Smooth Neighborhood Structure Mining on Multiple Affinity Graphs with Applications to Context-Sensitive Similarity

Song Bai[1], Shaoyan Sun[2], Xiang Bai[1], Zhaoxiang Zhang[3], Qi Tian[4]

[1]*Huazhong University of Science and Technology* [2]*University of Science and Technology of China*

[3]*CAS Center for Excellence in Brain Science and Intelligence Technology* [4]*University of Texas at San Antonio*

Goal

Input:

M affinity graphs $\mathcal{G}^{(v)} = (X, W^{(v)})_{v=1}^{M}$, where vertices of the graph $X = \{x_1, x_2, \ldots, x_N\}$ represent images, and W(v) is the edge weights of the v-st graph.

Output:

A more faithful similarity measure D

Sparse Contextual Activation (SCA)

A context-sensitive similarity can be defined

$$\hat{d}_J(x_q, x_p) = 1 - \frac{\sum_{i=1}^{N} \min\left(F_{q,i}, F_{p,i}\right)}{\sum_{i=1}^{N} \max\left(F_{q,i}, F_{p,i}\right)}$$

where $F \in \mathbb{R}^{N\times 1}$ is the sparse contextual activation defined as

$$F_{q,p} = \begin{cases} \exp\left(-d(x_q, x_p)\right) & \text{if } x_p \in \mathcal{N}_k(x_q) \\ 0 & \text{otherwise.} \end{cases}$$

The key to make it works lies that similar images select same neighbors with same responses.

Smooth Neighborhood

Select neighbors smoothly along the underlying manifold

$$\min_{Y} \sum_{i<j}^{N} w_{ij} \|Y_i - Y_j\|^2 + \mu \sum_{i=1}^{N} \|Y_i - I_i\|^2,$$

where $Y_i = [y_{i1}, y_{i2}, \ldots, y_{iN}] \in \mathbb{R}^{1\times N}$ is the indicator functions of x_i that describes the probability distribution of its neighbors.

When multiple affinity graphs are available, we want to learn a share indicator

$$\min_{\alpha, Y} \sum_{v=1}^{M} \alpha^{(v)^\gamma} \sum_{i<j}^{N} w_{ij}^{(v)} \|Y_i - Y_j\|^2 + \mu \sum_{i=1}^{N} \|Y_i - I_i\|^2,$$

$$s.t. \sum_{v=1}^{M} \alpha^{(v)} = 1, 0 \le \alpha^{(v)} \le 1,$$

where $\alpha = \{\alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(M)}\}$ is the weight of affinity graphs, and $\gamma > 1$ is the weight controller

Optimization: alternatively

- Update Y as $Y = \mu \left(\sum_{v=1}^{M} \alpha^{(v)^\gamma} L^{(v)} + \mu I \right)^{-1}$
- Update $\alpha^{(v)}$ as

$$\alpha^{(v)} = \frac{\left(Tr(Y^{\mathrm{T}} L^{(v)} Y)\right)^{\frac{1}{1-\gamma}}}{\sum_{v'=1}^{M} \left(Tr(Y^{\mathrm{T}} L^{(v')} Y)\right)^{\frac{1}{1-\gamma}}}.$$

Until convergence.

Experiments

MPEG-7 shape dataset

- Qualitative Evaluation: using Jaccard similarity between two neighborhood sets $S(x_q, x_p) = \frac{|\mathcal{N}(x_q) \cap \mathcal{N}(x_p)|}{|\mathcal{N}(x_q) \cup \mathcal{N}(x_p)|}$

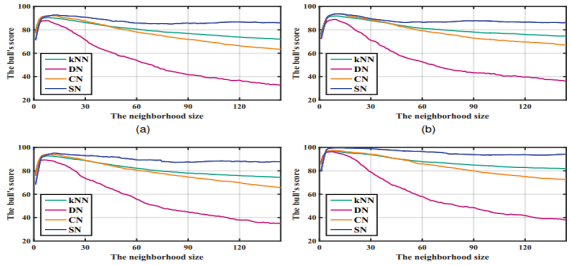

Fig. 1. The comparison using Eq. (10) on MPEG-7 dataset. The baseline similarities used are IDSC (a), SC (b), ASC (c) and IDSC+SC+ASC (d) respectively

- Improving Context-Sensitive Similarity: combining neighborhood selection techniques with SCA to provide retrieval performances

Table 2. The bull's eye scores of different methods on MPEG-7 dataset

Descriptors	Methods	Bull's eye score
IDSC	Contextual Dissimilarity Measure (CDM) [1]	88.30%
IDSC	Generic Diffusion Process (GDP)* [13]	90.96%
IDSC	Index-Based Re-Ranking [8]	91.56%
IDSC	Graph Transduction (GT) [2]	91.61%
IDSC	Locally Constrained Diffusion Process [19]	92.36%
IDSC	RL-Sim Re-Ranking [7]	92.62%
IDSC	Shortest Path Propagation (SSP) [3]	93.35%
IDSC	Mutual kNN Graph (mkNN) [15]	93.40%
IDSC	Sparse Contextual Activation (SCA) [9]	93.44%
IDSC	**Smooth Neighborhood (Ours)**	**93.52%**
SC	Generic Diffusion Process (GDP)* [13]	92.81%
SC	Graph Transduction (GT) [2]	92.91%
SC	Sparse Contextual Activation (SCA) [9]	95.21%
SC	**Smooth Neighborhood (Ours)**	**95.25%**
ASC	Generic Diffusion Process (GDP)* [13]	93.95%
ASC	Index-Based Re-Ranking [8]	94.09%
ASC	RL-Sim Re-Ranking [7]	95.75%
ASC	Locally Constrained DP (LCDP) [19]	95.96%
ASC	Tensor Product Graph (TPG) [5]	**96.47%**
ASC	**Smooth Neighborhood (Ours)**	95.98%
IDSC+SC	Co-Transduction [4]	97.72%
IDSC+SC	Locally Constrained Mixed Diffusion (LCMD) [20]	98.84%
IDSC+SC	Sparse Contextual Activation (SCA) [9]	99.01%
IDSC+SC	**Smooth Neighborhood (Ours)**	**99.25%**
AIR	Tensor Product Graph (TPG) [5]	99.99%
AIR	Generic Diffusion Process (GDP) [13]	100.00%
AIR	**Smooth Neighborhood (Ours)**	**100.00%**

Ukbench Image dataset

Table 3. The N-S scores of different methods on Ukbench dataset. Note that Query Adaptive Fusion uses 5 input similarities, and the last result of our method is produced by using all the 4 similarities implemented in this paper

Descriptors	Methods	N-S score
BoW (3.52)	kNN Re-ranking [35]	3.56
BoW (3.22)	Tensor Product Graph [5]	3.61
BoW (3.26)	Co-transduction [4]	3.66
BoW (3.50)	RNN Re-ranking [16]	3.67
BoW (3.54)	Graph Fusion [11]	3.67
BoW (3.33)	Contextual Dissimilarity Measure [1]	3.68
BoW (3.56)	Sparse Contextual Activation [9]	3.69
BoW (3.57)	**Smooth Neighborhood (Ours)**	**3.75**
CNN (3.44)	**Smooth Neighborhood (Ours)**	**3.66**
CNN (3.65)	**Smooth Neighborhood (Ours)**	**3.81**
HSV (3.17)	Graph Fusion [11]	3.28
HSV (3.40)	Sparse Contextual Activation [9]	3.56
HSV (3.40)	**Smooth Neighborhood (Ours)**	**3.56**
BoW (3.20,3.17,2.81)	Locally Constrained Mixed Diffusion [20]	3.70
BoW (3.54), HSV (3.17)	Graph Fusion [11]	3.77
BoW (3.54), HSV (3.17)	Graph Fusion [12]	3.83
BoW (3.58), CNN (3.40), *etc.*	Query Adaptive Fusion [36]	3.84
BoW (3.56), HSV (3.40)	Sparse Contextual Activation [9]	3.86
BoW (3.13), CNN (3.87)	ONE [37]	3.89
BoW (3.57), CNN (3.44), *etc.*	**Smooth Neighborhood (Ours)**	**3.98**

References

[1] M. Donoser and H. Bischof. "Diffusion processes for retrieval revisited ", In *CVPR*, 2013.

[2] S. Zhang, M. Yang, T. Cour, K. Yu, and D. N. Metaxas, "Query specific fusion for image retrieval", in ECCV, 2012

Acknowledgement

This work was supported in part by NSFC 61573160, NSFC 61429201 and China Scholarship Council. This work was supported in part to Dr. Qi Tian by ARO grants W911NF-15-1-0290 and Faculty Research Gift Awards by NEC Laboratories of America and Blippar.