



人工智能与信息社会

人工智能应用：语音识别——个人助理

陈斌 北京大学 gischen@pku.edu.cn

语音识别技术

- › 语音识别技术，也被称为自动语音识别 Automatic Speech Recognition(ASR)，其目标是将人类的语音中的词汇内容转换为相应的文字。



语音识别基本原理

- › 我们知道声音其实是一种波，如果要对于声音进行分析，就需要对于声音进行分帧。
- › 也就是把声音按照时间切成若干个小段，每一小段称为一帧。



语音识别基本原理

- › **单词的发音由音素构成。音素是语音中的最小的单位，依据音节里的发音动作来分析，一个动作构成一个音素。**
- › **对英语，一种常用的音素集是卡内基梅隆大学的一套由39个音素构成的音素集。汉语一般直接用全部声母和韵母作为音素集。**
- › **状态：比音素更细致的语音单位。通常把一个音素划分成3个状态。**

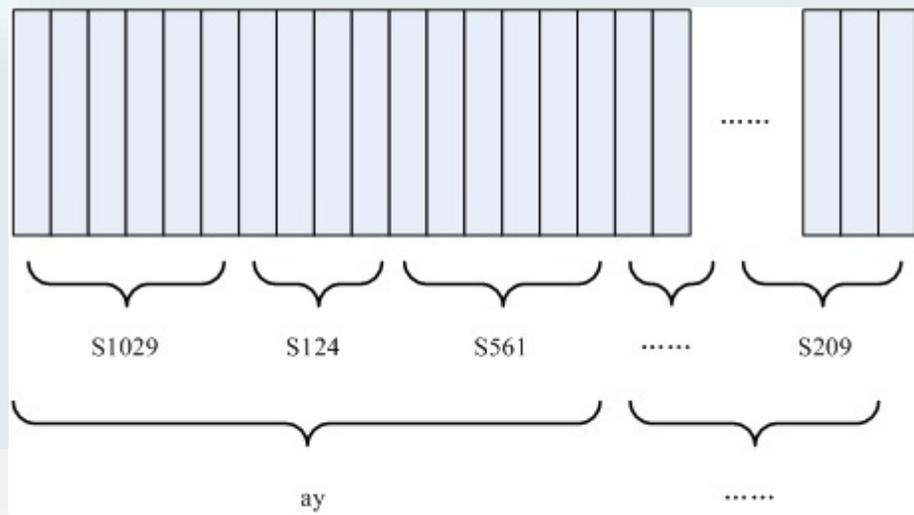
语音识别基本原理

- › 1、语音分帧 2、声学特征提取
- › 3、把帧识别成状态 4、把状态组合成音素
- › 5、把音素组合成单词



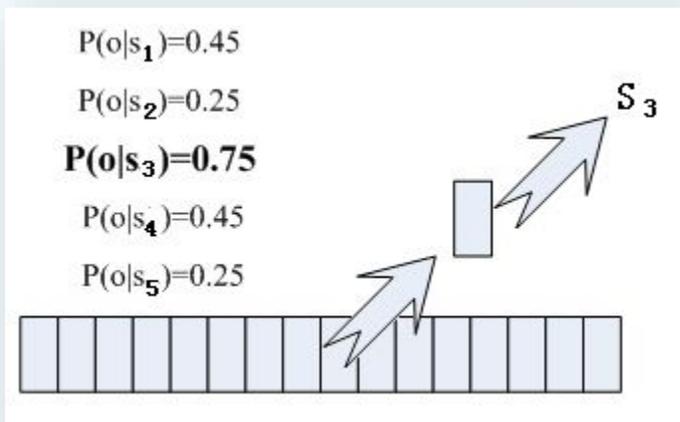
语音识别基本原理

- › 每个小竖条代表一帧，若干帧语音对应一个状态，每三个状态组合成一个音素，若干个音素组合成一个单词。
- › 一般认为某帧对应哪个状态的概率最大，那这帧就属于哪个状态。



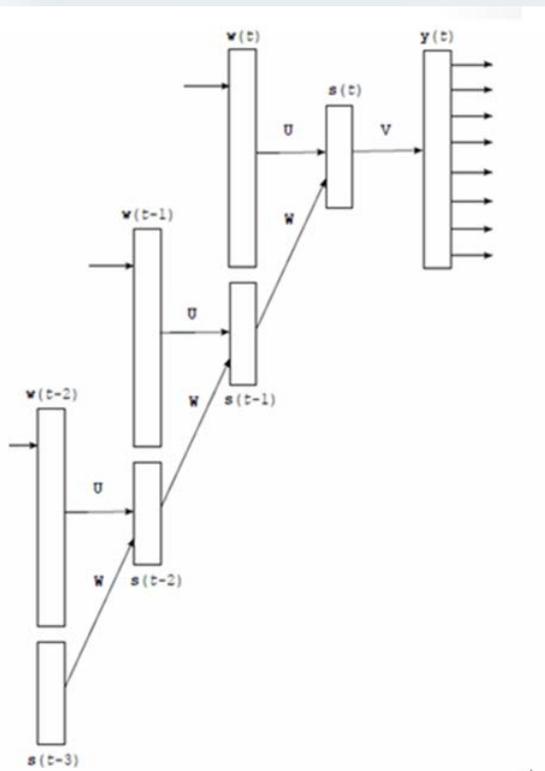
神经网络在语音识别中的应用

- › 通过巨大数量的语音数据，经过神经网络监督训练得到“声学模型”。
- › 通过声学模型里的参数，就可以知道帧和状态对应的概率。



神经网络在语音识别中的应用

- 神经网络不仅可以编码最近的几个词，还可以把前文中的所有词（称作“历史”）中的各种信息都作为输入特征。
- 由于历史是一个序列，可以采用递归神经网络（RNN）来建立声学模型。



语音识别应用

- › 语音识别技术的应用包括语音拨号、语音导航、室内设备控制等。
- › 语音识别技术与其他自然语言处理技术如机器翻译及语音合成技术相结合，可以构建出更复杂的应用，例如同声传译。



语音识别应用：语音输入法

- › 通过语音识别输入文字，最高速度能够达到1分钟400字，比普通键盘输入更加高效。
- › 科大讯飞的语音输入，不仅支持中文录入、中文转英文等功能，还支持粤语、四川话、东北话、上海话、闽南语等多种方言输入。



语音识别应用：个人助理



语音识别应用：个人助理

- › 小冰是微软推出的一个人工智能聊天机器人，已经可以创作诗歌、撰写新闻、主持节目，已在北京人民广播电台开播节目。
- › 无论从用户数量、活跃度还是交互流量来看，微软小冰均是目前全球最大规模流量的对话式人工智能产品。



语音识别应用：个人助理

- › Siri是一款内置在苹果iOS系统中的人工智能助理软件。
- › 利用自然语言处理技术，用户可以使用自然的对话与手机进行交互，完成搜索数据、查询天气、设置手机日历、设置闹铃等许多服务。



语音识别应用：个人助理

- › **Google Assistant 结合了谷歌积累多年的技术，其“持续性对话”功能让机器与人的交流更为自然。**
- › **智能语音助手和搜索引擎是相辅相成的，更聪明的搜索逻辑能够更快的帮助用户找到答案。**

