Visual Task Selection Algorithm

August 11, 2021

1 Algorithm

Algorithm 1: Visual Task Selection

Input: $\{w_i\}(i \in ||Q||)$ word embedding of natural language $Q_n l$.
I an input image.
Require: f: recurrent embedding module.
Require: *Agent*: visual selection system with policy π.
Require: $\mathcal{D} = \{Q_{nl}, I\}$: train set.

Build task pool accordingly;
while *epoch reaches maximum* **do**
  `// Perform Agent for all` $\mathcal{D}$
  `// Encoder`
  $h_i = f_{enc}(w_i, h_{i-1})$; ▷ $f_{enc}(\cdot)$ is nonlinear function for encoder
  `// Decoder with Attention`
  $e_{ij} = a(s_{j-1}, h_i)$; ▷ e_{ij} is the associate energy of probability a_{ij}, s_j is an RNN hidden state for time j, $a(\cdot)$ is the alignment model which is parametriz as a feedforward neural network
  $a_{ij} = \frac{exp(e_{ij})}{\sum_{k=1}^{||Q||} exp(e_{ik})}$; ▷ a_{ij} is the attention weight
  $c_i = \sum_{j=1}^{||Q||} a_{ij} h_{ij}$; ▷ c_i is the weighted context vector
  $s_i = f_{dec}(s_{i-1}, \mathrm{a}_{i-1}, c_i)$; ▷ $f_{dec}(\cdot)$ is nonlinear function for decoder
  `// Output Action Sequence`
  $p(\mathrm{a}_i|\{\mathrm{a}_1, ..., \mathrm{a}_{i-1}, c_i, Q_{nl}\}) = g(\mathrm{a}_{i-1}, s_i, c_i, Q_{nl})$; ▷ $g(\cdot)$ is a nonlinear function
  $A = [\mathrm{a}_1, ..., \mathrm{a}_i, ..., \mathrm{a}_t]$;
  `// Update Parameters`
  Update parameters by Loss $\mathcal{L} = \mathcal{L}_{policy} + \mathcal{L}_{\tau\alpha}$; ▷ Update parameters in $f_{enc}(\cdot), a(\cdot), f_{dec}(\cdot), g(\cdot)$
end