

阿里云天池牛年读书会

# 神经网络与深度学习

分享嘉宾：邱锡鹏教授  
复旦大学计算机学院

# 天池读书会

TIANCHI 天池



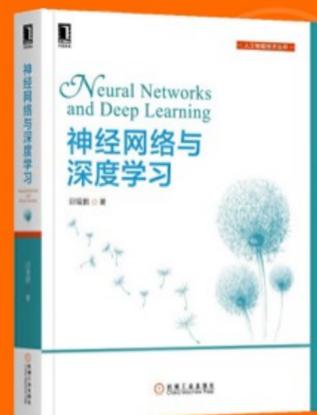
机械工业出版社  
华章公司

## 《神经网络与深度学习》

本书主要介绍神经网络与深度学习中的基础知识、主要模型、以及在计算机视觉、自然语言处理等领域的应用。

直播嘉宾：邱锡鹏

直播时间：3月25日20:00 ~ 21:00

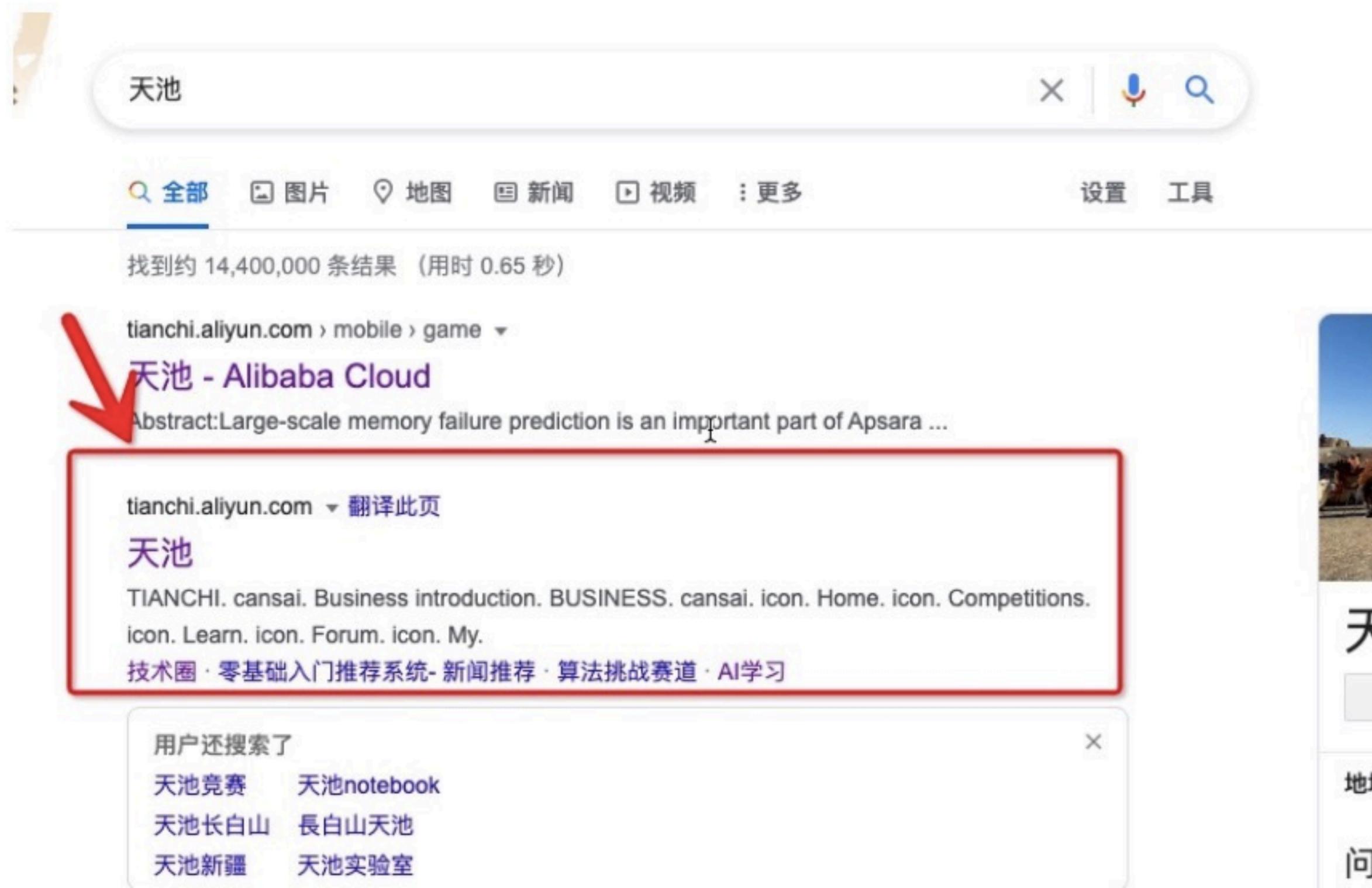


扫码领取  
读书会配套学习资源

bookclub



1) 首先需要进入天池官网，大家打开浏览器，搜索 天池，找到 tianchi.aliyun.com即可访问进入天池官



2) 在天池官网，将鼠标移到 天池学习 ，即可出现下拉列表，点击 天池读书会 ，即可进入天池读书会的页面。



3) 在天池读书会页面，你可以对对应的读书会图书进行提问，优秀的提问还有机会获得赠书，还可以点击配套的训练营或者课程资源进入学习，还有点击实践代码获取读书会的项目实践的代码，跟着我一起进行项目实践和代码学习，同时还有很多其他的读书会，大家也可以观看举办过的读书会的回放，或者预约还没开始的读书会。



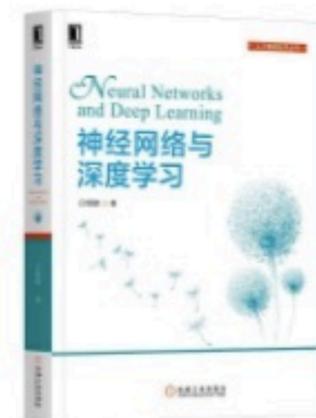
## 邱锡鹏 复旦大学计算机科学技术学院教授

直播主题 《神经网络与深度学习》

直播时间 2021年3月25日 20:00

学习资料 《神经网络与深度学习》课程

实践项目 fastNLP



[🗨️ 提问](#) | [📖 学习课程](#) | [🛒 购买地址](#) | [📄 PPT下载](#) | [👉 实践项目](#) | [📺 预约直播](#)

1. 分享嘉宾简介
2. 图书简介
3. 项目实践-fastNLP介绍
4. Q&A 答疑

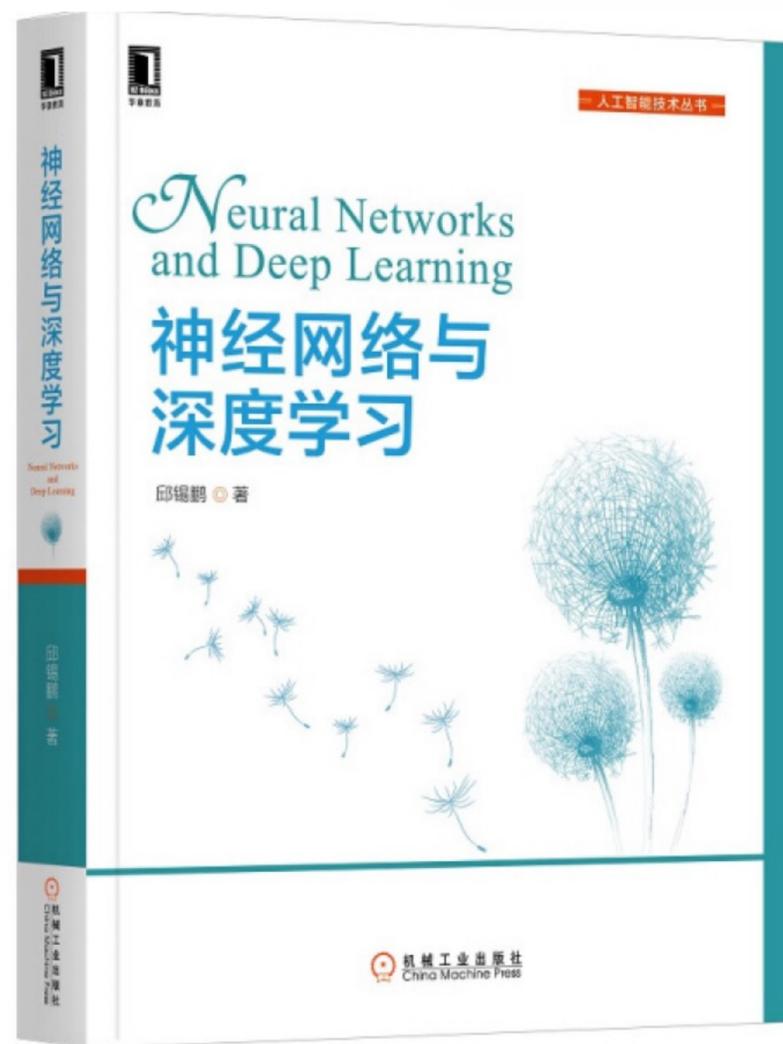


邱锡鹏

- 复旦大学教授、国家优青获得者
- 研究领域为**自然语言表示学习**，发表70余篇CCF-A/B类论文。获ACL 2017杰出论文奖、CCL 2019最佳论文奖、钱伟长中文信息处理科学技术奖青年创新一等奖，入选首届中国科协青年人才托举工程，2020年度AI 2000人工智能全球最具影响力提名学者。
- 主持开发了**自然语言处理开源系统FudanNLP和FastNLP**，被工业届和学术界广泛使用。
- 发表**开源教材《神经网络与深度学习》**，被多家高校选为教材。
- 培养**学生**曾获中国中文信息学会优博、中国人工智能学会优博、微软学者、百度奖学金、上海市优博等。

# 图书简介

TIANCHI 天池



邱锡鹏，神经网络与深度学习，机械工业出版社，2020。  
ISBN:9787111649687. <https://nndl.github.io/>



淘宝扫码



读者群

蒲公英封面:希望这本教材能够帮助更多的学生进入深度学习以及人工智能领域，他们会为人工智能领域注入新的生机与活力。

直播相关资料获取及回放查看地址：<https://tianchi.aliyun.com/specials/promotion/activity/bookclub>

# 最早的版本

邱锡鹏 V 

2015-12-11 22:08 来自 微博 weibo.com

分享最近写的《神经网络与深度学习讲义》，未完待续。也欢迎大家多提建议。

[神经网络与深度学习讲义20151211.pdf](#)



**神经网络与深度学习讲义20151211.pdf**  
分享者: 邱锡鹏

阅读 33.5万 推广 | 828 | 108 | 103

《神经网络与深度学习》讲义  
Notes on Artificial Neural Networks and Deep Learning

邱锡鹏  
xpqiu@fudan.edu.cn  
2015年12月11日

58 第七章 循环神经网络

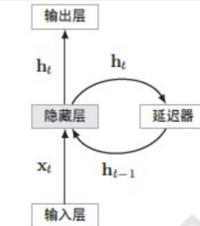


图 7.1: 循环神经网络

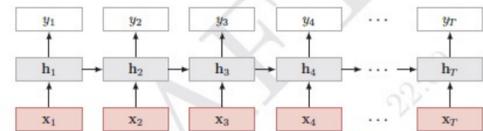


图 7.2: 按时间展开的循环神经网络

(long short term memory neural network, LSTM) [Hochreiter and Schmidhuber, 1997] 是训练神经网络的一个扩展。

### 7.1 简单循环网络

我们先来看一个非常简单的循环神经网络，叫简单循环网络 (Simple Recurrent Network, SRN) [Elman, 1990]。

假设时刻  $t$  时，输入为  $x_t$ ，隐层状态（隐层神经元活性）为  $h_t$ 。  $h_t$  不仅和当前时刻的输入相关，也和上一个时刻的隐层状态相关。

一般我们使用如下函数：

$$h_t = f(Uh_{t-1} + Wx_t + b), \quad (7.2)$$

这里，  $f$  是非线性函数，通常为 logistic 函数或 tanh 函数。

图 7.2 给出了按时间展开的循环神经网络。

邱锡鹏：《神经网络与深度学习》讲义 <http://nlp.fudan.edu.cn/dl-book/>

70页的讲义

# 感谢读者的反馈和建议！

## 《神经网络与深度学习》读者反馈

邱老师您好，在读的过程中有以下四点小疑问：

1. 第 19 页，动量法：迭代式是否应为：

$$\theta_t = \theta_{t-1} + (\rho \nabla \theta_{t-1} - \lambda g_t)$$

2. 第 37 页，平均感知机的式(4.15)： $c_k$  是否是多余的？

3. 第 45 页，

~~式(5.12)下面的描述：W 的 Forbenius 范数是否应是：~~

$$\|W\|_F^2 = \sum_{l=1}^L \sum_{j=1}^{n^{l-1}} \sum_{i=1}^{n^l} W_{ij}^{(l)2}$$

~~式(5.14)：正则化项  $\frac{1}{2} \lambda \|W\|_F^2$  (标量) 对  $W^{(l)}$  (矩阵) 求导的结果是否不应是  $\lambda W$  (具有三个维度的张量)？第 47 页的算法流程也有该式，~~

4. 第 61 页，式(7.7)：我的计算结果与讲义有一点区别：

$$\begin{aligned} \frac{\partial \mathbf{h}_i}{\partial \mathbf{h}_{i-1}} &= \frac{\partial f(U\mathbf{h}_{i-1} + W\mathbf{x}_i + \mathbf{b})}{\partial \mathbf{h}_{i-1}} \\ &= \frac{\partial(U\mathbf{h}_{i-1} + W\mathbf{x}_i + \mathbf{b})}{\partial \mathbf{h}_{i-1}} \frac{\partial f(U\mathbf{h}_{i-1} + W\mathbf{x}_i + \mathbf{b})}{\partial(U\mathbf{h}_{i-1} + W\mathbf{x}_i + \mathbf{b})} \\ &= U^\top \text{diag}'(U\mathbf{h}_{i-1} + W\mathbf{x}_i + \mathbf{b}) \end{aligned}$$

如果式(7.2)为  $\mathbf{h}_i = Uf(\mathbf{h}_{i-1}) + W\mathbf{x}_i + \mathbf{b}$ ，则可得到讲义的结果。

其他几个不影响阅读的地方：

1. 第 25 页，式(3.58)：softmax 的公式笔误了；

2. 第 26 页，

式(3.60)最后一个等号处、式(3.61)第三个等号处、式(3.65)：z 少一个上标  $\hat{z}$ ；

式(3.61)：第二个等号处的第二个求和号应去掉；

倒数第三行： $\mathbf{1}^\top$  的下标从  $K$  换成  $C$  (代表类别总数) 比较合适吧；

最后一行：求和号内的分子少了一个偏导符号，softmax 的输入应为  $\hat{z}^{(i)}$ ；

3. 第 27 页，

式(3.66)第一个等号处：z 少一个上标  $\hat{z}$ ；

式(3.66)第三、四个等号处： $\mathbf{1}^\top$  的下标从  $K$  换成  $C$  比较合适吧；

式(3.67)：式(3.66)省略了常系数  $1/N$ ，式(3.67)没有省略；

4. 第 45 页，式(5.15)第一个等号处：第二项的分子应为  $\partial J(W, \mathbf{b})$ ；

5. 第 50 页，式(6.1)：求和号的上标应是  $m$ ；



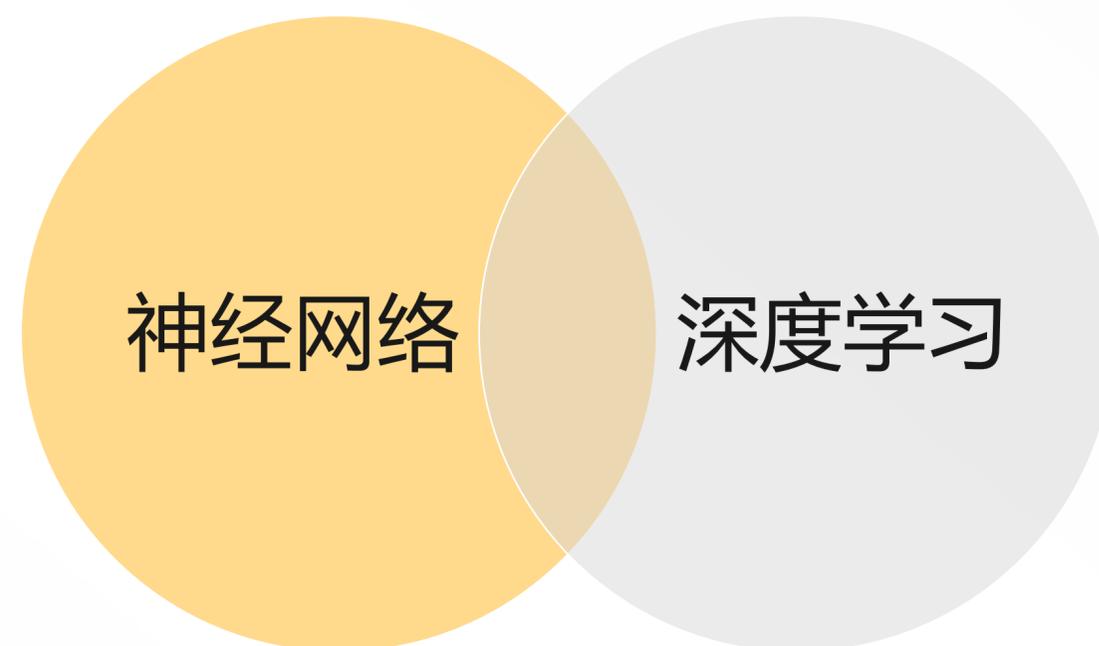
fandywang commented on 11 Mar 2017 • edited by xpqiu

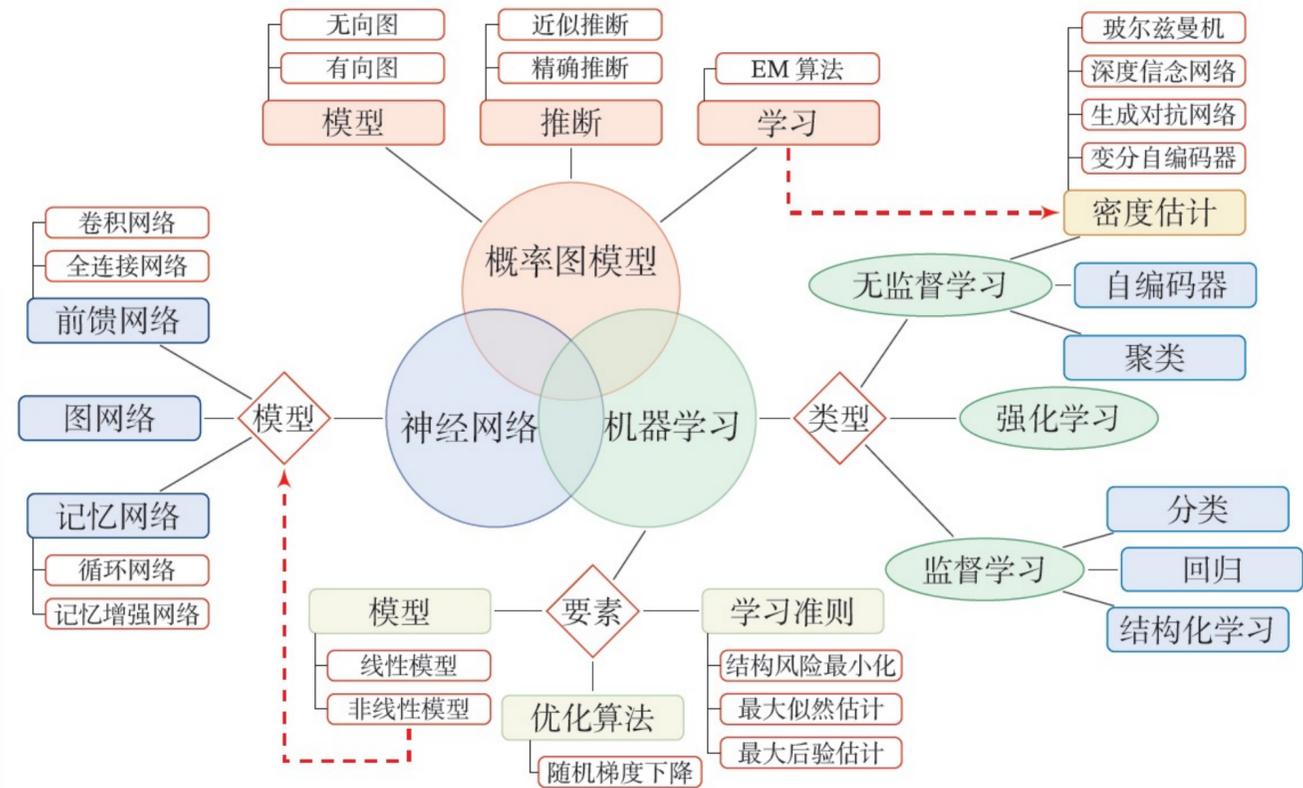
- 1. ch1 page1 "要通过真正地通过图灵测试" -> "要真正地通过图灵测试"
- 2. ch1 page2 "由大量的节点 (或称"神经元", 或"单元") 和之间相互连接构成" -> "由大量的节点 (或称"神经元", 或"单元") 之间相互连接构成"
- 3. ch1 page2 "甚至在很多的任务上比算法重要" -> "甚至在很多的任务上比算法更重要"
- 4. ch2 page9 "一个标量c与矩阵A乘积为A的每个元素是A的相应元素与c的乘积" -> "一个标量c与矩阵A乘积为cA的每个元素是A的相应元素与c的乘积"
- 5. ch2 page10 "一个n x n的对角矩阵矩阵A满足" -> "一个n x n的对角矩阵A满足"
- 6. ch2 page10 "n x n的对角矩阵矩阵A" -> "n x n的对角矩阵A"
- 7. ch2 2.1 建议增加下特征值/特征向量, 可逆矩阵, 行列式, 矩阵求导, 常见基本初等函数求导示例等内容 (有些内容在2.5中, 建议调整到2.1)
- 8. ch2 page12 2.24等式右边两项颠倒下顺序?
- 9. ch2 page12 2.25式 tr 在前面并未介绍
- 10. ch2 2.2章节 "数学优化" -> "数值优化"?尤其在第二段刚开始描述是"数值优化问题的定义为", 后面同样也出现了数值优化说法, 建议统一成更为常见的数值优化叫法
- 11. 对 R 和 Z (可能还有其他符号) 并未说明, 也许需要有一个字符集说明?
- 12. ch2 page13 "离散优化问题的求解一般都比较困难, 优化算法的复杂度都比较高。" -> "离散优化问题的求解一般都比较困难, 优化算法的复杂度都比较高。"
- 13. ch2 page13 整数规划部分也仿照组合优化部分给一些典型的例子?
- 14. ch2 page14 "有约束条件的约束问题常常可以通过拉格朗日乘数转化为非约束问题。" -> "有约束条件的约束优化问题常常可以通过拉格朗日乘子法转化为无约束优化问题。" (or 拉格朗日乘数法)
- 15. ch2 page15 "所对于所有的x" -> "对于所有的 x"
- 16. ch2 page15 定理2.2 二阶导表达式错误
- 17. ch2 page16 "有一阶必要性定理可知  $\nabla f(x)$  则" -> "有一阶必要性定理可知  $\nabla f(x)=0$ , 则"
- 18. ch2 page16 "如果顺利的话序列, (xn)收敛到局部最优解x\*。" -> "如果顺利的话序列(xn)收敛到局部最优解x\*。"
- 19. ch2 page17 如果可以对牛顿法多展开介绍下就更好了, 甚至涵盖 BFGS, LBFSG, OWLQN 等优化方法。另外, 共轭梯度法前面提了一句, 是否也加一些内容?
- 20. ch2 2.3章节中"实验"和"试验"说法统一吧
- 21. ch2 page18 "随机事件 (或简称事件) 指的是一个被赋予概率的事物集合" -> "随机事件 (或简称事件) 指的是一个被赋予概率的事物集合"
- 22. ch2 page18 "概率表示对一个随机事件发生的可能性大小" -> "概率表示一个随机事件发生的可能性大小"
- 23. ch2 page19 随机变量 X 的符号都用斜体吧
- 24. ch2 page21 "若随机变量X 服从一个位置参数为  $\mu$  和  $\sigma^2$ " -> "若随机变量X 服从一个未知参数为  $\mu$  和  $\sigma^2$ "
- 25. ch2 page23 多项分布定义部分重复给出了二项分布定义, 应该没必要, 删了吧, 直接到"把二项分布推广到随机向量, 就得到了多项分布。假设一个袋子中装了很 XXXX"
- 26. ch2 page23 式2.57 p 改为  $\theta$ ?
- 27. ch2 page23 式2.58 gamma 函数  $\Gamma(x)$  的定义给下? 比如在右侧补充下?
- 28. ch2 page23 "从这种表示形式和Dirichlet分布类似" -> "这种表示形式和Dirichlet分布类似"
- 29. ch2 page23 既然最后提到 Dirichlet分布, 不妨也在合适的位置介绍下?
- 30. ch2 2.4章节很多本来是下标, 但没有正确呈现出来, 如  $x_i, x_j$
- 31. ch2 page26  $l(x)$  是什么意思, 要给出解释吧?
- 32. ch2 page27 2.4.2章节英文描述翻译成中文吧?
- 33. ch2 page27 2.4.2章节DKL(p||q) 要给出解释吧?
- 34. ch2 page29 "在1844-1845年研究种群数量的增长模型时提出的命名的" -> "在1844-1845年研究种群数量的增长模型时提出命名的"

- 46. ch3 page38 "数据的原始表示转换为。" 转换成什么呢?
  - 47. ch3 page39 "在机器学习, 搜索步长  $\alpha$  中也叫作学习率 (Learning Rate) 。" -> "在机器学习中, 搜索步长  $\alpha$  也叫作学习率 (Learning Rate) 。"
  - 48. ch3 page40 "Mini-Batach" -> "Mini-Batch"
  - 49. ch3 page40 Mini-Batch 兼顾随机和批量方法优点, 但是前文并未提及批量方法优点是什么。另外, SGD 方法的不足也可以提下, 这样说 Mini-Batch 是一种折中方法逻辑更清晰
  - 50. ch3 page40 线性搜索和自适应学习率如果能有更详细介绍, 或者给出参考资料更好
  - 51. ch3 page41~42 可以介绍下为什么 动量法、adagrad、adadelat有效, 以及适用什么场景?
  - 52. ch3 page42 RMSprop、Adam等也许也可以提下?
  - 53. ch3 page43、47、50 "给定N 给样本" -> "给定N 个样本"
  - 54. ch3 page46 式3.47 分子少了右括号
  - 55. ch3 page47 式3.54 为  $\infty$
  - 56. ch3 page48 式3.59 使用  $\lambda$  表示学习率, 而前文中一直是 a, 最好统一下?
  - 57. ch3 page48 "空间中的存在一些区域" 不通顺
  - 58. ch3 page48 多分类通过二分类实现及存在缺陷, 如果通过直观的图呈现出来更佳
  - 59. ch3 page49 "对比公式3.65中的两类分类判别函数" 这里的公式3.65标错了吧
  - 60. ch3 page49 softmax 回归最后是否提一句和最大熵模型推到出来的后验概率形式一致
  - 61. ch3 page50 "采样梯度下降法" -> "采用梯度下降法"
  - 62. ch3 page51 "其中,  $l()$  为指示函数。" 频繁的出现, 也许没必要?
  - 63. ch3 page52 "决策树模型可以读性好, 具有描述性" -> "决策树模型可读性好, 具有描述性"
  - 64. ch3 page53 式3.82漏掉了  $p(c)$ , 先验概率还是不能省略的, 除非显式说明各类别出现概率均等
  - 65. ch3 page54 公式 3.83 给的太直接了, 很难理解, 且缺少一些解释
  - 66. ch3 page56 "正确率和召回率是广泛用于信息检索和统计学分类领域的两个度量" -> "正确率和召回率是广泛用于信息检索和统计学分类领域的两个度量"
  - 67. ch3 page56 "F1 值是根据正确率和召回率二者给出的一个综合的评价指标" -> "F1 值是根据正确率和召回率二者给出的一个综合的评价指标"
  - 68. ch3 page58 "1960年代," -> "1960年代,"
  - 69. ch3 page57 3.4 章节可以考虑加下奥卡姆剃刀原理?
  - 70. ch4 页面标号没有和 ch3 连上
  - 71. ch4 page49 "感知器也可以看出是线性分类器的一个经典学习算法。" -> "感知器也可以看做是线性分类器的一个经典学习算法。"
  - 72. ch4 page51 公式4.3, " $w \cdot x_i < 0$  当  $y_i < 0$ " -> " $w \cdot x_i <= 0$  当  $y_i < 0$ ."
  - 73. ch4 page51 "具体的学习过程如算法5.2所示。" -> "具体的学习过程如算法4.1所示。"
  - 74. ch4 page51 算法4.1 第6行多了一个逗号
  - 75. ch4 page52 "具体的学习过程如算法5.2所示。" -> "具体的学习过程如算法4.1所示。"
  - 76. ch4 page53 "如果训练集  $s$  是线性可分的" -> "如果训练集不是线性可分的"
  - 77. ch4 page53 "算法5.2" -> "算法4.1"
  - 78. ch4 page54 "比如输出是序列或来其它结构化的形式。" -> "比如输出是序列或者其它结构化的形式。"
  - 79. ch4 page57 公式4.15第二行少了一个左括号
  - 80. ch4 page58 算法4.3 第8、9行少了  $\phi$
  - 81. ch5 page75 公式5.13 未写完整
- []
- 未完待续

# 神经网络与深度学习的关系

- 人工智能的一个子领域
  - 神经网络：一种以（人工）神经元为基本单元的模型
  - 深度学习：一类机器学习问题，主要解决贡献度分配问题





## 概述

机器学习概述  
线性模型

## 基础模型

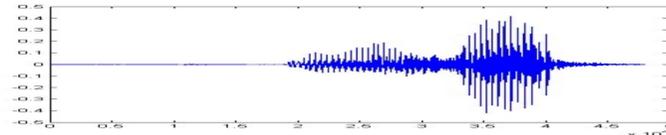
前馈神经网络  
卷积神经网络  
循环神经网络  
记忆与注意力机制  
网络优化与正则化  
无监督学习  
模型独立的学习方式

## 进阶模型

概率图模型  
深度信念网络  
深度生成模型  
深度强化学习  
序列生成模型

# 机器学习 ≈ 构建一个映射函数

语音识别  $f(\text{语音波形图}) = \text{“你好”}$



图像识别  $f(\text{猫咪照片}) = \text{“猫”}$



围棋  $f(\text{围棋棋盘}) = \text{“5-5”}$  (落子位置)



对话系统  $f(\text{“你好”}) = \text{“今天天气真不错”}$   
用户输入 机器

# 机器学习的四个要素

TIANCHI天池

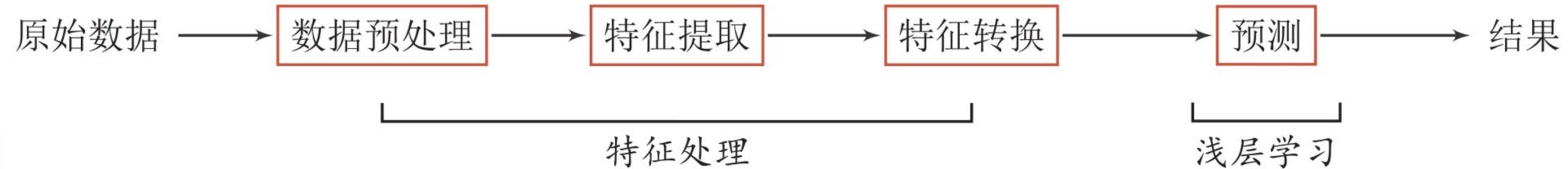
数据

模型

学习  
准则

优化  
算法

- 当我们用机器学习来解决一些模式识别任务时，一般的流程包含以下几个步骤：



## 特征工程 ( Feature Engineering )

- 浅层学习 ( Shallow Learning )：不涉及特征学习，其特征主要靠人工经验或特征转换方法来抽取。

# 语义鸿沟：人工智能的挑战之一

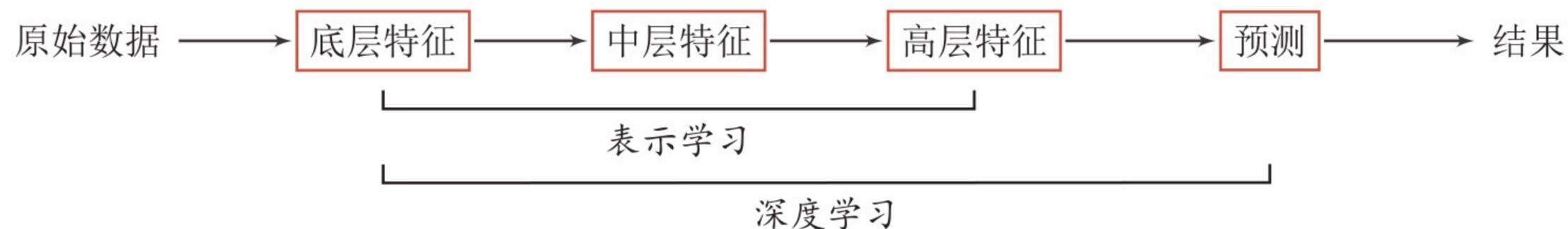
- 底层特征 VS 高层语义
  - 人们对文本、图像的理解无法从字符串或者图像的底层特征直接获得

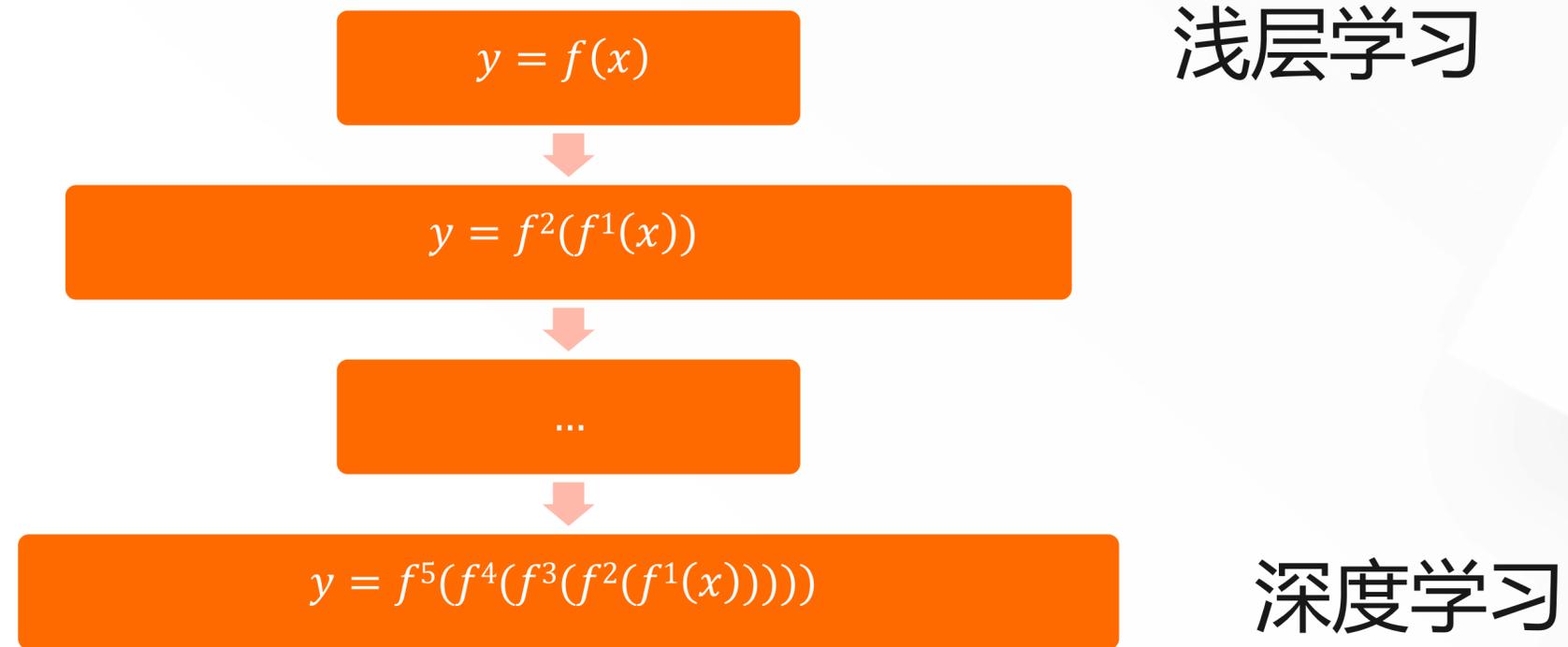


床前明月光，  
疑是地上霜。  
举头望明月，  
低头思故乡。

## 深度学习 = 表示学习 + 浅层学习

- 通过构建具有一定“深度”的模型，可以让模型来自动学习好的特征表示（从底层特征，到中层特征，再到高层特征），从而最终提升预测或识别的准确性。



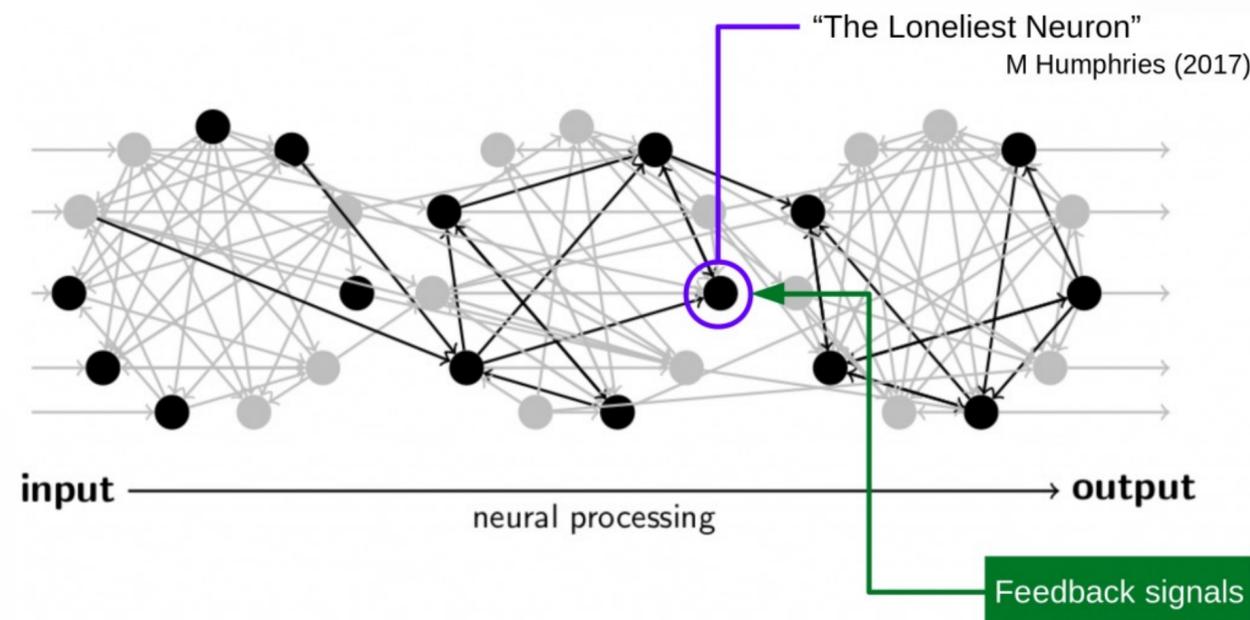


$f^1(x)$  为非线性函数，不一定连续。

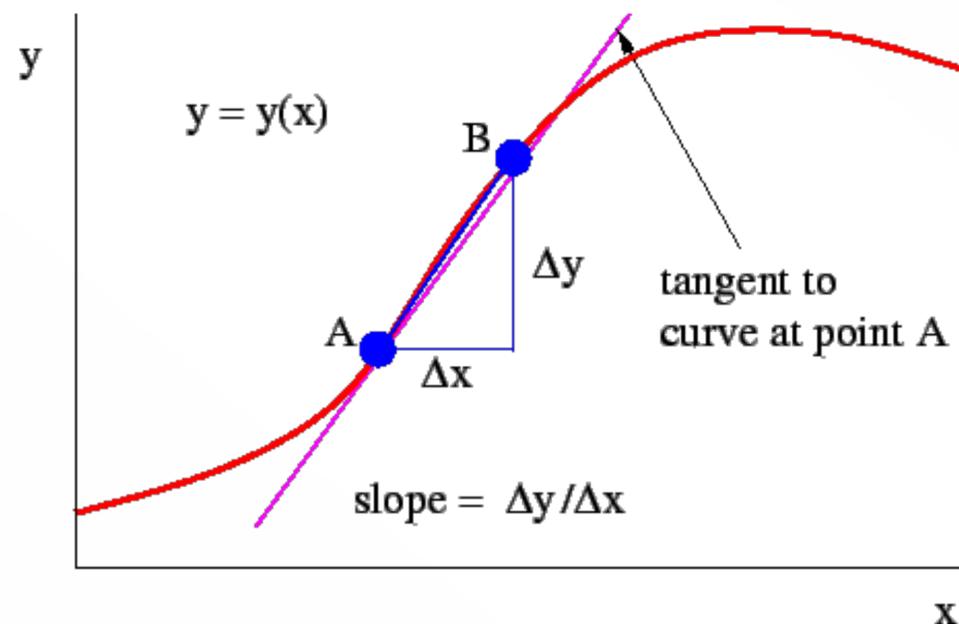
$$y = f^5(f^4(f^3(f^2(f^1(x))))))$$

$f^1(x)$ 为非线性函数，不一定连续。

**贡献度分配问题**：一个复杂系统中每个组件对最终评价的贡献。



# 如果解决贡献度分配问题？



偏导数

贡献度

$$\frac{\partial L}{\partial W^l} = \frac{L(W^l + \Delta W^l) - L(W^l)}{\Delta W^l}$$

当 $f^l(x)$ 连续时为神经网络！  
比如 $f^l(x) = \sigma(W^l f^{l-1}(\cdot))$

**神经网络天然不是深度学习，  
但深度学习天然就是神经网络！**

**系统性**：系统地整理了神经网络和深度学习的知识体系。鉴于深度学习涉及的知识点较多，本书从机器学习的基本概念、神经网络模型以及概率图模型三个层面来串联深度学习所涉及的知识点，使读者对深度学习技术的理解更具系统性、条理性和全面性。

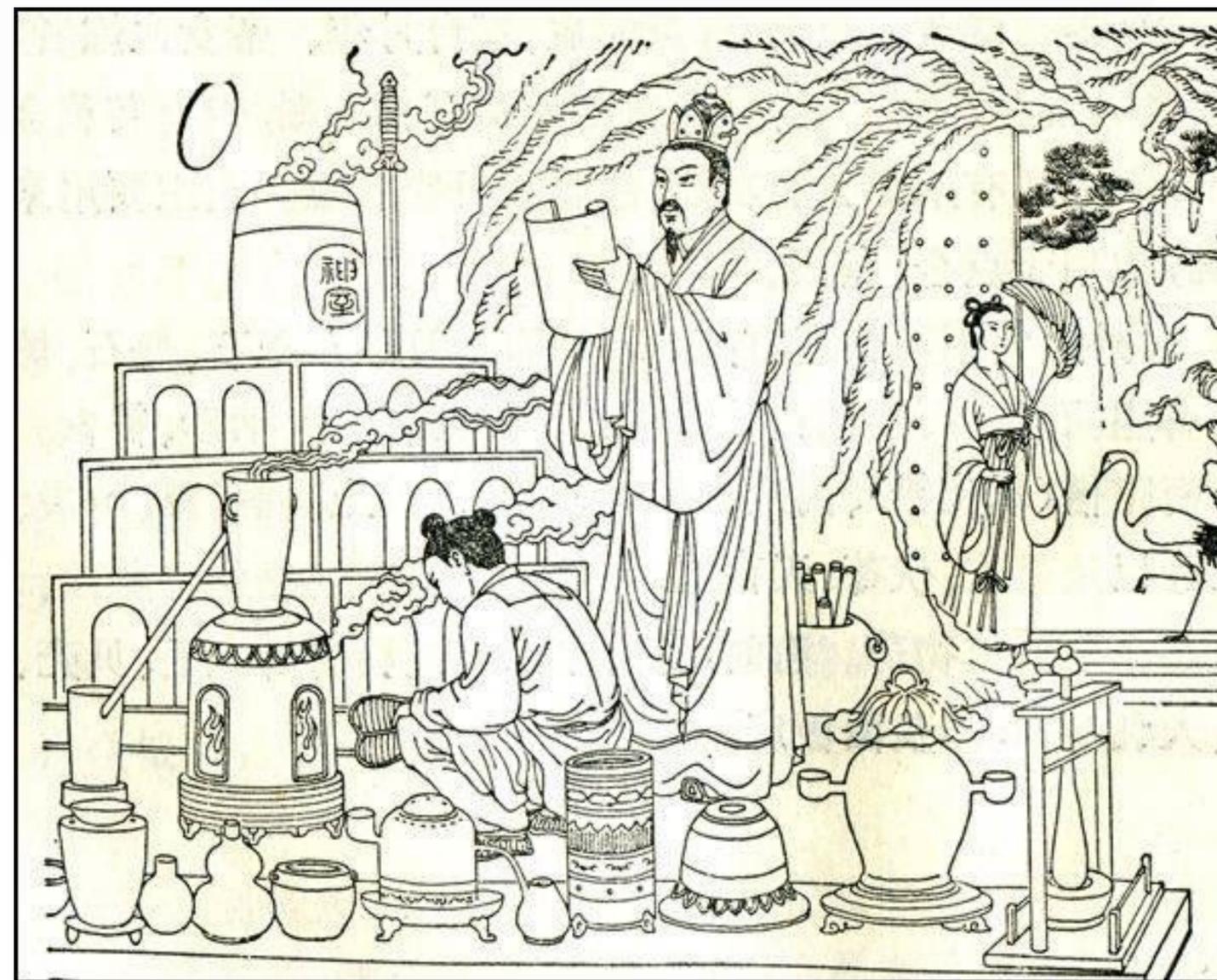
**可读性**：本书在编排上由浅入深，在语言表达上力求通俗易懂，并通过增加图例、示例以及必要的数学推导来理解抽象的概念。同时，附录简要介绍了本书所涉及的必要数学知识，便于读者查用。

**实践性**：本书在网站上配套了针对每章知识点的编程练习，使得读者在学习过程中可以将理论和实践密切结合，加深对知识点的理解，并具备分析问题和解决问题的能力。

# 学习方法

# 深度学习：真的是炼丹术吗？

- 理论支撑不足
- 调参一头雾水
- 模型无法解释
- 改进没有方向



# 学习目标

- 培养自己的深度学习思维，和解决实际应用问题的能力。
  - 掌握深度学习的基础理论与实践技巧
    - 理解并分析具体的深度学习算法
  - 更加强调学生动手能力
    - 独立地开发深度学习模型
  - 具备基本的问题分析能力
    - 学以致用，解决现实生活中真实存在的实际问题

到老一場空  
練拳不練功

# 需要具备哪些知识才能学习这本书？

## 前置知识？

- 数学：线性代数、微积分、变分法、概率论、优化、信息
- 简单了解的话可以从本书附录学习

## 编程语言？

- Python/NumPy/Pytorch

## 机器学习？

- 不需要，本书会介绍机器学习的基本知识

## 理论

- 广：覆盖面
- 深：核心知识点

理论和实践  
紧密结合

## 实践

- 理论需辅以实践项目
- 根据不同的基础来安排不同难度的实践

## 目标

- 知识点的掌握程度
- 解决实际问题的能力

## 前置知识多

- 数学：线性代数、微积分、变分法、概率论、优化、信息论
- 编程语言：Python

## 内容比较杂乱

- 领域多、知识点多

## 更新速度快

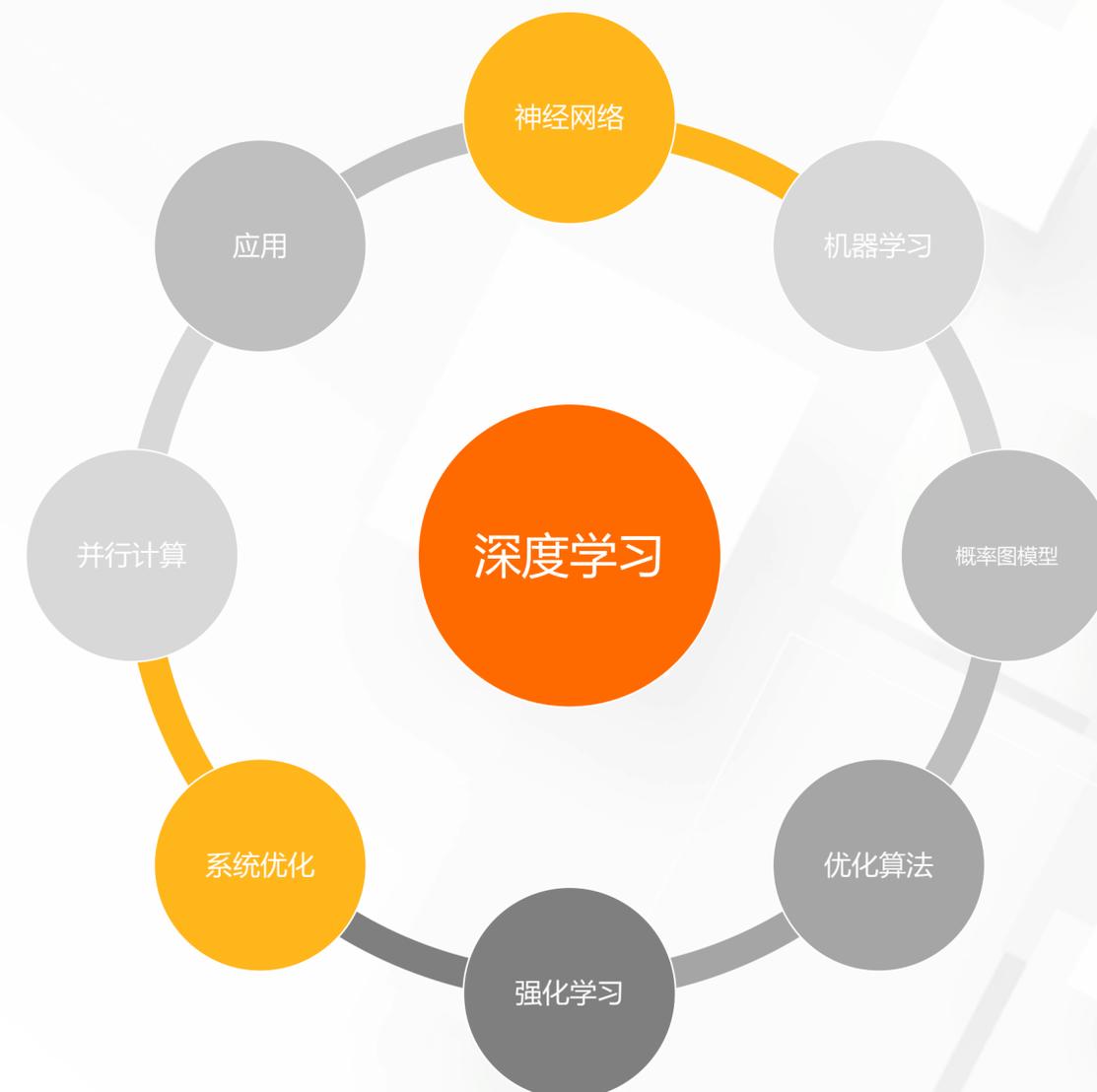
- 新论文层出不穷

## 和实践结合紧密

- 动手能力

# 构建自己的知识体系

- 如何省时省力地通过阅读构建一个知识体系？
  - 正统文献（教材）
    - 进入一个陌生领域，首先要从阅读“正统”文献或者作品开始
      - 正统不代表正确，也不代表面面俱到！
    - 网上材料很多，但寻找好内容的成本太高
  - 权威综述文章
    - 综述文章基本上会概括一个领域最近十年的成就
  - （领域）学术专著
    - 专著的范围一般都比较窄，而且观点会非常鲜明，甚至偏激
  - 学术前沿
    - 跟踪自己领域的最新研究成果
    - 充分利用开源代码



## 课时分配：概述（8~9课时）

- 课程概述（3课时）
  - 人工智能、机器学习、神经网络的介绍、发展历程
- 机器学习概述（3课时）
  - 以线性回归为例，讲授机器学习中的基本概念
  - 损失函数、优化、泛化
- 线性模型（3课时）
  - 传统机器学习的各种模型
  - Logistic回归、Softmax回归、感知器、SVM

## 课时分配：基础模型（15~24课时）

- 前馈神经网络（3课时）
  - 神经元、模型定义、反向传播
- 卷积神经网络（3课时）
  - 卷积定义、模型介绍、常用网络
- 循环神经网络（3课时）
  - 模型定义、LSTM、GRU
- 网络优化与正则化（3~5课时）
  - 优化：随机梯度下降、逐层归一化、超参优化
  - 正则化：L1/L2正则化、Dropout、标签平滑
- 记忆与注意力机制（3课时）
  - 注意力机制、外部记忆
- 无监督学习（3课时，选讲）
  - 编码、稀疏编码、自编码器、聚类、密度估计
- 模型独立的学习方式（3课时，选讲）
  - 集成学习、元学习、迁移学习、多任务学习

## 课时分配：进阶模型（15~21课时）

- 概率图模型（6~9课时）
  - 模型介绍（3课时）
  - 推断（3课时）
  - 学习（3课时）
- 深度信念网络（3课时，选讲）
  - 玻尔兹曼机、受限玻尔兹曼机、深度信念网络
- 深度生成模型（3课时）
  - 变分自编码器、GAN
- 深度强化学习（3课时，选讲）
  - 基于模型的方法、无模型方法、策略梯度
- 序列生成模型（3课时，选讲）
  - 语言模型
  - 条件生成模型

- 根据学生的基础来安排不同难度实践
- 通过GitHub来进行项目管理

拟在本学期会在大幅完善相关实践的编程内容

## Exercise

### 1. 热身练习 warmup

numpy是Python中对于矩阵处理很实用的工具包，本小节作业主要是熟悉基本的numpy操作。

### 2. 线性回归模型 Linear Regression

### 3. 线性模型

1. 支持向量机 support vector machine
2. Softmax回归 Softmax Regression

### 4. 前馈神经网络 Simple Neural Network

利用numpy实现全连接神经网络

### 5. 卷积神经网络 Convolutional Neural Network (CNN)

利用卷积神经网络，处理MNIST 数据集分类问题。

### 6. 循环神经网络 Recurrent Neural Network (RNN)

基于循环神经网络的唐诗生成问题

### 7. 注意力机制 Attention Mechanism

1. 使用sequence to sequence 模型将一个字符串序列逆置。
2. 使用attentive sequence to sequence 模型将一个字符串序列逆置。

### 11. 高斯混合模型 Gaussian Mixture Model

### 12. 受限玻尔兹曼机 Restricted Boltzmann Machine (RBM)

使用受限玻尔兹曼机 (Restricted Boltzmann Machine, RBM) ，对MNIST 数据集建模。

### 14. 深度强化学习 Deep Reinforcement Learning

强化学习：黑白棋

# fastNLP介绍

<https://github.com/fastnlp/fastNLP>  
<https://gitee.com/fastnlp/fastNLP>

# 天池读书会

边读书 边实践，沉浸式读书体验

## 第五期活动安排

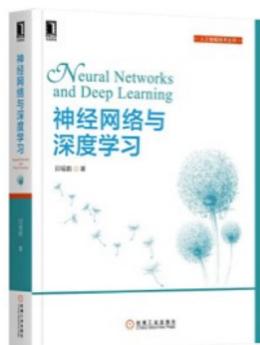
活动时间：3月24-3月31日晚八点



《神经网络与深度学习》

分享嘉宾 邱锡鹏

直播时间：3月25日20:00~21:00



扫码观看直播



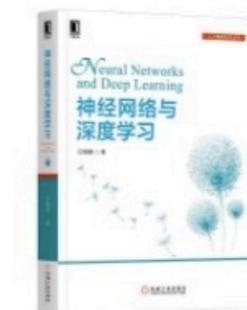
**邱锡鹏** 复旦大学计算机科学技术学院教授

直播主题 《神经网络与深度学习》

直播时间 2021年3月25日 20:00

学习资料 《神经网络与深度学习》课程

实践项目 fastNLP



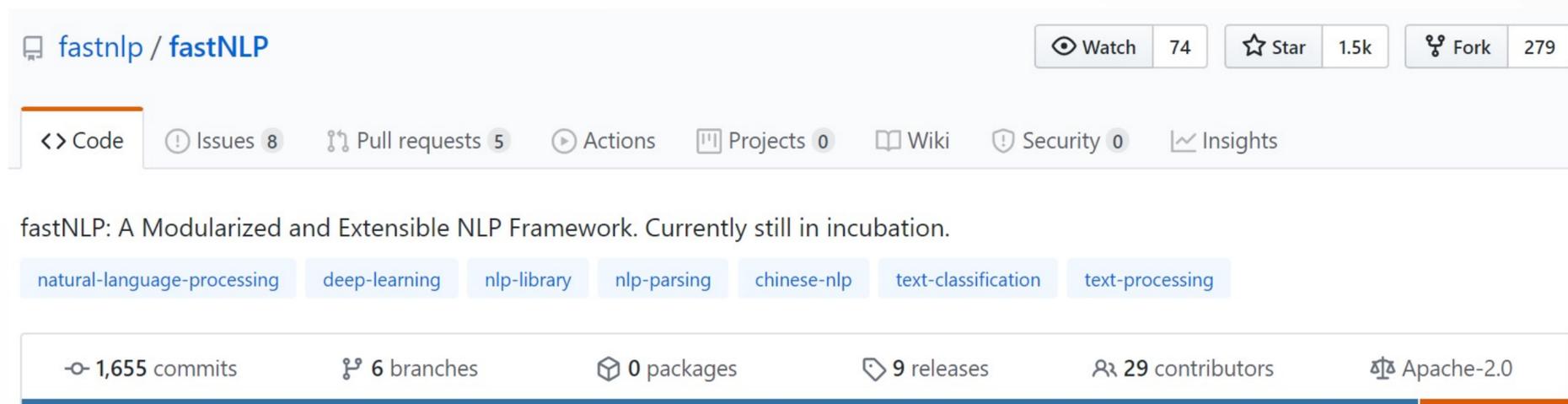
[🗨️ 提问](#) | [📖 学习课程](#) | [🛒 购买地址](#) | [📄 PPT下载](#) | [👉 实践项目](#) | [🕒 预约直播](#)

**fastNLP项目集** 是面向自然语言处理的开源项目集群，包含一系列基于fastNLP框架的开源项目：

1. [fastNLP](#)框架：中文友好的轻量级国产自然语言处理开源框架
2. [fastHan](#):中文自然语言处理基础任务项目，包括中文分词、词性标注、依存分析、命名实体识别等任务
3. [fitlog](#):用于辅助用户记录日志和管理代码的调参工具
4. fastSum:文本摘要工具
5. fastRE:关系抽取工具
6. 其它项目陆续增加中...

# fastNLP: A Modularized NLP Framework

TIANCHI 天池



fastnlp / fastNLP

Watch 74 Star 1.5k Fork 279

Code Issues 8 Pull requests 5 Actions Projects 0 Wiki Security 0 Insights

fastNLP: A Modularized and Extensible NLP Framework. Currently still in incubation.

natural-language-processing deep-learning nlp-library nlp-parsing chinese-nlp text-classification text-processing

1,655 commits 6 branches 0 packages 9 releases 29 contributors Apache-2.0

## fastNLP

build passing codecov 78% pypi v0.5.0 license Apache-2.0 docs passing

fastNLP是一款轻量级的自然语言处理（NLP）工具包，目标是快速实现NLP任务以及构建复杂模型。

fastNLP具有如下的特性：

- 统一的Tabular式数据容器，简化数据预处理过程；
- 内置多种数据集的Loader和Pipe，省去预处理代码；
- 各种方便的NLP工具，例如Embedding加载（包括ELMo和BERT）、中间数据cache等；
- 部分数据集与预训练模型的自动下载；
- 提供多种神经网络组件以及复现模型（涵盖中文分词、命名实体识别、句法分析、文本分类、文本匹配、指代消解、摘要等任务）；
- Trainer提供多种内置Callback函数，方便实验记录、异常捕获等。

## fastNLP教程

[中文文档、教程](#)

### 快速入门

- [0. 快速入门](#)

### 详细使用教程

- [1. 使用DataSet预处理文本](#)
- [2. 使用Vocabulary转换文本与index](#)
- [3. 使用Embedding模块将文本转成向量](#)
- [4. 使用Loader和Pipe加载并处理数据集](#)
- [5. 动手实现一个文本分类器I-使用Trainer和Tester快速训练和测试](#)
- [6. 动手实现一个文本分类器II-使用DataSetIter实现自定义训练过程](#)
- [7. 使用Metric快速评测你的模型](#)
- [8. 使用Modules和Models快速搭建自定义模型](#)
- [9. 快速实现序列标注模型](#)
- [10. 使用Callback自定义你的训练过程](#)

### 扩展教程

- [Extend-1. BertEmbedding的各种用法](#)
- [Extend-2. 使用fitlog 辅助 fastNLP 进行科研](#)



# 为什么我们要做fastNLP

TIANCHI天池

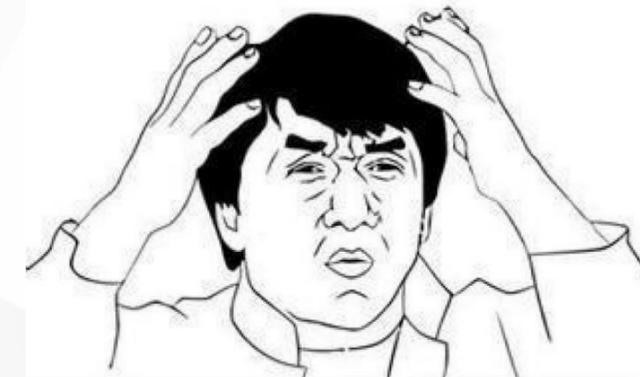
模型到部署上线存在间隙

通过与TorchServe衔接，我们尝试减小模型上线的难度（on developing）



性能复现

复现代码有时候会很困难，有一个参考可能会更容易

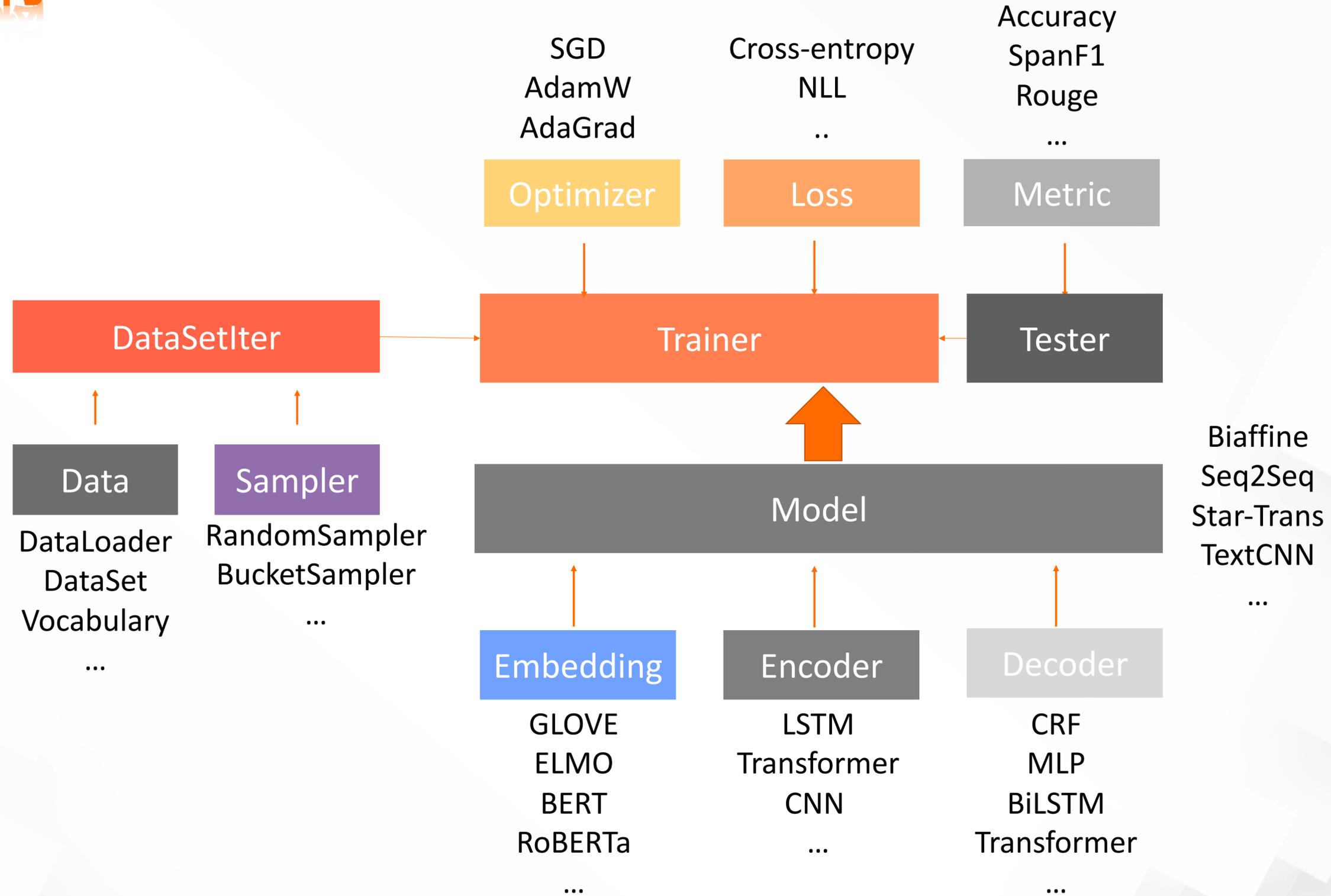


中文友好

fastNLP的文档，教程，issue甚至连代码注释都是中文的；同时fastNLP重视在中文任务上的性能

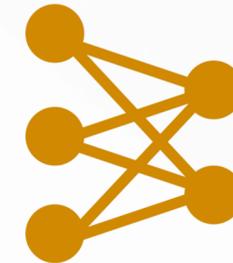
漢

# 整体架构



# Tabular式的数据结构

fastNLP.core.DataSet



instance 1  
instance 2

raw_words1	target1
raw_words2	target2
...	...

.....

人们在理解数据时，一般是通  
过一个一个的sample去理解，  
即通过行的方式

raw_words1	target1
raw_words2	target2
...	...

.....

批处理更喜欢以列的方式  
进行读取与处理

Batch1

# 高效切换Glove、ELMO、BERT等

## 代码示例：

```
# 同时使用两种embedding
embed = StackEmbedding([glove_embed, word2vec_embed])

# 使用character embedding和word embedding一样容易
char_embed = CNNCharEmbedding(vocab)
embed = StackEmbedding([char_embed, glove_embed])
```

## Contextual Embedding:

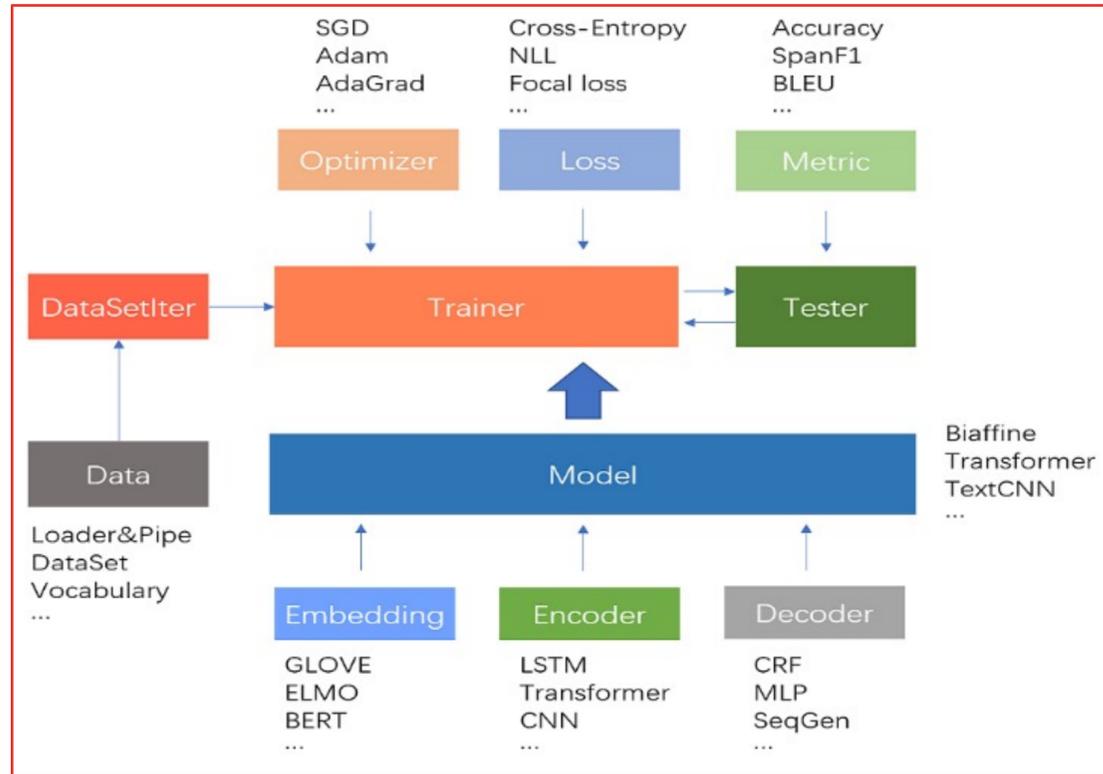
一行代码切换ELMO, BERT, RoBERTa

```
elmo_embed = ElmoEmbedding(vocab, model_dir_or_name='en-original')
bert_embed = BertEmbedding(vocab, model_dir_or_name='en')
roberta_embed = RoBERTaEmbedding(vocab, model_dir_or_name='en-base')

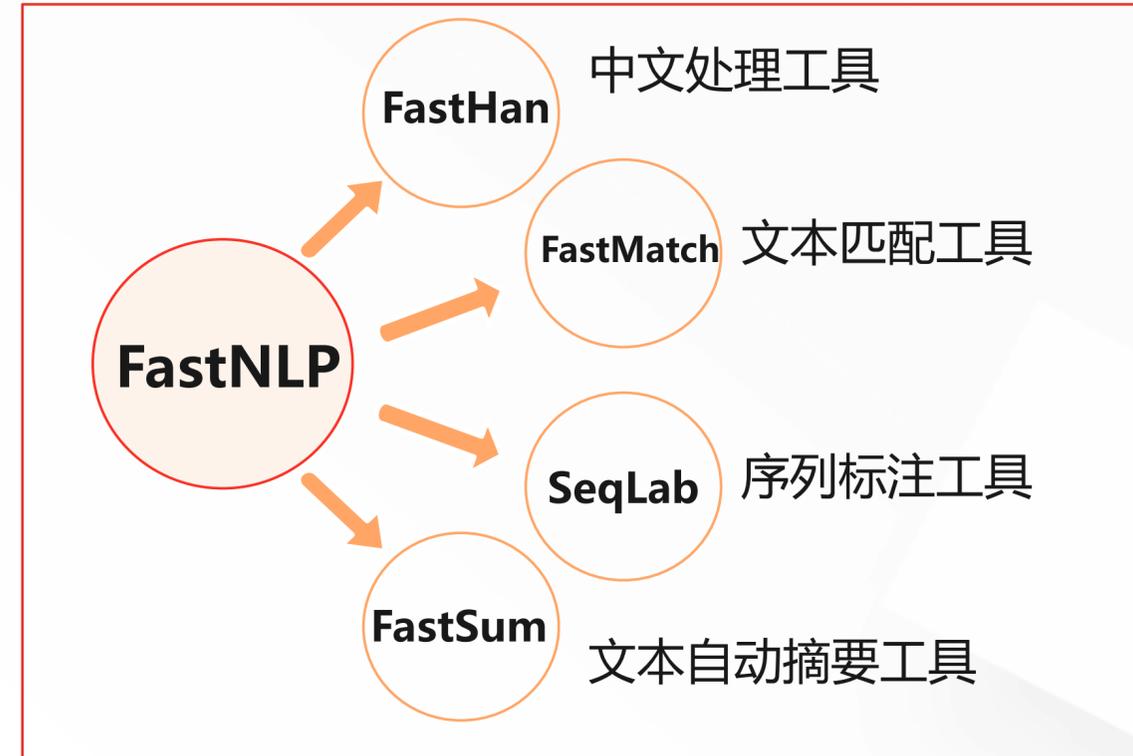
embed = StackEmbedding([glove_embed, elmo_embed, bert_embed])
```

这里的各种 Embedding 和 pytorch 的 nn.Embedding 是类似的，所以用法上基本没有区别，极大降低了使用门槛。

# 一套内核，多种应用



系统架构

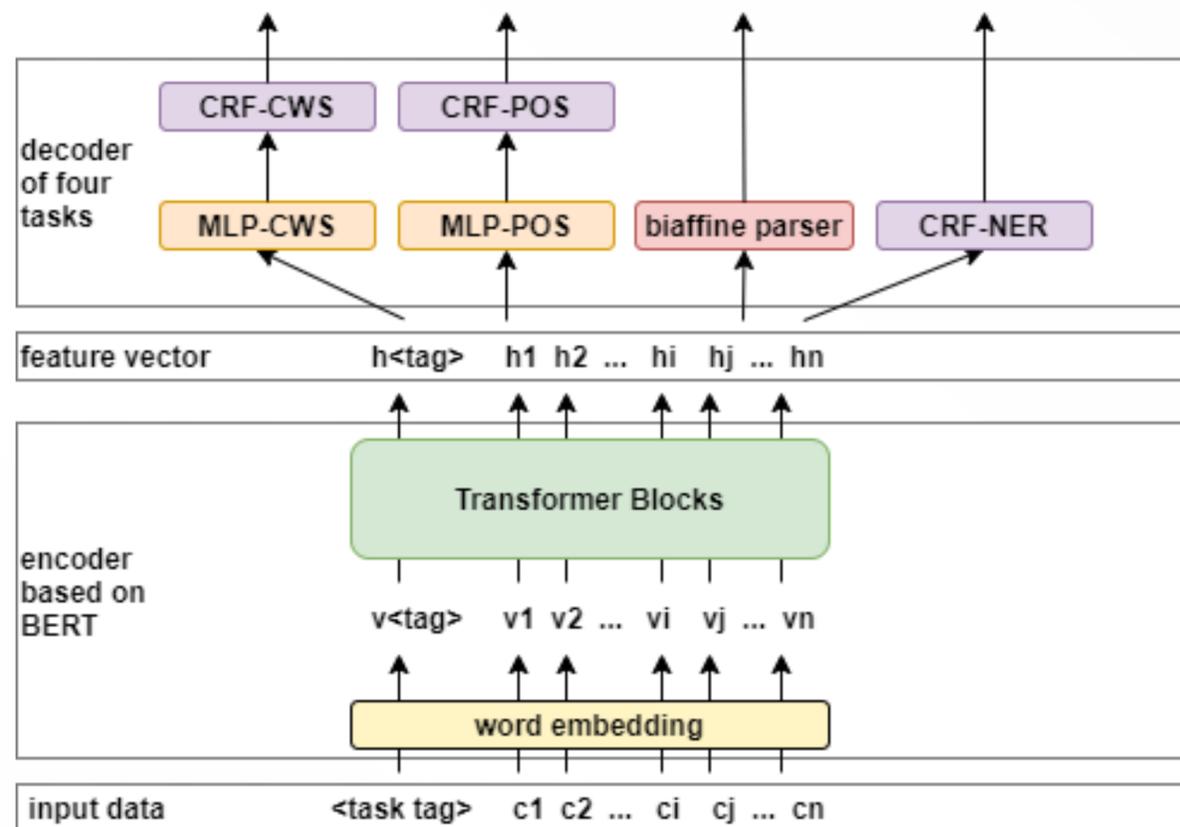


■ 根据PyPI.org统计，自2019年6月项目发布以来 FastNLP软件包**总下载量2万余次**

# fastHan:中文自然语言处理工具

TIANCHI天池

<https://github.com/fastnlp/fasthan>  
<https://gitee.com/fastnlp/fasthan>



```
from fastHan import FastHan
model=FastHan()
```

```
sentence=["我爱踢足球。","林丹是冠军"]
answer=model(sentence,'Parsing')
for i,sentence in enumerate(answer):
    print(i)
    for token in sentence:
        print(token,token.pos,token.head,token.head_label)
```

模型将输入如下内容:

```
0
我 PN 2 nsubj
爱 VV 0 root
踢 VV 2 ccomp
足球 NN 3 dobj
。 PU 2 punct
1
林丹 NR 2 top
是 VC 0 root
冠军 NN 2 attr
! PU 2 punct
```

任务	CWS	Parsing	POS	NER MSRA	NER OntoNotes
SOTA模型	97.1	85.66,81.71	93.15	95.25	79.92
base模型	97.27	81.22,76.71	94.88	94.33	82.86
large模型	97.41	85.52,81.38	95.66	95.50	83.82

在十多个数据集上达到目前的SOTA结果

直播相关资料获取及回放查看地址: <https://tianchi.aliyun.com/specials/promotion/activity/bookclub>

# 来自第三方的评测

TIANCHI天池

议下，我实验了fastHan分词，并取得了很好的效果。

我的实验数据为公司智能客服的历史客户提问，特别取了曾经真实出现过客户舆情热点的12次日志共14万条。

问法去重后使用base模式CWS任务分词，人工抽检200条客户提问，发现分词准确率很好，其中去除行业专业词汇准确率达到96.12%（Jieba为91.34%），包含行业专业词汇准确率在93.31%（Jieba在录入了专业词库后为92.61），fastHan的抽检准确率表现好于Jieba分词。

# fitlog = fast + git + log : 可视化、可交互调参工具

TIANCHI天池

以下每一行是一次实验记录

The screenshot displays the fitlog interface. At the top, there's a 'Statistics of metrics' popup window showing a table with columns: Metric name, max, min, avg, std/(N-1), and std/(N). The data row is: metric-SpanFPreRecMetric-f, 0.628966, 0.425844, 0.564206, 0.094222, 0.081598. Below this is a toolbar with buttons: Delete, Hide, Display, InvertCheck, ClearCheck, Columns, Setting, and Con. cols. A search bar and various icons are also present. The main table has columns: id, meta (fit\_msg, git\_msg, dataset, n\_heads, head\_dims, lr), hyper (epoch), metric (SpanFPreRecMetric, data\_test), memo, and action. The 'dataset' column in the first row is highlighted with a red box and labeled '下拉筛选'. The 'memo' column in the first row is highlighted with a red box and labeled '可直接编辑的备忘'. The 'action' column in the first row is highlighted with a green box and labeled '可视化收敛曲线'. A red arrow points from the 'Statistics of metrics' popup to the 'Support complex search syntax' label. A green arrow points from the 'memo' column to the 'Visualization of convergence curves' label.

meta		hyper					metric		memo	action		
fit_msg	git_msg	dataset	n_heads	head_dims	lr	epoch	SpanFPreRecMetric	data_test				
							f	SpanFPreRecMetric				
							f	f				
<input type="checkbox"/>	log_20200527_033121	Click to edit	Click to edit	weibo	6	48	0.0005	30	0.628966	0.544966	Click to edit	
<input type="checkbox"/>	log_20200527_021854	Click to edit	Click to edit	weibo	8	64	0.0015	83	0.618132	0.546875	Click to edit	
<input type="checkbox"/>	log_20200527_013632	Click to edit	Click to edit	weibo	8	48	0.0015	94	0.425844	0.350725	Click to edit	
<input type="checkbox"/>	log_20200527_010358	Click to edit	Click to edit	weibo	6	64	0.0015	97	0.583884	0.538462	Click to edit	
<input type="checkbox"/>	log_20200526_210347	Click to edit	Click to edit	resume	8	80	0.0015	47	0.860683	0.885276	Click to edit	
<input type="checkbox"/>	log_20200526_204911	Click to edit	Click to edit	resume	8	64	0.0015	42	0.945685	0.942149	Click to edit	
<input type="checkbox"/>	log_20200526_203500	Click to edit	Click to edit	resume	8	48	0.0015	18	0.945928	0.947888	Click to edit	
<input type="checkbox"/>	log_20200526_091311	Click to edit	Click to edit	ontonotes	6	64	0.001	27	0.702962	0.732482	Click to edit	

自动计算多行的统计信息

支持复杂搜索语法

可视化收敛曲线

可直接编辑的备忘

# fitlog可视化、可交互调参工具



## fitlog可视化、可交互调参工具

可自动化管理实验代码版本、超参数、实验结果的工具。使用时，在python代码中加入如下的内容即可。

### 代码示例：

```
import fitlog
# 设置log记录文件夹
fitlog.set_log_dir('logs')
# 记录超参数
fitlog.add_hyper(args)

for epoch in range(n_epochs):
    for step in range(num_step_per_epoch):
        # 记录loss
        fitlog.add_loss(loss, name='loss', step=step, epoch=epoch)
    # 记录metric
    fitlog.add_metric(f1, name='f1', epoch=epoch, step=step)
    if better_result:
        # 记录结果
        fitlog.add_best_metric(f1, name='f1')
```

### 代码示例：

```
import fitlog
# 设置log记录文件夹
fitlog.set_log_dir('logs')
# 记录超参数
fitlog.add_hyper(args)

# 直接使用FitlogCallback进行记录
Trainer(data_bundle.get_dataset('train'), model,
        dev_data=data_bundle.get_dataset('dev'),
        metrics=AccuracyMetric(),
        callback=FitlogCallback()).train()
```

<https://github.com/fastnlp/fitlog>  
<https://gitee.com/fastnlp/fitlog>

欢迎感兴趣的同学一起开发fastNLP!

# Q&A

学习资料、直播回放及交流

TIANCHI天池

# 天池读书会

TIANCHI天池

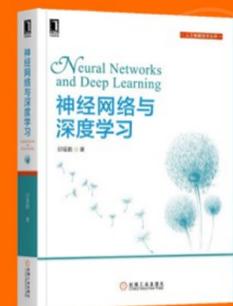
机械工业出版社  
华章公司

## 《神经网络与深度学习》

本书主要介绍神经网络与深度学习中的基础知识、主要模型、以及在计算机视觉、自然语言处理等领域的应用。

直播嘉宾：邱锡鹏

直播时间：3月25日20:00 ~ 21:00



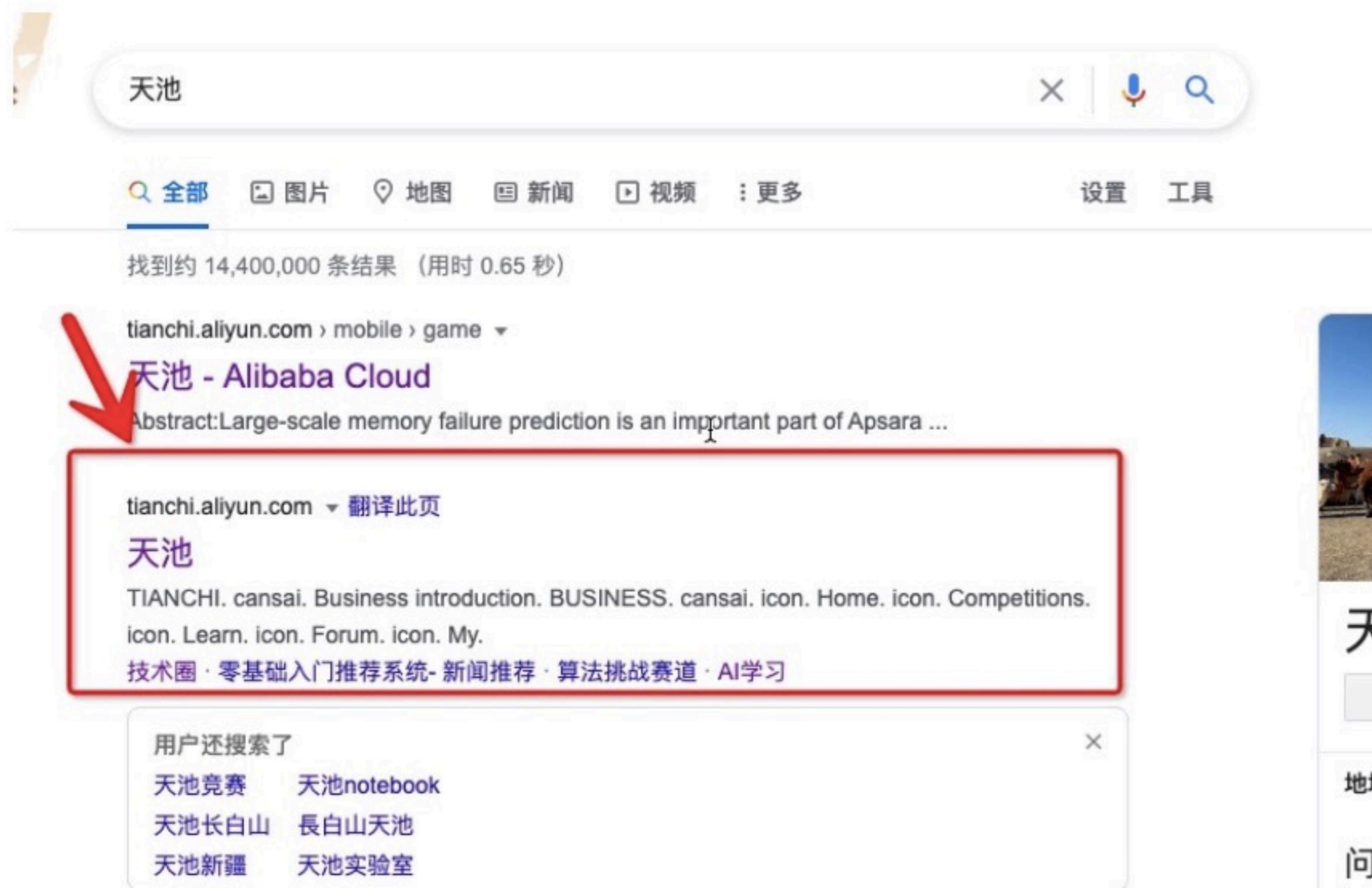
扫码领取  
读书会配套学习资源

bookclub



直播相关资料获取及回放查看地址：<https://tianchi.aliyun.com/specials/promotion/activity/bookclub>

1) 首先需要进入天池官网，大家打开浏览器，搜索 天池，找到 tianchi.aliyun.com即可访问进入天池官



2) 在天池官网，将鼠标移到 天池学习 ，即可出现下拉列表，点击 天池读书会 ，即可进入天池读书会的页面。



3) 在天池读书会页面，你可以对对应的读书会图书进行提问，优秀的提问还有机会获得赠书，还可以点击配套的训练营或者课程资源进入学习，还有点击实践代码获取读书会的项目实践的代码，跟着我一起进行项目实践和代码学习，同时还有很多其他的读书会，大家也可以观看举办过的读书会的回放，或者预约还没开始的读书会。



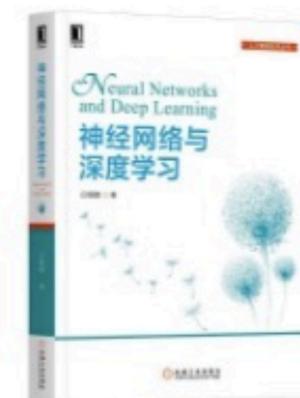
## 邱锡鹏 复旦大学计算机科学技术学院教授

直播主题 《神经网络与深度学习》

直播时间 2021年3月25日 20:00

学习资料 《神经网络与深度学习》课程

实践项目 fastNLP



[🗨️ 提问](#) | [📖 学习课程](#) | [🛒 购买地址](#) | [📄 PPT下载](#) | [👉 实践项目](#) | [🕒 预约直播](#)

谢谢观看，下次见

TIANCHI 天池

直播相关资料获取及回放查看地址：<https://tianchi.aliyun.com/specials/promotion/activity/bookclub>