Application Feedback in Guiding a Deep-Layered Perception Model

Itamar Arel[a] and Shay Berant[b]
[a]*Department of Electrical Engineering & Computer Science, University of Tennessee*
[b]*Binatix, Inc., Palo Alto, CA*

Abstract. Deep-layer machine learning architectures continue to emerge as a promising biologically-inspired framework for achieving scalable perception in artificial agents. State inference is a consequence of robust perception, allowing the agent to interpret the environment with which it interacts and map such interpretation to desirable actions. However, in existing deep learning schemes, the perception process is guided purely by spatial regularities in the observations, with no feedback provided from the target application (e.g. classification, control). In this paper, we propose a simple yet powerful feedback mechanism, based on adjusting the sample presentation distribution, which guides the perception model in allocating resources for patterns observed. As a result, a much more focused state inference can be achieved leading to greater accuracy and overall performance. The proposed paradigm is demonstrated on a small-scale yet complex image recognition task, clearly illustrating the advantage of incorporating feedback in a deep-learning based cognitive architecture.

Keywords. Deep-layered machine learning, perception, spatiotemporal inference.

Introduction

Perception is at the core of intelligent systems. The vast amount of information that humans (and advanced robotic systems) are exposed to every second of the day is driven by sensory inputs that span a huge observation space. The latter is due to the natural complexity of the world with which such systems interact. This inestimable amount of information must be somehow efficiently represented if one is to successfully function in the real-world. Deep machine learning (DML) is an emerging field [1] within cognitive computing which may be viewed as a framework for effectively coping with vast amounts of sensory information.

One of the key challenges facing the field of cognitive computing is perception in high-dimensional sensory inputs. An application domain in which this challenge clearly arises is pattern recognition in large images, where an input may comprise of millions of pixels. These millions of simultaneous input variables span an enormous space of possible observations. In order to infer the content perceived, a system is required to map each observation to a possible set of recognized patterns. However, due to a phenomenon known as the *curse of dimensionality* [2], the complexity of training a system to map observations to recognized pattern classes grows exponentially with the number of input variables. Such growth primarily pertains to the number of examples the system is required to be presented with prior to becoming adequately proficient.

A common approach to overcome the curse of dimensionality is to pre-process the data in a manner that reduces its dimensionality to such a level that can be effectively processed by a classification module, such as a multi-layer perceptron (MLP) artificial neural network. Such dimensionality reduction is often referred to as feature extraction. Its goal is to retain the key information needed to correctly classify the input within a lower-dimensional space. As a result, it can be argued that the intelligence behind many pattern recognition systems has shifted to human-engineered feature extraction processes, which at times are very difficult and highly application-dependent. Moreover, if incomplete, distorted or erroneous features are extracted classification performance may degrade significantly.

Some recent neuroscience [3][4] findings have provided insight into the principles governing information representation in the mammal brain, leading to new ideas for designing systems that represent information. One of the key findings has been that the neocortex, which is associated with many cognitive abilities, does not explicitly pre-process sensory signals, but rather allows them to propagate through a complex hierarchy of modules that, over time, learn to represent observations based on the regularities they exhibit. Such hierarchical representation offers many advantages, including robustness to diverse range of noise and distortions in the data, as well as the ability to cope with missing or erroneous inputs.

DML continues to emerge as a promising, biologically-inspired framework for complex pattern inference. A key assumption in DML is that representation is driven by regularities in the observations. As one ascends the hierarchical architecture of DML systems, more abstract notions are formed. Hence, in higher layers of the hierarchy scope is gained while detail is lost. This appears to be a pragmatic trade off, as well as a biologically plausible one. In the context of artificial general intelligence (AGI) [5], one can view perception as being identical to modeling data, in the sense that partial observations of a large visual field are utilized in inferring the state of the world with which the agent interacts.

In most existing deep learning schemes [6][7] there is either none or weak relationship between the (unsupervised) training of the model (DML) engine and the decision making modules. This forces DML systems to form a representation purely based on regularities in the observation rather than being driven also by the application at hand (e.g. visual pattern recognition). It is well known, for example, that neurons in layer IV of the neocortex receive all of the synaptic connections from outside the cortex (mostly from thalamus), and themselves make short-range, local connections to other cortical layers. This suggests that learning may not be driven exclusively by regularities in the observations, but rather co-guided by external signals.

In this paper we present an elegant methodology for guiding the representation of a DML system such that it serves as a more relevant perception engine, yielding improved classification accuracy. The approach is based on adjusting the DML sample presentation distribution as it is trained such that relevant salient features can be hierarchically captured.

The rest of this paper is structured as follows. In section 1 we outline the proposed deep learning system and its operational modes. Section 2 describes the proposed feedback-based scheme for guiding DML representation. Section 3 describes the simulation results while in Section 4 conclusions are drawn.

1. Deep-layered Inference Engine

The proposed DML architecture comprises of a hierarchy of multiple layers each hosting a set of identical cortical circuits (or nodes), which are homogeneous to the entire system, as illustrated in Figure 1. Each node models the inputs it receives in an identical manner to all other nodes. This modeling, which can be viewed as a form of lossy compression, essentially represents the inputs in a compact form that captures only the dominant regularities in the observations. The system is trained in an unsupervised manner by exposing the hierarchy to a large set of observations and letting the salient attributes of the inputs be formed across the layers. Next, signals are extracted from this deep-layered inference engine to a supervised classifier for the purpose of robust pattern recognition. Robustness here refers to the ability to exhibit classification invariance to a diverse range of transformations and distortions, including noise, scale, rotation, displacement, etc.

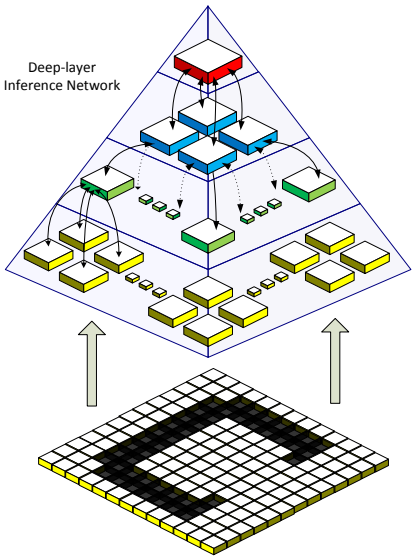

Figure 1. Deep-layered visual perception network comprising multiple layers hosting identical cortical circuits. The lowest layer of the hierarchy receives raw sensory inputs. Features generated by the cortical circuits are passed as input to a supervised classifier.

The internal signals of the cortical circuits comprising the hierarchy may be viewed as forming a feature space, thus capturing salient characteristics of the observations. The top layers of the hierarchy capture broader, more abstract, features of the input data, which are often most relevant for the purpose of pattern recognition.

The nature of this deeply-layered inference architecture involves decomposing high-dimensional inputs into smaller patches, representing these patched in a compact manner and hierarchically learning the relationships between these representations across multiple scales. The underlying assumption is that input signal proximity is coherent with the nature of the data structure that is being represented. As an example, two pixels in an image, which are in close proximity, are assumed to exhibit stronger correlation than that exhibited by two pixels that are very distant. This assumption

holds firmly for many natural modalities, including natural images and videos, radar images and frequency components of an audio segment. In a face recognition application, for example, the output of the classifier may be a single value denoting whether or not the input pattern corresponds to a particular person.

DML combined with a classifier may be viewed as a general semi-supervised learning framework. Training the system can be generalized as follows. During the first step, a set of unlabeled samples (i.e. inputs/observations that do not have a known class label associated with them) are provided as input to the DML engine. The latter will learn from such samples about the general structure of the sensory input space it is presented with. During the second step, a set of labeled samples (i.e. inputs that have a distinct class labels associated with them) is provided as inputs. Signals are extracted to a classifier, which is then trained in a supervised manner on the labeled set. Testing is then achieved by presented unseen observations and evaluating the output of the classifier relative to the actual image class.

2. Application Feedback for Improved Perception

As described above, the semi-supervised framework that applies to most deep machine learning schemes implies a strict decoupling between the model (i.e. unsupervised training of the DML architecture) and the application (i.e. classifier). Learning to represent the input space based purely on regularities in the observations appears elegant. However, a more pragmatic approach is to guide the learning process of the DML engine so that it is of greater relevance to the classifier. For example, if the observations exhibit regularities which are not pertinent to the classifier to perform well, they may as well be ignored or discarded by the model. Thus, we propose a feedback mechanism between the classifier and the deep learning engine such that representation is optimized for the classification process.

The feedback mechanism proposed involves adapting the sample (i.e. observation) presentation distribution to the DML engine based on results obtained from the classification process. To do so, the DML and classifier trainings are performed concurrently, rather than in succession. This is somewhat of a paradigm shift from existing DML methodologies, but one that is argued vital. The classifier considered is a simple MLP feedforward neural network. As opposed to uniformly presenting samples from the unlabeled set to the DML engine, the sample presentation distribution is modified such that observations which need to be reinforced are presented more frequently. The need to reinforce presentations is derived from the classification error measured such that observations (i.e. input samples) that yield relatively high errors will be more frequently presented to the DML engine.

3. Simulation Results

The simulation results pertain to a simple image classification scenario. The goal is to provide an example highlighting the advantage of adjusting the sample distribution in an online manner. A database consisting of a train and test sets of images was created synthetically. Each of the two set contained 500 images, belonging to 9 classes (the letters 'C','G','H','M','P','R','T','X','Z'). These classes were each represented by a template image. Every image in both sets of the database was created from one of the template

images, with random distortion applied. The distortion included scaling, rotation, erosion and application of additive noise. Sample images from the test set are illustrated in Figure 2.

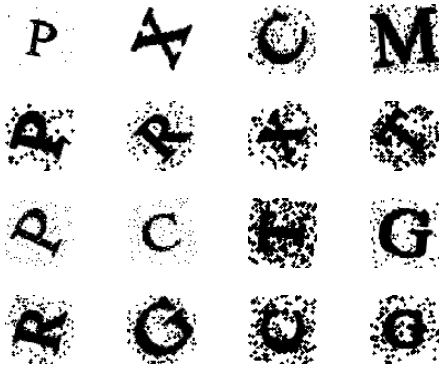

Figure 2. Examples of letter images distorted randomly and used in evaluating the proposed DML system.

The system was trained in a supervised manner on the training set, whereby the classifier was being targeted with a vector filled with '0's except for a single '1' in the location that corresponds to the image label. Testing was conducted on test set images, which are guaranteed to differ from the training set.

During the training phase, a model and a MLP neural network classifier were trained concurrently, in two modes. In the first mode, there was no classifier-model feedback involved, and the sample distribution remained uniform throughout the entire learning process. In the second mode, a feedback mechanism was applied in order to influence the sample presentation distribution, such that images with high classification error were presented more frequently, as described above.

In both modes, training was performed in batches. During each batch, 100 images were randomly preselected for presentation from the *N* images comprising the training set. In the first mode, these images were selected uniformly and independently at random. During the second mode, an adaptive presentation scheme was applied. At the end of each batch, after the DML parameters update, the classification error for each of the images in the training set has been evaluated. As a result, the sample presentation distribution was updated by applying a simple convex summation of the form

$$p_i^{t+1} = \alpha p_i^t + (1-\alpha)\frac{e_i^t}{\sum_{i=1}^{N} e_i^t},\ i = 1,..,N \quad t = 2,3,... \tag{1}$$

where p_i^t=1/*N* denotes the probability of selecting image *i* for presentation at batch *t*, e_i^t the classification error for the i^{th} image (calculated as the element mean on the absolute difference between the classifier output and the target vector) at batch *t*, and $0<\alpha<1$ is an coefficient set to 0.95.

Figure 3 illustrates the classification accuracy (on both the training and testing sets) as a function of the number of batches. Classification did not reach 100% accuracy intentionally, as the number of training samples was limited and the distortions vast. This was chosen in an aim to accentuate the difference of applying the

feedback mechanism on a simple task. As can be observed, on both training and testing sets, performance was consistently higher when the proposed feedback mechanism was applied, suggesting that guiding the representation of the DML engine by emphasizing underperforming samples leads to improved overall classification performance.

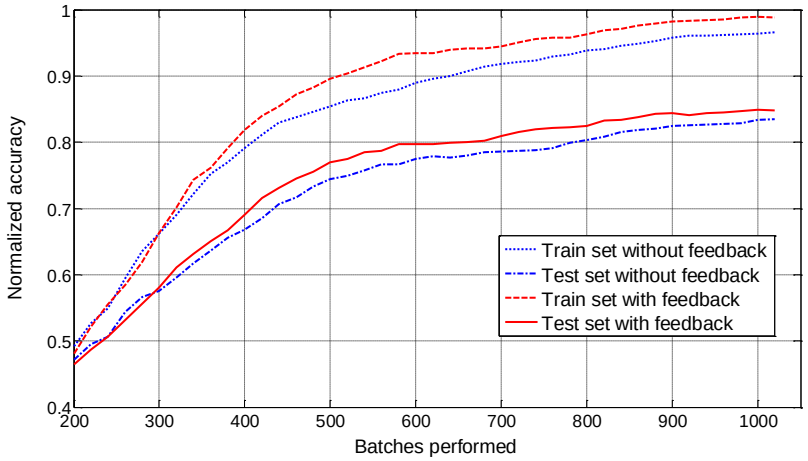

Figure 3. Classification accuracy as a function of the number of batches for both train and test set images.

4. Conclusions

In stark contrast to mainstream schemes, this paper presents a natural way in which data presentation in deep machine learning systems can be driven by the application, rather than purely by regularities in the observations. An online technique for adjusting the sample distribution based on classification error retains the hands-off attributes of DML, as it requires no human intervention. Simulation results clearly illustrate the benefits of a feedback mechanism from the application to the DML model. Moreover, the proposed approach can be broadly applied to other DML architectures and application domains.

References

[1] Y. Bengio, Learning Deep Architectures for AI, *Foundations and Trends in Machine Learning*, **2**,(2009), 1-127.

[2] R. Bellman, *Dynamic Programming*, Princeton University Press, 1957.

[3] T. Lee, D. Mumford, R. Romero, V. Lamme, The Role of the Primary Visual Cortex in Higher Level Vision, *Vision research*, **38** (1998), 2429-2454.

[4] G. Wallis, H. Bülthoff, Learning to recognize objects, *Trends in Cognitive Sciences*, **3** (1999), 23-31.

[5] I. Arel, S. Livingston, Beyond the Turing Test, *IEEE Computer*, **42** (2009), 104-105.

[6] M. Ranzato, F.J. Huang, Y. Boureau, Y. LeCun, Unsupervised Learning of Invariant Feature Hierarchies with Applications to Object Recognition, *Proc. Computer Vision and Pattern Recognition Conference*, 2007.

[7] G. E. Hinton, S. Osindero, Y. Teh. A Fast Learning Algorithm for Deep Belief Nets, *Neural Computation*, **18**, (2006), 1527-1554.