

## 第四章 自然语言分类任务



扫码试看/订阅

《NLP 实战高手课》视频课程

## 4.1 重新审视 Word Embedding

# 内容概述

- Negative Sampling
- Word Embedding 的一些问题
- 早期解决 Word Embedding 问题的一些方法

# Negative Sampling

# Word Embedding 的问题

# 早期解决 Word Embedding 问题的一些方法

- 加入 POS
- Context Embedding
  - <https://www.aclweb.org/anthology/K16-1006.pdf>
- 通过翻译加入上下文因素
  - <https://papers.nips.cc/paper/7209-learned-in-translation-contextualized-word-vectors.pdf>

# 补充材料

- 关于 Negative Sampling 的推导:
- <https://arxiv.org/pdf/1402.3722.pdf>



## 4.2 深度迁移学习模型：从 ELMo 到 BERT

# 内容概述

- 深度迁移学习模型的核心思想
- ELMo
- BERT

# 深度迁移学习的核心思想

- 解决 Word Embedding 不足
- 充分利用无标注的数据
- 使用较深的模型

# ELMo

- 架构和任务
- 使用方法

# BERT

- 架构和任务
- 使用方法

## 4.3 深度迁移学习模型：RoBERTa、XLNet、ERNIE 和 T5

# 内容综述

- RoBERTa
- XLNet
- ERNIE
- T5

# RoBERTa

- 核心思想：
  - 通过更好地训练 BERT 可以达到超过其他新的预训练语言模型的效果
- 核心改动：
  - 更大的 Batch Size
  - 去掉 Next Sentence Prediction
  - 采用更大的预训练语料
  - Dynamic Masking



# XLNet

- 主要改动
  - Permutation Language Modeling
  - Transformer-XL
- XLNet 还在很多地方进行了改动

# ERNIE

- 核心思想：
  - 使用 Multi-task Learning 提升模型效果

# T5

- 核心思想：
  - 将 Multi-task Learning 的任务变为 Seq2Seq 的任务
  - 测试了各类预训练语言模型设定的有效性

## 4.4 深度迁移学习模型：ALBERT 和 ELECTRA

# 内容综述

- 核心思想
- ALBERT
- ELECTRA

# 核心思想

- 只要是增加预训练模型的复杂度，人们往往可以得到更好的效果
- 但是：
  - 更大的模型意味着更大的算力消耗
  - 更大的模型意味着预训练语言模型过程的更大消耗
- 结论：应该在同等复杂度上进行模型比较

# ALBERT

- 核心思想
  - 权重共享
  - 输入层的优化
  - Sentence Order Prediction

# ELECTRA

- 核心思想：采用对抗训练提升模型训练效果
  - 通过 MLM 训练 Generator
  - Discriminator 负责区分 Generator 生成的 token 是否被替代
- 其他改进：
  - 采用权重共享





扫码试看/订阅

《NLP 实战高手课》视频课程