

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

语言模型 问答系统 文本生成 BERT

语义解析

模型蒸馏

GPT-3

注意力机制

迁移学习

知识推理

transformer

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

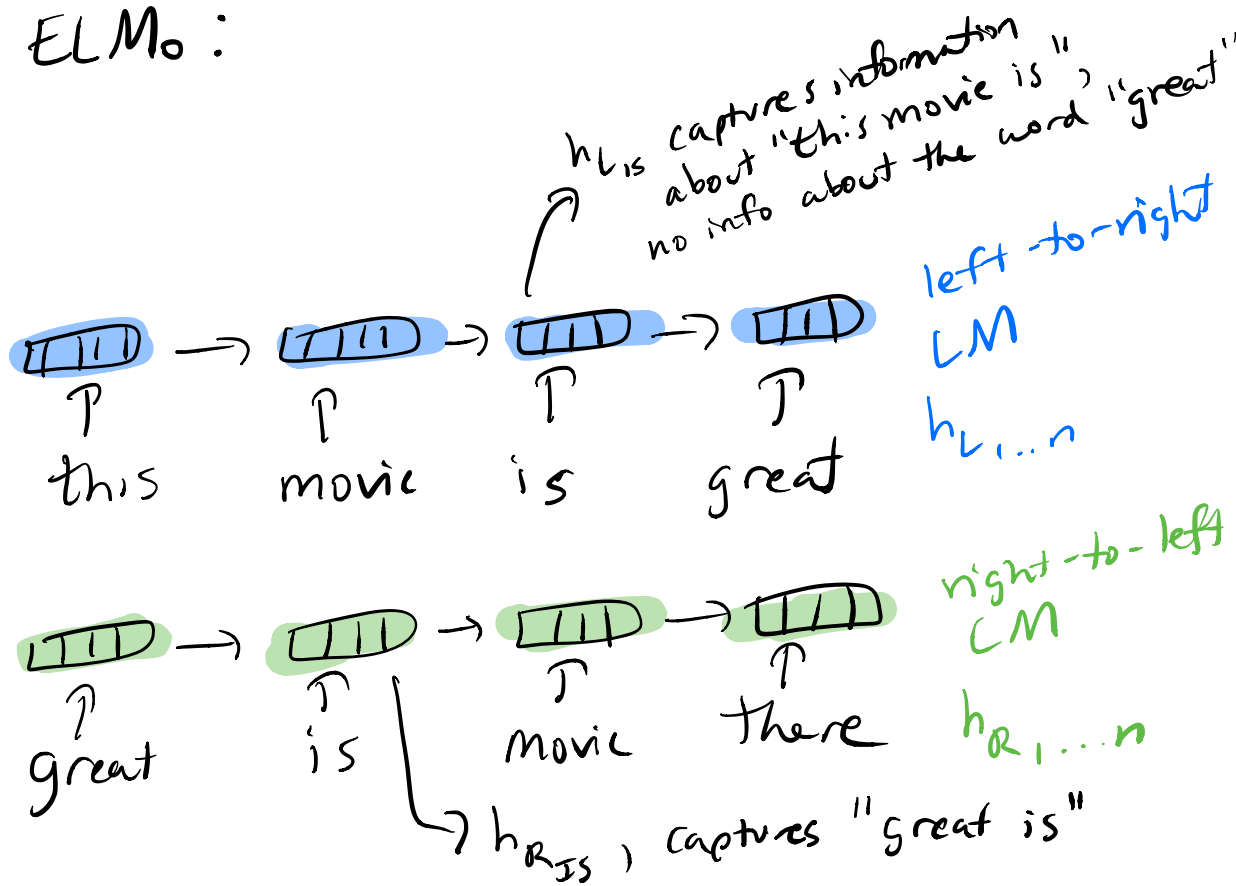
资料下载方式 2: 扫码点击**底部菜单栏**称为 **AI 内容创作者?** 回复 [添砖加瓦]

Today: - from ELMo to BERT

- from language modeling to masked LM

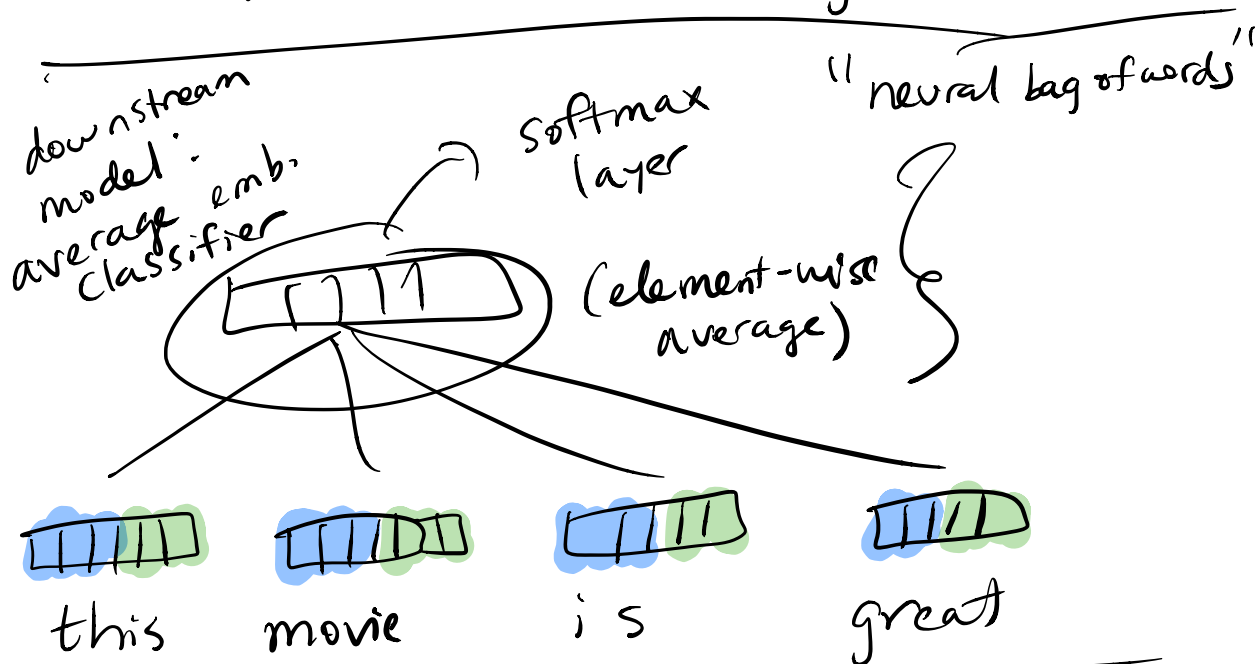
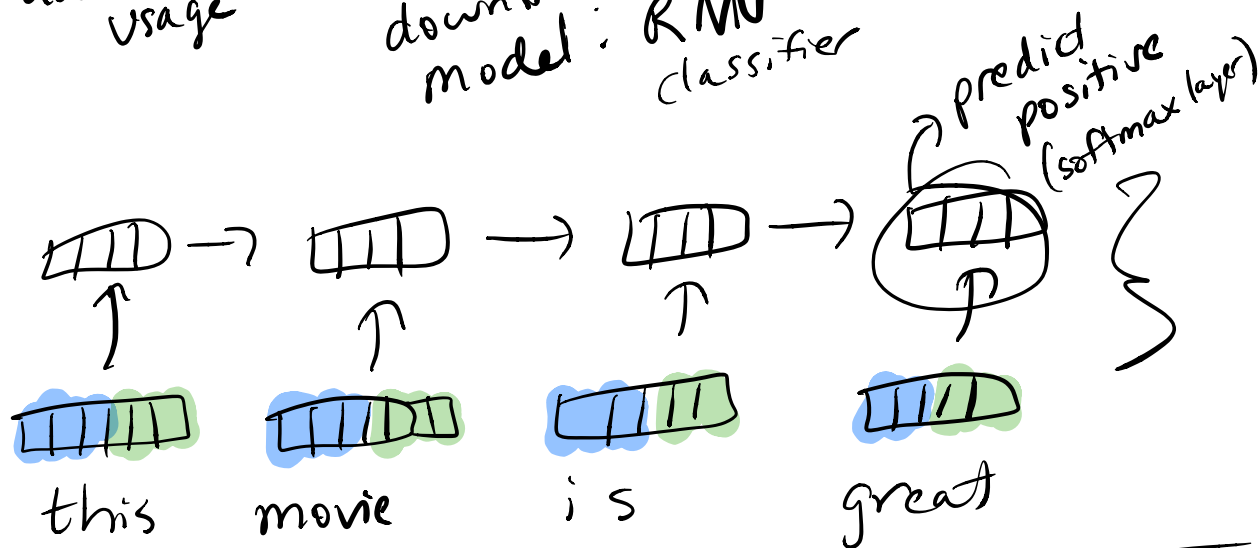
Goal of pretraining: use these big LMs as **text encoders**. Their goal is to enable downstream models to focus on the task at hand, instead of learning how language works.

ELMo:



downstream
usage

downstream
model: RNN
classifier



ex:

(in response to
question)

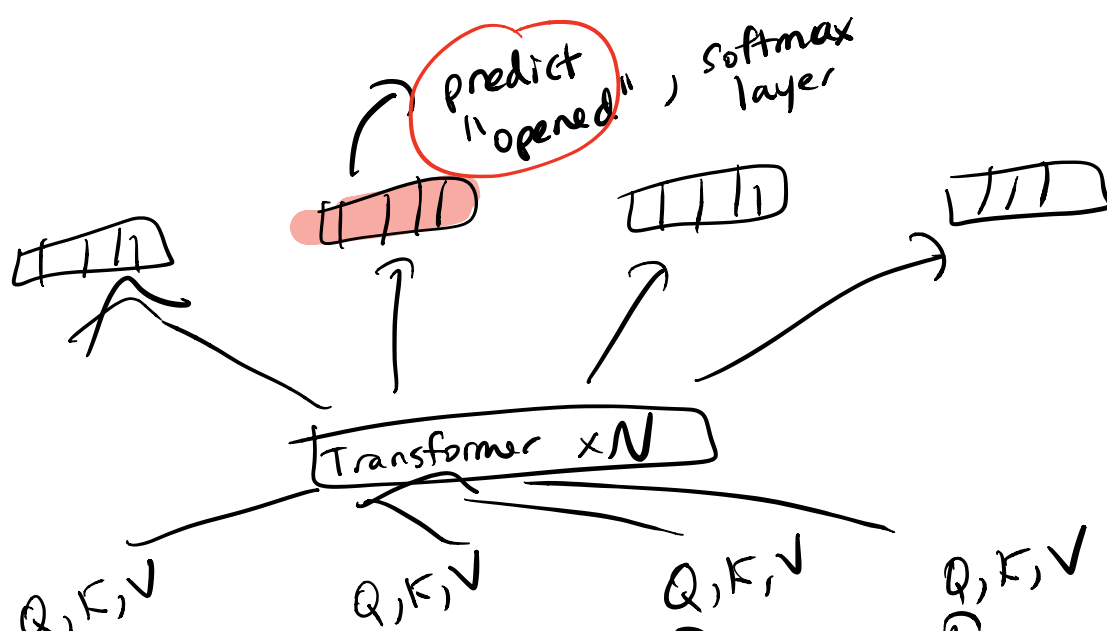
i loved the acting, but, (the rest of the
movie was terrible).

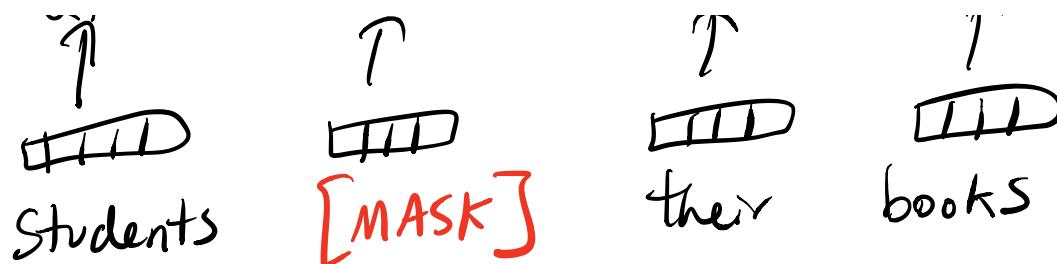
the ELMo approach of two separate LMs that are then concat together is a little hacky...

- can we accomplish the same goal within a single model
- change pretraining obj.
 - from LM to masked LMs

masked LM:

- given a full sequence of words (not just prefix) where $X\%$ of the words have been masked out
- instead of predicting the next word, we only predict masked words



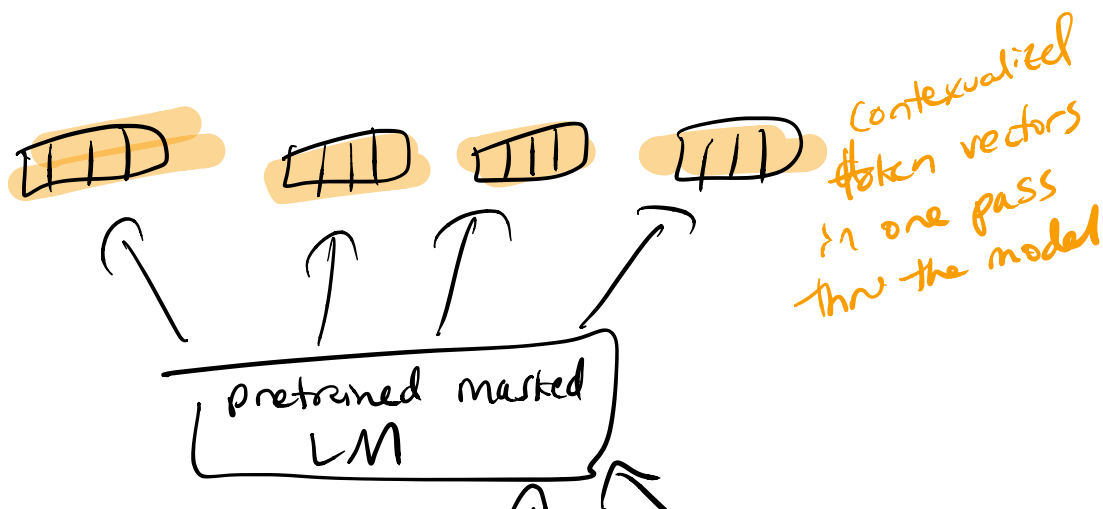


- all of the final layer representations are fully contextualized
 - "aware" of words in the past as well as words in the future

- same training loss as NLMs
 - minimizing neg. log likelihood of the ground-truth (unmasked) tokens

ELMo → BERT:

- 2 unidirectional LMs → 1 masked LM
- recurrent models to Transformers
- BERT was pretrained on a LOT more data

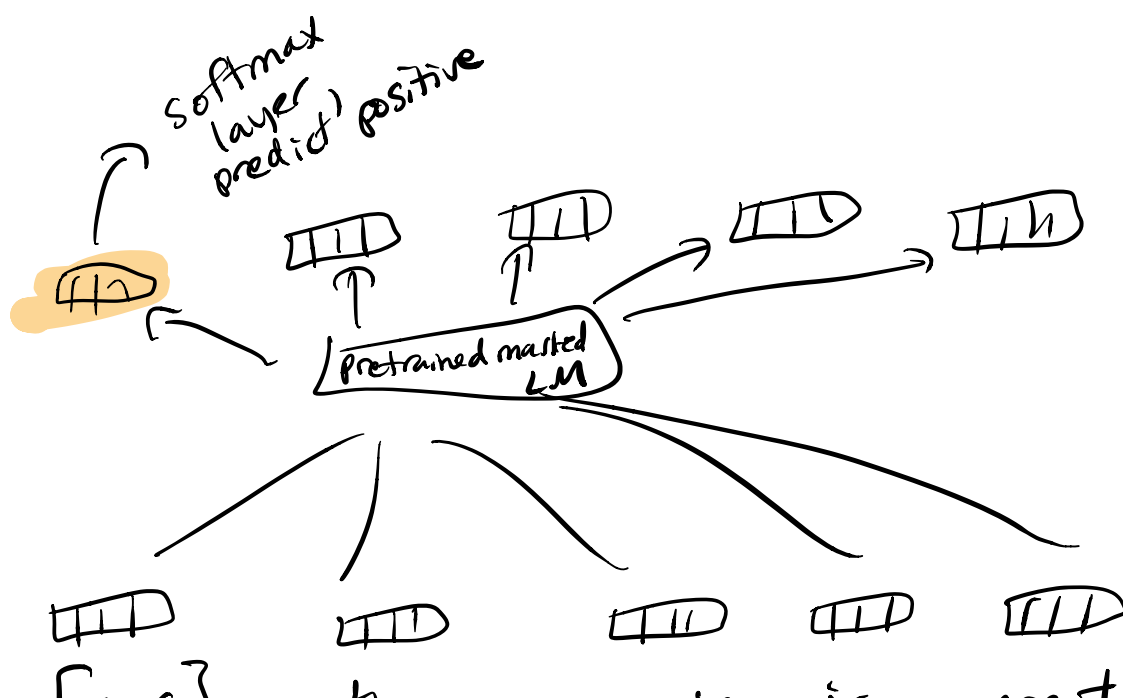


the movie is great

how do we use BERT for a downstream task? the pretrained architecture is almost the same as the downstream model

e.g. sentiment analysis

- add a special token to the beginning of every sequence
 - [CLS] token



[CLS] This movie is great

- backprop the error signal from the sentiment classifier through the entire pretrained masked LM
 - "fine-tuning"
- no external downstream model
 - only new component is a single softmax layer

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

语言模型 问答系统 文本生成 BERT

语义解析

模型蒸馏

GPT-3

注意力机制

迁移学习

知识推理

transformer

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]