Deep Learning Theory and Applications

RNNs for Text and Language

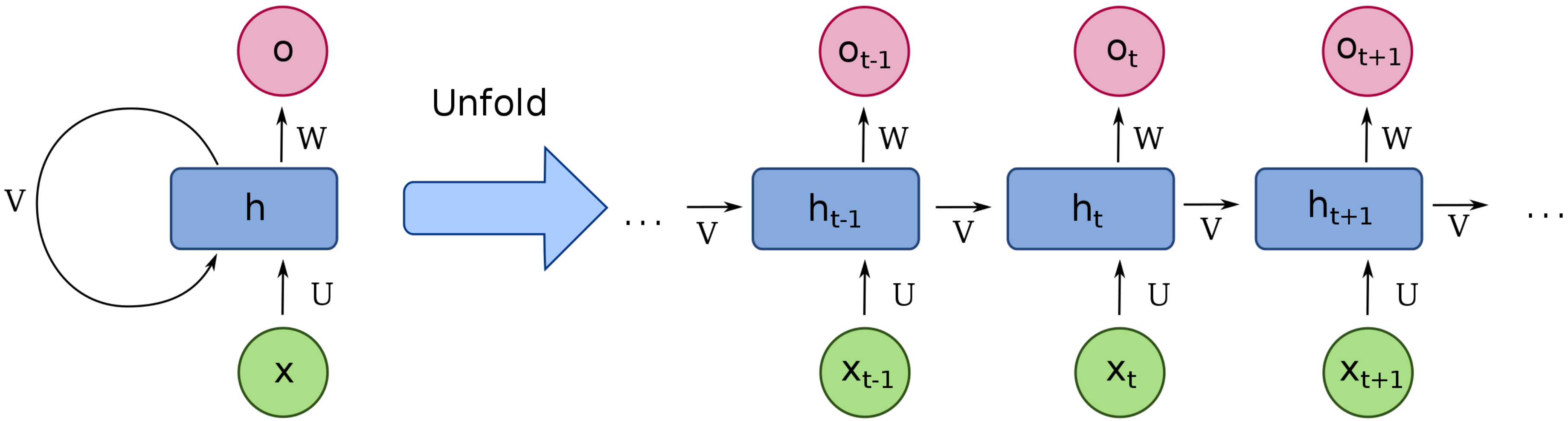
CPSC/AMTH 663

Yale

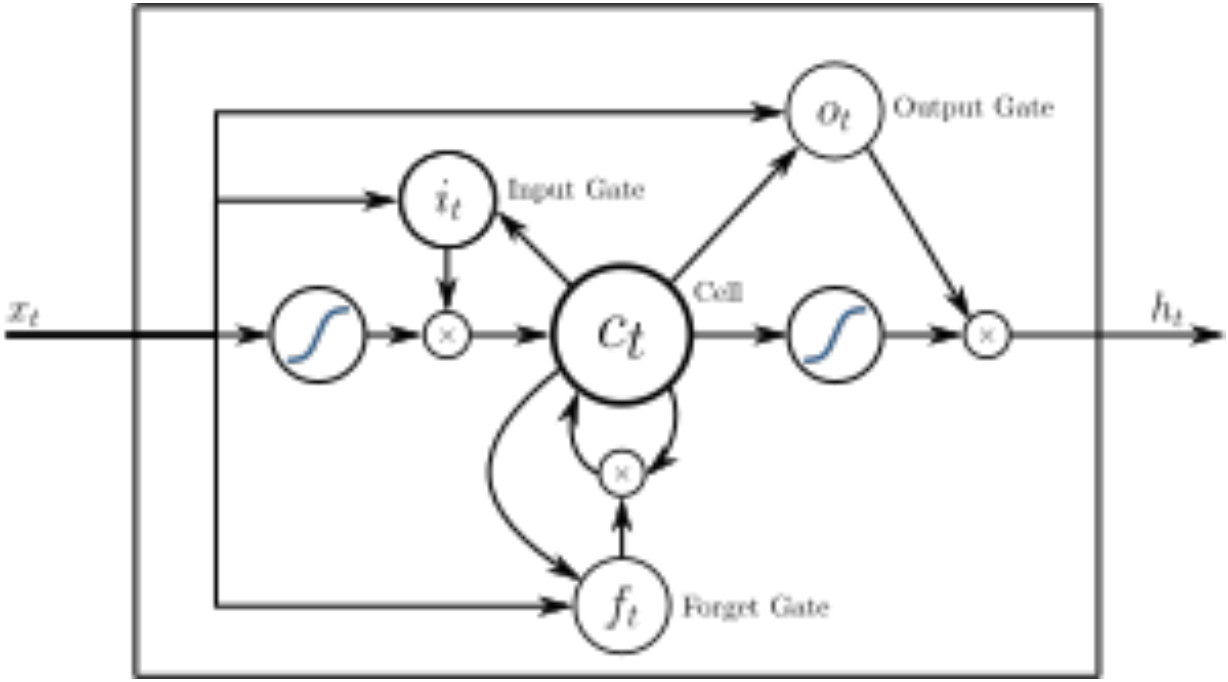
Outline

1. NLP
2. Machine Translation
3. Sentiment Analysis
4. Variable Length Inputs
5. Attention
6. Image captioning

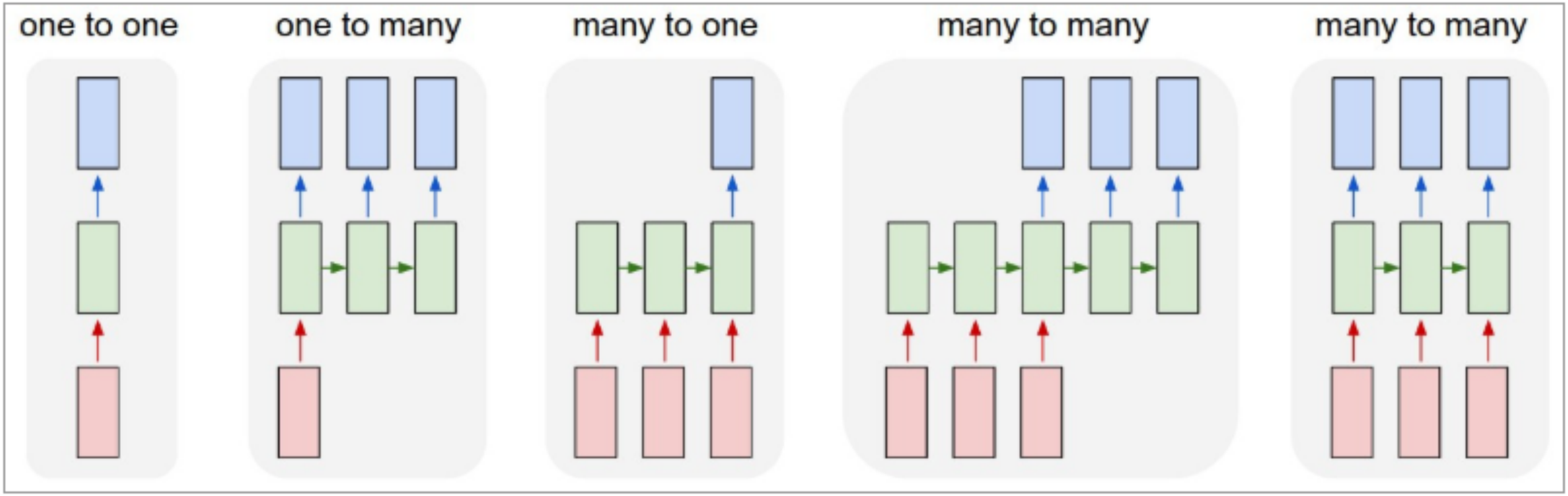
RNN



LSTM



RNNs on variable length Inputs

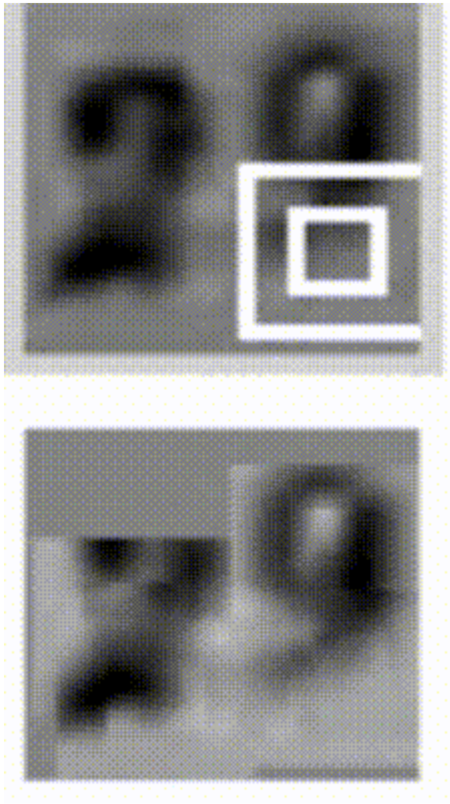


Variable length inputs, variable length outputs, variable length computation

If training vanilla neural nets is optimization over functions,

training recurrent nets is optimization over programs.

Non-sequential data



RNNs can still process them by scanning over the data

For images scanning over pixels

Natural Language Processing

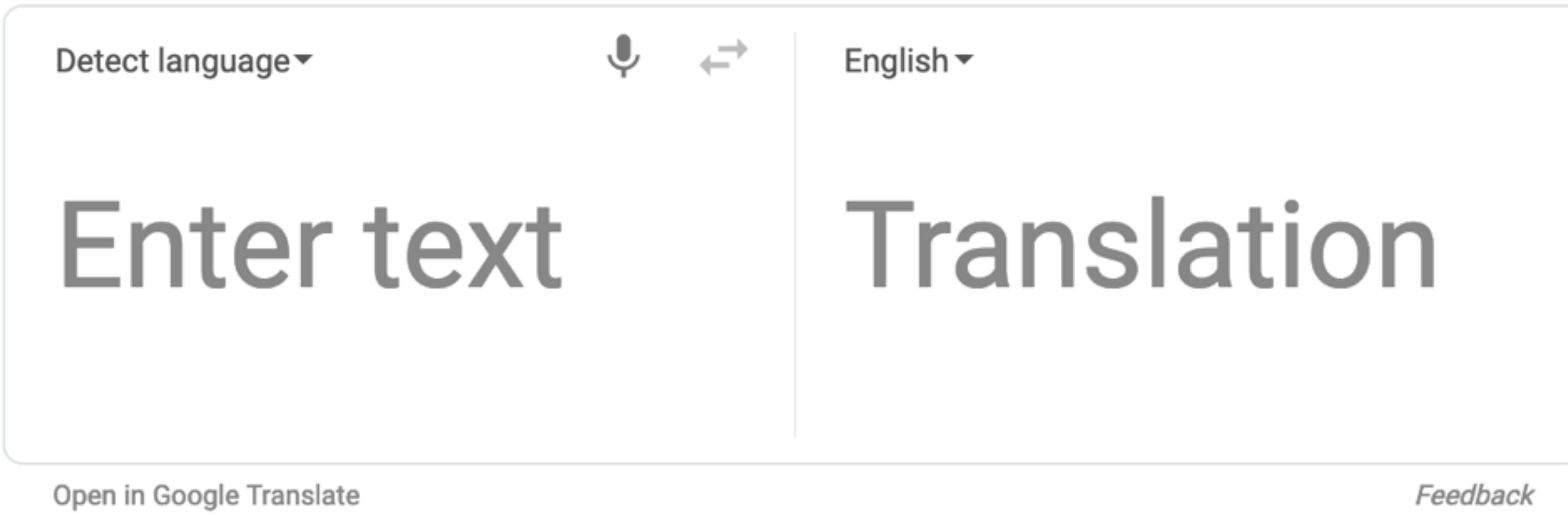
- Field of AI concerned with interactions between humans and computers
- How to program computers to process and analyze natural language data (as opposed to computer languages)

NLP Tasks

- Speech recognition
- Natural language generation
- Natural language understanding

- More specific tasks
  - Machine translation
  - Sentiment Analysis
  - Question Answering
  - Text generatoin

Machine Translation



Statistical machine translation:

P(e|f) e = english phrase f= foreign phrase

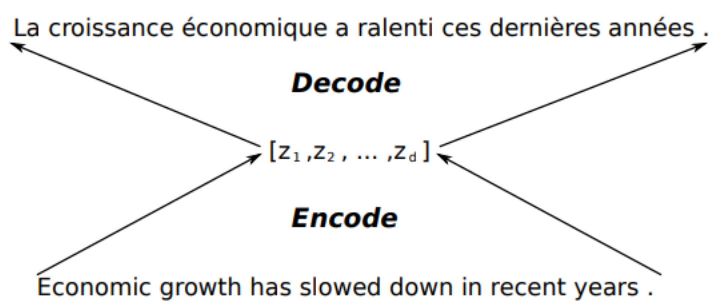
$p(e|f) \propto p(f|e)p(e)$ language model $p(e)$

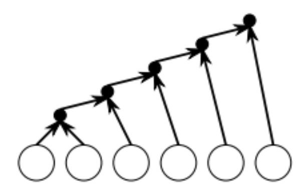
Translates n-grams phrases consisting of n-words

Neural Machine translation

Encoder-decoder architecture

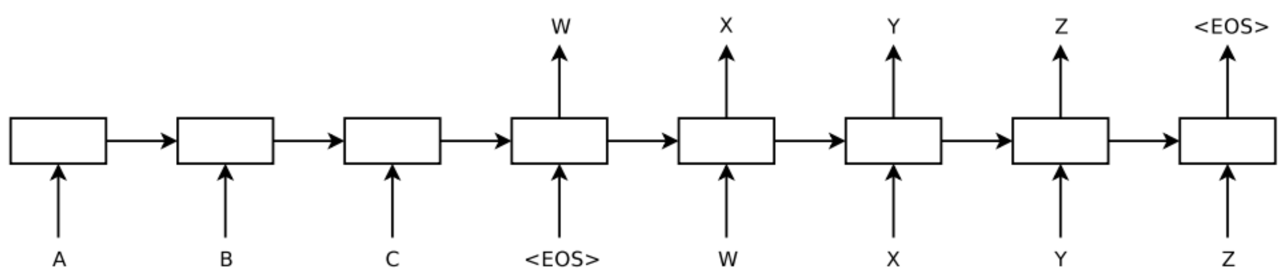
Each of these is a neural network



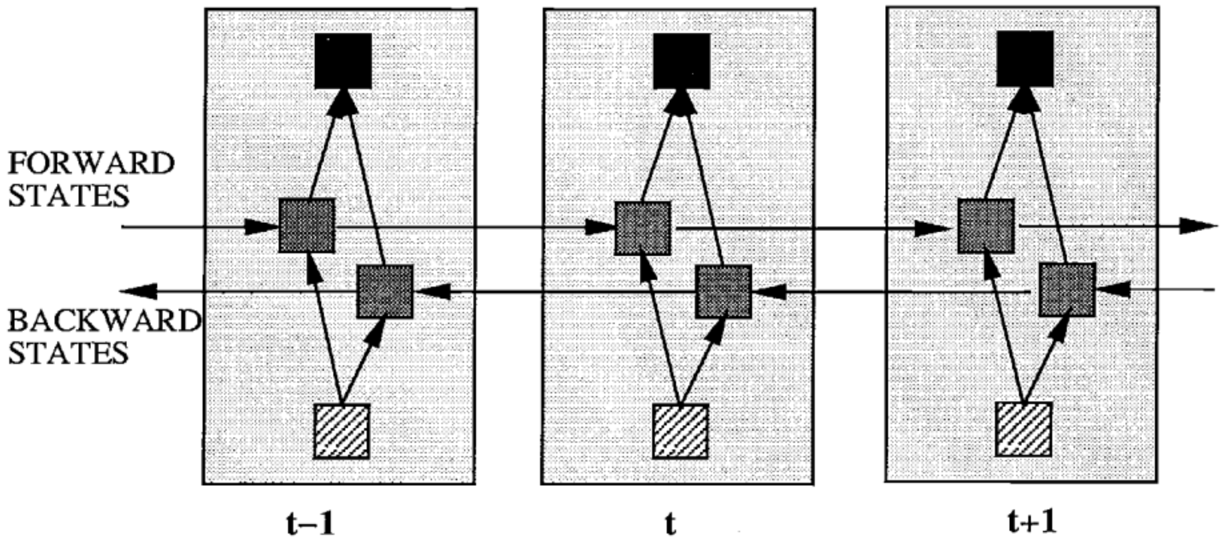


(a)

Cho et al 2014, Sutskever 2014

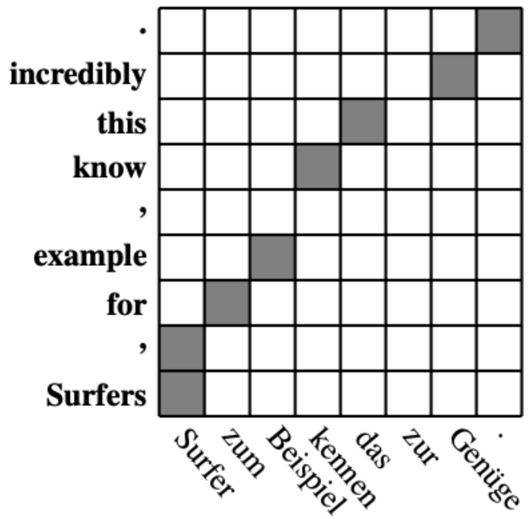


Bidirectional RNN

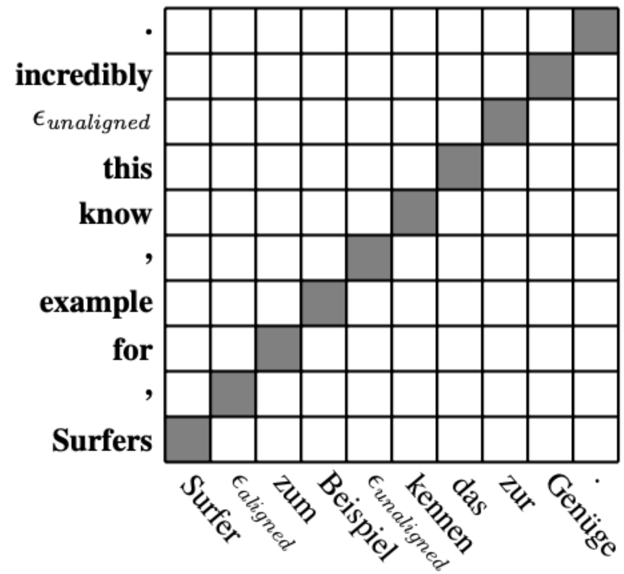


Each time the model generates a phrase, soft searches for a set of positions

Joint Alignment and Translation



(a) Original



(b) One-to-one alignment

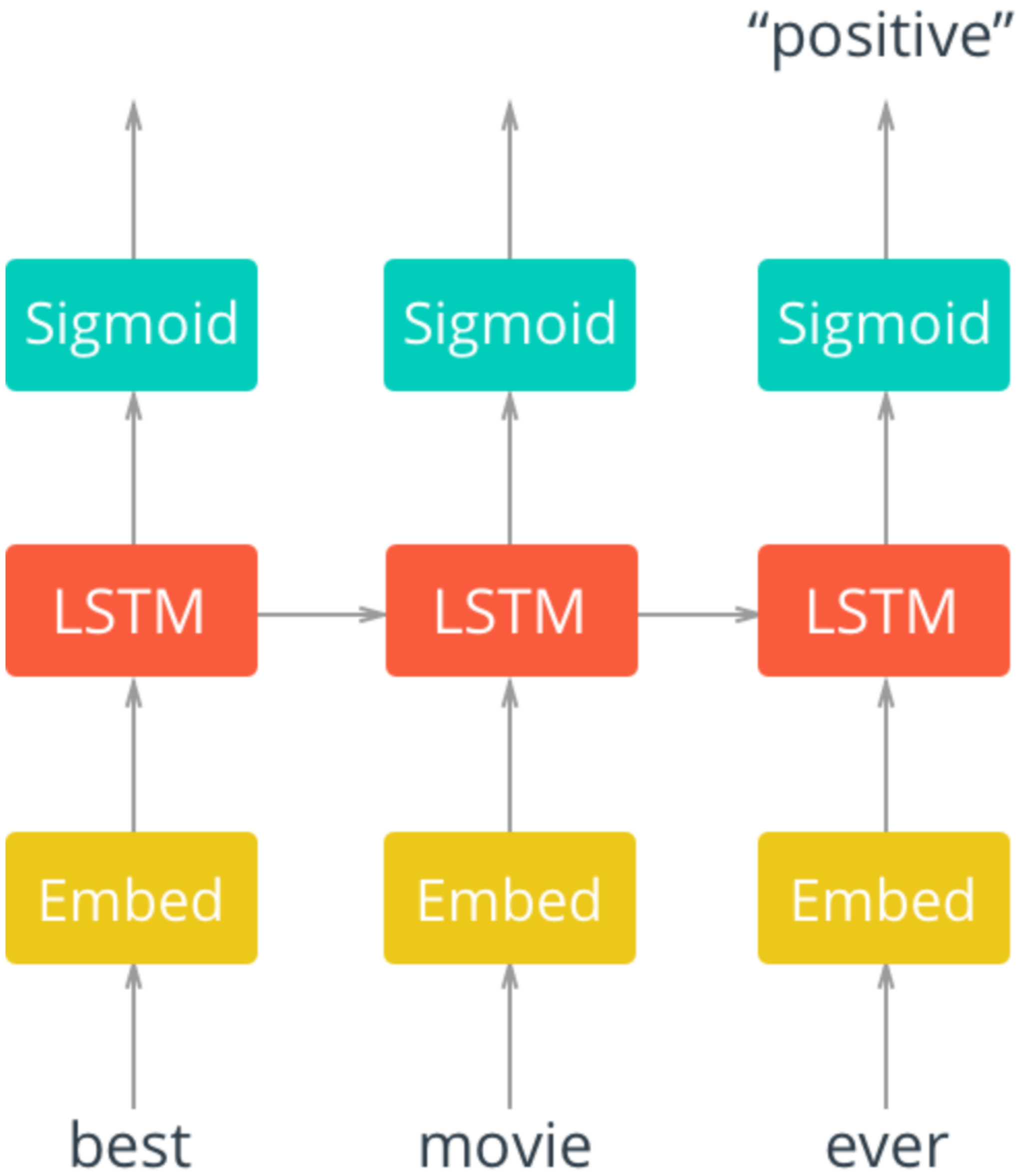
Sentiment Analysis

- Coronet has the best lines of all day cruisers.
- Bertram has a deep V hull and runs easily through seas.
- Pastel-colored 1980s day cruisers from Florida are ugly.
- I dislike old cabin cruisers.

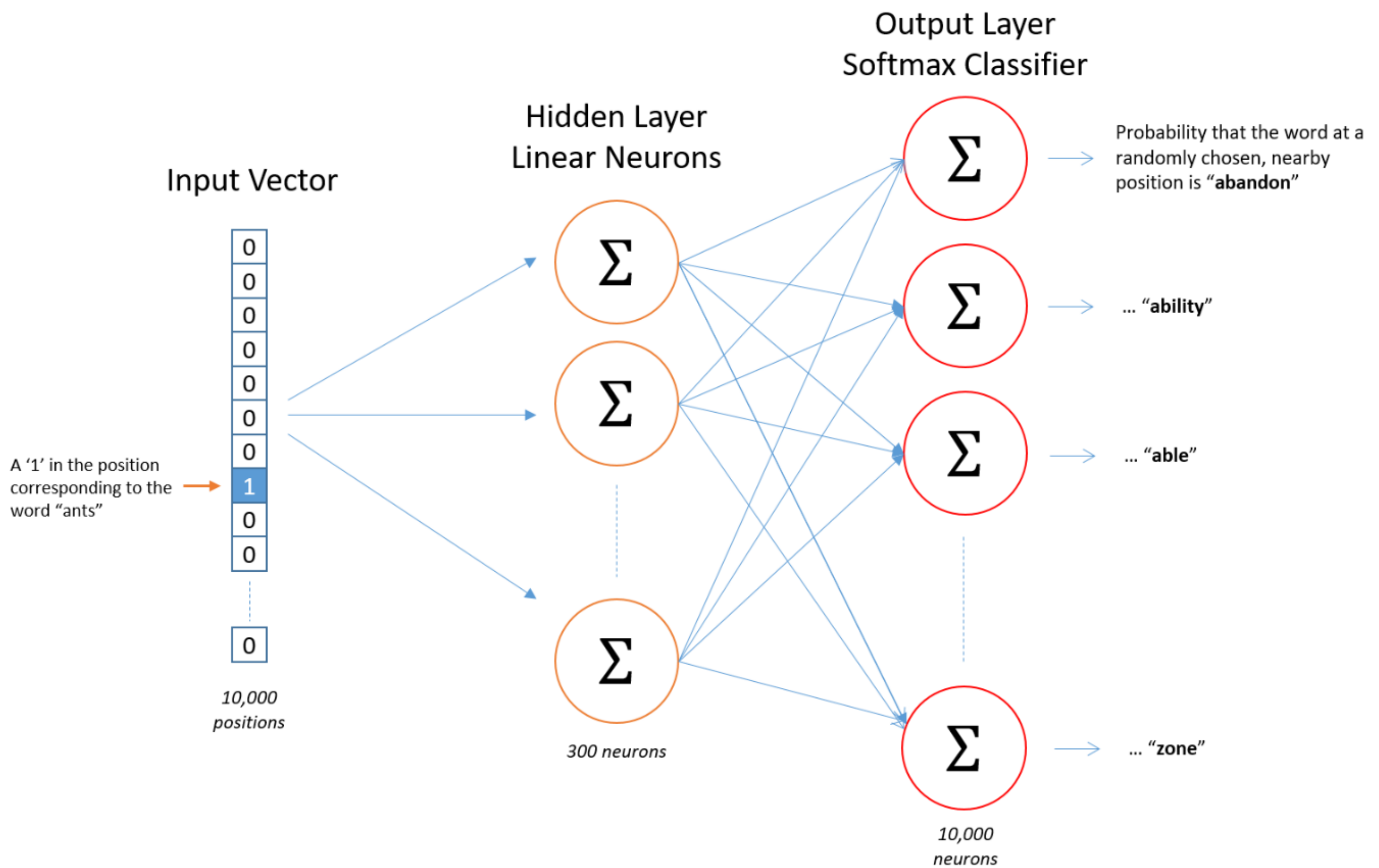
I do not dislike cabin cruisers. (negation handling)

Polarity:

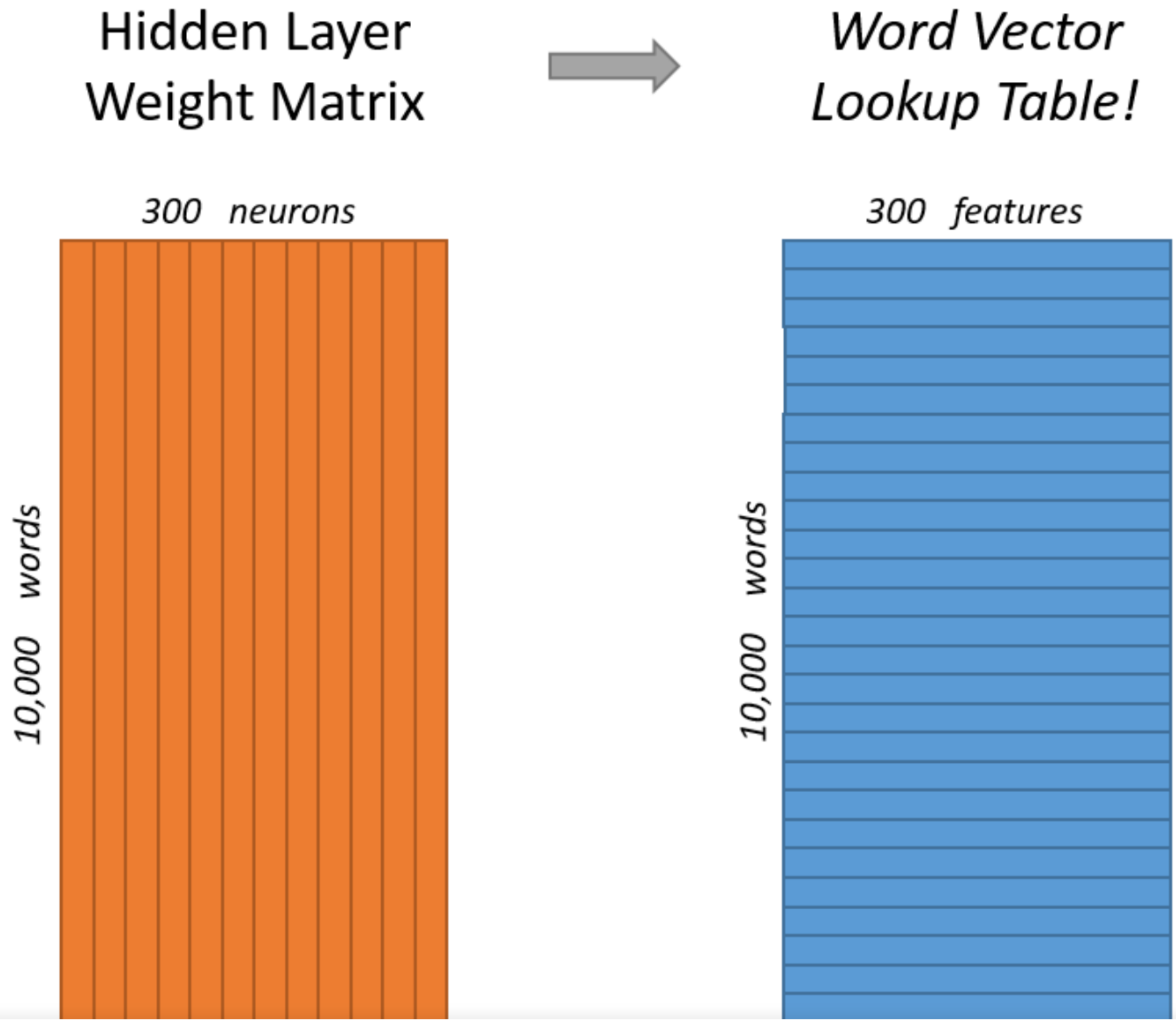
Beyond Polarity


“positive”
Sigmoid
Sigmoid
Sigmoid
LSTM
LSTM
LSTM
Embed
Embed
Embed
best
movie
ever

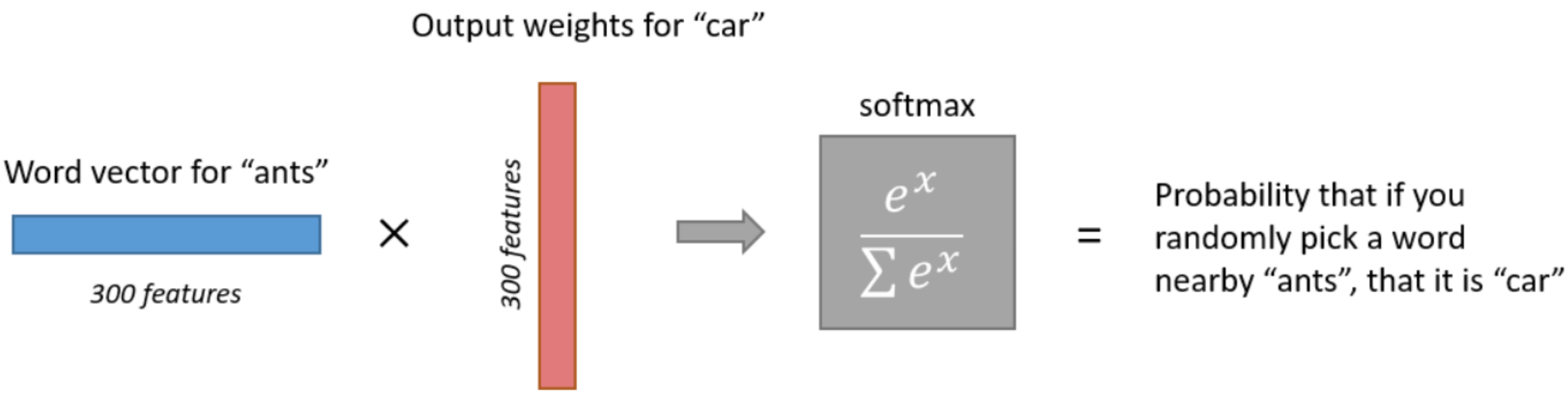
Word vector Embedding



Word vectors

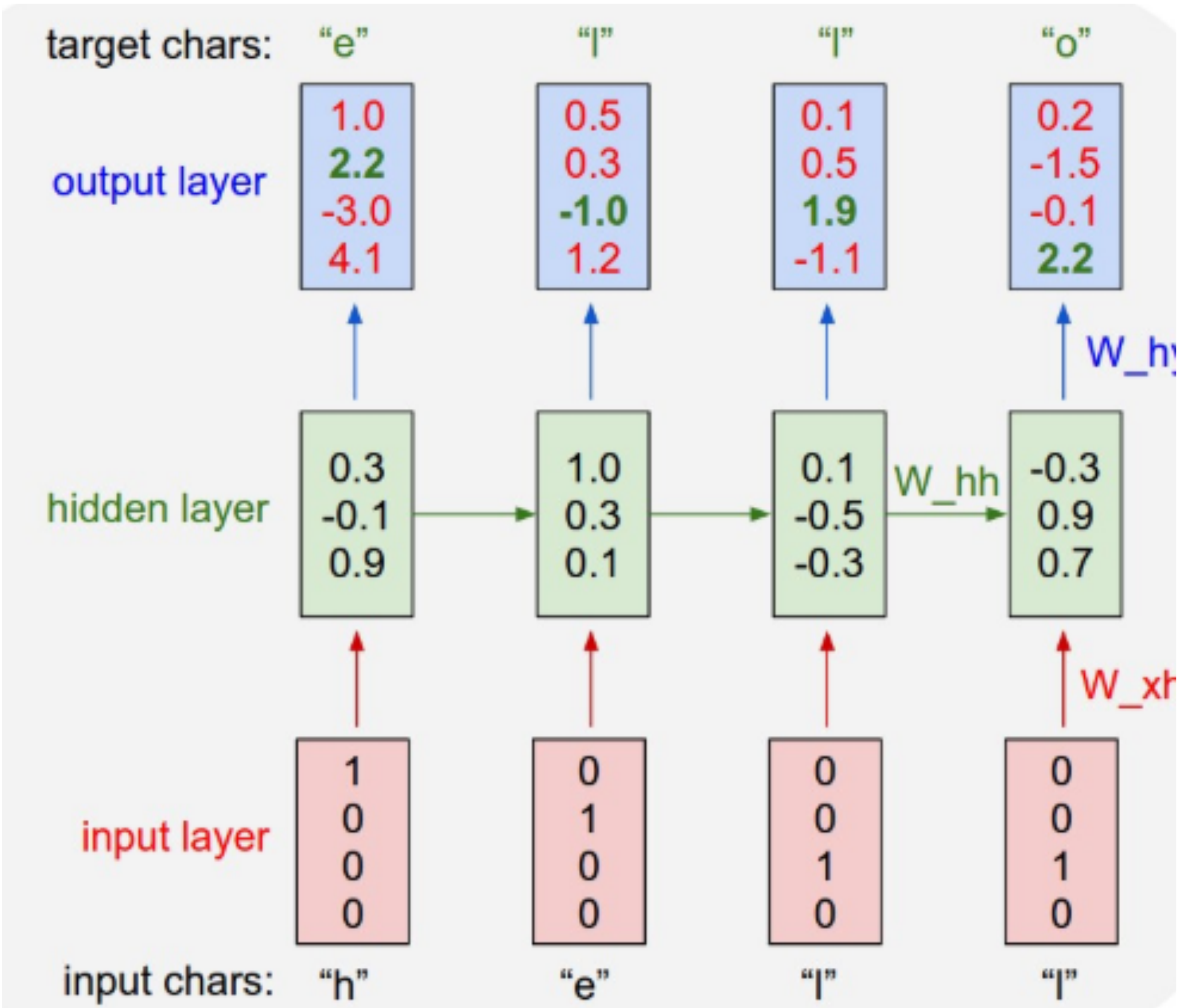


Output calculation



Training on negative samples improves this.

Character Level Language Model



Text Generation

CHARISMA / POWER

January 2017

People who are powerful but uncharismatic will tend to be disliked. Their power makes them a target for criticism that they don't have the charisma to disarm. That was Hillary Clinton's problem. It also tends to be a problem for any CEO who is more of a builder than a schmoozer. And yet the builder-type CEO is (like Hillary) probably the best person for the job.

I don't think there is any solution to this problem. It's human nature. The best we can do is to recognize that it's happening, and to understand that being a magnet for criticism is sometimes a sign not that someone is the wrong person for a job, but that they're the right one.

Text generation databases: Paul Graham Essays, Project Gutenberg

RNN for Text Generation

2-layer LSTM

"The surprised in investors weren't going to raise money. I'm not the company with the time there are all interesting quickly, don't have to get off the same programmers. There's a super-angel round fundraising, why do you can do. If you have a different physical investment are become in people who reduced in a startup with the way to argument the acquirer could see them just that you're also the founders will part of users' affords that and an alternation to the idea. [2] Don't work at first member to see the way kids will seem in advance of a bad successful startup. And if you have to act the big company too."

Temperature

- **Temperature of softmax**

$$P_i = \frac{e^{\frac{y_i}{T}}}{\sum_{k=1}^{n} e^{\frac{y_k}{T}}}$$

- lower number (e.g. 0.5) makes the RNN more confident,
- higher temperatures will give more diversity but at cost of more mistakes (e.g. spelling mistakes, etc). I
- In particular, setting temperature very near zero will give the most likely thing that Paul Graham might say:
-

"is that they were all the same thing that was a startup is that they were all the same thing that was a startup is that they were all the same thing that was a startup is that they were all the same"

Depth helps!

- 3-layer LSTM

```
PANDARUS:
Alas, I think he shall be come approached and the day
When little srain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:
They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:
Well, your wit is in the care of side and that.

Second Lord:
They would be ruled after this chamber, and
my fair nues begun out of the fact, to be conveyed,
Whose noble souls I'll have the heart of the wars.

Clown:
Come, sir, I will make did behold your worship.

VIOLA:
I'll drink it.
```

Baby Names

Rudi Levette Berice Lussa Hany Mareanne Chrestina Carissy Marylen Hammine Janye Marlise Jacacrie Hendred Romand Charienna Nenotto Ette Dorane Wallen Marly Darine Salina Elvyn Ersia Maralena Minoria Ellia Charmin Antley Nerille Chelon Walmor Evena Jeryly Stachon Charisa Allisa Anatha Cathanie Geetra Alexie Jerin Cassen Herbett Cossie Velen Daurenge Robester Shermond Terisa Licia Roselen Ferine Jayn Lusine Charyanne Sales Sanny Resa Wallon Martine Merus Jelen Candica Wallin Tel Rachene Tarine Ozila Ketia Shanne Arnande Karella Roselina Alessia Chasty Deland Berther Geamar Jackein Mellisand Sagdy Nenc Lessie Rasemy Guen Gavi Milea Anneda Margoris Janin Rodelin Zeanna Elyne Janah Ferzina Susta Pey Castina

New names that have not been sampled before!

Evolution of a language model: Iteration 100

```
tyntd-iafhatawiaoihrdemot  lytdws  e ,tfti, astai f ogoh eoase rrranbyne 'nhthnee e
plia tklrgd t o idoe ns,smtt   h ne etie h,hregtrs nigtike,aoaenns lng
```

What did the network learn so far?

Words and spaces.

300 Iterations

```
"Tmont thithey" fomesscerliund
Keushey. Thom here
sheulke, anmerenith ol sivh I lalterthend Bleipile shuwy fil on aseterlome
coaniogennc Phe lism thond hon at. MeiDimorotion in ther thize."
```

Punctuation

Iteration 500

```
we counter. He stutn co des. His stanted out one ofler that concossions and was
to gearang reay Jotrets and with fre colt otf paitt thin wall. Which das stimn
```

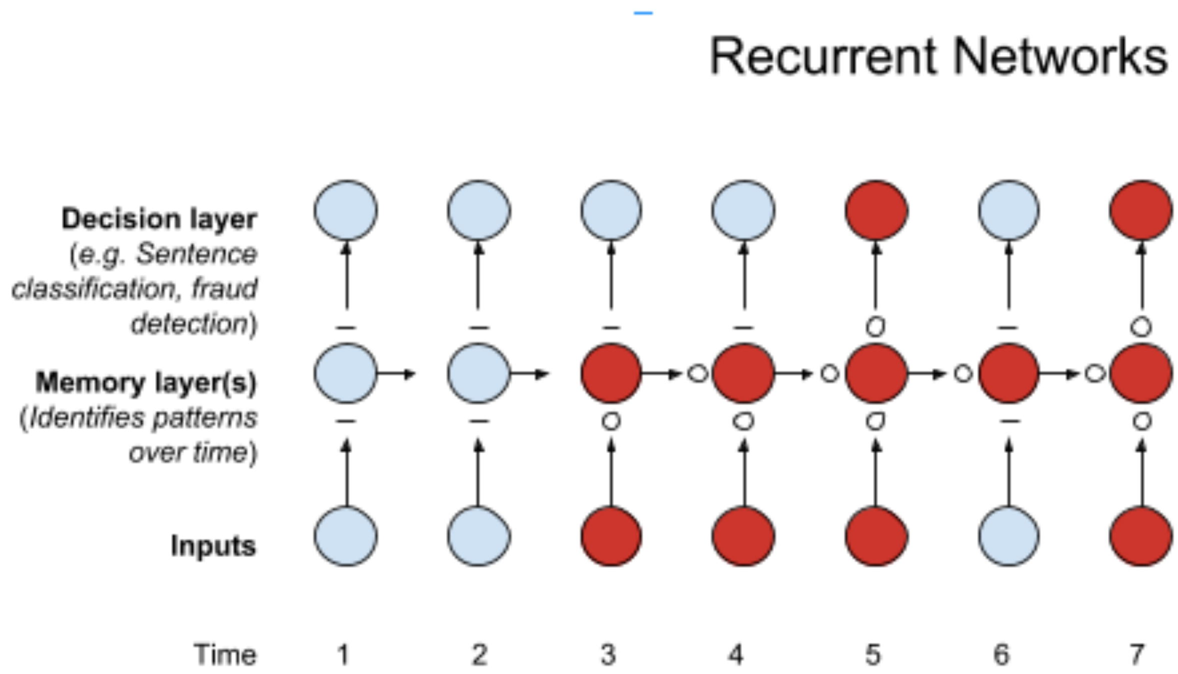
Spelling short words

Iteration 700

```
Aftair fall unsuch that the hall for Prince Velzonski's that me of
her hearly, and behs to so arwage fiving were to it beloge, pavu say falling misfort
how, and Gogition is so overelical and ofter.
```

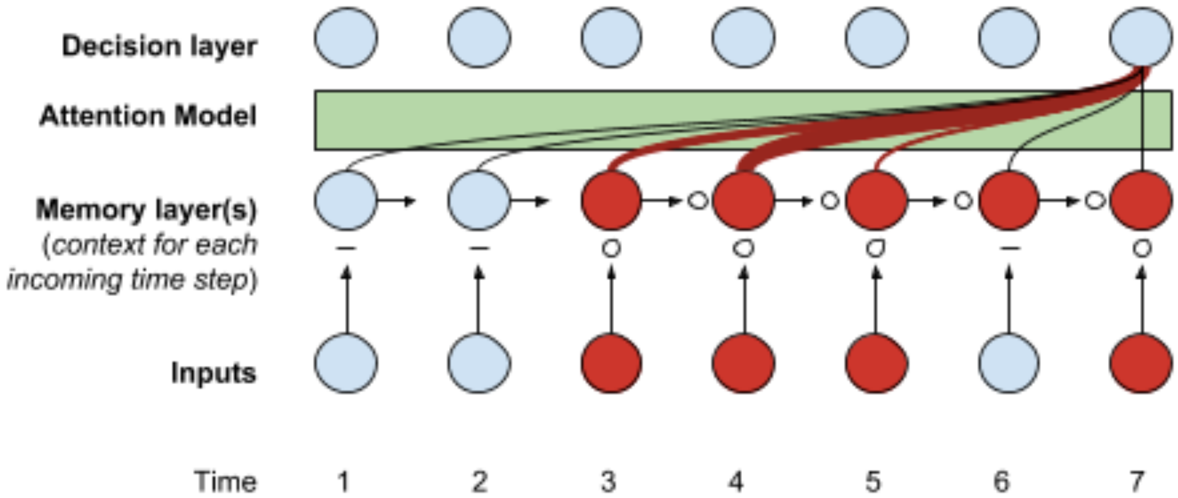
Word ordering

Memory in RNN



- All information has to be in previous hidden state

Attention Mechanism



$$\mathbf{c}_i = \sum_j a_{ij}\mathbf{s}_j$$

$$\mathbf{a}_i = \mathrm{softmax}(f_{att}(\mathbf{h}_i, \mathbf{s}_j))$$

Image Captioning

Attention for image captioning

Further reading

- http://karpathy.github.io/2015/05/21/rnn-effectiveness/
- Neural machine translation https://arxiv.org/pdf/1409.0473.pdf
- Attention: https://arxiv.org/pdf/1706.03762.pdf