TAFE-Net: Task-Aware Feature Embeddings for Low Shot Learning

Purvanshi Mehta

October 17, 2019

1 Summary

1.1 Problem

In current few and zero shot learning scenario feature embeddings are re-used from existing convolution neural networks trained on large-scaled labeled training datasets. The embeddings are common irrespective of the prediction task. Often the image embedding may be out of domain or sub-optimal for the task.

1.2 Innovation

The authors propose the idea of using task - aware feature embedding (TAFE-Net). The generic image features are transformed to task aware embeddings through meta learning.

1.3 Contributions

The TAFE-Net model consists of one task aware meta learner G and a prediction network F. The extracted generic features of an image from a pre-trained model are fed through the dynamic feature layers of F whose parameters are meta - learned by G. F predicts whether the image is relevant to the current task.

To generate feature weights for F the output dimension of g_i must be equal to the weight size of i-th feature layer in F. Since the parameterization of weight generators g can consume a large amount of energy and memory, the authors propose a *Weight Factorization technique*. This takes on the assumption that channels of a convolution layer may have orthogonal functionality. The weight of ith layer W_i is decomposed into a weight W_S^i which is shared among all tasks, into W_t which is the task specific parameter and into $*c_out$ which is the grouped convolution. Thus G only needs to generate the task specific weight W_i^t.

Since F gives a binary prediction for every image relevance to the current task, to distinguish which task the image belongs to a multi-class cross entropy loss is used. For the embedding loss, a metric based learning approach is used for positive inputs of a given task. Cosine similarity is used as the distance metrics. The over all loss becomes $L_{cls} + \beta.L_{emb}$ where β is the hyperparameter to balance two terms.

1.4 Evaluation

The experiment work has been done on - zero shot learning, unseen attribute-object composition task and few shot learning.

- **Zero shot learning** Results are presented on 5 datasets, with state of the art results on 3 tasks. Analysis has also been done on with and without using the embedding loss which indicates clear improvemenst when the embedding loss is used.
- **Unseen Visual-attribute Composition** The most challenging task taken up by the paper is on predicting the attributes given an image in a zero shot way. The results are also compared by ablation studies using different type of feature embeddings. TAFE -net achieves state of the art results on both MITStates and StanfordVSD dataset.
- **Few shot image classification** - Unlike other few shot learning basemarks which expriment with few classes and low resolution images, the evaluation on their model is done on ImageNet images which contains hundreds of classes divided into base classes and novel classes.The goal is to obtain high accuracy on the novel classes without sacrificing the performance on the base classes. The comparison has been done prototypical networks, Logistic regression and Matching Networks. TAFE-Net performs comparable to the State-of-the-Art methods without using hallucinator.

1.5 Substantiation

The authors claim to generate task aware embeddings instead of using common embeddings for all tasks. They approach this by adding an additional embedding loss to improve the performance of the meta learner and generate the weight factorization scheme to generate the meta learned weights more efficiently. The authors provide t-SNE visualization for zero shot and unseen attribute prediction tasks where the task aware embeddings distinguish from each other. The authors achieve state of the art results which validate their approach empirically and their claim to use task aware embessings.

2 Review

- The authors emphasize the importance of task aware embeddings but they do not lay an emphasis on how their generated embeddings differ from each other. Simply saying they do not provide an emphasis on **Interpretability** of the images that are ignored by the current task. The model F is a predictor which gives a bianry prediction of which image is relevnt or irrelevant to the current task. The question to ask here is **What images are being ignored by F and why**, do these images share a common feature?

For solving this an analysis could be done on the images that are taken by the model into consideration for a task and the images that are being ignored. One way to analyse is compare their embedding space. Various methods have been proposed like similarity metrics to compare two objects in latent space. The analysis would not only help us in understanding how does a Neural Network form task relevant features but also what are these features which common to all selected images which it keeps.

- The authors used **cosine similarity metrics** as the distance measure. In [1] the authors justify the use of class mean as prototypes when distances are computed with a Bregman divergence, such as squared **Euclidean distance**. Experiments could be conducted to test the efficiency while using Euclidean distance as metrics. Theoretically the accuracy must increase if we do so.

- The problem described in the paper i.e of learning task specific features is not specific to few shot learning but also to other computer vision and language processing domains. An extension of this method could be in other CV domains which learn task specific embeddings. The problem could specipically be applied to multi task settings where in the ideal scenario same embeddings are used to different tasks. The *dynamic* embeddings could be meta learned in a similar way as described in the paper, but here our tasks could be different regression or classifcation task rather than classification of zero or few shot examples. The embedding loss would be replaced the loss of each task.

- Today pre trained weights from Imagenet are being used the most for transfer learning. Is every feature from Imagenet necessary to the current transfer learning task? The concept of meta learning task specific features could be applied here as a fine tuning mechanism.

References

[1] Jake Snell, Kevin Swersky, and Richard S. Zemel. Prototypical networks for few-shot learning. *CoRR*, abs/1703.05175, 2017.