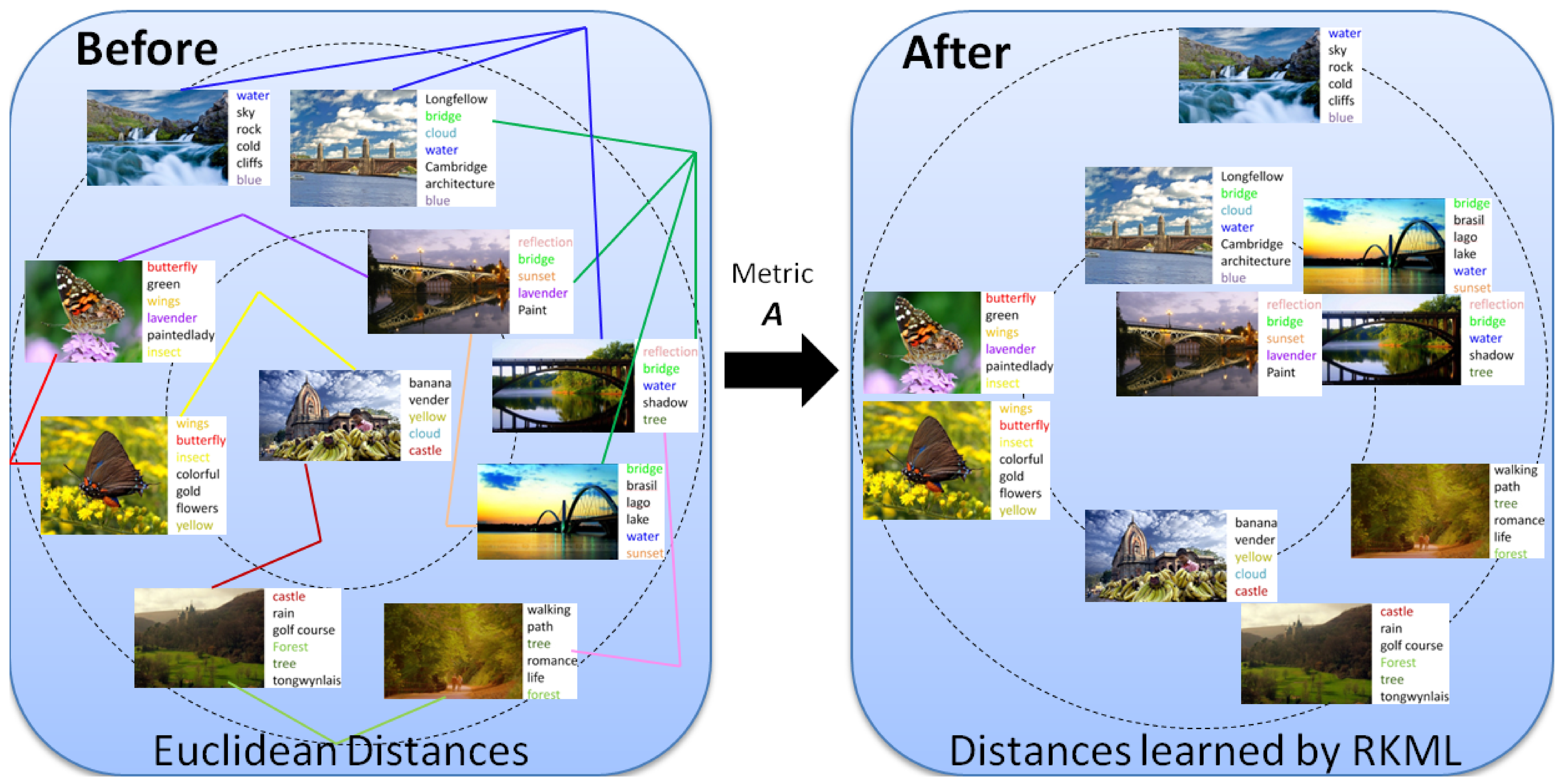
Large-scale Image Annotation by Efficient and Robust Kernel Metric Learning

Zheyun Feng, Rong Jin, and Anil Jain

Department of Computer Science and Engineering, Michigan State University, MI, United States

Motivation

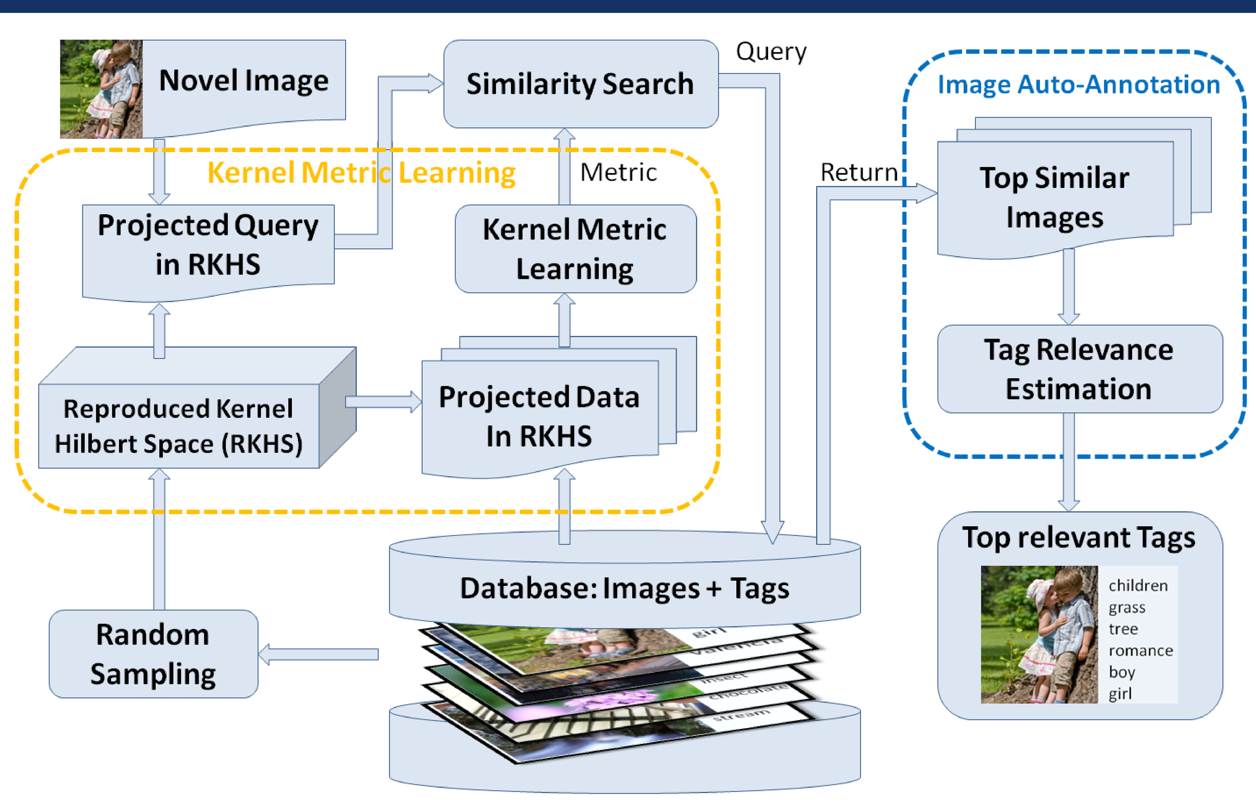
Learn a kernel (nonlinear) distance metric (**KML**) that reflects the semantic similarity between images:



Limitations of existing KML algorithms:

✧ Limited to binary constraints, and unable to exploit soft constraints;
✧ High computational cost due to the PSD constraint;
✧ Data overfitting due to the flexibility of nonlinear kernel function.

Proposed RKML Algorithm



☞ Effectively incorporate soft constraints in KML;
☞ Exploit the regression technique to avoid projection to PSD cone;
☞ Reduce data overfitting by regularizing the rank of learned metric;
☞ Apply Nyström method to improve computational efficiency;
☞ Provide the theoretical guarantee for KML for the first time.

Notation

- $\{(\mathbf{x}_1,\mathbf{y}_1),\cdots,(\mathbf{x}_n,\mathbf{y}_n)\}$: training set, where $\mathbf{x}_i \in \mathbb{R}^d$ and $\mathbf{y}_i \in \{0,1\}^m$;
- $T:\mathcal{H}_\kappa \mapsto \mathcal{H}_\kappa$: a linear operator in kernel space;
- $A \in \mathbb{R}^{n\times n}$: PSD kernel metric learned from training data;
- $K = [\kappa(\mathbf{x}_i,\mathbf{x}_j)]_{n\times n}$: the kernel matrix;
- $\mathcal{S} = [s_{i,j}]_{n\times n}$: Pairwise semantic similarities;
- $K^b \in \mathbb{R}^{n\times n_s}$, $K^s = [\kappa(\hat{\mathbf{x}}_i,\hat{\mathbf{x}}_j)]_{n_s\times n_s}$, with n_s random samples.

Regression based Kernel Metric Learning

Learn a linear operator T in functional space from the training examples

$$\widehat{T} = \mathop{\arg\min}_{T\succeq 0} \sum_{i,j=1}^n \frac{1}{2}\left(s_{i,j} - \langle \kappa(\mathbf{x}_i,\cdot), T[\kappa(\mathbf{x}_j,\cdot)]\rangle_{\mathcal{H}_\kappa}\right)^2.$$

Simplify using $\widehat{T}[f](\cdot) = \sum_{i,j=1}^n \kappa(\mathbf{x}_i,\cdot)A_{i,j}f(\mathbf{x}_j)$ due to the representer theorem of kernel learning [1],

$$\boxed{\min_{A\succeq 0}\ \mathcal{L}(A) = \frac{1}{2}|\mathcal{S} - KAK^\top|_F^2}$$

☞ Rank regularization: $A = K_r^{-1}\mathcal{S}K_r^{-1}$
☞ PSD guarantee: When $\mathcal{S}$ is PSD, so is A.
☞ Nyström approximation [2]: $\tilde{K}_r = K^b[K_r^s]^{-1}[K^b]^\top$;
☞ $Sim(\mathbf{x}_a,\mathbf{x}_b) = \langle \kappa(\mathbf{x}_a,\cdot), \widehat{T}[\kappa(\mathbf{x}_b,\cdot)]\rangle = \sum_{i,j=1}^n \kappa(\mathbf{x}_a,\mathbf{x}_i)\kappa(\mathbf{x}_b,\mathbf{x}_j)A_{i,j}$.

Theoretical Guarantee of RKML

Define

$$T_* = \mathop{\arg\min}_{T'} \mathrm{E}_{(\mathbf{x}_a,\mathbf{x}_b,\mathbf{y}_a,\mathbf{y}_b)}\left[(\mathbf{y}_a^\top\mathbf{y}_b - \langle\kappa(\mathbf{x}_a,\cdot), T'[\kappa(\mathbf{x}_b,\cdot)]\rangle_{\mathcal{H}_\kappa})^2\right]$$

Let $T_*(r)$ be the best r-rank approximation of T_*. Then, with overwhelming probability

$$\|\widehat{T} - T_*(r)\|_2 \le O\left(\frac{1}{\sqrt{n}(\lambda_r - \lambda_{r+1})}\right)$$

When the eigengap $\lambda_r - \lambda_{r+1}$ is sufficiently large, the approximation error is reduced at rate of $1/\sqrt{n}$.

Reference

[1] B. Schölkopf, A. J. Smola, Learning with kernels: support vector machines, regularization, optimization and beyond, MIT Press, 2002.

[2] P. Drineas, M. W. Mahoney, On the nyström method for approximating a gram matrix for improved kernel-based learning, JMLR 6 (2005) 2153–2175.

Datasets

	No. of Images	No. of Tags	Tags per image (Mean/maximum)	Images per tag (Mean/maximum)
ESP Game	20,768	268	4.69 / 15	363 / 5,059
IAPR TC12	19,627	291	5.72 / 23	386 / 5,534
Flickr 1M	999,764	1,000	5.98 / 202	5,976 / 76,531

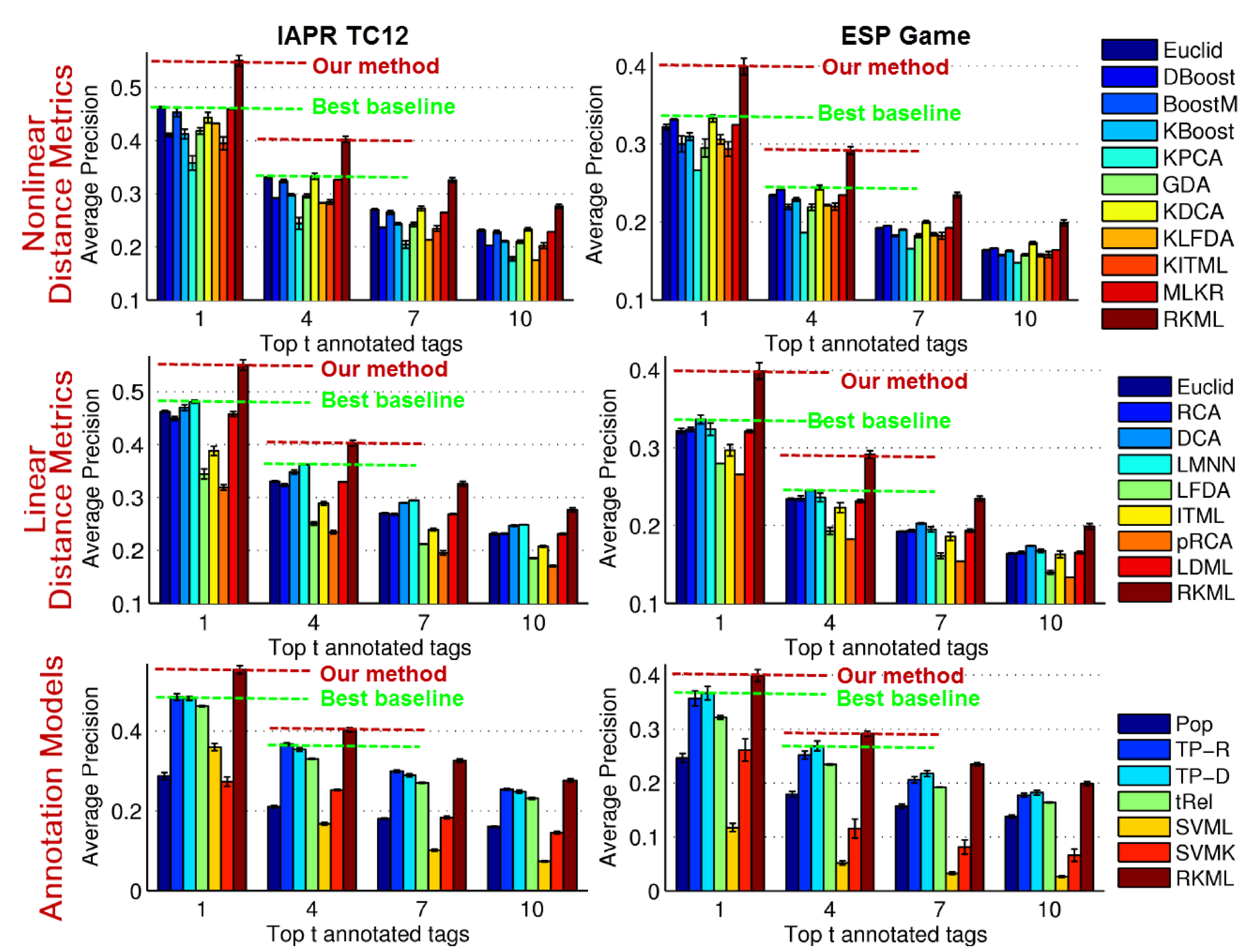
Effects of Nonlinear ML and Soft Constraints

	ESP Game				Flickr 1M			
AP@t(%)	t=1	t=3	t=5	t=8	t=1	t=3	t=5	t=8
RKML	40±1.1	32±0.4	27±0.4	22±0.4	24±0.1	18±0.1	15±0.2	13±0.2
RKMLH	34±1.0	28±0.5	24±0.4	20±0.3	20±0.2	16±0.2	14±0.2	11±0.1
RLML	36±0.8	28±0.7	24±0.5	20±0.4	13±0.3	11±0.2	10±0.1	9±0.1

RKMLH runs RKML using binary constraints generated by topic model;
RLML is the linear version of RKML.

Comparison with State-of-the-art Baselines

☞ Effectiveness



☞ Efficiency

	Kernel Metric Learning						Boosting DML		Our
Time(s)	KPCA	GDA	KDCA	KLFDA	KITML	MLKR	DBoost	BoostM	RKML
IAPR TC12	2.8e4	4.8e4	2.2e4	8.8e4	5.3e4	2.2e3	1.7e4	1.1e6	4.6e2
ESP Game	3.3e4	5.4e4	3.7e4	3.2e5	6.8e4	3.5e4	4.3e4	1.2e6	1.3e3
Flickr 1M	7.343	3.3e4	1.3e5	1.0e5	3.7e6	7.9e3	1.2e4	3.2e5	3.4e3

	Linear Metric Learning				TagProp		Voting	SVM	
Time(s)	DCA	LMNN	ITML	LDML	TP-R	TP-D	tRel	SVML	SVMK
IAPR TC12	1.5e4	1.4e4	4.2e5	4.2e5	9.1e2	4.6e2	1.0e1	2.5e3	4.0e5
ESP Game	2.3e4	1.7e4	5.8e4	5.5e5	2.7e2	1.5e2	1.5e1	1.6e2	8.9e4
Flickr 1M	8.1e4	6.0e4	3.0e4	5.2e5	1.6e5	9.9e4	5.7e3	-	-

Annotation Examples Generated by RKML and Baselines