

Welcome and Introduction

Miguel Rios

Universiteit van Amsterdam

March 29, 2019

Content

- ① Introduction
- ② Natural Language Processing
- ③ Course Topics

Course details

- Github course page <https://uva-slpl.github.io/nlp2/>

Course details

- Github course page <https://uva-slpl.github.io/nlp2/>
- Syllabus

Course details

- Github course page <https://uva-slpl.github.io/nlp2/>
- Syllabus
 - Slides

Course details

- Github course page <https://uva-slpl.github.io/nlp2/>
- Syllabus
 - Slides
 - Reading material

Course details

- Github course page <https://uva-slpl.github.io/nlp2/>
- Syllabus
 - Slides
 - Reading material
- Projects

Course details

- Github course page <https://uva-slpl.github.io/nlp2/>
- Syllabus
 - Slides
 - Reading material
- Projects
- Posts

Course details

- Github course page <https://uva-slpl.github.io/nlp2/>
- Syllabus
 - Slides
 - Reading material
- Projects
- Posts
- Grading

Course details

- Github course page <https://uva-slpl.github.io/nlp2/>
- Syllabus
 - Slides
 - Reading material
- Projects
- Posts
- Grading
 - Report in groups of 3

Course details

- Github course page <https://uva-slpl.github.io/nlp2/>
- Syllabus
 - Slides
 - Reading material
- Projects
- Posts
- Grading
 - Report in groups of 3
 - Project 1 **50%**

Course details

- Github course page <https://uva-slpl.github.io/nlp2/>
- Syllabus
 - Slides
 - Reading material
- Projects
- Posts
- Grading
 - Report in groups of 3
 - Project 1 **50%**
 - Project 2 **50%**

Course details

- Github course page <https://uva-slpl.github.io/nlp2/>
- Syllabus
 - Slides
 - Reading material
- Projects
- Posts
- Grading
 - Report in groups of 3
 - Project 1 **50%**
 - Project 2 **50%**
- Lab starts **April 10th** check out the Posts for more info.

What is NLP?

- Goal understanding of language
Not only string or keyword matching

What is NLP?

- Goal understanding of language
Not only string or keyword matching
- End systems

What is NLP?

- Goal understanding of language
Not only string or keyword matching
- End systems
 - [Classification](#): Text categorization, sentiment classification

What is NLP?

- Goal understanding of language
Not only string or keyword matching
- End systems
 - **Classification**: Text categorization, sentiment classification
 - **Generation**: Question answering, Machine Translation

What is NLP?

- Goal understanding of language
Not only string or keyword matching
- End systems
 - **Classification**: Text categorization, sentiment classification
 - **Generation**: Question answering, Machine Translation
- Computational methods to learn more about how language works
(Computational Linguistics)

Natural language inference

- Textual entailment is defined as a directional relation between pairs of text expressions, the T Text, and the H Hypothesis.

Natural language inference

- Textual entailment is defined as a directional relation between pairs of text expressions, the T Text, and the H Hypothesis.
- Systems decide for each entailment pair whether T entails H or not.

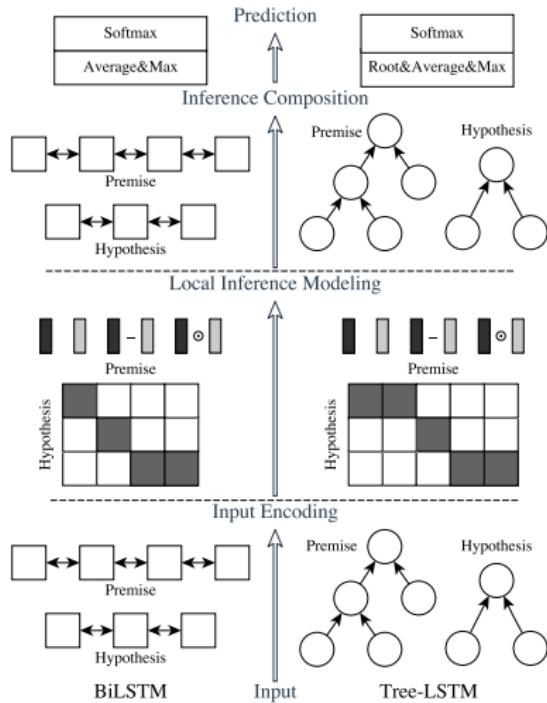
Natural language inference

- Textual entailment is defined as a directional relation between pairs of text expressions, the T Text, and the H Hypothesis.
- Systems decide for each entailment pair whether T entails H or not.

T: The purchase of Houston-based LexCorp by BMI for \$2Bn prompted widespread sell-offs by traders as they sought to minimize exposure.

H: BMI acquired an American company.

Natural language inference



⁰[Chen et al., 2016]

Machine translation

- [Bahdanau et al., 2015]

Machine translation

Question answering

Task 1**Story**

john went to the kitchen
daniel travelled to the kitchen
sandra journeyed to the kitchen
john went to the bedroom
mary went to the bedroom
sandra went back to the bedroom
john journeyed to the garden
daniel went back to the bedroom

Question

where is john

Answer

garden

Confidence

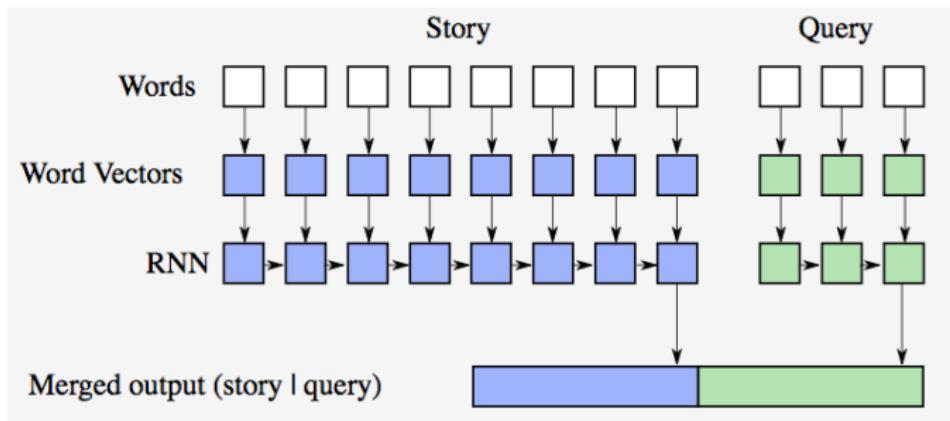
100.0%

Correct answer

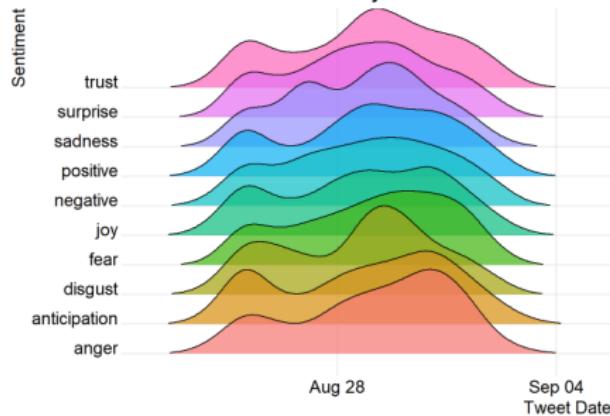
garden

Sentence	Hop 1	Hop 2	Hop 3
john went to the kitchen	0.000	0.000	0.000
daniel travelled to the kitchen	0.000	0.000	0.000
sandra journeyed to the kitchen	0.000	0.000	0.000
john went to the bedroom	0.025	0.000	0.000
mary went to the bedroom	0.000	0.000	0.000
sandra went back to the bedroom	0.000	0.000	0.000
john journeyed to the garden	0.973	1.000	1.000
daniel went back to the bedroom	0.000	0.000	0.000

Question answering



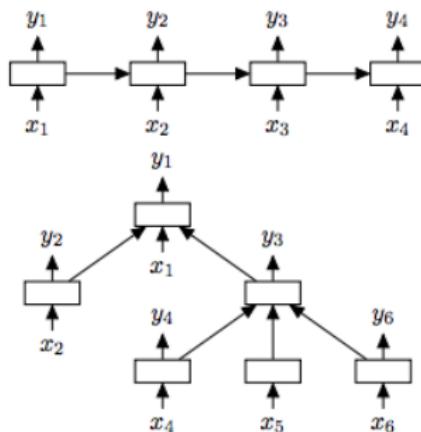
Sentiment classification



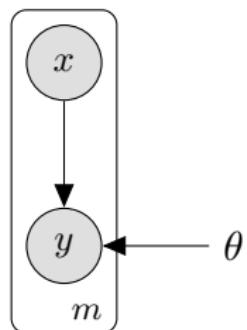
A word cloud visualization showing common positive and neutral words. Large words include fun, love, excited, teach, create, improve, and simplify. Smaller words include finally, clean, helpful, true, good, safe, etc.

⁰<https://www.edgarsdatalab.com/2017/09/04/sentiment-analysis-using-tidytext/>

Sentiment classification



Graphical Models



Supervised learning

- We have data inputs $X = \langle x_1, \dots, x_n \rangle$, and the corresponding outputs $Y = \langle y_1, \dots, y_n \rangle$ generated by some unknown procedure

Supervised learning

- We have data inputs $X = \langle x_1, \dots, x_n \rangle$, and the corresponding outputs $Y = \langle y_1, \dots, y_n \rangle$ generated by some unknown procedure
- which we assume can be captured by a probabilistic model with known probability (mass/density) function e.g.

$$p(y|x, \theta) = \text{Cat}(y|f(x; \theta)), \quad (1)$$

Supervised learning

- We have data inputs $X = \langle x_1, \dots, x_n \rangle$, and the corresponding outputs $Y = \langle y_1, \dots, y_n \rangle$ generated by some unknown procedure
- which we assume can be captured by a probabilistic model with known probability (mass/density) function e.g.

$$p(y|x, \theta) = \text{Cat}(y|f(x; \theta)), \quad (1)$$

- y classes computed with a neural network f parameterised by θ

Supervised learning

- We have data inputs $X = \langle x_1, \dots, x_n \rangle$, and the corresponding outputs $Y = \langle y_1, \dots, y_n \rangle$ generated by some unknown procedure
- which we assume can be captured by a probabilistic model with known probability (mass/density) function e.g.

$$p(y|x, \theta) = \text{Cat}(y|f(x; \theta)), \quad (1)$$

- y classes computed with a neural network f parameterised by θ
- estimate parameters that assign maximum likelihood to observations

Supervised learning

	x	y
Parsing	Sentence	Syntactic tree
Machine translation	Source	Target translation
NLI	Text and Hypothesis	Entailment relation

Supervised learning

Dependency:

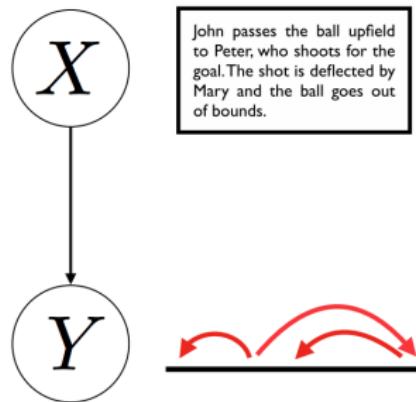


Parts-of-speech:

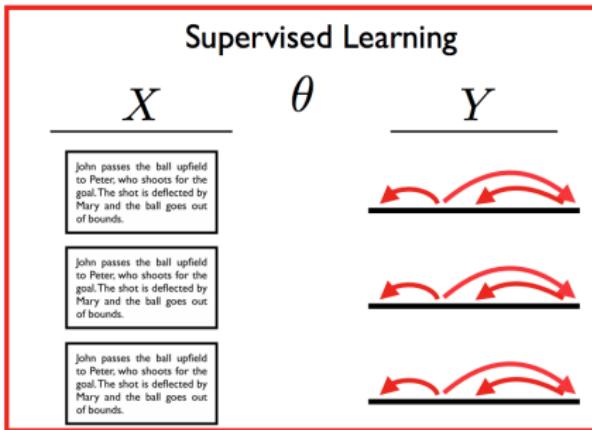
DT NN VBD IN DT JJ NN

The cat sat on a green wall

Supervised learning



Supervised learning



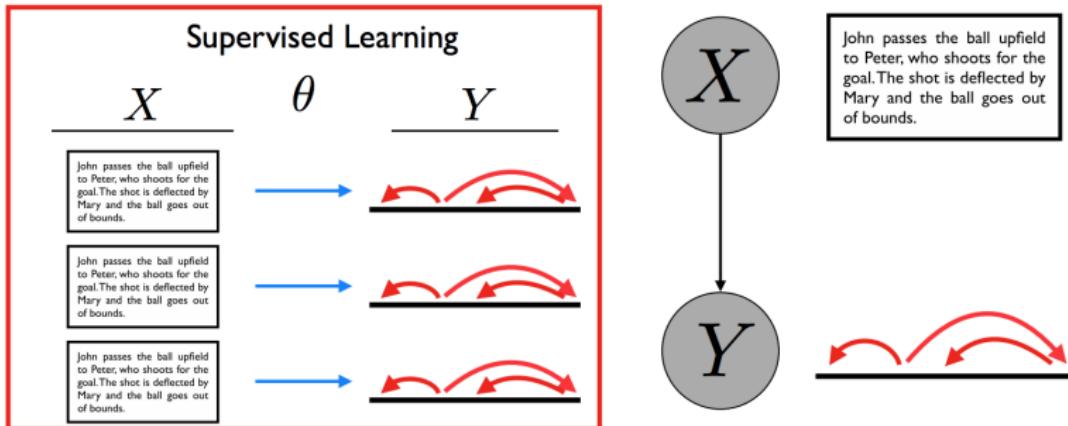
X

John passes the ball upfield to Peter, who shoots for the goal. The shot is deflected by Mary and the ball goes out of bounds.

Y



Supervised learning



Supervised learning

- Maximum likelihood estimation tells you which loss to optimise (i.e. negative log-likelihood)

Supervised learning

- Maximum likelihood estimation tells you which loss to optimise (i.e. negative log-likelihood)
- Automatic differentiation (backprop) chain rule of derivatives: give a tractable forward pass and get gradients

Supervised learning

- Maximum likelihood estimation tells you which loss to optimise (i.e. negative log-likelihood)
- Automatic differentiation (backprop) chain rule of derivatives: give a tractable forward pass and get gradients
- Stochastic optimisation powered by backprop general purpose gradient-based optimisers

Latent variable approach

- Because NN models work but they may struggle with:

Latent variable approach

- Because NN models work but they may struggle with:
- lack of training data

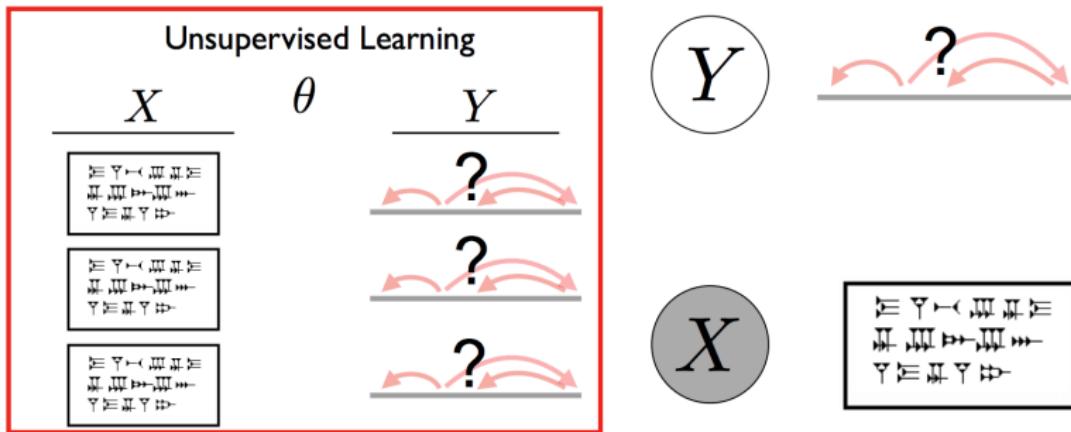
Latent variable approach

- Because NN models work but they may struggle with:
- lack of training data
- partial supervision

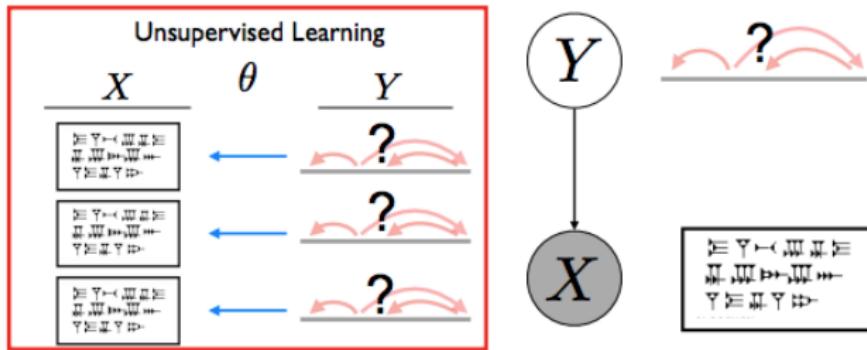
Latent variable approach

- Because NN models work but they may struggle with:
- lack of training data
- partial supervision
- lack of inductive bias

Latent variable approach

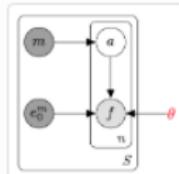


Latent variable approach



What is this course?

Lexical alignment



IBM 1 and 2: Models over words and MLE via EM for categorical distributions

2019-04-04.

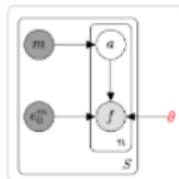
Abstract

Slides

Class material

Background reading

Further reading



Cont. IBM 1 and 2: Models over words and MLE via EM for categorical distributions

2019-04-08.

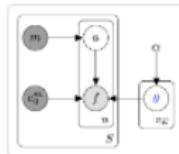
Abstract

Slides

Class material

Background reading

Further reading



Bayesian IBM1: Dirichlet priors and posterior inference

2019-04-11.

Abstract

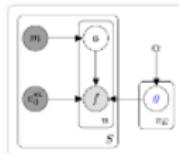
Slides

Class material

Background reading

Further reading

Discussion

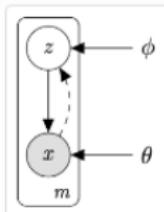


Neural IBM Models

2019-04-15.

What is this course?

Deep generative models for NLP

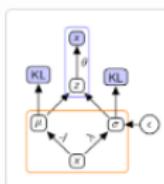


Probabilistic modelling for NLP

2019-04-18.

Abstract

Background reading



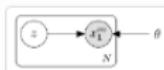
Variational auto-encoders

2019-04-25.

Abstract

Background reading

Further reading



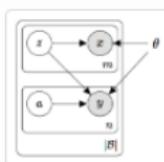
Generative language models

2019-04-29.

Abstract

Background reading

Discussion



Generative models of word representation

2019-05-02.

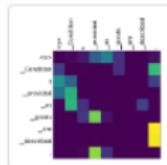
Abstract

Background reading

Further reading

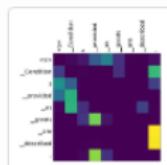
Discussion

What is this course?



Generative models for natural language inference and machine translation

2019-05-06.



Continuous relaxations of discrete variables

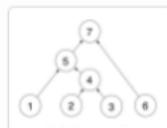
2019-05-09.

Abstract

Background reading

Further reading

Discussion



Discrete latent variable models and generative models for morphology

2019-05-13.

Abstract

Background reading

Further reading

Discussion



Guest Lecture

TBA. 2019-05-16.

Goals

- go through current literature

Goals

- go through current literature
- define probabilistic models

Goals

- go through current literature
- define probabilistic models
- start combining probabilistic models and NN architectures

Next class

- Probabilistic Graphical Models

Next class

- Probabilistic Graphical Models
- Introduction to Word Alignment

Questions?

References I

- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *ICLR*, 2015.
URL <http://arxiv.org/abs/1409.0473>.
- Qian Chen, Xiaodan Zhu, Zhen-Hua Ling, Si Wei, and Hui Jiang.
Enhancing and combining sequential and tree LSTM for natural
language inference. *CoRR*, abs/1609.06038, 2016. URL
<http://arxiv.org/abs/1609.06038>.
- Steve Merity. Question answering on the facebook babi dataset using
recurrent neural networks and 175 lines of python keras, 2015. URL
https://smerity.com/articles/2015/keras_qa.html.
- Graham Neubig. Learning with latent linguistic structure, 2018. URL
<http://www.phontron.com/slides/neubig18blackbox.pdf>.

References II

Kai Sheng Tai, Richard Socher, and Christopher D. Manning. Improved semantic representations from tree-structured long short-term memory networks. *CoRR*, abs/1503.00075, 2015. URL
<http://arxiv.org/abs/1503.00075>.