

Visual Comparison of Images Using Multiple Kernel Learning for Ranking

Amr Sharaf¹

amr.sharaf@alexu.edu.eg

Mohamed E. Hussein^{2,1}

mohamed.e.hussein@ejust.edu.eg

Mohamed A. Ismail¹

maismail@alexu.edu.eg

¹ Alexandria University

El-Shatby, Alexandria 21526, Egypt

² Egypt-Japan University of Science and Technology (E-JUST)

New Borg Al-Arab, Alexandria 21934, Egypt

Abstract

Ranking is the central problem for many applications such as web search, recommendation systems, and visual comparison of images. In this paper, the multiple kernel learning framework is generalized for the learning to rank problem. This approach extends the existing learning to rank algorithms by considering multiple kernel learning and consequently improves their effectiveness. The proposed approach provides the convenience of fusing different features for describing the underlying data. As an application to our approach, the problem of visual image comparison is studied. Several visual features are used for describing the images and multiple kernel learning is adopted to find an optimal feature fusion. Experimental results on three challenging datasets show that our approach outperforms the state-of-the-art and is significantly more efficient in run-time.

Proposed Approach (RankMKL)

Given two images, it is required to learn which image exhibits a particular visual attribute more than the other. Our approach works on a per attribute basis, thus a separate model is learned for each visual attribute. Figure 1 demonstrates the outline of our approach. The first step is to extract a set of features from each image. Several feature sets are selected to capture different visual cues in the image. To capture the image texture, we extract Local Binary Patterns (LBP) [3] and compute the response from a set of Gabor filters. For capturing the shape and appearance of the images, GIST [4] and HoG [1] descriptors are used. Finally, a color histogram is computed in the LAB color space to capture the color information.

The second step is to fuse the different feature sets and learn the ranking model. For this task, a separate kernel function is computed for each set of features (i.e. we compute five different kernels). The computed kernels are considered as base kernels for our multiple kernel learning module. Using the multiple kernel learning algorithm, we learn the optimal weights for creating a linear combination from the base kernels together with the optimal parameters for the ranking model.

Instead of using a single kernel matrix (K) for learning the ranking model, an optimal combination from several base kernels is learned, and the combination of the base kernels matrix (K_d) is used for training the ranking model, where $k_d(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T_d \phi(\mathbf{x}_j)_d$ represents the dot product in feature space ϕ and is parametrized by \mathbf{d} such that:

$$k_d(\mathbf{x}_i, \mathbf{x}_j) = f_d(\{k_i(\mathbf{x}_i, \mathbf{x}_j)\}_{i=1}^t), \quad (1)$$

where t is the number of base kernels, $\mathbf{d} \in \mathbb{R}^t$ is the optimal kernel weights to be learned, and the combination function f_d can be a linear or a non-linear function for combining the base kernels. Our goal is to learn the optimal values for (\mathbf{d}) together with the optimal values for the Lagrange multipliers (α) representing the learned ranking model. Accordingly, the standard rankSVM [2] objective function is updated as follows:

$$\begin{aligned} & \underset{\alpha}{\text{maximize}} \quad \{\mathbf{1}^T \alpha - \frac{1}{2} \alpha^T Q_d \alpha + r(\mathbf{d})\} \\ & \text{subject to} \quad 0 \leq \alpha_{i,j} \leq C, \forall (i, j) \in P, \\ & \quad \mathbf{d} \geq 0, \end{aligned} \quad (2)$$

$$Q_{d,(i,j),(u,v)} = k_d(\mathbf{x}_i, \mathbf{x}_u) + k_d(\mathbf{x}_j, \mathbf{x}_v) - k_d(\mathbf{x}_i, \mathbf{x}_v) - k_d(\mathbf{x}_j, \mathbf{x}_u), \quad (3)$$

where both the regularizer r and the kernel k_d can be any general differentiable functions of \mathbf{d} with continuous derivatives and P represents the set of preference pairs such that: $P = \{(i, j) | \mathbf{x}_i \succ \mathbf{x}_j\}$. In our approach,

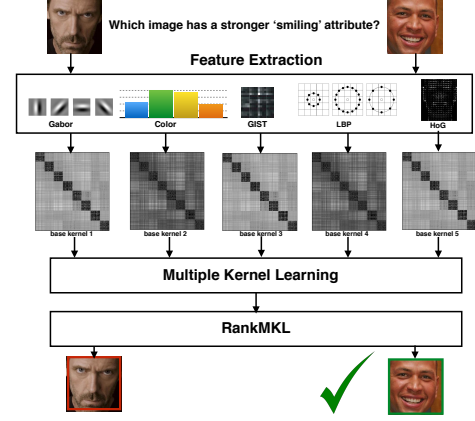


Figure 1: Illustration of our proposed approach. Given two images, it is required to detect which image has a stronger visual attribute than the other. Different features are extracted and Multiple Kernel Learning (MKL) is used for fusing the kernels from each feature set. RankMKL is used for ranking the images.

five base kernels are used, one for each of the five feature sets (LBP, HoG, Gabor, GIST, and Color). The kernel function k_d is selected as a linear combination from the five base kernels: $k_d(\mathbf{x}_u, \mathbf{x}_v) = \sum_{i=1}^5 d_i k_i(\mathbf{x}_u, \mathbf{x}_v)$ and L2 regularization function is used for $r(\mathbf{d})$. Gradient descent is used for solving Eq. (2) using the same algorithm in [5].

Conclusion

In this paper, the standard multiple kernel learning formulation is extended to the learning to rank problem. Effectiveness of the proposed approach is demonstrated on the visual image comparison task. Although MKL has been extensively used for object recognition and image categorization, this is the first time it has been used for image comparison. Through extensive experiments, the advantage of our approach is clearly demonstrated both in terms of accuracy and runtime efficiency. Future work includes exploring more applications of multiple kernel learning for ranking, such as web search and recommendation systems.

- [1] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.
- [2] Thorsten Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 133–142. ACM, 2002.
- [3] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, 2002.
- [4] Aude Oliva and Antonio Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, 42(3):145–175, 2001.
- [5] Manik Varma and Bodla Rakesh Babu. More generality in efficient multiple kernel learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 1065–1072. ACM, 2009.