Sparse-coded Net Model and Applications

Y. Gwon, M. Cha, W. Campbell, H.T. Kung, C. Dagli

IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2016)

September 16, 2016

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

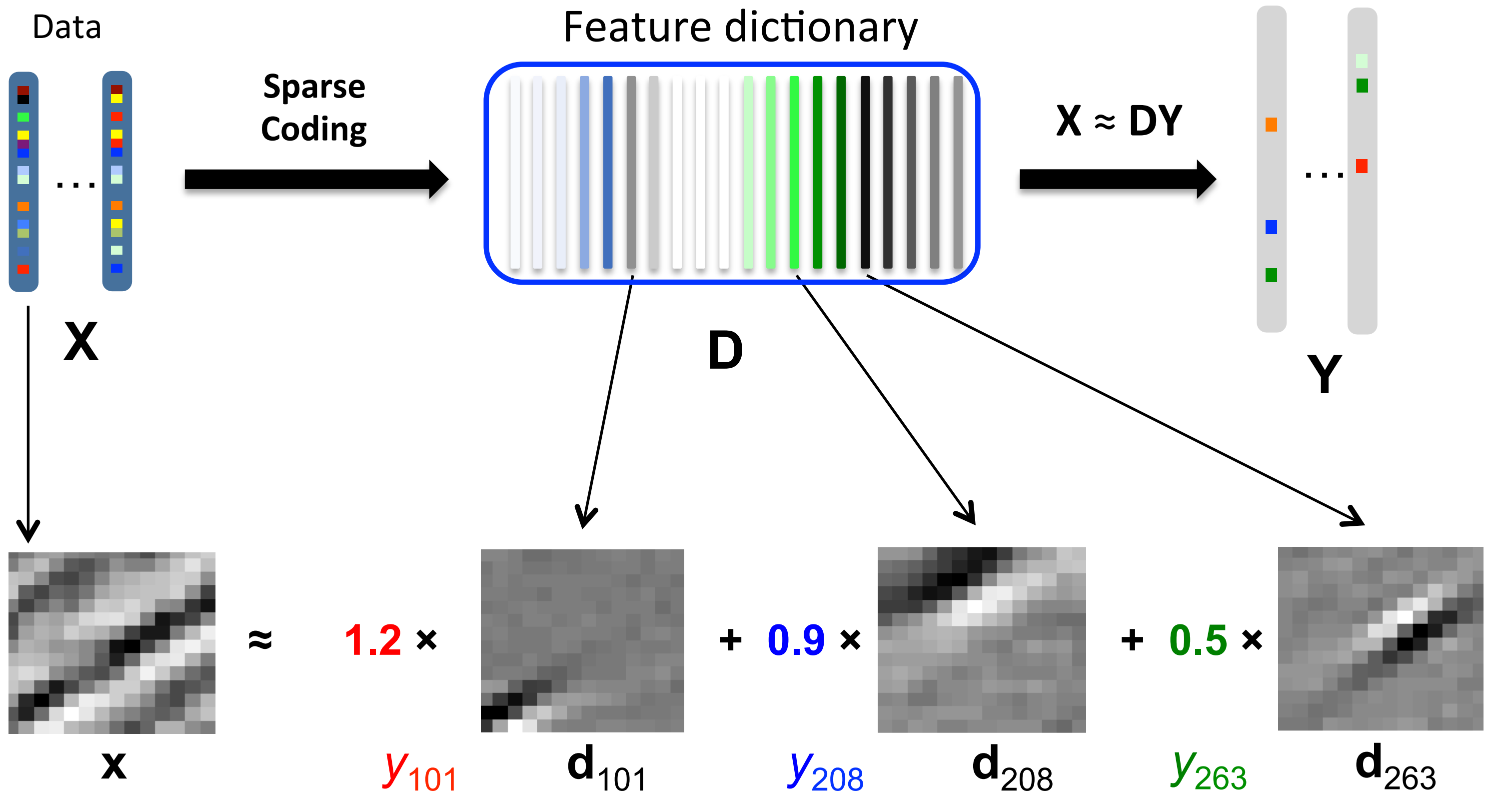
This work is sponsored by the Defense Advanced Research Projects Agency under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

Outline

- Background – Sparse Coding
- Semi-supervised Learning with Sparse Coding
- Sparse-coded Net
- Experimental Evaluation
- Conclusions and Future Work

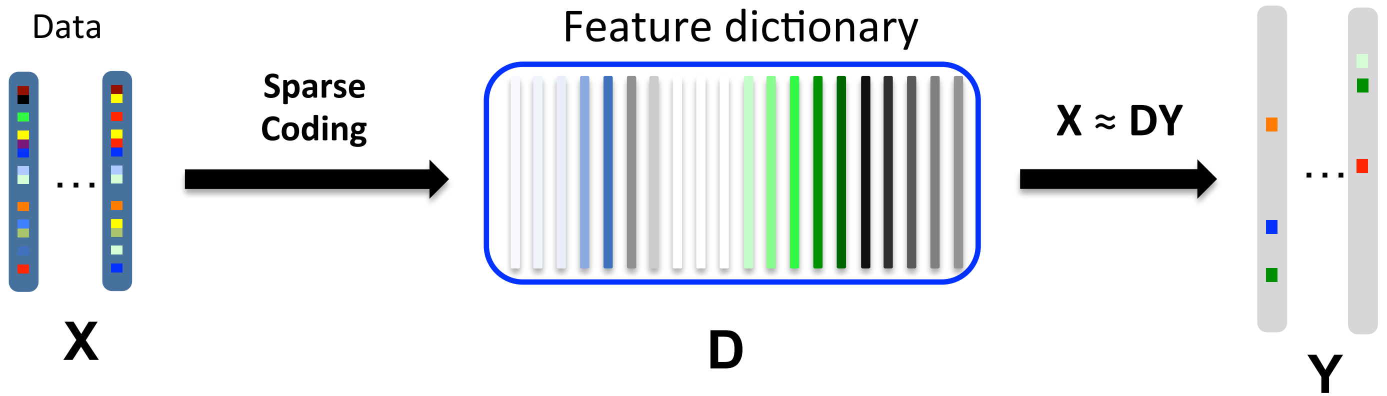
Background: Sparse Coding

- **Unsupervised method to learn representation of data**
  - **Decompose data into sparse linear combination of learned basis vectors**
  - **Domain transform: raw data ⟶ feature vectors**



Background: Sparse Coding (cont.)

- **Popularly solved as L_1-regularized optimization (LASSO/LARS)**
  - **Optimizing on L_0 pseudo-norm is intractable $\Longrightarrow$ greedy-L_0 algorithm (OMP) can be used instead**



$$\min_{\{\mathbf{D},\mathbf{y}\}} \|\mathbf{x} - \mathbf{D}\mathbf{y}\|_2^2 + \lambda\|\mathbf{y}\|_1$$

Convex relaxation

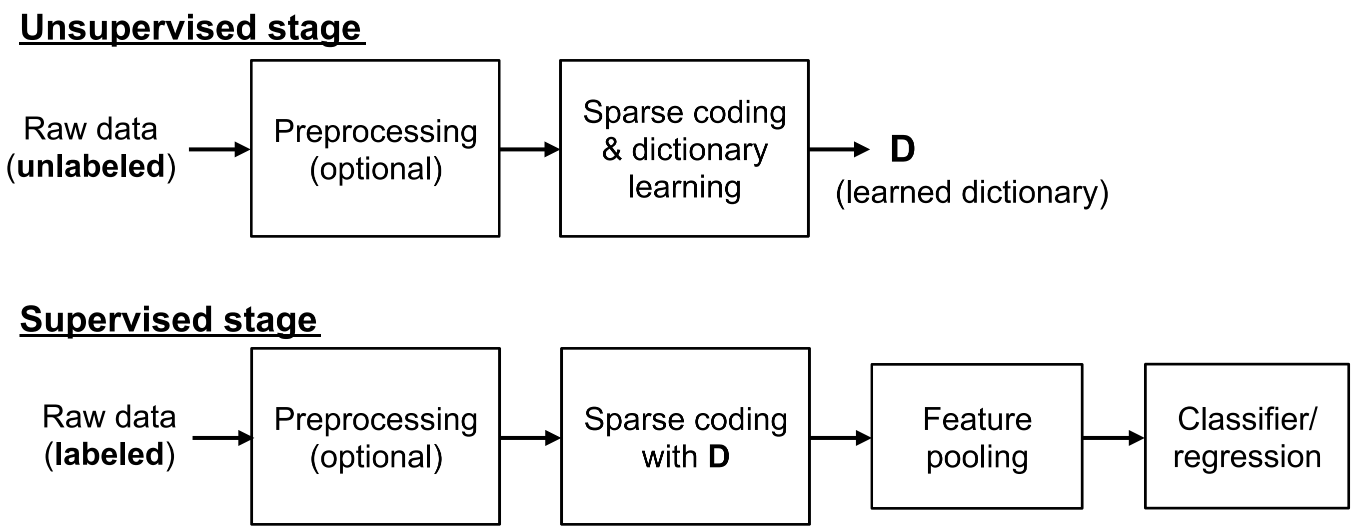
$$\min_{\{\mathbf{D},\mathbf{y}\}} \|\mathbf{x} - \mathbf{D}\mathbf{y}\|_2^2 + \lambda\|\mathbf{y}\|_0$$

Outline

- Background – Sparse Coding
- Semi-supervised Learning with Sparse Coding
- Sparse-coded Net
- Experimental Evaluation
- Conclusions and Future Work

Semi-supervised Learning with Sparse Coding

- **Semi-supervised learning**
  - Unsupervised stage: learn feature representation using unlabeled data
  - Supervised stage: optimize task objective using learned feature representations of labeled data
- **Semi-supervised learning with sparse coding**
  - Unsupervised stage: sparse coding and dictionary learning with unlabeled data
  - Supervised stage: train classifier/regression using sparse codes of labeled data



Outline

- Background – Sparse Coding
- Semi-supervised Learning with Sparse Coding
- Sparse-coded Net
- Experimental Evaluation
- Conclusions and Future Work

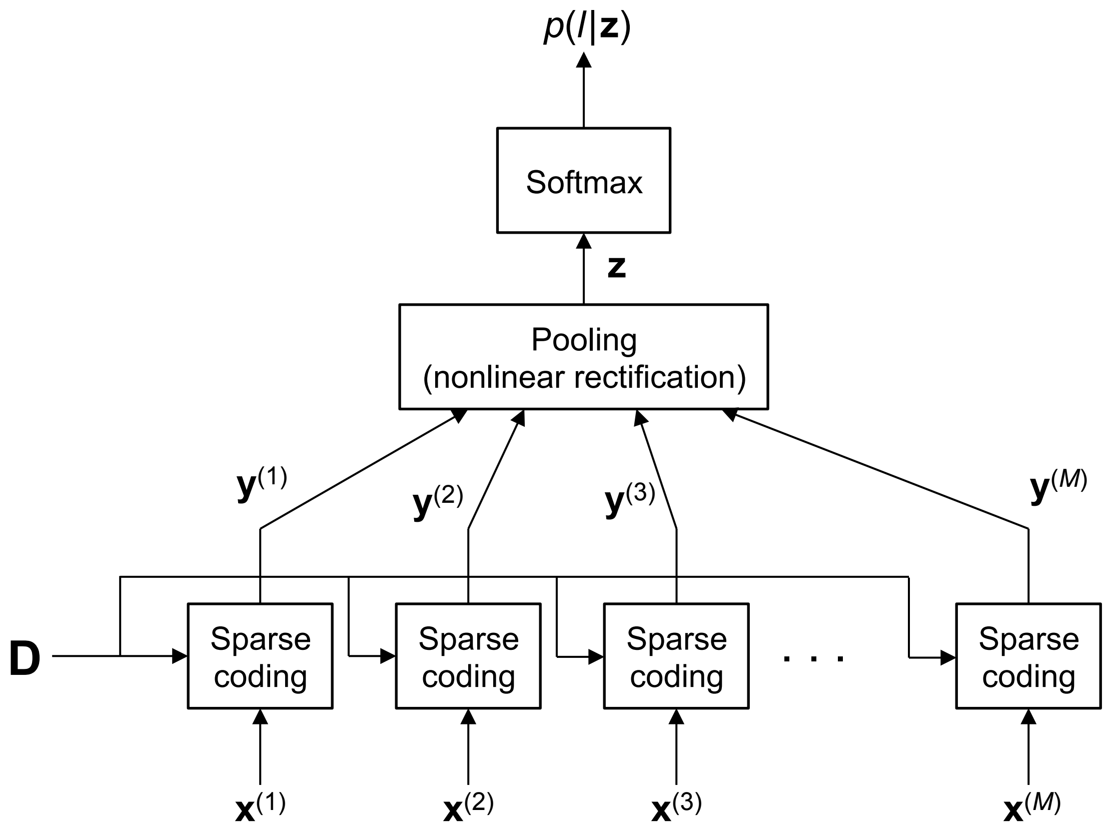
Sparse-coded Net Motivations

- Semi-supervised learning with sparse coding cannot jointly optimize feature representation learning and task objective

- Sparse codes used as feature vectors for task cannot be modified to induce correct data labels
  - No supervised dictionary learning $\Longrightarrow$ sparse coding dictionary is learned using only unlabeled data

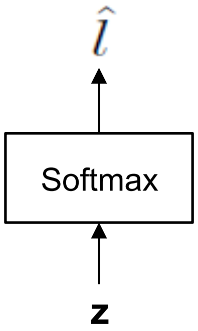
Sparse-coded Net

- **Feedforward model with sparse coding, pooling, softmax layers**
  - **Pretrain: semi-supervised learning with sparse coding**
  - **Finetune: SCN backpropagation**



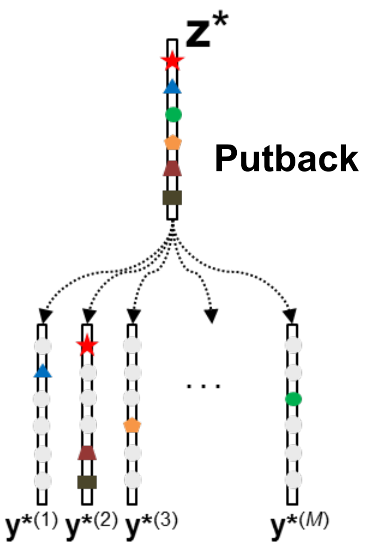
SCN Backpropagation

- **When predicted output does not match ground truth, hold softmax weights constant and adjust pooled sparse code by gradient descent**
  - **$\mathbf{z} \longrightarrow \mathbf{z}^*$**
- **Adjust sparse codes from adjusted pooled sparse code by putback**
  - **$\mathbf{z}^* \longrightarrow \mathbf{Y}^*$**
- **Adjust sparse coding dictionary by rank-1 updates or gradient descent**
  - **$\mathbf{D} \longrightarrow \mathbf{D}^*$**
- **Redo feedforward path with adjusted dictionary and retrain softmax**
- **Repeat until convergence**



Rewrite softmax loss as function of z

$$J(\mathbf{z}) = \frac{1}{2}\left\|\hat{l} - l\right\|^2$$



Outline

- Background – Sparse Coding
- Semi-supervised Learning with Sparse Coding
- Sparse-coded Net
- Experimental Evaluation
- Conclusions and Future Work

Experimental Evaluation

- Audio and Acoustic Signal Processing (AASP)
  - 30-second WAV files recorded in 44.1kHz 16-bit stereo
  - 10 classes such as bus, busy street, office, and open-air market
  - For each class, 10 labeled examples
- CIFAR-10
  - 60,000 32x32 color images
  - 10 classes such as airplane, automobile, cat, and dog
  - We sample 2,000 images to form train and test datasets
- Wikipedia
  - 2,866 documents
  - Annotated with 10 categorical labels
  - Text-document is represented as 128 LDA features

Results: AASP Sound Classification

Sound Classification Performance on AASP dataset

Method	Accuracy
Semi-supervised via sparse coding (LARS)	73.0%
Semi-supervised via sparse coding (OMP)	69.0%
GMM-SVM	61.0%
Deep SAE NN (4 layers)	71.0%
Sparse-coded net (LARS)	**78.0%**
Sparse-coded net (OMP)	**75.0%**

- Sparse-coded net model for LARS achieves the best accuracy performance of 78%
  - Comparable to the best AASP scheme (79%)
  - Significantly better than the AASP baseline† (57%)

†D.Stowell, D.Giannoulis, E.Benetos, M.Lagrange, and M.D.Plumbley, "Detection and Classification of Acoustic Scenes and Events," *IEEE Trans. on Multimedia*, vol. 17, no. 10, 2015.

Results: CIFAR Image Classification

Image Classification performance on CIFAR-10

Method	Accuracy
Semi-supervised via sparse coding (LARS)	84.0%
Semi-supervised via sparse coding (OMP)	81.3%
GMM-SVM	76.8%
Deep SAE NN (4 layers)	81.9%
Sparse-coded net (LARS)	**87.9%**
Sparse-coded net (OMP)	**85.5%**

- Again, sparse-coded net model for LARS achieves the best accuracy performance of 87.9%
  - Superior to RBM and CNN pipelines evaluated by *Coates et al.*†

†A. Coates, A. Ng, and H. Lee, "An Analysis of Single-layer Networks in Unsupervised Feature Learning," in AISTATS, 2011.

Results: Wikipedia Category Classification

Text Classification performance on Wikipedia dataset

Method	Accuracy
Semi-supervised via sparse coding (LARS)	69.4%
Semi-supervised via sparse coding (OMP)	61.1%
Deep SAE NN (4 layers)	67.1%
Sparse-coded net (LARS)	**70.2%**
Sparse-coded net (OMP)	**62.1%**

- We achieve the best accuracy of 70.2% with sparse-coded net on LARS
  - Superior to 60.5 – 68.2% by existing approaches[†1,†2]

†1 K. Duan, H. Zhang, and J. Wang, "Joint learning of cross-modal classifier and factor analysis for multimedia data classification," Neural Computing and Applications, vol. 27, no. 2, 2016.

†2 L. Zhang, Q. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Ensemble Manifold Regularized Sparse Low-rank Approximation for Multi- view Feature Embedding," *Pattern Recognition*, vol. 48, no. 10, 2015.

Outline

- Background – Sparse Coding
- Semi-supervised Learning with Sparse Coding
- Sparse-coded Net
- Experimental Evaluation
- Conclusions and Future Work

Conclusions and Future Work

Conclusions

- Introduced sparse-coded net model that jointly optimizes sparse coding and dictionary learning with supervised task at output layer
- Proposed SCN backpropagation algorithm that can handle mix-up of feature vectors related to pooling nonlinearity
- Demonstrated superior classification performance on sound (AASP), image (CIFAR-10), and text (Wikipedia) data

Future Work

- More realistic larger-scale experiments necessary
- Generalize hyperparameter optimization techniques for various datasets (e.g., audio, video, text)