LANGUAGE MODELING

Natural Language Processing

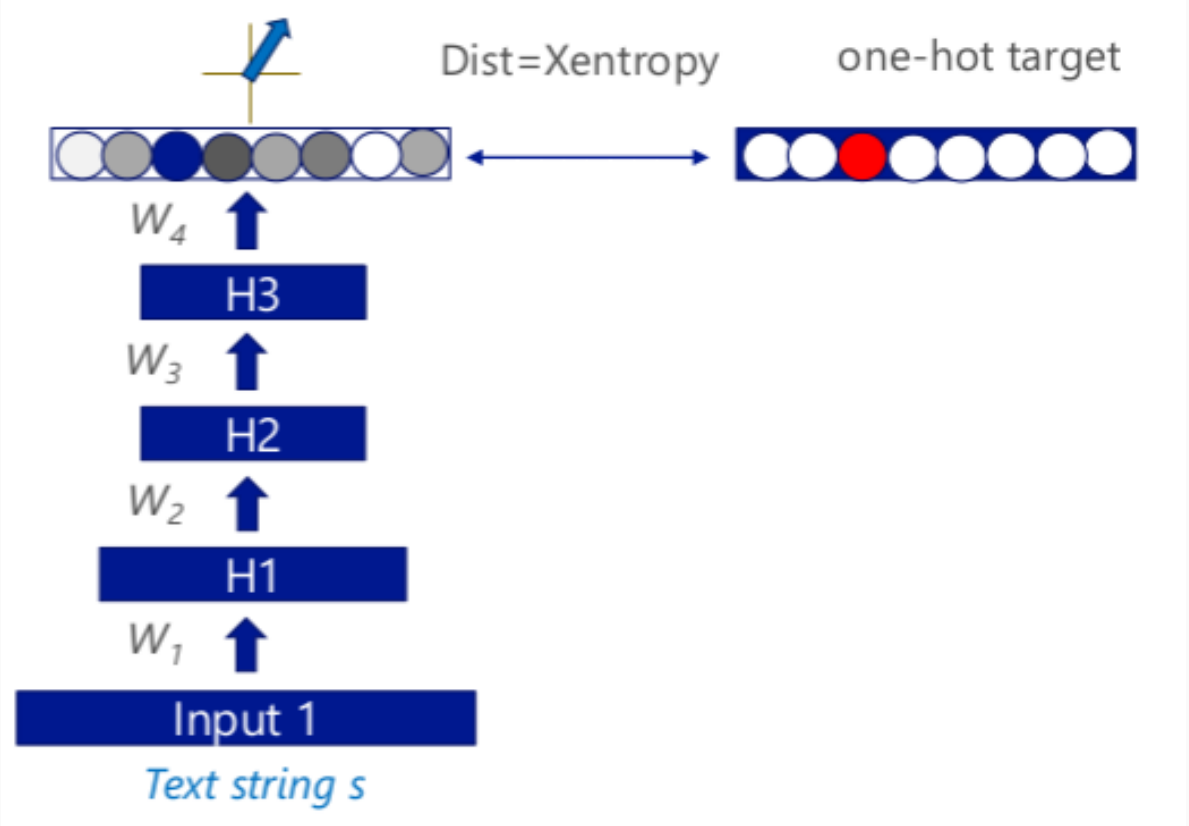
Sergey Nikolenko

Harbour Space University, Barcelona, Spain
January 17, 2018

A GENERAL APPROACH: DSSM

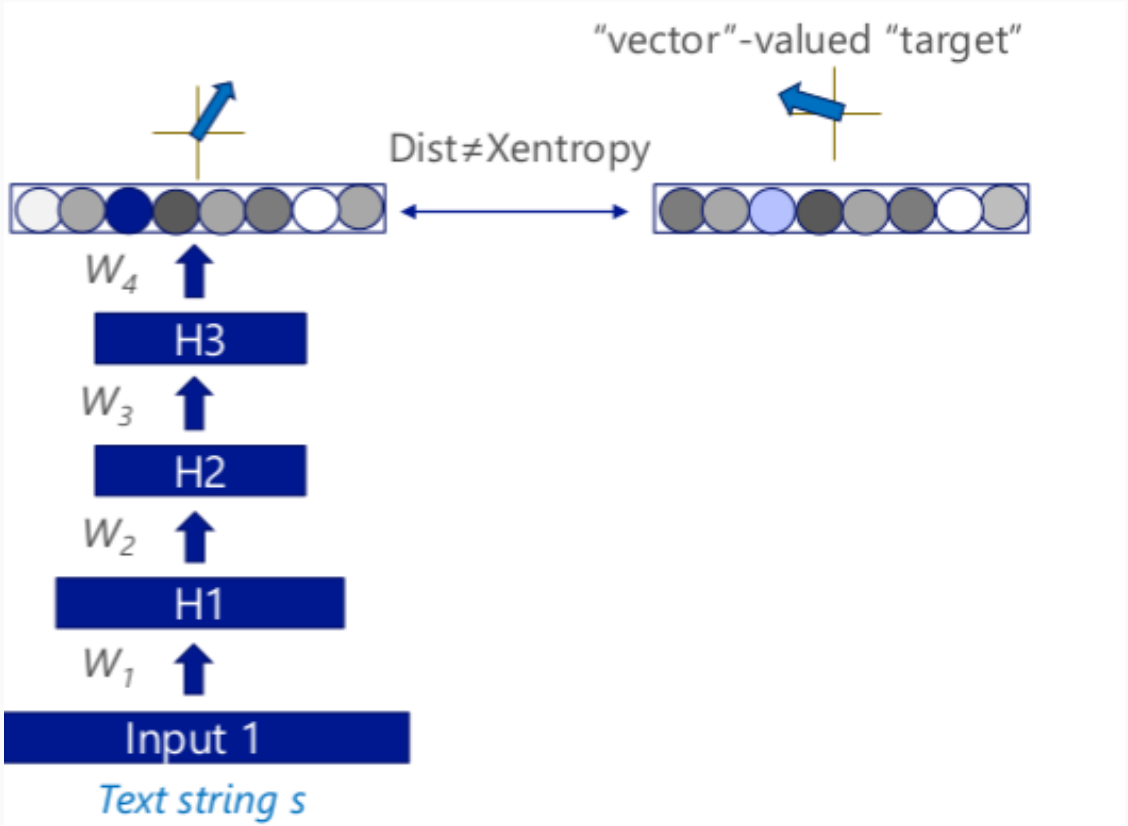
DSSM

- A general approach to NLP based on CNNs is given by *Deep Structured Semantic Model* (DSSM) (Huang et al., 2013; Gao et al., 2014a; 2014b):
  - one-hot target vectors for classification (speech recognition, image recognition, language modeling).



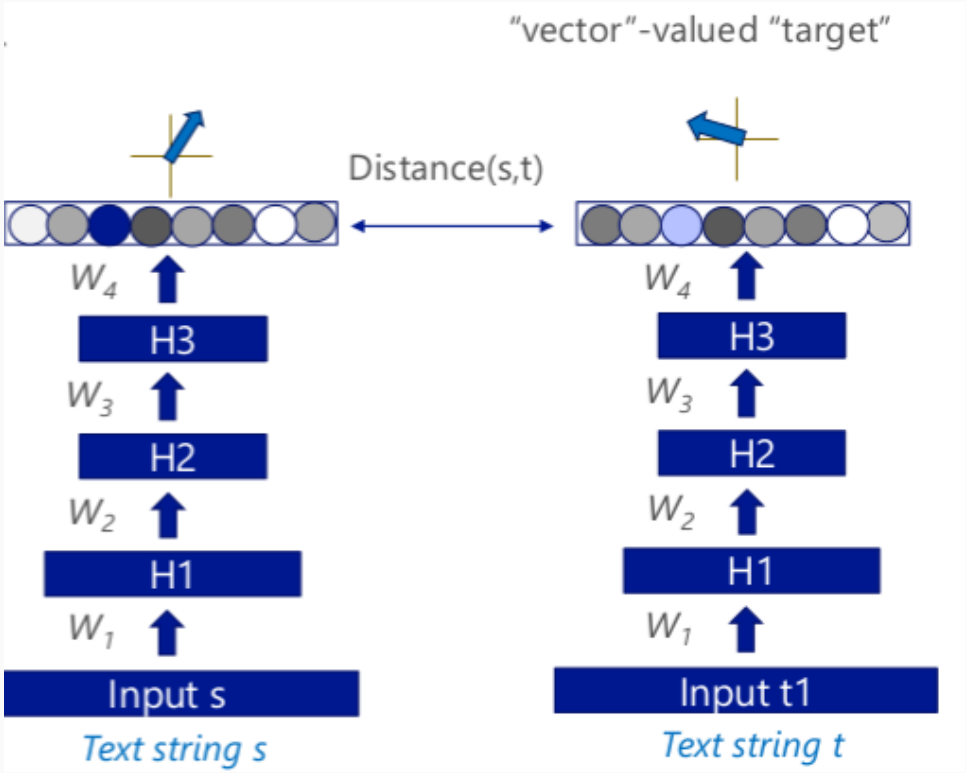
DSSM

- A general approach to NLP based on CNNs is given by *Deep Structured Semantic Model* (DSSM) (Huang et al., 2013; Gao et al., 2014a; 2014b):
  - vector-valued targets for semantic matching.



DSSM

- A general approach to NLP based on CNNs is given by *Deep Structured Semantic Model* (DSSM) (Huang et al., 2013; Gao et al., 2014a; 2014b):
  - can capture different targets (one-hot, vector);
  - to train with vector targets – reflection: bring source and target vectors closer;
  - *siamese networks*.



DSSM

- A general approach to NLP based on CNNs is given by *Deep Structured Semantic Model* (DSSM) (Huang et al., 2013; Gao et al., 2014a; 2014b):
- DSSMs can be applied in a number of different contexts when we can specify a supervised dataset:
  - semantic word embeddings: word by context;
  - web search: web documents by query;
  - question answering: knowledge base relation/entity by pattern;
  - recommendations: interesting documents by read/liked documents;
  - translation: target sentence by source sentence;
  - text/image: labels by images or vice versa.
- Basically, this is an example of a general architecture that can be trained to do almost anything.

DSSM

- A general approach to NLP based on CNNs is given by *Deep Structured Semantic Model* (DSSM) (Huang et al., 2013; Gao et al., 2014a; 2014b):
- *Deep Structured Semantic Models* (DSSM) (Huang et al., 2013; Gao et al., 2014a; 2014b): a deep convolutional architecture trained on similar text pairs.
- Can be used for information retrieval: model relevance by bringing relevant documents closer to their queries (both document and query go through the same convolutional architecture).
- November 2, 2016: a post by Yandex saying that they use (modified) DSSM in their new *Palekh* search algorithm.

LANGUAGE MODELING

LANGUAGE MODELING

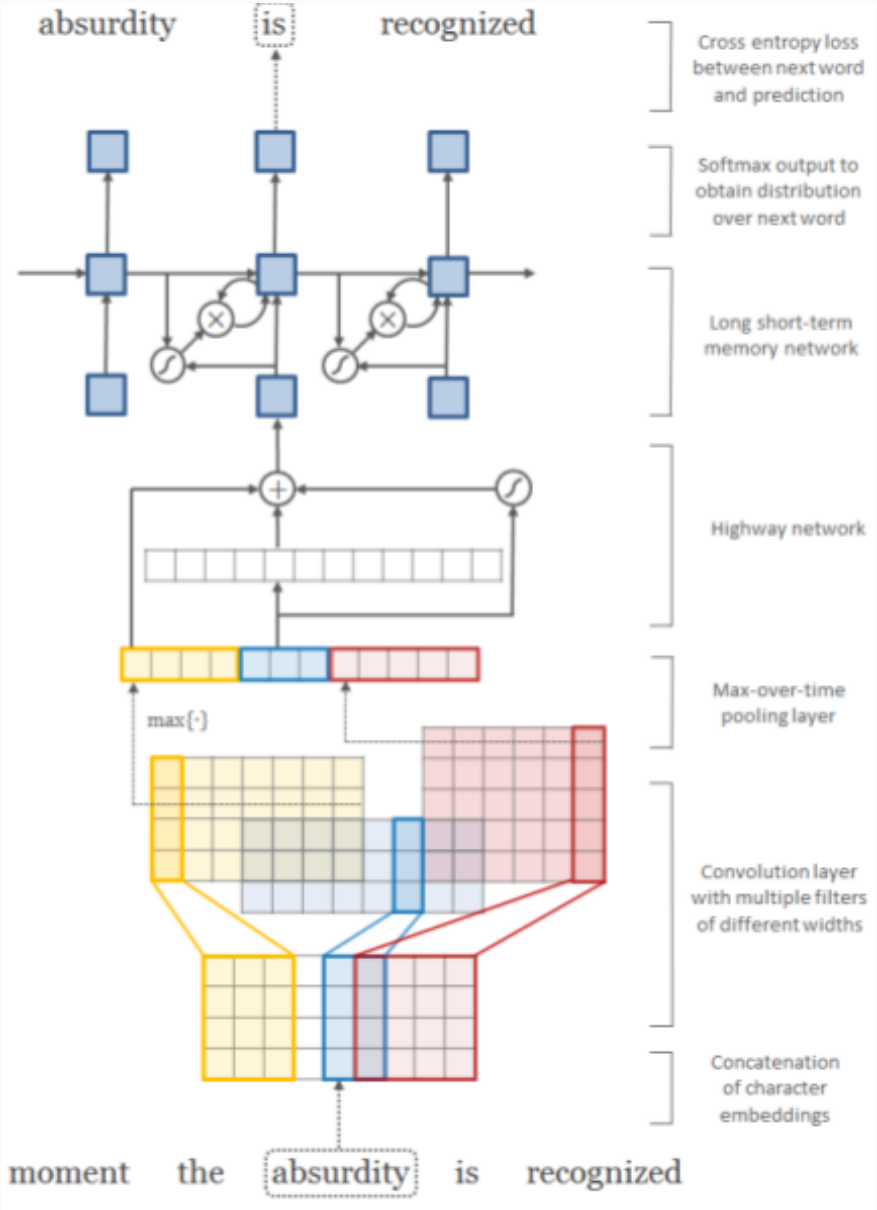
- General problem: predict next word in a text by previous words.
- Classical approaches:
  - count n-grams of words;
  - smoothing, fallback on smaller n-grams;
  - e.g., Kneser-Ney smoothing;
  - computational problems (lots of n-grams).
- But now, of course, it's deep learning all the way down.

TEXT GENERATION WITH RNNS

- Language modeling and text generation is a natural direct application of NN-based NLP; word embeddings started as a "neural probabilistic language model" (Bengio et al., 2003).
- First idea – sequence learning with RNNs/LSTMs.
- Surprisingly, simple RNNs can produce quite reasonably-looking text even by training character by character, with no knowledge of the words ("The Unreasonable Effectiveness..."), including the famous example from (Sutskever et al. 2011):

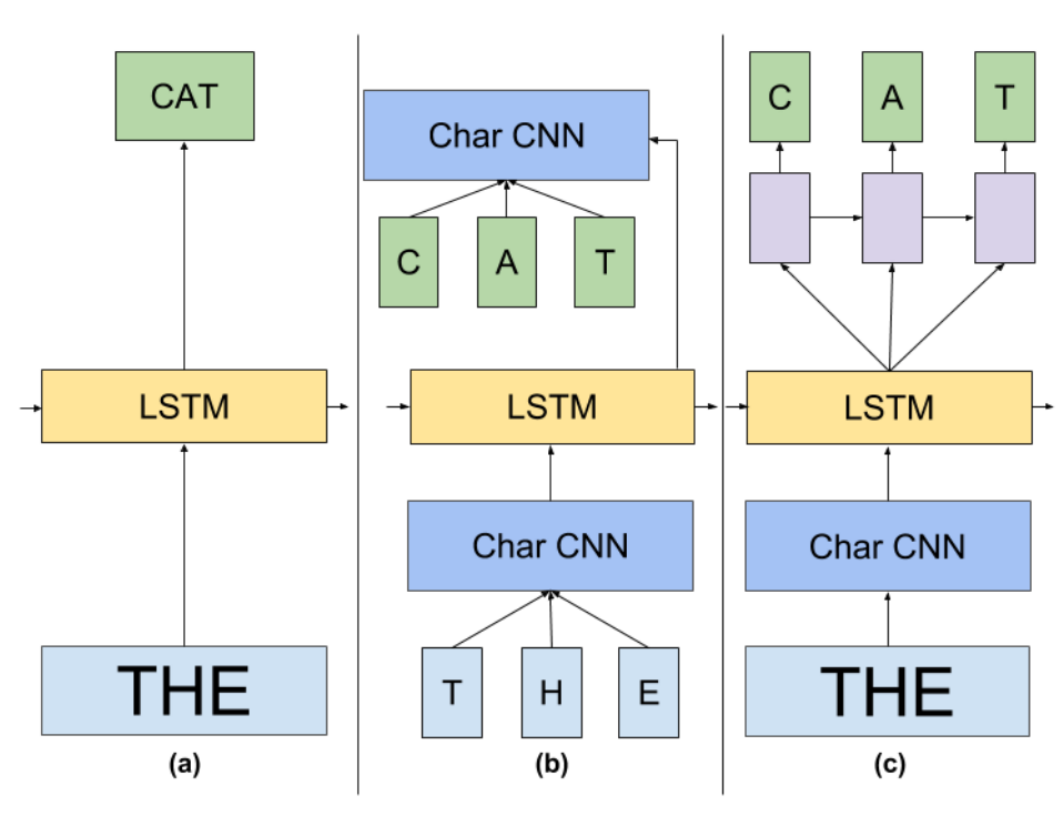
  The meaning of life is the tradition of the ancient human reproduction: it is less favorable to the good boy for when to remove her bigger...
- This is, of course, not "true understanding" (whatever that means), only short-term memory effects.
- We need to go deeper in terms of both representations and sequence modeling.

MODERN CHAR-BASED LANGUAGE MODEL: KIM ET AL., 2015



A MODERN LANGUAGE MODEL

- Jozefowicz et al., 2016: a language model with large LSTMs

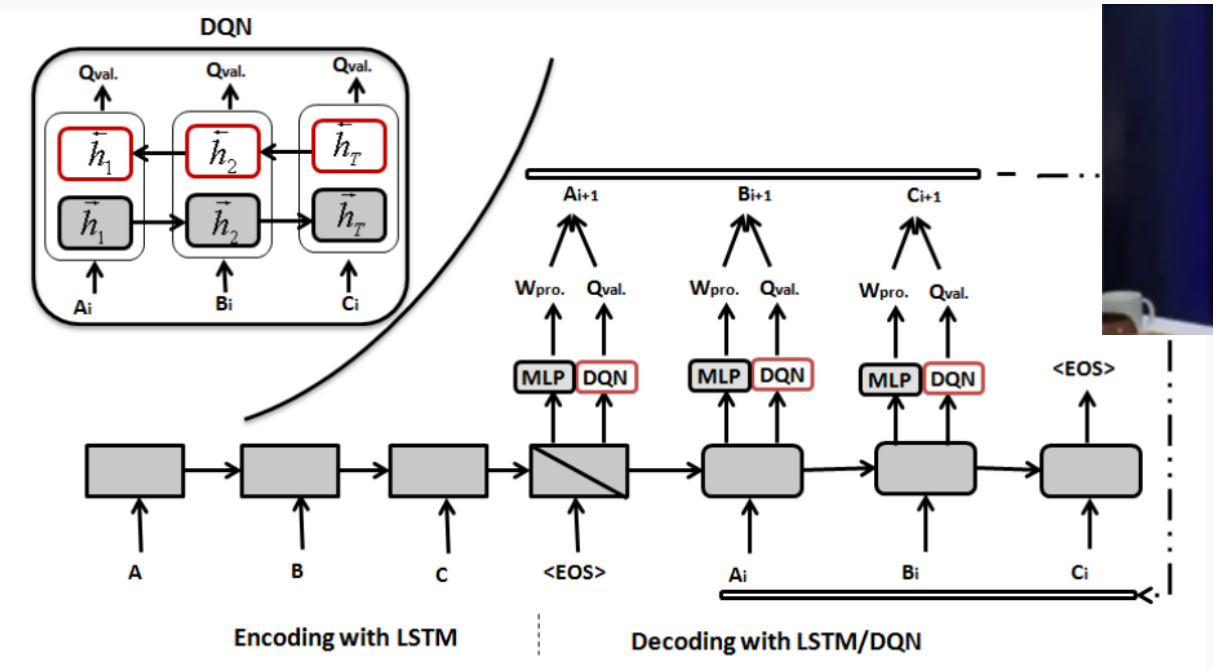


A MODERN LANGUAGE MODEL

- Basically, more data and larger model trumps more complicated models:

DQN

- (Guo, 2015): generating text with deep reinforcement learning.
- Begin with easy parts, then iteratively decode the hard parts with DQN.



- Next, we proceed to specific NLP problems that have led to interesting developments.

THANK YOU!

Thank you for your attention!