Bandit Structured Prediction

Algorithm 1: Bandit Structured Prediction
Input: Sequence of learning rates \(\gamma_k\)
Output: Optimal parameters \(\hat{\theta}\)
Initialize parameters \(\theta_0\)
for \(k = 0, \ldots, K\) do
  Observe input structure \(x_k\)
  Sample output structure \(\bar{y}_k \sim p_{\theta}(y|x_k)\)
  Obtain feedback \(\Delta(\bar{y}_k)\)
  Compute stochastic gradient \(s_k\)
  Update parameters \(\theta_{k+1} = \theta_k - \gamma_k s_k\)
Choose a solution \(\hat{\theta}\) from the list \(\{\theta_0, \ldots, \theta_K\}\).

Objectives

- **Expected Loss (EL):**
  Expectation of a task loss \(\Delta(\bar{y})\) over all input and output structures:
  \[ L^{\text{EL}}(\theta) = \mathbb{E}_{p(x)p(y|x)}[\Delta(\bar{y})]. \]
  Stochastic gradient:
  \[ s^\text{EL}_{k} = \Delta(\bar{y}_k) \frac{\partial \log p_{\theta}(\bar{y}|x_k)}{\partial \theta}. \]
  Output structures \(\bar{y}\) are sampled word by word from the distribution resulting from the softmax transformation in the output layer of the network.

- **Pairwise Preference Ranking (PR):**
  Transfer EL to pairs of structures \((\bar{y}, \tilde{y})\):
  \[ L^{\text{PR}}(\theta) = \mathbb{E}_{p(x)p(y|x)}[\Delta(\bar{y}, \tilde{y})]. \]
  Stochastic gradient:
  \[ s^\text{PR}_k = \Delta(\bar{y}_k, \tilde{y}_k) \left( \frac{\partial \log p_{\theta}(\bar{y}|x_k)}{\partial \theta} + \frac{\partial \log p_{\theta}(\tilde{y}|x_k)}{\partial \theta} \right). \]
  To rank \(\bar{y}_k\) over \(\tilde{y}_k\) with pairwise feedback, either continuous (cont)
  \[ \Delta(\bar{y}_k, \tilde{y}_k) = \Delta(\bar{y}_k) - \Delta(\tilde{y}_k), \]
  or binary (bin)
  \[ \Delta(\bar{y}_k, \tilde{y}_k) = \begin{cases} 1 & \text{if } \Delta(\bar{y}_k) > \Delta(\tilde{y}_k), \\ 0 & \text{otherwise}. \end{cases} \]
  Draw negative sample \(\tilde{y}_k\) from distribution \(p_{\theta}\), one word per output structure (chosen randomly):
  \[ p_{\theta}(\tilde{y}_k = w|x) \propto \exp(-\omega_w) \sum_{-1}^{\infty} \exp(-\omega_w). \]

Bandit Seq2Seq

Bandit structured prediction [1] is a stochastic optimization framework where learning is performed from partial feedback. This feedback is received in the form of task loss evaluation of a predicted output structure, without having access to gold standard structures.

In this work, we advance the framework by lifting linear bandits to neural seq2seq learning using attention-based RNNs, and incorporating control variates for variance reduction and improved generalization.

Experiments

- Neural machine translation domain adaptation:
  - Adapt a pre-trained model (Europarl, fr-en) to new domains (News Commentary and TED).
  - Simulated feedback with GLEU on references
  - Encoder-decoder architecture with attention
  - Full-information baselines: maximum likelihood estimation on reference translations

Strategies for handling of unknown words:
- attention-based replacement of UNKs for word-based models [4]
- sub-word models with Byte-Pair-Encoding (BPE) [5]

Findings

- Successful training of NMT with weak feedback
- Large improvements for domain adaptation, outperforming linear models
- Control variates improve generalization, see [6]

Acknowledgements

This research was supported in part by the German research foundation (DFG), and in part by a research cooperation grant with the Amazon Development Center Germany.

References