Beyond Comparing Image Pairs: Setwise Active Learning for Relative Attributes

Lucy Liang and Kristen Grauman
University of Texas at Austin

Abstract

It is useful to automatically compare images based on their visual properties—to predict which image is brighter, more feminine, more blurry, etc. However, comparative models are inherently more costly to train than their classification counterparts. Manually labeling all pairwise comparisons is intractable, so which pairs should a human supervisor compare? We explore active learning strategies for training relative attribute ranking functions, with the goal of requesting human comparisons only where they are most informative. We introduce a novel criterion that requests a partial ordering for a set of examples that minimizes the total rank margin in attribute space, subject to a visual diversity constraint. On three challenging datasets and experiments with "live" online annotators, the proposed method outperforms both traditional passive learning as well as existing active rank learning methods.[1]

1. Motivation and Challenges

While vision research has long focused on *categorizing* visual entities (e.g., recognizing objects or activities), there is increasing interest in *comparing* them. For example, whereas the presence or absence of an attribute in an image may not be clear-cut, whether one image exhibits the attribute more or less than another may be more informative [6]. Similarly, while a user doing image search may have difficulty declaring certain images as entirely irrelevant, he may more easily decide whether one image is more or less relevant than another [7].

In such settings, methods to learn ranking functions are a natural fit. Rather than labels, training a ranking function requires ground truth *comparisons* that relate one instance to another (e.g., person A is *smiling more* than person B; image X is *more relevant* than image Y). At a glance, this means that the labeling load could grow quadratically with the number of images, raising important scaling concerns. Yet, exhaustive pairwise comparisons should not be necessary to learn the concept, as some will be redundant.

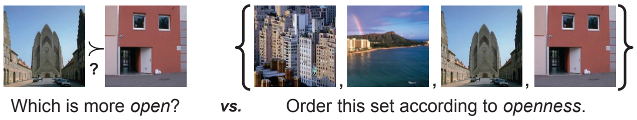

Figure 1: To learn relative attribute ranking functions, we propose an efficient active selection criterion that asks annotators to partially order a set of diverse yet informative images. Whereas a pairwise approach (left) gets just one bit of information, the setwise approach (right) amortizes annotator effort by getting (implicitly) all mutual comparisons.

Our goal is to leverage human supervision only where it is needed most when training relative attributes, such as *more/less bright*, *more/less feminine*, etc. To this end, we explore active learning for ranking functions. Active learning empowers the system to select those examples a human should label in order to most expedite learning. While its use for classification is fairly mature in both the learning and vision communities, it is much less studied for ranking. In vision, prior work for active learning with attributes focuses solely on classification problems [2, 4, 1].

Active rank learning presents three distinct technical challenges. First, hard comparisons for the system can also be hard for a human labeler due to their visual similarity. Second, restricting labeling tasks to solely paired comparisons can be wasteful; the human labeler spends time interpreting the attributes in two images, yet the system gets only one bit of information in return (that is, which image has the property more than the other). Third, the quadratic number of possible comparisons poses a scalability challenge for any but the most simplistic criteria, since active selection typically entails scanning through all yet-unlabeled data to select the optimal request.

2. Overview of Proposed Approach

In light of these challenges, we explore a series of increasingly complex active selection criteria for learning to rank. We start with a pairwise margin-based criterion that selects pairs with high uncertainty. Then, we consider a setwise extension [8] that requests a partial order on multiple examples at once. Finally, we introduce a novel setwise criterion that both amortizes human perceptual effort and promotes diversity among selected images, thereby avoiding uninformative comparisons that may be too close for even humans to reliably distinguish. See Fig. 1.

[1] Per the call for papers, we are submitting single-blind because this work appears in the main conference at CVPR 2014.

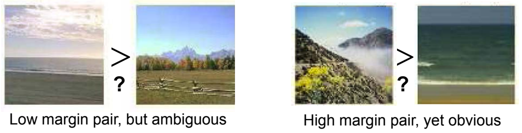

Figure 2: Which image in each pair exhibits more *diagonal plane*? Focusing on either extreme—low or high rank margins—can thwart active learning, requesting comparisons that are too hard or easy for both the human and learning algorithm.

To model a relative attribute [6], e.g., *fuzziness*, we train a large-margin ranking function r that predicts the relative strength of that attribute in an image: $r(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x}$, where $\boldsymbol{x}$ is an image descriptor. The linear ranking function parameters $\boldsymbol{w}$ are optimized to both satisfy the ground truth ordering among any supplied training images as well as maximize the *rank margin* $|\boldsymbol{w}^T\boldsymbol{x}_i - \boldsymbol{w}^T\boldsymbol{x}_j|$ between nearest-ranked pairs. Whereas past work assumes the ground truth orders are supplied as pairs, we explore the use of *partial orders* among larger tuples of images.

We follow a pool-based active learning strategy to train relative attribute rankers. At each iteration, the system must examine a pool $\mathcal{P}$ of unlabeled images and predict what comparisons will most benefit its current ranking functions. After it makes a selection, the comparisons are posed to annotators, and their (aggregated) responses are used to augment the ordered training set. Then, the learned attribute rankers are retrained, and the process repeats.

When applied to ranking, we find that traditional margin-based active selection methods tend to make uninformative requests of human annotators. Intuitively, the mutually close set of examples may be hard for a human annotator to compare relatively. See Fig. 2.

To combat this problem, we introduce a new approach called the *diverse setwise low margin* (DSLM) criterion. Our goal is to select the set of image examples that minimize the mutual rank margin in attribute space, subject to a visual diversity constraint in the original image feature space. To capture visual diversity, we first cluster all the image descriptors $\boldsymbol{x}_i$ (e.g., GIST, color) in $\mathcal{P}$. This establishes the primary modes among the unlabeled examples. Let c_i denote the cluster to which image i belongs. Our selection objective is:

$$\mathcal{S}^* = \operatorname*{argmin}_{\mathcal{S}\subseteq\mathcal{P}} \sum_{(i,j)\in\mathcal{S}} |\boldsymbol{w}^T\boldsymbol{x}_i - \boldsymbol{w}^T\boldsymbol{x}_j|, \qquad (1)$$
$$s.t. \quad c_i \neq c_j, \forall i \neq j,$$

where $|\mathcal{S}|$ is the given target set size. In other words, the most useful set is the one that has examples difficult enough to be informative to the ranker, yet not "too" difficult for the human to make unambivalent decisions (since each is from a different cluster). This balances *exploiting* the margin uncertainty with *exploring* the feature space.

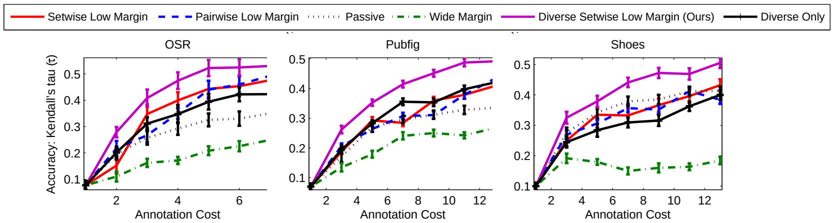

Figure 3: Learning curves summarizing the "live" active learning results on 3 datasets and 27 total attributes. Best viewed in color.

To optimize Eqn. 1, we develop an efficient search strategy. The idea is as follows. Only a strictly rank-contiguous set will minimize the total margin; yet, there may not be a rank-contiguous set for which diversity holds. Thus, we scan contiguous sets in sequence, always maintaining the current best margin score. If the current best is not diverse, we perturb it using the next nearest sample until it is. The key to efficiency is to exploit the 1D ordering inherent in attribute ranks, even though the clusters are in the high-dimensional descriptor space. We refer to our CVPR 2014 paper for details.

3. Example Results

We use 3 challenging public datasets: Outdoor Scenes, PubFig Faces, and Shoes. We show that with an active approach, a system can learn accurate relative attributes with less human supervision. This in itself is a contribution, as no prior work examines active training of comparative visual models. Furthermore, we show that the proposed setwise strategy consistently outperforms existing methods.

Figure 3 shows the "live" results, where we push active requests to MTurk workers and iteratively update the model. We compare DSLM to five baselines: 1) passive [6]; 2) diverse only, which selects samples based on the same diversity constraint as DSLM, but ignores margins; 3) wide margin, which chooses pairs with the widest margins; 4) pairwise low margin, which chooses the pairs with the lowest margins; and 5) setwise low margin [8]. We see our method outperforms all the alternatives. The practical impact is significant: across all attributes, our method requires 39% fewer annotations to attain the same accuracy reached by the passive learner in the last iteration.

References

[1] A. Biswas and D. Parikh. Simultaneous active learning of classifiers & attributes via relative feedback. In *CVPR*, 2013.

[2] S. Branson, C. Wah, F. Schroff, B. Babenko, P. Welinder, P. Perona, and S. Belongie. Visual recognition with humans in the loop. In *ECCV*, 2010.

[3] P. Donmez and J. Carbonell. Optimizing estimated loss reduction for active sampling in rank learning. In *ICML*, 2008.

[4] A. Kovashka, S. Vijayanarasimhan, and K. Grauman. Actively selecting annotations among objects and attributes. In *ICCV*, 2011.

[5] B. Long, O. Chapelle, Y. Zhang, Y. Chang, Z. Zheng, and B. Tseng. Active learning for ranking through expected loss optimization. In *SIGIR*, 2010.

[6] D. Parikh and K. Grauman. Relative Attributes. In *ICCV*, 2011.

[7] B. Siddiquie, R. S. Feris, and L. S. Davis. Image ranking and retrieval based on multi-attribute queries. In *CVPR*, 2011.

[8] H. Yu. SVM selective sampling for ranking with application to data retrieval. In *KDD*, 2005.