CS 6501 Natural Language Processing

Text Generation

Yangfeng Ji

November 14, 2018

Department of Computer Science
University of Virginia

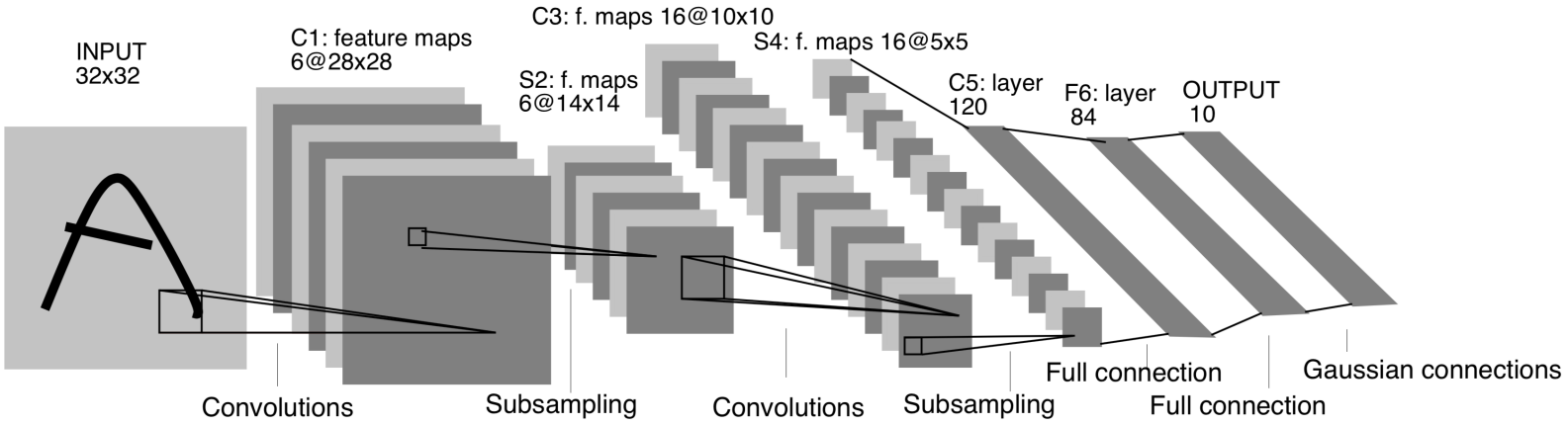
Overview

1. Basic Architecture of CNN

2. Applications in NLP

Basic Architecture of CNN

LeNet-5



[LeCun et al., 1998]

Two Basic Components

- Convolutional operations

$$s(t) = \int x(a)w(t-a)da \tag{1}$$

$$s(t) = \sum_{a=-\infty}^{\infty} x(a)w(t-a) \tag{2}$$

- Pooling

Convolutional Operations

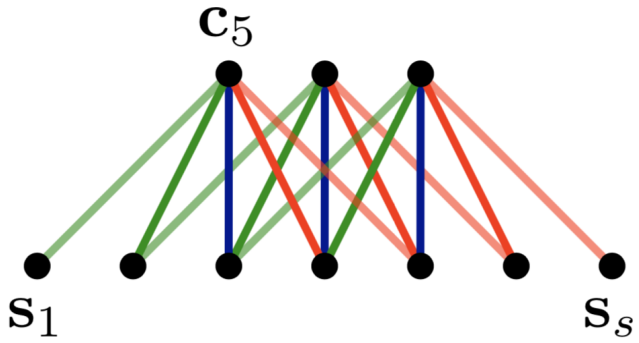
1-D convolutional operations

$$c_j = \boldsymbol{m}^\top \boldsymbol{s}_{j-m+1:j} \tag{3}$$

- $\boldsymbol{m} \in \mathbb{R}^n$: vector, filter
- $\boldsymbol{s} \in \mathbb{R}^T$: vector, input signal

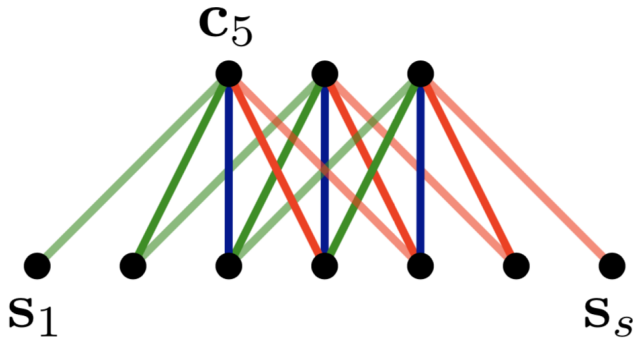
Convolutional Operations (II)

- Narrow type of convolution

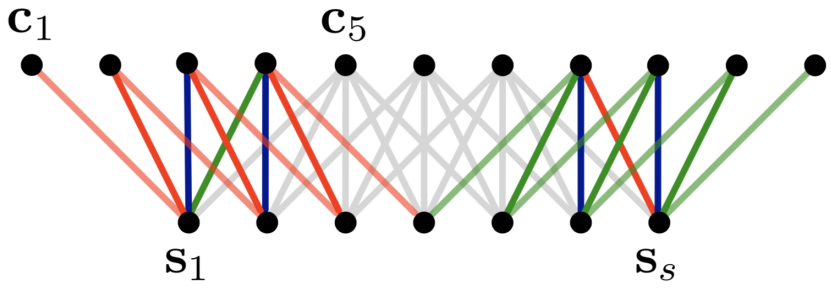


Convolutional Operations (II)

- Narrow type of convolution

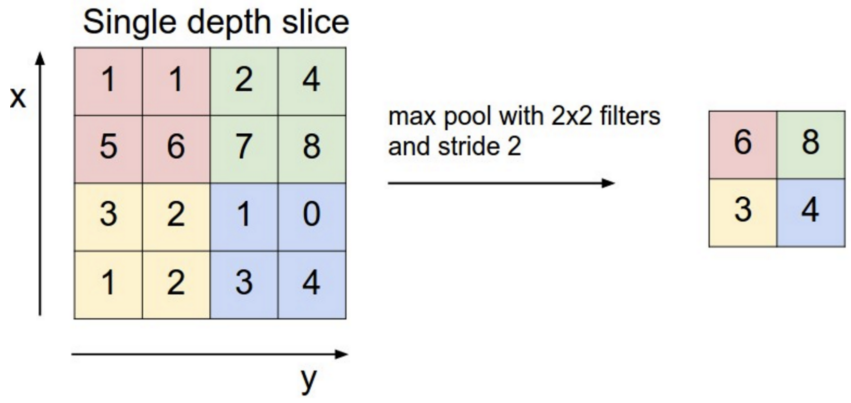


- Wide type of convolution



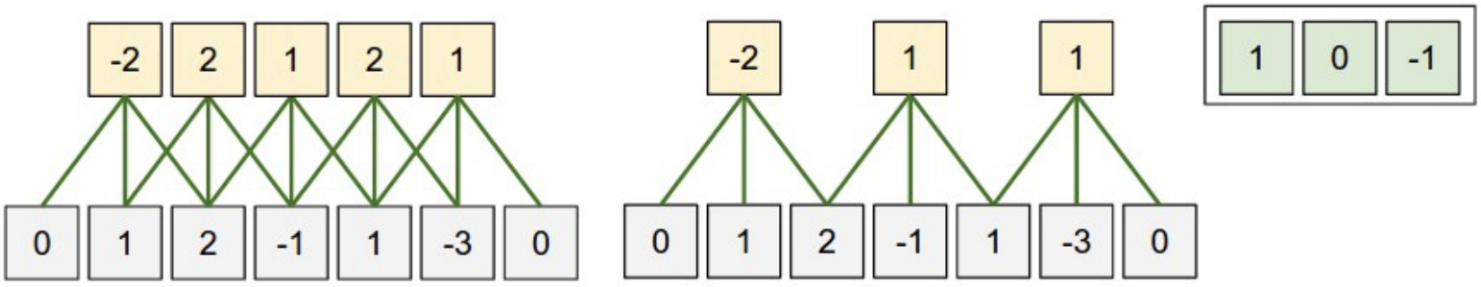
Pooling

- Max pooling



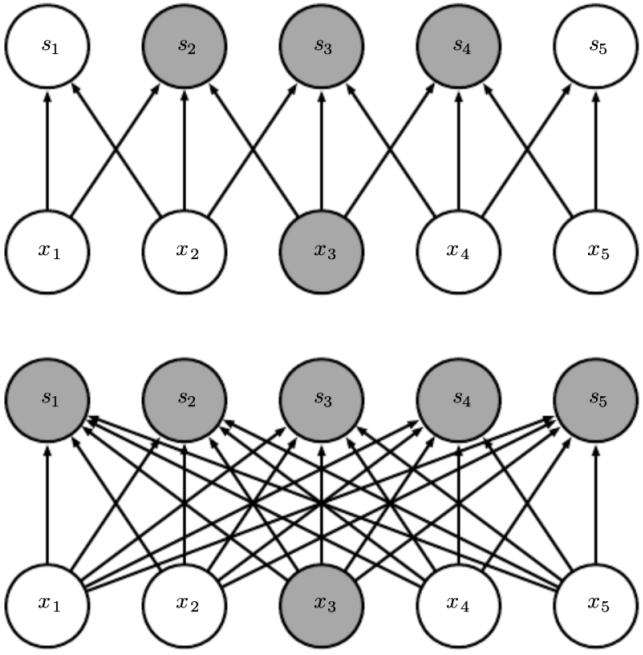
- Average pooling [LeCun et al., 1998]
- Min pooling

Stride Size



Advantages of CNNs

Comparing to Feed-forward NNs: Parameter sharing, sparse connections



[Goodfellow et al., 2016]

Applications in NLP

Example: Time-Delay Neural Networks

- Input

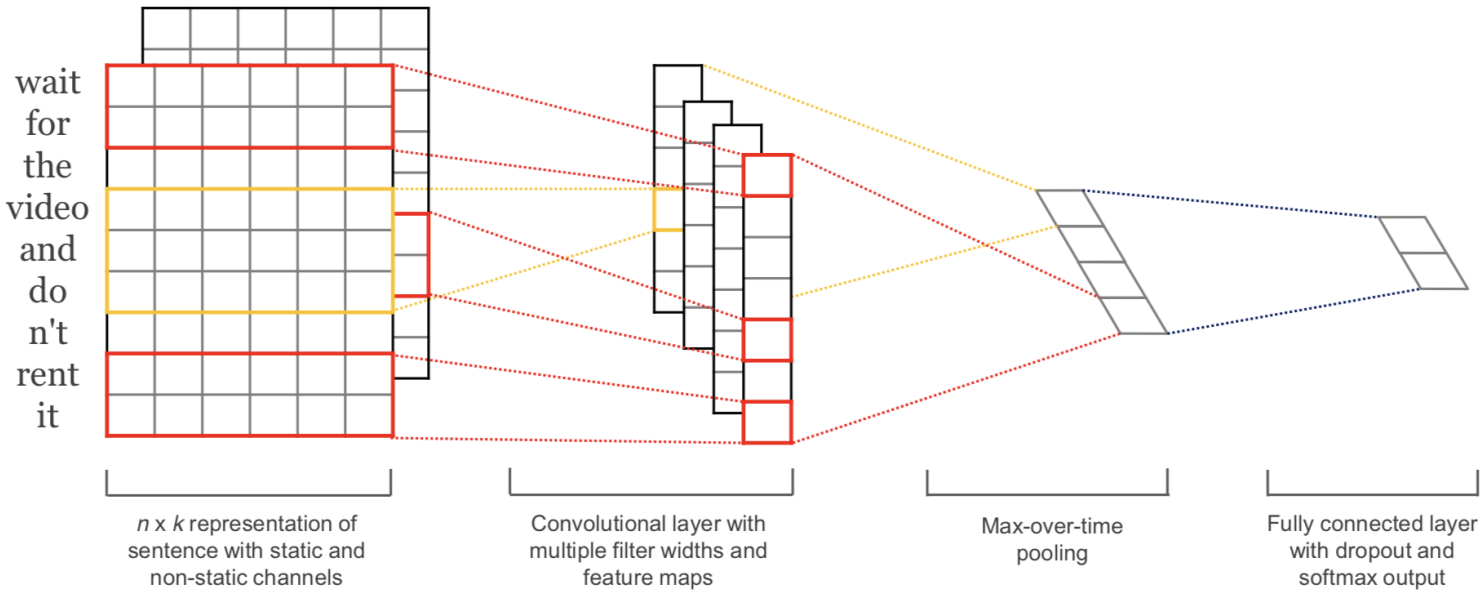
$$\mathbf{s} = \begin{bmatrix} | & | & | \\ \mathbf{w}_1 & \dots & \mathbf{w}_s \\ | & | & | \end{bmatrix}$$

- 1-D convolutional operations
- Max-pooling

$$\mathbf{c}_{max} = \begin{bmatrix} \max(\mathbf{c}_{1,:}) \\ \vdots \\ \max(\mathbf{c}_{d,:}) \end{bmatrix}$$

[Kalchbrenner et al., 2014]

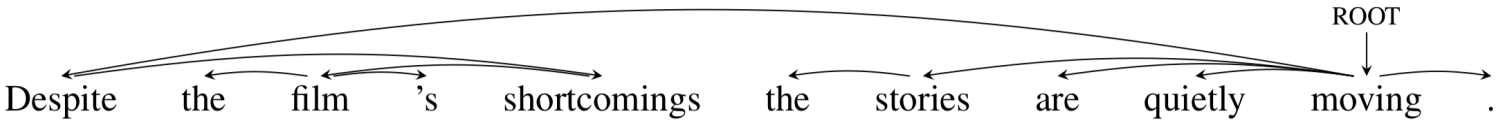
Example: 2-D Convolutional Operations



$$c_i = f(\boldsymbol{w}\boldsymbol{x}_{i:i+h-1} + b) \qquad (4)$$

[Kim, 2014]

Dependency-based CNNs

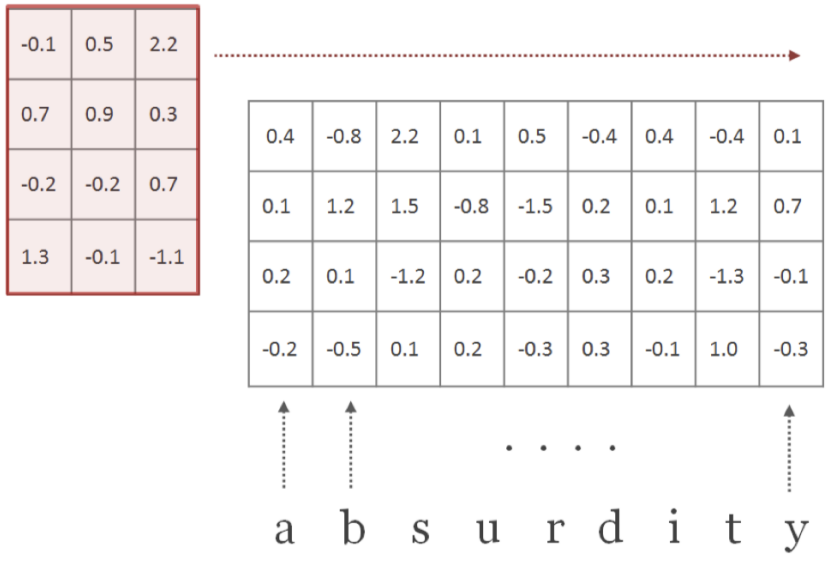


$$\boldsymbol{x}_i^{(k)} = [\boldsymbol{x}, \boldsymbol{x}_{p(i)}, \ldots, \boldsymbol{x}_{p^{k-1}(i)} \quad (5)$$

$$c = f(\boldsymbol{w}\boldsymbol{x}_{i,k} + b) \quad (6)$$

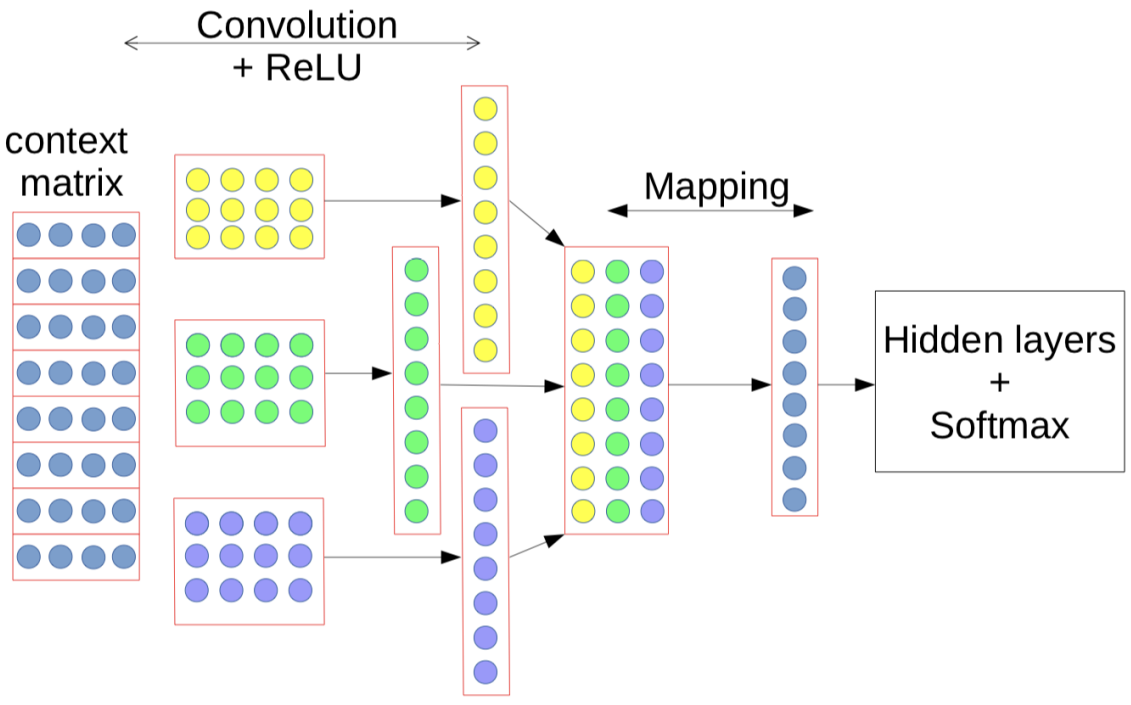
[Ma et al., 2015]

CharCNN



[Kim et al., 2016]

CNN for Language Modeling



[Gehring et al., 2016]

Summary

1. Basic Architecture of CNN

2. Applications in NLP

Reference

Gehring, J., Auli, M., Grangier, D., and Dauphin, Y. N. (2016).
A convolutional encoder model for neural machine translation.
arXiv preprint arXiv:1611.02344.

Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016).
Deep Learning, volume 1.
MIT press Cambridge.

Kalchbrenner, N., Grefenstette, E., and Blunsom, P. (2014).
A convolutional neural network for modelling sentences.
arXiv preprint arXiv:1404.2188.

Kim, Y. (2014).
Convolutional neural networks for sentence classification.
arXiv preprint arXiv:1408.5882.

Kim, Y., Jernite, Y., Sontag, D., and Rush, A. M. (2016).
Character-aware neural language models.
In *AAAI*.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998).
Gradient-based learning applied to document recognition.
Proceedings of the IEEE, 86(11):2278–2324.

Ma, M., Huang, L., Xiang, B., and Zhou, B. (2015).
Dependency-based convolutional neural networks for sentence embedding.
arXiv preprint arXiv:1507.01839.