VSB TECHNICAL UNIVERSITY OF OSTRAVA | FACULTY OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE | DEPARTMENT OF COMPUTER SCIENCE

Algorithms for Big Data

Language modeling

Jan Platoš

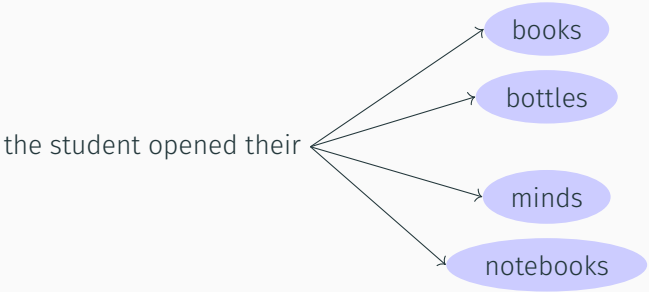
November 6, 2020

Department of Computer Science
Faculty of Electrical Engineering and Computer Science
VŠB - Technical University of Ostrava

Language modeling

Language modeling

Language modeling is a task of predicting what word comes next.



Language modeling

Language modeling is a task of predicting what word comes next.

- Given a sequence of words $x_1, x_2, \ldots, x_t$, compute the probability distribution of the next word x_{t+1}:

$$P(x_{t+1} = w_j | x_t, \ldots, x_1)$$

- Where w_j is a word in the vocabulary $V = \{w_1, \ldots, w_{|V|}\}$.

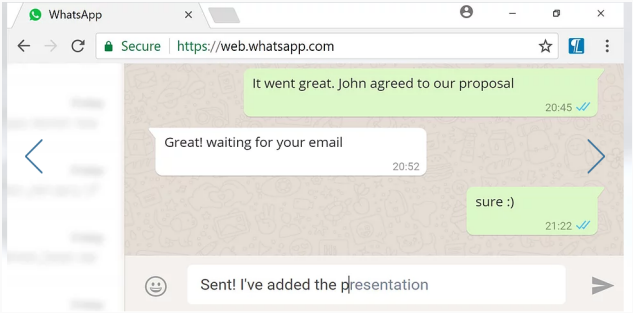
Language modeling

Language modeling is a task of predicting what word comes next.

- Given a sequence of words $x_1, x_2, \ldots, x_t$, compute the probability distribution of the next word x_{t+1}:

$$P(x_{t+1} = w_j | x_t, \ldots, x_1)$$

- Where w_j is a word in the vocabulary $V = \{w_1, \ldots, w_{|V|}\}$.



Language modeling

Language modeling is a task of predicting what word comes next.

- Given a sequence of words $x_1, x_2, \ldots, x_t$, compute the probability distribution of the next word x_{t+1}:

$$P(x_{t+1} = w_j | x_t, \ldots, x_1)$$

- Where w_j is a word in the vocabulary $V = \{w_1, \ldots, w_{|V|}\}$.

n-Gram Language model

- An **n-gram** is a chunk of n consecutive words:
  - **uni**grams: "the", "students", "opened", "their"
  - **bi**grams: "the students", "students opened", "opened their"
  - **tri**grams: "the students opened", "students opened their"
  - **4-**grams: "the students opened their"
- Idea is to collect a statistics about how frequently different n-grams are and use them to predict next word.
- We assume that a word x_{t+1} depends only on the preceding $(n-1)$ words.

$$P(x_{t+1} = w_j | x_t, \ldots, x_{t-n+2}) = \frac{P(x_{t+1}, x_t, \ldots, x_{t-n+2})}{P(x_t, \ldots, x_{t-n+2})}$$

- The values may be computed from the corpora.

n-Gram Language model

The n-Gram language model may be used to generate text.

today the ...

n-Gram Language model

The n-Gram language model may be used to generate text.

today the price ...

n-Gram Language model

The n-Gram language model may be used to generate text.

today **the price** ...

n-Gram Language model

The n-Gram language model may be used to generate text.

today **the price** of ...

n-Gram Language model

The n-Gram language model may be used to generate text.

today the **price of** ...

n-Gram Language model

The n-Gram language model may be used to generate text.

today the **price of** gold ...

n-Gram Language model

The n-Gram language model may be used to generate text.

today the price of gold per ton , while production of shoe lasts and shoe industry , the bank intervened just after it considered and rejected an imf demand to rebuild depleted european stocks , sept 30 end primary 76 cts a share

n-Gram Language model

The n-Gram language model may be used to generate text.

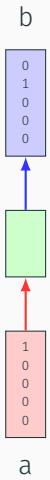
today the price of gold per ton , while production of shoe lasts and shoe industry , the bank intervened just after it considered and rejected an imf demand to rebuild depleted european stocks , sept 30 end primary 76 cts a share

- The result is incoherent. More than two words need to be taken into account!!!
- The increasing of *n* leads to the sparsity problem and increase the model size.
- Sparsity problem - the sequence never appears in the data.

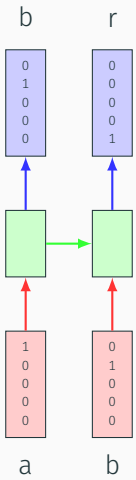
Neural Language Model

- The task:
  - Input: sequence of tokens: $x_1, \ldots, x_t$
  - Output: Probability of next token $P(x_{t+1} = w_j | x_t, \ldots, x_1)$
- A window approach may work similarly as for n-grams.
  1. Input is one-hot-vectors
  2. Compute token embedding for each token and concatenate as input.
  3. Define a hidden layer.
  4. Set output as **softmax** function over the hidden layer.
- This solves the problem of sparsity and reduces the size of the model to linear.
- Some problems remain:
  - The fixed window limits the precision and is never large enough.
  - The weights are not shared between words in a window.

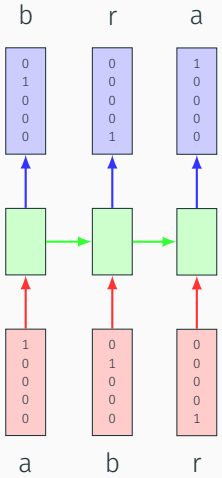
Neural Language Model



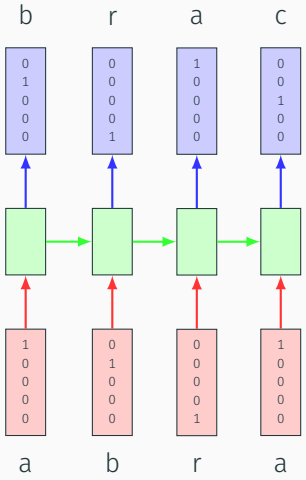
Neural Language Model



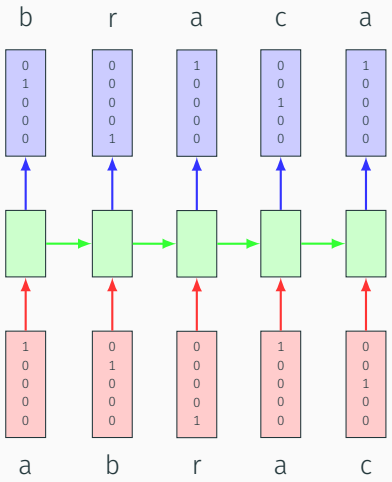
Neural Language Model



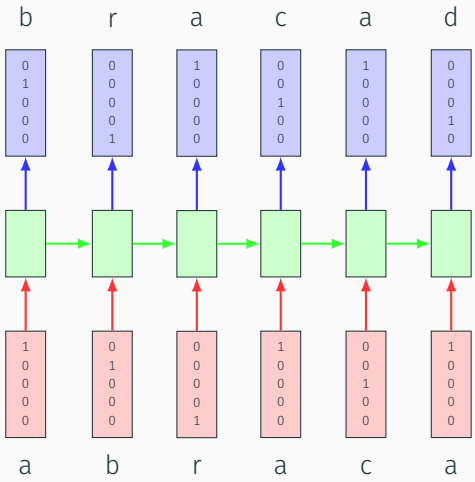
Neural Language Model



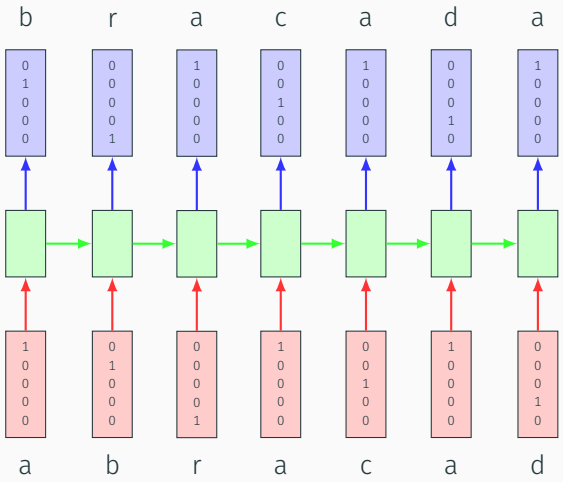
Neural Language Model



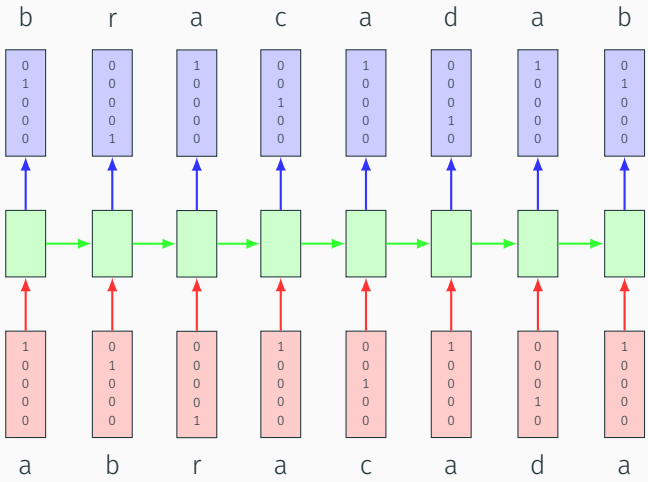
Neural Language Model



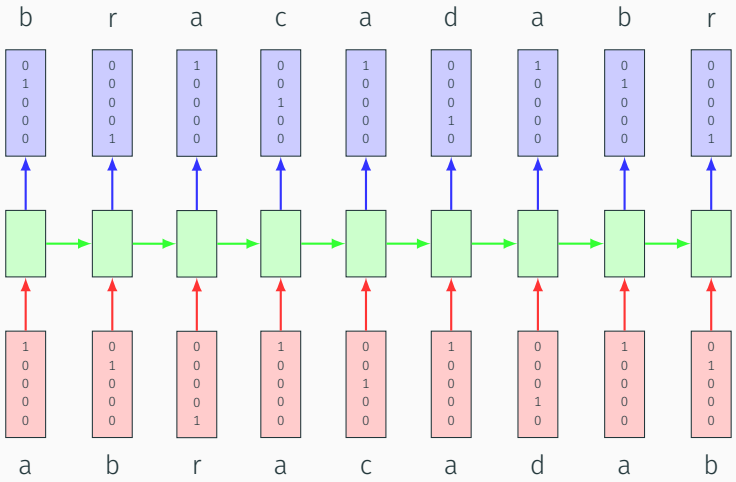
Neural Language Model



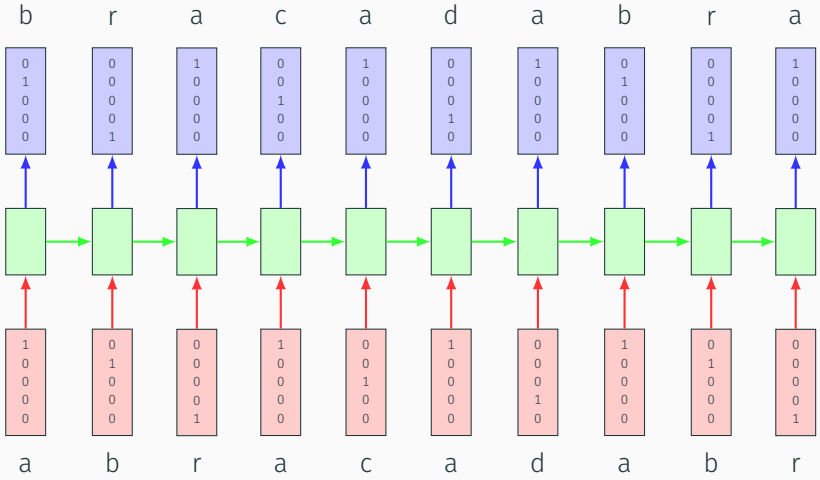
Neural Language Model



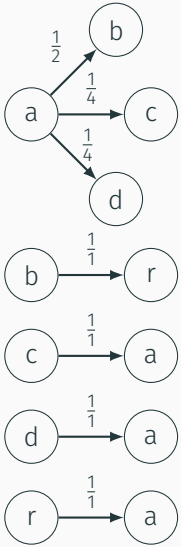
Neural Language Model



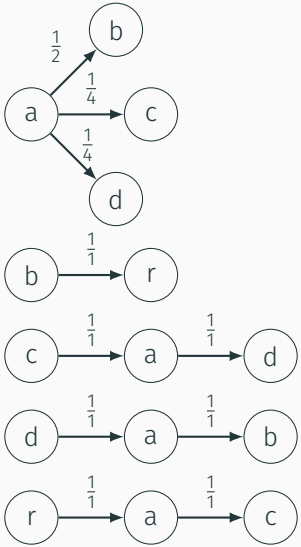
Neural Language Model



Neural Language Model



Neural Language Model



Neural Network Model

- As was seen in the example, even the simple text needs context information.
- Simple dense network is not able to deal with text in a token-by-token manner.
- Network with memory may deal with context information well.
- LSTM layers are used frequently with text data.
- Characters or Words may be used to model the text.
- Many neurons/layers need to be used to create a large enough capacity for text modeling.
- Training, a good model, takes much time.

Recurrent Neural Network (RNN) - Char. Examples

RNN as a political speech writer (input phrase Jobs)[1]

Good afternoon. God bless you.
The United States will step up to the cost of a new challenges of the American people that will share the fact that we created the problem. They were attacked and so that they have to say that all the task of the final days of war that I will not be able to get this done. The promise of the men and women who were still going to take out the fact that the American people have fought to make sure that they have to be able to protect our part. ...

[1] https://medium.com/@samim/obama-rnn-machine-generated-political-speeches-c8abd18a2ea0

Recurrent Neural Network (RNN) - Char. Examples

LSTM as a novelist[2]

”The Malfoys!” said Hermione.
Harry was watching him. He looked like Madame Maxime.
When she strode up the wrong staircase to visit himself.
”I’m afraif I’ve definitely been suspended from power, no chance - indeed?” said Snape. He put his head back behind them and read groups as they crossed a corner and fluttered down onto their ink lamp, and picked up his spoon. The doorbell rang. It was a lot cleaner down in London...

[2] https://medium.com/deep-writing/harry-potter-written-by-artificial-intelligence-8a9431803da6

Questions?