Supplementary Material: Discovering Video Clusters from Visual Features and Noisy Tags

Arash Vahdat, Guang-Tong Zhou, and Greg Mori

School of Computing Science, Simon Fraser University, Canada
{avahdat,gza11,mori}@cs.sfu.ca

1 Optimization

The Flip MMC framework proposed in this paper jointly optimizes the model parameters that describe each cluster, finds the best assignment of videos to clusters, and refines the tag labeling to reduce the noise in tag annotation. Similar to MMC, the Flip MMC optimization is a challenging non-convex optimization problem due to the discrete optimization that assigns videos to clusters and refines tag labels.

Here this non-convex optimization problem is rewritten in unconstrained format as:

$$\min_{w} \frac{\lambda}{2}||w||_2^2 + R_w \tag{1}$$

where R_w is the the risk function defined in the form of an assignment problem as:

$$R_w = \min_{y_n} \sum_{n=1}^{N} R'_w(y_n) \tag{2}$$

$$\text{s.t.} \quad L \leq \sum_{n=1}^{N} \mathbb{1}_{(y_n=k)} \leq U$$

where $R'_w(y_n)$ computes the "mis-clustering" cost of assigning the n-th video to the cluster y_n using:

$$R'_w(y_n) = \min_{\mathbf{t}'_n} \max_{y,\mathbf{t}} \left(w^\top \phi(x_n, \mathbf{t}, y) + \Delta^{y,y_n}_{\mathbf{t},\mathbf{t}'_n} - w^\top \phi(x_n, \mathbf{t}'_n, y_n) + \gamma \Delta'_{\mathbf{t}'_n,\mathbf{t}_n} \right). \tag{3}$$

In Eq. 3 annotated tags change to $\mathbf{t}'_n$ such that the error of assigning the video x_n to y_n is minimal while number of changes are being penalized by $\Delta'_{\mathbf{t}'_n,\mathbf{t}_n}$.

In order to address the unconstrained optimization problem in Eq. 1, we develop a coordinate descent-style approach shown in Algorithm 1. This algorithm alternates between finding the parameters of each cluster (w) and finding an assignment of videos to clusters. The algorithm mainly consists of three steps performed iteratively. First, "mis-clustering" cost is computed in Eq. 3, and then it is used for computing risk function by solving the assignment problem in Eq. 2. Finally, the model parameters are updated given the risk values. The following explains these steps in detail.

Algorithm 1 Flip MMC Optimization

Input : $\{x_n, \mathbf{t}_n\}_{n=1}^{n=N}$, K, ϵ
Output : parameters w
Initialize w_1
for $\tau \leftarrow 1$ **to** τ_{max} **do**
  for $n \leftarrow 1$ **to** N **do**
    for $y_n \leftarrow 1$ **to** K **do**
      Compute $R'_w(y_n)$ using Eq. 3
    end for
  end for
  Solve the assignment problem in Eq. 2
  compute $\frac{\partial R_w}{\partial w}\big|_{w_\tau}$, from Eq. 4
  Compute $[w_{\tau+1}, w^*_\tau, gap]$, from [1], Alg. 1
  if $gap \leq \epsilon$ or $\tau == \tau_{max}$ **then**
    return w^*_τ
  end if
end for

Computing mis-clustering cost: In this step, the mis-clustering costs for each video and cluster, $R'_w(y_n)$ is computed by solving the integer min-max problem in Eq. 3. Here, the heuristic proposed in Vahdat and Mori [2] is used to find an approximate solution to this problem. In a nutshell, the heuristic solves an approximation of the min-max problem in two steps. First, an approximate solution to the inner maximization is computed and the refined tag labels are found by solving the outer minimization given the approximate solution of inner maximization. Second, the exact inner maximization is computed for the fixed refined tag labels from the previous step. Due to the simple structure of the loss function and clustering model, the min-max optimization can be solved efficiently by performing the so-called loss augmented inference [3] twice. In this work, given decomposable hamming loss function and our simple model, loss augmented inference becomes inferring the unary potential functions on tags that can be done efficiently for each tag independently.

Note that the refined tag label, $\mathbf{t}'_n$ is found for each cluster separately. This enables us to flexibly find cluster-specific tags while making sure that refined tag labels and the annotated tags are not very different.

Risk function computation: Given the mis-clustering costs for all clusters and video, the assignment problem in Eq. 2 becomes a linear integer programming problem. This problem in general is an NP-hard problem, and here an approximated solution is found using linear programming (LP) relaxation, using GNU Linear Programming Kit (GLPK).

Updating w: The optimization problem in Eq. 1 is a non-convex optimization problem. For solving this problem, we use the NRBM approach of Do and Artières [1], which is a non-convex extension of the cutting plane algorithm. This approach starts from an initial w, and at each iteration it creates a piecewise linear approximation of the objective function by adding a cutting plane

at the optimum discovered from the previous iteration. The approach requires the computation of the risk function and its gradient at the current optimum, which can be efficiently computed for our linear clustering model given the optimal solutions for the risk function. Assuming that from Eq. 2 the n-th video is assigned to y_n^*, its tags are refined to $\mathbf{t}_n^*$, and the most-violated labels in the inner maximization in Eq. 3 are y_n^{**} and $\mathbf{t}_n^{**}$, the gradient is computed simply using:

$$\frac{\partial R_w}{\partial w} = \sum_{n=1}^{N} \phi(x_n, \mathbf{t}_n^{**}, y_n^{**}) - \phi(x_n, \mathbf{t}_n^*, y_n^*) \tag{4}$$

The final assignment of videos to clusters is produced by solving the assignment problem in Eq. 2 using the optimal parameters, w_τ^* obtained in Algorithm 1.

References

1. Do, T.M.T., Artières, T.: Large margin training for hidden markov models with partially observed states. In: ICML. (2009)
2. Vahdat, A., Mori, G.: Handling uncertain tags in visual recognition. In: ICCV. (2013)
3. Taskar, B., Chatalbashev, V., Koller, D., Guestrin, C.: Learning structured prediction models: A large margin approach. In: ICML. (2005)