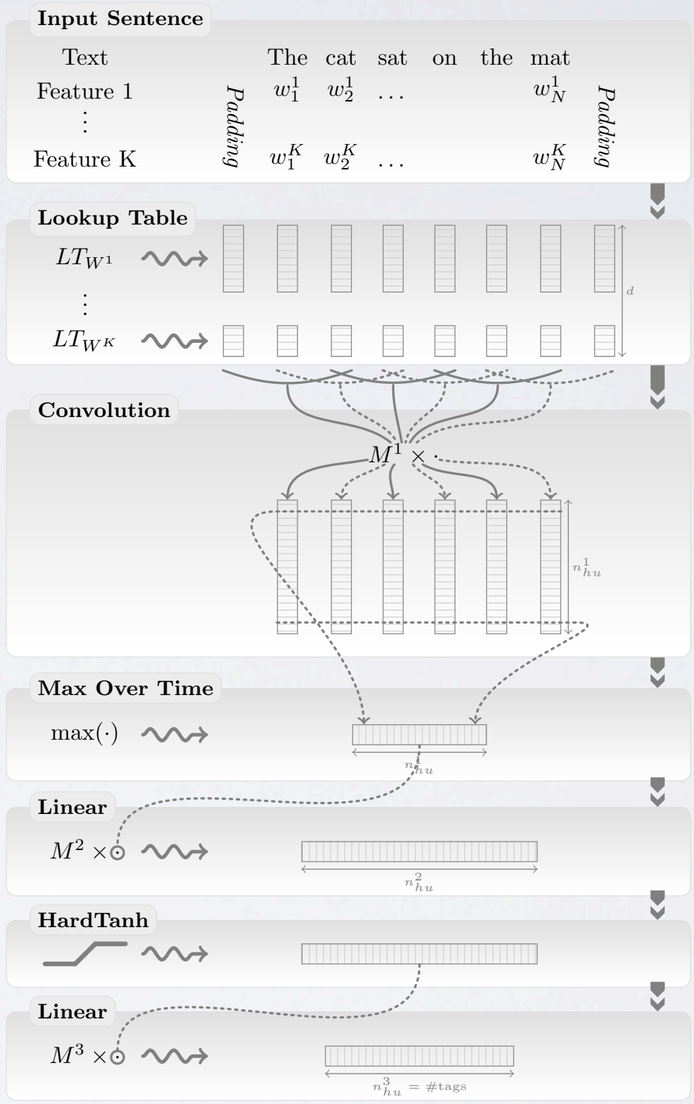
Neural networks

Natural language processing - multitask learning

SENTENCE NEURAL NETWORK

Topics: sentence convolutional network

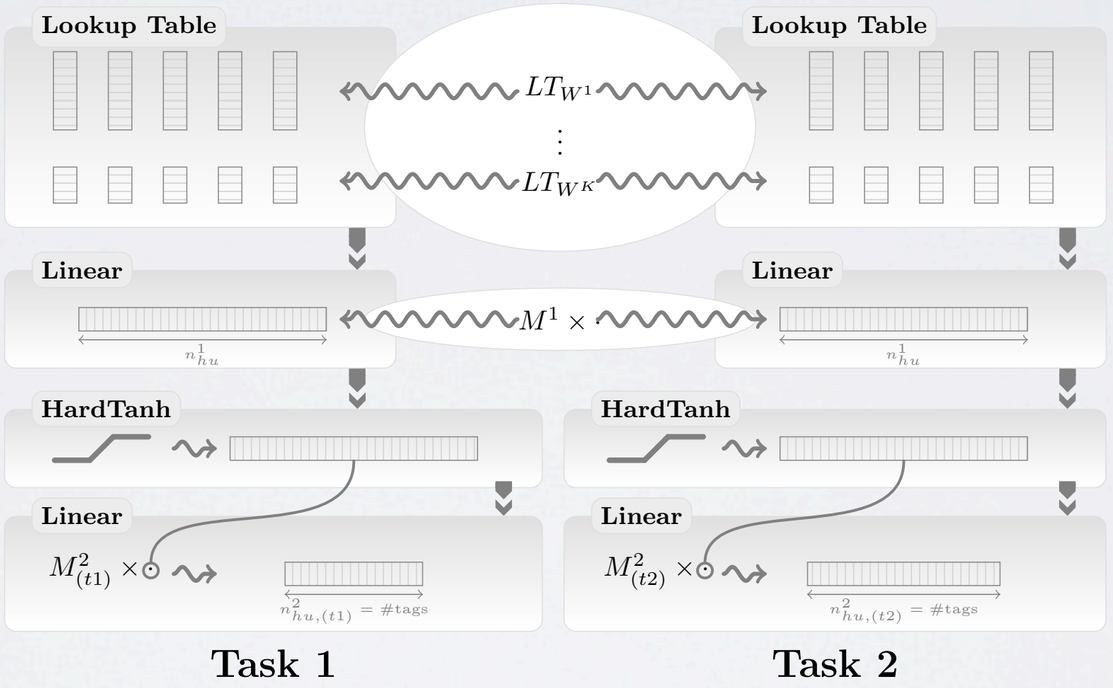
- How to model each label sequence
  - could use a CRF with neural network unary potentials, based on a window (context) of words
    - not appropriate for semantic role labeling, because relevant context might be very far away
  - Collobert and Weston suggest a convolutional network over the whole sentence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



SENTENCE NEURAL NETWORK

Topics: multitask learning

- Could share vector representations of the features across tasks
  - simply use the same lookup tables across tasks
  - the other parameters of the neural networks are not tied



- This is referred to as multitask learning
  - the idea is to transfer knowledge learned within the word representations across the different task

SENTENCE NEURAL NETWORK

Topics: language model

- We can design other tasks without any labeled data
  - identify whether the middle word of a window of text is an "impostor"

    "**`cat sat on the mat`**" vs "**`cat sat think the mat`**"

  - can generate impostor examples from unlabeled text (Wikipedia)
    - pick a window of words from unlabeled corpus
    - replace middle word with a different, randomly chosen word
  - train a neural network (with word representations) to assign a higher score to the original window

original window

$$\max\left\{0, 1 - f_\theta(x) + f_\theta(x^{(w)})\right\}$$

impostor window with word w

  - similar to language modeling, except we predict the word in the middle

SENTENCE NEURAL NETWORK

Topics: experimental comparison

- From Natural Language Processing (Almost) from Scratch by Collobert et al.

Approach	**POS** (PWA)	**CHUNK** (F1)	**NER** (F1)	**SRL** (F1)
Benchmark Systems	97.24	94.29	89.31	77.92
NN+SLL	96.37	90.33	81.47	70.99

SENTENCE NEURAL NETWORK

Topics: experimental comparison

- From Natural Language Processing (Almost) from Scratch by Collobert et al.

Approach	POS (PWA)	CHUNK (F1)	NER (F1)	SRL (F1)
Benchmark Systems	97.24	94.29	89.31	77.92
NN+SLL	96.37	90.33	81.47	70.99
NN+SLL+LM2	97.12	93.37	88.78	74.15
NN+SLL+LM2+MTL	97.22	93.75	88.27	74.29

SENTENCE NEURAL NETWORK

Topics: experimental comparison

- From Natural Language Processing (Almost) from Scratch by Collobert et al.

Approach	POS (PWA)	CHUNK (F1)	NER (F1)	SRL (F1)
Benchmark Systems	97.24	94.29	89.31	77.92
NN+SLL	96.37	90.33	81.47	70.99
NN+SLL+LM2	97.12	93.37	88.78	74.15
NN+SLL+LM2+MTL	97.22	93.75	88.27	74.29
NN+SLL+LM2+Suffix2	97.29	–	–	–
NN+SLL+LM2+Gazetteer	–	–	89.59	–
NN+SLL+LM2+POS	–	94.32	88.67	–
NN+SLL+LM2+CHUNK	–	–	–	74.72

SENTENCE NEURAL NETWORK

Topics: experimental comparison

- Nearest neighbors in word representation space:

FRANCE	JESUS	XBOX	REDDISH	SCRATCHED	MEGABITS
454	1973	6909	11724	29869	87025
AUSTRIA	GOD	AMIGA	GREENISH	NAILED	OCTETS
BELGIUM	SATI	PLAYSTATION	BLUISH	SMASHED	MB/S
GERMANY	CHRIST	MSX	PINKISH	PUNCHED	BIT/S
ITALY	SATAN	IPOD	PURPLISH	POPPED	BAUD
GREECE	KALI	SEGA	BROWNISH	CRIMPED	CARATS
SWEDEN	INDRA	PSNUMBER	GREYISH	SCRAPED	KBIT/S
NORWAY	VISHNU	HD	GRAYISH	SCREWED	MEGAHERTZ
EUROPE	ANANDA	DREAMCAST	WHITISH	SECTIONED	MEGAPIXELS
HUNGARY	PARVATI	GEFORCE	SILVERY	SLASHED	GBIT/S
SWITZERLAND	GRACE	CAPCOM	YELLOWISH	RIPPED	AMPERES

- For a 2D visualization: http://www.cs.toronto.edu/~hinton/turian.png