Algorithm to initialize action labels for region-based actions using frame-level action annotations

November 1, 2015

1 Problem

In M videos, we have Q_m temporal annotations of actions (action stamps) in each video, in the form of a time interval where the action is performed. The actions can overlap in time. We wish to split each frame into R regions, and select a single action label for each region in each frame. As we do not know in advance which regions perform the actions, we plan to build a model to associate each action interval with one or more regions.

The structure of the problem allow us to infer the following relations:

- Each action interval q must appear at least one time in the video. The action label of the action interval q is defined as a_q.
- Two overlapping action intervals can't appear in the same region.

We define $v^m_{r,q} = 1$ when the action interval q appears in region r in the video m, and $v^m_{r,q} = 0$ otherwise. We assume we have pose labels for each frame, independent for each region. For an action interval q, we use the histogram of pose labels for each region in the action interval, defined for the video m as $h^m_{r,q}$. We can solve the problem of finding the correspondence between action intervals and regions in a formulation similar to k-means, using the structure of the problem as constraints in the labels.

$$
\begin{aligned}
P1) \quad \min J = & \sum_{m=1}^{M} \sum_{r=1}^{R} \sum_{q=1}^{Q_m} v^m_{r,q} ||h^m_{r,q} - \mu^r_{a_q}||_2^2 - \frac{1}{\lambda} v^m_{r,q} \\
\text{s. to} \quad & \sum_{r=1}^{R} v^m_{r,q} \geq 1 \\
& v^m_{r,q_1} + v^m_{r,q_2} \leq 1 \text{ if } q_1 \cap q_2 \neq \emptyset \\
& v^m_{r,q} \in \{0, 1\}, \ \forall m
\end{aligned}
\tag{1}
$$

$\mu^r_{a_q}$ are computed as the mean of the descriptors with the same action label within the same region, and μ^r_{idle} are computed using all the non-as. The general idea is to solve $P1$ iteratively as k-means, finding the cluster centers for each region r, μ^r_a using the labels $v^m_{r,q}$, and then finding the best labeling given the cluster centers.

2 Solving $P1$

2.1 Finding the optimal labels given μ_a^r

The most difficult step is to find the labels $v_{r,q}^m$ that minimizes $P1$ satisfying the constraints.

Exhaustive search One alternative is to make an exhaustive search over the set of valid labeling. For instance, if there is Q non-overlapping actions (the worst case), then the number of different labeling matrices is 16^Q, which can be handled if Q is small. When action intervals overlap, the possibilities decrease and possibly the problem can still be handled using exhaustive search. If there is small overlapping of action intervals, and Q is moderate, then alternative methods should be used.

Relaxing the problem Other alternative is relaxing the labels to lie in the $[0, 1]$ interval, transforming the integer program into a linear program, that can be solved in polynomial time. Also, a sparsity regularization term added to the relaxed variables could be useful to search for better solutions, keeping the problem as a linear program formulation. There is some alternatives to refine the solutions if needed:

- Branch and bound techniques for fractional values
- Cutting plane method for integer programming

2.2 Finding the cluster centers

Given the labels, finding the cluster centers seems to be straightforward. Nevertheless, as some frames "turned off" if their label become 0, there is still some doubts about the convergence rate of this algorithm, that should be explored with caution.

3 Using classifiers scores instead of distortion metric

We could change the objective function to learn classifiers instead of cluster centers, maximizing the score of the classifiers instead of minimizing the distortion. One disadvantage is that J is unbounded, and we are maximizing, so we would need tor regularize W. For instance, a related formulation could be given by

$$P2) \quad \max J = \sum_{m=1}^{M} \sum_{r=1}^{R} \sum_{q=1}^{Q_m} v_{r,q}^m {W_{a_q}^r}^\top h_{r,q}^m - \lambda\Omega(W)$$

$$\begin{aligned} \text{s. to} \quad & 1 \leq \sum_{r=1}^{R} v_{r,q}^m \leq o_q^m \\ & \sum_{q=1}^{Q_m} v_{r,q}^m \leq t_m \\ & v_{r,q_1}^m + v_{r,q_2}^m \leq 1 \text{ if } q_1 \cap q_2 \neq \emptyset \\ & v_{r,q}^m \in \{0,1\},\ \forall m \end{aligned} \tag{2}$$