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

Sequence of length n

"Vacuum cleaners are noisy."

Sequence of length m

"Staubsauger sind laut."

- 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:

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

\[ y^t = V \cdot h^t + c \]

\( t = \tau \) ... final step

Fixed-size vector that represents the entire input seq.
Decoder RNN

\[ h^t = \tanh(W' \cdot h^{t-1} + R \cdot x + b') \]

\[ y^t = \text{softmax}(V' \cdot h^t + c') \]

Fixed-size vector used to generate entire output sequence
Training

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

\( <\text{EOS}> \) End of sequence marked with <EOS>

Example:

\( x_1 = \text{Staubsauger, sinn, laut} \), \( y_1 = \text{Vacuum, cleaners, arc, noisy} \), <EOS>

Goal of training:

\[
\text{Minimize } \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} - \log \Pr(y_{it} | x_i)
\]

where \( T \) = length of output sequences

\( \Pr (y_{it} | x_i) \) = probab. of word \( y_{it} \) given input sequence \( x_i \)
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
Example

\( k = 2 \)

- Vacuum, \( P_r = 0.2 \)
- Cleaners, \( P_r = 0.18 \)
- ... (More actions)

- Vacuum, \( P_r = 0.05 \)
- Are, \( P_r = 0.15 \)
- Noisy, ... (Actions)

- Are, ... (Actions)
- Vacuum, \( P_r = 0.07 \)
- Clean, ... (Actions)

... (Continuation)

until reaching \(<\text{EOS}>\)
Quiz

Which of the 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).