Neural networks

Natural language processing - convolutional network

WORD TAGGING

Topics: word tagging

- In many NLP applications, it is useful to augment text data with syntactic and semantic information
  - we would like to add syntactic/semantic labels to each word
- This problem can be tackled using a conditional random field with neural network unary potentials
  - we will describe the model developed by Ronan Collobert and Jason Weston in:

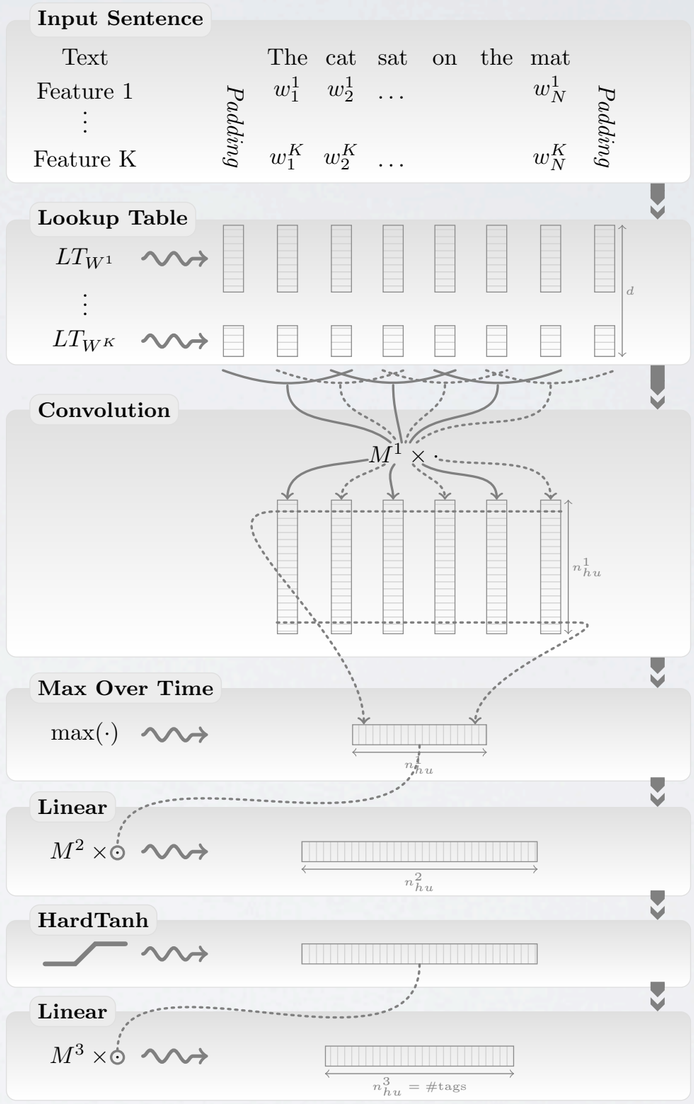
    A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning
    Collobert and Weston, 2008

    (see Natural Language Processing (Almost) from Scratch for the journal version)

SENTENCE NEURAL NETWORK

Topics: sentence convolutional network

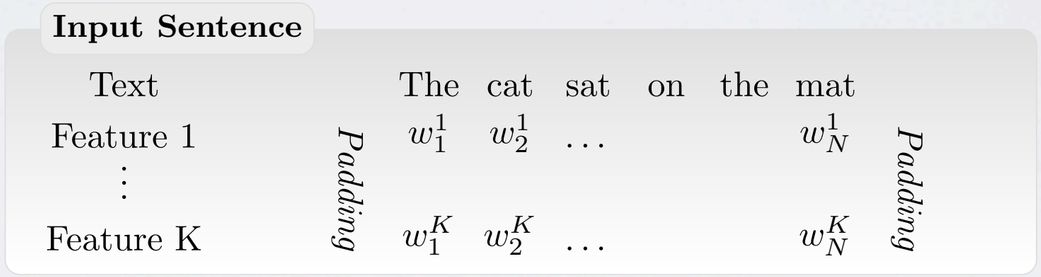
- How to model each label sequence
  - could use a CRF with neural network unary potentials, based on a window (context) of words
    - not appropriate for semantic role labeling, because relevant context might be very far away
  - Collobert and Weston suggest a convolutional network over the whole sentence
    - prediction at a given position can exploit information from any word in the sentence



SENTENCE NEURAL NETWORK

Topics: sentence convolutional network

- Each word can be represented by more than one feature
  - feature of the word itself
  - substring features
    - prefix: " `eating` " → " `eat` "
    - suffix: " `eating` " → " `ing` "
  - gazetteer features
    - whether the word belong to a list of known locations, persons, etc.



- These features are treated like word features, with their own lookup tables

SENTENCE NEURAL NETWORK

Topics: sentence convolutional network

- Feature must encode for which word we are making a prediction
  - done by adding the relative position $i\text{-}pos_w$, where pos_w is the position of the current word
  - this feature also has its lookup table

Input Sentence

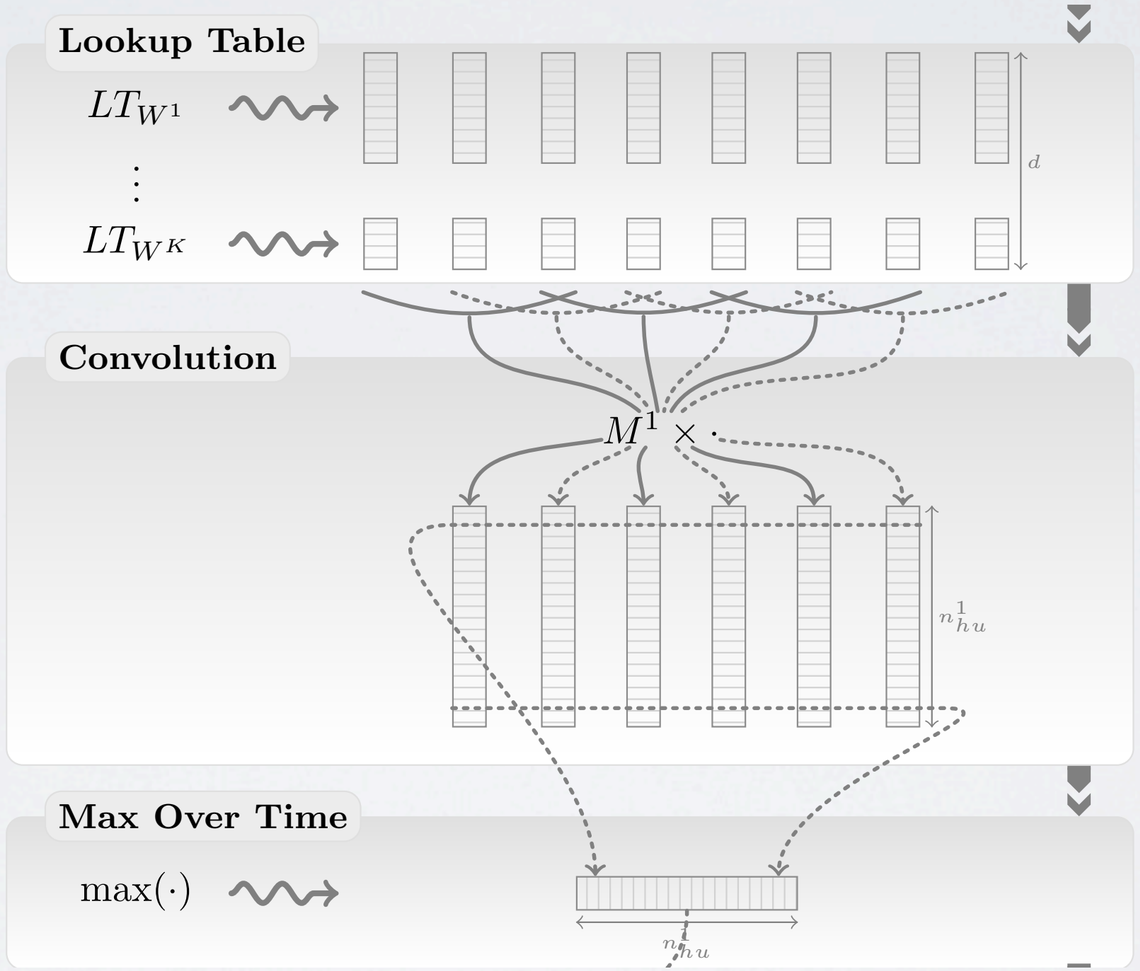
Text		The	cat	sat	on	the	mat	
Feature 1	*Padding*	w_1^1	w_2^1	...			w_N^1	*Padding*
⋮								
Feature K		w_1^K	w_2^K	...			w_N^K	

- For SRL, must know the roles for which verb we are predicting
  - also add the relative position of that verb $i\text{-}pos_v$

SENTENCE NEURAL NETWORK

Topics: sentence convolutional network

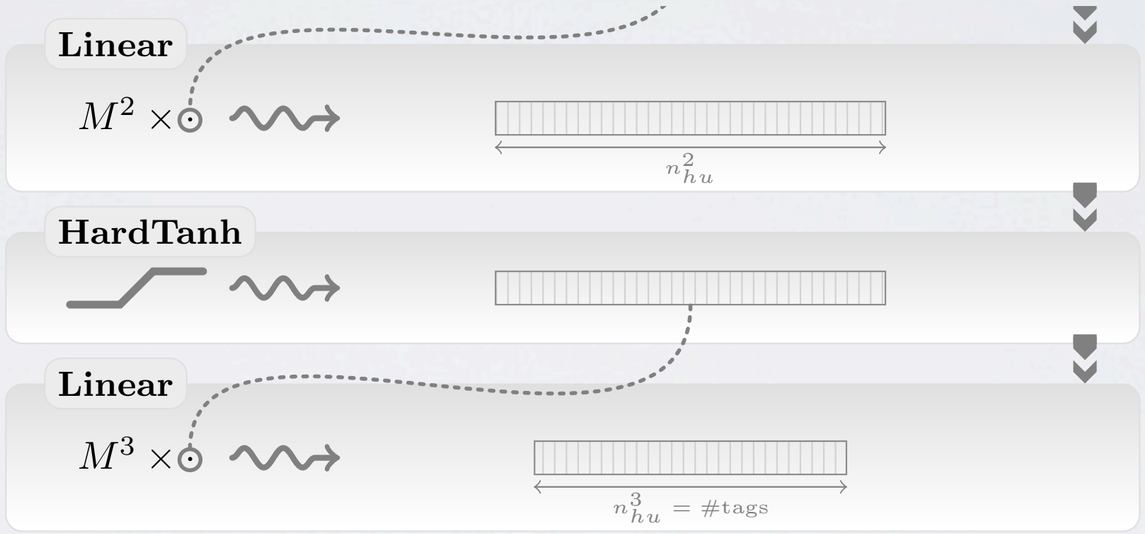
- Lookup table:
  - for each word concatenate the representations of its features
- Convolution:
  - at every position, compute linear activations from a window of representations
  - this is a convolution in 1D
- Max pooling:
  - obtain a fixed hidden layer with a max across positions



SENTENCE NEURAL NETWORK

Topics: sentence convolutional network

- Regular neural network:
  - the pooled representation serves as the input of a regular neural network
  - they proposed using a "hard" version of the tanh activation function



- The outputs are used as the unary potential of a chain CRF over the labels
  - no connections between the CRFs of the different task (one CRF per task)
  - a separate neural network is used for each task