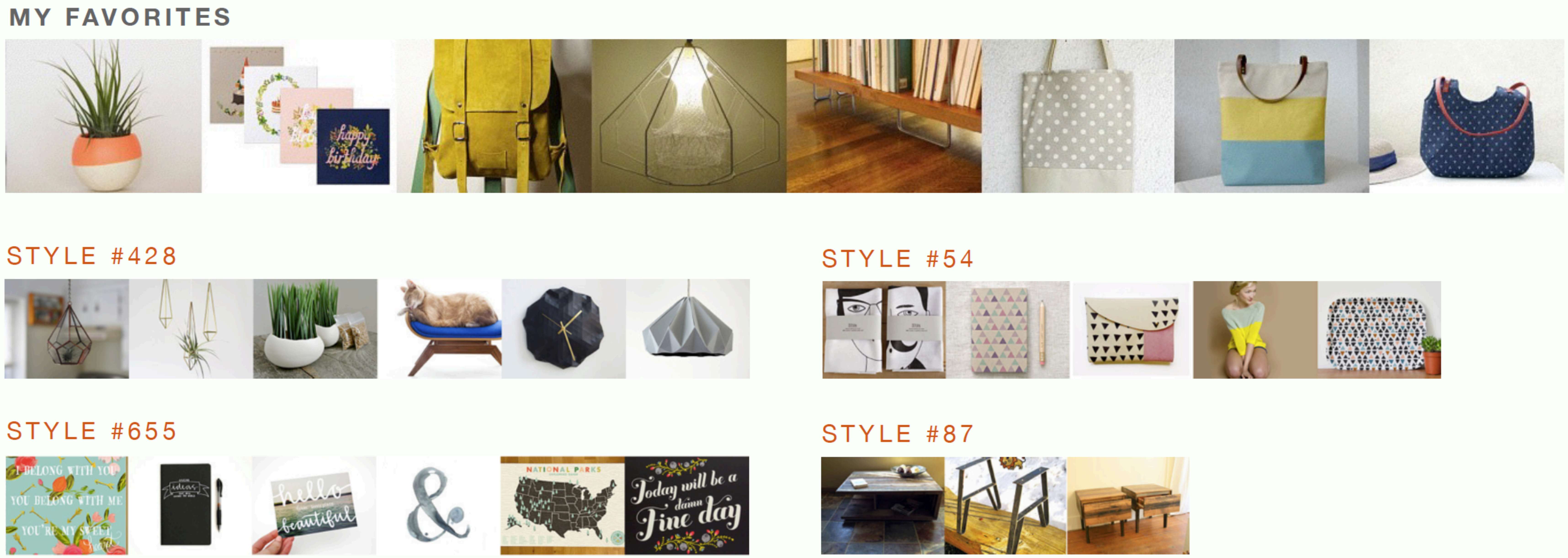
G = Cities

H = Vintage

Etsy – Personalization and Recommendation

Given that each user has an style profile:

Recommend N listings from most highly weighted styles



Etsy – Personalization and Recommendation

Given that each user has an style profile:

USER RECOMMENDATION

- Use LSH on MapReduce to do fast, approximate KNN
- Get top N users with most similar interest profiles
- *Recommend these top N users*

LISTING RECOMMENDATION

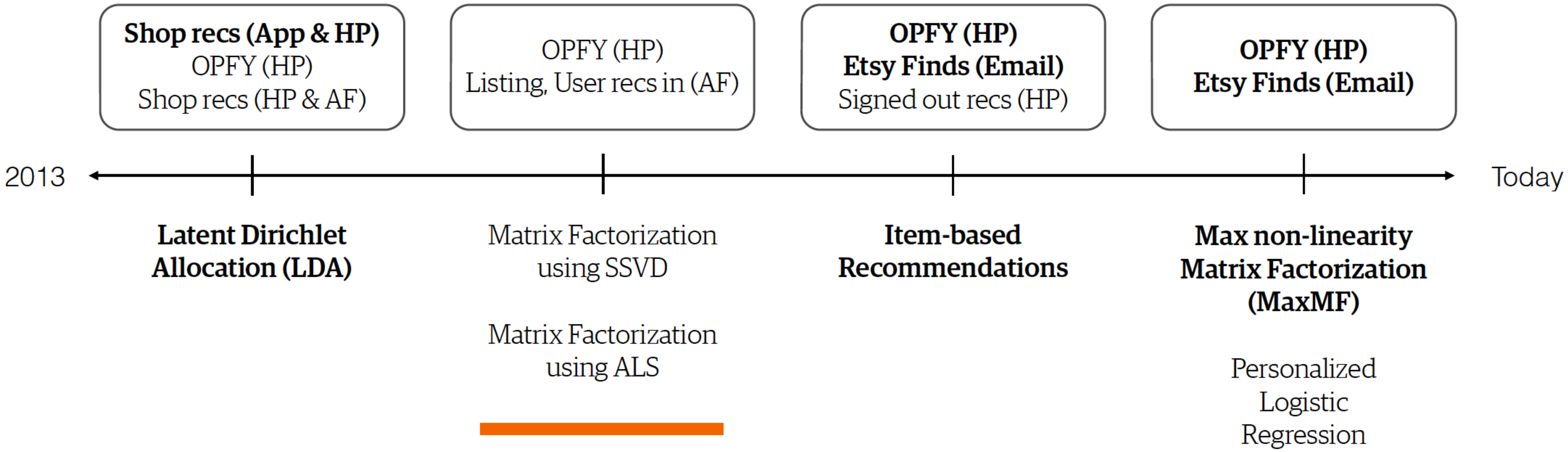
- Multiply user style profile by topic-item matrix
- Return list of ranked items for each user
- *Recommend top N listings*

SHOP RECOMMENDATION

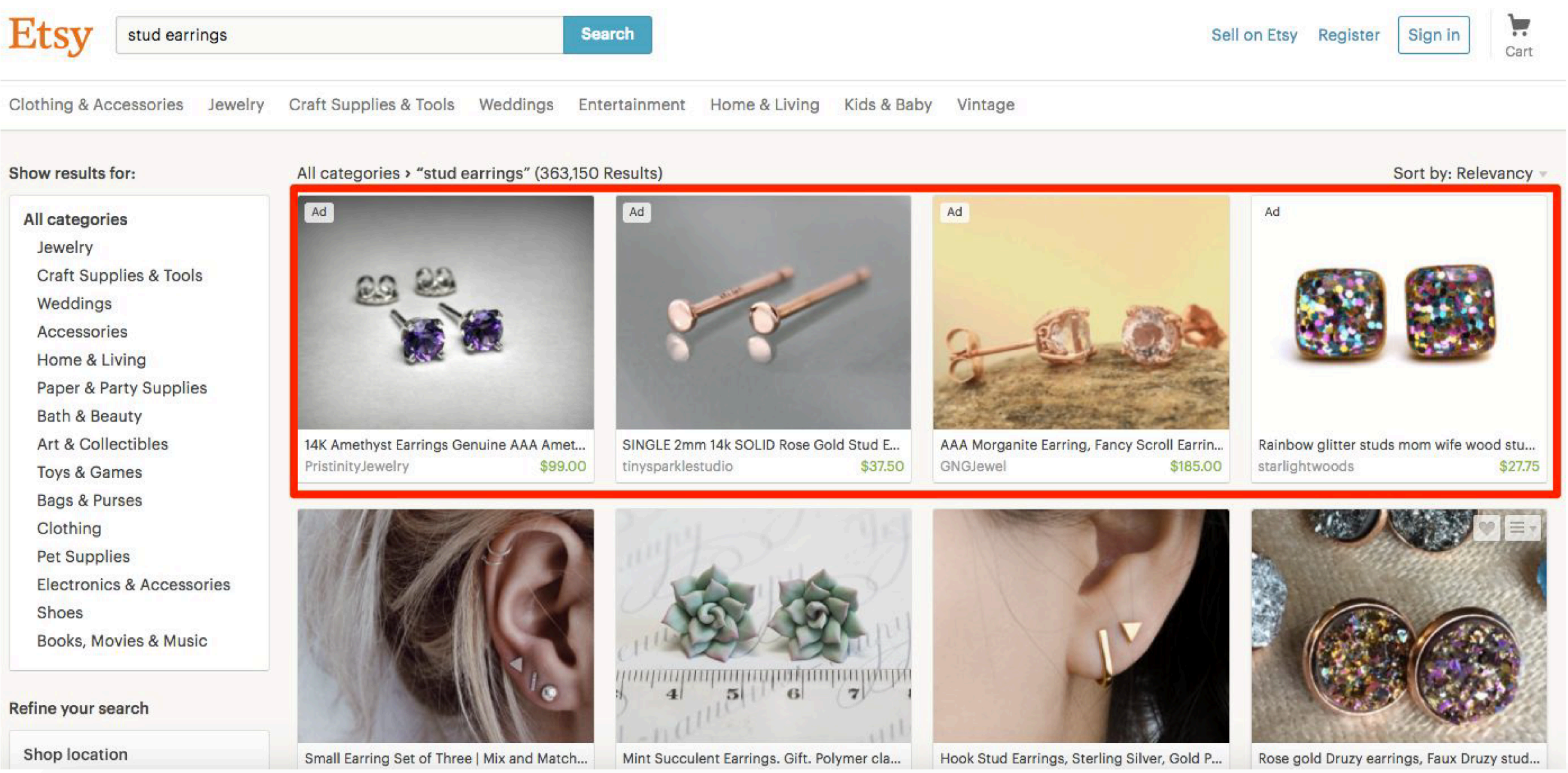
- Rebuild topic model replacing items with corresponding shop
- Get list of ranked shops, as in item recommendations
- *Recommend top N shops*

Etsy – Personalization and Recommendation

A Brief History



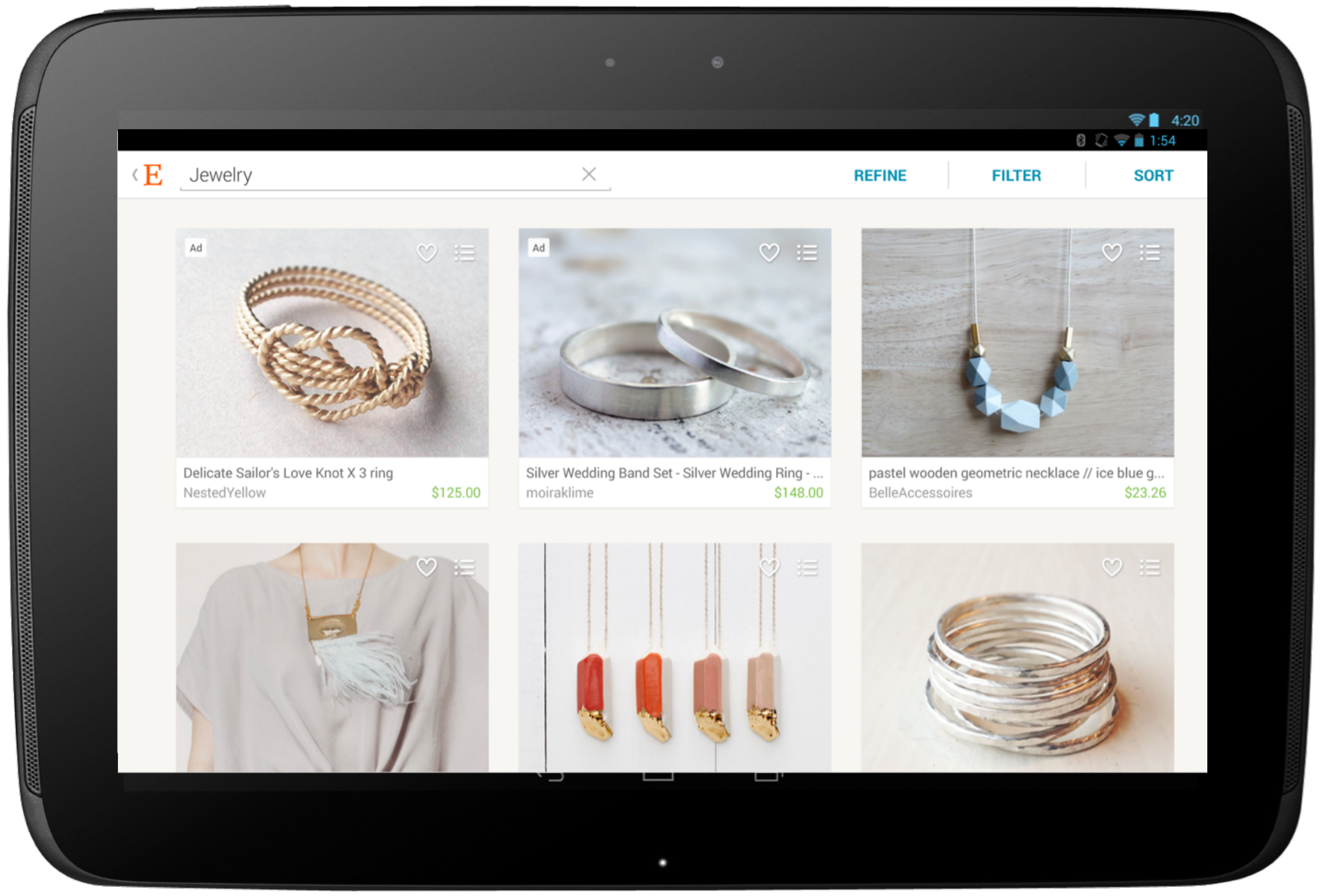
Etsy – Computational Advertising



Etsy – Computational Advertising

Large-scale Ads CTR Prediction System

- 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 – Computational Advertising

Large-scale Ads CTR Prediction System

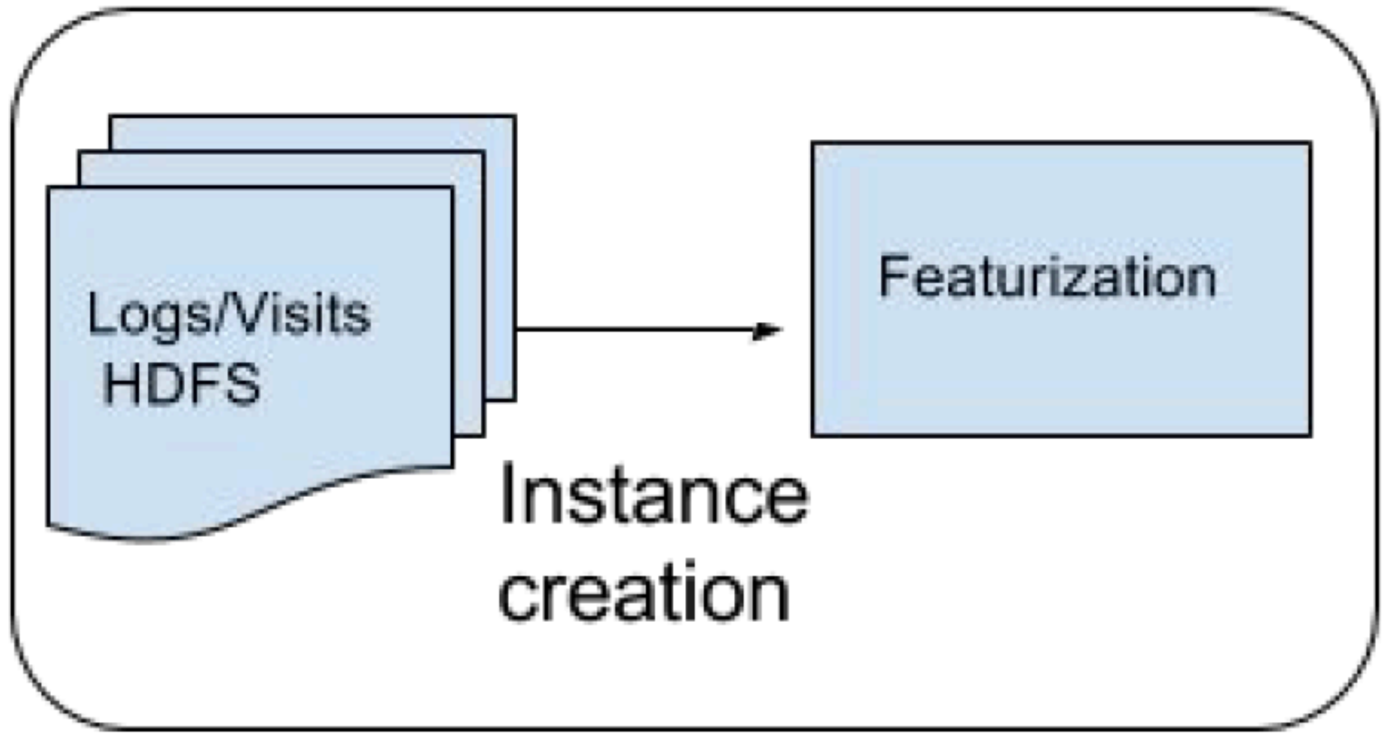
- Data Aggregation
- Feature Engineering
- Learning Algorithm
- Ensemble Learning

Etsy – Computational Advertising

Training Data: 14 days Organic Search Data or 30 Days Promoted Listings Data

Balanced Sampling

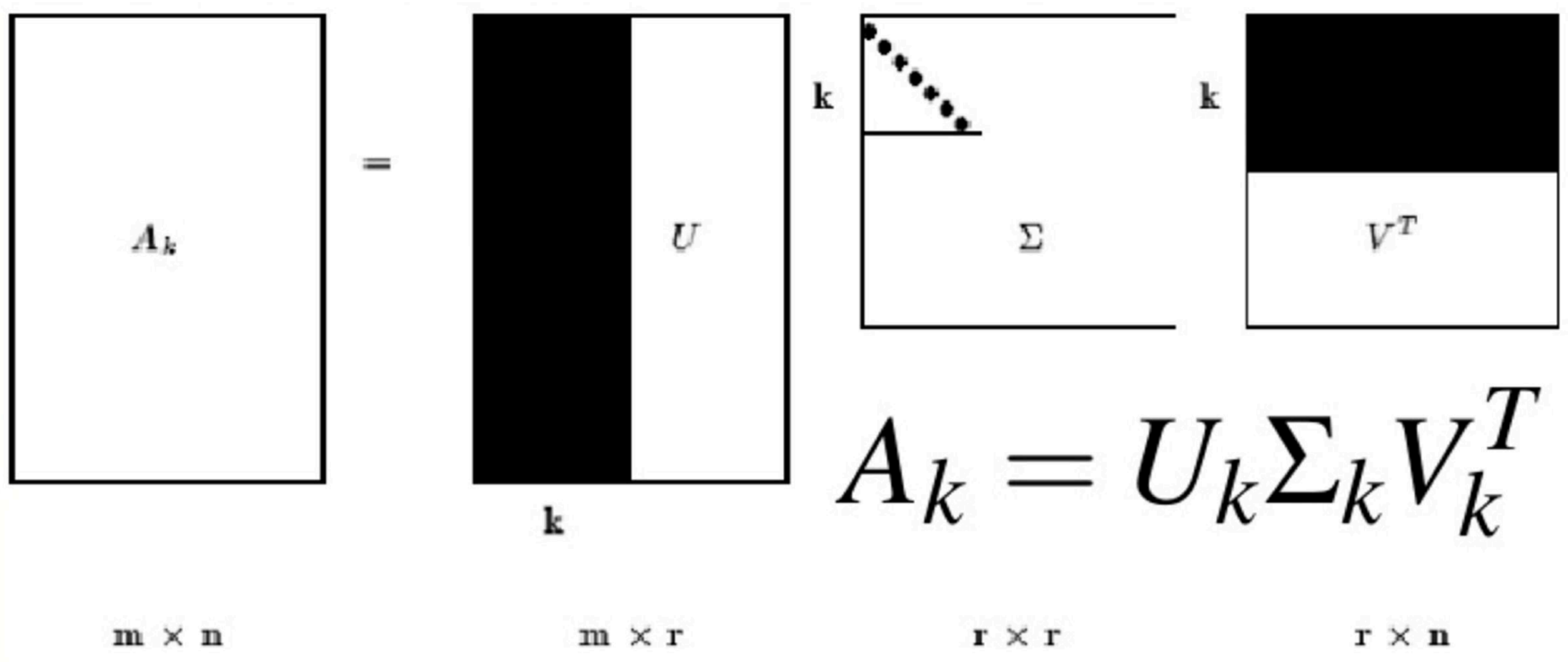
Evaluation Data: Previous Day Promoted Listings Data



Etsy – Computational Advertising

Feature Engineering

- Simple unigram text features
- Seller, ads, user historical information
- Ads information like price, inventory amount
- SVD-based features
- Utilizing feature hashing



Etsy – Computational Advertising

Logistic Regression with FTRL-Proximal Background

Features: vector $\mathbf{x}_t \in \mathbb{R}^d$

model parameters $\mathbf{w}_t$

Prediction = $\sigma(\mathbf{w}_t \cdot \mathbf{x}_t)$ where $\sigma(a) = 1/(1 + \exp(-a))$

we observe the label $y_t \in \{0, 1\}$

$$\ell_t(\mathbf{w}_t) = -y_t \log p_t - (1 - y_t) \log(1 - p_t),$$

Etsy – Computational Advertising

Logistic Regression with FTRL-Proximal

- **FTRL-Proximal is equivalent to Online (Stochastic) Gradient Descent when no regularization is used**
- **The key is re-expressing the gradient descent as simply as possible:**

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

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

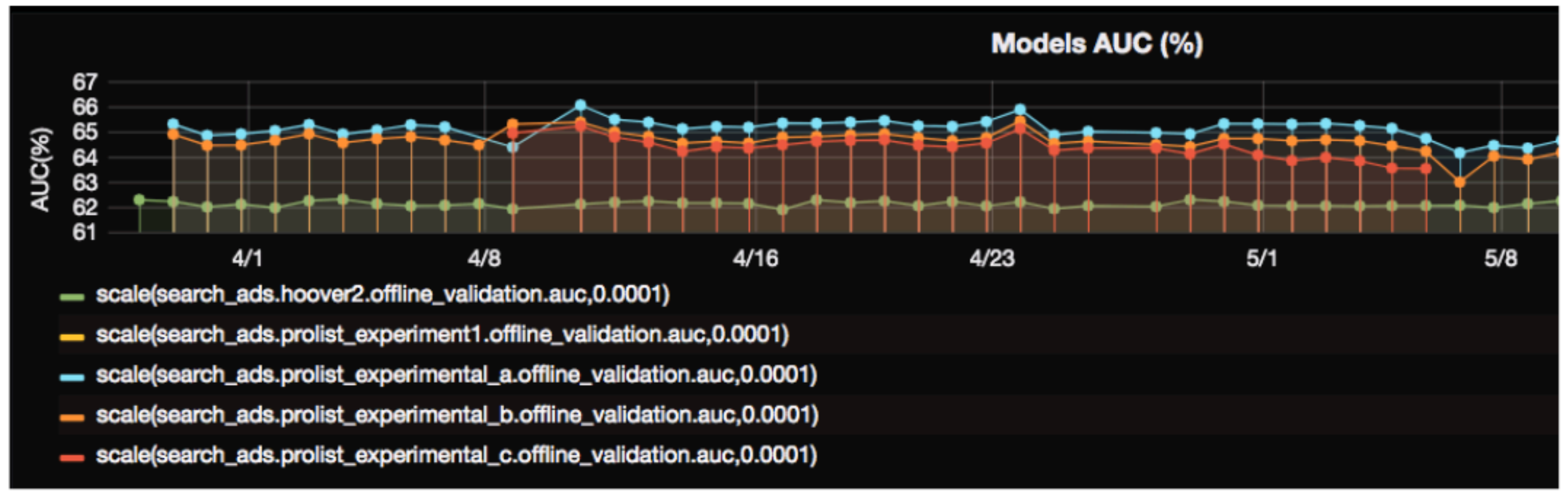
Etsy – Computational Advertising

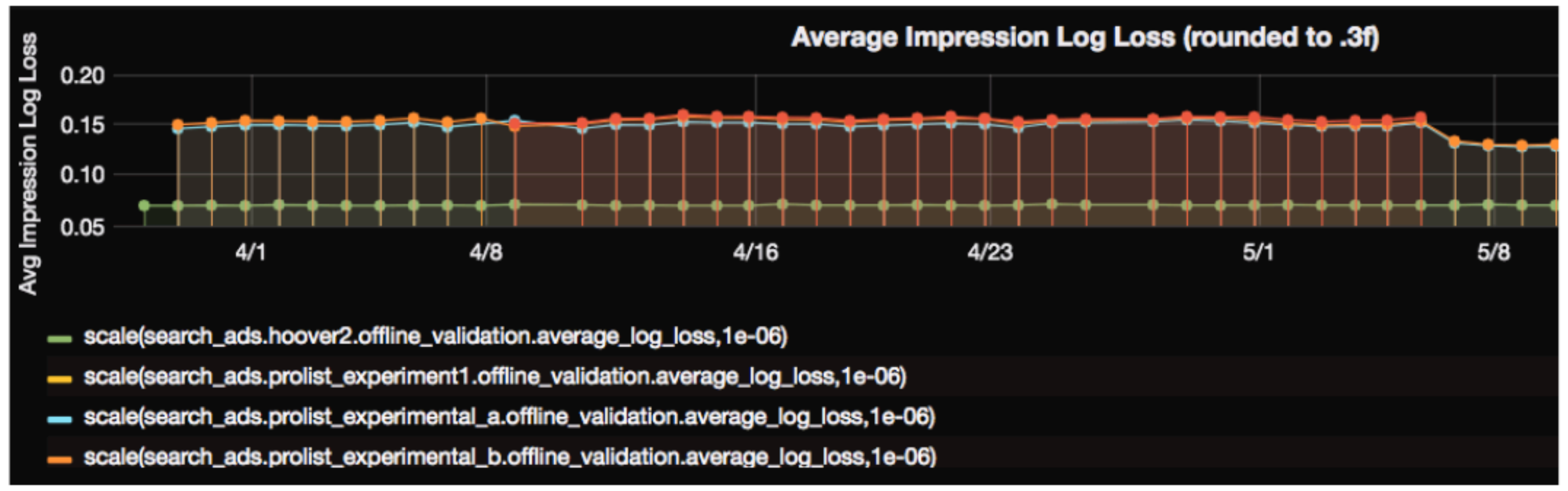
Ensemble Learning

- Cold Start
- Warm Start
- Mix

Etsy – Computational Advertising

	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**





AI in E-commerce

AI in E-commerce at Etsy

- Multi-modal Deep-learning based Search Solution (KDD 2016)
- Probabilistic Graphical Model based Personalization Recommendation (KDD 2014)
- Ensemble Learning based CTR Prediction Solution (AdKDD 2017/KDD 2017)

Thanks!