Distributed Representations

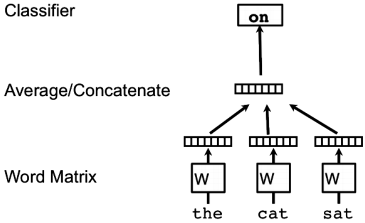
of meaning and knowledge

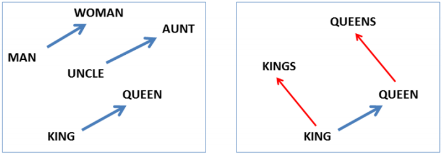
Petr Baudiš ⟨pasky@ucw.cz⟩

May 2015

word2vec

Task: Language model — predict likely and unlikely next words in sentence (useful for speech recognition)
Side-effect: Numerical 300D vector representation for each word
(It seems the distributional semantics hypothesis is true.)





(Mikolov et al., NAACL HLT, 2013)

Distributed Representations

Word vectors encode their semantic meaning.
Multi-word expressions and phrases also carry meaning.

When we think about something,
particular patterns of neurons in our brain activate.
Natural language: Serialization (dump) of our thought vectors designed to evoke similar vectors in similar brains.

Activations in many kinds of neural networks: thought vectors

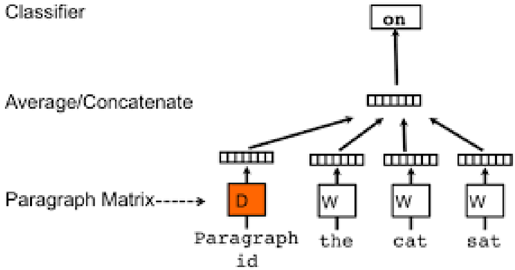
Conjecture: True meanings of things
can be represented by vectors.

Conjecture 2: Thinking about things (reasoning)
may be represented by vector operations.

From Words — Higher! (Flat)

Represent meaning of sentences, paragraphs.

Unsupervised: Context-dependent (like word2vec). To predict the next word, aside of context word vectors add a paragraph word.

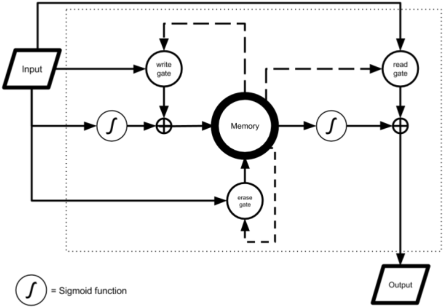


From Words — Higher! (Recurrent)

Build a neural network (RNN) that can learn more complex sentence structure to extract meaning.

RNN: **Recurrent** neural network; reads sentence word by word and learns which words affect the meaning in what way.
It has **memory**.

Long short-term memory (LSTM): Popular type of RNN.

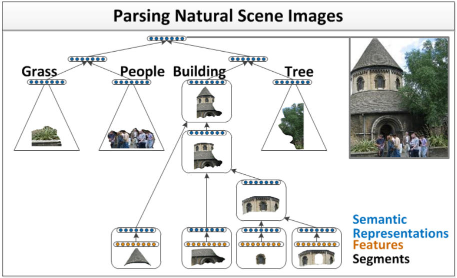


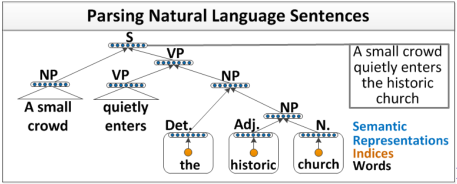
From Words — Higher! (Recursive)

Build a neural network (RNN) that can learn more complex sentence structure to extract meaning.

RNN: **Recursive** neural network; walks the parse tree of a sentence and gradually builds up represnetations.

Parsing Natural Scenes and Natural Language with Recursive Neural Networks, Socher et al., 2011





Applications

Document classification (sentiment analysis).

Question answering.

Automatic image captions.

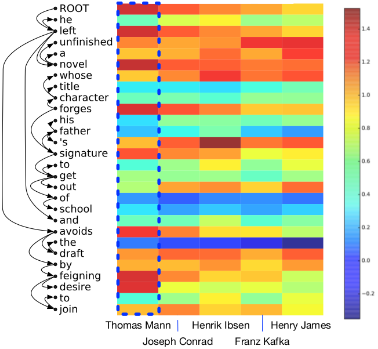
Machine translation.

QANTA

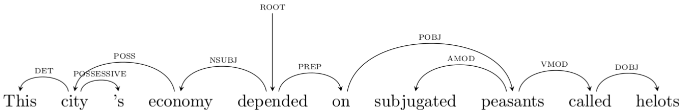
Iyyer et al., 2014: A Neural Network for Factoid Question Answering over Paragraphs

A question answering neural network with trans-sentential averaging.

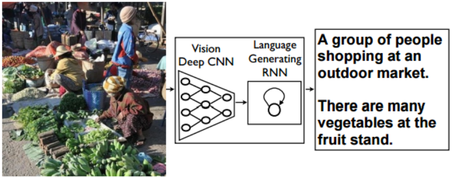
Later in its existence, this polity's leader was chosen by a group that included three bishops and six laymen, up from the seven who traditionally made the decision. Free imperial cities in this polity included Basel and Speyer. Dissolved in 1806, its key events included the Investiture Controversy and the Golden Bull of 1356. Led by Charles V, Frederick Barbarossa, and Otto I, for 10 points, name this polity, which ruled most of what is now Germany through the Middle Ages and rarely ruled its titular city.



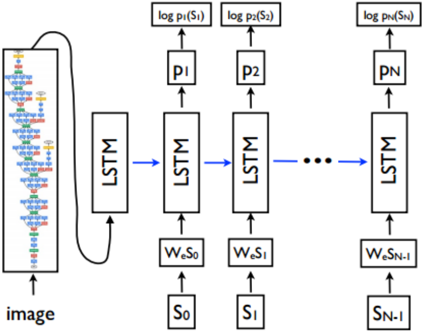
DT-RNN, sentence level distributed representations.



Show and Tell: Automatic Image Captions



[1]



---

[1] Show and tell: A neural image caption generator, Vinyals et al. 2014

Demo Time!

Demo 1: LISA machine translation.[2]
BiRNN encoder, RNN decoder with alignment model.

Demo 2: DefGen question answering (word by definition).[3]
RNN definition encoder, final internal state to be the word vector.

Demo 3 (offtopic): YodaQA — live.ailao.eu (or google YodaQA)

[2] Neural Machine Translation by Jointly Learning to Align and Translate, Bahdanau et al. 2015

[3] Learning to Understand Phrases by Embedding the Dictionary, Hill et al. 2015

Conclusion

- We are rapidly developing the ability to create and manipulate **artificial thought vectors**
- What next: Complex thought systems, temporal events and processes.
- Left out: Learning relations, reasoning.
- Left out: Memory Networks

`pasky@ucw.cz`
`petr.baudis@gmail.com`

Thank you for your attention!

If you have good ideas, good data and fast computers, you can do almost anything. – Geoffrey Hinton