Revision of week 3 and 4

Nhung Nguyen and Viktor Schlegel

Week 3 - Distributional semantics

> We shall know a word by the company it keeps.
>
> Firth (1957)

Week 3 - Count-based approach

- Term-document matrix

tf-idf

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	0	7	13
good	114	80	62	89
fool	36	58	1	4
wit	20	15	2	3

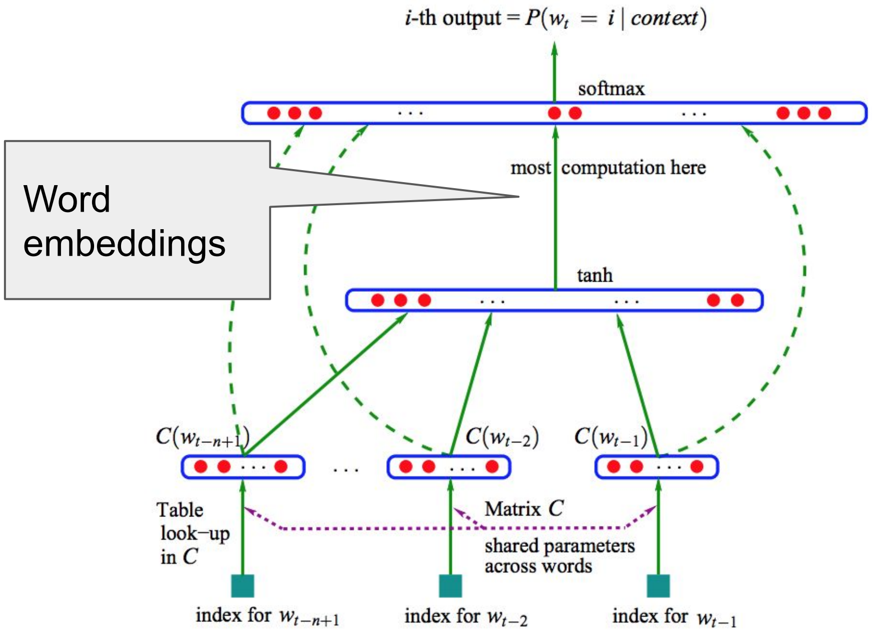
- Term-term matrix

smoothing

	aardvark	computer	data	pinch	result	sugar	...
apricot	0	0	0	1	0	1	
pineapple	0	0	0	1	0	1	
digital	0	2	1	0	1	0	
information	0	1	6	0	4	0	

Week 3 - Prediction-based approach

- Bengio's language model



$\Pr(w_t|w_{t-1},\ldots,w_{t-m+1}) = \text{softmax}(\mathbf{Wy})$

$\mathbf{y} = \tanh(\mathbf{Vx})$

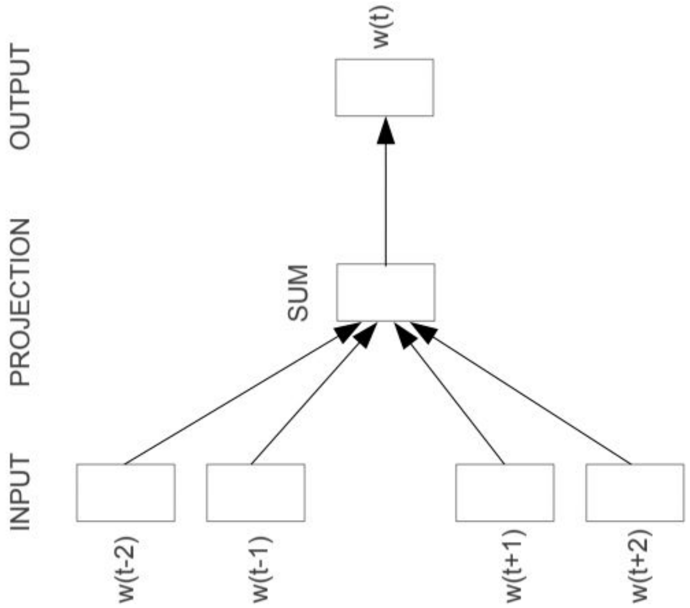
$\mathbf{x} = \text{concat}(\mathbf{w}_{t-1},\ldots,\mathbf{w}_{t-m+1})$

Each *contextual* word w_{t-j} is represented a column of matrix **C**

Week 3 - Prediction-based approach (Cont.)

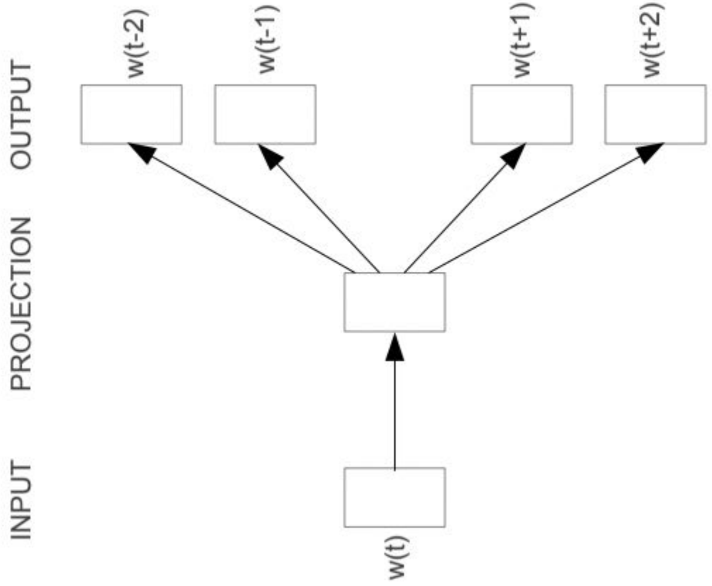
- Mikolov's CBOW

$Pr(w_t|w_{t-1},\dots,w_{t-m+1})$ = softmax(**Wy**)



- Mikolov's Skip-gram

$Pr(.|w_t)$ = softmax(**Wy**)



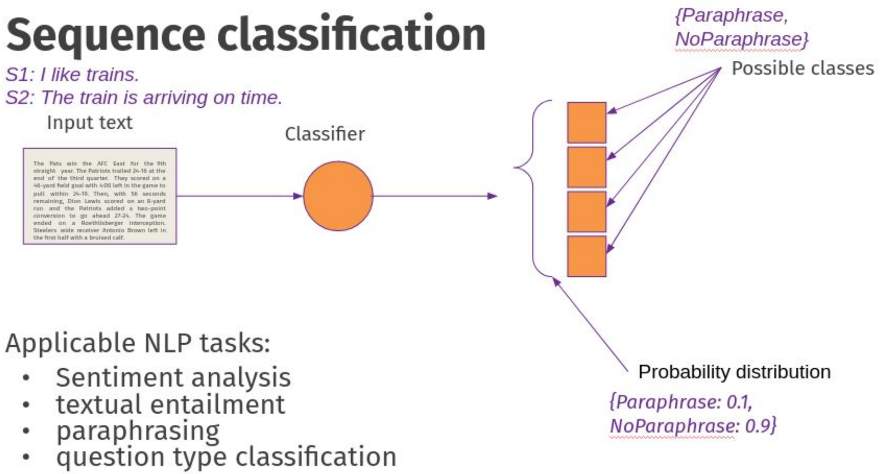
Week 3 - Word Sense Disambiguation

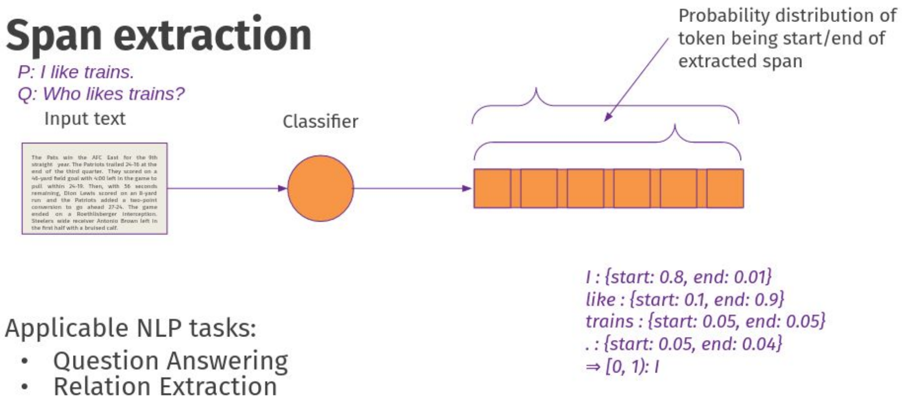
- WordNet: a database of lexical relations
  - A word has different senses

mouse^1 : .... a *mouse* controlling a computer system in 1968.
mouse^2 : .... a quiet animal like a *mouse*
bank^1 : ...a *bank* can hold the investments in a custodial account ...
bank^2 : ...as agriculture burgeons on the east *bank*, the river ...

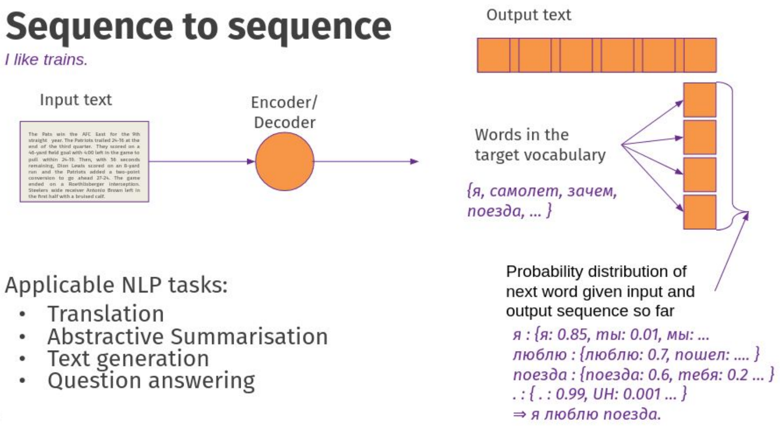
- Disambiguating word sense using Lesk's algorithm and supervised learning

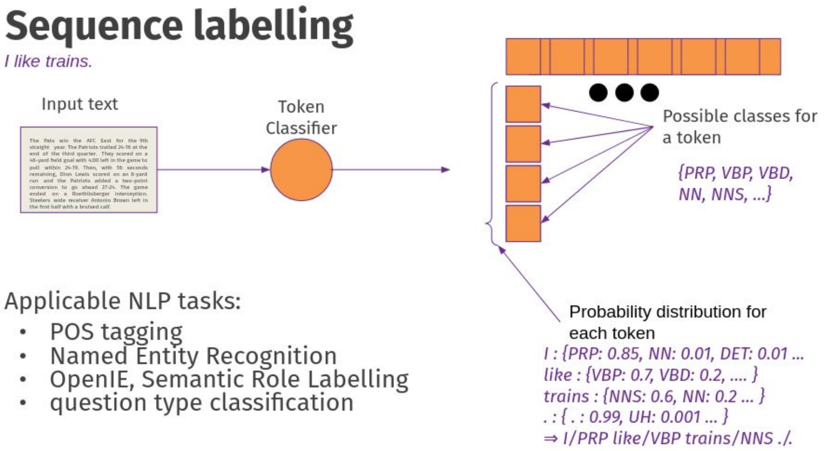
Week 4 - Sequence modelling





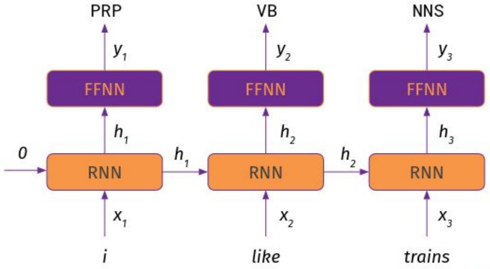
Week 4 - Sequence modelling





Week 4: RNN, LSTM, BiLSTM

RNN: Forward run

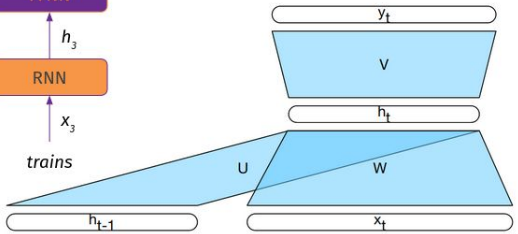


$$h_0 = 0$$
$$h_t = g(\boxed{Uh_{t-1}} + Wx_t + b)$$
$$y_t = f(Vh_t + b_z)$$

RNN: Backward run

Backpropagation through time

h_t depends on h_{t-1}
Unroll computation graph and do standard backpropagation, because t is not infinite.

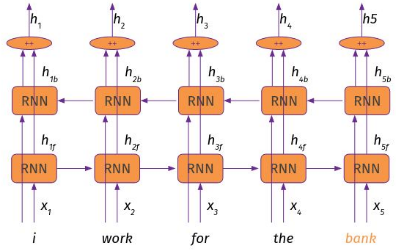


RNN: vanishing gradients

< 0.25! Lots of multiplications!

BiRNN

Idea: If we can go from left to right (“past”), why not also just go right to left (“future”)



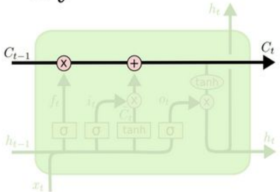
concatenation

$$h_t = [h_{t_f}; h_{t_b}]$$
$$dim(h_t) = dim(h_{t_f}) + dim(h_{t_b})$$

16

LSTM: Context vector

- Context, or memory vector C_t in addition to h_t
- Context information from “the past” calculations
- At any step t, LSTM learns how much of h_t is to be “added” to C_t and how much of C_{t-1} is “kept”

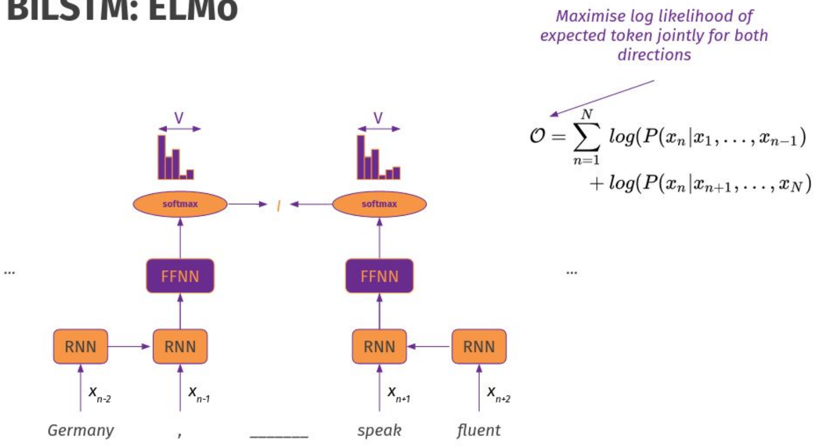


For example: Information about grammatical gender of subject

Week 4 - Contextualised embeddings

```
d[word] = vector
f(word, context) = contextualised_vector
```

Language modelling with BiLSTM: ELMo



Task: probability of next word given *n* previous words.

$$P(x_{n+1}|x_1,...,x_n)$$

"I grew up in Germany, i speak fluent ____"

- syntax information ✔
- semantic information ✔
- self supervision ✔

Week 4 - NER

NER as a tagging problem (BIO scheme)

	Adam	Smith	works	for	IBM	in	London	.
POS tagging	NNP	NNP	VBZ	IN	NNP	IN	NNP	.
Entity recognition	B_PER	I_PER	O	O	B_ORG	O	B_LOC	O

Begin the mention

Inside the mention

- # classes = 2 * # entity types + 1

Week 4 - NER approaches

- Local approach: tags are ***independent*** each other
  - Any classifiers can be used
    - SVM: truly local
    - RNN, LSTM, BiLSTM: not truly local

- Global approach: tags are ***dependent*** each other
  - Hidden Markov Model (HMM)
  - Conditional Random Fields (CRF)

Week 4: CRF vs. Neural networks

CRF

- Feature engineering
- Do not need pre-trained vectors
- Models are roughly interpretable
- *Perform well with datasets that have many NE categories(*)*

Neural networks

- Do not need features
- Need pre-trained vectors from big language models
- Models use implicit features (created by hidden layers) → not easy to interpret
- *Perform not so well with datasets that have many NE categories(*)*

(*) This observation is only based on my personal experiences

Week 4 - NEL

Mentions are given

Mr. Trump discussed Brexit with Mrs. May .

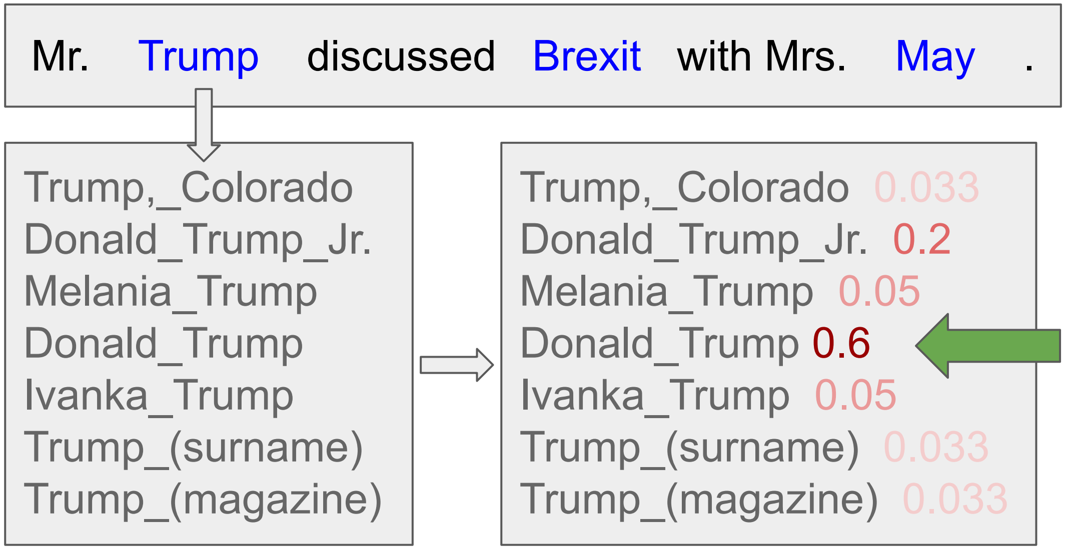
en.wikipedia.org/wiki/Donald_Trump

en.wikipedia.org/wiki/Brexit

en.wikipedia.org/wiki/Theresa_May

Looks like ***classification*** but very ***"fine-grained"***

Week 4 - NEL basic steps



Candidate generation
(A short list of candidates, < 100)

Candidate ranking

Thank you very much

Good luck with your exams!