Informatics Project Proposal - Automatic Curriculum Learning for Deep Models Using Active Learning

Ian McWilliam s0904776

Contents

Abstract

In this paper we propose a project researching how active learning methods can be used to automatically construct training curricula, with the aim being to test the hypothesis that training deep models with such 'active curricula' can improve upon other training paradigms such as random sampling, pre-training or traditional curriculum learning. The work will investigate a variety of methodologies for constructing a curriculum using active learning metrics, focusing on how the different training methods affect the performance of convolutional networks on a range of image classification tasks. The output of the project will hopefully be a set of novel ways to improve the training of deep models, as well as a more thorough exploration of the relationship between active and curriculum learning than currently exists in the literature.

1 Introduction

1.1 Active Learning

Active learning refers to a learning paradigm wherein machine learning algorithms actively select, or 'query', the samples from which it learns; in the case of deep learning this contrasts with the usual approach of randomly sampling labeled training samples. The motivation for active learning is that, by allowing it to intelligently select the samples from which it learns, the algorithm can achieve superior generalization performance from a smaller number of training samples than if the samples had been chosen randomly. In domains where unlabeled data is abundant but obtaining labels is expensive, active learning can be used to reduce the cost associated with training a deep model, as the active approach allows the designer to obtain labels only for the samples which will be most beneficial to learning.

There are a variety of methods by which the algorithm can query datapoints, however they generally focus on finding the points in the input space that the algorithm is most 'uncertain' about, allowing the algorithm to fill what could be seen as gaps in its knowledge of the domain. OUTLINE SOME MAIN ACTIVE LEARNING METHODS, THEN DISCUSS DEEP MODEL SPECIFIC METHODS SUCH AS IN YARIN GAL ETC.

1.2 Curriculum Learning

A related field is that of *curriculum learning*, which explores how the learning process can be improved by presenting training samples to the algorithm in a meaningful order (with the order defining a 'curriculum'), again in contrast to simply sampling randomly from a training set. Motivated by the way in which humans and animals learn, a curriculum generally begins with 'easy' examples, before transitioning to more challenging ones, with the aim being to improve generalization performance and convergence speed. TALK ABOUT BENGIO PAPER ETC. It is interesting to note that, while similar, active and curriculum learning are somewhat opposed in their learning philosophies, with the former focusing on learning from uncertain/difficult samples and the latter focusing beginning with easy samples before continuing to more difficult ones.

Bengio et al [2] find a theoretical motivation for curriculum learning by suggestings that it can be seen as a *continuation method* (a method of optimisation non-convex functions). Curriculum learning has also drawn similarities with *transfer learning*, as the early, 'easier' stages of training can be seen as a separate task, with the network weights then being used as the initial weights for training on the more difficult samples. Similiarly curriculum learning can be seen as a form of pre-training, comparing the method to unsupervised pre-training methods as in [4], where pre-training allows the supervised learning to begin in a region of parameter space that led to superior generalization performance.

2 Purpose

2.1 Curriculum Construction

One of the main difficulties with implementing curriculum learning is that the curriculum must be constructed prior to training, requiring some predefined measure of difficulty with which to order the training samples. In certain domains there may be a natural ordering of difficulty, for example in the geometric shapes example in [2], however in many tasks manually constructing a curriculum is not as straight forward. This motivates the development of methods to automatically construct learning curricula without requiring expert domain knowledge; in this paper we suggest that using active learning metrics may be used to automatically construct such curricula, potentially providing an efficient way to improve the training of deep learning algorithms.

2.2 Active Curricula

Active learning methodologies generally involve calculating a measure of the algorithm's 'uncertainty' in predicting the label of a given input sample. In the traditional active learning setting, the algorithm then queries the label of the sample(s) about which it has the most uncertainty and is then trained using the selected sample(s). An alternative approach is to view the uncertainty measure as a way of approximating the 'difficulty' of a given input sample, which can then be used to construct a learning curriculum, which we term an 'active curriculum'.

2.3 Curriculum Construction

As discussed in the introduction, active and curriculum learning offer a somewhat dichotomous view on how best to order training samples. In active learning the samples about which the algorithm is most uncertain are chosen to train on, while in curriculum learning training focuses on training on easier samples, before progressing to more difficult ones. This trade off is one that has not been explored in the literature, and it is a dimension on which we will focus in the proposed work.

Given an uncertainty/difficulty metric using active learning methods, we have the option of constructing a curriculum beginning with training on the least uncertain samples, or alternatively training on the sa

3 Background

3.1 Automatic Curriculum Construction

3.2 Curriculum Learning for Deep Models

4 Methods

4.1 Active Learning Metrics

4.2 Curriculum Construction Methods

Constructing curricula with a set of Active Learning methods. Various curriculum constructions, select easy, select hard, easy to hard, hard to easy

5 Evaluation

Compare to pre-constructed curricula, random sampling, pre-training. Test on several tasks, i.e. geometric shapes, MNIST, CIFAR. Also discuss the ease with which it can be applied vs pre-training or standard curriculum learning.

6 Outputs

New training paradigms, hopefully generic ways to consistently improve on random sampling

7 Workplan

References

[1] E.L.Allgower and K.Georg, "Numerical continuation methods. An introduction", *Springer-Verlag*, 1980

[2] Y.Bengio, J.Louradour, R.Collobert and J.Weston, "Curriculum Learning", *Proc. Int'l Conf. Machine Learning*, 2009

[3] D.Cohn, L.Atlas and T.Ladner, "Improving Generalization with Active Learning", *Machine Learning*, Issue 15, p201-221, 1994

[4] D.Erhan, P.A.Manzagol, Y.Bengio, S.Bengio and P.Vincent, "The difficulty of training deep architectures and the effect of unsupervised pre-training", *AI & Statistics*, 2009

[5] Y.Gal, R.Islam, Z.Ghahramani, "Deep Bayesian Active Learning with Image Data", *NIPS, Bayesian Deep Learning workshop*, 2016

[6] A. Graves, M. G. Bellemare, J. Menick, R. Munos, and K. Kavukcuoglu. "Automated curriculum learning for neural networks", *Proc. Machine Learning Research*, 70, p1311-1320, 2017

[7] A.Katharopoulos and F.Fleuret, "Not All Samples are Created Equal: Deep Learning with Importance Sampling", arXiv preprint, arXiv:1803.00942, 2018

[8] N.Justesen, S.Risi, "Automated Curriculum Learning by Rewarding Temporally Rare Events", arXiv preprint, arXiv:1803.0713, 2018

[9] M.Pawan Kumar, B.Packer, D.Koller, "Self-Paced Learning for Latent Variable Models", *Advances in Neural Information Processing Systems 25*, 2010

[10] B.Settles, "Active Learning Literature Survey", Computer Sciences Technical Report 1648, 2009

[11] D.Weinshall, G.Cohen, "Curriculum Learning by Transfer Learning: Theory and Experiments with Deep Networks", arXiv preprint, arXiv:1802.03796, 2018