Action Prediction from Videos via Memorizing Hard-to-Predict Samples

Yu Kong, Shangqian Gao, Bin Sun, and Yun Fu
Northeastern University, Boston, MA, USA
(yukong, yunfu)@ece.neu.edu

Objective: inferring the action label before the entire action execution has been observed.

Classifying actions from temporally incomplete data. Actions may appear similarly in the beginning

Basic idea: Memorize hard-to-predict samples for prediction.

Challenges
- What “experiences” shall we memorize?
- How shall we use them for prediction?
- How to maintain a compact experience memory?

Definitions
Uniformly segment a video into $K$ segments

Segment

A video $x$

Progress level $g = k = 3$

Observation ratio $r = k/K = 0.3$

Partial video $x^{(k)}$

Observation ratio $r$

Progress level $k$

Visual features: Residual networks for frames

Partially Observed Video

RGB frames

Optical Flow frames

Memory Key-Value Pair

Memory

A key (progress level)

A value (partial video)

Memory definition

$M = (K, V, A)$

$K$: key (data matrix)

$V$: (label, progress level) pair

$A$: age

Loss function

$max\{q - K[n_a] + q - K[n_a] = q - K[n_a] + \xi, 0\}$

Correct $a$: correct action label, correct progress level

Error $b$: incorrect action label

Error $c$: incorrect progress level

Memory update

If error occurs: $n' = arg\max A[z] + r$

$K[n_a] = K[n_a] + q$

Average item $K[n]$ and $q$

No error occurs

<table>
<thead>
<tr>
<th>Methods</th>
<th>UCF-101</th>
<th>Sports-1M</th>
</tr>
</thead>
<tbody>
<tr>
<td>MESC</td>
<td>34.3%</td>
<td>46.7%</td>
</tr>
<tr>
<td>Dynamic BoW</td>
<td>36.2%</td>
<td>43.4%</td>
</tr>
<tr>
<td>Integral BoW</td>
<td>36.2%</td>
<td>43.4%</td>
</tr>
<tr>
<td>MSDA</td>
<td>39.5%</td>
<td>49.2%</td>
</tr>
<tr>
<td>CDR + SVM</td>
<td>40.0%</td>
<td>50.8%</td>
</tr>
<tr>
<td>MTSSVM</td>
<td>40.0%</td>
<td>49.9%</td>
</tr>
<tr>
<td>DeepCN</td>
<td>45.6%</td>
<td>55.0%</td>
</tr>
<tr>
<td>RGB Bi-LSTM</td>
<td>50.0%</td>
<td>48.7%</td>
</tr>
<tr>
<td>Our method</td>
<td>51.6%</td>
<td>57.4%</td>
</tr>
</tbody>
</table>

Reference