Multi-Task Transfer Methods to Improve One-Shot Learning for Multimedia Event Detection

Wang Yan
wyang@sfu.ca
Jordan Yap
jjyap@sfu.ca
Greg Mori
mori@cs.sfu.ca

School of Computing Science
Simon Fraser University
Burnaby, BC, CANADA

This paper proposes a new multi-task learning method with implicit inter-task relevance estimation, and applies it to complex Internet video event detection, which is a challenging and important problem in practice, yet seldom has been addressed. In this paper, “detection” means to detect videos corresponding to the event of interest from a (large) video dataset, not to localize the event spatially or temporally in a video. In the problem definition, one positive and plenty of negative samples of one event are given as training data, and the goal is to return the videos of the same event from a large video dataset. In addition, we assume samples of other events are available.

Fig. 1 shows an overview of the proposed methods. The widths of the lines between the one-exemplar event and others represent the inter-event relevance, which is unknown a priori in our problem settings. However, the proposed method can implicitly infer the relevance and utilize the most relevant event(s) more in multi-task learning, where the shared information from the relevant events helps to build a better model from the one exemplar. The proposed method does not assume the relevance between other events, as indicated by the red line. Although the learning algorithm outputs models of all input events, only that of the one-exemplar event is expected from the relevant events helps to build a better model from the one exemplar. The proposed method can implicitly infer the relevance and utilize the most relevant tasks, but softmax is used in the representation, i.e.

\[ \Omega(W) = \sum_{t=1}^{T} \|w_t - w_{r_t}\|_2^2 \]  

This penalty focuses more on the smallest inter-model distance, which is slightly different from the former one with equally weighted \( K \) smallest distances. The experiments show that we can get good results with the smooth penalty. In addition to squared \( l_2 \) distance, one can also use the penalty term defined by correlations. The term still focuses more on most relevant tasks, but softmax is used in the representation, i.e.

\[ \Omega(W) = -\log \sum_{t=1}^{T} \exp(-\|w_t - w_{r_t}\|_2^2). \]  

We use the first option in the following experiments. All penalties in this subsection make the objective non-convex, but one can still get good results empirically. The objective is optimized by Quasi-Newton Soft-Threshold (QNST) method [2].
