Finding Task-Relevant Features for Few-Shot Learning by Category Traversal

Purvanshi Mehta

October 10, 2019

1 Summary

1.1 Problem

Metric based few shot learning works by learning a model which can separate the feature space into several clusters. Given a query we embed it, compare the distance to the support samples and assign the labels of the support class closest to the query. These approaches treat the support class independently from each other and do not look at the entire task as a whole. This constraints to use the same set of features for all possible tasks which in turn hinders the ability to distinguish the most relevant dimensions for the task at hand.

1.2 Innovation

The paper proposes a Category Traversal Module(CTM) which traverses across the entire support set at once, identifying task-relevant features based on both intra-class commonality and inter-class uniqueness in the feature space.

1.3 Contributions

The major contribution of the paper is in identifying the problem of using task dependant inter and intra-class support set features.

They do so by proposing a CTM module. The first part of CTM is the *concentrator* which finds the universal features shared by all instances for one class. This is achieved by a CNN architecture followed by downsampling.

The second component is the *projector* which removes the irrelevant features and selects the most relevant ones by looking at the concentrator features from all support categories at the same time. The class prototypes are concatenated and a CNN is applied followed by a softmax to obtain a mask.

1.4 Evaluation

The experiments have been conducted on miniImageNet and teiredImageNet. The baseline has been compared on 1-shot and 5-shot both shallow (4 layer NN)

and deep network (ResNet 18), with clear improvements when both concentrator and projector modules are used.

They also embed their CMT module inside - Matching Networks, Relation Networks and Prototypical Networks and show consistent improvements when CMT module is used.

The authors also show t-SNE visualisations of relation network with a CTM module and without a CTM module.

1.5 Substantiation

The authors claim to investigate their approach by comparing with the state of the art techniques, deploying CTM as a plug and play module in any metric learning technique and lastly they investigate if their method makes features more discriminative and representative.

Through their experiments and comparisons with previous metric based learning techniques all their claims were proved correct. The major selling point of the paper that is to form distinctive representations was satisfied by t-SNE visualizations.

2 Review

One paper critic and two paper extensions have been provided.

- The authors propose a valid problem in the scenario of metric based learning. They also propose a method to distinguish and eventually find task relevant features. But no emphasis has been made upon the embedding of the support class itself.

  Another way to see the problem stated in the paper is as an embedding learning problem. With techniques such as Word2Vec being highly successful in the domain of NLP, we can leverage the task of similar and distinct feature learning for forming more distinguished features. The loss would make feature embeddings of two similar train samples closer to each other and increase the distance between two dissimilar train samples.

- The distance metrics could be mapped by Euclidean distance or another Neural Network. In fact the concept of **Relation Networks could be extended** to another module which contains inter class embeddings. Instead of just forming the intraclass embedding of the support set X_i in the support set S we can form two embeddings - inter (By using a CNN or LSTM) and intra class. The three metrics now can be combined instead of just two

  $$C(f_\phi(x_{query}), f_\phi(x_{interclass}), f_\phi(x_{intraclass}))$$

  Where C is the concatenation operation as defined in the paper. The combined embeddings should be mapped to a value between 0 to 1. Where matched pairs have a similarity 1 and unmatched have a similarity 0.

- The problem stated in the paper can be applied to other fields - Like *Multimodal Learning*. In multimodal Learning a big problem is of the data being noisy and redundant in several modalities. The question which comes to my mind is - *Can a similar masking technique help in extracting inter modality relevant features and give us more distinct representations?* This could be achieved by concatenating last layer features of all modalities, applying a CNN and finally applying a softmax to produce a mask.