Deep Learning for Natural Language Processing

Autoregressive Sequence Models

UNIVERSITY OF GOTHENBURG

CHALMERS

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

Richard Johansson

richard.johansson@gu.se

structured prediction: basic terminology

- sequence labeling is a **structured prediction** task
- input: a sequence $\boldsymbol{x}$
- output: a sequence $\boldsymbol{y}$ of the same length as $\boldsymbol{x}$

$$\begin{array}{ccccc} w_1 & w_2 & w_3 & \cdots & w_n \\ \downarrow & \downarrow & \downarrow & & \downarrow \\ l_1 & l_2 & l_3 & & l_n \end{array}$$

Algorithmic approaches

- **Exhaustive search**

  Cast structured prediction as a combinatorial optimisation problem over the set of target representations.

  Viterbi algorithm, Eisner algorithm

- **Greedy search**

  Cast structured prediction as a sequence of classification problems: at each point in time, predict one of several options.

  window-based part-of-speech tagging, arc-standard algorithm

Algorithmic approaches

- **Exhaustive search**

  Cast structured prediction as a combinatorial optimisation problem over the set of target representations.

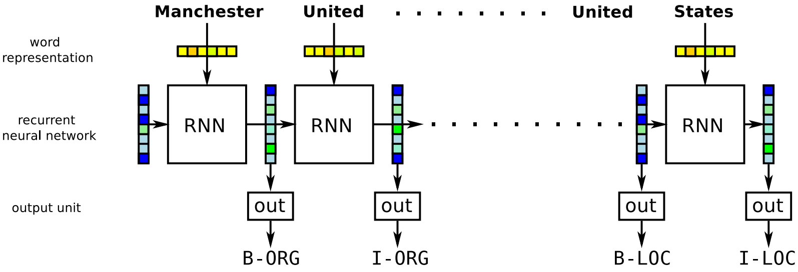
  Viterbi algorithm, Eisner algorithm

- **Greedy search**

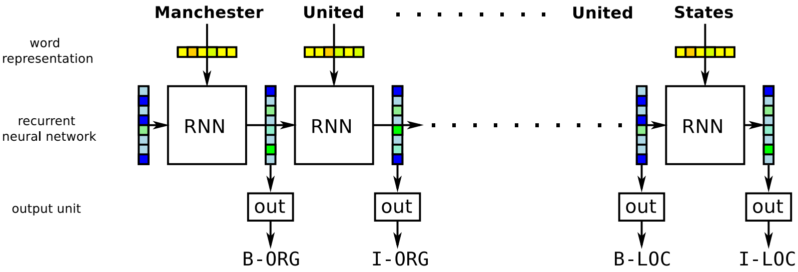
  Cast structured prediction as a sequence of classification problems: at each point in time, predict one of several options.

  window-based part-of-speech tagging, arc-standard algorithm

RNN-based sequence labeling



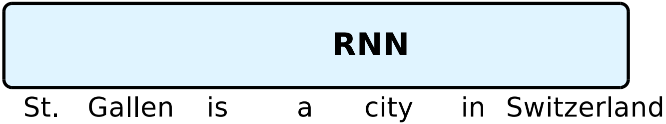
a limitation of our current model



- our output decisions don't affect each other
- can we model the **interdependency** between labels?
  - for instance, that B-LOC+I-LOC is good
  - but B-LOC+I-ORG is bad

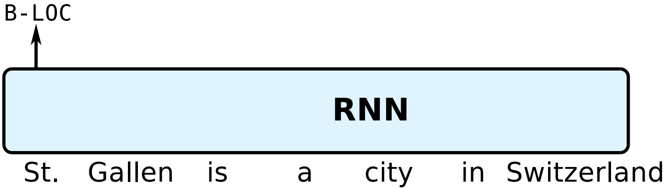
autoregressive sequence models

- an **autoregressive** sequence model is one where each prediction depends on previous outputs



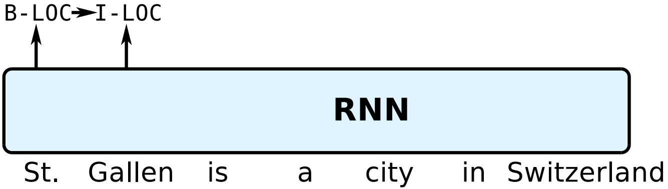
autoregressive sequence models

- an **autoregressive** sequence model is one where each prediction depends on previous outputs



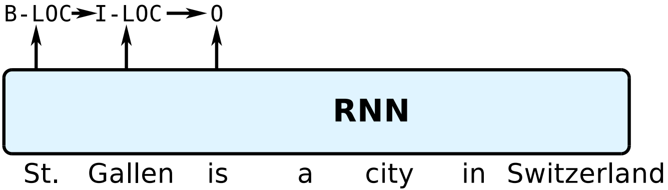
autoregressive sequence models

- an **autoregressive** sequence model is one where each prediction depends on previous outputs



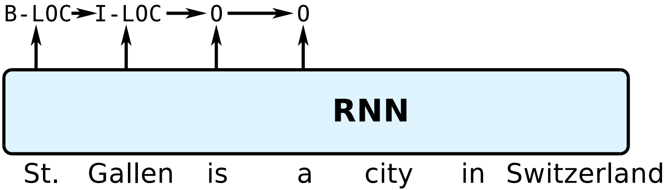
autoregressive sequence models

- an **autoregressive** sequence model is one where each prediction depends on previous outputs



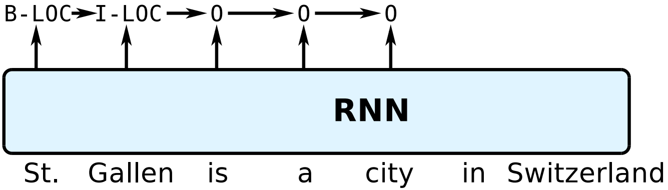
autoregressive sequence models

- an **autoregressive** sequence model is one where each prediction depends on previous outputs



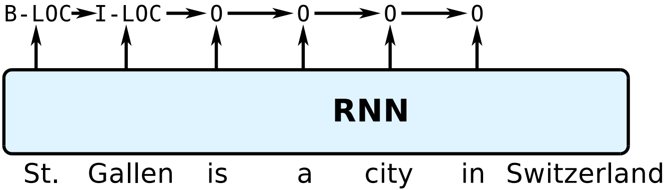
autoregressive sequence models

- an **autoregressive** sequence model is one where each prediction depends on previous outputs



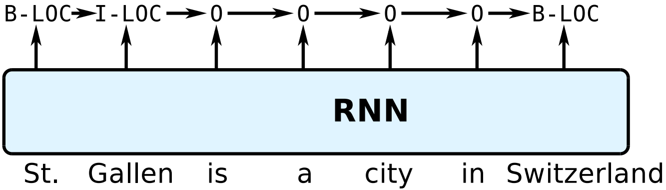
autoregressive sequence models

- an **autoregressive** sequence model is one where each prediction depends on previous outputs



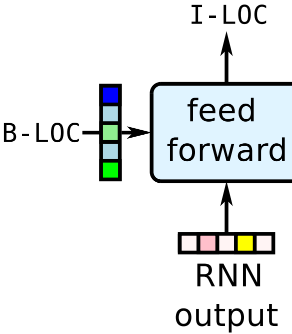
autoregressive sequence models

- an **autoregressive** sequence model is one where each prediction depends on previous outputs



implementation of autoregressive sequence models

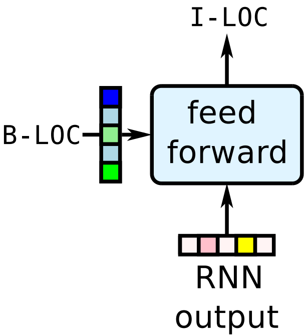
- example of a model that depends on the previous output



- if the prediction model is an RNN, it depends on the full history

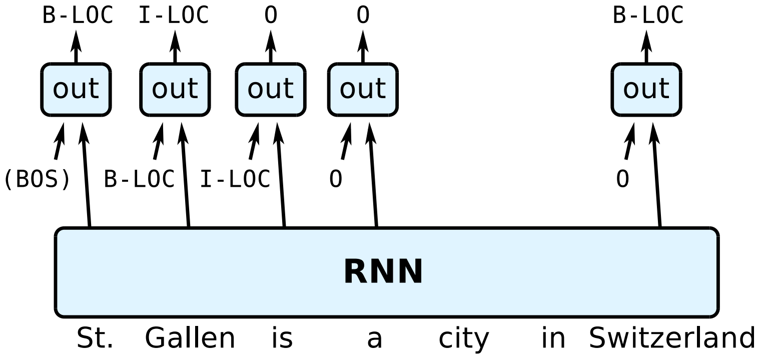
training autoregressive sequence models

- how do we train a model that depends on its own predictions?



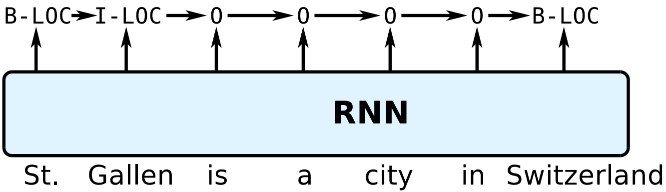
- the classical solution is to use the **gold-standard** label
- this idea is called **teacher forcing**

implementing teacher forcing



after training: running the system

- at prediction time, we run the system incrementally
- in this case, the previous label is a **predicted** label



limitations of teacher forcing

- training-time and prediction time data distributions are different
  - at prediction time, some of the previous labels will be incorrect
- if we make a mistake, the system might be in a situation it has never seen before!
  - this is called **exposure bias**
- risk of compounding errors

limitations of autoregressive models

- the predictions are influenced by past predictions but not by future predictions

Paris	**Hilton**	**is**	**a**	**media**	**celebrity**
B-LOC	I-PER	O	O	O	O

- because the prediction algorithm is **greedy**, the model can't change its mind!
- in the next lecture, we will see a non-greedy approach

exercise 2

- we will continue our NER experiments

Manchester	United	will	return	to	the	United	States
B-ORG	I-ORG	O	O	O	O	B-LOC	I-LOC

- we will investigate autoregressive models and **conditional random fields** (next lecture)