Online Asymmetric Similarity Learning for Cross-Modal Retrieval
Yilong Wu, Shuhui Wang, Qingming Huang
1Key Lab of Intell. Info. Process., Inst. of Comput. Tech, Chinese Academy of Sciences, China
2University of Chinese Academy of Sciences, China

Motivations:
- The critical problem in cross-modal retrieval task is how to measure the similarity between data from different modalities.
- The relations between images and texts are highly asymmetric.
- There are two kinds of relative similarities that can be used.
- CNN features are state-of-the-art features, but there are many CNN layers. Choosing which layer to use is a difficult problem.

Contributions:
- We propose an online learning method to learn the similarity function between heterogeneous modalities by preserving the bi-directional relative similarity in the training data.
- We extend it to an online multiple kernel learning method to address the problem of combining different layers of CNN features for cross-retrieval.

Learning Bi-direction Relative Similarity:
- Consider learning a bilinear similarity function $s(v_i, t_j) = v_i^T W t_j$.
- Bi-directional relative similarity constraints are indispensable for modeling the cross-modal relation.

We expect the similarity function to satisfy the following two conditions simultaneously:

$\theta(W, v_i, t_j, c_i) = \max \{0, s(v_i, t_j) - s(v_i, t_j^c) + 1\}$

$\theta(W, t_j, v_i, c_j) = \max \{0, s(v_i, t_j) - s(v_i^c, t_j) + 1\}$

$L(W; D_{train}) = \sum_{i,j} \left( \theta(W; v_i, t_j, c_i) + \theta(W; t_j, v_i, c_j) \right)$

The model is trained by the Passive-aggressive (PA) algorithm. We call this method Cross-Modal Online Similarity function learning (CMOS).

Conclusions:
- We have proposed CMOS and its multiple kernel extension CMOMKS to learn a similarity function between heterogeneous data modalities by preserving relative similarity constraints from two directions.
- The CMOS online model is learned by the Passive-Aggressive algorithm. Multiple kernelized similarity function is further combined in CMOMKS by the Hedging algorithm.
- Experimental results on three public cross-modal datasets have demonstrated that the proposed methods outperform state-of-the-art approaches.