Supplementary Materials for SODA: Story Oriented Dense Video Captioning Evaluation Framework

Soichiro Fujita[1], Tsutomu Hirao[2], Hidetaka Kamigaito[1], Manabu Okumura[1], and Masaaki Nagata[2]

[1] Institute of Innovative Research, Tokyo Institute of Technology
{fujiso, kamigaito, oku}@lr.pi.titech.ac.jp
[2] NTT Communication Science Laboratories, NTT Corporation
{tsutomu.hirao.kp, masaaki.nagata.et}@hco.ntt.co.jp

A The Pseudo Code of Dynamic Programming Algorithm

Algorithm 1 shows the pseudo code that fills out the $|\mathcal{G}| \times |\mathcal{P}|$ tables S and B, where B stores information required for backtracking. $B[i][j]$ is set to LEFT, UPPER, and DIAGONAL when $S[i][j]$ is computed by using $S[i-1][j]$, $S[i][j-1]$, and $S[i-1][j-1]$, respectively. Since the maximum score of the sum of IoU by considering temporal ordering is stored in the $|\mathcal{G}|$-th line and $|\mathcal{P}|$-th column of S, we can derive the optimal matching by backtracking the path starting from that cell in B by following DIAGONAL. Note that when the k-th generated proposal is matched with the ℓ-th reference proposal, the next matching is with the $u(> k)$-th generated proposal and the $v(> \ell)$-th reference proposal.

Algorithm 1 Dynamic programming algorithm for matching captions

procedure DynamicProgramming
  for $i = 0$ to $|\mathcal{P}|$ **do**
    $S[i][0] \leftarrow 0, B[i][0] \leftarrow 0$
  for $j = 0$ to $|\mathcal{G}|$ **do**
    $S[0][j] \leftarrow 0, B[0][j] \leftarrow 0,$
  for $i = 1$ to $|\mathcal{G}|$ **do**
    for $j = 1$ to $|\mathcal{P}|$ **do**
      if $\max(S[i-1][j], S[i-1][j-1] + C_{i,j}, S[i][j-1]) = S[i-1][j]$ **then**
        $S[i][j] = S[i-1][j]$
        $B[i][j] =$ Left
      if $\max(S[i-1][j], S[i-1][j-1] + C_{i,j}, S[i][j-1]) = S[i-1][j-1] + C_{i,j}$ **then**
        $S[i][j] = S[i-1][j-1] + C_{i,j}$
        $B[i][j] =$ Diagonal
      if $\max(S[i-1][j], S[i-1][j-1] + C_{i,j}, S[i][j-1]) = S[i][j-1]$ **then**
        $S[i][j] = S[i][j-1]$
        $B[i][j] =$ Upper
procedure BackTrack
  $k \leftarrow |\mathcal{G}|, \ell \leftarrow |\mathcal{P}|$
  while $k \geq 0$ and $\ell \geq 0$ **do**
    if $B[k][\ell] =$ Left **then**
      $\ell \leftarrow \ell - 1$
    if $B[k][\ell] =$ Upper **then**
      $k \leftarrow k - 1$
    if $B[k][\ell] =$ Diagonal **then**
      $a(g_k) = \ell$
      $k \leftarrow k - 1, \ell \leftarrow \ell - 1$