

# 科大龙明康访谈学习

作者: [flowaters](#)

原文链接: <https://ld246.com/article/1527492911245>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

# 背景

今天被一篇《科大讯飞两代AI平台演进之路--讯飞云计算研究院副院长龙明康访谈》的文章刷屏了。这篇文章的作者是龙明康老师。

本文记录一下读书笔记。

# 关键词

讯飞语音云

- 日均PV：从百万到千亿
- 平台化：AIUI（智能人机交互系统）、AIoT（智能物联网）平台
- 开发者数量：80万+
- 代表：讯飞语音输入法、锤子发布会语音输入

# 关键词解释

UI是user interface的缩写，即人机交互界面。AIUI即AI的UI。

如语音UI为VUI(voice-user interface)。

# 语音识别

## 现状

语音识别如果在理想的环境下是比较好做的，业界现在在理想场景下的语音识别率差距越来越小。

## 难点

1. 如何解决复杂的环境噪音、用户带口音、远场等情况下的识别率低的问题
2. 做通用领域的识别相对好做，而专业领域就比较难做，比如医疗、法律等方向
3. 做用户级的个性化声学模型、语言模型，深度学习当前大多数还是有监督学习，所以如何结合应用场景来降低标注成本也是很有挑战的事情

## 技术

1. 自研CNN结构的声学模型
2. Encoder-Decoder方案

## 坑

1. 识别率的优化问题

早期大家喜欢在实验环境做大量的优化，达到85%的识别率后，才上线。实际上线后统计，发现只有6

%识别率。所以后来吸取了这些经验，采取先上线，利用真实数据快速迭代的方法。

## 产品

### 平台化、标准化

语音云大概是在2009年由于继栋大大提出并启动项目，早期讯飞的AI能力面向toB销售，现在看来，时的模式就是私有云的方式。随着对接越来越多，需求定制和技术支持的工作已经不堪重负，大家意识到平台化、标准化才是出路。

### 语音输入典型产品

- 讯飞输入法
- [讯飞开放平台：魔飞](#)

## 技术

### 算法

- 神经网络
- 决策树
- 支持向量机等
- 深度学习

### 框架

- TensorFlow：生态很全面，上手简单，但是性能不够好
- MXnet：性能优化的比较好，节省显存，运算效率高。另外MXnet只做训练

## 各领域难点

- 智能客服、机器人：偏语义理解领域

## AI三次浪潮

### 第一次

出现了很多顶级算法，但是这些算法只能解决狭窄领域的问题，而且当时的计算能力是严重不足的，以进入第一次冬天。

### 第二次

出现了语音识别、机器翻译、专家系统、类神经网络，但是效果完全达不到人们对AI的预期。

## 第三次

依赖大数据技术、深度学习技术的成熟，且在计算能力大幅增加。

当前应该已经到了AI应用爆发的时期，随着大家对AI的理解越来越深刻，AI已经被应用到越来越多的分领域，从讯飞开放平台的开发者增长可见一斑。我认为当前的主要瓶颈还是AI人才短缺，这也包括AI的业务专家，他们才是把AI带到世界每个角落的发动机。

## AI工程师之路

### 入门

#### 了解现状

大概清楚当前AI的一些现状，能干什么，大概能做到什么程度

要了解这个也比较简单，从讯飞开放平台上就可以了解到很多

#### 了解生产环节

理论研究、模型训练、引擎工程化、服务化，训练的部分还有个深度学习平台需要建设

### 进阶

#### 分布式服务经验

有一些分布式服务经验的工程师可以从服务化入手，逐渐了解引擎的一些特性和运作机制。

#### 算法功底经验

如果已经具备一些算法功底，包括图论、概率论相关知识就可以转做引擎工程化。

引擎工程对性能有非常高的要求，所以需要深入计算机体系结构，结合CPU/GPU/memory优化。

在这个过程中就可以深入探索下模型是怎么来的。

通常这个时候可以去尝试使用TensorFlow在一些开源的模型上进行调参，要调好参其实还是非常有门槛，需要具备设计模型的能力，这样的人在讯飞AI研究院都是非常资深的级别了。

### 再进阶

搞特征工程了，偏理论研究，需要深入研究泛函、矩阵分析、概率图模型、随机过程、优化理论等。

对于这个进阶过程，讯飞内部都有各个阶段的培训材料，在讯飞开放平台的AI大学中也有部分分享课。

### 参考

- 科大讯飞两代AI平台演进之路--讯飞云计算研究院副院长龙明康访谈
- 浅说语音用户界面：VUI+GUI