

Deep Learning for Natural Language Processing

Dr. Minlie Huang (黃民烈)

aihuang@tsinghua.edu.cn

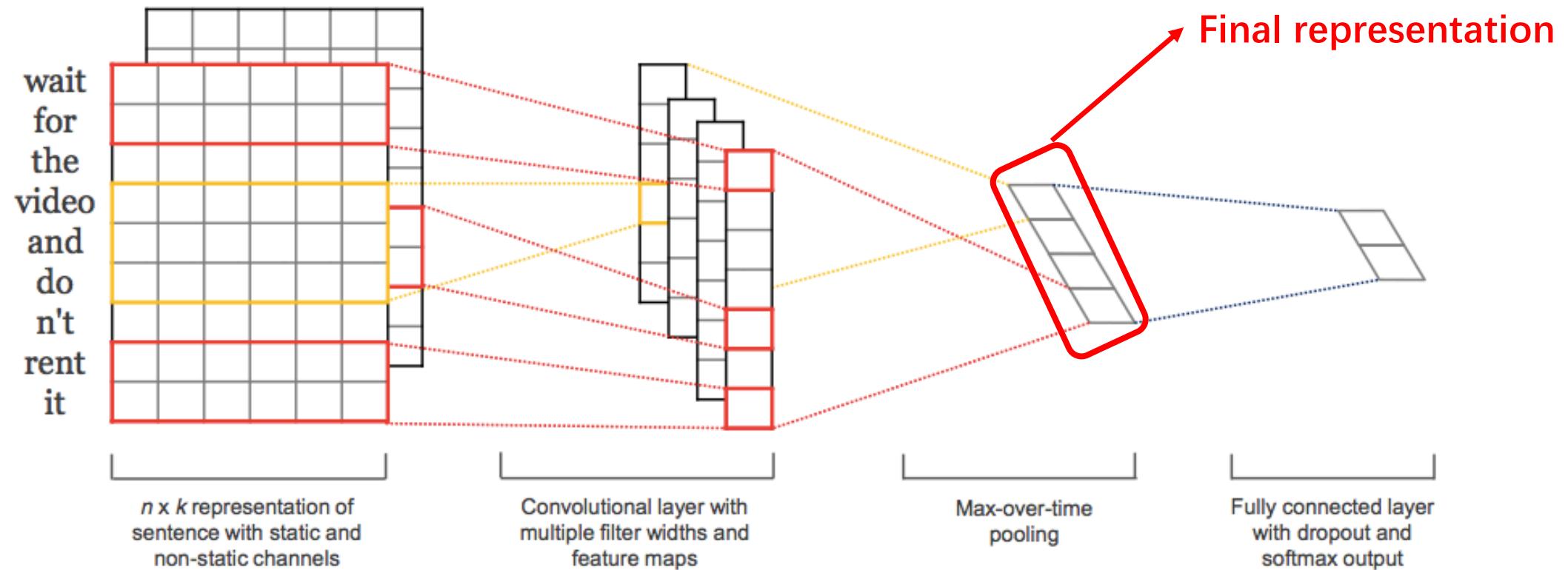
Computer Science Department

Tsinghua University

Homepage: <http://coai.cs.Tsinghua.edu.cn/hml>

Convolutional Network for NLP

Convolutional Network for NLP



Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In *EMNLP*. Association for Computational Linguistics, 1746–1751.

Convolution Layer

- Word vectors:

$$\mathbf{x}_i \in \mathbb{R}^k$$

- Sentence (concatenating words sequentially):

$$\mathbf{x}_{1:n} = \mathbf{x}_1 \oplus \mathbf{x}_2 \oplus \dots \oplus \mathbf{x}_n$$

- Words in range (a vector):

$$\mathbf{x}_{i:i+j}$$

- Convolutional filter (a vector), h is a window of h words:

$$\mathbf{w} \in \mathbb{R}^{hk}$$

Convolution Layer

- The feature value of position i:

$$c_i = f(\mathbf{w}^T \mathbf{x}_{i:i+h-1} + b)$$

- All possible windows:

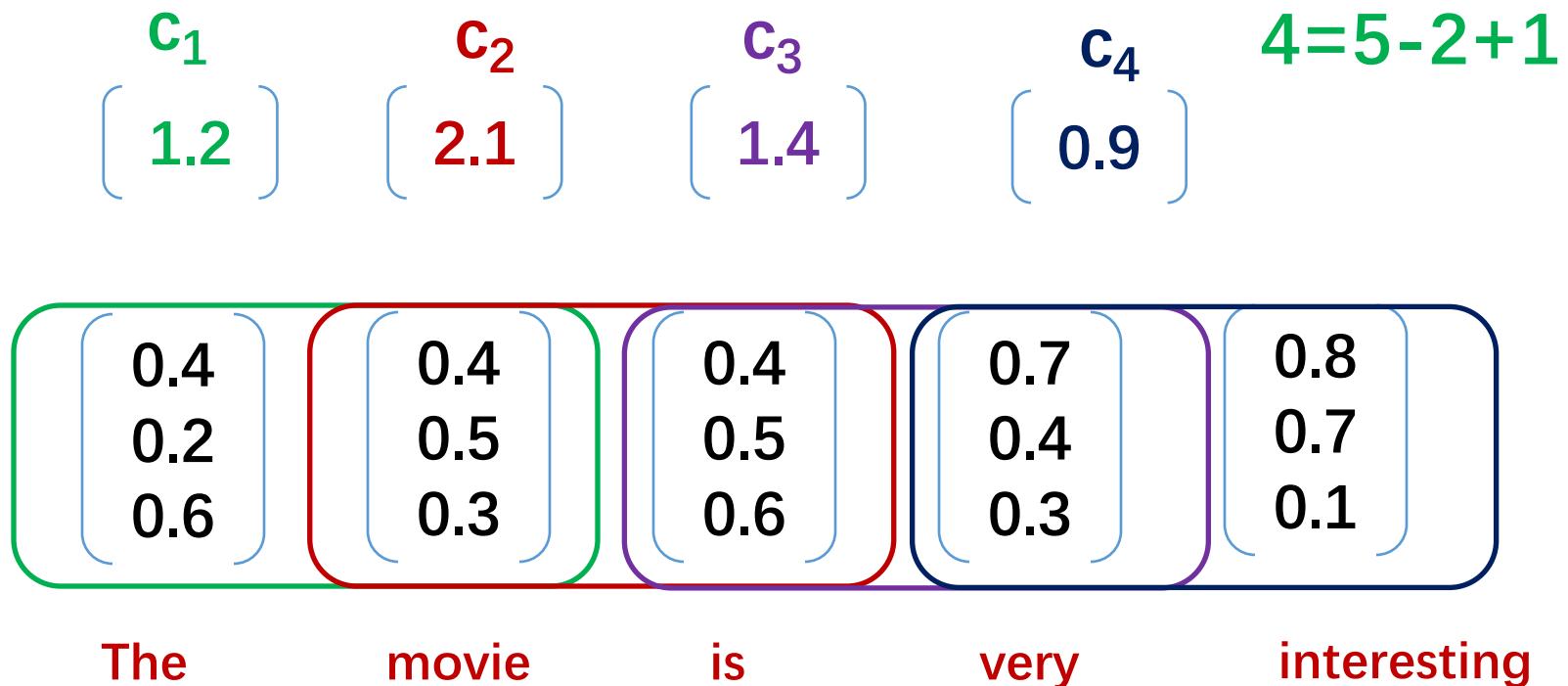
$$\{\mathbf{x}_{1:h}, \mathbf{x}_{2:h+1}, \dots, \mathbf{x}_{n-h+1:n}\}$$

- The resulting feature map:

$$\mathbf{c} = [c_1, c_2, \dots, c_{n-h+1}] \in \mathbb{R}^{n-h+1}$$

Convolution Layer

When the window width $h=2$



Pooling Layer

- Feature map with \mathbf{W} :

$$\mathbf{c} = [c_1, c_2, \dots, c_{n-h+1}] \in \mathbb{R}^{n-h+1}$$

- Pooling this vector to a single scalar (max, min, mean):

$$\hat{c} = \max\{\mathbf{c}\}$$

- But only **ONE** value here? Machine learning generally requires more input values

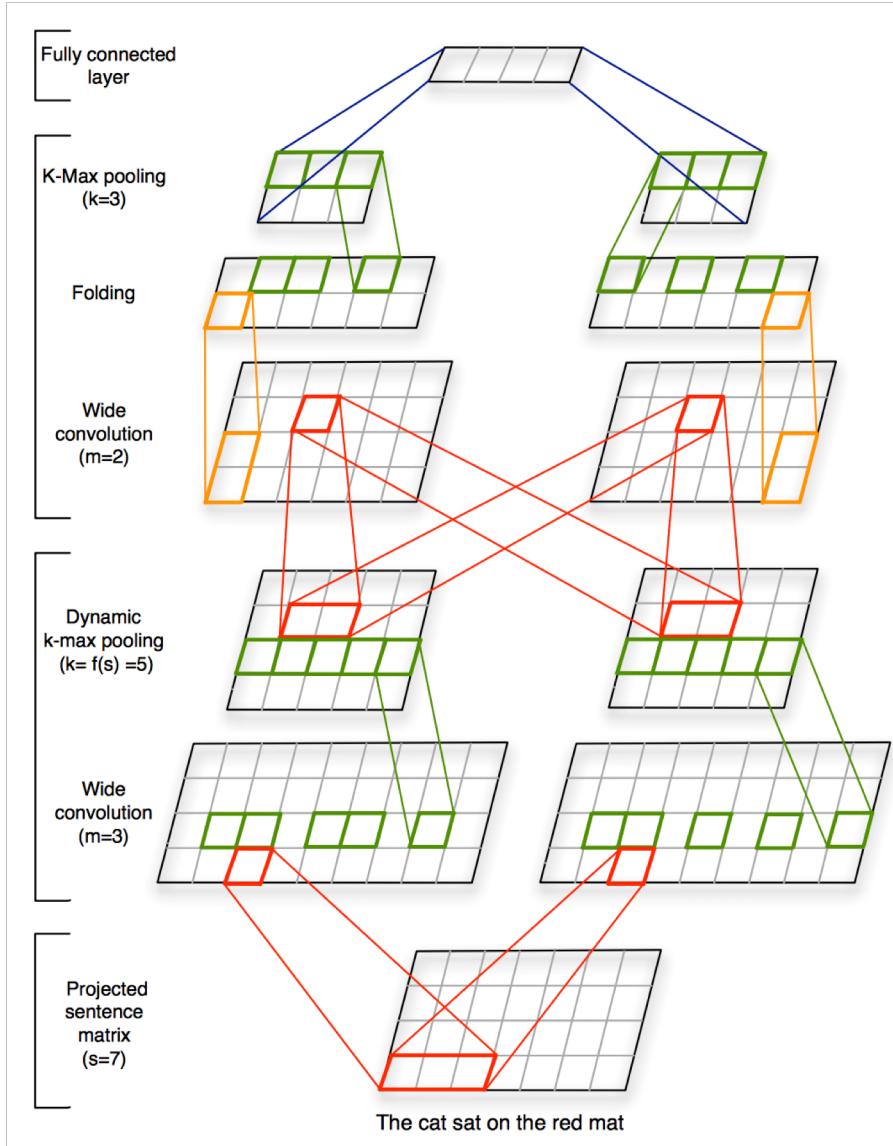
Final Softmax Layer

- Using **multiple filters W** (~ 100), and **multiple window h**
 - The final vector is of $N_w \times N_h$
- The final feature vector
$$Z = [c_{h_1,1}, c_{h_1,2}, \dots, c_{h_1,N_w}, c_{h_2,1}, c_{h_2,2}, \dots, c_{h_2,N_w} \dots]$$

h_i : how many types of windows
 N_w : how many filters
- The final softmax layer (or another MLP layer) takes z as input:

$$y = \text{softmax} \left(W^{(S)} z + b \right)$$

Advanced CNN Models for NLP

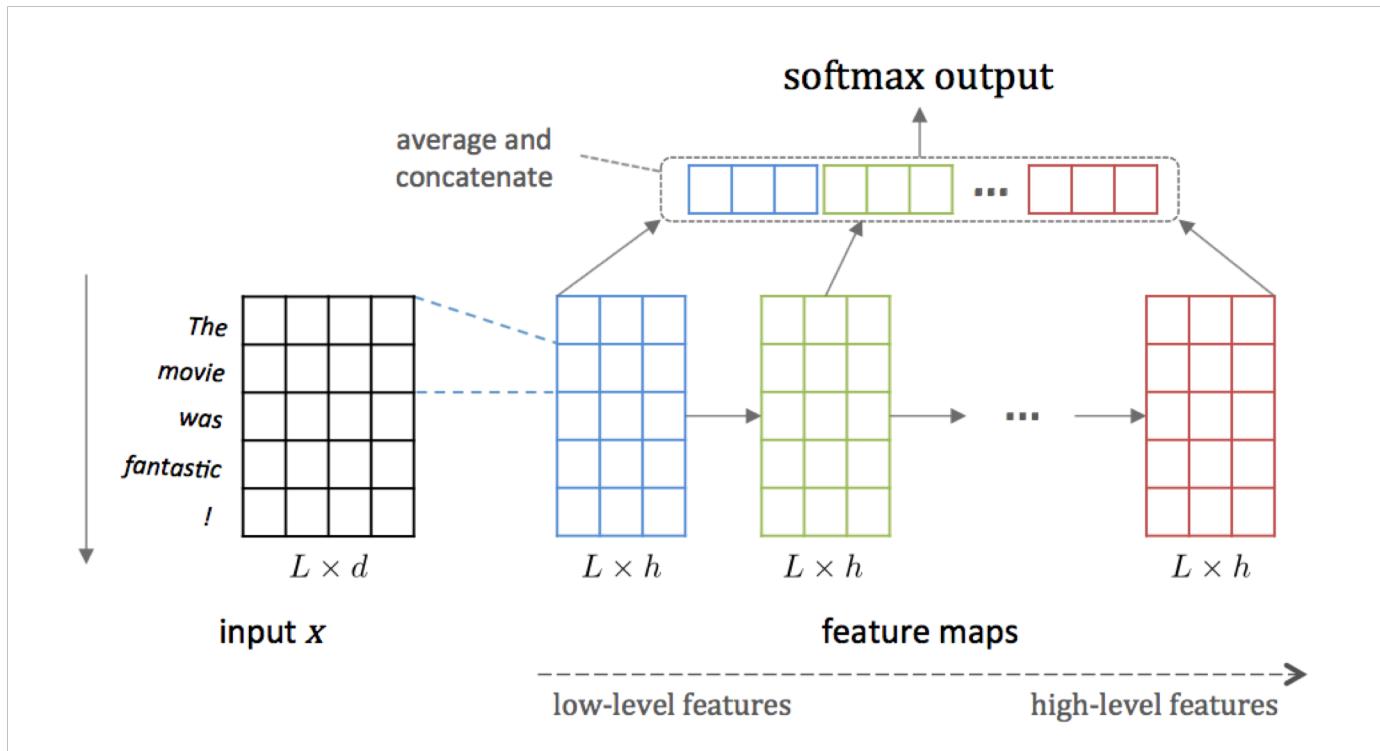


- One-dimensional convolution
- K-max pooling over time
- Multiple filters and multiple channels
- Very similar to image processing

N. Kalchbrenner, E. Grefenstette, and P. Blunsom. A Convolutional Neural Network for Modelling Sentences . In: Proceedings of ACL. 2014.

Advanced CNN Models for NLP

- Convolutional Neural Network



Lei T, Barzilay R, Jaakkola T. Molding cnns for text: non-linear, non-consecutive convolutions[J]. EMNLP, 2015.

| Model | Fine-grained | | Binary | |
|-----------------|--------------|-------------|--------|-------------|
| | Dev | Test | Dev | Test |
| RNN | | 43.2 | | 82.4 |
| RNTN | | 45.7 | | 85.4 |
| DRNN | | 49.8 | | 86.8 |
| RLSTM | | 51.0 | | 88.0 |
| DCNN | | 48.5 | | 86.9 |
| CNN-MC | | 47.4 | | 88.1 |
| CNN | 48.8 | 47.2 | 85.7 | 86.2 |
| PVEC | | 48.7 | | 87.8 |
| DAN | | 48.2 | | 86.8 |
| SVM | 40.1 | 38.3 | 78.6 | 81.3 |
| NBoW | 45.1 | 44.5 | 80.7 | 82.0 |
| Ours | 49.5 | 50.6 | 87.0 | 87.0 |
| + phrase labels | 53.4 | 51.2 | 88.9 | 88.6 |

Comments on CNN for NLP

- Suitable for **longer** text
- **Deep structures** (multiple con/pool operations) seem to be NOT that helpful
- **Not as strong as** other models such as recursive/recurrent models

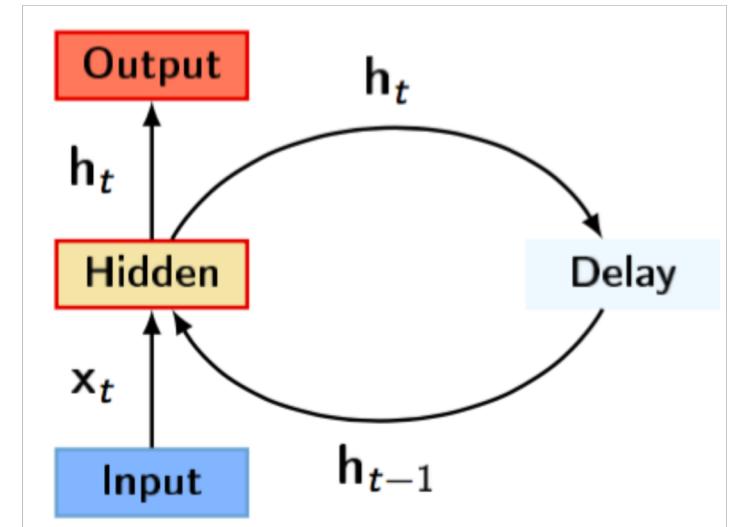
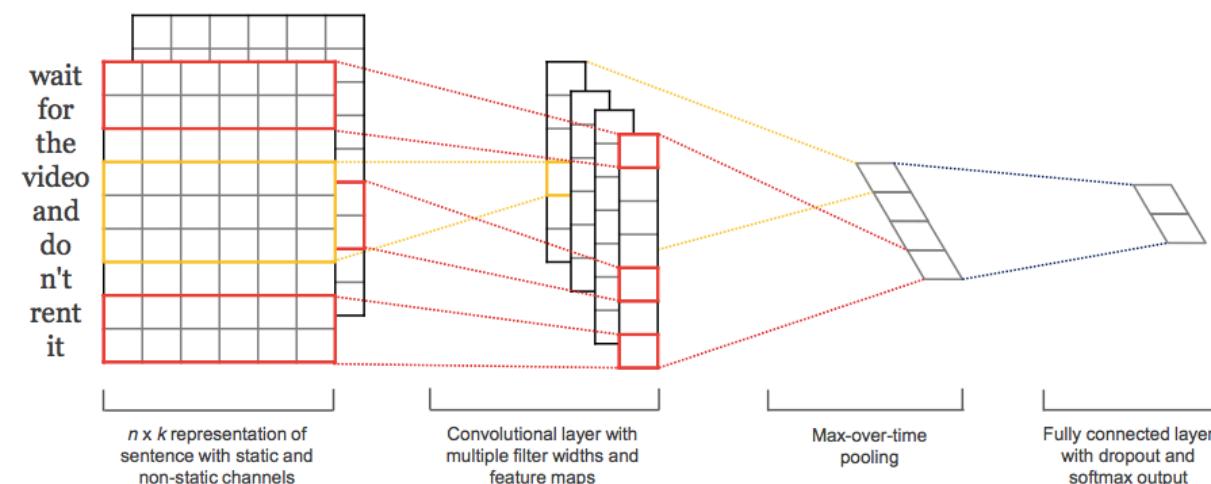
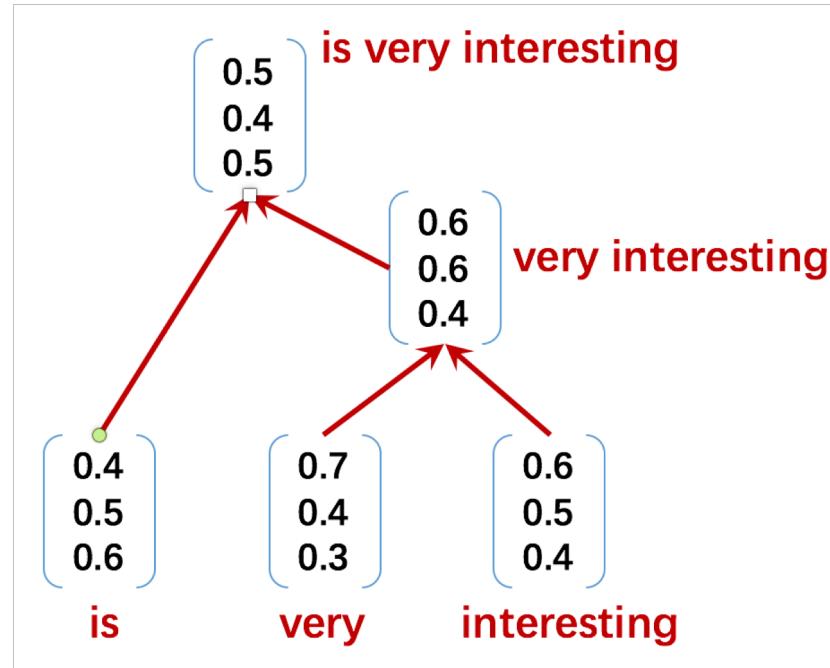
Summary

- Word representation
 - Word vectors
- Phrase/sentence representation
 - Recursive models
 - Recurrent models
 - CNN
- Document representation
 - CNN
 - Hierarchical models (RNN)
 - Hybrid models

Summary

0.123
0.243
0.789
-0.150
0.893
0.163
-0.876

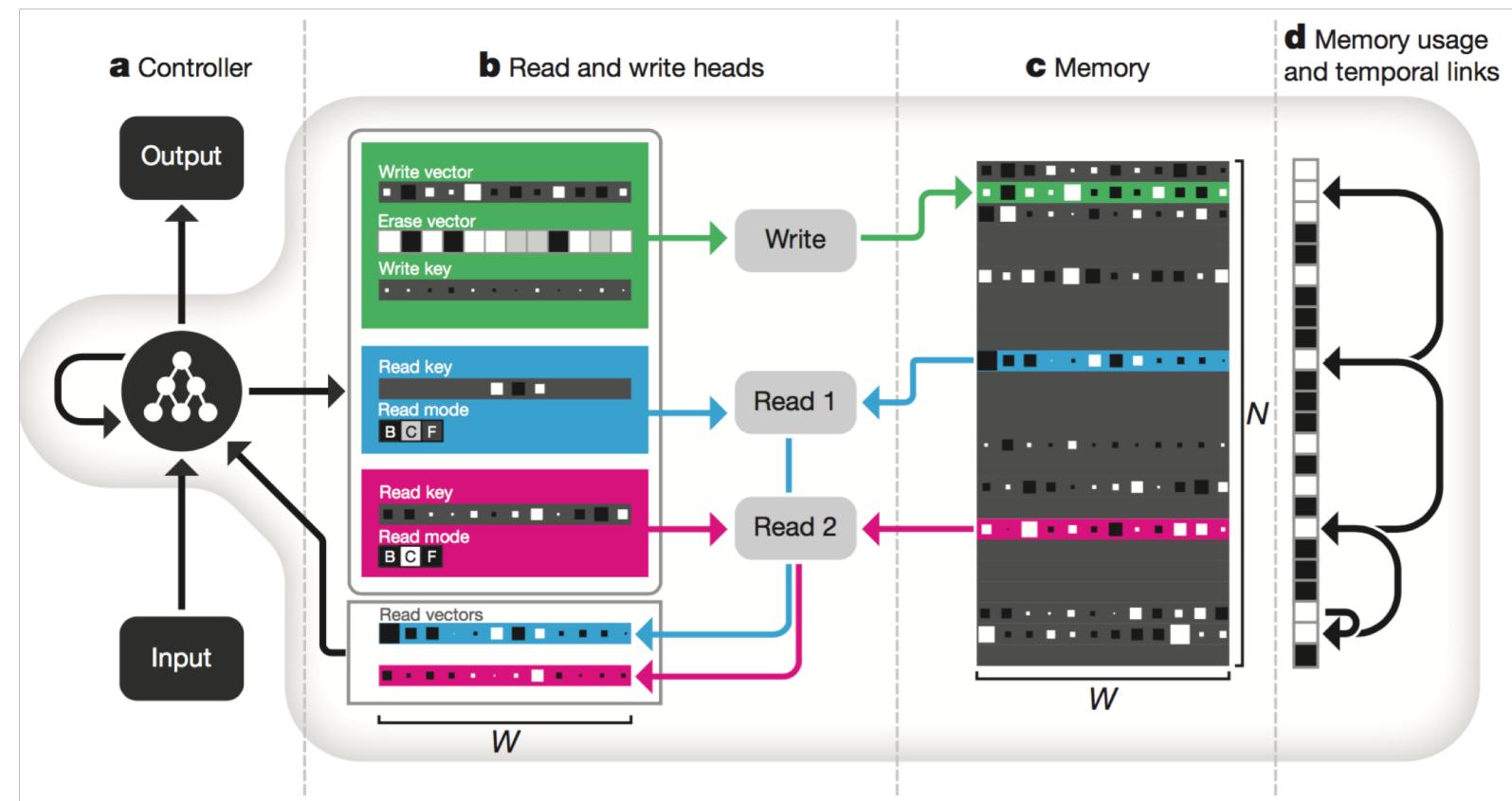
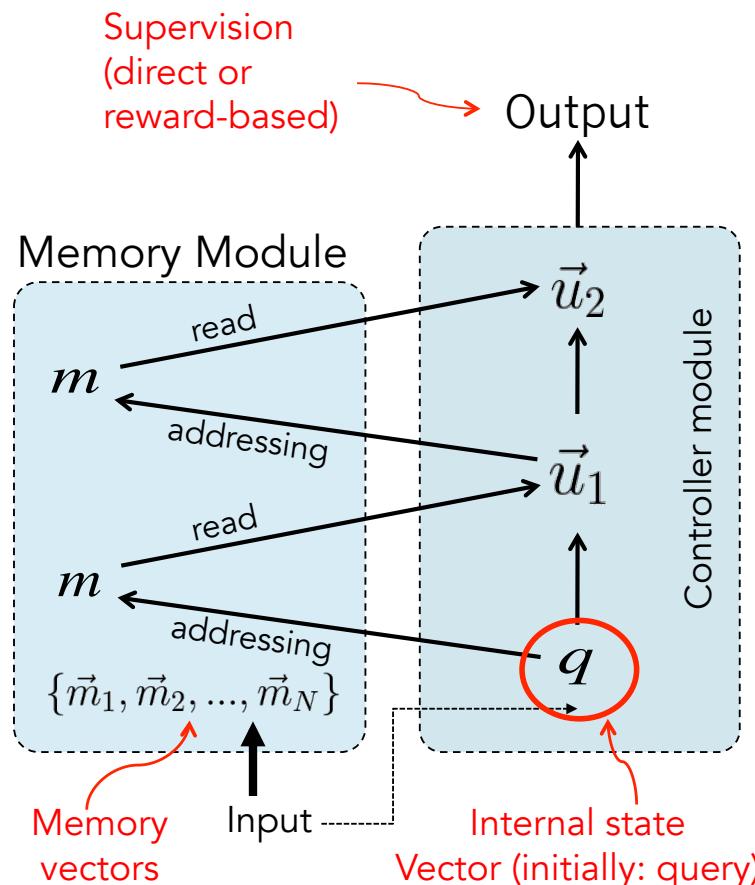
Deep



Advance Topics of Deep Learning for NLP

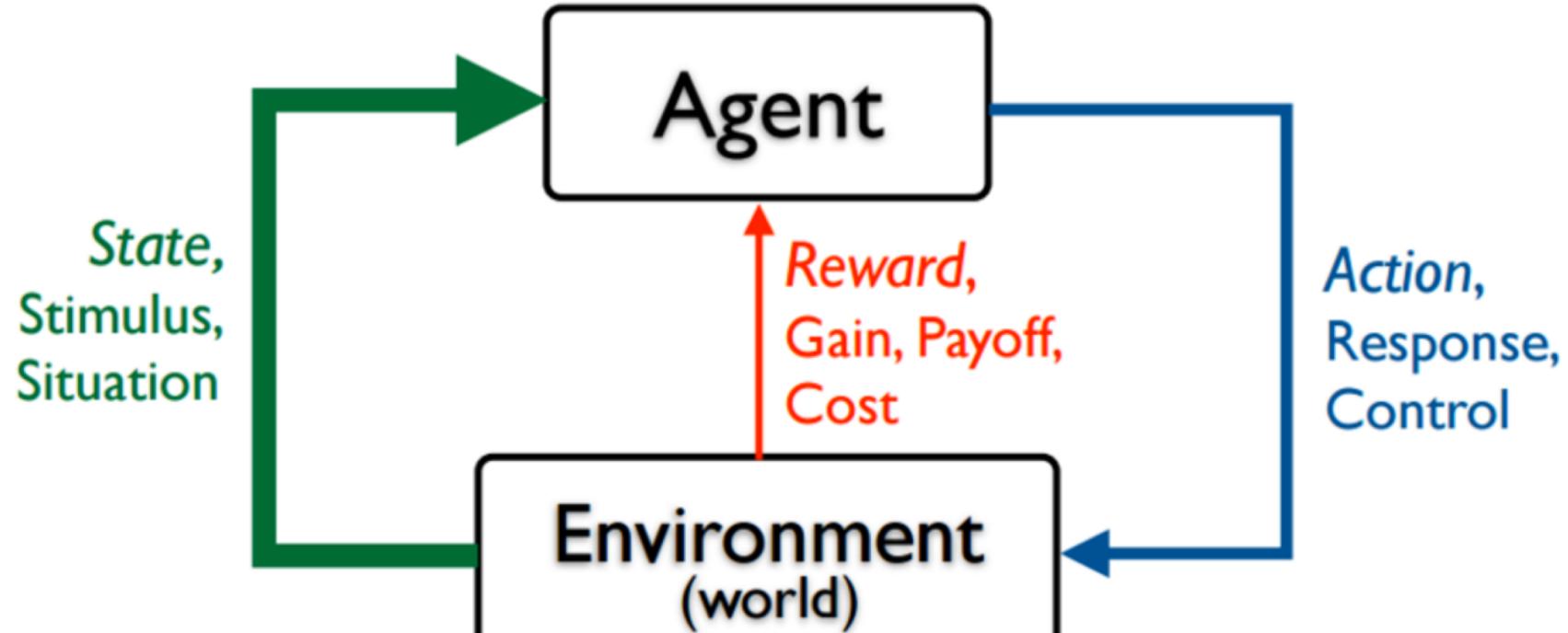
- Memory network
- Deep reinforcement learning
- Sequence2Sequence learning

Memory Network Models



Neural Differentiable Computer

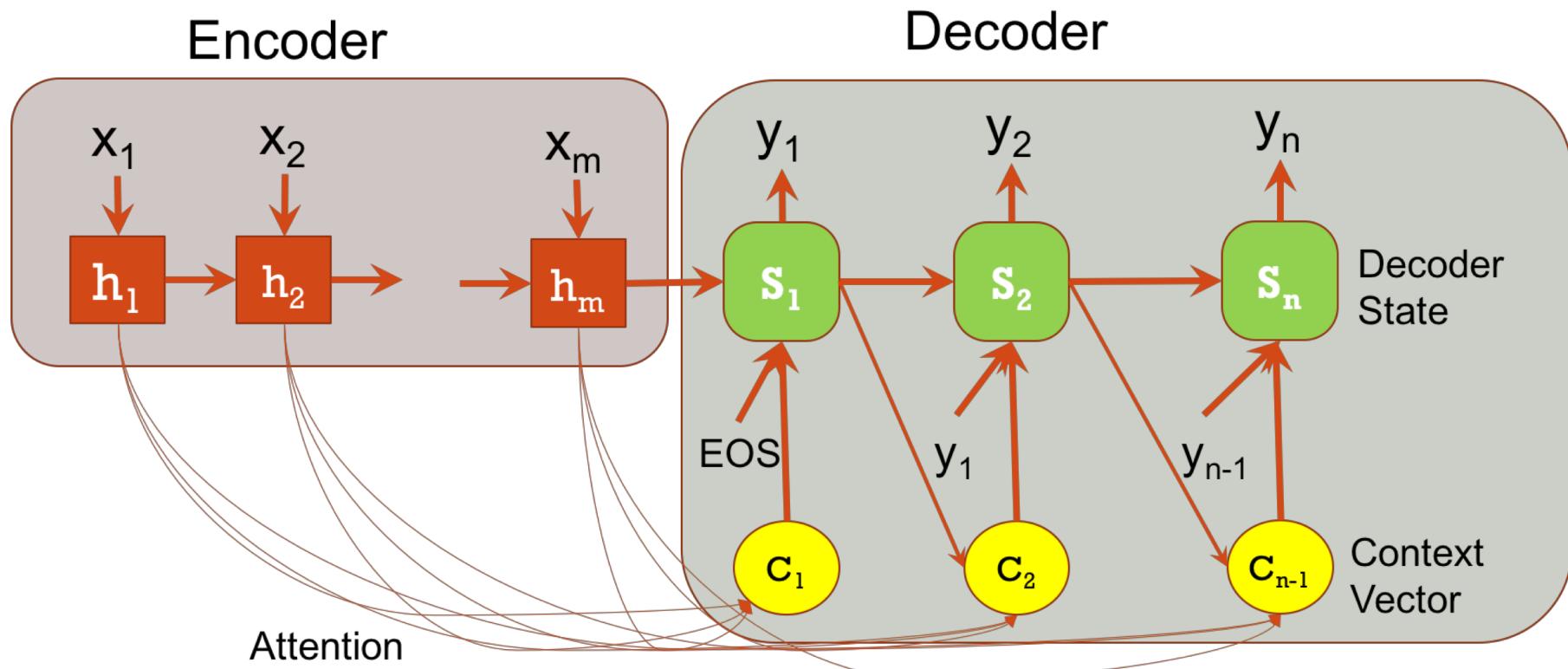
Deep Reinforcement Learning



Deep learning to represent **states**, **actions**, or **policy functions**

Sequence to Sequence Learning

- From Sequence $X_1 X_2 \cdots X_m$ to $Y_1 Y_2 \cdots Y_n$
- Machine translation, dialogue generation



Thanks for Attention

- Dr. Minlie Huang
- aihuang@tsinghua.edu.cn
- Homepage: <http://coai.cs.tsinghua.edu.cn/hml>