Learning Discrete Representations via Information Maximizing Self-Augmented Training

Weihua Hu
UTokyo, RIKEN
hu@ms.k.u-tokyo.ac.jp

Takeru Miyato
Preferred Networks
miyato@preferred.jp

Seiya Tokui
Preferred Networks
tokui@preferred.jp

Eiichi Matsumoto
Preferred Networks
matsumoto@preferred.jp

Masashi Sugiyama
RIKEN, UTokyo
sugi@k.u-tokyo.ac.jp

The task of unsupervised discrete representation learning is to obtain a function that maps *similar* data into similar discrete representations, where the *similarity* of data is defined according to applications of interest. It is a central machine learning task because of the compactness of the representations and ease of interpretation. The task includes two important machine learning tasks as special cases: clustering and unsupervised hash learning.

Deep neural networks are promising to be used because they can model the non-linearity of data and scale to large datasets. However, their model complexity is huge, and therefore, we need to carefully regularize the networks in order to learn useful representations that exhibit intended invariance for applications of interest.

To this end, we propose a method called Information Maximizing Self-Augmented Training (IMSAT). In IMSAT, we use data augmentation to impose the invariance on discrete representations, which is illustrated as red arrows in Figure 1. As depicted, we encourage the predicted representations of augmented data points to be close to those of the original data points in an end-to-end fashion. We term such regularization *self-augmented training (SAT)*. SAT is flexible to impose various types of invariances on the representations predicted by neural networks. For example, it is generally preferred for data representations to be locally invariant, i.e., remain unchanged under local perturbations on data points. Using SAT, we can impose the local invariance on the representations by pushing the predictions of perturbed data points to be close to those of the original data points. For image data, it may also be preferred for data representations to be invariant under affine distortion, e.g., rotation, scaling and parallel movement. We can similarly impose the invariance via SAT by using the affine distortion for the data augmentation.

We then combine the SAT with the regularized information maximization (RIM) for clustering (Gomes, Krause, and Perona, 2010), and arrive at our Information Maximizing Self-Augmented Training (IMSAT), an information-theoretic method for learning discrete representations using deep neural networks. We illustrate the basic idea of IMSAT in Figure 1. Following the RIM, we maximize information theoretic dependency between inputs and their mapped outputs, while regularizing the mapping function. IMSAT, however, differs from the original RIM in two ways. First,

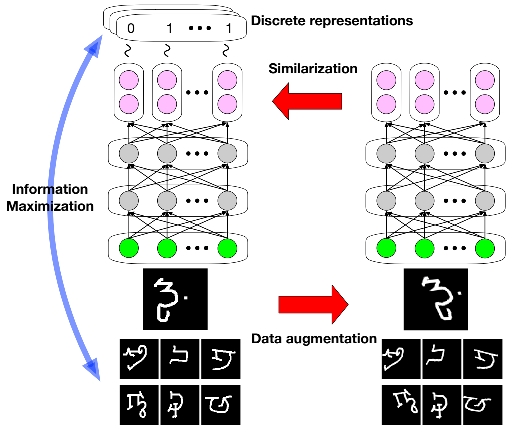

Figure 1: Basic idea of our proposed method for unsupervised discrete representation learning. We encourage the prediction of a neural network to remain unchanged under data augmentation (Red arrows), while maximizing the information-theoretic dependency between data and their representations (Blue arrow).

IMSAT deals with a more general setting of learning discrete representations; thus, is also applicable to hash learning. Second, it uses a deep neural network for the mapping function and regularizes it in an end-to-end fashion via SAT. Learning with our method can be performed by stochastic gradient descent (SGD); thus, scales well to large datasets.

Extensive experiments on eight benchmark datasets show that IMSAT produces state-of-the-art results for both clustering and unsupervised hash learning. For instance, IMSAT achieved 98.4% clustering accuracy on the MNIST dataset, which significantly outperforms DEC with 84.3% (Xie, Girshick, and Farhadi, 2016), which is the state-of-the-art clustering method. Full version of the paper is available at `https://arxiv.org/abs/1702.08720`.

References

Gomes, R.; Krause, A.; and Perona, P. 2010. Discriminative clustering by regularized information maximization. In *NIPS*.

Xie, J.; Girshick, R.; and Farhadi, A. 2016. Unsupervised deep embedding for clustering analysis. In *ICML*.