Title

Attribute Prototype Network for Zero-Shot Learning

1st Author

Jiuniu Wang - Wenjia Xu

Submission Date

19 Aug 2020

Publisher

ACM

Abstract:

From the beginning of zero-shot learning research, visual attributes have been shown to play an important role. In order to better transfer attribute-based knowledge from known to unknown classes, we argue that an image representation with integrated attribute localization ability would be beneficial for zero-shot learning. To this end, we propose a novel zero-shot representation learning framework that jointly learns discriminative global and local features using only class-level at tributes. While a visual-semantic embedding layer learns global features, local features are learned through an attribute prototype network that simultaneously regresses and decorrelates attributes from intermediate features. We show that our locality augmented image representations achieve a new state-of-the-art on three zero-shot learning benchmarks. As an additional benefit, our model points to the visual evidence of the attributes in an image, e.g., for the CUB dataset, confirming the improved attribute localization ability of our image representation.

My View:

In zero-shot-learning we need to design a model which can predict properly on the test dataset that contains unseen classes. This paper proposes an approach so that the model (APN model) extracts some local features from images in the train dataset and by using these features, it can learn some prototype attributes which are useful at test time with unseen classes. Because, these attributes are adopted from local features not global features, so the model do not learn likelihood Ly. On the other hands, the model is learning some semantic attributes that prevent model to be biased toward seen classes. In addition, it aims to improve the localization of image representation.

Previous Idea:

- Visual attention
- Rely on pretrained image representation
- Focus on learning compatibility function between image representation and attributes

Previous Idea weakness:

Previously, model learns some global features which describe the global content of image. This approach accurately focuses on some object parts, often the discovered parts and attributes are biased toward training and seen classes due to the learned correlation. In addition, some features co-occur frequently which may be learned as a shortcut to maximize the likelihood of training data and therefor fail to deal with unknown combination or configuration of features which have not been seen in training time.

Paper solution:

The author aims to improve the localization of image representation and learn prototype attributes based on local features as it can. Due learning process, it tries to collect unrelated and semantic attributes by injecting losses on intermediate layers and enforce the local features to encode semantic attributes. These attributes are class-level and improve the class embedding. By making these attributes unrelated, it prevents increasing the likelihood of training data, in addition, by localizing the extracted features, the model can decide and work well on combination of some unobserved features on unseen classes.

Result:

This model (APN) is trained on three different dataset and by comparing to BaseMod(which is a part of APN) it is more accurate but not very good.

In my opinion, it is better that we execute some preprocess on data, such as, semantic segmentation. If we first cut the object and then give it to image encoder (backbone), the prototype attributes which model learn, are more related to object parts and unrelated to background. So, I think it may make the model more robust.