

本期论文主题:Elmo

导师: Yamada



《Deep contextualized word representations》

基于深度上下文的词表征

作者: Matthew E. Peters

单位: Allen Institute for Artificial Intelligence

发表会议及时间: NAACL, 2018

前期知识储备

Pre-knowledge reserve



概率论

了解基本的概率论知识，
掌握条件概率的概念和公式

语言模型

掌握语言模型的原理，了解语言模型的评价标准

Char CNN

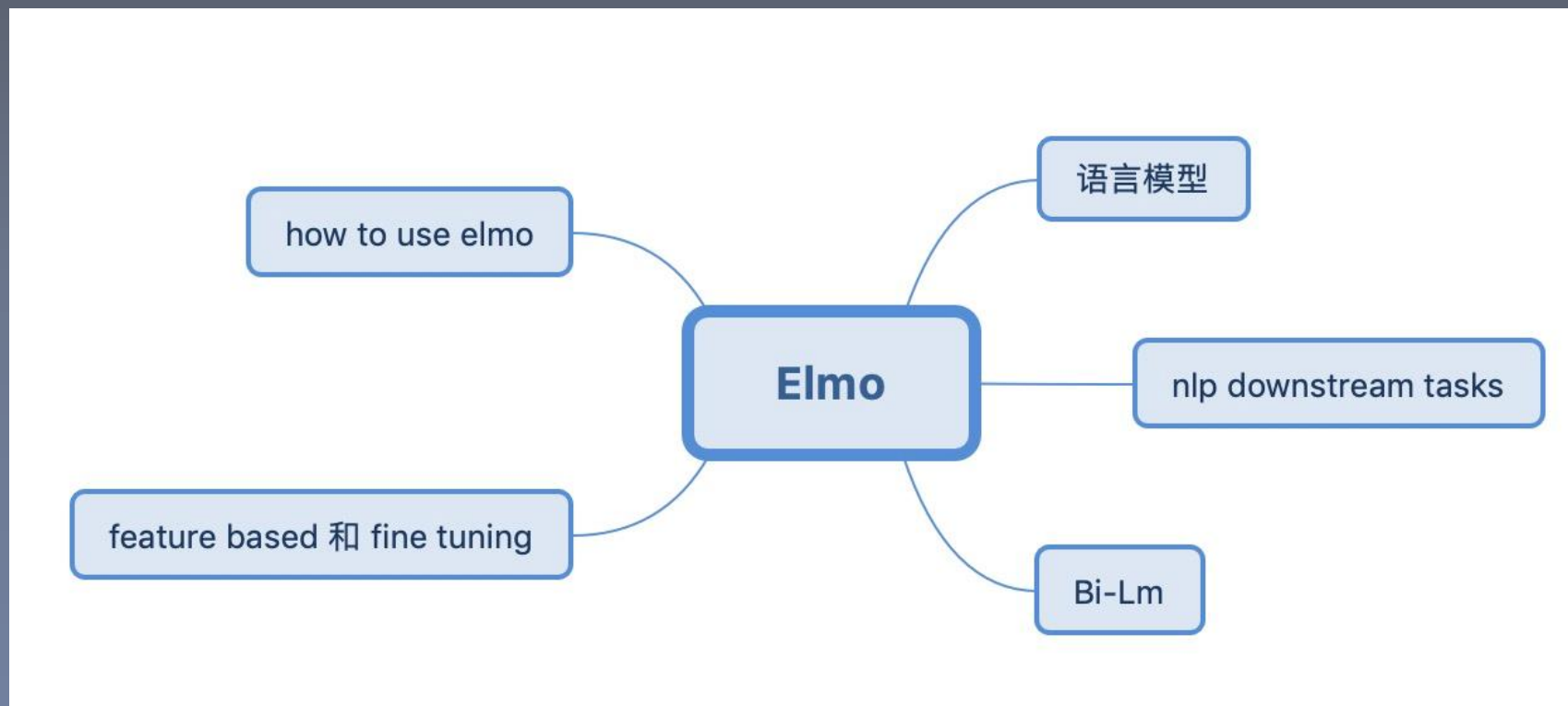
掌握Char CNN的基本工作原理。

注意力机制

了解注意力机制的思想，
掌握注意力机制的分类和实现方式

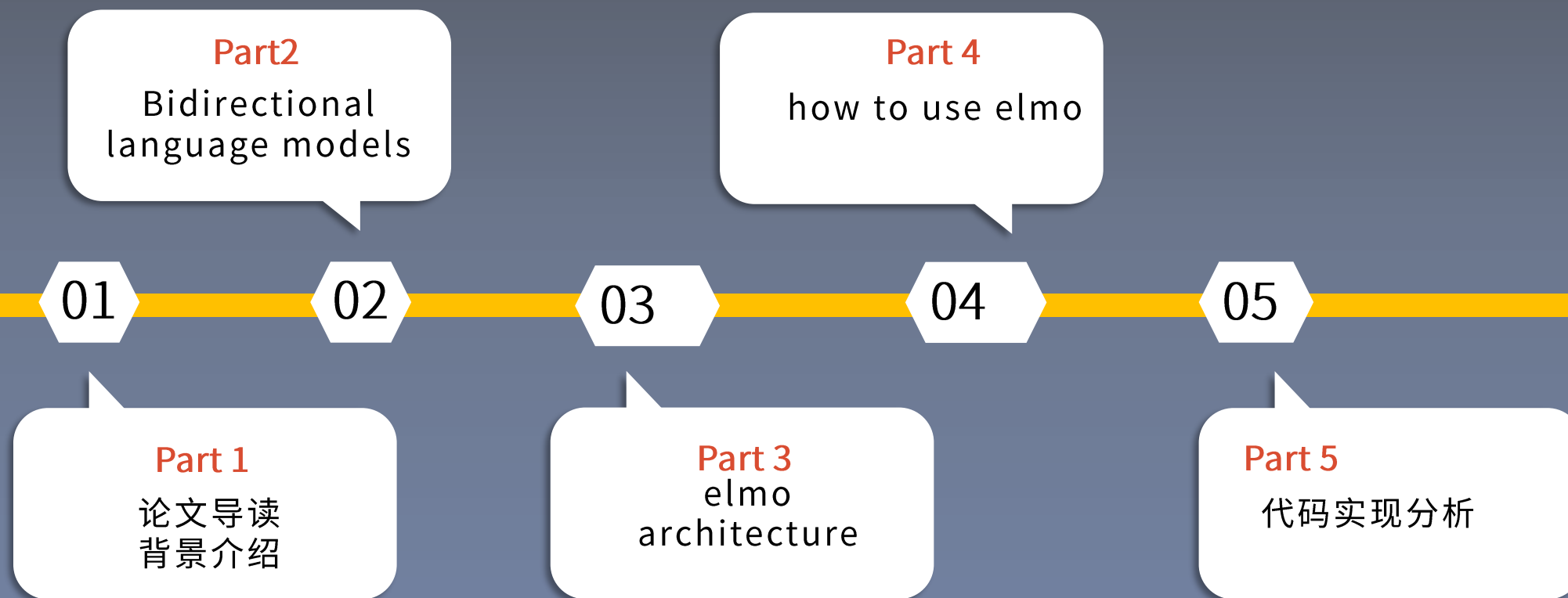
学习目标

Learning objectives



课程安排

The schedule of course



第一课：论文导读

The first lesson: the paper guide

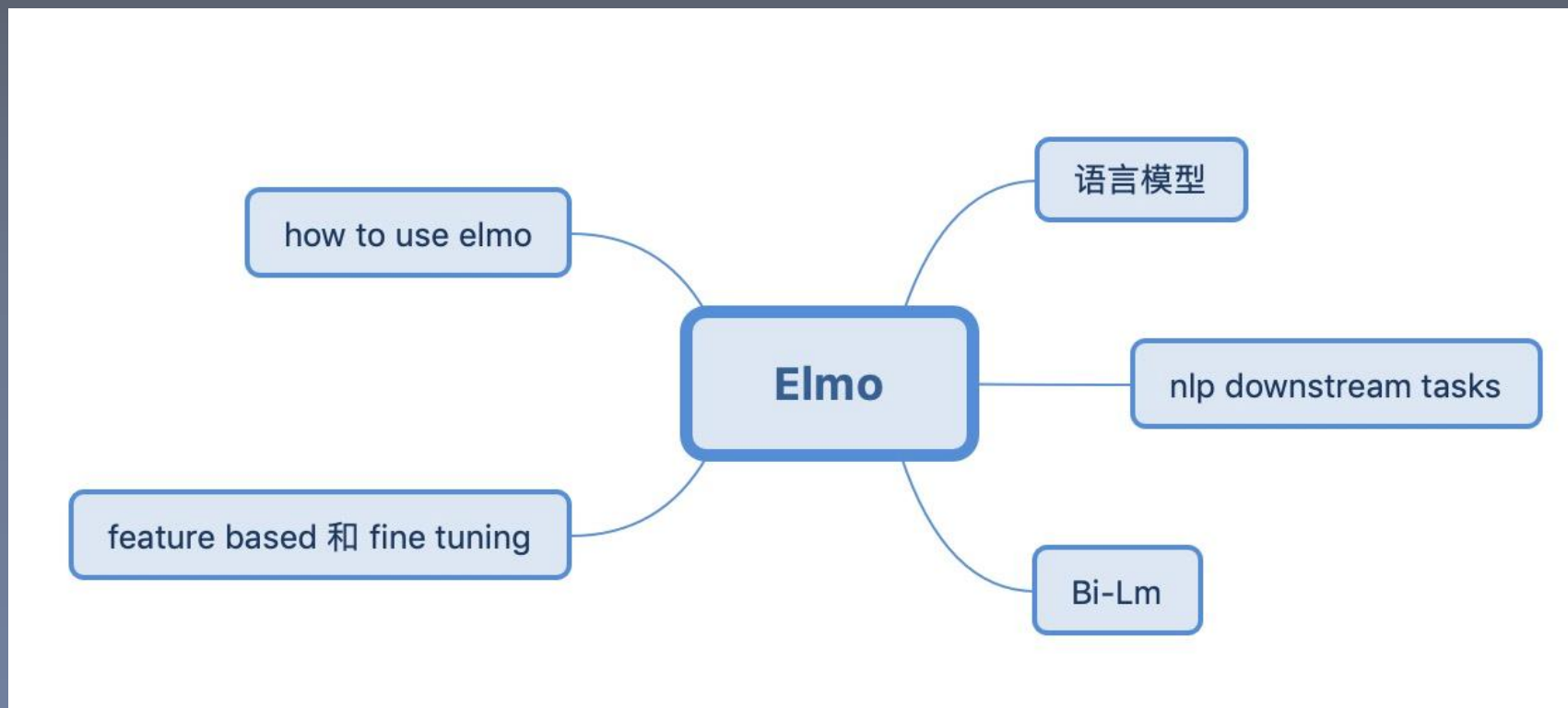
目录

1/ 论文研究背景、成果及意义

2/ 论文泛读

3/ Word2vec 和Char CNN回顾

4/ 本课回顾及下节预告



论文研究背景、成果及意义

研究背景

Research background



深度之眼
deepshare.net



重点 重点来了!

nlp下游任务

SQuAD:阅读理解任务

SNLI:自然语言推理任务

SRL:语义角色标注任务

Coref:指代消解任务

NER:命名实体识别任务

SST-5:情感分析任务

研究背景

Research background

语言模型

给定句子(词语序列): 今天早上我去食堂吃饭电视

$$S = W_1, W_2, \dots, W_k$$

语言模型概率:

$$P(S) = P(W_1, W_2, \dots, W_k) = p(W_1)p(W_2|W_1)\dots P(W_k|W_1, W_2, \dots, W_{k-1})$$

研究背景

Research background



深度之眼
deepshare.net



重点 重点来了!

Feature-Based and Fine-tuning

Feature-Base(特征提取):

- 1: 首先在big data A上训练语言模型，训练完毕得到语言模型（用作embedding）。
- 2: 然后构造task-specific model，采用有label的语料B训练该Model, **将语言模型的参数固定**，语料B的训练数据经过语言模型得到LM embedding，作为task-specific model的额外特征。

Fine-tuning(微调):

- 1: 构造语言模型，采用big data A来训练语言模型。
- 2: 在语言模型基础上增加少量神经网络层来完成specific task，然后采用有label的语料B来训练模型，这个过程中语言模型的**参数不固定**，**依然是trainable variables**。

研究背景

Research background

Feature-Based and Fine-tuning

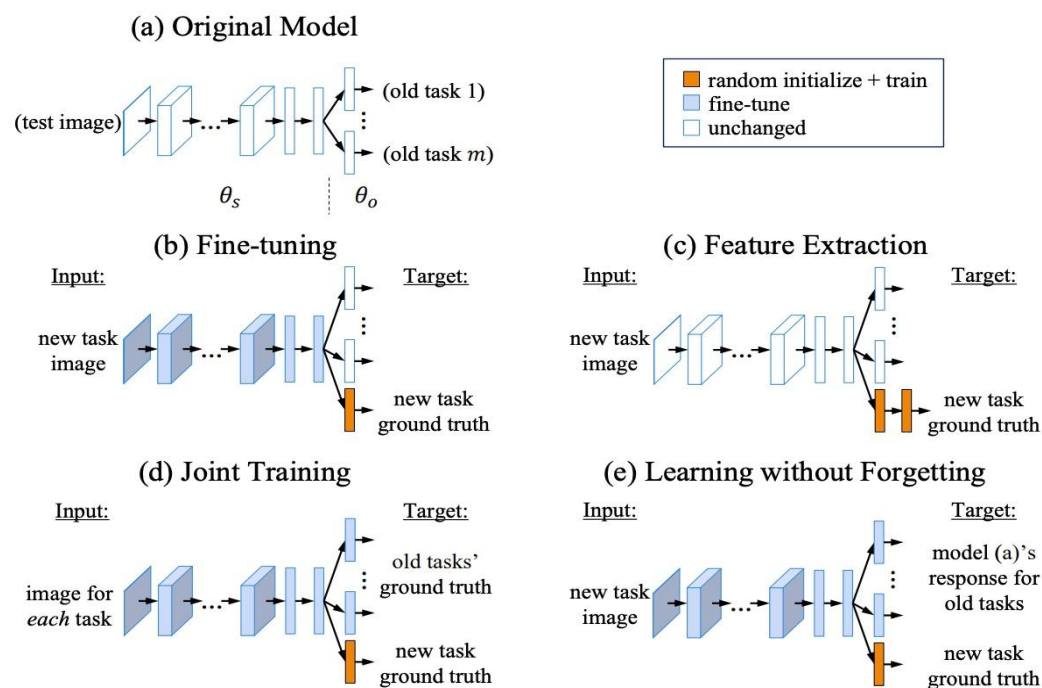
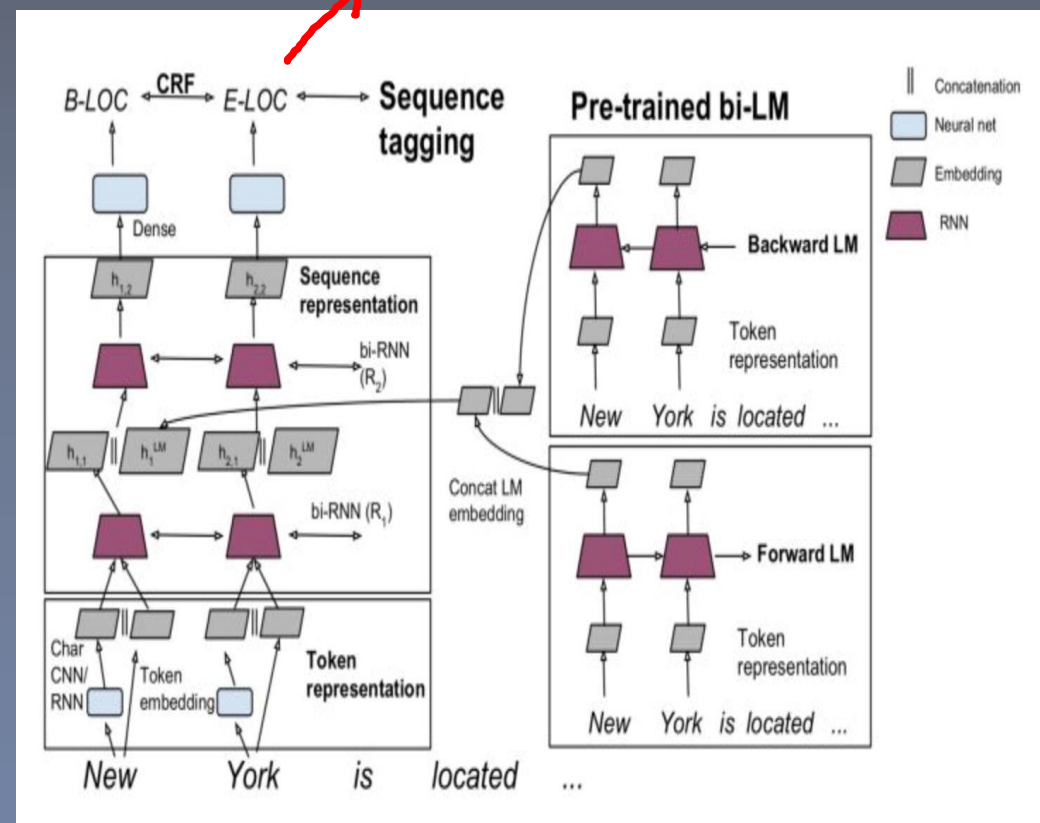


Fig. 2. Illustration for our method (e) and methods we compare to (b-d). Images and labels used in training are shown. Data for different tasks are used in alternation in joint training.



研究成果

Research Results

E > 1 M

TASK	PREVIOUS SOTA		OUR <u>BASELINE</u>	ELMo + BASELINE	INCREASE (ABSOLUTE/ RELATIVE)
<u>SQuAD</u>	Liu et al. (2017)	<u>84.4</u>	81.1	<u>85.8</u>	4.7 / 24.9%
<u>SNLI</u>	Chen et al. (2017)	88.6	88.0	88.7 ± 0.17	0.7 / 5.8%
SRL	He et al. (2017)	81.7	81.4	84.6	3.2 / 17.2%
Coref	Lee et al. (2017)	<u>67.2</u>	67.2	<u>70.4</u>	3.2 / 9.8%
NER	Peters et al. (2017)	<u>91.93 ± 0.19</u>	90.15	<u>92.22 ± 0.10</u>	2.06 / 21%
SST-5	McCann et al. (2017)	53.7	51.4	54.7 ± 0.5	3.3 / 6.8%

在6项nlp下游任务中都取得了不俗的表现。



研究意义

Research Meaning



重点 重点来了!

Elmo历史意义

- 提出了动态词向量，能让词语学习到当前语境信息。
- 拉开了预训练模型的序幕

nlp领域

模型的词向量表征

Word2vec为代表

2018

Elmo

nlp领域

提出elmo这种动态词向量

研究意义

Research Meaning



深度之眼
deepshare.net



重点 重点来了!

Elmo历史意义

- 提出了动态词向量，能让词语学习到当前语境信息。
- 拉开了预训练模型的序幕

自从elmo模型提出来后，预训练模型正式被nlp各项下游任务中采用，下游任务中的各种模型都被预训练模型+给替代。

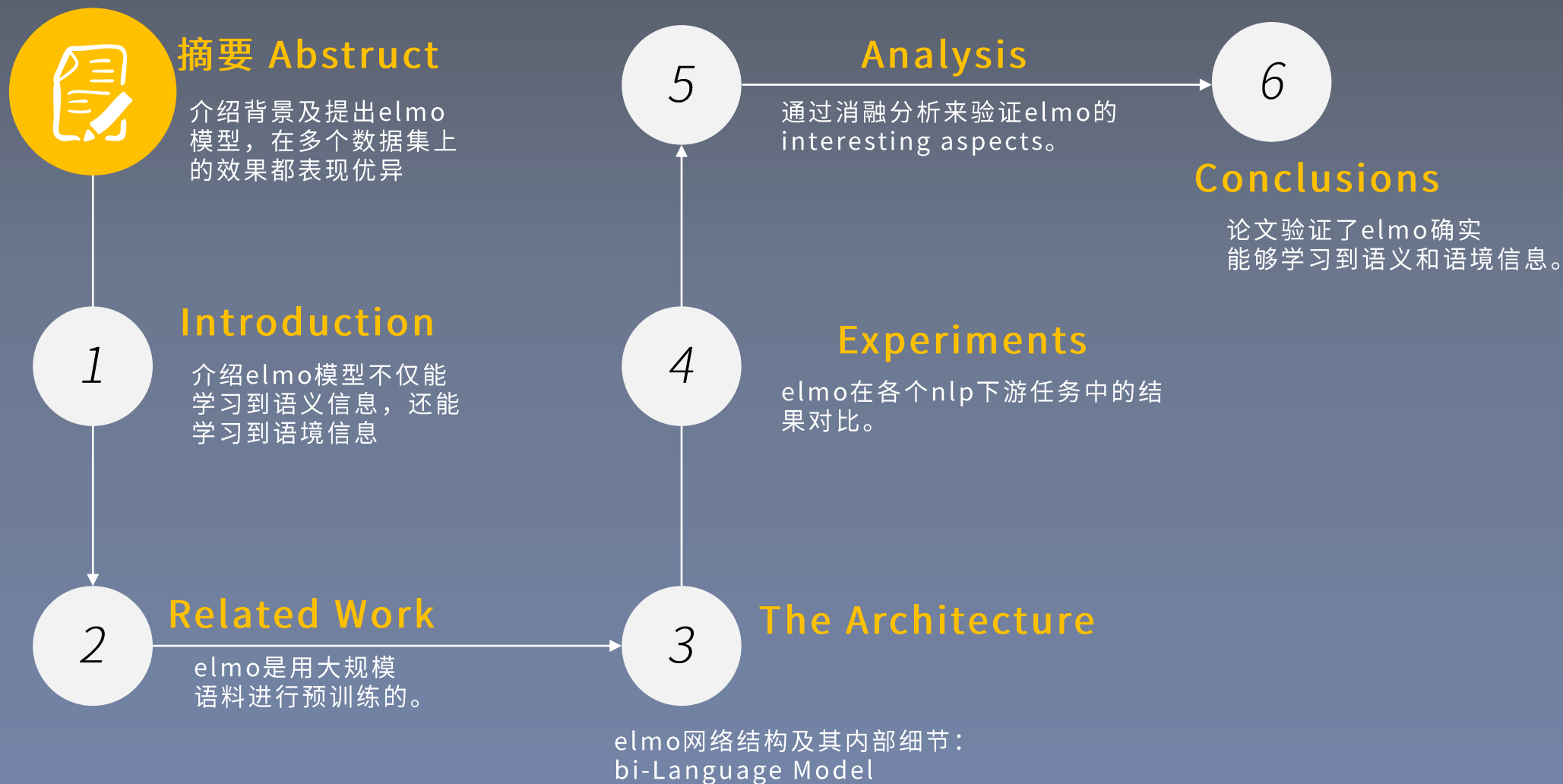


论文泛读

Strcuture of Paper

论文结构

Structure of Papers



摘要

abstract

摘要核心

1. 我们介绍了一种新的词向量表征可以解决词语义特征和语境特征。
2. 我们的向量是在大规模语料通过bidirectional language model学习到的。
3. elmo模型在各项nlp下游任务中都表现得良好。
4. 暴露预训练的深层内部是至关重要的，允许下游模型混合不同类型的半监督信号。

论文小标题

Paper title

1. Introduction

2. Related Work

3. ELMo: Embeddings from Language Models

3.1 Bidirectional Language Models

3.2 ELMo

3.3 Using biLMs for supervised NLP tasks

3.4 Pre-trained bidirectional language
model architecture

4. Evaluation

5. Analysis

5.1 Alternate layer weighting schemes

5.2 Where to include ELMo

5.3 What information is captured by the
biLM representations

5.4 Sample efficiency

5.5 Visualization of learned weights

6. Conclusion

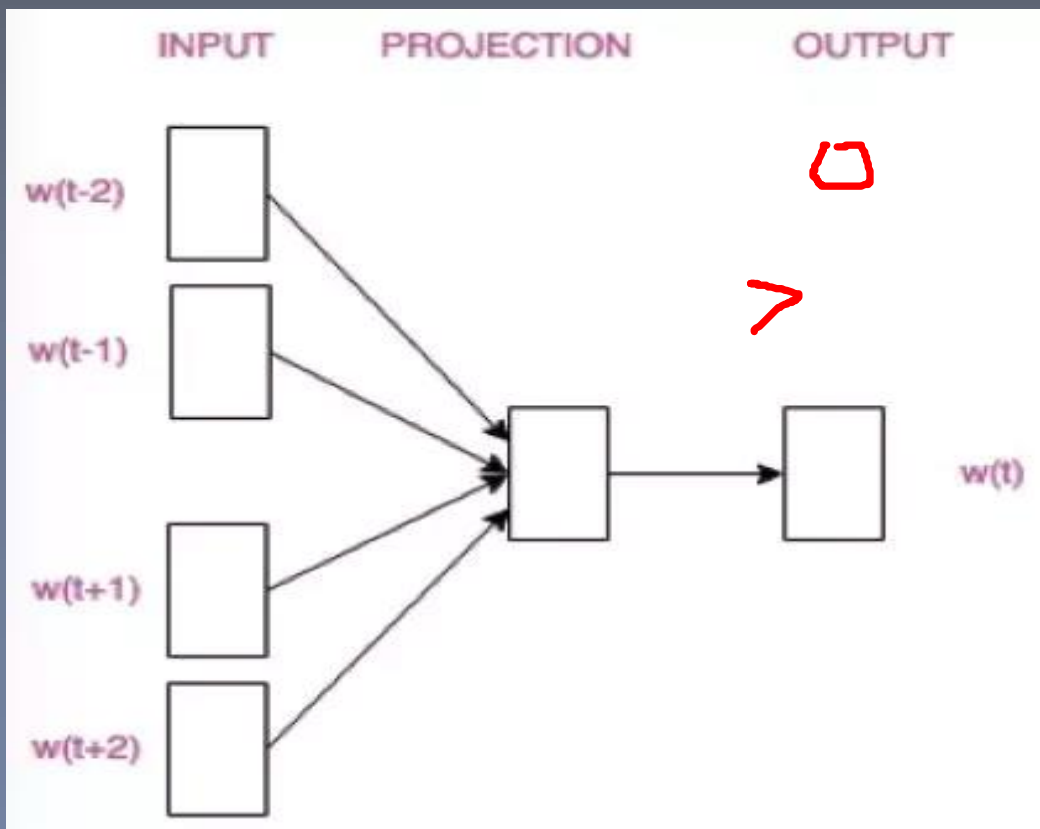
Word2vec以及 Char CNN的回顾

Strcuture of Paper

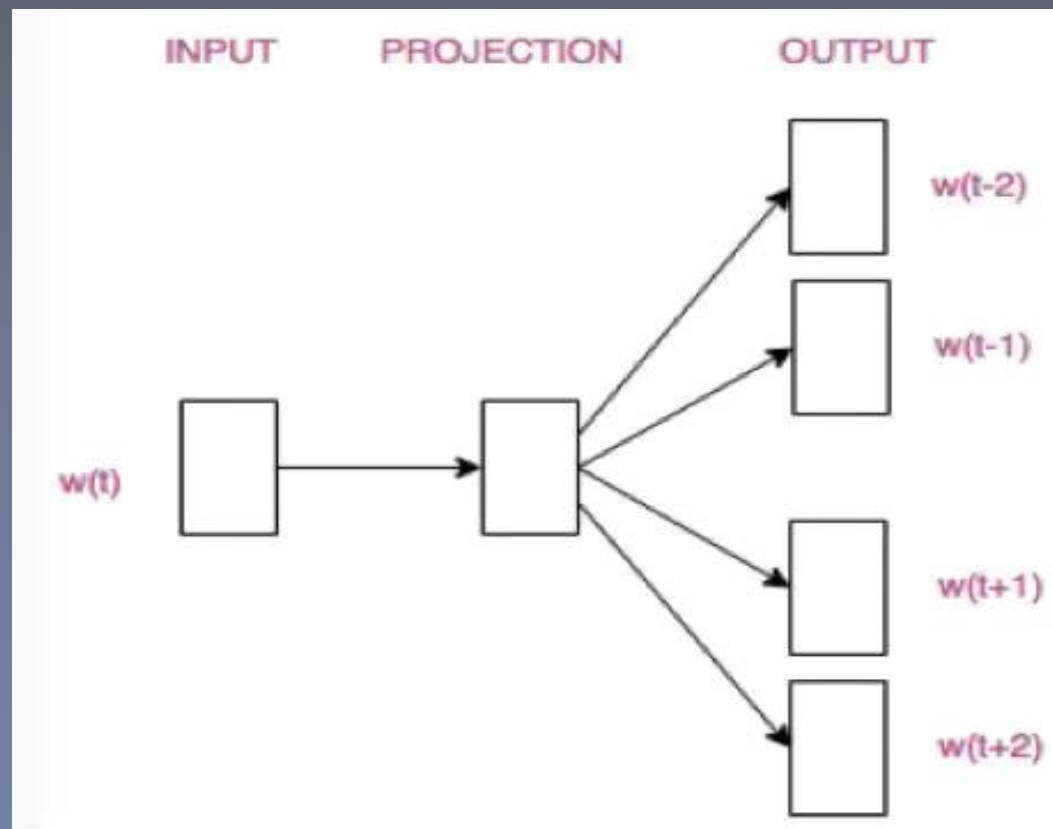
论文结构

Structure of Papers

Word2vec



cbow

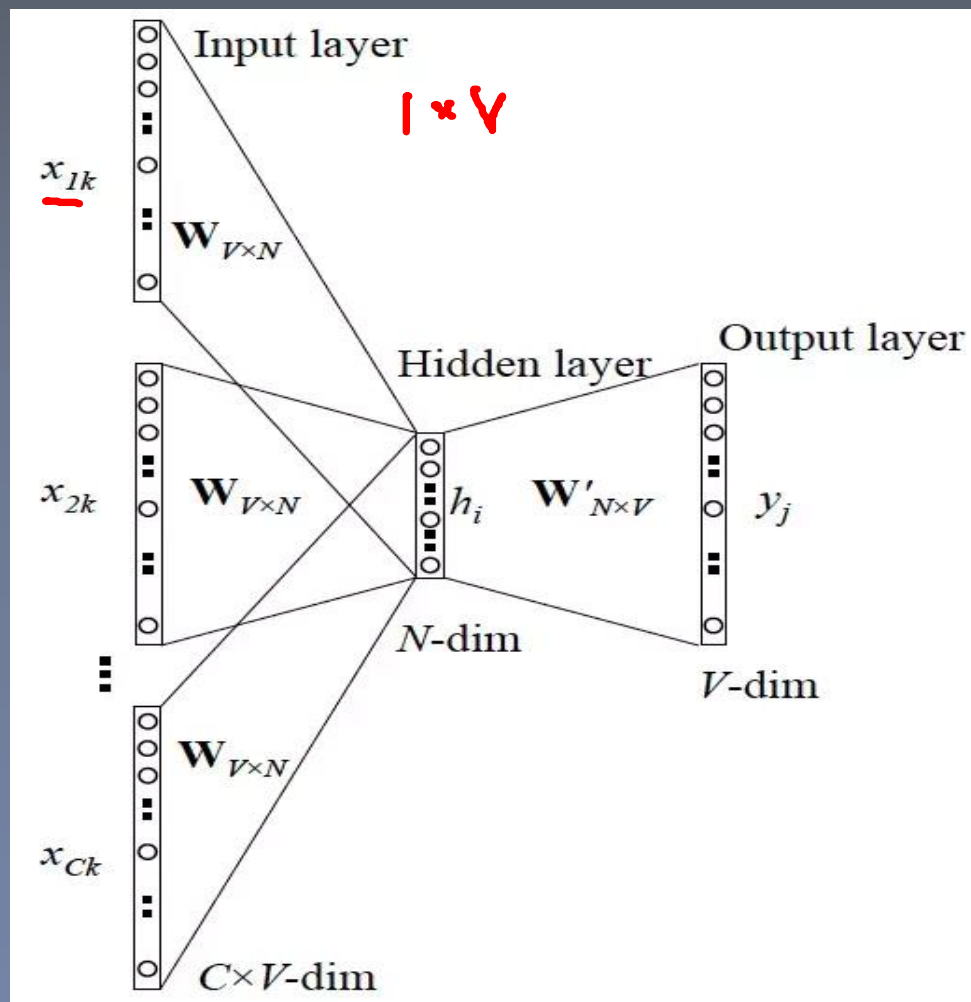


skip-Gram

论文结构

Structure of Papers

Word2vec



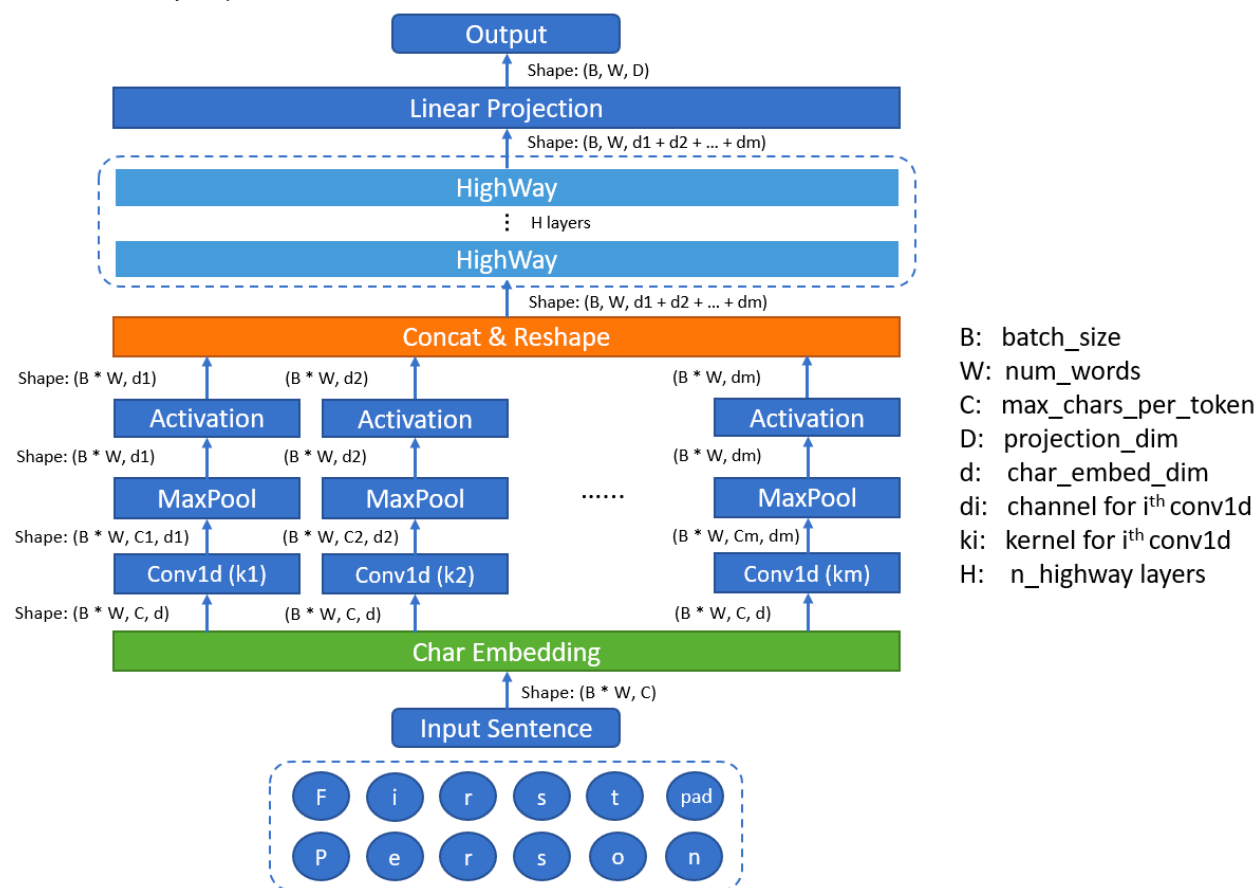
1. 输入层:上下文单词的onehot {假设单词向量空间dim为 V , 上下文单词个数为 C }。
2. 所有onehot分别乘以共享的输入权重矩阵 W { VN 矩阵}。
3. 所得的向量concat平均作为隐层向量, size为 $1 \times N$
4. 乘以输出权重矩阵 $W1$ { $N \times V$ }。
5. 得到向量 $1 \times V$, 映射到词典, 直接过softmax。

论文结构

Structure of Papers

Char CNN

Char Encode Layer架构



本课回顾及下节预告

Review in the lesson and Preview of next lesson

本课回顾

Review in the lesson



01 研究背景及成果意义

学习了nlp下游任务以及概念feature-based和fine-tuning、了解了论文的实验结果。

02 论文总览

论文总共包含6个部分，论文主要介绍elmo的结构。

03 回顾Word2vec以及Char CNN

回顾了Word2vec的流程以及学习了Char CNN的结构。

下节预告

Preview of next lesson



01 Bidirectional Language Model

学习Bidirectional Language Model

02 elmo

学习elmo的主要结构，了解怎么使用elmo结构。

03 实验设置及结果分析

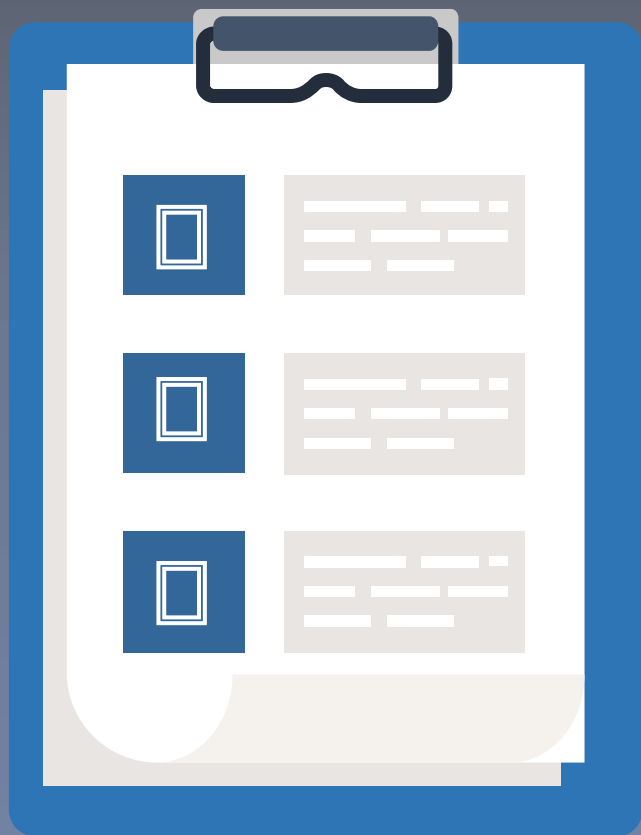
比较了模型在几个数据集上的表现情况。

04 论文总结

总结论文中创新点、关键点及启发点

下节课前准备

Preview of next lesson



- 下载论文
- 泛读论文
- 筛选出自己不懂的部分，带着问题进入下一课时

—— 结 语 ——

循循而进，欲速则不达也。





深度之眼
deepshare.net

联系我们：

电话：18001992849

邮箱：service@deepshare.net

QQ：2677693114



公众号



客服微信

