Research Proposal:

Semi-supervised learning of deep hierarchical hidden representations

Miquel Perelló Nieto, email: miquel@perellonieto.com

March 4, 2015

Abstract – During last decade, models that learn a hierarchical representation of the data have been achieving state-of-the-art performance in image classification. However, it is still unclear if purely supervised methods are better than semi-supervised. Besides, the proportion of unlabeled data is always larger than labeled.

This research proposes to investigate new algorithms to exploit all the available data for a pattern recognition tasks.

Motivation and Background

Deep Learning emerged as a new and promising field in the machine learning community. The term gained popularity in 2006 when Hinton et.al. [1] proposed an unsupervised method to pre-train a Deep Belief Network (DBN). At that time, Hinton demonstrated the need of unsupervised learning to find a good initialization of the parameters. This pre-training – and consequently the weight initialization – was shown to be close enough to a good local minima. The idea was to train a stack of Restricted Boltzmann Machines (RBMs) and extract a hierarchy of hidden representations from the input space. This method allowed to pre-train the network with unlabeled images if necessary [4], and shown to be useful for datasets with small input size like MNIST or CIFAR10 (28x28 and 32x32 pixels respectively). Though, the pre-training was computationally expensive for large scale datasets like ImageNet (natural images about 480x410 pixels).

During the following years, several research groups explored new training methods, initializations and activation functions [2]. Some of the analysis probed that the unsupervised pre-training phase was not necessary for object recognition tasks.

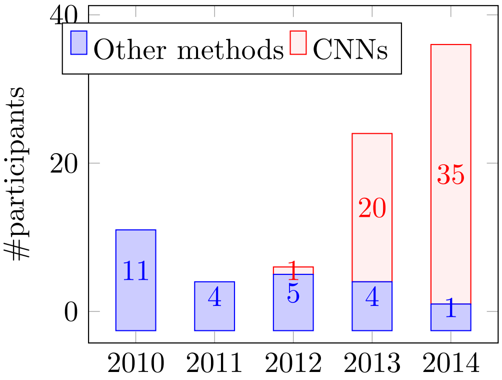

Figure 1: Proportion of participants using CNNs and other models in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

After these outcomes, one of the most successful deep architectures [3] was able to beat the state-of-the-art models on natural image classification by a large margin. Earlier models required complex hand-crafted methods to find an appropriate set of features. In contrast, training a deep Convolutional Neural Network (CNN) from raw pixels solved the task without apparent problems. Consequently, almost all computer vision research groups became interested in the hidden representation extracted by CNNs.

Since 2012, most of the image classification tasks are being solved using supervised learning and CNNs (See example on Figure 1). However, these models rely on large amounts of labeled data, and whilst the accessible data is considerable, labeled data is expensive and consequently not that large. Therefore, it is necessary to find algorithms and methods that exploit this asymmetry.

Proposal

If we assume that the data lies in a manifold with a lower dimensionality than the input space, then it must be possible to explore new techniques able to learn these shapes using all the available data.

The focus of this research is not to improve the accuracy on classification challenges like the ILSVRC, but to find new algorithms that generalize to large test sets with a realistic amount of unlabeled data.

The initial step is to create sizable artificial datasets to test the algorithms, or to use large datasets with small input dimensionality. If good results are achieved, there exist datasets where most of the samples do not include label. For example, video datasets are still in an early phase and are not densely annotated (eg. Hollywood2, YouTube Faces, UCF datasets). Other possible candidates are audio datasets (music, speech conversations or audio from videos). In some of these examples the use of new architectures that exploit the temporal or spatial invariance can be preferred. Consequently, it is advisable to try architectures that incorporate convolutions.

One possible idea for semi-supervised learning will consist on: (1) train a discriminative model to find a hidden representation of the data, (2) use a generative model to learn the statistical distribution of the hidden representation at different hierarchical levels, (3) use this information to assign certain levels of confidence on the classification of unlabeled data, (4) incorporate the new labels in the training with a different learning rate (or "confidence level").

With this idea in mind, I propose to use as a discriminative models: CNNs, Multilayer Feedforward Neural Networks (MFNNs), or Recurrent Neural Networks RNNs. And as a generative models Denosing Autoencoders (dAEs), RBMs, Neural Autoregressive Distribution Estimators (NADEs), or once more RNNs.

Conclusion

It is still unclear if purely supervised methods are a better learning approach than semi-supervised methods. Heretofore, both methods have shown to achieve state-of-the-art results. However, only semi-supervised learning considers the real proportion of labeled and unlabeled data. In addition, we know empirically that humans are able to learn from unlabeled data.

This research proposes to elaborate a study of new algorithms and methods to exploit unlabeled data, without loss of generality.

References

[1] GE Hinton, S Osindero, and YW Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 1554:1527–1554, 2006.

[2] Kevin Jarrett, Koray Kavukcuoglu, Marc' Aurelio Ranzato, and Yann LeCun. What is the best multi-stage architecture for object recognition? *2009 IEEE 12th International Conference on Computer Vision*, pages 2146–2153, September 2009.

[3] Alex Krizhevsky, I Sutskever, and GE Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *NIPS*, pages 1–9, 2012.

[4] QV Le, MA Ranzato, R Monga, and Matthieu Devin. Building high-level features using large scale unsupervised learning. *arXiv preprint arXiv: ...*, 2011.