
Learning Discrete Structured Representations by Adversarially Maximizing Mutual Information : Supplementary Material

Karl Stratos¹ Sam Wiseman²

A. Forward Algorithm

The forward algorithm is shown in Algorithm 1. We write $[m]$ to denote the set of m integers $\{1, \dots, m\}$ and $[[A]] = 1$ if A is true and 0 otherwise.

B. Dataset Construction for Predictive Document Hashing

We take pairs from the Who-Did-What dataset (Onishi et al., 2016). The pairs in this dataset were constructed by drawing articles from the LDC Gigaword newswire corpus. A first article is drawn at random and then a list of candidate second articles is drawn using the first sentence of the first article as an information retrieval query. A second article is selected from the candidates using criteria described in Onishi et al. (2016), the most significant of which is that the second article must have occurred within a two week time interval of the first. We filtered article pairs so that each article is distinct in all data. The resulting dataset has 104404, 8928, and 7326 article pairs for training, validation, and evaluation.

To follow the standard setting in unsupervised document hashing, we represent each article as a TFIDF vector using a vocabulary of size 20000. We use SentencePiece (Kudo & Richardson, 2018) to learn an optimal text tokenization for the target vocabulary size.

REFERENCES

Kudo, T. and Richardson, J. Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 66–71, 2018.

Onishi, T., Wang, H., Bansal, M., Gimpel, K., and McAllester, D. Who did what: A large-scale person-

¹Rutgers University ²Toyota Technological Institute at Chicago. Correspondence to: Karl Stratos <karlstratos@gmail.com>.

Algorithm 1 Forward

Input: $p(z_i|i, z_{i-o:i-1})$ for $z_{i-o:i} \in \{0, 1\}^{o+1}$ and $i \in [m]$

Output: For $z_{i-o:i-1} \in \{0, 1\}^o$ and $i \leq m$

$$\pi(z_{i-o:i-1}|i) = \sum_{\bar{z} \in \{0,1\}^{i-1}: \bar{z}_{-j}=z_{i-j} \forall j \in [o]} p(\bar{z})$$

Runtime: $O(m2^o)$

Base: $\pi(z_{-o:-1}|i) \leftarrow [[z_{-o:-1} = 0^o]]$ for $i \leq 1$

Main: For $i = 2 \dots m$, for $z_{i-o:i-1} \in \{0, 1\}^o$,

$$\begin{aligned} \pi(z_{i-o:i-1}|i) \leftarrow & p(z_{i-1}|i-1, z^{(0)}) \times \pi(z^{(0)}|i-1) \\ & + p(z_{i-1}|i-1, z^{(1)}) \times \pi(z^{(1)}|i-1) \end{aligned}$$

where $z^{(b)} = (b, z_{i-o}, \dots, z_{i-2}) \in \{0, 1\}^o$

centered cloze dataset. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 2230–2235, 2016.