school of ai

Welcome to Beijing School of AI !

第五课：RNN与NLP

Siraj Raval
Directeur, The School of AI
AI Educator, Best-Selling Author, Youtuber

by Siraj Raval

SoAI项目Github地址

BIG DATA DIGEST
大数据文摘

清华大学 Tsinghua University 清华-青岛数据科学研究院

school of ai Beijing

内容提要

一．课程回顾-神经网络与CNN

二．内容简介

a. RNN系列

b. NLP基础

c. 应用

a. 文本挖掘

b. 聊天机器人

三．集中讲解-问题解答

四．总结

章节	题目	预习资源	正式解读
介绍	深度学习入门指南	请完成ppt里提到的作业，长期	这节课长达2h，内容非常多，覆盖深度学习入门几乎全部知识点
第一课	宠物图片分类,英文视频,中文版笔记	Github代码	安装fastai环境并动手实现，代码解析：Github本地,Colab,Kaggle
第二课	特征工程及SGD,英文视频,中文版笔记	Github代码	请提前预习左侧资源
第三课	多标签分类,英文视频,中文版笔记	Github代码	-
第四课	NLP&推荐系统,英文视频,中文版笔记	Github代码	-
第五课	从反向传播到神经网络,英文视频,中文版笔记	Github代码	-
第六课	正则化卷积,英文视频,中文版笔记	Github代码	-
第七	Resnets、GAN等,英文视频,中	Github代码	-

一 课程回顾

- 神经网络与CNN-作业
  - 纯Python实现前馈神经网络，求解异或
  - 实现CNN分类（minist-fashion）
- 测试
  - 神经网络典型结构有哪些
  - 前馈神经网络的基本组件
    - 激活函数,BP,SGD,loss
  - 常见工具包及区别

第三课作业

温故

- 3brown1blue的神经网络系列视频，至少看一遍，建议做笔记，记录到自己的github上，把链接贴上来
- 参考课堂上的示例代码，亲自用python实现神经网络，求解异或问题
  - 最好把nilson的神经网络与深度学习这本书买下来，好好研究
- 学一门深度学习工具包：tensorflow、pytorch、keras三选一，实现mnist分类

知新

- 主题：RNN与NLP，对应fastai第四课
- 案例：
  - 文本挖掘
  - 聊天机器人

其它

- 第一课的宠物分类已经有6人提交作业，好几个人不只是执行代码，还做了进一步的优化：正则表达式、学习率精调、分组实验等，最终王瑞华的准确率最高95.66%
- 优化思路还有不少，比如上次课提到的inception v4网络
- 欢迎大家继续比拼下去，看谁最终取胜

school of ai Beijing

一 课程回顾

- 神经网络与CNN-作业
  - 纯Python实现前馈神经网络，求解异或
  - 实现CNN分类（minist-fashion）
- 测试
  - 神经网络典型结构有哪些
  - 前馈神经网络的基本组件
    - 激活函数,BP,SGD,loss
  - 常见工具包及区别

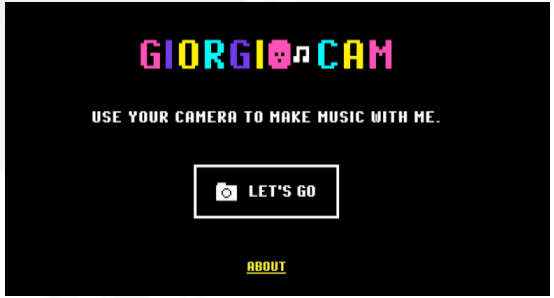
准确率比拼

- 已经有6人提交作业，好几个人不只是执行代码，还做了进一步的优化：正则表达书、学习率精调、分组实验等，最终王瑞华的准确率最高95.66%

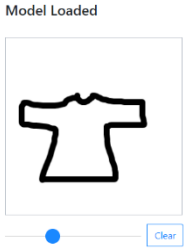
学员	准确率	其它
李婧	94.7%	code,多个维度，3组实验
李婧华	94%	-
周启红	-	只有代码没有数据
ztq222-周天奇	mnist上实现	-
王瑞华	95.66%	正则表达式，最高分
jllstone	94.8%	调学习率

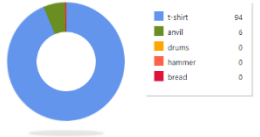
- 优化思路还有不少，比如上次课提到的inception v4网络
- 目前王瑞华的宠物分类模型准确率最高95.66%
- 谁能进一步提升?
- 欢迎大家继续比拼下去，看谁最终取胜

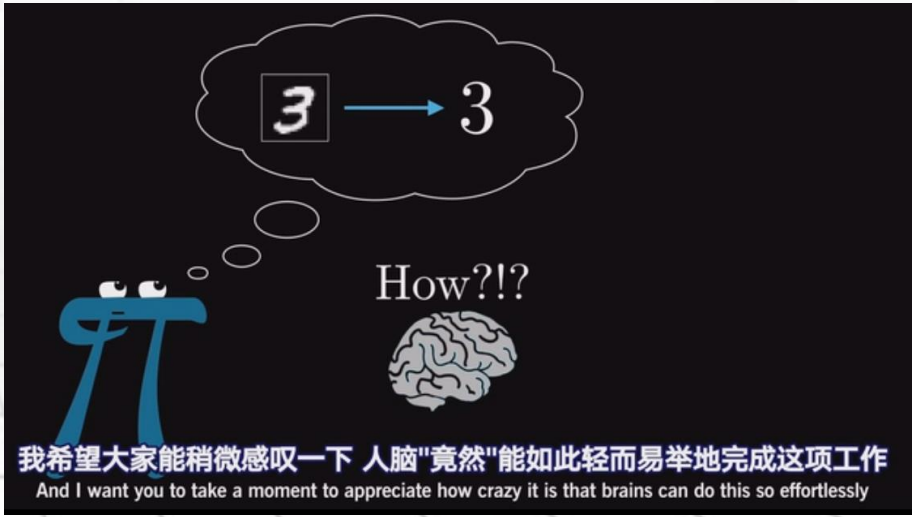
一 课程回顾



- 神经网络与CNN-作业
  - 纯Python实现前馈神经网络，求解异或
  - 实现CNN分类（minist-fashion）
- 测试
  - 神经网络典型结构有哪些
  - 前馈神经网络的基本组件
    - 激活函数,BP,SGD,loss
  - 常见工具包及区别

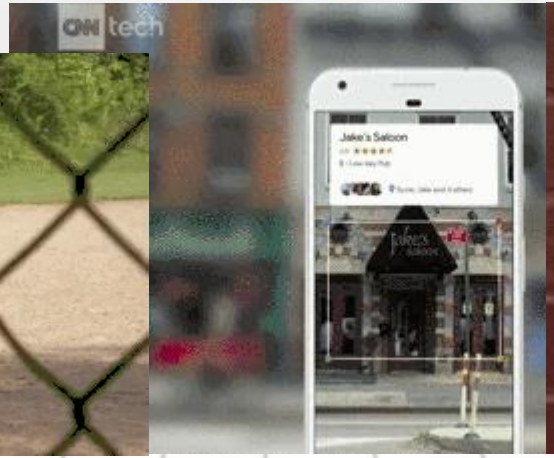


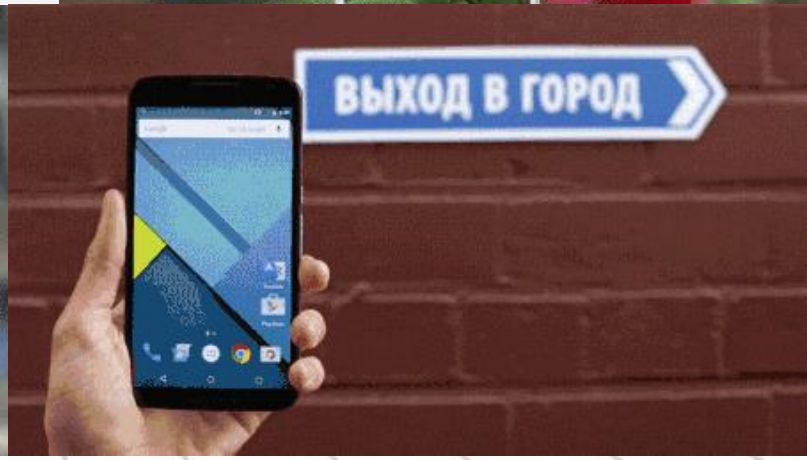




AI应用-计算机视觉

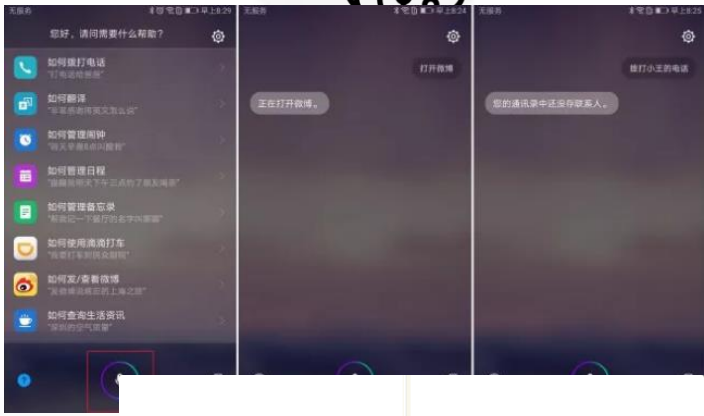
- 应用（AI改变生活）
  - 图像、语音、文字、游戏、无人驾驶等领域
  - NLP：2018-5-8，Google I/O大会
    - Google Assistant秒杀Siri和Cortana

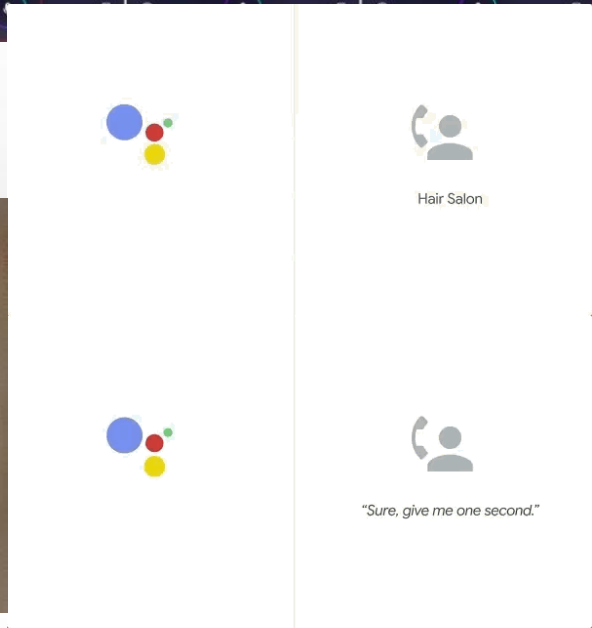




AI应用-机器人

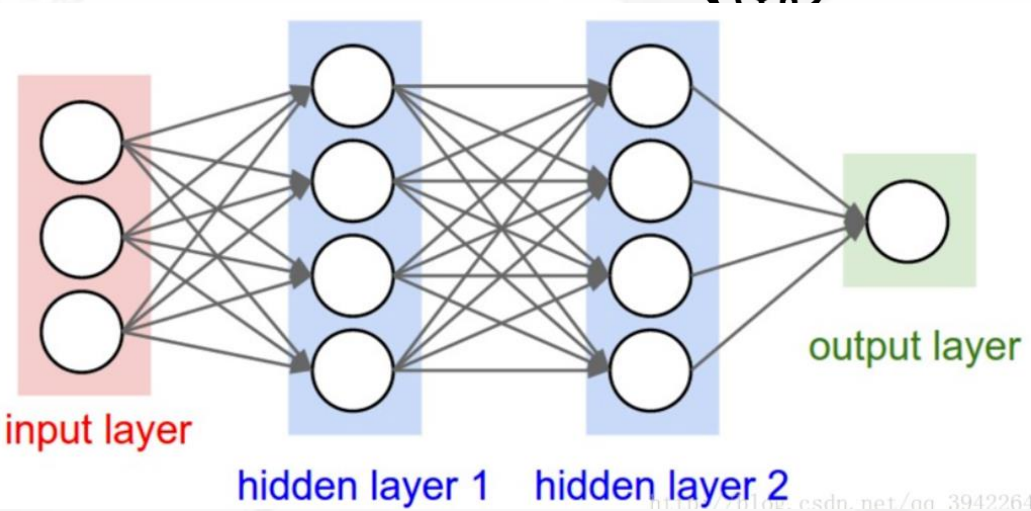
- 应用（AI改变生活）
  - 图像、语音、文字、游戏、无人驾驶等领域
  - NLP：2018-5-8，Google I/O大会
    - Google Assistant秒杀Siri和Cortana
  - 2018-7-4，百度，语音机器人





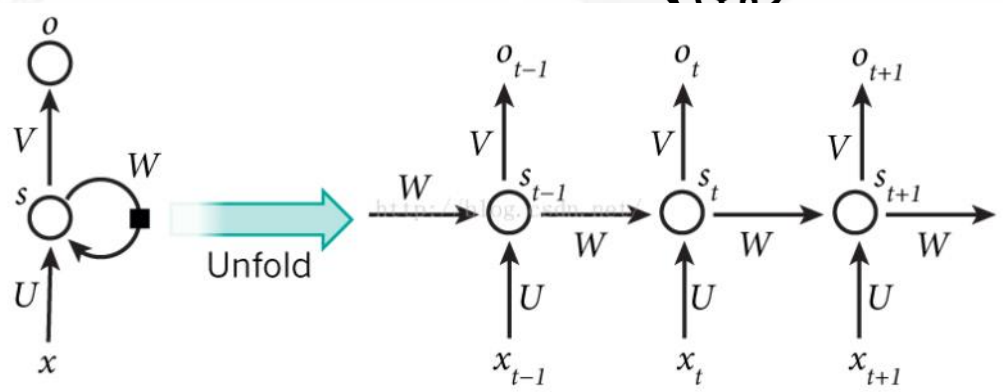
二 内容简介

- **为什么需要RNN?**
- 假设
  - 元素之间相互独立
  - 神经元，输入与输出
- 问题
  - 现实生活中，不一定成立
  - 股票预测
  - 语音识别
- 案例
  - 我出生于黄冈，长江边上，到处山川河流，湖泊密布，作为___人，不太习惯帝都的干燥
  - 这几天好热，我是___人，帝都太干燥，受不了
  - 本质：像人一样拥有记忆能力

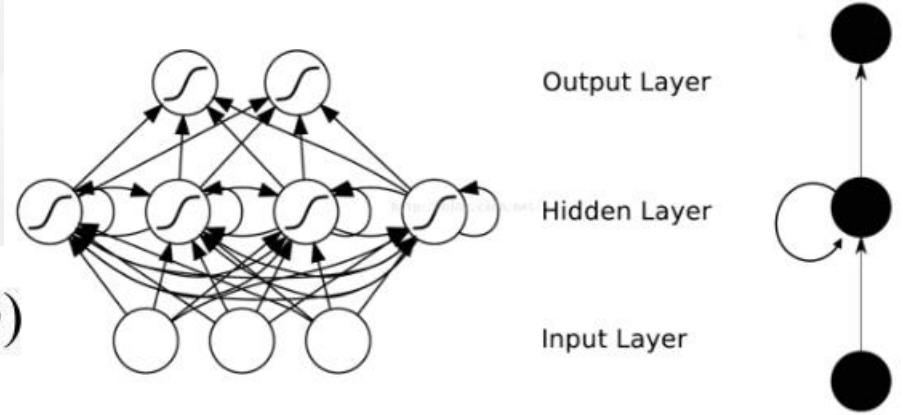


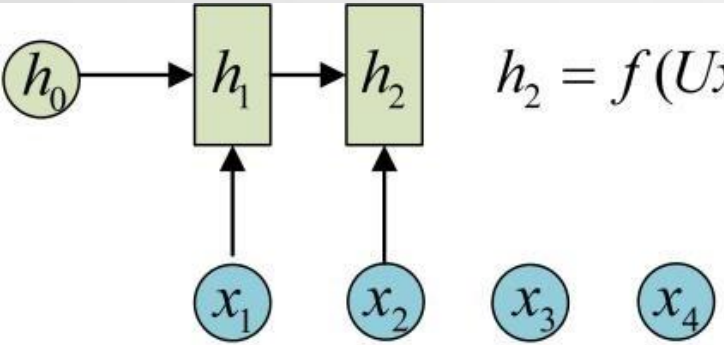
二 内容简介

- **RNN基本结构**
- 假设
  - 元素之间相互独立
  - 神经元，输入与输出
- 结构：前后依赖
- 参考：
  - 循环神经网络（RNN）原理通俗解释
  - TensorFlow中RNN实现的正确打开方式



$$S_t = f(U * X_t + W * S_{t-1})$$

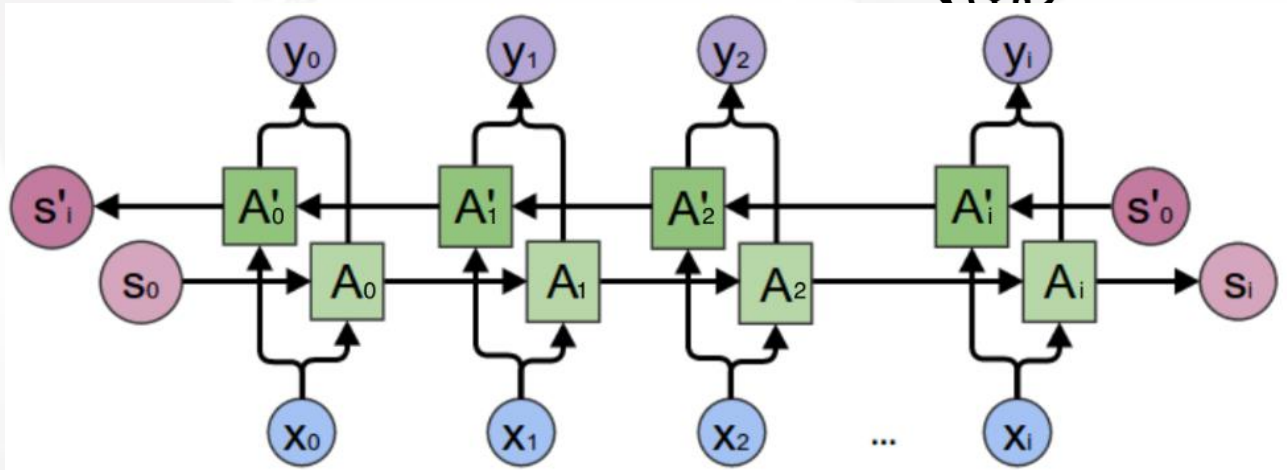




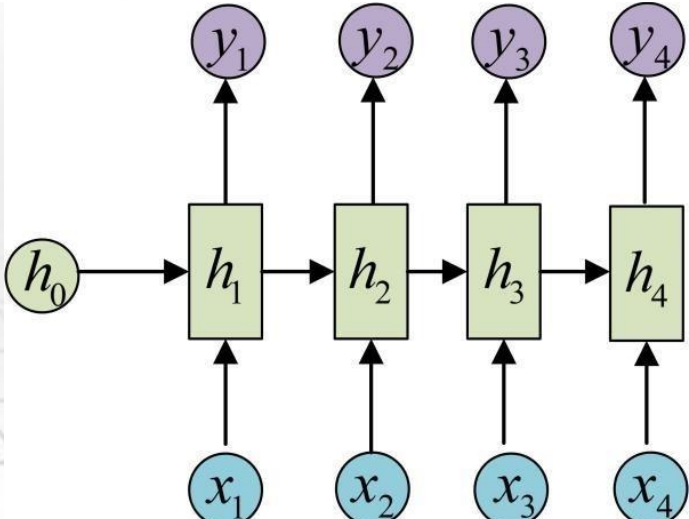
$$h_2 = f(Ux_2 + Wh_1 + b)$$

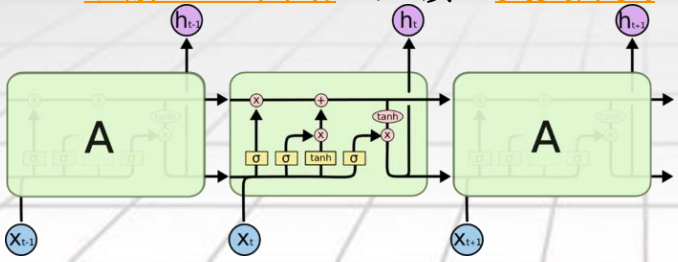
二 内容简介

- **RNN基本结构**
- 结构：前后依赖
- 类型：
  - 简单RNN
  - 双向RNN
  - LSTM，GRU。。。
- 参考：
  - 循环神经网络（RNN）原理通俗解释
  - TensorFlow中RNN实现的正确打开方式
  - 理解LSTM网络（必读）英文原文



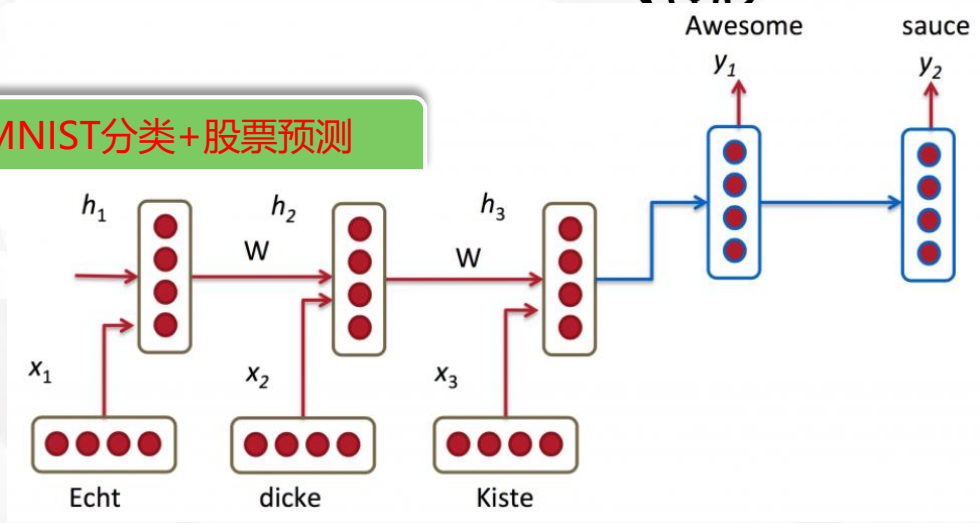
作业：学习RNN+LSTM理论知识

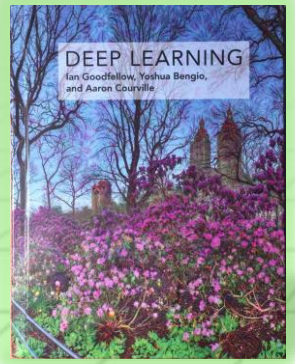


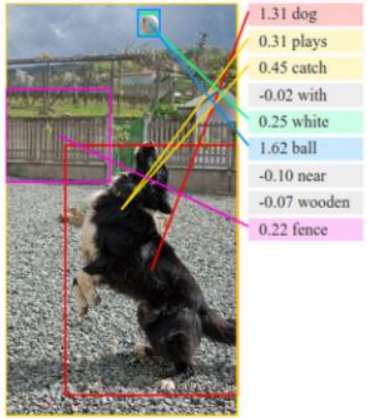


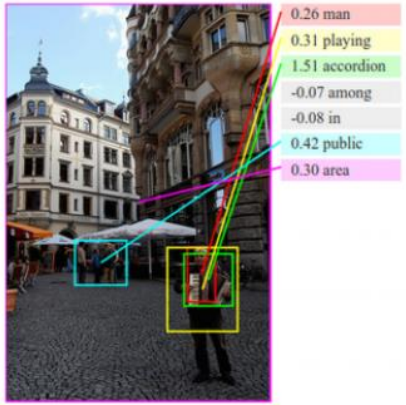
- **RNN应用**
  - 聊天机器人
  - 机器翻译
  - 语音识别
  - 图像描述
- 实战示例
  - RNN用于MNIST分类
  - LSTM股票预测

作业：动手实现MNIST分类+股票预测



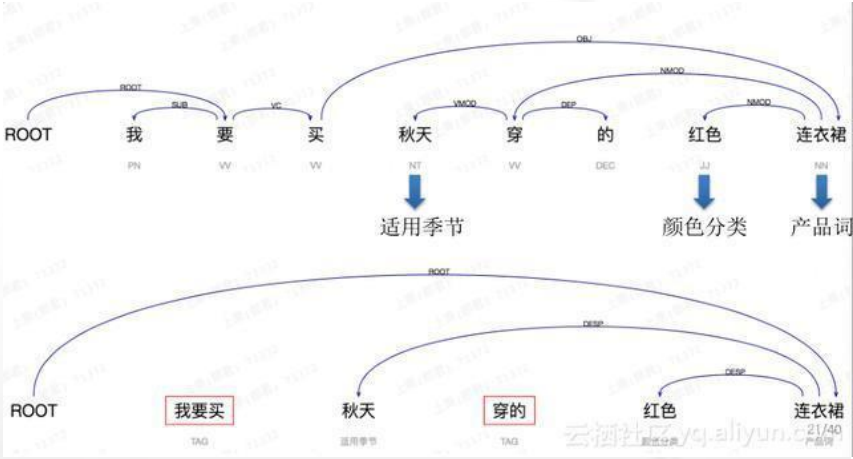


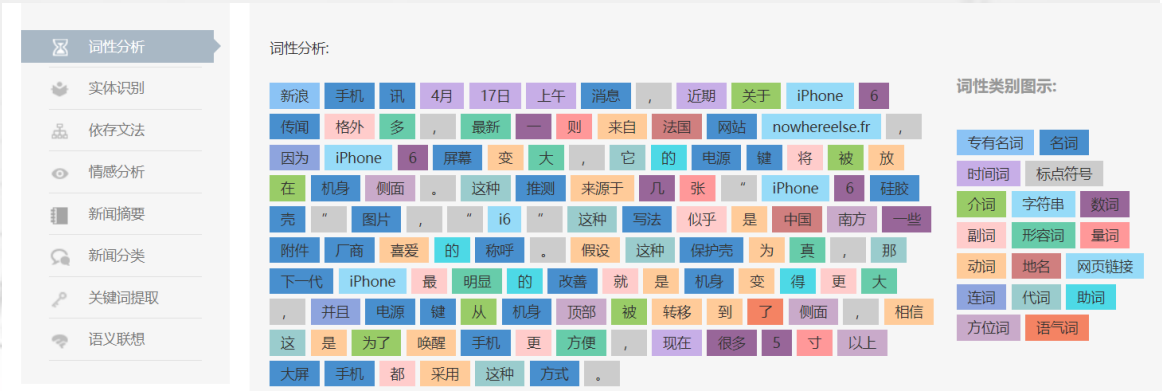




三 NLP

- 自然语言处理（NLP）基本问题
  - 分词（word segmentation）
  - 词性标注（part of speech）：POS
  - 命名实体识别：NER
  - 指代消解
  - 自动摘要
  - 文本分类
  - 句法分析
  - 问答系统
- 参考
  - NLP学习总结
  - 十分钟NLP概述
  - BosonNLP在线示例
  - Stanford CoreNLP

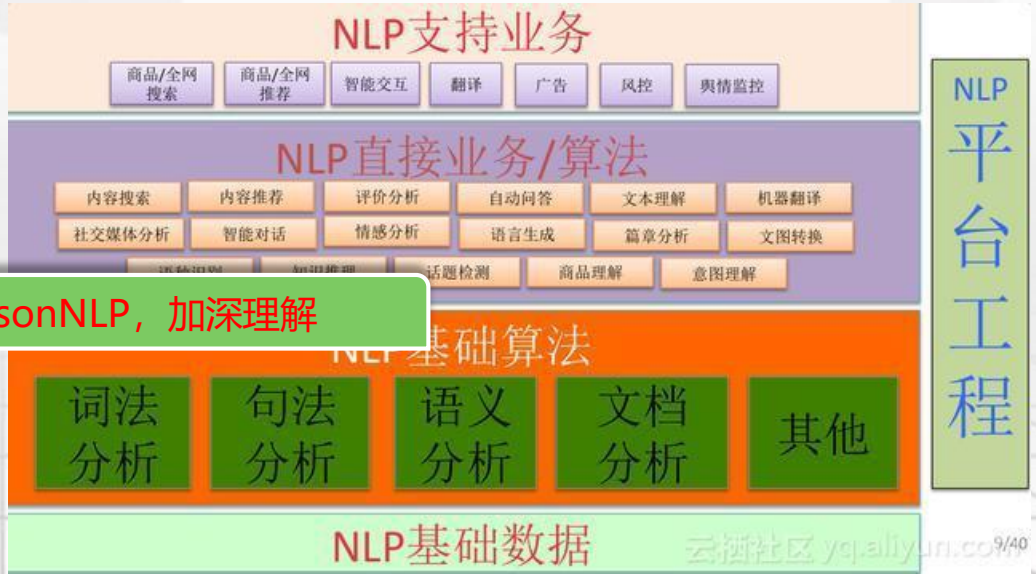




三 NLP

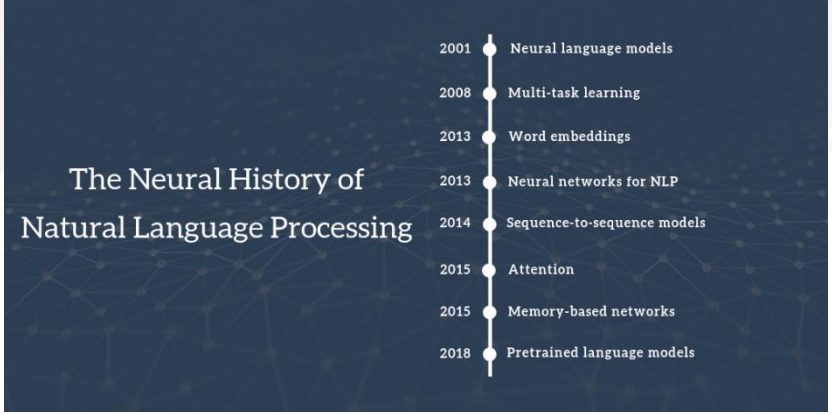
- 自然语言处理（NLP）基本问题
  - 分词（word segmentation）
  - 词性标注（part of speech）：POS tagging
  - 命名实体识别：NER
  - 指代消解
  - 自动摘要
  - 文本分类
  - 句法分析
  - 问答系统
- 参考
  - NLP学习总结
  - 十分钟NLP概述
  - BosonNLP在线示例
  - Stanford CoreNLP

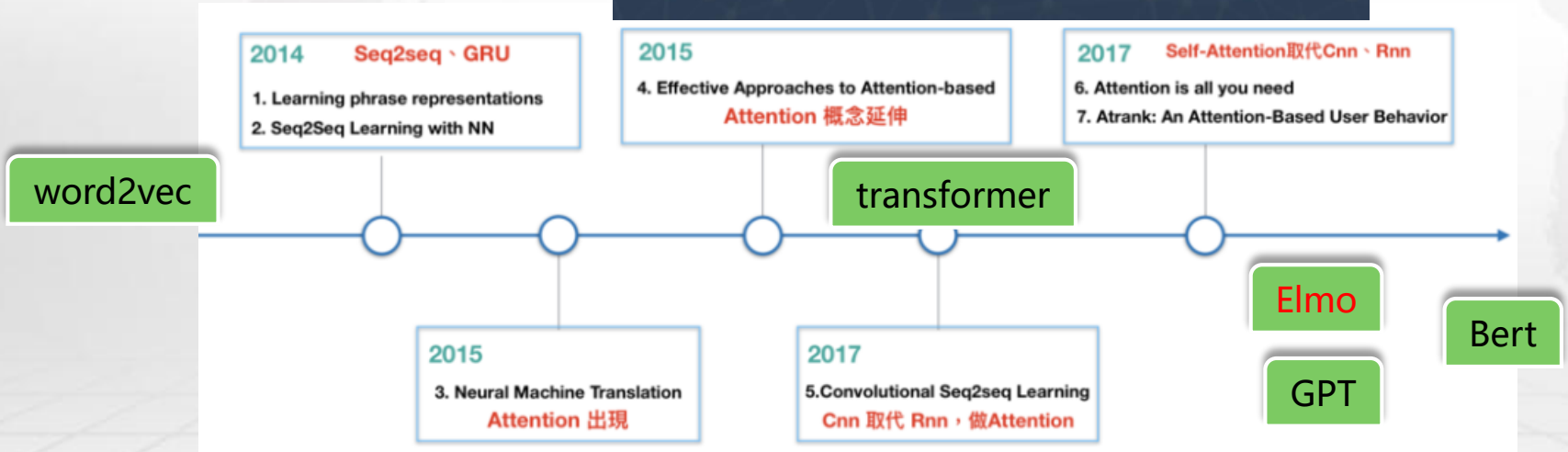
作业：体验bosonNLP，加深理解



三 NLP

- NLP技术发展





三 NLP

词性分析: 查看文档 结果不正确

大舅 去 二 舅家 找 三 舅 说 四 舅 被 五 舅 骗 去
六 舅 家 偷 七 舅 放在 八 舅 柜子 里 九 舅借十舅 发给
十一 舅 工资 的 1000 元 ?

词性类别图示:

名词 动词 数词
介词 方位词 助词
量词 标点符号

- NLP挺难。。。
  - 究竟谁是小偷？大舅去二舅家找三舅说四舅被五舅骗去六舅家偷七舅放在八舅柜子里九舅借十舅发给十一舅工资的1000元？

— Text to annotate —

大舅去二舅家找三舅说四舅被五舅骗去六舅家偷七舅放在八舅柜子里九舅借十舅发给十一舅工资的1000元?

— Annotations —

parts-of-speech × named entities × dependency parse × openie ×

— Language —

Chinese

Submit

Part-of-Speech:

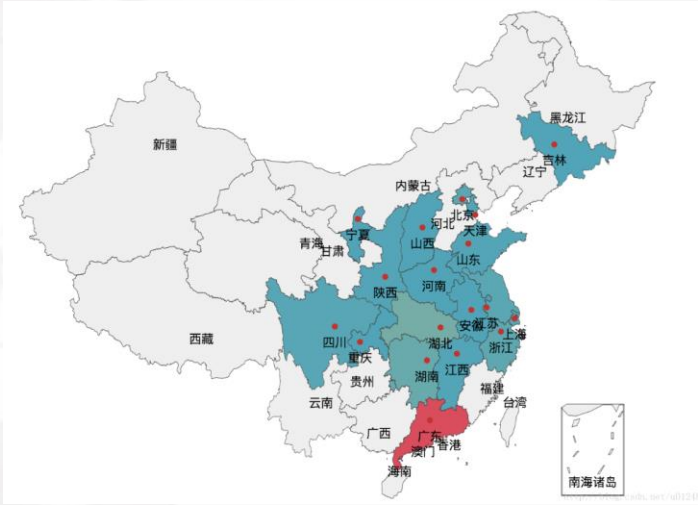
1 大舅 去 二 舅家找三舅说四舅 被 五 舅骗 去 六 舅家 偷 七 舅 放在 八 舅 柜子 里 九 舅 借十舅 发给 十一 舅 工资 的 1000 元 ?

Named Entity Recognition:

1 大舅 去 二 舅家找三舅说四舅 被 五 舅骗 去 六 舅家 偷 七 舅 放在 八 舅 柜子 里 九 舅 借十舅 发给 十一 舅 工资 的 1000 元 ?

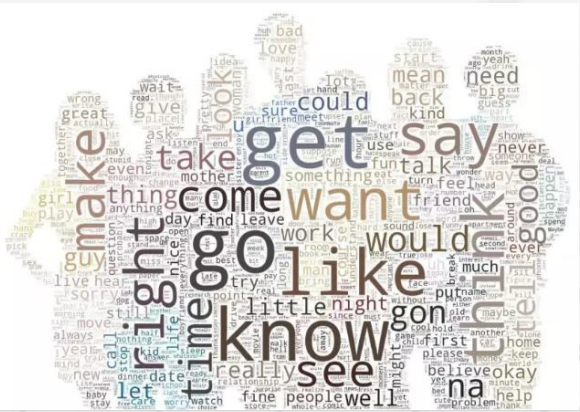
四 应用-文本挖掘

- 微信朋友圈分析
  - 通过itchat抓取朋友圈数据
  - 性别、地狱、签名分析
  - python分析微信朋友圈
  - Python解密微信大数据



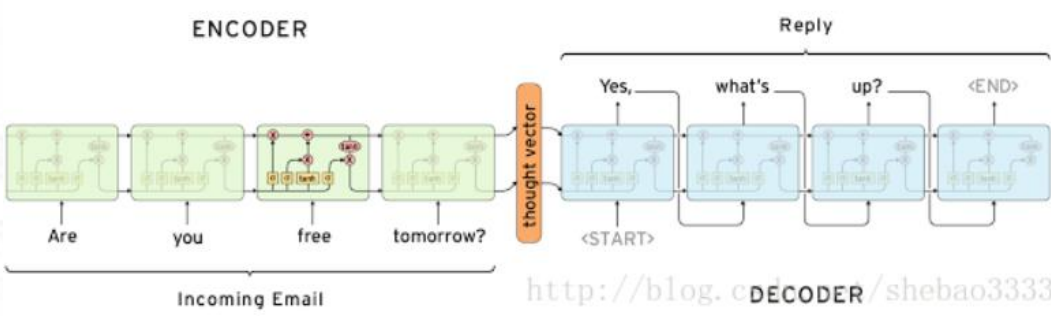
四 应用-文本挖掘

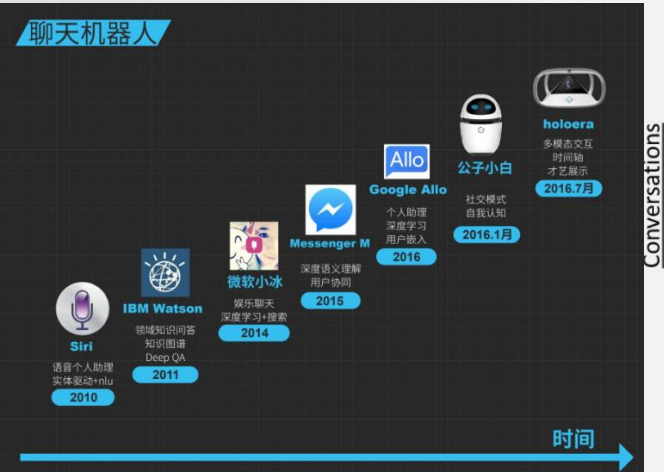
- 大数据文摘近期文章
  - 卅年春秋，谁主沉浮？从400篇任正非演讲稿分析中，一探华为
  - 临别给《生活大爆炸》做个台词数据分析，你猜谢耳朵最爱说什么？
- 代码
  - Big bang挖掘（结果展示），华为讲话稿挖掘

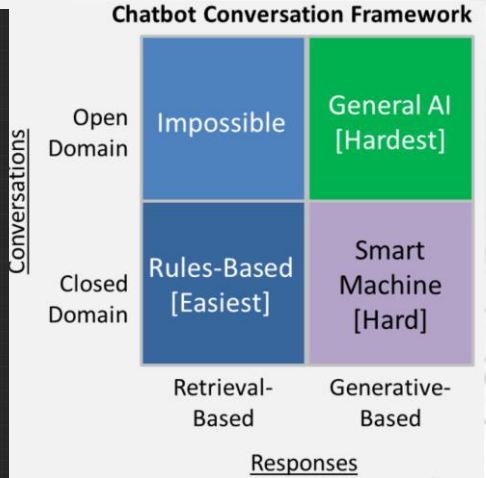


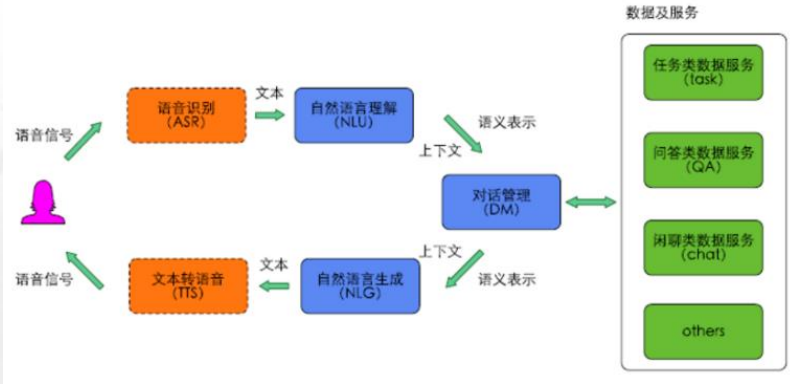
四 应用-聊天机器人

- 实现方法
  - 索引式，生成式
- 聊天机器人进阶
  - 微信朋友圈抓取









四 应用-聊天机器人

- 估值一个亿的AI核心代码

```
"""
AI 核心代码，估值 1 个亿
"""
while 1:
    print('AI 说：' + input().strip("吗？？") + "!")
```

while 1

test (2)

```
在吗？
AI 说：在!
你好
AI 说：你好!
能听懂汉语吗？
AI 说：能听懂汉语!
真的吗？
AI 说：真的!
```

AiMain.java

```
/**
 * AI核心代码，估值1个亿
 */
public class AiMain {
    public static void main(String[] args) {
        Scanner sc = new Scanner(System.in);
        String str;
        while (true) {
            str = sc.next();
            str = str.replace( target: "吗",  replacement: "");
            str = str.replace( target: "?",  replacement: "!");
            str = str.replace( target: "？",  replacement: "!");
            System.out.println(str);
        }
    }
}
```

AiMain › main()

Run: AiMain AiMain AiMain

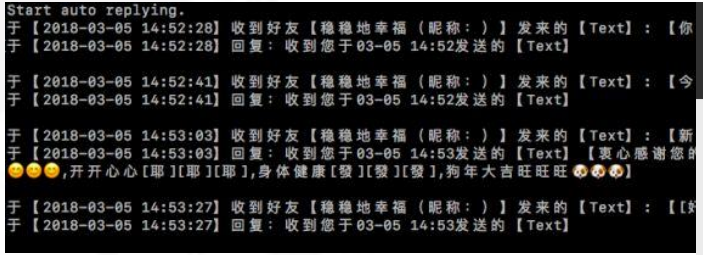
```
在吗?
在!
你好
你好
能听懂汉语吗?
能听懂汉语!
真的吗?
真的!
```

四 应用-聊天机器人

- AI公司估值泡沫-《大腕儿》

四 聊天机器人

- 微信自动回复
  - 接入图灵机器人（或微软小冰）
  - Itchat微信自动回复



作业：分析自己的朋友圈

四 聊天机器人

自己实现自动回复

- 聊天机器人对战
  - 优酷地址
  - 天猫精灵、小爱音箱
- Seq2seq训练
  - charRNN-TF
- Transformer chatbot


```
PS C:\Users\wqw\Desktop\tantan\fastai\好玩儿的python\Python-learning\courses\chapter_8> python .\chatbot.py
请主人输入话题：(随便什么词) 0
轮数    图灵(问)                                        青云客(答)
--------------------------------------------------------------------------------
1       图灵： 1
                                                        菲菲：3
2       图灵： 4
                                                        菲菲：你说个“4”是什么意思
3       图灵： 你说的话我听不懂说明白点
                                                        菲菲：听不懂就算啦，只能说你笨
4       图灵： 我也是这么说我
                                                        菲菲：哦哦，知道了
5       图灵： 你知道的太多了，哈哈。
                                                        菲菲：这样就算多了？菲菲可不认同网友的话哦
6       图灵： 嘻嘻，能用菲菲这种称呼的关系就比较亲密啦！
                                                        菲菲：有什么好笑的
7       图灵： 看到你，开心就想笑的
                                                        菲菲：那你看屁
8       图灵： 对，我在静静的看你。
                                                        菲菲：哗啦啦啦的
                                                        菲菲：疯狂的遊戲需要誰准許
12      图灵： 你这是怎么了？
                                                        菲菲：心情不太好，你别问了好吗
13      图灵： 你总要说出来嘛，我来安慰你。
                                                        菲菲：不来了
14      图灵： 报销路费
                                                        菲菲：我都係
15      图灵： 系啥啊你
                                                        菲菲：呵，你试试，
16      图灵： 咋试
                                                        菲菲：大爷您没问题啊，甭复习了啊！
17      图灵： 昨晚预习今早复习还真勤奋
                                                        菲菲：快去读书吧，好好学习哦
18      图灵： 天天向上
                                                        菲菲：你个笨猪会学习么
19      图灵： 非要教坏人家你才满意？诛心哟人家才不回答你这个问题呢
                                                        菲菲：生死相許
20      图灵： 自古多情空余恨！
                                                        菲菲：{face:51}说点高智商的话吧，再怎么着，我也是梅州行的大美人
21      图灵： 梅州往大的范围讲属于哪个地方呢？
```

作业：自己尝试训练seq2seq模型，并对接到以上代码

反思

- 机器会取代人类吗?
  - Hanson：一定会，未来30年
  - 小度：不会，机器只是辅助
    - 参考：**艾诚中英文切换超自如 对话网红机器人获赞**

人类还很无知，连大脑工作原理还没研究清楚
不要过于迷信AI，但也不要妄自菲薄
人类是迷雾中前行的勇者，未来充满变数

二 分组讨论

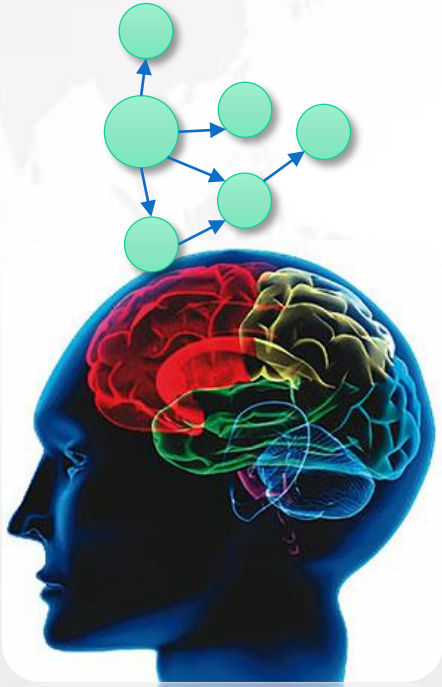
- 课程概要：https://github.com/wqw547243068/school-of-ai-beijing/blob/master/README.md
- 要点
  - 成员：成员背景、编程功底、意图、组长
  - 温故：上次课讲了什么，印象最深的几点
  - 知新：分享本次自习心得
  - 思考：遇到哪些问题，怎么解决
- 注意
  - 时间控制在10min以内
  - 每个子类目不超过3个问题
  - 尽量不要重复——先到先得
- 组长
  - 协调、监督组员进度
  - 结伴学习，共同成长

三 集中讲解

- 知识点
  - 机器学习流程
  - GPU
  - 图像处理
  - 卷积神经网络
  - 。。。

school of ai
Beijing

如何高效学习

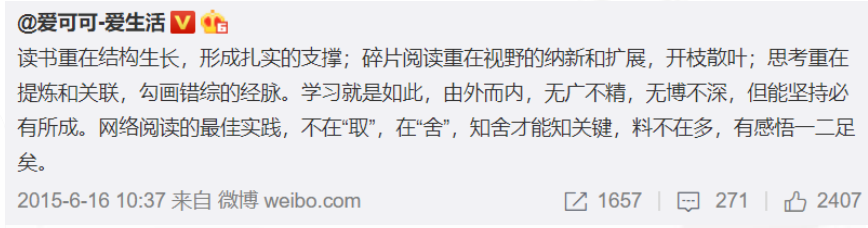


互联网时代学习之道：

- 多看：知识图谱
  - 系统阅读：结构生长
  - 碎片阅读：开枝散叶
- 多动手：验证
  - 消化理解，提升留存
  - 开花结果，学以致用
- 多思考：关联推理
  - 提炼关联，查缺补漏
  - 不断完善知识图谱
  - 沉淀：自己做笔记
- 其他：好奇心+分享+上进心

大脑的学习之道：

- 图谱结构+注意力+联想记忆+推理反思



school of ai
Beijing

翻转课堂

- 约法三章
  - 随机分组：全部学员5-6人一组，随机分配，尽量均衡，组团学习
  - 组内互助：选组长，督促学员学习，相互帮助
  - 组间竞争：每次课会对表现优秀的组加分，动态排名
  - 奖惩分明：最后一名自觉给第一名买奖品（零食、红包等）
- 准备工作
  - 分组：报数
  - 去中心化：围着讲台散开
  - 积分榜
  - QQ学习群（左图）：The School of AI
    - 资料共享、作业发布
  - QQ讨论组（右图）
    - 如果需要共享自己的屏幕，需要使用讨论组
  - 添加小助手alpha微信：**xiniuedu5（右上角二维码）**

THANKS

王奇文-wqw547243068@163.com