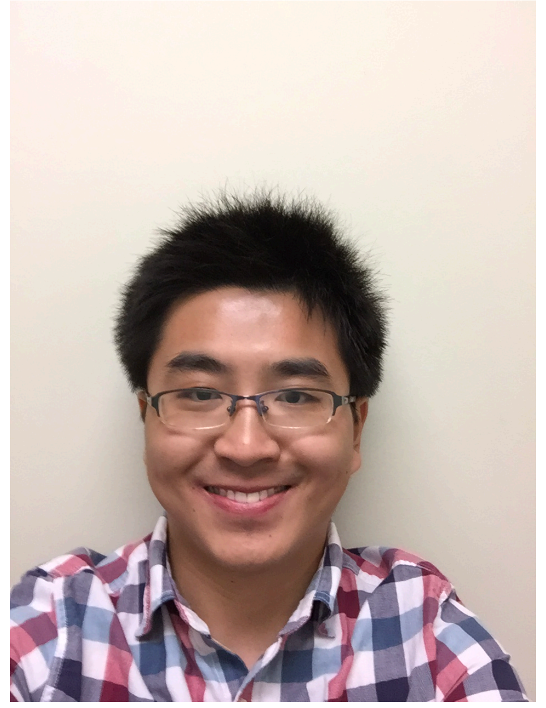
Convolutional Neural Networks with Recurrent Neural Filters

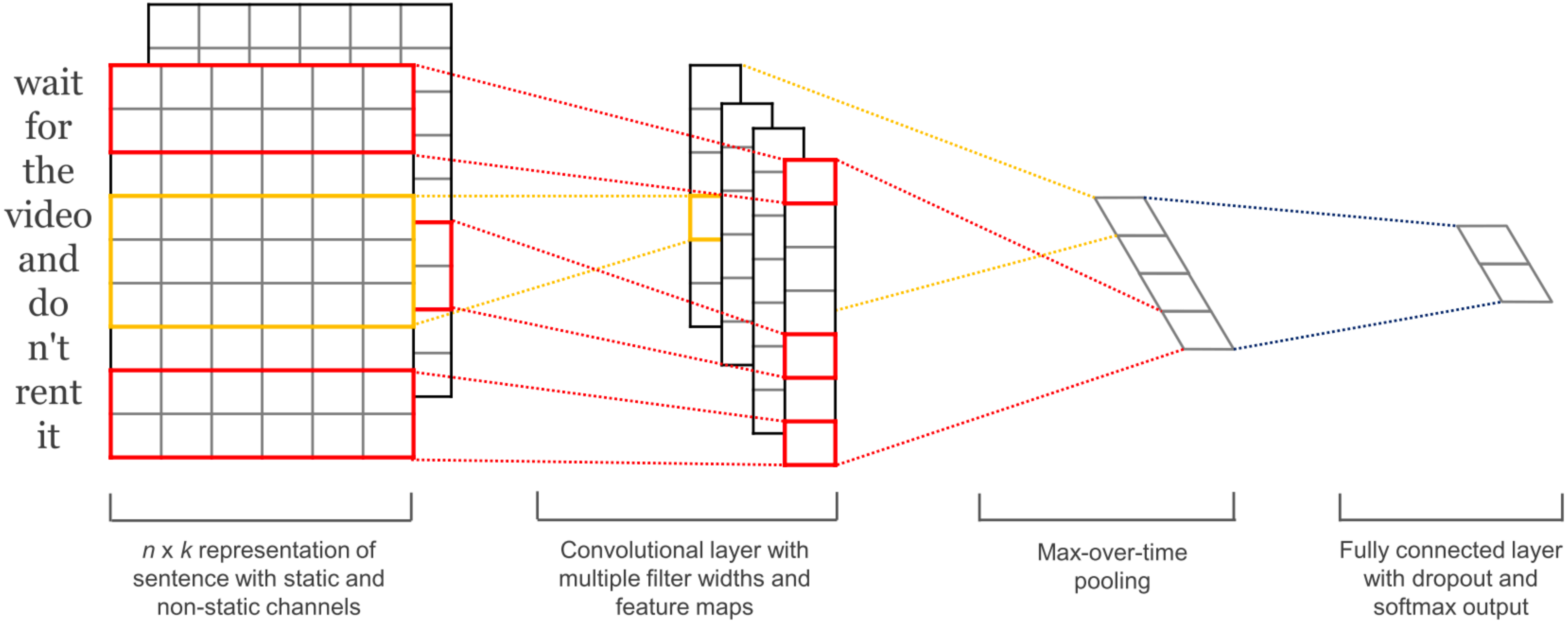


Yi Yang
ASAPP

Chunyang Xiao
Bloomberg

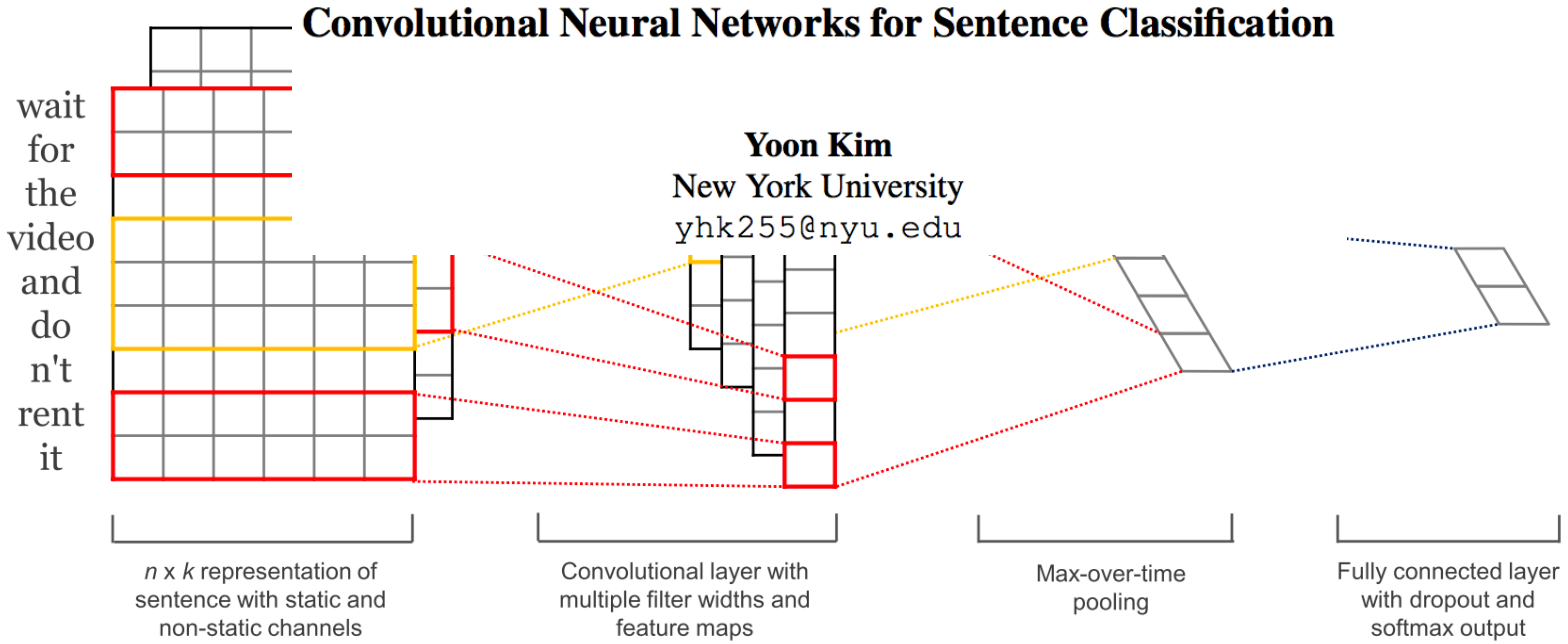
CNNs for NLP problems

[Yoon Kim (2014)]



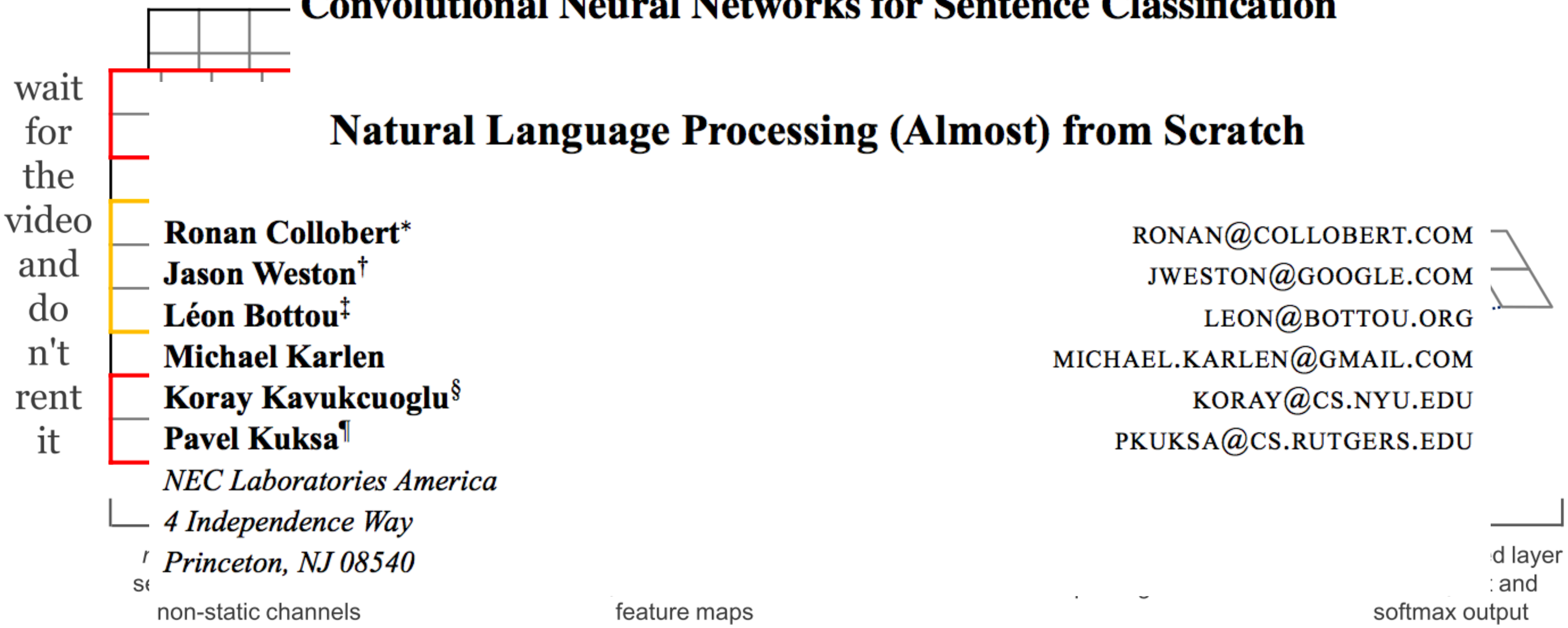
CNNs for NLP problems

[Yoon Kim (2014)]



CNNs for NLP problems

[Yoon Kim (2014)]



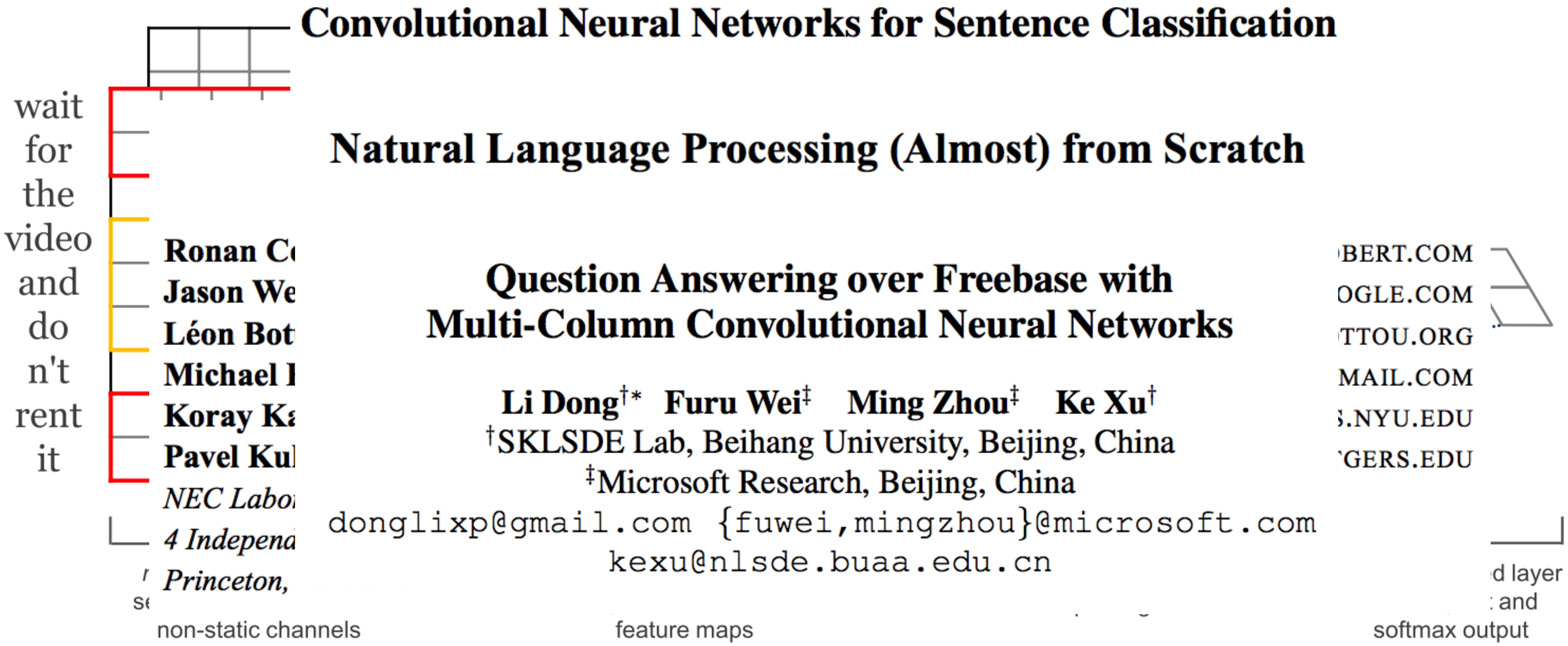
Convolutional Neural Networks for Sentence Classification

Natural Language Processing (Almost) from Scratch

Ronan Collobert* RONAN@COLLOBERT.COM
Jason Weston† JWESTON@GOOGLE.COM
Léon Bottou‡ LEON@BOTTOU.ORG
Michael Karlen MICHAEL.KARLEN@GMAIL.COM
Koray Kavukcuoglu§ KORAY@CS.NYU.EDU
Pavel Kuksa¶ PKUKSA@CS.RUTGERS.EDU
NEC Laboratories America
4 Independence Way
Princeton, NJ 08540

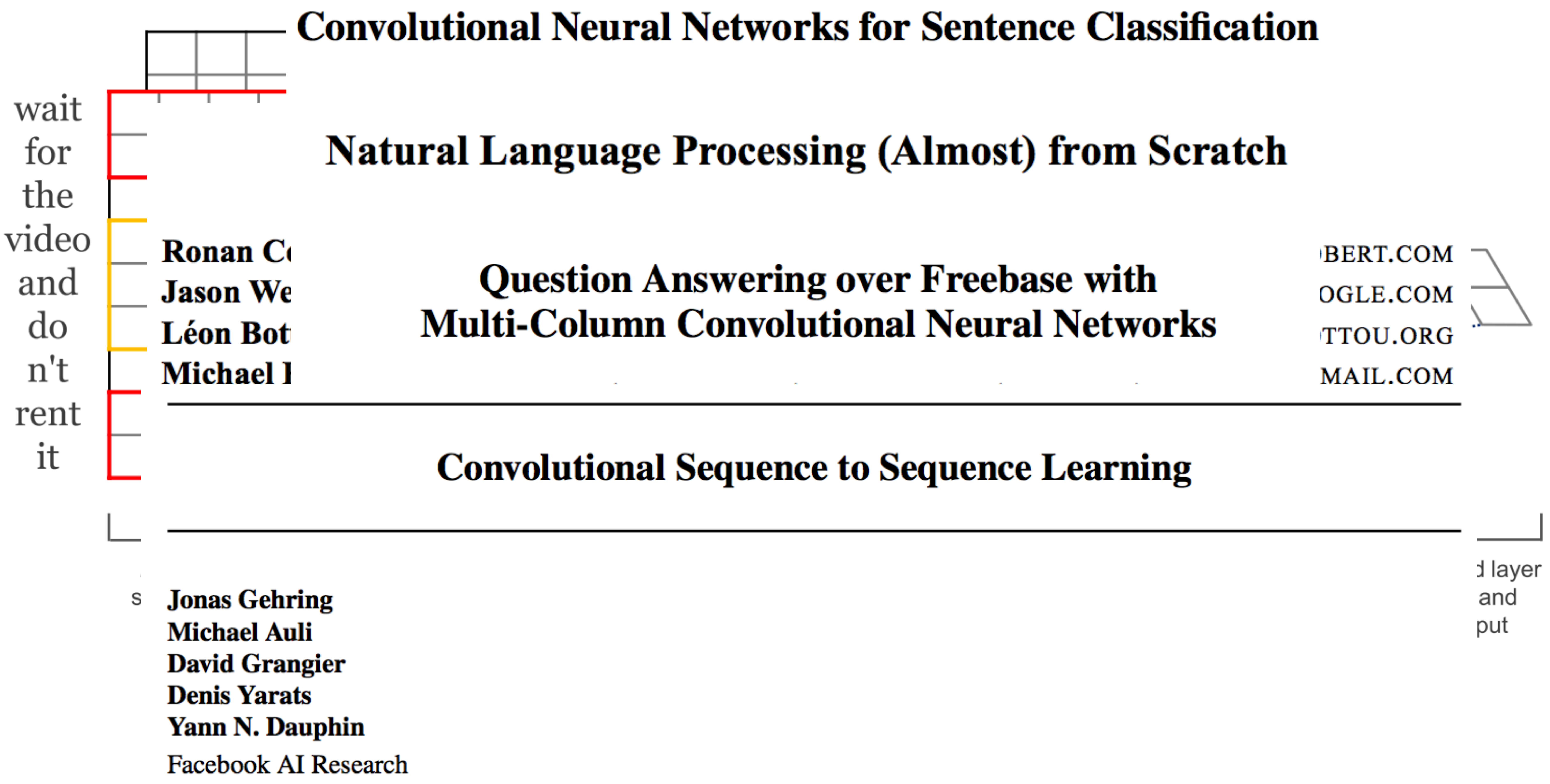
CNNs for NLP problems

[Yoon Kim (2014)]

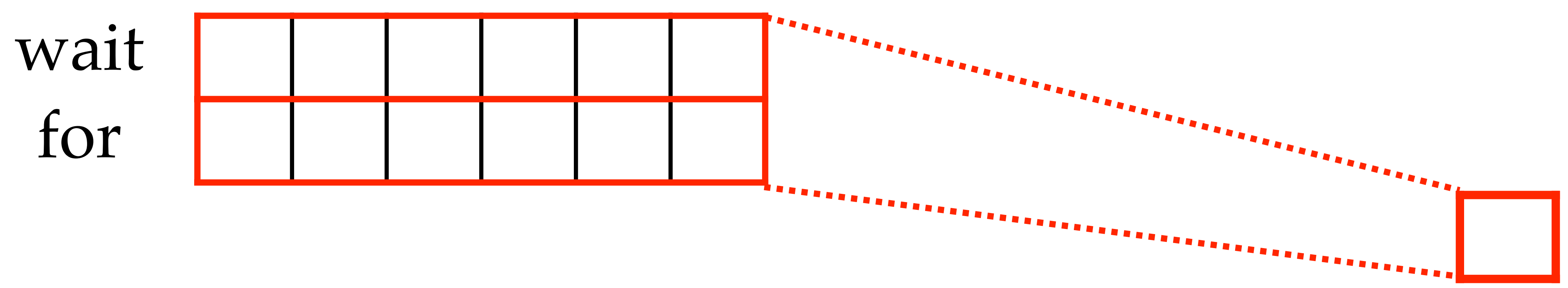


CNNs for NLP problems

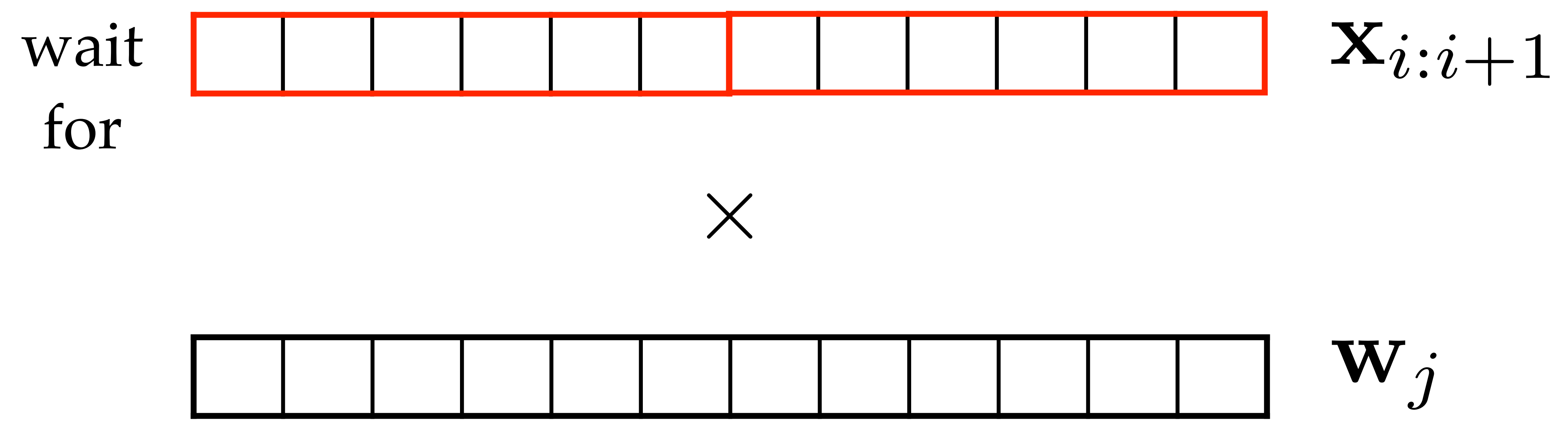
[Yoon Kim (2014)]



Linear convolution filters



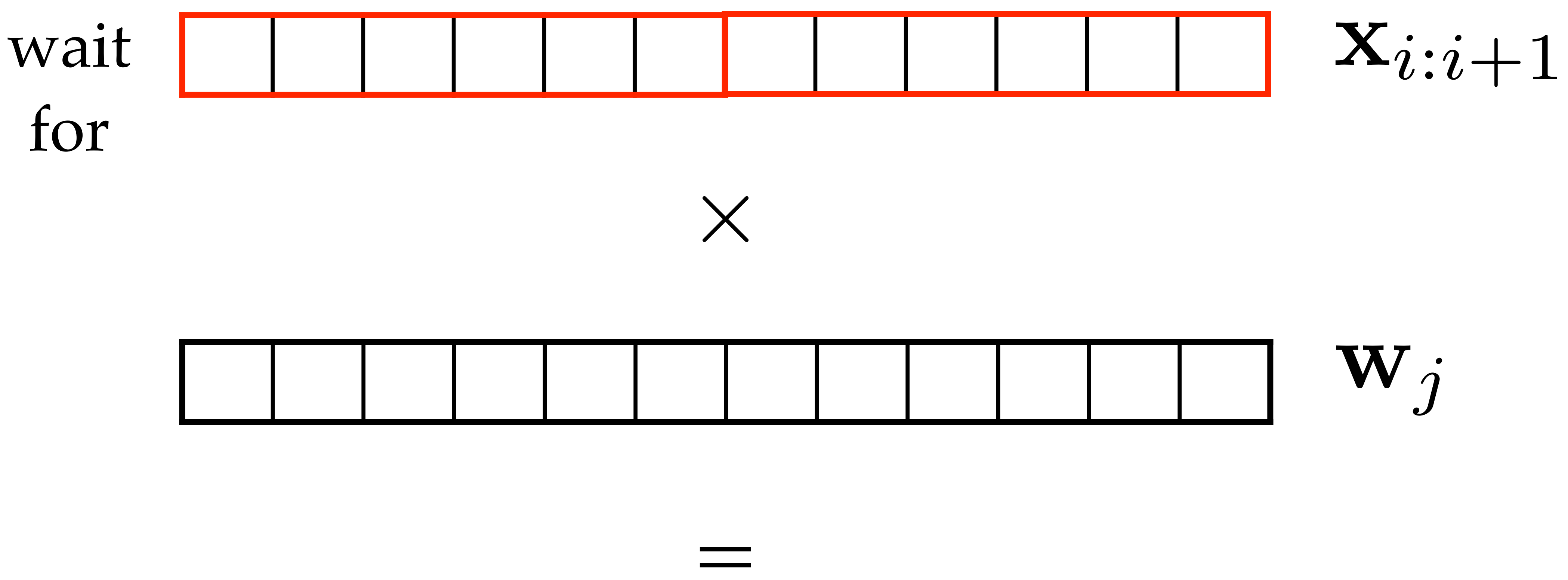
Linear convolution filters



$$=$$

$$c_{i,j} = f(\mathbf{w}_j \mathbf{x}_{i:i+1} + b_j)$$

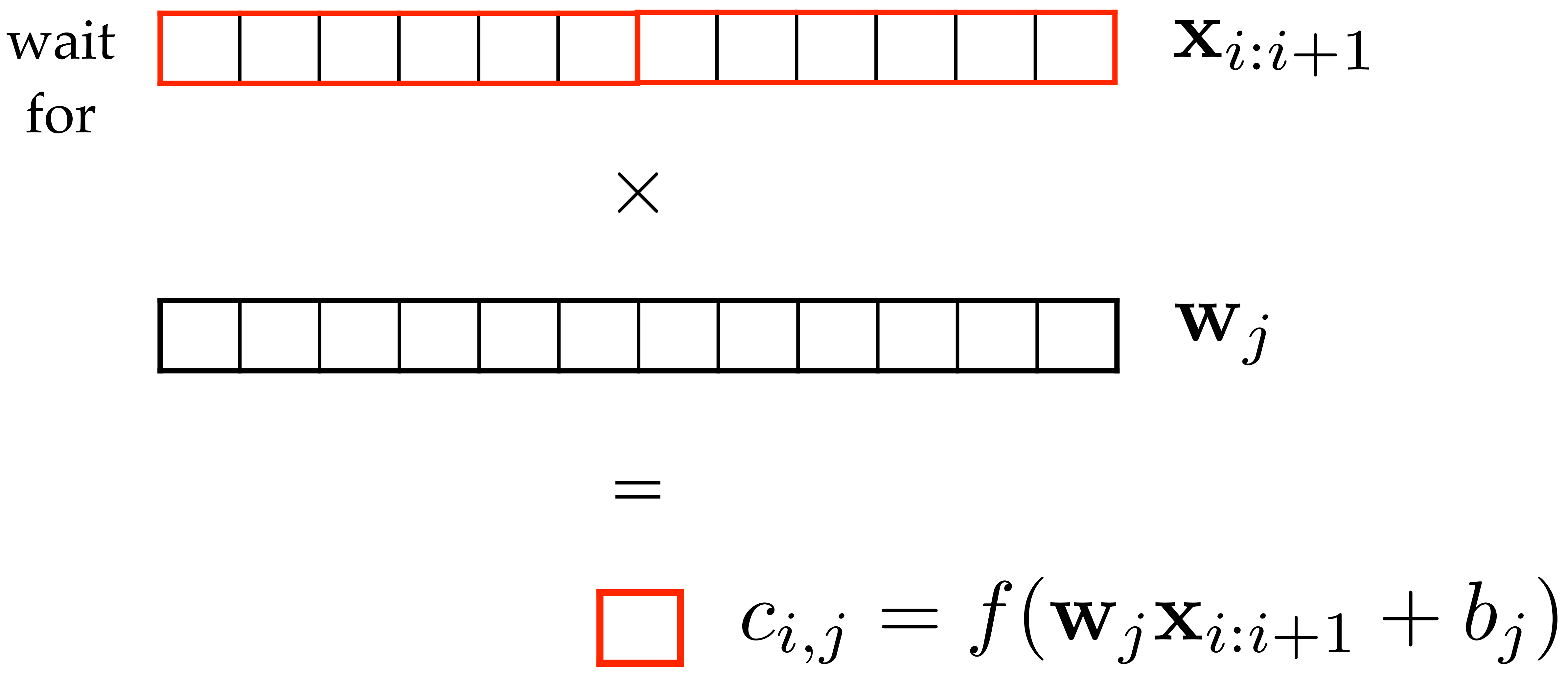
Linear convolution filters



$$c_{i,j} = f(\mathbf{w}_j \mathbf{x}_{i:i+1} + b_j)$$

- Limit high-order filters
  - compositionality
  - long-term deps.

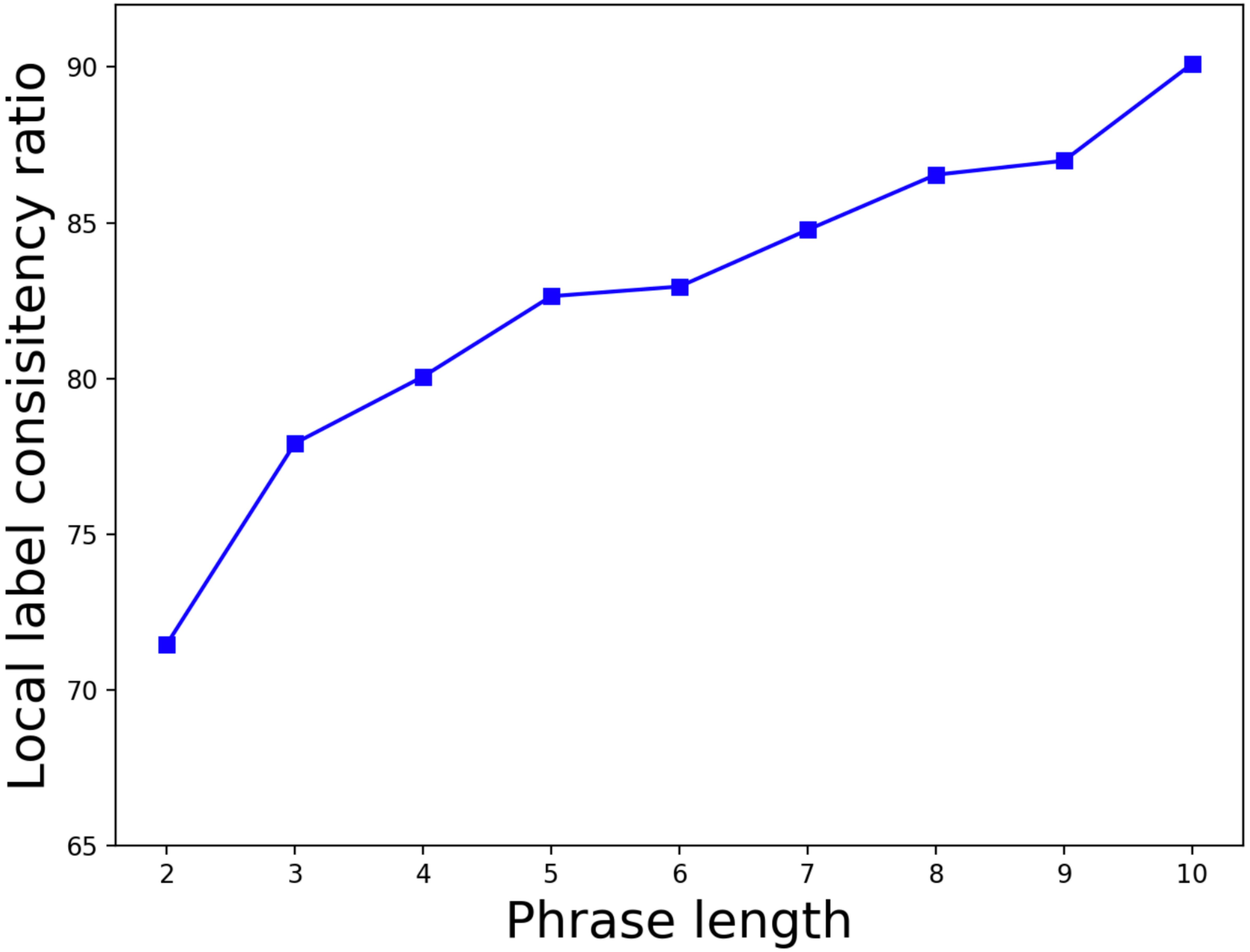
Linear convolution filters



- Limit high-order filters
  - compositionality
  - long-term deps.
- Filters are independent
  - duplication

Local label consistency ratio

Ratio of m-grams that share the same labels as the original sentences.

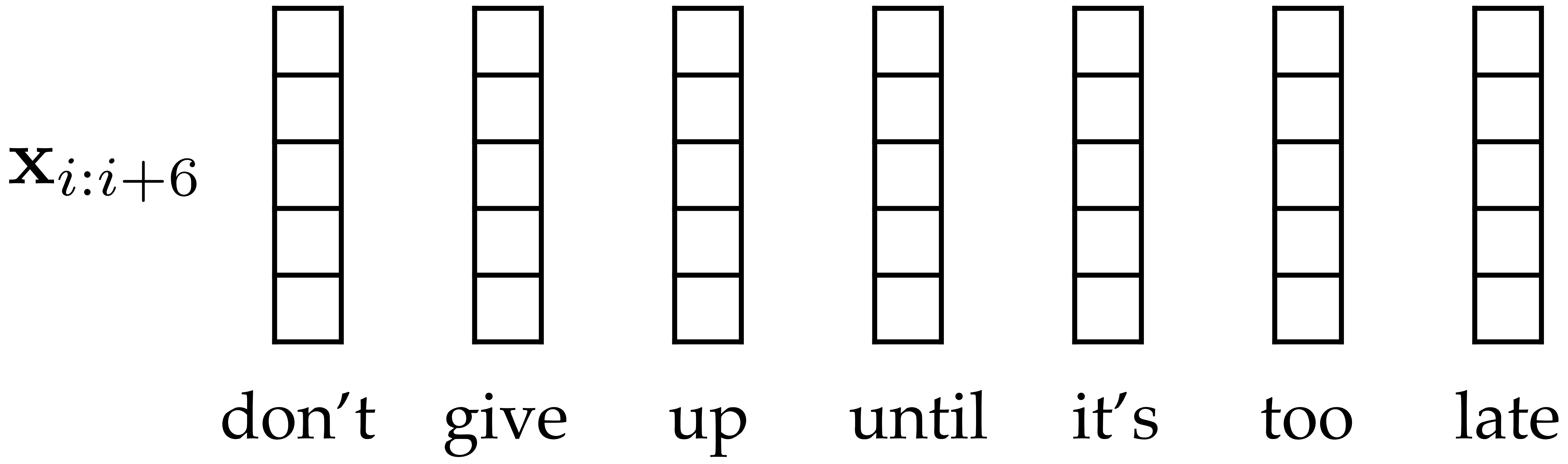


Socher et al. (2013)

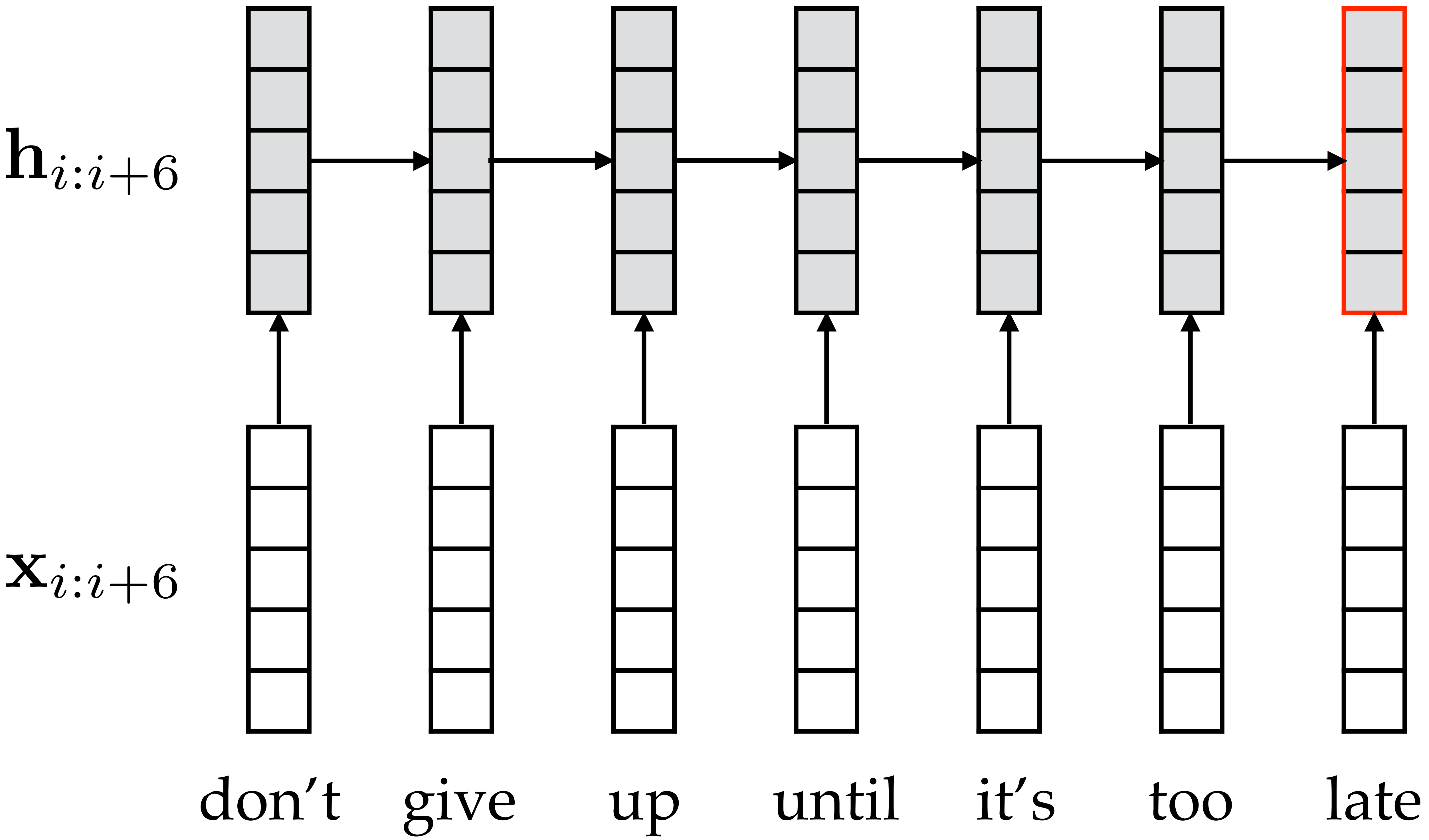
Recurrent neural filters

don’t give up until it’s too late

Recurrent neural filters



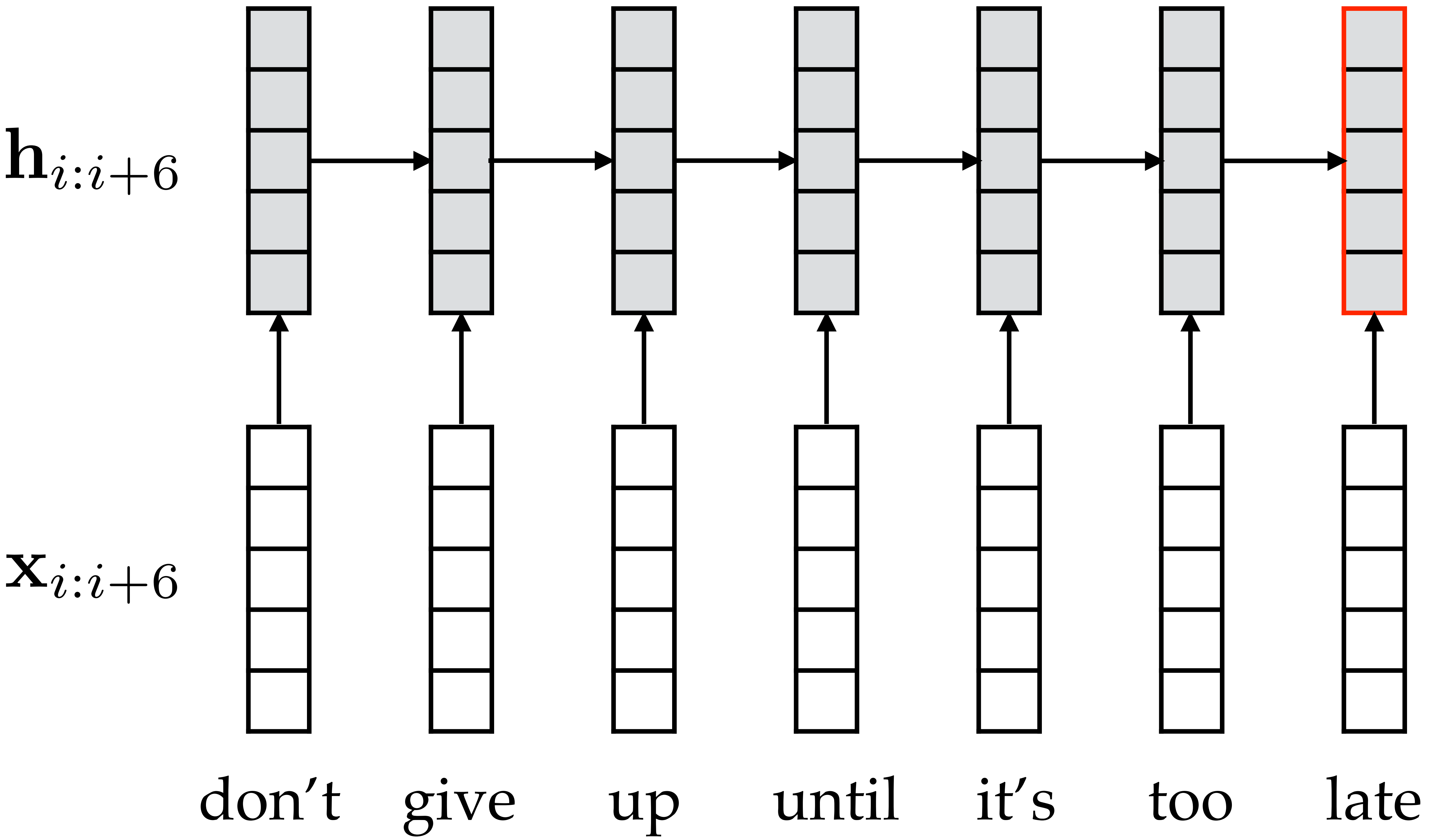
Recurrent neural filters



$$\mathbf{h}_t = \mathrm{RNN}(\mathbf{h}_{t-1}, \mathbf{x}_t)$$

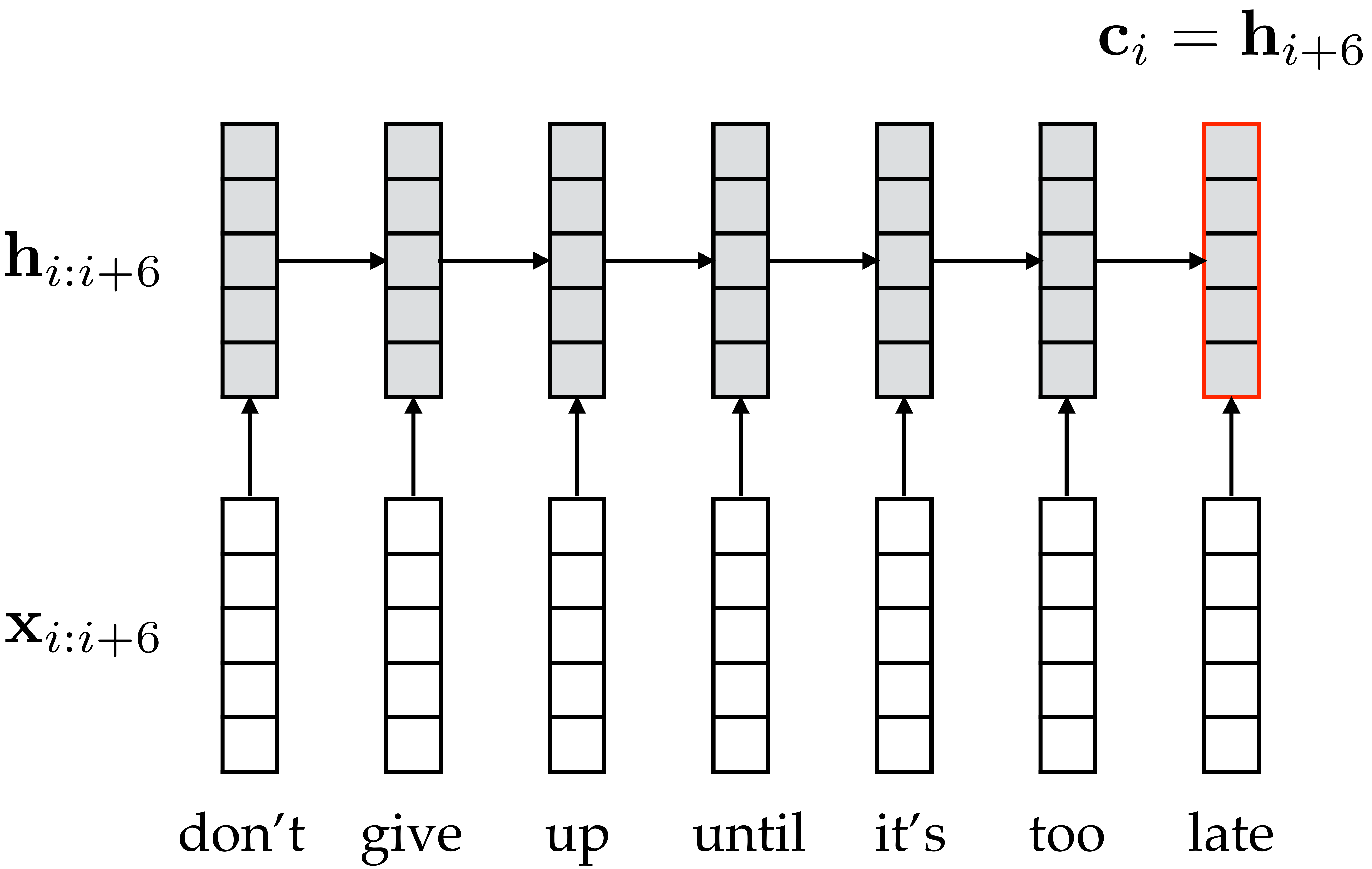
Recurrent neural filters

$$\mathbf{c}_i = \mathbf{h}_{i+6}$$



$$\mathbf{h}_t = \mathrm{RNN}(\mathbf{h}_{t-1}, \mathbf{x}_t)$$

Recurrent neural filters



$$\mathbf{h}_t = \mathrm{RNN}(\mathbf{h}_{t-1}, \mathbf{x}_t)$$

- RNN is implemented as
  - gated recurrent unit
  - LSTM unit

Cho et al. (2014); Hochreiter and Schmidhuber (1997)

CNN architectures

- CNN sentence encoder

$$\mathbf{v} = \max\{\mathbf{C}\}, \text{ where } \mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_{n-m+1}]$$

Yoon Kim (2014); Yang et al. (2015)

CNN architectures

- CNN sentence encoder

$$\mathbf{v} = \max\{\mathbf{C}\}, \text{ where } \mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_{n-m+1}]$$

- Sentence classification (e.g., sentiment classification)

$$p(y|\mathbf{v}) = \mathrm{Softmax}(\mathbf{W}_v \mathbf{v})$$

Yoon Kim (2014); Yang et al. (2015)

CNN architectures

- CNN sentence encoder

$$\mathbf{v} = \max\{\mathbf{C}\}, \text{ where } \mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_{n-m+1}]$$

- Sentence classification (e.g., sentiment classification)

$$p(y|\mathbf{v}) = \mathrm{Softmax}(\mathbf{W}_v \mathbf{v})$$

- Sentence matching (e.g., answer sentence selection)

$$p(y|\mathbf{v}_1, \mathbf{v}_2) = \mathrm{Sigmoid}(\mathbf{v}_1^\top \mathbf{W}_v \mathbf{v}_2)$$

Yoon Kim (2014); Yang et al. (2015)

Data

- Sentence classification
  - Stanford Sentiment Treebank (SST)
    - Binary classification / fine-grained classification

Socher et al. (2013); Wang et al. (2007); Yang et al. (2015)

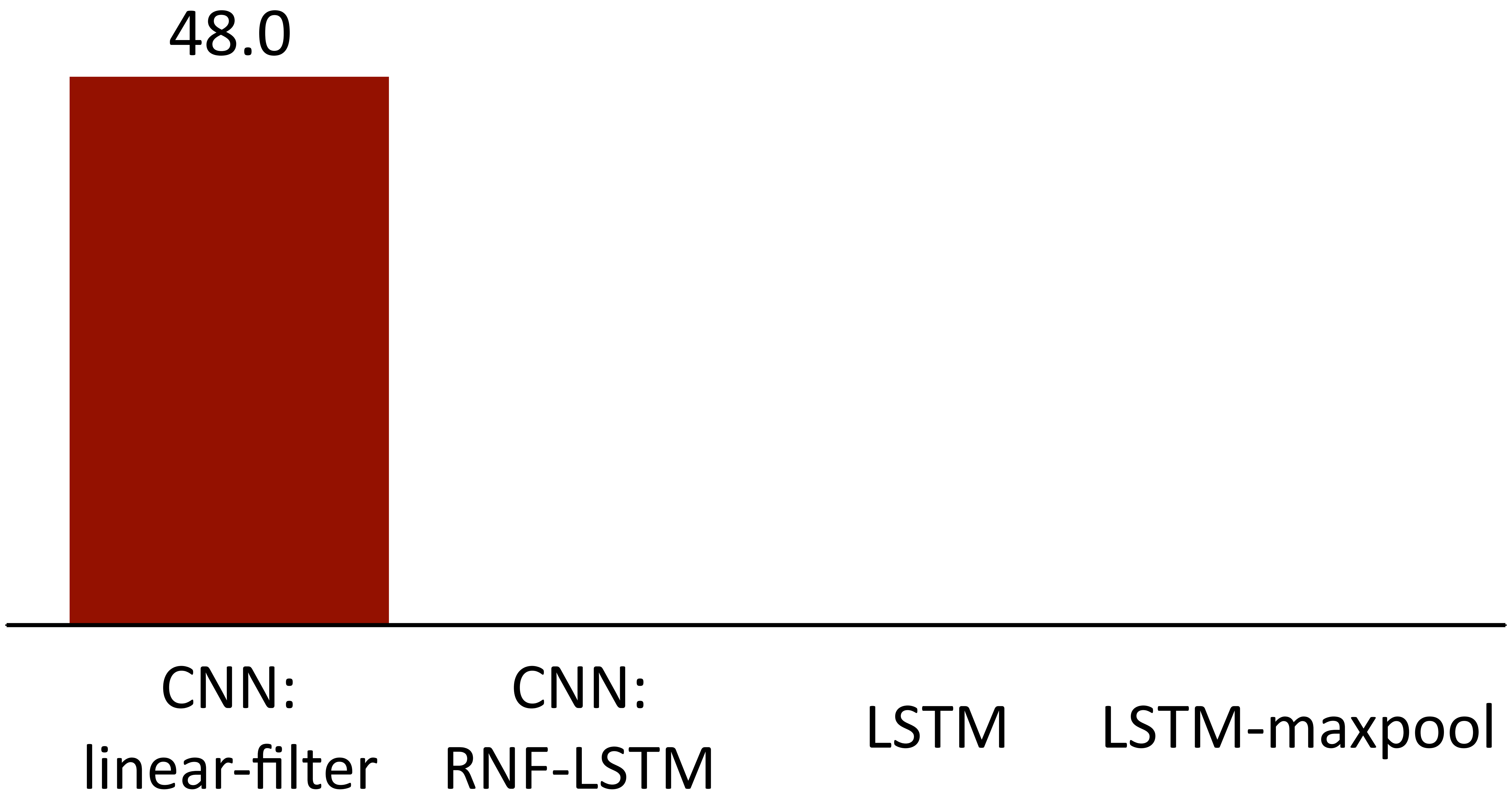
Data

- Sentence classification
  - Stanford Sentiment Treebank (SST)
    - Binary classification / fine-grained classification
- Sentence matching
  - QASent
  - WikiQA

Socher et al. (2013); Wang et al. (2007); Yang et al. (2015)

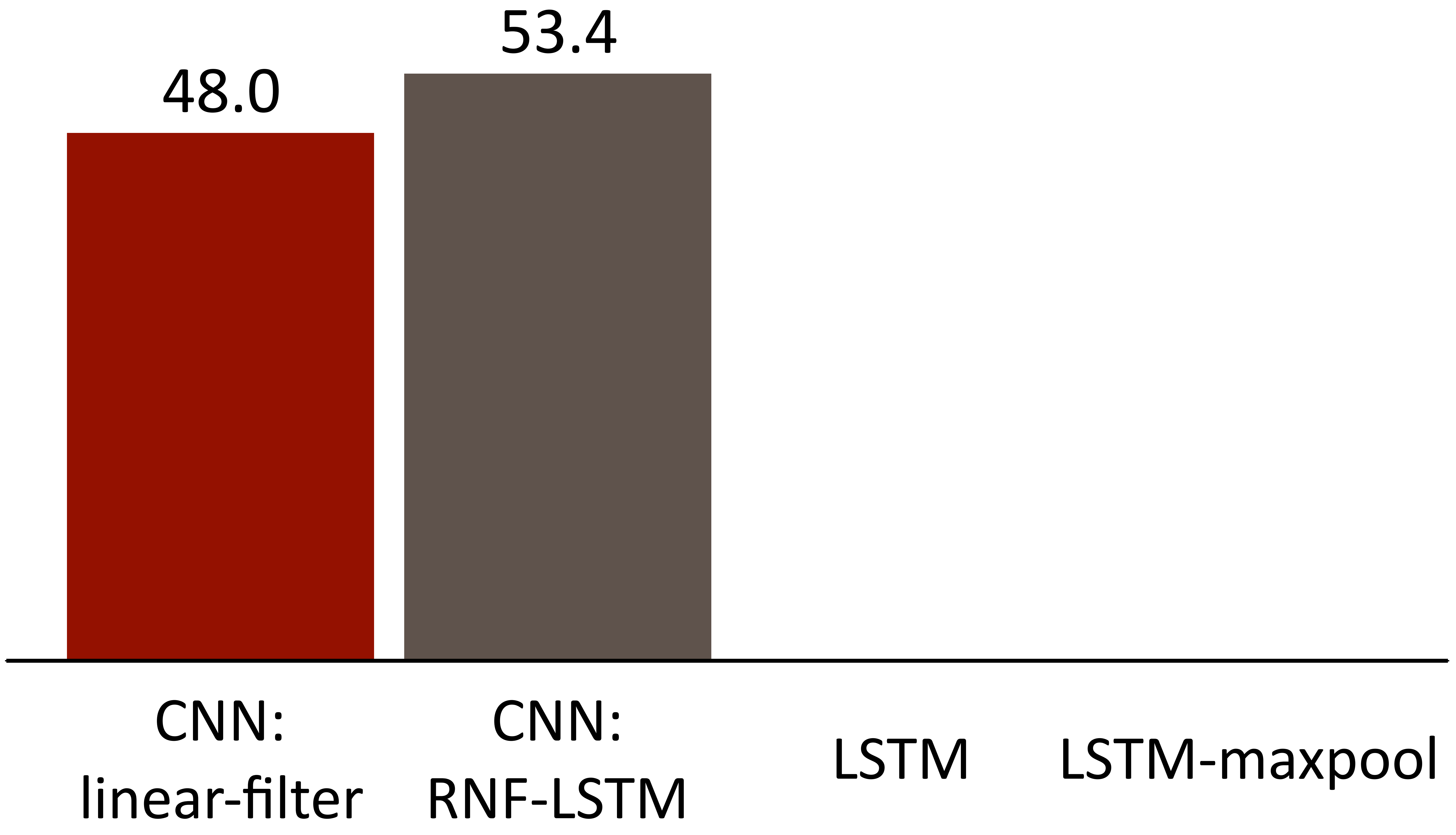
Results: sentence classification

- Accuracy results for fine-grained sentiment classification



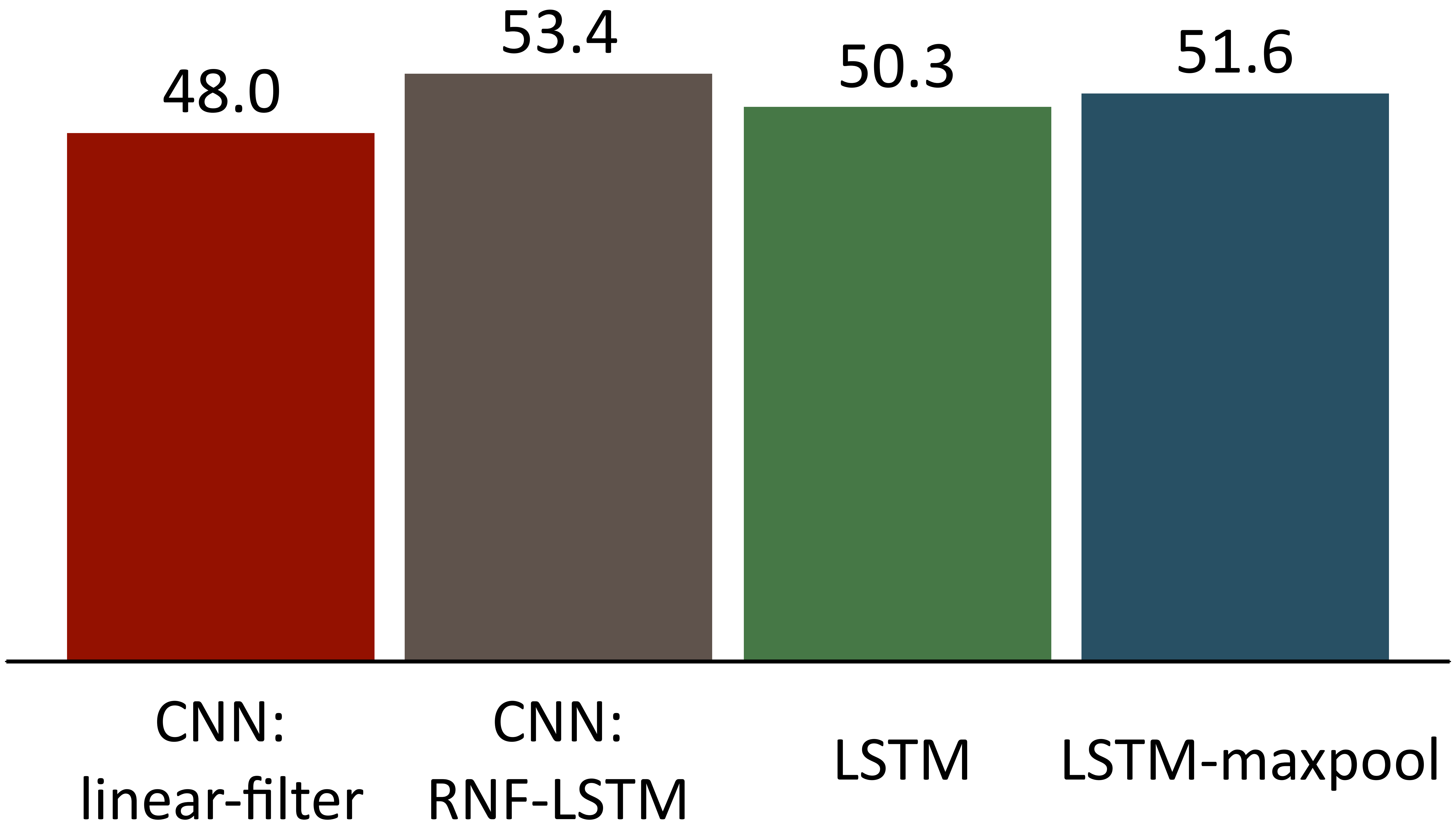
Results: sentence classification

- Accuracy results for fine-grained sentiment classification



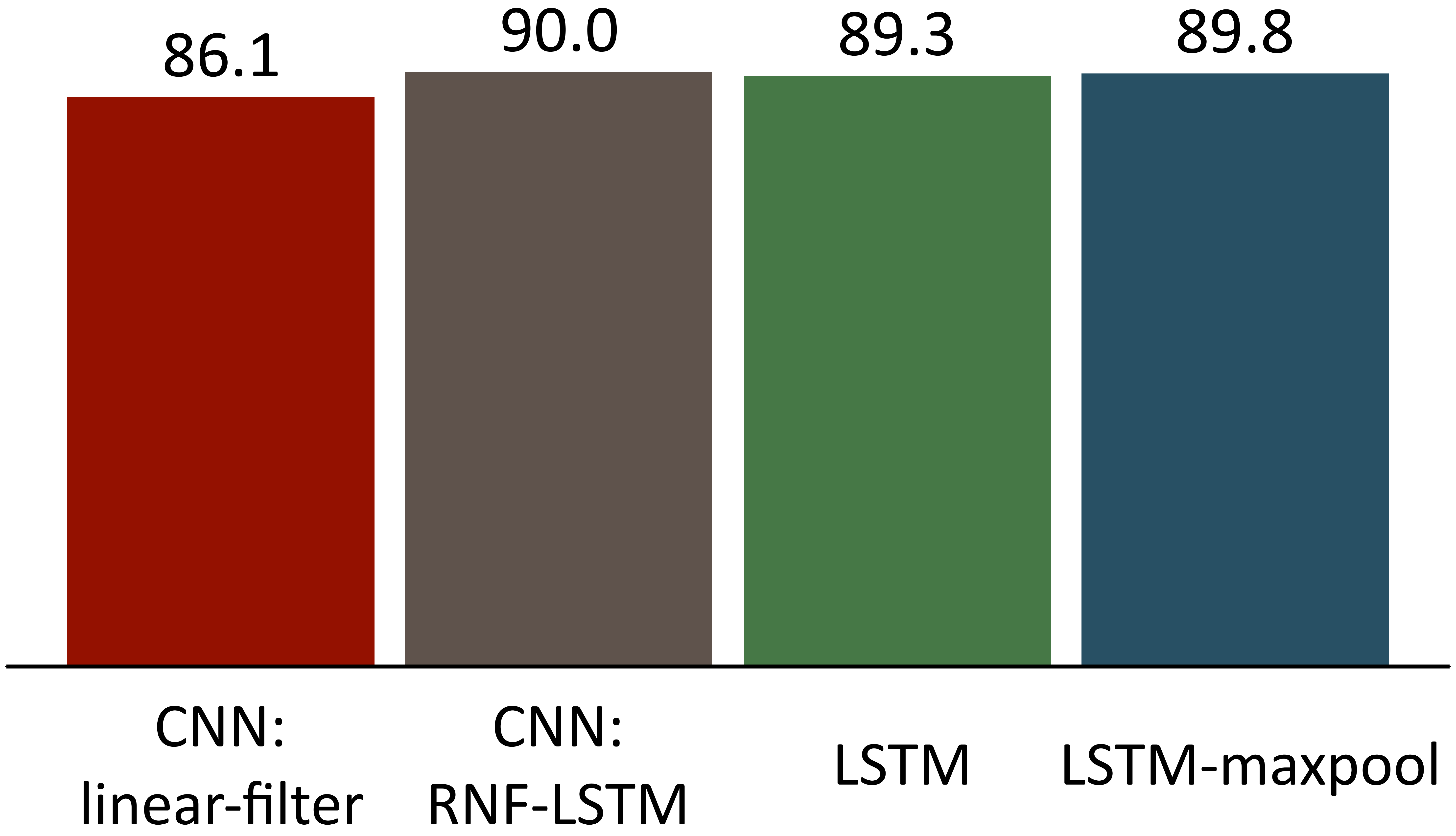
Results: sentence classification

- Accuracy results for fine-grained sentiment classification



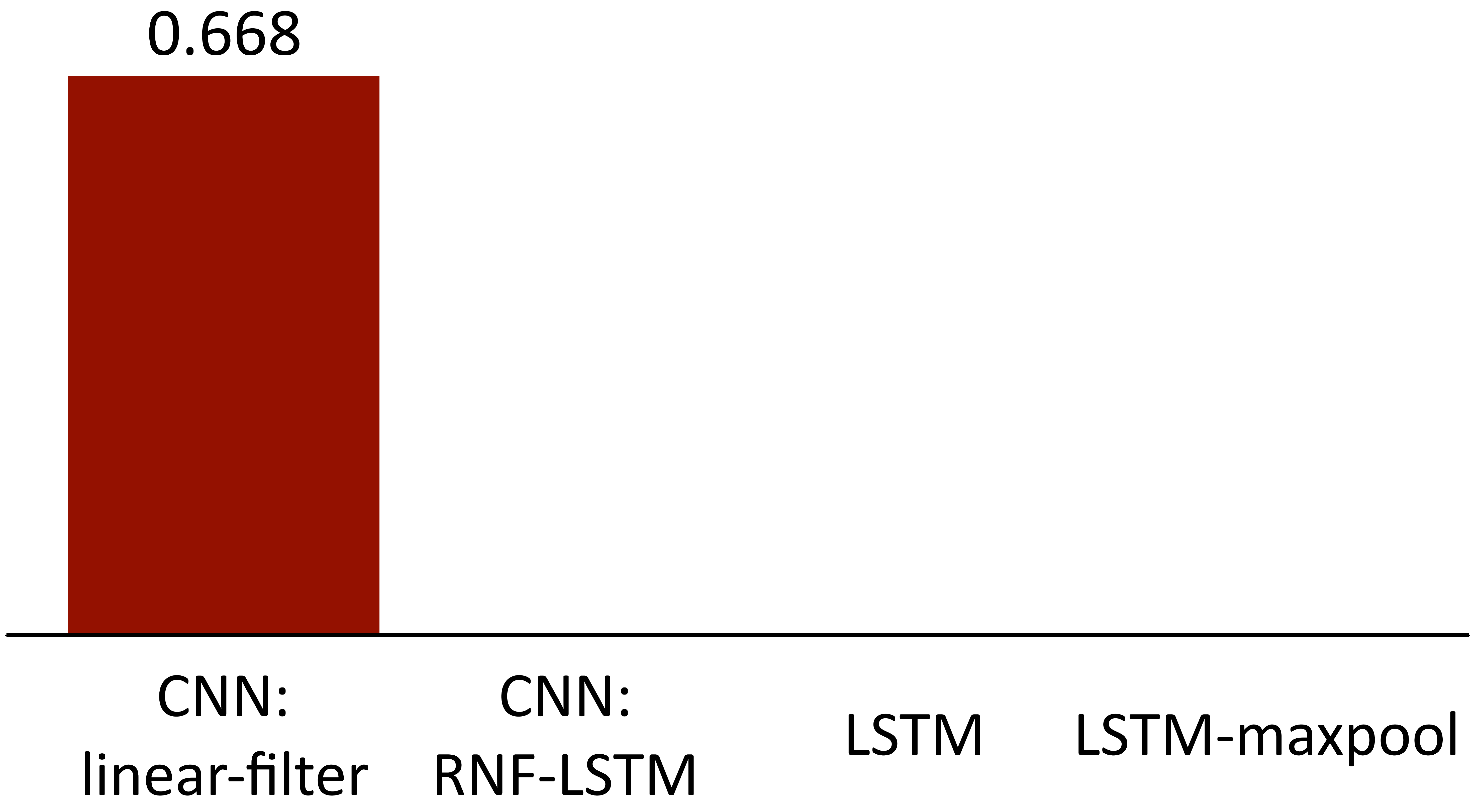
Results: sentence classification

- Accuracy results for binary sentiment classification



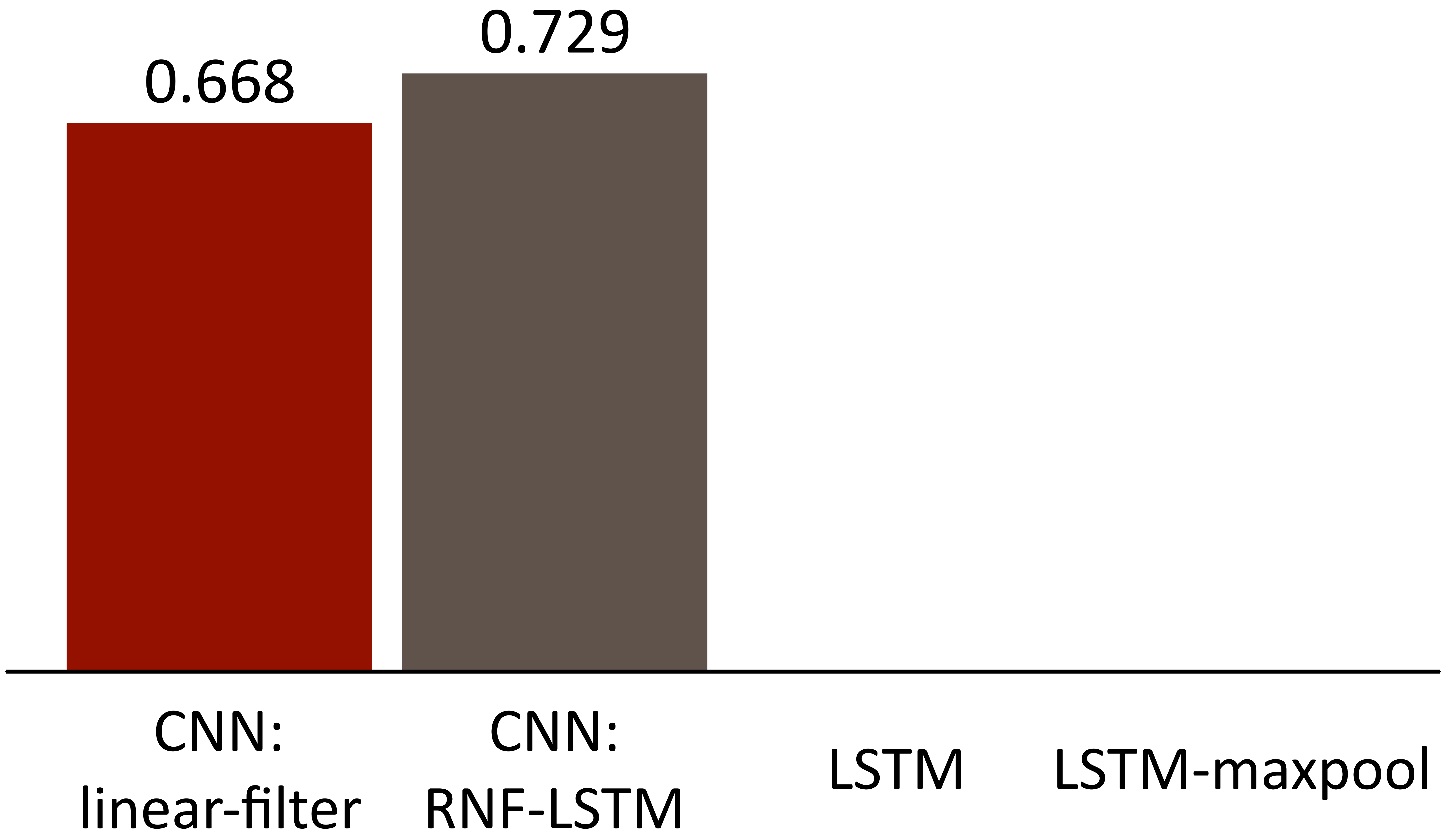
Results: sentence matching

- MAP results on the WikiQA dataset



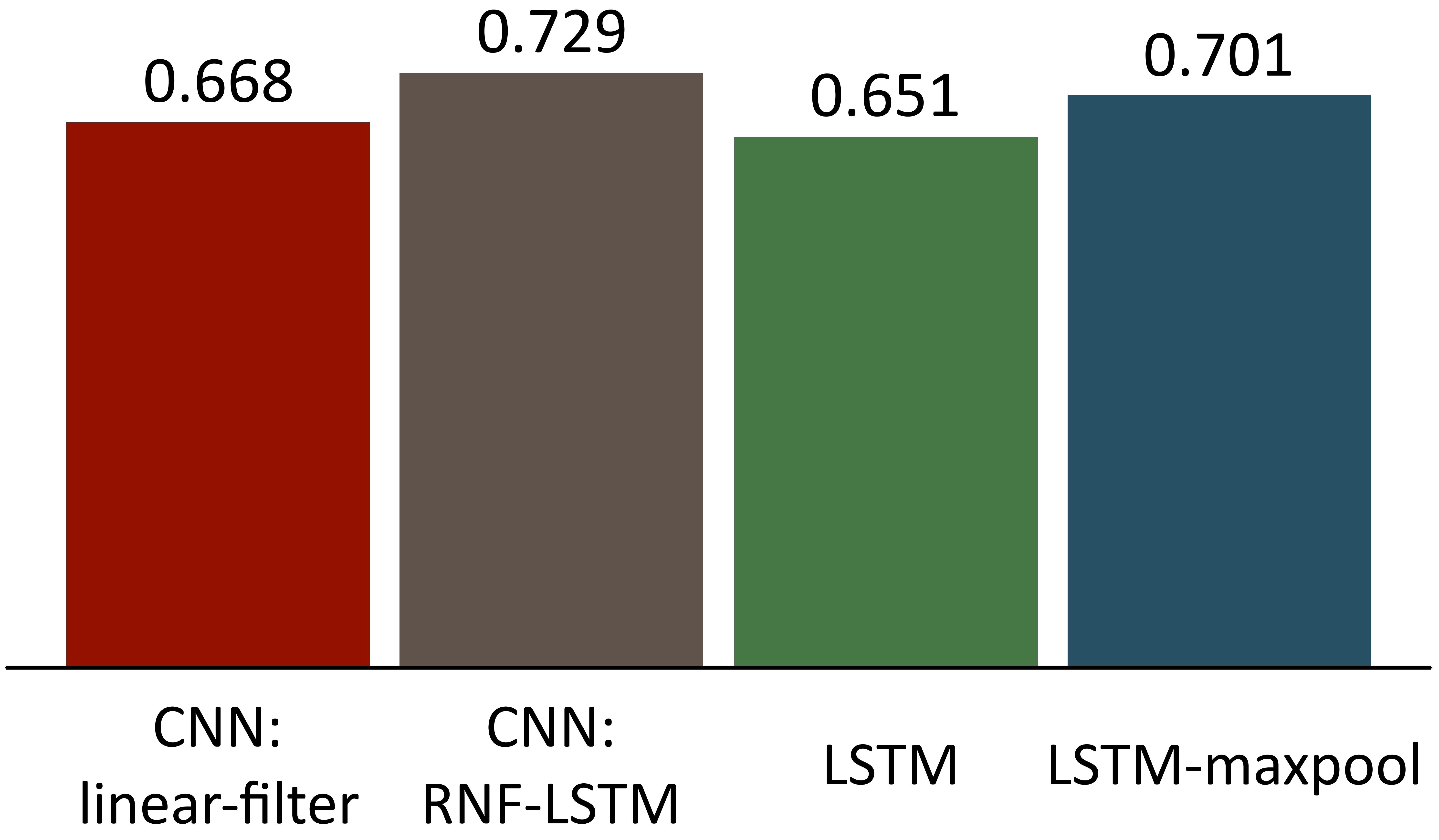
Results: sentence matching

- MAP results on the WikiQA dataset



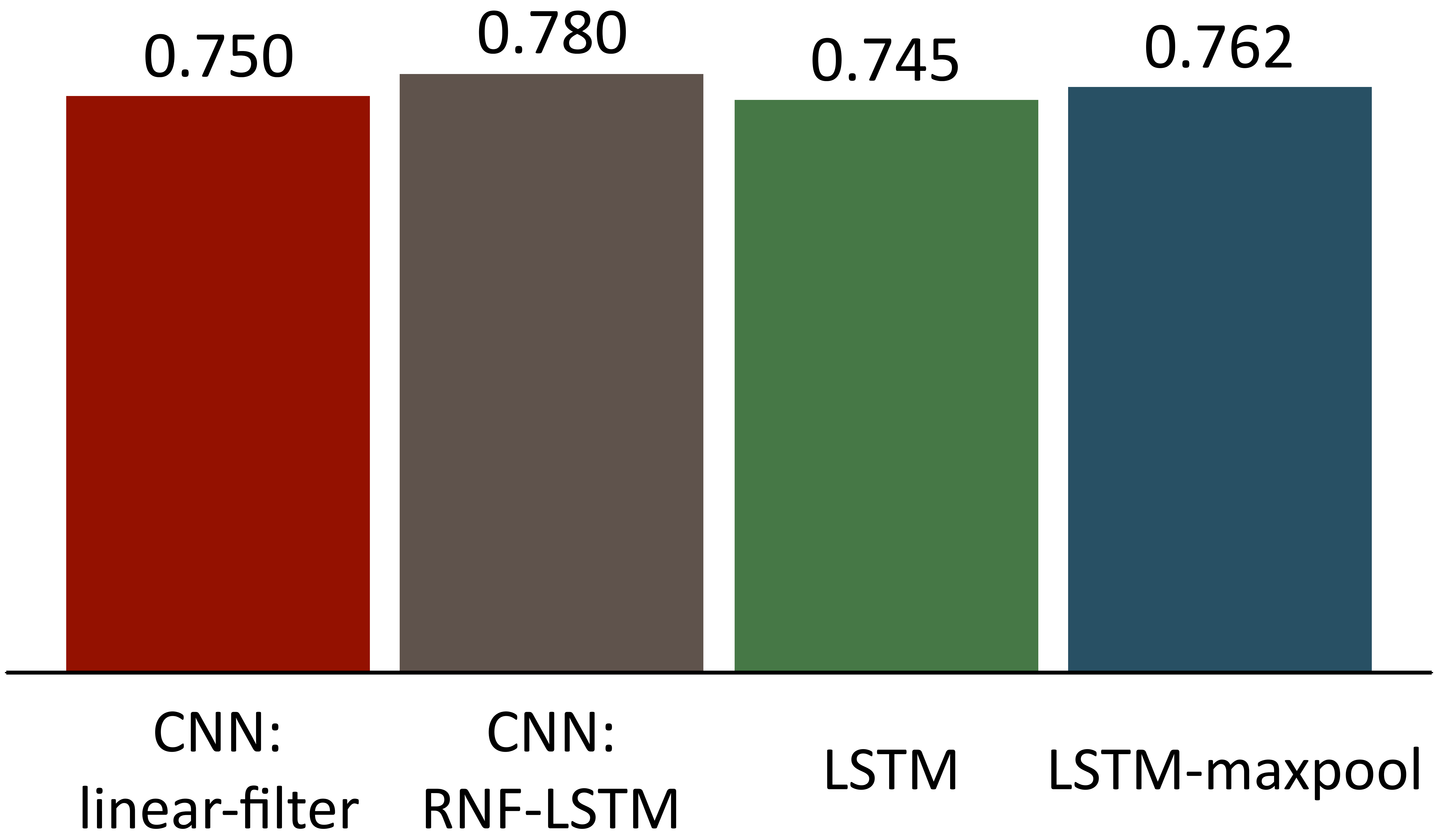
Results: sentence matching

- MAP results on the WikiQA dataset



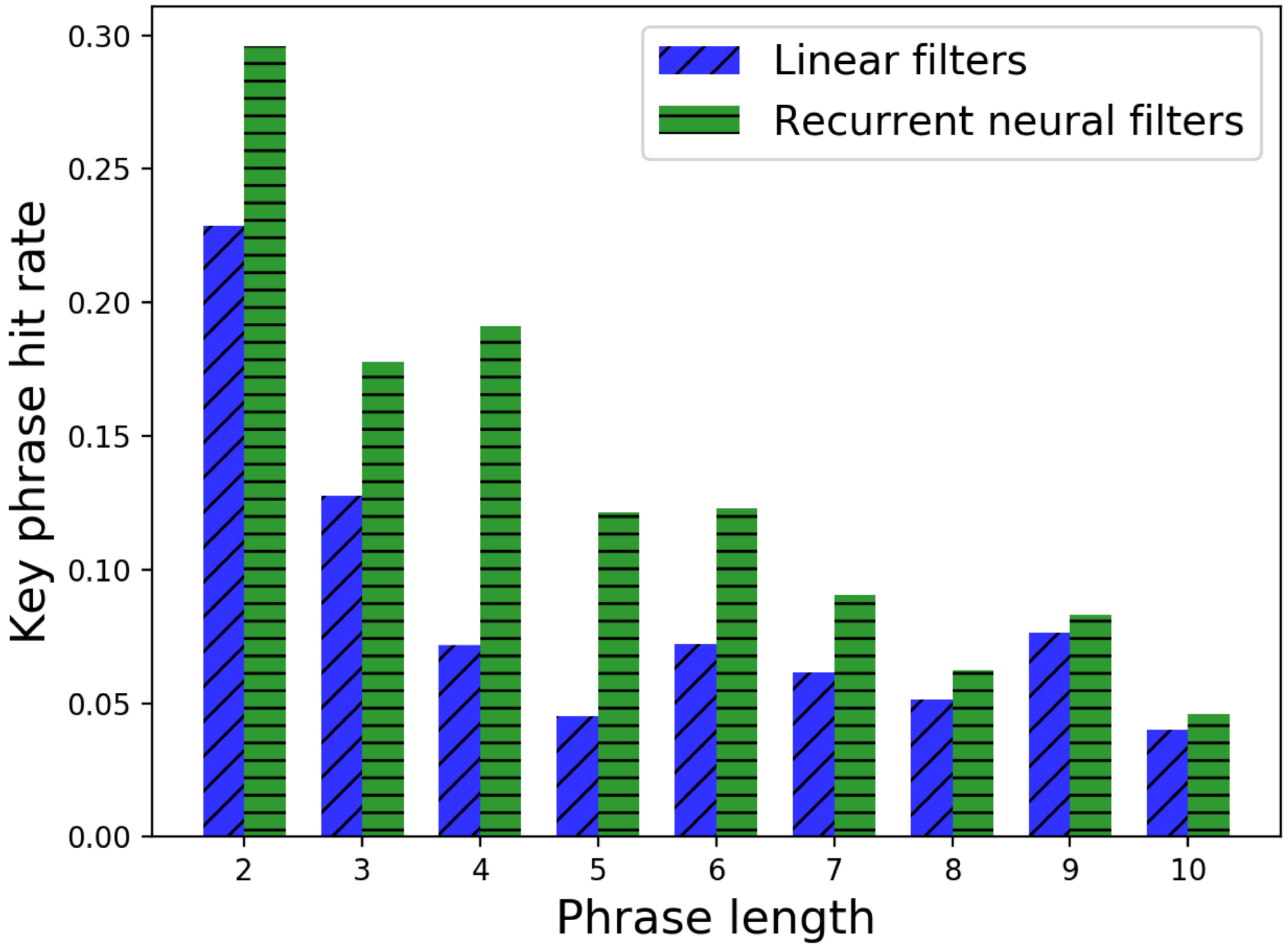
Results: sentence matching

- MAP results on the QASent dataset



Key phrase hit rate

Key phrases: the phrase label is the same as the sentence label (SST)



Conclusions

- Conventional CNNs adopt linear convolution filters that fails to account for language compositionality.

- Recurrent neural filters (RNFs) yield much better results than linear filters on many NLP tasks.

- Code: https://github.com/bloomberg/cnn-rnf