Weakly Supervised Learning of Visual Size and Fit in Fashion Images

Nour Karessli
Romain Guigourès
Reza Shirvany
Finding clothes that fit is the biggest problem for customers shopping online and offline.
Supporting customers on their size and fit purchase decision is a challenging problem:

- Thousands of new articles everyday with short lifetime
- Return process takes a few days to few weeks
- Zero or few sales and returns for new articles
We present a novel teacher-student approach:

- Demonstrate the rich value of fashion images in inferring size characteristics of fashion apparel
- Effectively tackle the challenging cold start problem of providing size advice for new articles using images
- Generate large scale confidence-weighted weak annotations from crowd's subjective feedback → control weak annotations influence on the final model
Related Work

Teacher-Student Transfer Learning

Transferring knowledge from privileged information space to decision space [Vapnik et al. JMLR15]:

- Teacher leverages privileged historical weakly annotated data of sales and returns
- Student uses this knowledge to learn from images in decision space
Approach

Teacher Statistical Model

\[ \mathcal{L} = \binom{n}{k} p^k (1 - p)^{n-k} \]

\[ s = -\ln(\mathcal{L}) \]

Confidence score

Predicted Label

Backbone Feature Extractor

MLP

Fashion Images

Weakly labeled data

Sales

Returns

zalando
Approach - Teacher

Data
Weakly annotated data from customers subjective feedback provided in the return process

Statistical Modeling
Binomial classifier considering two factors:

● Article category
  Different categories show different size return rate (high heels vs. sneakers)

● Article lifetime
  Article sales period influence return rate (seasonality, sales, etc.)
Binomial Likelihood

\[ \mathcal{L} = \binom{n}{k} p^k (1 - p)^{n-k} \]

- \( p \): expected size return rate of article category over the sales period
- \( k \): size returns of the item
- \( n \): sales of the item

Estimator Score

Based on negative logarithm of likelihood

\[ s = - \ln(\mathcal{L}) \]
Student - SizeNet

CNN Backbone Feature Extractor

Transfer knowledge using bottleneck features of pre-trained network

- Resnet [He et al. CVPR16] pre-trained on ImageNet dataset [Deng et al. CVPR09]
- FashionDNA [Bracher et al. KDD16] pre-trained on in-house rich fashion dataset of 1.3 million articles
CNN Backbone Feature Extractor

Transfer knowledge using bottleneck features of pre-trained network

- Resnet [He et al. CVPR16] pre-trained on ImageNet dataset [Deng et al. CVPR09]
- FashionDNA [Bracher et al. KDD16] pre-trained on in-house rich fashion dataset of 1.3 million articles

Multi-Layer Perceptron

4 fully connected layers with nonlinear activations

Use binary cross entropy loss weighted based on estimator confidence score

\[ w = \ln(1 + s) \]

logarithmic transformation of score allows us to reduce the skewness
Dataset

- 127K articles of women textile including 12 categories such as: dresses, blouses, jeans, skirts, jackets, etc.
- Config SKU level
  - Manufacturers use different fabrics depending on the dying technique
  - Customers don’t perceive size and fit the same way depending on the color of clothes

<table>
<thead>
<tr>
<th>Class</th>
<th>#Articles</th>
<th>#Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>size issue</td>
<td>68,892</td>
<td>69,064</td>
</tr>
<tr>
<td>no size issue</td>
<td>58,152</td>
<td>58,321</td>
</tr>
<tr>
<td>total</td>
<td>127,044</td>
<td>127,385</td>
</tr>
</tbody>
</table>
Evaluation - Baseline

Attributes
Replace article images with sparse k-hot encoding of *human annotated* binary fashion attributes

- Neckline
- Sleeve length
- Pattern
- Length
Evaluation - Baseline

![ROC Curve](image1)

- Attributes-AUC = 0.78
- SizeNet-AUC = 0.74
- ResNet-AUC = 0.70

![Precision-Recall Curve](image2)

- Attributes-AP = 0.81
- SizeNet-AP = 0.77
- ResNet-AP = 0.74
Evaluation - Weights Importance

\[ \tau \] threshold applied on weights
Evaluation - Size Issue Probability vs. Weights

- **Bottom right**
  almost no samples are misclassified by SizeNet when Teacher is certain of no size issue;

- **Top left**
  high density of correctly predicted samples by SizeNet where Teacher is unsure

- **Top right**
  samples show that SizeNet has learned accurately from Teacher

- **Bottom left**
  SizeNet misclassifies fewer samples where Teacher is unsure
Evaluation - Size Issue Explanations

Generate Explanations using RISE [Petsiuk et al. BMCV19]

- Randomly generate masked input
- Use the corresponding outputs to assess region saliency

Insights

- True positives show more localized heatmaps
- False positives show more expanded maps
- False positives are affected by article design
Conclusion

- Fashion images in fact contain information about article size and fit issues
- Fashion images are valuable assets in tackling the challenging cold start problem
Future Work

- Include expert-labeled data
- Explore generalization capacity to fashion images in the wild
- Evaluate SizeNet explanations to understand if they correspond to actual customer problems
Thanks!

SizeNet: Weakly Supervised Learning of Visual Size and Fit in Fashion Images

Nour Karessli
Romain Guigourès
Reza Shirvany

zalando