Do "Also-Viewed" Products Help User Rating Prediction?

Chanyoung Park, Donghyun Kim, Jinoh Oh, Hwanjo Yu*
Data Mining Lab., POSTECH, Korea

* Corresponding author

User Rating Prediction

Given: Ratings given by users on products

Goal: Predict users' ratings for other products

How to accurately predict ratings?

- We need to obtain **high-quality product embedding vector**
- Various <u>*side information*</u> is helpful
  - Social network, Review text, temporal dynamics
- Recently, the **appearance of a product** has been shown to be helpful for learning product embedding [AAAI 16, IJCAI 16, RecSys 16]
  - Product image as *side information*
  - Visually-aware recommender system

Motivation of visually-aware RS

- Product appearance is very influential to users in **clothing domain**

Never buy a T-shirt without looking at how it looks like

What about in **other domains** such as pet supplies / office products?

POSTECH

Limitation of visually-aware RS

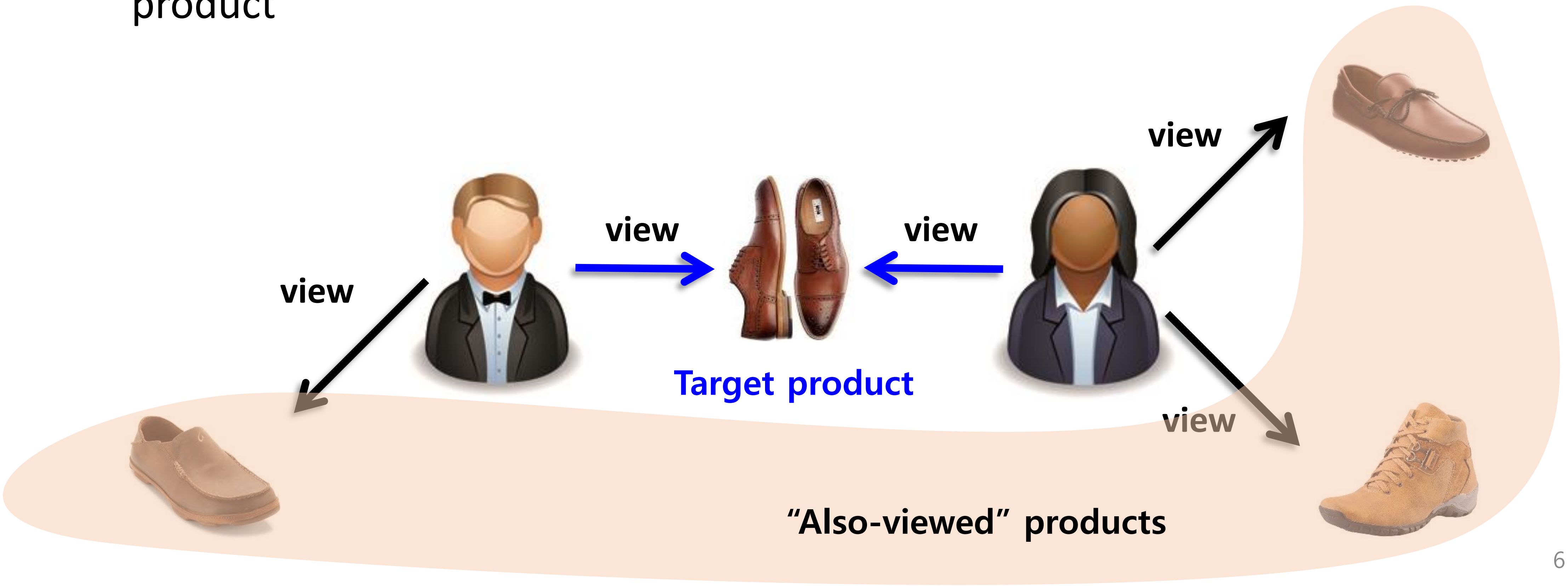
- **In other domains, variety of product aspects** besides appearance are more important
  - e.g., functionality and specification

- Even for clothing, the appearance of a product is not the only factor that influences users

- **Let’s consider other product aspects as well!**

Data exploration: *"Also-viewed"* Products

- What is ***"also-viewed"*** products?
  - A list of products that has also been viewed by users who have viewed a product



Data exploration: *“Also-viewed”* Products

Users pay more attention to different aspects of products in different product domains

- *“Also-viewed”* Products
  - Clothing domain
    - Visually look similar
  - Non-Clothing domain
    - Visually not similar but functionally related

Example

- Target product
  - **Liquid** Flea repellent
- *“Also-viewed”* products
  - flea killing capsules and flea traps

Related Work: Visually-aware recommender system

- VBPR: Visual Bayesian Personalized Ranking from Implicit Feedback [AAAI 16]
  - Incorporates visual features extracted from product images into Matrix Factorization

Modeling rating

$$\hat{r}_{ij} = U_i^T V_j + P_i^T Q_j$$
$$= U_i^T V_j + P_i^T \boldsymbol{E} f_j$$

r_{ij}: ratings given by user i on product j

$\mathrm{U_i}$: user latent vector ($1 \times L$)

V_j: product latent vector ($1 \times L$)

$\mathrm{P_i}$: user visual latent vector ($1 \times D$)

Q_j: product visual latent vector ($1 \times D$)

$\boldsymbol{E}$: Embedding kernel (for dimension reduction K → D)

f_j: visual feature vector extracted from CNN ($1 \times F$)

Related Work: Visually-aware recommender system

- Limitation
  1. Showed only a slight improvement in domains where aspects such as functionality and specifications are significant for user ratings
     - **Cannot be generalized to *visually non-aware* product domains**
  2. Models users' visual preferences only based on the images of products rated by them in the past, which are very few.
     - **Cannot handle the *data sparsity* issue**

Solution?

Visual feature + ***"also-viewed" products*** information!

Why *"also-viewed" products*?

1. Encode not only visual similarity, but also functional or specification-related similarity
   - Clothing domain
     - Captures aspects overlooked by visual features of a product
   - Non-Clothing product domain
     - Explicit relationships among products expressed through *"also-viewed"* information are even more helpful than visual features (visual features are not important anyways)

2. Solve the data sparsity problem (Most products are unrated)
   - Explicit relationships between each **rated product** and its **unrated *"also-viewed"* products** help us reflect various aspects of products in the product embedding

Method

Building Product-affinity network

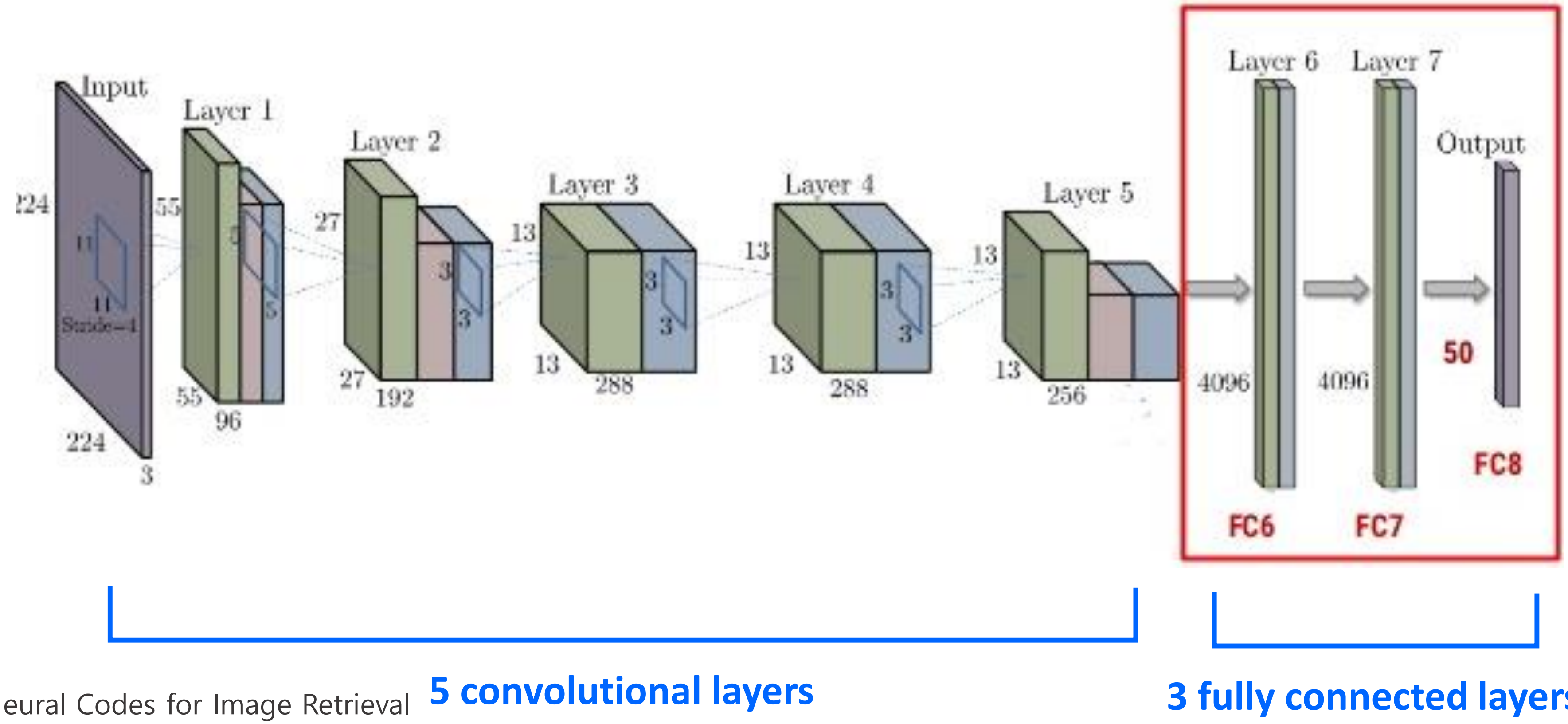
- Directed graph

$$s_{jk} = \begin{cases} 1 & k \in N_j. \\ 0 & \text{otherwise.} \end{cases}$$

S: Product-affinity matrix
N_j: *"Also-viewed"* products of product *j*

Extracting Visual Features

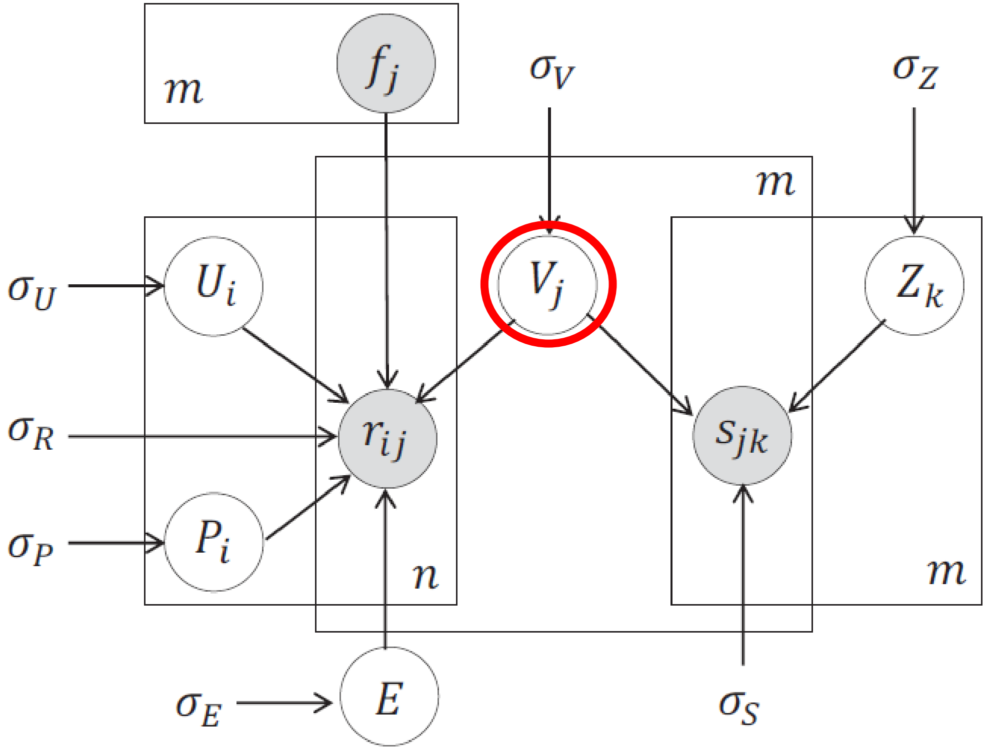
- Deep Convolutional Neural Network [NIPS 12]
  - For each product image, extract output of FC7 of length $F = 4096$



Source: Babenko et.al, Neural Codes for Image Retrieval

Graphical Model

- Matrix co-factorization by sharing product embedding V



- Modeling rating

$$\hat{r}_{ij} = U_i^T V_j + P_i^T \boldsymbol{E} f_j$$

Extracted from CNN

- Modeling relationship

$$\hat{s}_{jk} = V_j^T Z_k$$

Visual Matrix Co-Factorization (VMCF)

- Loss function: VMCF

$$L(R, S, U, V, P, Z, \mathbf{E}) =$$

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m} I_{ij}^{R}\Big(r_{ij} - g(U_i^T V_j + P_i^T \mathbf{E} f_j)\Big)^2 + \frac{\lambda_S}{2}\sum_{j=1}^{m}\sum_{k=1}^{m} I_{jk}^{S}\Big(s_{jk} - g(V_j^T Z_k)\Big)^2$$

$$+ \frac{\lambda_U}{2}||U||_F^2 + \frac{\lambda_V}{2}||V||_F^2 + \frac{\lambda_P}{2}||P||_F^2 + \frac{\lambda_Z}{2}||Z||_F^2 + \frac{\lambda_{\mathbf{E}}}{2}||\mathbf{E}||_F^2$$

Reduced Model (MCF)

- When images of products are not available, the final objective function is reduced to:

$$L(R,S,U,V,Z) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m} I_{ij}^{R}\Big(r_{ij} - g(U_i^T V_j)\Big)^2 + \frac{\lambda_S}{2}\sum_{j=1}^{m}\sum_{k=1}^{m} I_{jk}^{S}\Big(s_{jk} - g(V_j^T Z_k)\Big)^2$$

$$+ \frac{\lambda_U}{2}||U||_F^2 + \frac{\lambda_V}{2}||V||_F^2 + \frac{\lambda_Z}{2}||Z||_F^2$$

- **In visually non-aware domains,**
  - *Importance:* Functionality/specification >> Appearance
  - Therefore, MCF is still beneficial

Experiments

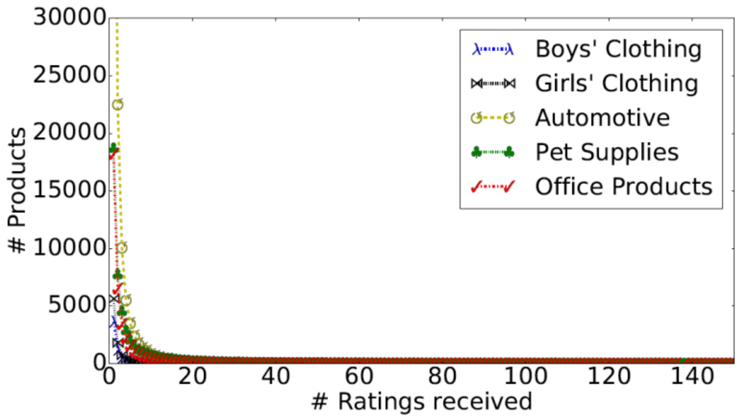
Dataset

- Data statistics
  - 5 public datasets extracted from *Amazon.com*
    - http://jmcauley.ucsd.edu/data/amazon/
  - Includes URLs for product image, price, a list of also bought product, a list of also viewed products, and etc.
  - We use the lists of *also viewed products* to construct product-affinity network

Dataset	#Users	#Products	#Ratings	#Relations
Boys' Clothing	4,496	6,391	15,997	31,370
Girls' Clothing	5,941	9,549	22,524	51,990
Automotive	84,418	126,934	406,852	2,162,853
Pet Supplies	85,115	49,048	427,543	1,066,131
Office Products	50,570	40,181	240,146	672,586

Evaluation Protocol

- "All" evaluation
  - Tested on all samples in the test dataset
- "Cold-start" evaluation
  - Ratings in the test dataset are sampled such that each product (cold-product) has fewer than four ratings in the training dataset.



Most products received very few ratings

→ *Evaluations on cold-products are in fact more crucial than evaluations on every product*

Comparison Methods

- ItemCF
- PMF [NIPS 08]
  - Matrix factorization-based recommendation method
- VMF [AAAI 16]
  - A visually-aware matrix factorization-based method

Baselines	Personalized?	Visually-Aware?	Incorporate *"also-viewed"*?
ItemCF	X	X	X
PMF	O	X	X
VMF	O	O	X
MCF	O	X	O
VMCF	O	O	O

POSTECH

Comparison Methods

Baselines	Personalized?	Visually-Aware?	Incorporate *"also-viewed"*?
ItemCF	X	X	X
PMF	O	X	X
VMF	O	O	X
MCF	O	X	O
VMCF	O	O	O

- PMF vs. VMF
  - To verify that **appearance is more significant in clothing domain** than other domains

Comparison Methods

Baselines	Personalized?	Visually-Aware?	Incorporate *"also-viewed"*?
ItemCF	X	X	X
PMF	O	X	X
VMF	O	O	X
MCF	O	X	O
VMCF	O	O	O

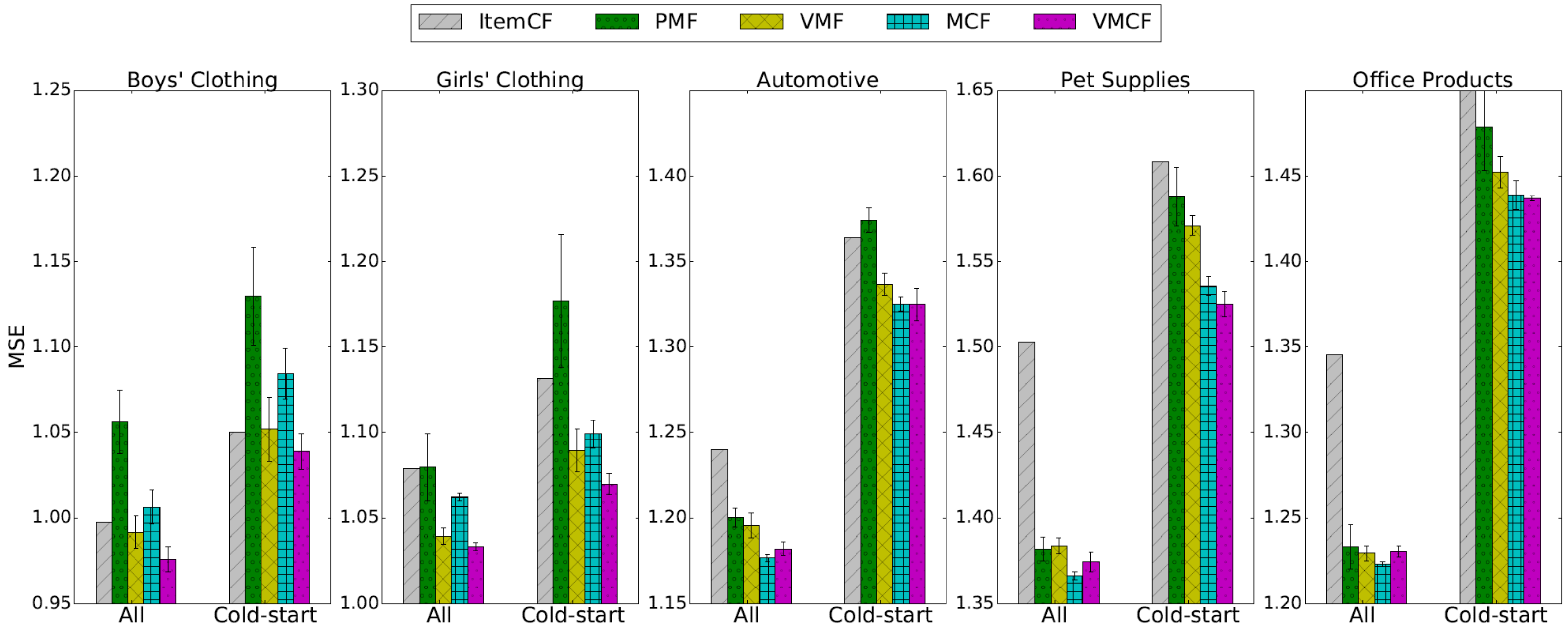
- VMF vs. MCF
  - To verify that **other aspects besides appearance are more significant in visually non-aware product domain** than in visually-aware product domain

Comparison Methods

Baselines	Personalized?	Visually-Aware?	Incorporate *"also-viewed"*?
ItemCF	X	X	X
PMF	O	X	X
VMF	O	O	X
MCF	O	X	O
VMCF	O	O	O

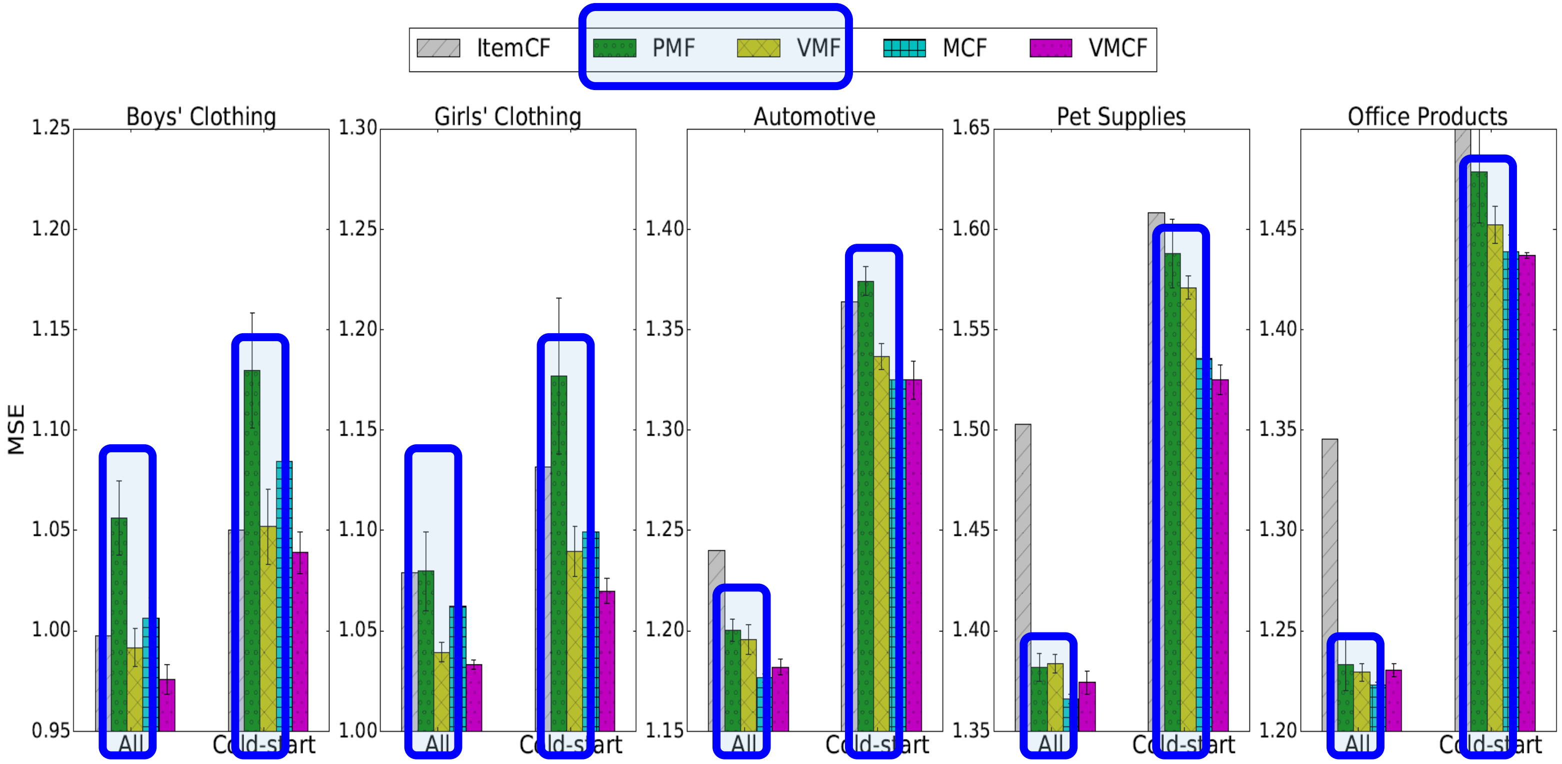
- VMF, MCF vs. VMCF
  - To demonstrate the **benefit of jointly modeling product images and *"also-viewed"* product information** in both visually-aware and visually non-aware product domain

Performance Analysis



Performance Analysis

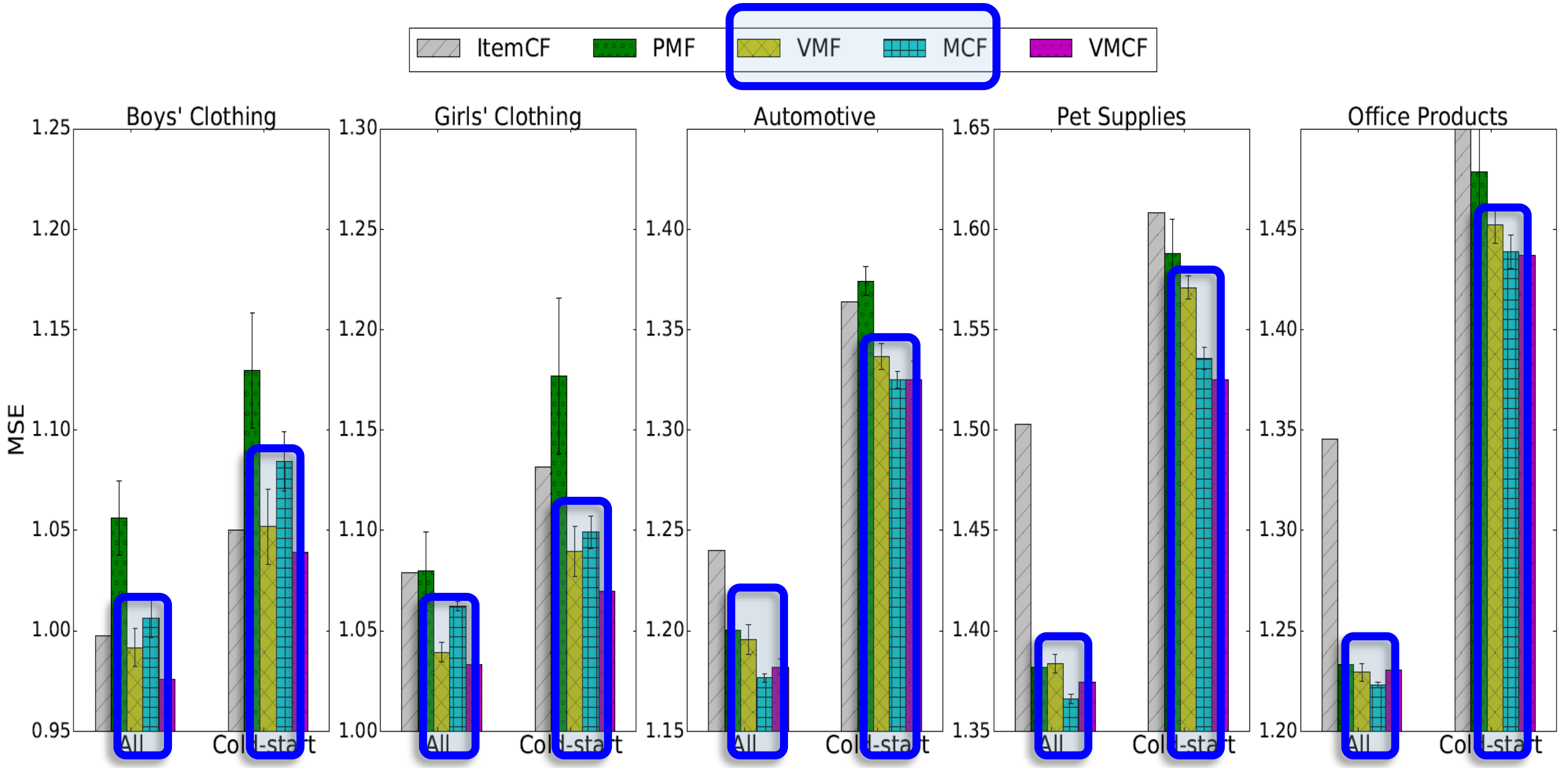
1. PMF vs. VMF



- Observation
  - Benefit is more significant in visually-aware product domain
- Analysis
  - Appearance plays a more significant role in visually-aware product domain
  - Although other product aspects are more influential in visually non-aware product domain, visual features are indeed still helpful

Performance Analysis

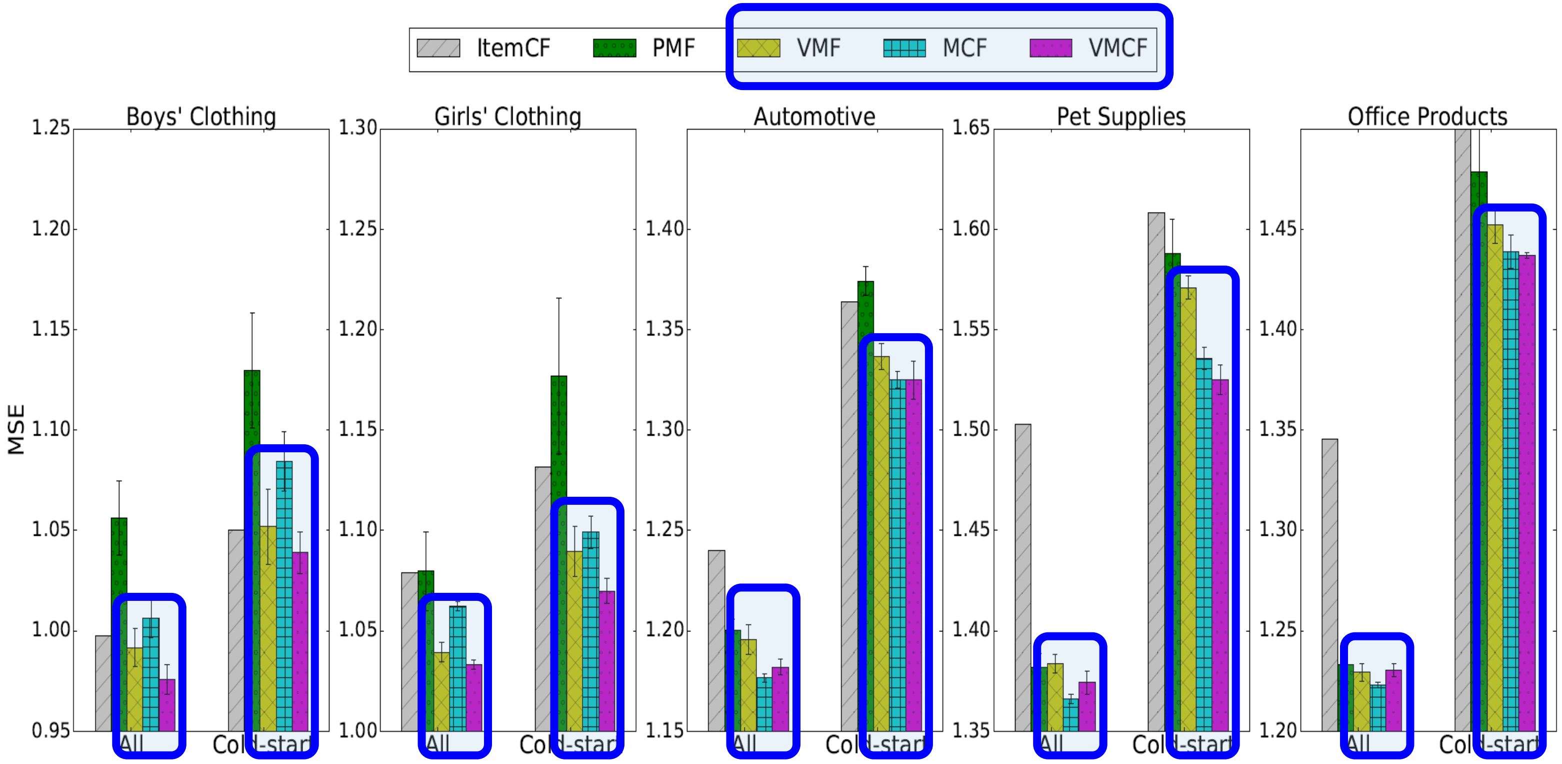
2. VMF vs. MCF



- Observation
  - **Visually-aware product domain**
    - ✓ **VMF > MCF**
  - **Visually non-aware product domain**
    - ✓ **MCF > VMF**
- Analysis
  - **Visually-aware product domain**
    - Appearance > Others
  - **Visually non-aware product domain**
    - Others > Appearance

Performance Analysis

3. VMF, MCF vs. VMCF



- Observation
  - **Visually-aware product domain**
    - ✓ All: **VMCF > VMF, MCF**
    - ✓ Cold-start: **VMCF >> VMF, MCF**
  - **Visually non-aware product domain**
    - ✓ All: **MCF > VMCF**
    - ✓ Cold-start: **VMCF > MCF**
- Analysis
  - **Visually-aware product domain**
    - Other aspects beyond appearance are taken into account by modeling the *"also-viewed"* relationship
  - **Visually non-aware product domain**
    - Given sufficient data, visual features are noise
    - If not, visual features are still useful

Quality of Product Embedding

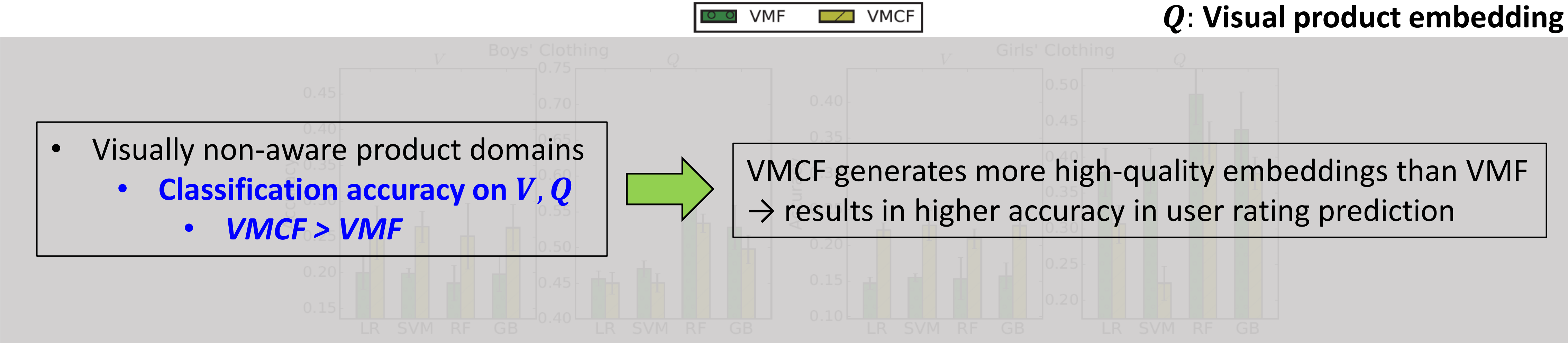
<u>Q. Do we indeed obtain high-quality product embedding by our method?</u>

Proof) Perform classification with product embedding vectors to see if vectors generated by our model represents the product well

- Competitor: (Visual) Product embedding (V, Q) of *VMF*
- Classification algorithms
  - Logistic regression , Support vector machine, Random forest, Gradient boosting
- Input: Products in top-10 most frequently appeared categories
- Label: Corresponding category
- Goal: Classify which category a product belongs given product embeddings

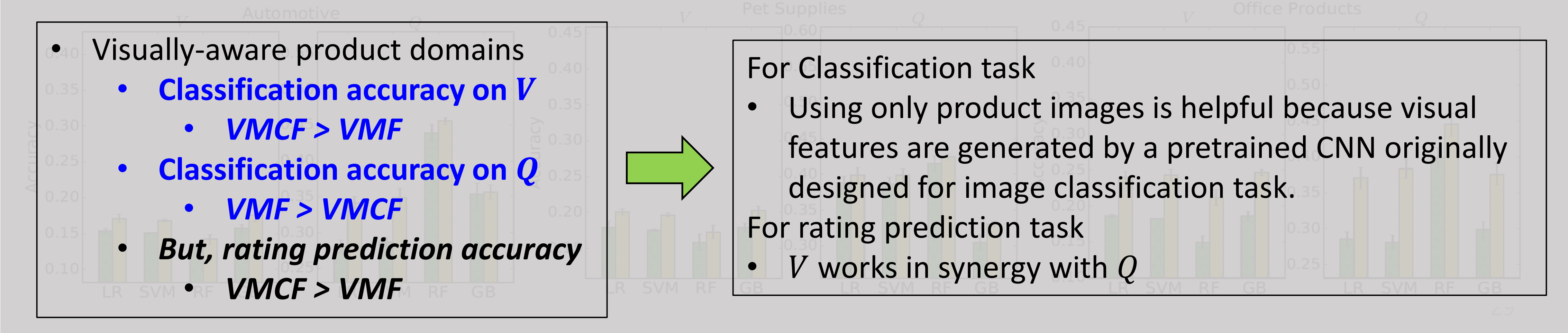
Quality of Product Embedding

$\boldsymbol{V}$: **Product embedding**
$\boldsymbol{Q}$: **Visual product embedding**



- Visually non-aware product domains
  - **Classification accuracy on V, Q**
    - ***VMCF > VMF***

VMCF generates more high-quality embeddings than VMF
→ results in higher accuracy in user rating prediction



- Visually-aware product domains
  - **Classification accuracy on V**
    - ***VMCF > VMF***
  - **Classification accuracy on Q**
    - ***VMF > VMCF***
  - ***But, rating prediction accuracy***
    - ***VMCF > VMF***

For Classification task
- Using only product images is helpful because visual features are generated by a pretrained CNN originally designed for image classification task.

For rating prediction task
- V works in synergy with Q

Conclusion

- Every product domain has dominant aspects that are more influential to user ratings than others.
  - Clothing domain → Appearance
  - Other domains → Functionality, Specification
- Proposed a matrix co-factorization framework that jointly factorizes user ratings data and “also-viewed” product information
  - Reflects various product aspects that are varyingly influential to user ratings in different product domains

Thank you!

References

- [AAAI 16] He, Ruining, and Julian McAuley. "VBPR: visual bayesian personalized ranking from implicit feedback." *arXiv preprint arXiv:1510.01784* (2015).
- [IJCAI 16] He, Ruining, et al. "Sherlock: sparse hierarchical embeddings for visually-aware one-class collaborative filtering." *arXiv preprint arXiv:1604.05813* (2016).
- [RecSys 16] He, Ruining, et al. "Vista: A Visually, Socially, and Temporally-aware Model for Artistic Recommendation." *arXiv preprint arXiv:1607.04373* (2016).
- [NIPS 08] Salakhutdinov, Ruslan, and Andriy Mnih. "Probabilistic Matrix Factorization." *Nips*. Vol. 1. No. 1. 2007.