Siamese Neural Networks for One-Shot Image Recognition

Yakuncheva Irina
Higher School of Economics
Applied mathematics and informatics
Moscow, Russia
irka1804@mail.ru

Abstract **— As a rule, with a deep study we need a large amount of data, and the more we have, the better the results. However, there are cases when available data is too small, and the process of machine learning becomes difficult. Therefore, there is an idea to train a model consisting of only a few copies. This article explores a method for studying Siamese neural networks that determine the similarity between inputs. Thanks to a convolutional architecture, you can achieve results which exceed those of other deep learning models with near state-of-the-art performance on One-Shot classification tasks.**

Keywords**—Siamese Neural Networks; One-Shot; Omniglot; machine learning.**

INTRODUCTION

People have the ability to memorize and recognize new patterns. In particular, people can also quickly understand new concepts and then determine their various variations in the future. This principle applies more broadly than just to generalizing to unseen images of a particular concept class. This type of inference can therefore be very useful but it is not necessarily an inherent property in our models, which may be closely fit to the available classes without learning useful structure for other classes. Classification under the restriction that we may only observe a single example of each possible class before making a prediction about a test instance is particularly interesting task. This is called One-Shot learning and it is the primary focus of our model presented in this work.

The problem of One-Shot learning can be directly addressed by developing domain-specific features or inference procedures which possess highly discriminative properties for the target task. As a result, systems which incorporate these methods tend to excel at similar instances but fail to offer robust solutions that may be generally applied to other types of problems. In this paper, was 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.

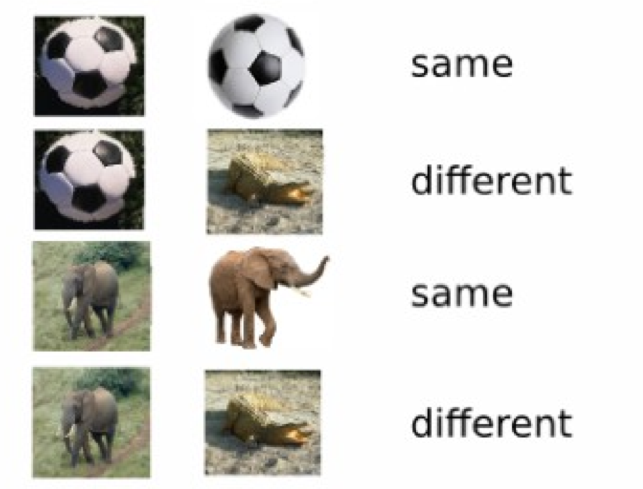
MAIN PART

Model

For One Shot classification models, only one training grade for each class is sufficient. The Siamese network consists of two identical neural networks, each of which takes one of two inputs. They have the same configuration with the same parameters and weights. Updating the settings is the same in both networks.

The basic idea is that the Siamese network receives two inputs. Each input is subjected to a dimensional reduction transformation, implemented as a neural network. As the number of parameters decreases, the neural network becomes less prone to retraining. As a result, we get two vectors that are somehow compared. Thus, instead of classifying their inputs, the neural network learns to distinguish between them, or, conversely, to calculate similarities.

For example, if the input data is two animals, then the neural network determines whether they belong to the same type.



Also, the input can be submitted in the form of a question-answer. It possible when one input is a questionable sentence, another input is the answer, and the result is how relevant the answer to the question is. The questions and answers do not look exactly the same, but if the goal is to extract similarities or the relationship between them, the Siamese architecture can also work well.

Experiment

Siamese network was trained on a subset of the Omniglot data set, which will escribed. The Omniglot data set was collected by Brenden Lake and his collaborators at MIT via Amazon's Mechanical Turk to produce a standard benchmark for learning from few examples in the handwritten character recognition domain. Omniglot contains examples from 50 alphabets ranging from well-established international languages like Latin and Korean to lesser known local dialects. To train this verification network, were put together three different data set sizes with 30,000, 90,000, and 150,000 training examples by sampling random same and different pairs. Was setted aside sixty percent of the total data for training. Also, affine distortions were added. An additional copy of the data set corresponding was produced to the augmented version of each of these sizes. Eight transforms were added for each training example, so the corresponding data sets have 270,000, 810,000, and 1,350,000 effective examples. In the table below (Table 1), list the final verification results for each of the possible training sets, where the listed test accuracy is reported at the best validation checkpoint and threshold. Results were reported across different training runs, varying the training set size and toggling distortions.

Table 1. Accuracy on Omniglot verification task (siamese convolutional neural net)

Method	Test
30k training	
no distortions	90.61
affine distortions x8	91.90
90k training	
no distortions	91.54
affine distortions x8	93.15
150k training	
no distortions	91.63
affine distortions x8	**93.42**

Conclusion

This article introduced a strategy for implementing One Shot classification. For this, deep convolutional Siamese neural networks were studied. During the research, new results were obtained that compared the performance of the received networks with the existing modern classifier developed for the Omniglot data set. For a more qualitative analysis, the article examined training for the verification task by processing pairs of images and their distortions using a global affine transformation. Networks have shown strong performance, which indicates not only that the accuracy of this level can be achieved through this approach to metric learning, but also that this approach should extend to teaching tasks in other areas. For example, to classify images. In the future, it is planned to find functions that will be better adapted to the variations encountered in the new examples.

References

[1] Srivastava, Nitish. Improving neural networks with dropout. Master's thesis, University of Toronto, 2013.

[2] Lake, Brenden M, Salakhutdinov, Ruslan, Gross, Jason, and Tenenbaum, Joshua B. One shot learning of simple visual concepts. In Proceedings of the 33rd Annual Conference of the Cognitive Science Society, volume 172, 2011.

[3] Fei-Fei, Li, Fergus, Robert, and Perona, Pietro. One-shot learning of object categories. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 28(4):594– 611, 2006.