

UMass · CS685 | Advanced Natural Language Processing (2020)

CS685 (2020) · 课程资料包 @ShowMeAI



视频

中英双语字幕



课件

一键打包下载



笔记

官方笔记翻译



代码

作业项目解析



视频 · B 站 [扫码或点击链接]

<https://www.bilibili.com/video/BV1BL411t7RV>



课件 & 代码 · 博客 [扫码或点击链接]

<http://blog.showmeai.tech/umass-cs685>

NLP

迁移学习

语言模型 问答系统 文本生成 BERT

语义解析

知识推理

模型蒸馏

transformer

GPT-3

注意力机制

Awesome AI Courses Notes Cheatsheets 是 [ShowMeAI](#) 资料库的分支系列，覆盖最具知名度的 **TOP50+** 门 AI 课程，旨在为读者和学习者提供一整套高品质中文学习笔记和速查表。

点击课程名称，跳转至课程**资料包**页面，**一键下载**课程全部资料！

机器学习	深度学习	自然语言处理	计算机视觉
Stanford · CS229	Stanford · CS230	Stanford · CS224n	Stanford · CS231n
# Awesome AI Courses Notes Cheatsheets · 持续更新中			
知识图谱	图机器学习	深度强化学习	自动驾驶
Stanford · CS520	Stanford · CS224W	UCBerkeley · CS285	MIT · 6.S094



微信公众号

资料下载方式 2：扫码点击**底部菜单栏**

称为 **AI 内容创作者**？回复 [添砖加瓦]

course introduction

CS 685, Fall 2020

Advanced Natural Language Processing

<http://people.cs.umass.edu/~miyyer/cs685/>

Mohit Iyyer

College of Information and Computer Sciences

University of Massachusetts Amherst

Course logistics

This class will be completely asynchronous with the exception of office hours!

- Each Monday, new videos and readings will be released (see course website).
- There will normally be a short quiz about the week's topics, to be submitted on Gradescope (none for the first week!)
- Occasionally, we will release optional practice problems to help you prepare for the exam. Feel free to discuss these during office hours!
- Gradescope for all assignment submissions

TAs:

Tu Vu

Simeng Sun

Kalpesh Krishna

The TAs are my own PhD students and are very experienced with NLP research!

email all of us (including me!) at
cs685instructors@gmail.com

course website:

<https://people.cs.umass.edu/~miyyer/cs685>

Zoom office hours! (all times Eastern)

Monday w/ Tu: 8-9am

Tuesday w/ Mohit: 8-9am

Wednesday w/ Simeng: 8-9am

Thursday w/ Kalpesh: 2-3pm

All office hours will begin on 8/31 (i.e., none the first week)

If necessary, office hours will be extended by one hour during homework / exam weeks

anonymous questions / comments?

- submit questions/concerns/feedback to <https://forms.gle/wtSgjAQ3aa9z29ux5>
- or use Piazza (you should all be enrolled already)
- Mohit will include responses to some/all of these questions (as well as Piazza posts) in the weekly videos

No official prereqs, but the following will be useful:

- comfort with programming
 - We'll be using Python (and PyTorch) throughout the class
- comfort with probability, linear algebra, and mathematical notation
- Some familiarity with matrix calculus
- Excitement about language!
- Willingness to learn

Please brush up on
these things as needed!

Grading breakdown

- (10%) weekly quizzes
- (30%) Problem sets
 - Written: math and concepts
 - Programs: in Python
 - All HWs will be on Google Colab
- (25%) Midterm (mid-October, open book/internet)
- (35%) Final projects (groups of 4)
 - Choose any topic you want
 - Project proposal (10%)
 - Final report (25%)

Readings

- No need to buy any textbooks!
- Readings will be provided as PDFs on website
 - Usually NLP research papers / notes

natural language processing

natural language processing

languages that evolved naturally through human use
e.g., Spanish, English, Arabic, Hindi, etc.

NOT: controlled languages (e.g., Klingon)

NOT: programming languages

Levels of linguistic structure

Characters

A	l	i	c	e		t	a	l	k	e	d		t	o		B	o	b	.
---	---	---	---	---	--	---	---	---	---	---	---	--	---	---	--	---	---	---	---

Levels of linguistic structure

Morphology

Characters

talk-ed [VerbPast]

A	l	i	c	e		t	a	l	k	e	d		t	o		B	o	b	.
---	---	---	---	---	--	---	---	---	---	---	---	--	---	---	--	---	---	---	---

Levels of linguistic structure

Words

Alice talked to Bob .

Morphology

talk -ed [VerbPast]

Characters

A l i c e t a l k e d t o B o b .

Levels of linguistic structure

Syntax: Part of Speech

Words

Morphology

Characters

Noun

VerbPast

Prep

Noun

Punct

Alice

talked

to

Bob

.

talk

-ed

[VerbPast]

Alice talked to Bob.

Levels of linguistic structure

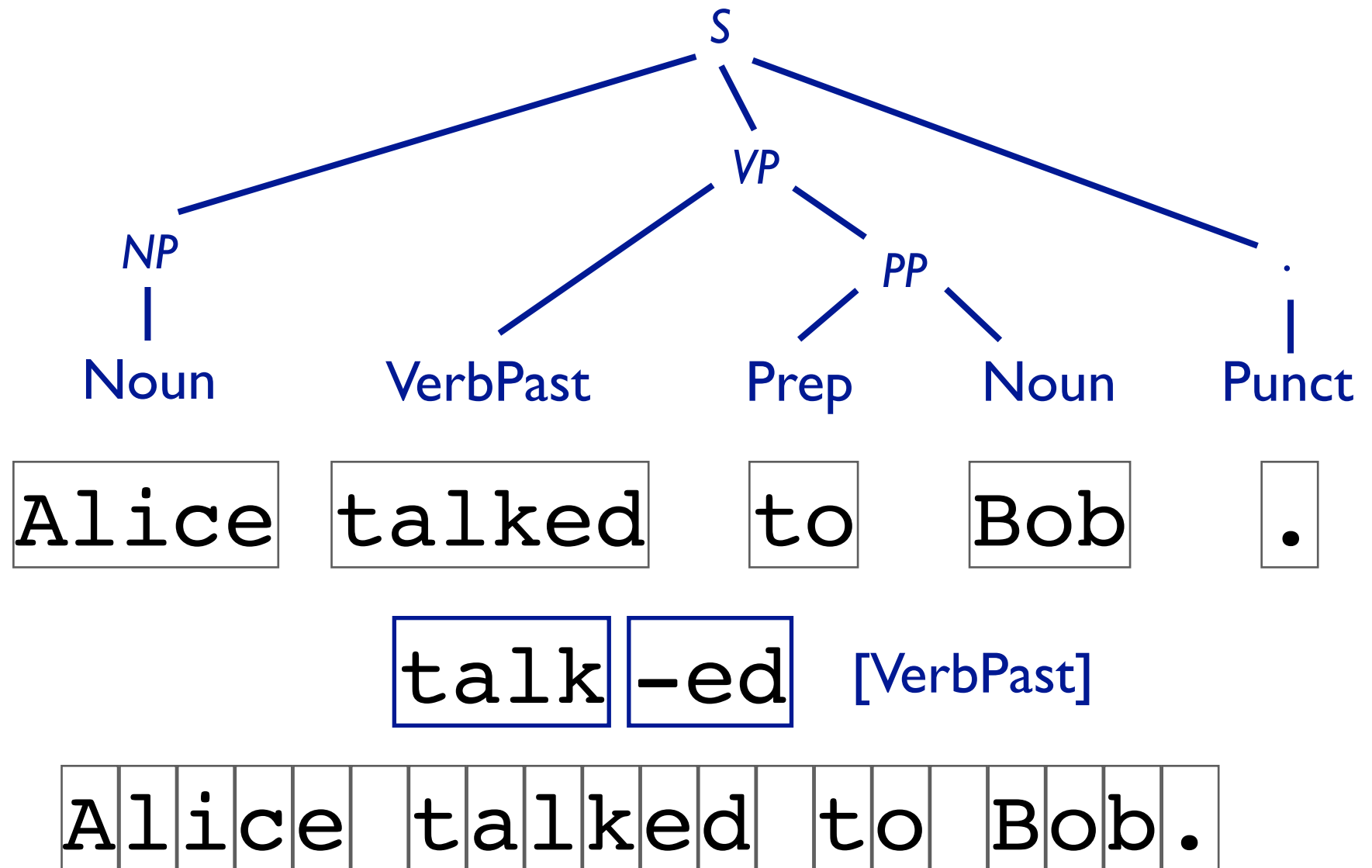
Syntax: Constituents

Syntax: Part of Speech

Words

Morphology

Characters



Levels of linguistic structure

Discourse

Semantics

Syntax: Constituents

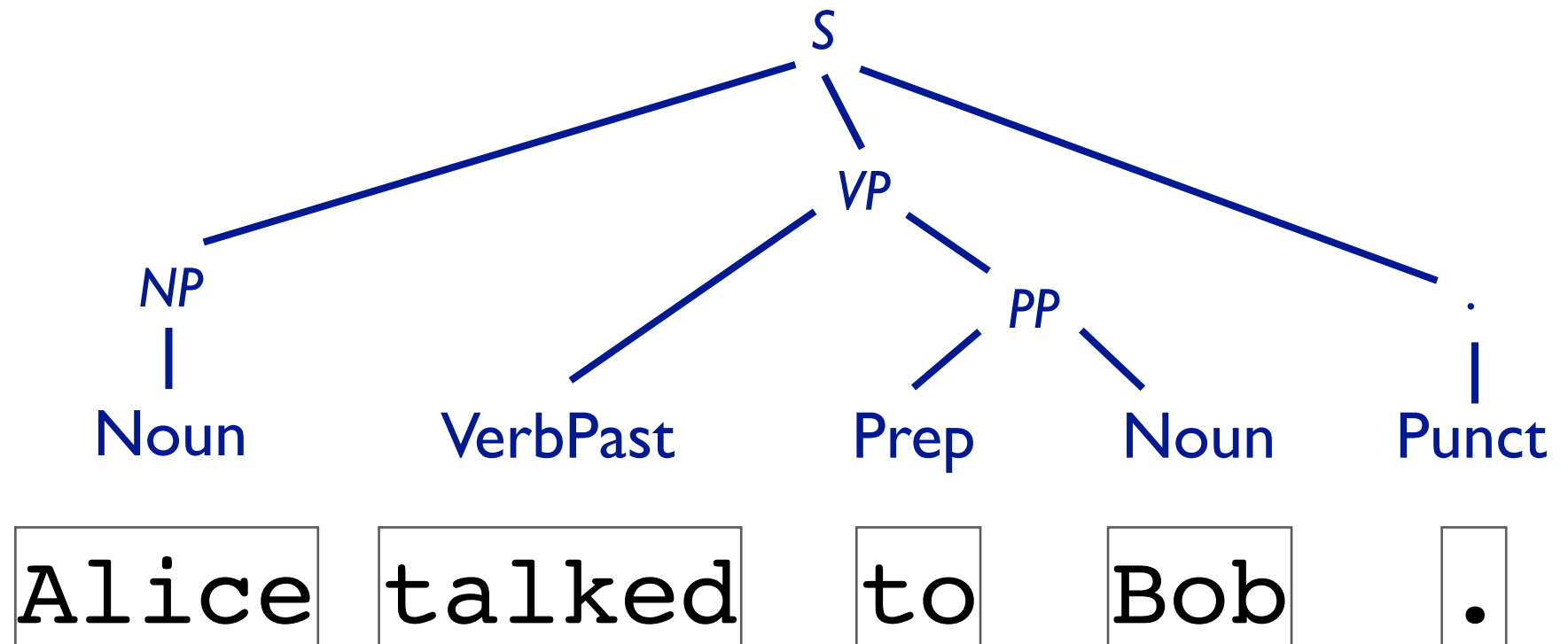
Syntax: Part of Speech

Words

Morphology

Characters

CommunicationEvent(e) SpeakerContext(s)
Agent(e, Alice) TemporalBefore(e, s)
Recipient(e, Bob)



talk -ed [VerbPast]

Alice talked to Bob.

supervised learning: given a collection of **labeled** examples (each example is a document X paired with a label Y), learn a mapping from X to Y

Tasks commonly tackled in a supervised setting:

- **Sentiment analysis:** map a product review to a sentiment label (positive or negative)
- **Question answering:** given a question about a document, provide the location of the answer within the document
- **Textual entailment:** given two sentences, identify whether the first sentence entails or contradicts the second one
- **Machine translation:** given a sentence in a source language, produce a translation of that sentence in a target language

self-supervised learning: given a collection of just text (no extra labels), create labels out of the text and use them for *representation learning*

- **Language modeling:** given the beginning of a sentence or document, predict the next word
- **Masked language modeling:** given an entire document with some words or spans masked out, predict the missing words

How much data can we gather for these tasks?

representation learning: given some text, create a representation of that text (e.g., real-valued, low-dimensional vectors) that capture its linguistic properties (syntax, semantics)

<i>word</i>	dim0	dim1	dim2	dim3
<i>today</i>	0.35	-1.3	2.2	0.003
<i>cat</i>	-3.1	-1.7	1.1	-0.56
<i>sleep</i>	0.55	3.0	2.4	-1.2
<i>watch</i>	-0.09	0.8	-1.8	2.9

transfer learning: **pretrain** a large self-supervised model, and then **fine-tune** it on a small downstream supervised dataset

- Transfer learning has recently (last ~2 years) become the method of choice for most downstream NLP tasks.
- Consequently, most of this class will focus on new research in transfer learning for NLP!

This course will be divided into ~4 high-level units, each of which will last 3 weeks

1. **Background:** language modeling and neural networks
2. **Transfer learning:** applications, modeling objectives, and analysis
3. **Text generation:** translation, paraphrasing, few-shot learning, the role of retrieval in generation
4. **Datasets, evaluation, security, and ethics,** and possibly other topics that I find interesting or you suggest!

This course will be divided into ~4 high-level units, each of which will last 3 weeks

1. **Background:** language modeling and neural networks
2. **Transfer learning:** applications, modeling objectives, and analysis
3. **Text generation:** translation, paraphrasing, few-shot learning, the role of retrieval in generation
4. **Datasets, evaluation, security, and ethics,** and possibly other topics that I find interesting or you suggest!

Be on the lookout for:

- Homework 0, to be released by Wednesday, due 9/4 (it's a math/coding review)
- Videos on language modeling, also to be released on Wednesday

Any technical issues? Registration issues?
Complaints or comments? Please use any of
{Piazza, instructors gmail, anonymous form,
or office hours} to let us know!

demos!
(allennlp.org)

UMass · CS685 | Advanced Natural Language Processing (2020)

CS685 (2020) · 课程资料包 @ShowMeAI



视频

中英双语字幕



课件

一键打包下载



笔记

官方笔记翻译



代码

作业项目解析



视频 · B 站 [扫码或点击链接]

<https://www.bilibili.com/video/BV1BL411t7RV>



课件 & 代码 · 博客 [扫码或点击链接]

<http://blog.showmeai.tech/umass-cs685>

NLP

迁移学习

语言模型 问答系统 文本生成 BERT

语义解析

知识推理

模型蒸馏

transformer

GPT-3

注意力机制

Awesome AI Courses Notes Cheatsheets 是 [ShowMeAI](#) 资料库的分支系列，覆盖最具知名度的 **TOP50+** 门 AI 课程，旨在为读者和学习者提供一整套高品质中文学习笔记和速查表。

点击课程名称，跳转至课程**资料包**页面，**一键下载**课程全部资料！

机器学习	深度学习	自然语言处理	计算机视觉
Stanford · CS229	Stanford · CS230	Stanford · CS224n	Stanford · CS231n
# Awesome AI Courses Notes Cheatsheets · 持续更新中			
知识图谱	图机器学习	深度强化学习	自动驾驶
Stanford · CS520	Stanford · CS224W	UCBerkeley · CS285	MIT · 6.S094



微信公众号

资料下载方式 2：扫码点击**底部菜单栏**

称为 **AI 内容创作者**？回复 [添砖加瓦]