Attribute Adaptation for Personalized Image Search

Adriana Kovashka **Kristen Grauman**
The University of Texas at Austin

Problem

- Existing methods assume monolithic attributes are sufficient
  [Lampert et al. CVPR 2009, Farhadi et al. CVPR 2009, Branson et al. ECCV 2010, Kumar et al. PAMI 2011, Scheirer et al. CVPR 2012, Parikh & Grauman ICCV 2011, …]
- However, there are real perceptual differences between annotators

Formal?  User labels: 50% "yes" 50% "no"

More ornamented?  or User labels: 50% "first" 20% "second" 30% "equally"

- Further, attribute terms can be imprecise

Is  *formal*?

= *formal* wear for a conference? OR

= *formal* wear for a wedding?

Is *blue* or *green*?

English: "blue"

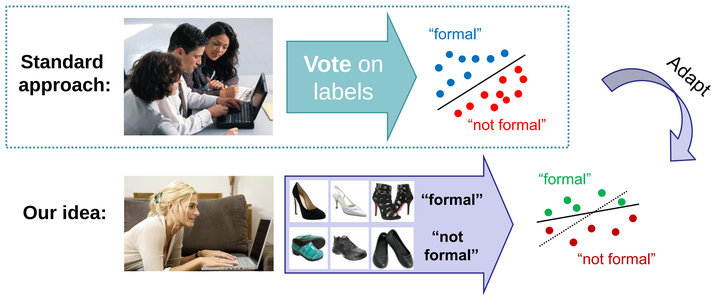
Russian: "neither" ("голубой" vs. "синий")

Japanese: "both" ("青" = blue *and* green)

Overweight? or just *Chubby?*

Our Idea

1) Treat learning of perceived attributes as an adaptation problem.



We adapt a generic attribute predictor trained with a large amount of majority-voted data with a small amount of user-labeled data.

2) Obtain labels implicitly from user's search history.

Impact: Capture user's perception with minimal annotation effort. Personalization makes attribute-based *image search* more accurate.

Learning Adapted Attributes

Training data $D_r = \{(\boldsymbol{x}_{i_1}, \boldsymbol{x}_{i_2})\}_{i=1}^{N}$ $\boldsymbol{x}_{i_1}$ is more [attribute] than $\boldsymbol{x}_{i_2}$

Learning
$$\min_{\boldsymbol{w}_r} \frac{1-\delta}{2}\|\boldsymbol{w}_r\|^2 + \frac{\delta}{2}\|\boldsymbol{w}_r - \boldsymbol{w}_r'\|^2 + C\sum_{i=1}^{N}\xi_i$$
$$\text{subject to } \boldsymbol{w}_r^T\boldsymbol{x}_{i_1} - \boldsymbol{w}_r^T\boldsymbol{x}_{i_2} \geq 1-\xi_i, \quad \xi_i \geq 0, \quad \forall i$$

Prediction
$$f_r(\boldsymbol{x}) = \delta f_r'(\boldsymbol{x}) + \sum_{i=1}^{N}\beta_i \boldsymbol{x}^T(\boldsymbol{x}_{i_1} - \boldsymbol{x}_{i_2})$$

B. Geng, L. Yang, C. Xu, and X.-S. Hua. "Ranking Model Adaptation for Domain-Specific Search." IEEE TKDE, March 2010.

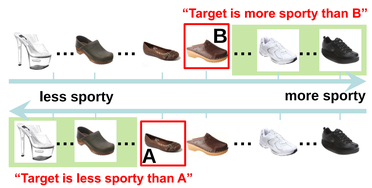
- Similar formulation for binary classifiers (Yang et al. 2007)

Inferring Implicit User-Specific Labels

- Transitivity

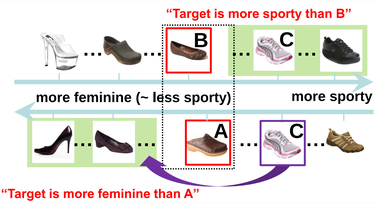
$$f_r(T) > f_r(A) \cap f_r(T) < f_r(B) \implies f_r(A) < f_r(B)$$

- Contradictions



Feedback implies no images satisfy all constraints.

Contradiction implies attribute models are inaccurate.



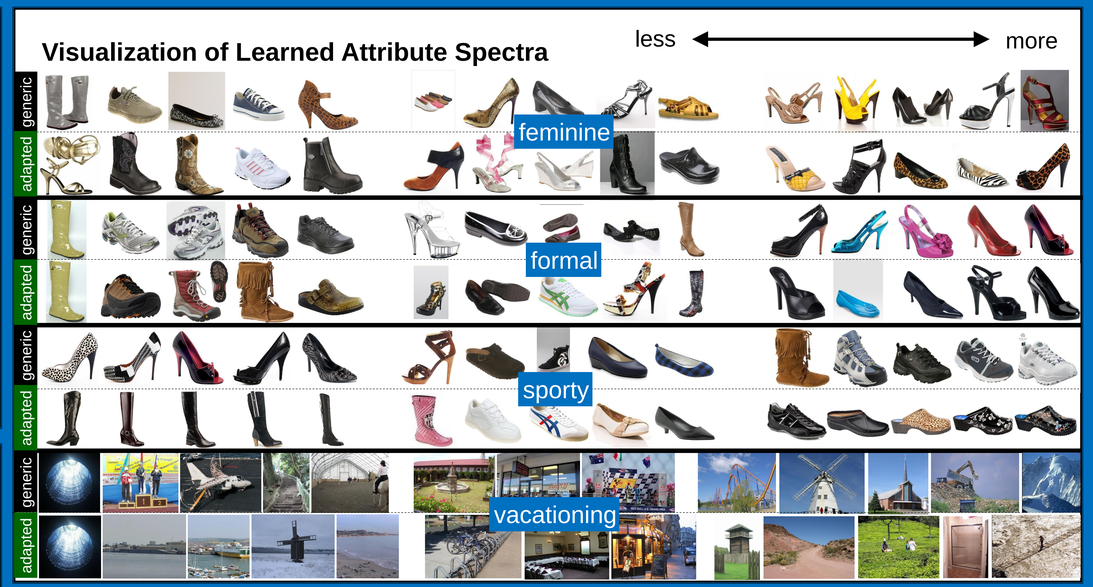
Relax conditions for contradiction.

Adjust models using new ordering on some image pairs.

Datasets

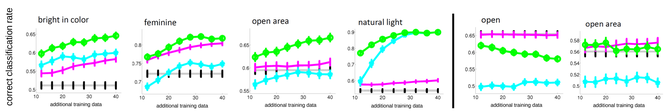
Shoes[Berg10, Kovashka12] attributes: pointy, open, bright, shiny, ornamented, high-heeled, long, formal, sporty, feminine
SUN[Patterson12] attributes: sailing, vacationing, hiking, camping, socializing, shopping, vegetation, clouds, natural light, cold, open area, far-away horizon
Size: 14k images each; **Features:** GIST, color, HOG, SSIM

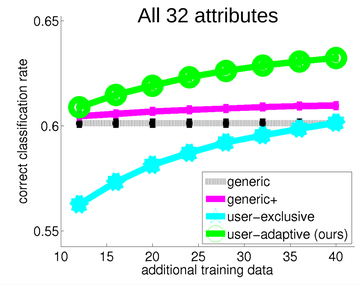
Visualization of Learned Attribute Spectra



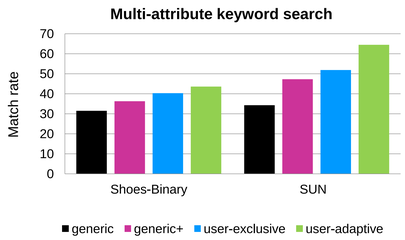
Adapted Attribute Accuracy

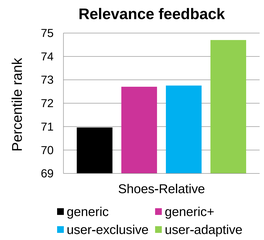
- **Generic:** status quo of learning from majority-voted data
- **Generic+:** like above, but uses more generic data
- **User-exclusive:** learns a user-specific model from scratch

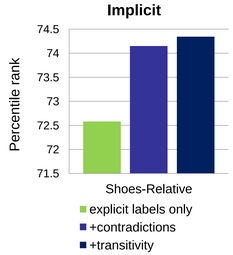




Impact of Adapted Attributes for Personalized Search







The personalized attribute models allow the user to more quickly find his/her search target. Implicitly gathering labels for personalization saves the user time, while producing similar results.