Deep Sparse-coded Network (DSN)

Youngjune Gwon, Miriam Cha, and H.T. Kung

HARVARD John A. Paulson School of Engineering and Applied Sciences

Motivation

Motivation

- Representational power of single-layer feature learning is limited to simple tasks
- Deep architectures allow us to decompose hierarchy of complex data (e.g., human face) into layers of features, with a feature at each layer using the features of the layer below
  - ➢ Best current practices use deep architectures based on autoencoder, restricted Boltzmann machine (RBM), and convolutional neural network (CNN)

Deep Architecture Based on Sparse Coding

- Single-layer sparse coding performance (according to Coates *et al*., 2011) is better than or on par with RBM and CNN
  - ➢ Sparse coding, due to its regularization on sparsity, gives probably the most effective (unsupervised) clustering method known to date
- Motivated by superior feature learning performance of single-layer sparse coding, we build a deep architecture based on sparse coding

Objective

Challenges

1) As going up layers, sparse coding increases dimensionality
2) Sparse coding makes an inherent assumption on the input being non-sparse
3) It is difficult to optimize all layers of sparse coding jointly

Objective

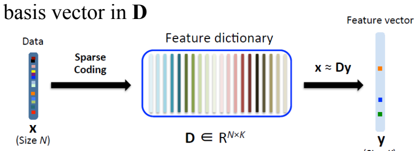
We propose Deep Sparse-coded Network (DSN), a deep architecture for sparse coding as a principled extension from its single-layer counterpart

- We propose a novel backpropagation algorithm that is specific to multi-layer sparse coding interlaced by spatial max pooling
- Using max pooling, we avoid linear cascade of dictionaries and keep the effect of multi-layering in tact
  - ➢ Remedy problem of too many feature vectors and preserve translational invariance
- We consider both l_1-regularized LASSO/ LARS and greedy-l_0 OMP for sparse coding methods

Approach

Sparse Coding

Represent **x** as a sparse linear combination of basis vector in **D**

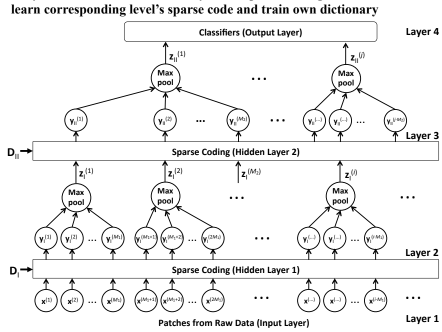


$$\hat{\mathbf{y}} = \arg\min_{\mathbf{y}} \underbrace{\|\mathbf{x} - \mathbf{D}\mathbf{y}\|_2^2}_{\text{Reconstructive error}} + \underbrace{\lambda \cdot \rho(\mathbf{y})}_{\text{Regularization}}$$

- Greedy algorithm such as OMP can minimize l_0 pseudo-norm, $\lambda\|\mathbf{y}\|_0$
- LASSO/LARS can be used to minimize l_1-norm, $\lambda\|\mathbf{y}\|_1$

Deep Sparse-coded Network (DSN)

4-layer DSN with two hidden layers of sparse coding, each of which can learn corresponding level's sparse code and train own dictionary



Training algorithms for DSN

1) Pretraining via layer-by-layer sparse coding and dictionary learning

- **Compute sparse codes while learning dictionary**

For hidden layer I:

$$\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(M_1)}\} \xrightarrow{\mathbf{D}_\mathrm{I}} \{\mathbf{y}_\mathrm{I}^{(1)}, \ldots, \mathbf{y}_\mathrm{I}^{(M_1)}\}$$

$$\{\mathbf{x}^{(M_1+1)}, \ldots, \mathbf{x}^{(2M_1)}\} \xrightarrow{\mathbf{D}_\mathrm{I}} \{\mathbf{y}_\mathrm{I}^{(M_1+1)}, \ldots, \mathbf{y}_\mathrm{I}^{(2M_1)}\}$$

- **Max pooling to aggregate sparse codes**

$$\{\mathbf{y}_\mathrm{I}^{(1)}, \ldots, \mathbf{y}_\mathrm{I}^{(M_1)}\} \xrightarrow{\text{max pool}} \mathbf{z}_\mathrm{I}^{(1)}$$

- **Pooled sparse code is passed to next layer**

$$\{\mathbf{z}_\mathrm{I}^{(1)}, \ldots, \mathbf{z}_\mathrm{I}^{(M_2)}\} \xrightarrow{\mathbf{D}_\mathrm{II}} \{\mathbf{y}_\mathrm{II}^{(1)}, \ldots, \mathbf{y}_\mathrm{II}^{(M_2)}\}$$

2) Training classifiers at the output layer

- **Train weights in classifier/regressor**

$$\hat{l} = h_{\mathbf{w}}(\mathbf{z}_\mathrm{II}) = \mathbf{w}^\top \cdot \mathbf{z}_\mathrm{II} + w_0$$

3) Backpropagation

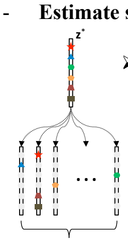
- **Compute classifier error**

$$\epsilon^{(i)} = l^{(i)} - h_{\mathbf{w}}(\mathbf{z}_\mathrm{J}^{(i)}) \quad \forall i$$

- **Compute updated pooled sparse codes**

$$z_{\mathrm{J},k}^{(i)} := z_{\mathrm{J},k}^{(i)} + \alpha \cdot \epsilon^{(i)} \cdot w_k \quad \forall i, k$$

- **Estimate sparse code via *putback***



Putback yields corrected sparse codes y*'s

- ➢ Updated pooled sparse codes are put back to locations at the original sparse codes
- **Down-propagate for all hidden layers**
- **Up-propagate using Steps 1&2**

Experiments

- Sample 20,000 images uniformly from CIFAR-10 dataset
  - ➢ 2,000 images per class
  - ➢ Five folds for cross validation
- Preprocess patches by ZCA-whitening before sparse coding

Average 1-vs-all classification accuracy for single-layer sparse coding and autoencoder

	Classification accuracy
Autoencoder	69.8%
OMP ($S = 0.1N$)	75.3%
OMP ($S = 0.2N$)	76.9%
LARS ($\lambda = 0.2$)	78.4%
LARS ($\lambda = 0.1$)	**80.1%**

**S*: number of non-zero items

In single-layer setting, sparse coding based on LARS (with λ=0.1) achieved the best single-layer accuracy at 80.1%

Average 1-vs-all classification accuracy comparison between DSN and deep stacked autoencoder (SAE)

	Classification accuracy
Deep SAE (pretraining only)	71.8%
Deep SAE (pretraining+backprop)	78.9%
DSN-OMP (pretraining only)	79.6%
DSN-OMP (pretraining+backprop)	84.3%
DSN-LARS (pretraining only)	83.1%
DSN-LARS (pretraining+backprop)	**87.5%**

-Multi-layering improves accuracy performance for sparse coding by 3%, and the proposed backpropagation additional 4%
-DSN-OMP with only pretraining is already 0.7% better than the backpropagation-finedtuned deep SAE

Conclusion and Future Work

Conclusion

- Introduce Deep Sparse-coded Network (DSN), a deep architecture for sparse coding
- Discuss benefit and training methods of DSN including a novel backpropagation algorithm that effectively traverses and optimizes multiple layers of sparse coding and max pooling
- Report good classification performance on medium-sized setup with CIFAR-10

Future Work

- Experiment with DSN in larger datasets (CIFAR-100, Caltech-101, and Caltech-256)
- Test in broader scope for text, sounds, and wireless signal classification tasks

Acknowledgement

This work is supported in part by gifts from the Intel Corporation and in part by the Naval Supply Systems Command award under the Naval Postgraduate School Agreements No. N00244-15-0050 and No. N00244-16- 1-0018.