Analyzing Software using Deep Learning

Sequence-to-Sequence Networks and their Applications (Part 1)

Prof. Dr. Michael Pradel

Software Lab, University of Stuttgart

Summer 2020

Overview

- **Sequence-to-sequence networks**
- **API usage sequences for natural language queries**

  Based on ”Deep API learning” by Gu et al., 2016
- **Interpreting Python programs**

  Based on ”Learning to execute” by Zaremba and Sutskever, 2014

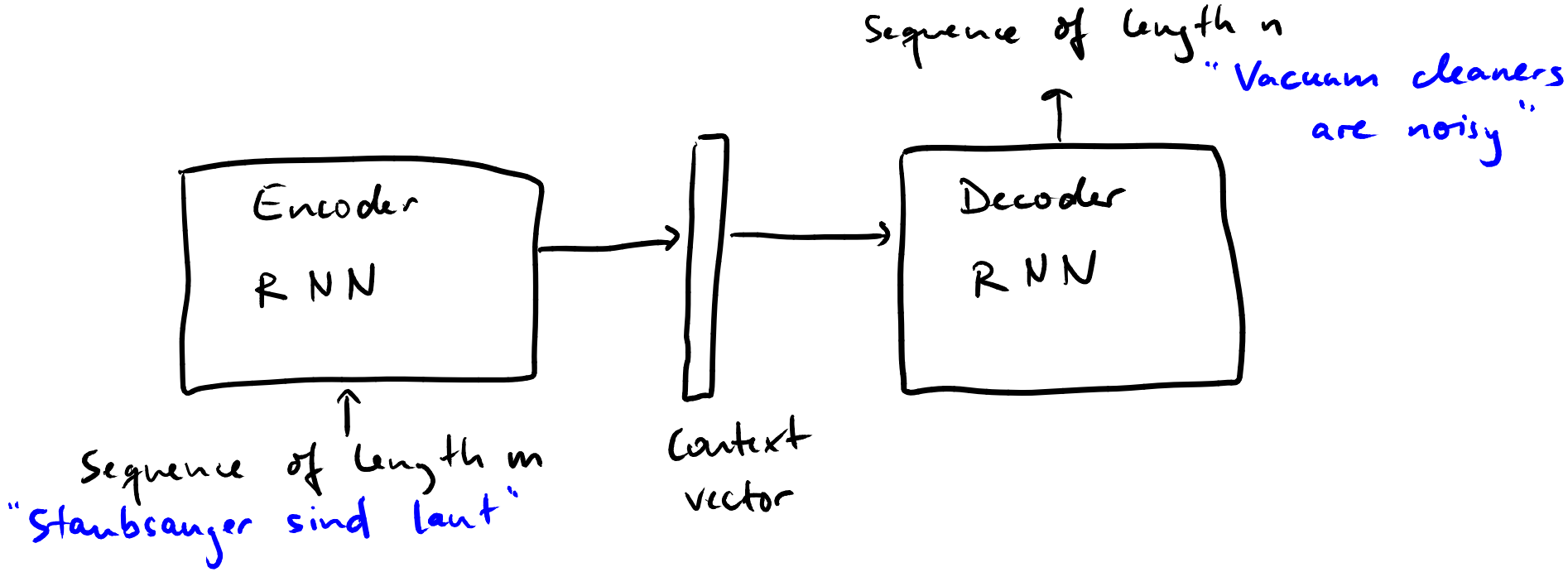
Sequence-to-Sequence

Goal: Translate sequence of items into another sequence of items

Various applications

- Translation between natural languages
- Generate image captions
- Summarize videos into text
- Answer natural language questions

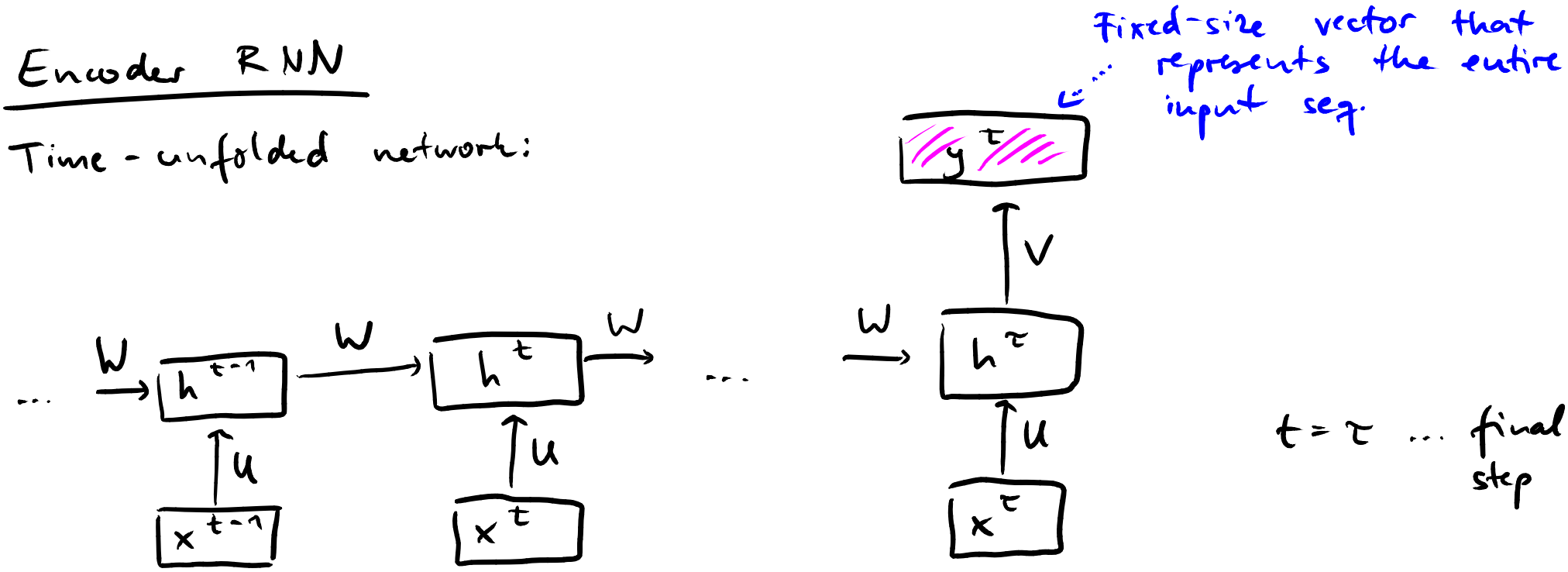
Overview to Sequence-to-Sequence Architecture



- m may be different from n
- both networks are trained jointly
- context vector summarizes input in a way suitable to generate output

Encoder RNN

Time-unfolded network:

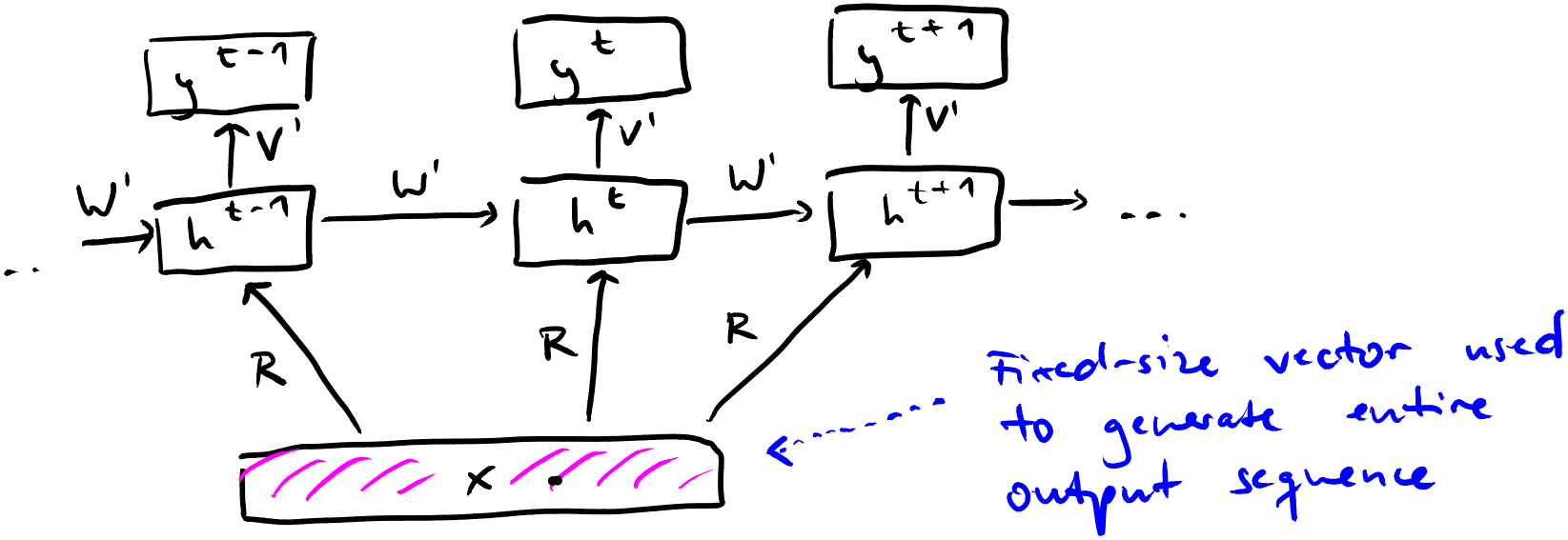


$$h^t = \tanh(W \cdot h^{t-1} + U \cdot x^t + b)$$

or any other activation fct.

$$y^\tau = V \cdot h^\tau + c$$

Decoder RNN



$$h^t = \tanh\left(W' \cdot h^{t-1} + R \cdot x + b'\right)$$

$$y^t = \mathrm{softmax}\left(V' \cdot h^t + c'\right)$$

Training

Training data: N pairs of sequences (x_i, y_i) for $i = 1, \dots, N$

↳ End of sequence marked with <EOS>

Example:

x_1 = Staubsauger, sind, laut, <EOS>

y_1 = Vacuum, cleaners, are, noisy, <EOS>

Goal of training:

Minimize $$\frac{1}{N} \sum_{i=1}^{N} \cdot \sum_{t=1}^{T} -\log \Pr(y_{it} \mid x_i)$$

where T .. length of output sequences

$\Pr(y_{it} \mid x_i)$ .. probab. of word y_{it} given input sequence x_i

Translation

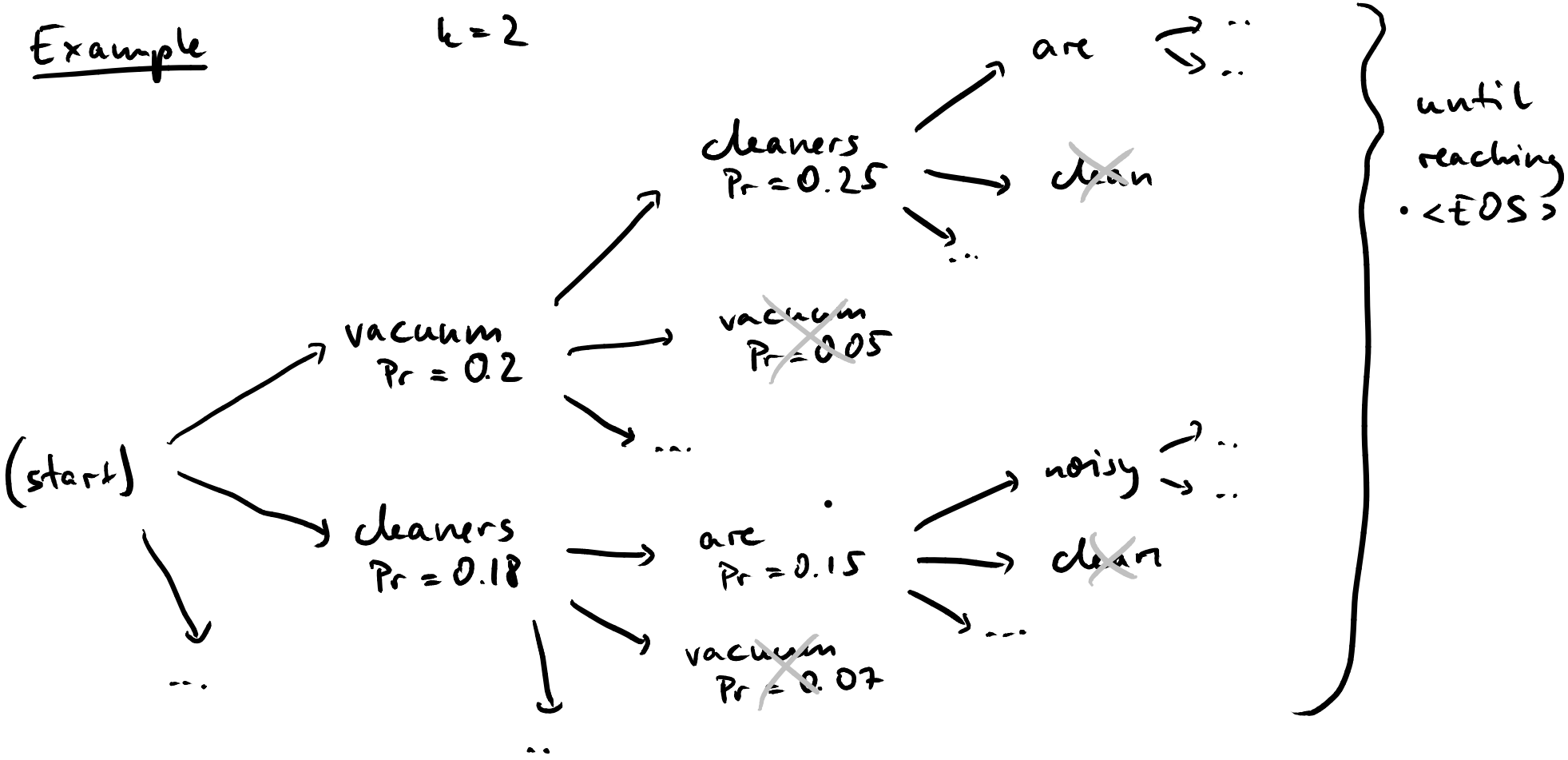
For many applications, want k most likely translations

Use left-to-right beam search

- For every word, consider k most likely alternatives
- Extend partial sentence in k ways
- After each time step, keep only k most likely partial sequences

<u>Example</u>

$k = 2$



Quiz

Which of following sentences is correct (multiple sentences may be correct)?

- The context vector is a potential bottleneck that may prevent the network from effective learning.
- The length of the input sequence must be the same across all instances of the training set.
- The length of the output sequence must be the same across all instances of the training set.
- Each instance in the training set must contain two sequences (input and output).

Quiz

Which of following sentences is correct (multiple sentences may be correct)?

- The context vector is a potential bottleneck that may prevent the network from effective learning.
- The length of the input sequence must be the same across all instances of the training set.
- The length of the output sequence must be the same across all instances of the training set.
- Each instance in the training set must contain two sequences (input and output).