

Prototypical Networks for Few-shot Learning

Björn Bebensee (2019–21343)
Topics in Artificial Intelligence

November 12, 2019

In the few-shot classification task a classifier must classify elements from new unseen classes (query set) not seen in the training set while only given a few labeled examples of each new class (support set). Snell, Swersky and Zemel [1] propose a new approach to few-shot classification using *Prototypical Networks*. On a high-level, their approach aims to address overfitting by learning an embedding for the inputs in which points cluster around a prototype representation for each class.

More specifically, given a support set $S = \{(x_1, y_1), \dots, (x_N, y_N)\}$ of features $x_i \in \mathbb{R}^D$ and their corresponding labels $y_i \in \{1, \dots, K\}$, Prototypical Networks compute an M -dimensional prototype representation as the mean of the embeddings of the support set using a learned embedding function $f_\phi : \mathbb{R}^D \rightarrow \mathbb{M}$. Then, given some distance metric d and a query point x we can obtain a distribution over classes p_ϕ as its distance in the embedding space to the different class prototypes. During training, we aim to minimize $J(\phi) = -\log p_\phi(y = k|x)$ for the true class k . The authors point out the importance of the choice of the distance function as Prototypical Networks are equivalent to mixture density estimation on the support set for *regular Bregman divergences*, a class of distance functions. Furthermore, for Bregman divergences, the prototype computation by the mean achieves optimal cluster representatives, when a Bregman divergence is used as a distance function. Snell et al. decide to focus on Euclidean distance, a Bregman divergence, here as they assume that any non-linearity can be learned in the embedding function. They also extend their model to the zero-shot learning setting in the following way: given a class meta-data vector v_k for each class k , define the class prototype as an embedding of v_k .

The authors evaluate their approach on Omniglot and the miniImageNet version of ILSVRC-2012 for the few-shot learning setting, and the 2011 Caltech UCSD bird dataset for the zero-shot setting. They achieve state-of-the-art results in both settings with this much more simple model and show that a careful choice of distance metric can further improve results not only for Prototypical Networks but also for Matching Networks (a competing method). Overall, Snell et al. believe the simplicity of Prototypical Networks to be their greatest advantage. They suggest further study of Bregman divergences other than Euclidean distance for Prototypical Networks.

References

- [1] Snell, Jake, Kevin Swersky, and Richard Zemel. "Prototypical networks for few-shot learning." Advances in Neural Information Processing Systems. 2017.