Paper: *Siamese Neural Networks for One-shot Image Recognition*

Nirajan Koirala

CSC 9010_001

Summary

Machine Learning methods have been successfully used to achieve state-of-art results in a variety of real-world applications from web-search to image recognition. These methods and algorithms however often break down when the data available for training them is very low. Authors in this paper explore a method for learning siamese neural networks which can help us address one-shot learning by ranking similarities between the inputs. They present a novel approach which limits assumptions on the structure of the inputs while automatically acquiring features which enable the model to generalize successfully from few examples. They experiment with character recognition tasks and their siamese convolutional NN are capable of predicting unknown class distributions given very few examples. Also, they can be trained using standard optimization techniques and this approach doesn't depend upon domain-specific knowledge.

First the neural network is trained to discriminate between the class-identity of image pairs. This model learns to identify input pairs according to the probability that whether they belong to the same or different class. It can be then used to evaluate new images, exactly one per novel class in a pairwise manner and the pair with the highest score according to the verification network is awarded the highest probability for the one-shot task. In section 3.1, they provide description of the model they use which is a siamese convolutional NN with L layers each with different units. RELU units are used in the first L-2 layers and sigmoidal units in the remaining layers. They provide description of the convolution layer later in the section with appropriate figures and formulas. Regularized cross-entropy objective is imposed on the binary classifier and this objective is combined with standard backpropagation algorithm for the twin networks to learn. They describe further techniques they've used like learning schedule and hyperparameter optimization. Additionally, they augment the training set with small affine distortions as well.

For the experiments the model is trained on a subset of Omniglot dataset. The dataset contains examples from 50 alphabets ranging from Latin to local dialects and also contains some fictitious character sets. About 10 to 40 characters are present in each of the alphabets and dataset is split into background and evaluation set. The background set is used for developing a model by learning hyperparameters and feature mappings and the evaluation set is used for measuring one-shot classification performance. For verification, they train the verification network using 3 different datasets with varying training examples by sampling random same and different pairs. They describe the strategy they used to monitor the performance of network during training and provide a table which lists the final verification results for each of the training sets. Once their siamese network is optimized for the verification task, they use it for one-shot learning. They describe the process of its evaluation and it is used to calculate the classification accuracy. Their method is able to produce 92% accuracy and unlike other models it is able to achieve it without including any extra prior knowledge about the characters or strokes which is the primary advantage of their model. They also try their approach on the

MNIST dataset to monitor how well their model which is trained on Omniglot can generalize on MNIST. They found that the network is still able to achieve reasonable generalization from the features learned on Omniglot without training at all on the MNIST.

Using the strategy for performing one-shot classification by first learning deep convolutional siamese neural network, the network is able to outperform all the available baselines by a significant margin. Authors argue that this approach should extend to one-shot learning tasks in other domains. They mention that they have been experimenting with an extended algorithm that exploits the data about individual stroke trajectories to produce final computed distortions and by imposing local affine transformations on these strokes they home that they can learn features which are better adapted to the variations commonly seen in new examples.

Strengths

- A novel approach is used for building models which do not require a lot of labelled data.
- Paper is well organized and points made are clear.

Weaknesses

- I think the architecture of the model used could have been elaborated a further.

Points of Confusion

- I am confused when they mention that the network is symmetric in section 3.
- What is a Bayesian Hyperparameter optimization framework?

Discussion Questions

- Can one-shot learning be used for other tasks like Reinforcement Learning and NLP?
- If we use this approach for large datasets CIFAR or ImageNet could it be able to find similarities from random real world images?
- A discussion on section 4.2 which discusses about the verification of the networks.