阿里AI Labs王刚解读9小时卖出百万台的“天猫精灵” | 高山大学（GASA）

原标题：阿里AI Labs王刚解读9小时卖出百万台的“天猫精灵” | 高山大学（GASA）

新智元推荐

来源：高山大学

【新智元导读】 11月14日的高山大学（GASA）思享课II期，阿里巴巴人工智能实验室首席科学家王刚教授为在场学员解读了“天猫精灵”这款产品以及阿里巴巴在人机交互上的突破，同时还就商业变现、与阿里生态系统的衔接、用户体验、语音交互大规模商用、竞争与合作等方面的问题，与学员们进行了深入的交流。在技术方面，他介绍了阿里巴巴的AliGene，特别是技术上的创新，比如CNN和LSTM的整合、自适应网络和意图识别和槽填充的联合模型等等。

*以下根据王刚2017年11月14日在高山大学（GASA）思享课II期的分享整理而成

在刚刚过去的“双十一”购物狂欢节中，短短9个小时之内“天猫精灵”智能音箱的销量突破了100万台，阿里掀起的这场价格战背后足以看出其对智能音箱市场的重视。在11月14日的高山大学（GASA）思享课II期，阿里巴巴人工智能实验室首席科学家王刚教授为在场学员解读了“天猫精灵”这款产品以及阿里巴巴在人机交互上的突破，同时还就商业变现、与阿里生态系统的衔接、用户体验、语音交互大规模商用、竞争与合作等方面的问题，与学员们进行了深入的交流。

语音成人工智能重要交互方式

“天猫精灵” 是阿里集团人工智能实验室今年八月份推出的一款智能音箱，也是我们在摸索下一代人机交互的第一款硬件产品。它拥有我们研发的语音系列技术，能够识别用户的语音并帮助人们转换为文字，同时还有自然语言理解的技术，能将用户的意图给出相应的结构化信息，然后再提供给用户相应的服务。

“天猫精灵” 虽然是一个很小的产品，可能跟我们的小手机差不多，但它的功能是非常丰富的，因为它的后台结合了非常多的服务，基本上包括了生活中的方方面面。从早上起来为你控制灯、窗帘，整理一天的安排，新闻的播放，出门的时候帮你打车，预定会议室，点外卖，查询航班信息，酒店里的快捷查询，到你的健身教练等等，帮助我们大大简化了生活中对服务和内容的获取。当然最重要的是作为一个音箱它不需要被禁歌，所以我们也有现在互联网上最全的版权，我们和腾讯达成了合作，基本上所有的歌的正版我们都能够拿到，所以这也是我们一些独特的优势。有了这样的产品之后，我们手机里面的APP等很多内容都可以被复制了。

同样是交互方式，为什么语音交互跟手机APP比更有优势？我们可以做一个对比：用手机APP听歌我们的步骤是需要打开手机解锁，找到APP，再用文字去输入歌名然后再点击播放，这个过程可能会耗时要一分钟或是更久；而天猫精灵可能只需要五秒就够了，它的快捷性和在效率提高方面的优势还是非常明显的。

语义理解平台（AliGenie 开放平台）

这是我们的**语音交互系统的架构图**，从设备端传来的语音信号在我们的网关收到后，开始一系列的理解和识别。从主链路来看，我们先通过语音识别，将声音转成文字，然后我们有一套语义理解的平台。这个平台会根据开发者的配置，对领域的设计、语料和词典的数据，对用户的这句话进行理解。如果需要调用第三方服务的意图，就会通过一个服务代理的机制，调用接入的第三方服务。这之后就会根据各个技能的配置，封装相应的答复语句，然后通过语音合成的服务，播报给用户。

值得一提的是，“天猫精灵”还支持**声纹识别的技术**，可以根据声音识别出不同的使用者，以此保证其安全性和私密性。这个技术在支付和用户个性化识别上都发挥了重要的作用。同时我们还提供了一整套的从端到云的解决方案，所以我们这套自然语言交互系统未来不会局限于“天猫精灵”，而是会集成到各种终端设备上，包括汽车，耳机等等。

用深度学习解决自然语义理解难题

自然语言人机交互是一种交互的媒介，这里面我们需要各行各业的开发者，把他们原来的服务，语音化，让这些服务可以更好的触达用户。每一次交互的革命，都会带来一个服务业的大升级，所以今天我也会简单介绍下我们的AliGenie开放平台，来帮助各行各业的开发者，进来做自己的对

话机器人。我们做自然语言理解的一个初衷也是让各行各业的人可以把自己的服务语音化。因此除了“天猫精灵”的终端产品以外，我们希望把技术赋能给第三方合作伙伴，包括：语音唤醒、语音识别、声纹识别、语义理解、语音合成等。

这里我放大讲下**自然语言语义理解的系统结构**。我们把自然语言理解拆解成两个任务，意图识别和槽填充。在语义理解的部分，意图和槽的定义来自开发者在AliGenie开放平台上的配置。在这里我们利用积累很长时间的知识图谱和用户的画像来帮助我们进行更好的用户语义的理解。这个理解的结果，会传给对话引擎，这里预先加载了开发者定义的对话策略，和第三方服务的调用策略。在这里，有的时候会进行主动发问，例如针对关灯的指令，会主动发问，“现在就关掉卧室的灯吗？”，也会调用第三方的服务，去生成对话的结果。最终变成一个自然语言的回复，答复给用户。

在整个语音交互系统里面，我认为最困难的一个部分是自然语言理解，因为语言是人创造的数据类型。机器在做一件人擅长的事情，人对机器的理解能力的期待是很高的。这也就为什么很多现在所谓的人工智能系统，很多用户会认为是人工智障。

自然语言理解因为人在用的时候有很大的多样性和模糊性，比如说我们要想问天气怎么样，很多用户可能问的是“明天是不是要洗车”、“明天是不是要穿秋裤”这样的一些非常千奇百怪的问法。但是我们为了给用户更好的体验，希望能够把它们精确地识别出来，所以具有很大的挑战性。另外除了刚刚提到的多样性，还有模糊性，比如说要想问苹果多少钱，不同的用户他问这个问题的时候，他的意图是不一样的，有的科技爱好者可能问的是电子设备，但也有人可能问的是水果。所以为了解决这样的一些模糊性，我们也把上下文的信息都拷贝进去，以便于做更好的决策和判断。

定义自然语言理解及挑战

那么在这里，我们定义的是一种去完成某个task的自然语言理解的场景，这也是今天课程的核心问题，也就是如何让机器去帮助我们完成某个任务。首先要去定义什么叫理解，理解是怎么表达的。在这里我们采用框架语义的表达方式，用技能、意图和槽值来结构化表示一句话的意思。这样我们在这个场景下，把自然语言理解抽象成一个机器学习的命题，也就是从一句自然语言，转变成这三种结构化的表达。

问题定义清楚了之后，我们就来看怎么解决这个问题。这里面有很多很有意思的挑战。例如，自然语言的句式和用词是很多样的，然而通常情况下训练语料又非常有限，用户的期望又很高。这都对模型和开放者平台提出了非常大的挑战。另外一个比较大的挑战是，大量的开发者还不是NLP和机器学习专家，如何让他们在不了解自然语言技术和机器学习技术的情况下，实现与NLP专家的协作，把一个技能的理解配置好，也是很有挑战的事情。

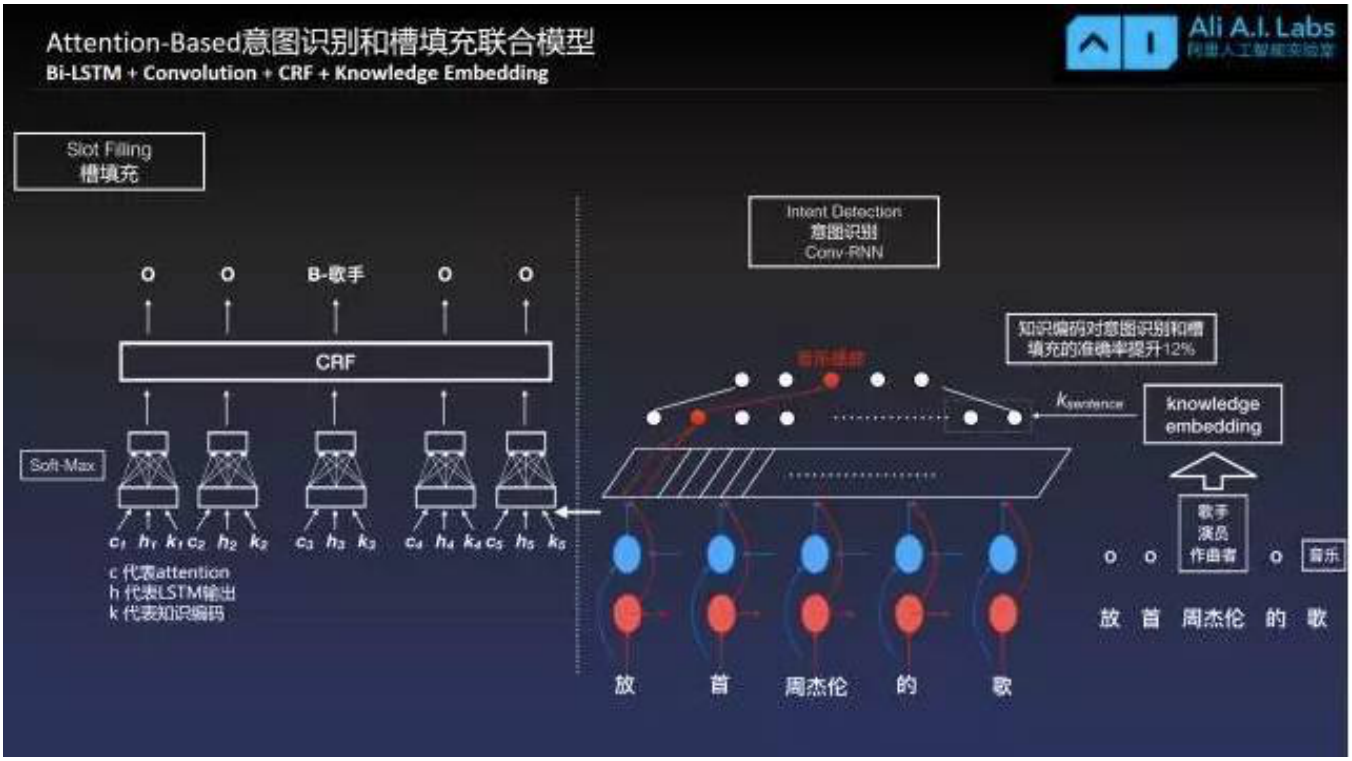
在这里我想做一个简单的总结，因为我们这套自然语言理解的技术的一个特点是大量地采用了深度学习方法。现在深度学习在各行各业都具有统治性的地位，因为它能够从数据里面去发现有哪些特征是最本质、最重要的信息。而深度学习技术在最近的2-3年的时间里，对自然语言理解有着非常大的推动作用。可能对于深度学习在NLP上的应用，大家之前更多是在

翻译等 seq2seq的问题上。所以今天我想从自然语言理解的角度，向大家介绍下，这部分比较重要的几个工作。

我们在AliGene系统里面使用了大量的深度学习技术，其中比较传统的是CNN， CNN是基于卷积神经网络，使用卷积方法这样的一个计算，去找本地最有效的那一些信息。另外就是LSTM，因为CNN可能更多是关注本地的东西，自然语言里处理句子一定是具有上下文关系，每个词都不是独立的，所以我们需要有这样的一个方法叫LSTM，它能够确保上下文的信息能够结合在一起。相比CNN在浅层特征抽取的时候仅关注n-gram信息不同，LSTM可以比较好地实现句子中的长短关系的提取，以及现在在这里使用的attention机制，让这种关联性更加直接。LSTM能更好的编码文本序列之间的关系。

我们一个比较简单的延伸是把CNN和LSTM进行了深度的整合，就训练一个端到端的系统，这样的系统不仅能够表达本地的信号，同时它也能够表达上下文的信息，把这两个信息通过一个优化包含在我们的处理之中。

另外我们也做了一些自适应神经网络，目前发展下的神经网络有一个缺点，它在训练好之后它能够针对不同的需要处理数据，但它用同样的一个结构去提取信息，这个模式并不是最优。因为每个输入的句子都有自己的特点，我们提取了自适应神经网络，重点就在于它能够自适应的根据信号特点去调整网络结构，从而找到最有用、最重要的信息。



Attention-Based意图识别和槽填充联合模型

这是目前我们在AliGene系统中使用的意图识别和槽填充的联合模型。对于意图识别的部分，我们采用了上面的convRNN的网络。然后这个BiLSTM的输出，也就是这个的h，我们用于序列标注的任务，也就是槽填充。

正如我前面提到的，这里用了基于attention机制的分类的方法进行序列的标注。因为序列标注的解空间非常大，所以在这之上添加了CRF层，利用viterbi解码的方式计算一个全局最大可能的序列标注结果。这里我们利用知识图谱的能力，对句子上可能的知识信息进行编码，例如周杰伦，在我们的知识库中既可能是歌手、也可能是演员、也可能是作曲者。这样的一个知识信息可以帮助我们更好的理解这句话的语义，同时和神经网络学习上的语义特征进行整体的优化求解。

例如这个例子，前面有“想放首”，“歌”这样的关键词，那么神经网络会认为，这里周杰伦是歌手的标注的可能性会大很多。同时知识编码在有的情况下很有用，例如，“来个黄焖鸡米饭”，和“来个七里香”，在句式上是一样的，但是因为里面知识内容的不同，我们可以理解成不同的意图。所以我们在实验中也发现，这部分对效果的提升非常明显。基于这套模型，我们的意图识别和槽填充的准确率都很高。我们同时也在公开数据集上测试了我们这套方法的有效性，在文本分类和问答匹配的数据集上都达到了目前业界最好的效果。

不能错过的精彩Q&A

Q：科大讯飞会不会沦为一家技术公司？

A：现在大家已经形成一个共识，AI的发展，人才和算法很重要，另外一个很重要的因素就是数据。如果没有数据，后劲和发展能够就会受到限制。天猫推出这个端，在数据方面是有优势的。基于大的销量，就会建立自己的生态系统，形成循环。刚才您提到科大讯飞，我们对于高新技术一直是有需求的，数据的量虽然可以带来提升，但是慢慢也会收敛，在后期不会像早期有那么强的数据增量，它的曲线大概是这样子的，所以我不觉得技术在后期会被弱化，只能说我们可能需要更多更具革命性的技术。我们也希望像科大讯飞这样的公司能够在探索技术前沿方面作出更大的创

新，引领我们。

Q：我有两个问题，一个是关于“天猫精灵”商业化的问题，你们实现盈利？另外一个，我们双十一去美国学习的时候那个教授也讲，他们大概有50%的智能音箱其实放在家里当音箱用，智能不智能其实没有那么大的关系，对于“天猫精灵”来讲，用户中有多大的比例是把它当成一个智能设备每天跟它有交互的，有多少人就把它当成一个音箱用？

A：第一个问题其实是超出我的范围，因为我是一个做算法、做研究的人，所以在商业化思考上，比如以后怎么赚钱，其实不算是我最关心的事情。但是我觉得假如一个产品，并且是一个人工智能产品，它在早期我觉得我们可能先不用想太多，先让它去自由的发展，不要给它加太多商业化的条条框框，在后面它可能会带动我们很多惊喜，比如阿里的云业务。我们先尽力把我们的产品做好，把技术做好，能够抵达到更多的用户，这是我们现在最大的KPI，最希望能够达到的。

您刚刚讲的第二个问题，首先我觉得就算是听音乐这个事情，实际上它也体现了我们的智能性，就刚刚我举的这个例子，我们听音乐的时候用手机APP，其实也是要花很长的时间去跟手机交互，但是用智能音箱可能只需要五秒，这个事情本身就是一个智能的表现，让人机交互相当于上了一个新的台阶，让人机交互更自然、更便利，当用户适应了这样的交互方式之后，它又可以把它延伸到别的方面，比如说在厨房里面做菜的时候我想要买一个调料，以前我先要把手洗干净，打开手机APP，现在只要喊一嗓子它就可以了，所以我觉得它是可以有延伸性的，当用户习惯了，我们培养了用户这样一种方式，用户适应了这样一种方式，他又会把它延伸到别的领域里面去，去进行更多的操作。

从我们的结果来看，听音乐是用户最大的一个需求，因为它毕竟是个音箱，但是用户也用了很多的服务，包括像定闹钟、听新闻、问日程的安排还有甚至包括订外卖，用户实际上更多是在探索我们提供的服务。还有很多情况，用户可能不知道我们在提供服务，所以我们也在不时地去推送给他们，让他们了解到我们这一周又有些新的功能上线了，用户还是觉得非常兴奋，感觉像一个未知的世界，他们每周都有新的惊喜。

Q：我们也在看这个东西，我们的看法是这样的：它实际从长远来讲，是一个智能管家，比如说今天我老婆要过生日，它会自动提醒你，“别忘

了，这事很大”，而且它会根据你老婆的习惯嗜好以及她已经有了什么，她缺什么，会自动推荐给你们，这样一来它整个和阿里的生态系统就完全衔接在一起了，所以这是我们理解的长远的想象。近期来讲，把它当一个音箱，刚开始都是一些大餐前的开胃小菜，那个只是说来验证我们说整个东西是不是有一个小闭环，那个和商业模式不一定有直接关系，但是说到第三个问题，说到技术，科大讯飞是不是会沦为技术公司，我非常同意王刚博士的说法，这里面有太多技术要解决。比如我们刚才说到CNN+LSTM非常好，但是一个大问题，刚才我们说的例子，说怎么让计算机记住我老婆的生日，提前多少天的时候要提示我说“老婆要过生日了，以前买过香水了，这次就不要买香水了，这次出了一个新款的什么口红你要不要买一个”，这种东西实际上是要和它的一些常识要挂在一起的，是不是CNN加上LSTM就能解决这个问题呢，我看是不行的。

A：对，不行的。

Q（接上一提问）：所以有太多的问题需要处理了，所以这个东西来讲的话，它会有太多的商业变现和场景的应用的关系，但是说商业变现倒是有一个事情我很好奇，就是一些简单的东西是可以买了，不限于听歌，太复杂的可能不能买，你说我买一个洗衣粉这种可以买了，这个时候马上就呈现出来一个问题，你和阿里旺旺是怎么衔接的，我很想了解这个方面的问题?

A：大量的语音交互这个新的模式，它能够让交互的效率在很多情况下能够提升，并且也能够让它对一些特定的人群，比如说老人它更加地友好，是这种交互方式的体验，但它的本质上提供的这些服务可能还是说是可以是一样的。刚才您讲的技术这块讲得特别好，未来在自然语义理解方面我们肯定是要考虑到更多世界知识，体验知识，所以我们要建立知识图谱，能够让语言理解特别在模糊性这块，因为很多人说话的时候，我们人和人交流的时候，他说话很多时候被省略掉很多词，因为我们假设对方是了解相应的知识的，所以我们也希望音箱能够有相似的能力，在自然语言理解这个环节就会提高它的效率。

Q：能介绍一下目前产品的用户体验情况吗？以及未来的改进方向?

A：我们对这个系统有很多衡量指标：一是达成率，因为我们现在定义是帮助用户完成任务的智能助手，比如我要听歌，我要播放相应的音乐。所

以我们就需要去看这个音箱返回的结果，它是不是帮助用户达成了想要的东西，这就是达成率。另外一个是性能，以及就它返回的时间有多久，这也是一个重要指标。当然还包括其他的一些指标，比如说听音乐，这个用户听音乐听了多久，他是不是中间切断了，这些我们也是在关注。从现在业务的情况来看，我们的达成率稳步上升。因为最开始第一个版本出来后，经过测试、评比，达成率还不是特别的理想。所以我们在后面经过了好几次技术的迭代，现在来看，我们的达成率还是不错的。不过我们还在优化我们的性能，这个产品的链条还是很长的，我们认为在速度的优化上还有空间，希望可以在1秒之内完成用户的需求。

Q：现在在人工智能领域分为两个，一个是图象的识别和相应的人工智能，一个是像语音、文本性的，你是怎么看待这两个不同的纬度以及它们之间的关系？未来它们最终在往前演化的时候，当逐渐实现了感知和理解的功能以后，它最终在智能端会进行汇集还是说会完全平行的两个方向的发展？如果能汇集的话，汇集的那个点是什么，它的核心的能力是什么？

A：我们马总（马云）也曾经说过，人的智能跟机器的智能是不一样的。我觉得从应用的场景来讲，图像人工智能与语音文本人工智能是可以并行的，比如说语音可以用在智能音箱，视觉可以用在自动驾驶、人脸识别、安防等方面。从应用的产品来看，我觉得它们更多是一个可以并行的关系。对于人来说，我们的大脑其实处理不同的信号也是有不同区域的，但是它总有一些方法让不同区域之间进行沟通，比如深度神经网络，它作为一种对知识要求不这么高的方法，它可以从数据里面自动的挖掘出哪些特征是重要的特征。所以这样的方法它同样可以在图象和语音上面都能够得到应用，因为这两者都是一个信号处理信号分析的问题，所以它们的方法是可以比较通用的，只是输入和输出不一样。

Q：根据我自己理解，我觉得以前鼠标操作电脑是一种操作方式，到现在手机屏幕用手指触摸是一种方式，我希望用语音来控制设备将来能作为一种全新的交互方式。我自己认为手指触摸从苹果到现在大概花了十年时间才比较成熟，对于这样的一个语音的交互方式，离真正能够大规模投入使用大概要多长时间？第二个问题是，现在都在讲智能家居的入口，如果家庭的这种管理方式真的变成现实了，可能大概需要多长时间才可能变成大规模投入使用的方式？

A：语音交互的大规模使用我觉得在美国已经发生了，比如说像Echo，它

每年大概是一千万的销售，背后可能就是几千万的家庭，所以在美国的覆盖面还是非常大的。在中国，首先从技术的成熟度来讲的话，我们希望语音交互能够更加可靠和稳定。我很乐观，我觉得在一年到一年半时间内，可以实现95%以上的满足能力。

Q：您怎么看待人工智能分别在硬件和软件上的发展？人工智能硬件比如说像“天猫精灵”，音箱是一种形态，那怎么样分别看待软件和硬件的发展方向？

A：我觉得它们两者后面是越来越融合的关系，端可能只是一种产品的形态，比如像我们的音箱，有这样一个硬件放在用户的家里面，但实际上人工智能都是在云上就进行处理，因为云上有比较多的计算的资源，所以我觉得它只是两种不同的表现形态。有的人工智能像美图秀秀里面做一些自动的化妆，它可能是以软件的形式显现出来，我们现在的“天猫精灵”的形态是以硬件的形式显现出来，但是它两者实际上用的智能能力我觉得是差不多的，没有太大的区别。

Q：我问一个技术问题一个非技术问题。技术问题，我看到一个应用的场景是让我比较兴奋的，您认为在技术层面对比其他平台阿里的优势在哪里？

A：这是一个挺有意思的问题，刚刚讲因为做技术的，可能也非常难去评价到底说哪一家的技术一定比另外一家好，并且技术也是不断地在发展迭代的，像我们的系统也是经常在更新，我相信其他的公司也是在不断地去更新技术，所以我是很难评价的。对于我们来讲，像知识图谱，像这样的东西能够跟我们的战略理解能够结合在一起，就是说补现在自然语言理解的短板，这是我们非常感兴趣的方向，别的公司我就不太清楚了，他们会不会往这方面做，包括我们做了之后跟它们的差别多大，这个我现在是没办法评论的，因为很多时候技术发展是一个一定要反复去试，试了之后到底能提高多少，预先没办法做判断。

Q（接上一提问）：第二个非技术的问题，如果我是产品公司的，我想做个音箱，我非常想跟阿里合作，但你们做的这个产品让我有几个担忧，这个产品很好卖，结果你自己做，你又做的很便宜，数据全给你了，那我怎么跟你合作？

A：我们现在为止没有跟音箱的公司合作，我们更多的是跟内容提供商，跟性能的开发商，包括其他的像IoT的设备商合作，所以我的建议可能是大家做我们需要的东西，以后不要再做音箱了，做内容，做性能，这是我们非常欢迎的。

Q：现在音箱在国内也有一家在做，喜马拉雅它们也有自己的内容，它的技术是猎豹给它提供的，我也是在使用，但现在包括你刚才说到"天猫精灵"里面的一些音乐版权是来自于腾讯，腾讯其实也在发展AI，也在做语音识别这一块，竞争会是怎样的？我想听听您的想法。

A：阿里音乐它有自己的版权，腾讯音乐它也有它的版权，所以我们做了一些版权的交换，就是双方都得益的，不是我们拿了腾讯的版权。你刚刚讲的这个问题，因为我们是处于非常早期智能音箱语音交互这个市场，可能大家更多想的是说一起把这个事情先做大了，至于最后的格局会怎么分配我觉得现在也是早了一点，还不能够做评论，反正我觉得大家一起先把用户，因为用户买了这个音箱他希望有更全的内容，我觉得我们应该先合作把这个事情先做起来，后面再说。

Q：你的技术肯定要不停的迭代，我想阿里有没有可能性比如说未来技术方面跟别人去合作，因为音箱是一个入口，因为这的确是一个场景足够广泛，在整个居室里面也好、在整个办公室，很多地方它都可以被用，但是有些会出现专门的居住场景，比如在厨房里面，厨房里面其实很多地方可以承载这种技术，它可以植入到电器里面去，比如说我是做电冰箱的，技术可以快速可以跟它去连接去购物，我不一定是听音乐，但是我可以植入菜谱，比如说我在卫生间里面我有一面镜子，其实镜子背后是有很多的东西可以连接，你可以通过人脸识别的方法告诉我身体需要什么，更多的健康性，因为这种东西抓取数据是很多，你们会愿意跟这些专门的垂直领域的企业去合作，输出你们的技术，然后更多的场景大家一起协同发展吗？

A：其实我们在云栖大会上发布了一个叫做AliGenie，跟"天猫精灵"这个端比，它就是一个开放的平台，它包括了刚才我提到的像SR、LP，包括后面整合的一百多个服务，我们是希望把这样的能力输出来跟第三方的厂家合作，像一些跟智能家居的厂商，我们是非常乐意的，其实我们也有很多在谈，另外我们也希望能够输出到不同的垂直行业里面，比如南方航空，它们就跟我们合作定制了一些针对它们贵宾室的音箱，总结起来我们的心态是非常开放的，合作这个主题是非常鲜明的，我们真的希望跟所有的能

够有关系的合作伙伴一起把这个事情做大，这个是毫无疑问的。

Q：我想请教一下，美国的那个场景跟中国的场景也有很多的不同，就是说他们的家都比较大，或者说主妇在家里花的时间比较长，相对于咱们中国可能小家庭比较多一些，都不太做饭，在家里时间少，您看一下未来这个市场发展，您这个产品跟美国的区别有没有完全的考虑进去，还是您预测中国的市场发展会跟美国对Echo的接受度是一样的成功，还是就是大家考虑了这个市场的差别呢?

A：我们在做的时候首先我们坚信很多用户的习惯是培养出来的，刚才我也说语音交互确实能够提高便利性，这种便利性很多用户现在是不知道的，交互这个东西其实跟我们的载体是有关系的，我们也可以把我们的能力输出到别的设备上，比如像家居产品、镜子，我觉得语音交互它带来这种革命性的或者说这种巨大的提升，不管是中美都是有需求的，可能只是说不同的市场我们可能需要不同的优化方法能够让它以更好的方式去被用户所接受，所以我觉得是这样的。

第二点，虽然大家说中美有差别，但是我们却在九个小时里卖了一百万台，我觉得这也是一个事实的证据，说明中国的家庭是有这种需求的，只是说以前大家可能比较悲观，反正我自己是比较乐观的，我们觉得这个市场是可以起来的，并且也能够像Echo在美国一样取得巨大的成功，我们是坚信这一点的。

Q：其实市场两派声音，一派认为智能音箱会成为下一个家庭入口，就是一个大步走的机会，还有一派认为这东西是伪命题，就觉得有没有智能音箱无所谓，你把它放在其他东西上面，甚至它就是一个你镶在墙上的一个接收器，或者说就在手机里面，它真正要连接的它就干三件事，一个就是控制你的智能家居，第二就是连接服务，比如我们叫滴滴叫外卖，你直接用语音叫就可以了，不用打开手机APP，还有一个就是你怎么样去能够提供好的内容，就是我在听音乐、听书或者说看电视，我通过它能够获取内容，真正谁能够整合这三样东西成为一个新的APP Store，那你才是真正下一个入口的掌握者，但是现在大家都搞音箱，其实我从Echo、像问问还有叮咚一堆我都买了，我买了这么长时间我真的用不起来，至少从我来说，我还是在用手机，不好用，可能对于老年人也许他会觉得好玩，你觉得未来它一定是音箱的形态，还是说它就是一个新的语音交互平台?

A：我可能更相信它会是一个比较通用的语音交互的平台，最后它应该是无所不在，然后又没有明显的存在感，你要它的时候它就来了，你不需要它的时候你看不见它，我觉得这可能是一个比较理想的情况。但是在早期的时候，我们也希望有这样的一个案例来证明它的可用性，音箱就作为一个载体被选中了，不管是在美国还是在中国，我们现在看它这个载体还是能够证明语音交互的价值。在未来不管是亚马逊Echo，还是我们的AliGenie，实际上我们的愿景都是希望能够让语音无处不在，不要局限于音箱这样的一种形式，而是以更灵活、更自由的方式和生活里面所有的联网的电子设备能够整合在一起，也是我们的一个信念。

Q：我大概在十多年前做过一些年的语音识别，我想请教一下，因为我们那个时候做发现在美国，因为它语音识别用的很早，最早汽车就是标配的说不能离开方向盘也不能去拿手机，所以它是在法律上强制在那个年代推语音识别，在中国其实很多汽车也装了蓝牙和语音识别的东西，但是为什么我们当时推不动？第一要教育消费者为什么要语音识别，第二是人和机器我想问的问题是，交互的情感上中国人好象比较保守和含蓄和害羞，他不想跟机器交互，你说为什么Echo在美国一年卖一千多万，在中国它的销量是怎么样的，或者说您看到的未来比如说2019年或者明年爆发，像我们现在手机上siri用了很多，华为的手机都有语音识别，但是有几个人真正在跟机器互动，明天早上问“几点了，天气怎么样”，就是中国人的这种文化习惯和使用习惯会不会造成对语音识别接受度和推广度的一个滞后。以前还有人跟我讲过，男人和女人对于语音识别使用都不一样的，以前像SoftBank做的机器人，很多类似的放在银行里面，有的是全身的有的是半身的，银行做了一个统计，说跟机器互动的大部分是女人，因为她们比较有耐心，男人进去就觉得很害羞或者觉得老爷们跟机器说话是不是挺傻的，你们有没有研究过这个东西，这实际上决定了我们产品最后能不能成功，您对这个事情有什么看法?

A：挺有意思的一个问题。我们当时没有研究过受众的性格这样的一些情况，我感觉以前大家用语音交互用的不多，可能主要的问题还是技术不够成熟，比如说在车里面，比如我们想用它，因为有噪声各种因素不准，用了两次之后用户有挫败感，就觉得还不如用手机，我的理解更多的是因为技术上的不成熟所以导致用户失去了耐心。我们如果看近几年来因为深入学习、大数据的使用，语音识别在语言理解它的效果是越来越好了，所以慢慢达到了能够让大家没有觉得很灰心丧气，现在我们慢慢到了这个临界

点，可能十句里面有八句九句它都能够回答出来了，那用户就会接受了，我觉得是一个技术的问题。

Q：关于多模态交互，阿里大概做了哪些东西，请您跟我们分享一下?

A：我们做了像表情，我们也做了手势，我们也做了这样一些技术的研发，但至于会不会有这样的一种交互方式出来、产品出来，这个还不太确定。

Q：在家庭里面现场环境非常吵，比如有小朋友，家人聊天的时候会不会出现误判，现在是怎么解决的?

A：这种情况是存在的，比如我们现在的产品里面在噪声比较大，比较吵的情况下会对我们造成困扰，就是会识别错误，后面可能我们会用更多的比如像背景噪声的建模，用声音的位置等方式，希望能够尽量把造成的因素压制下来，这是我们现在要做的事情，我们希望能够在明年解决这个问题。

Q：我特别想了解咱们在做这个产品最开始那个概念究竟从哪里来的，就我一定要做这个事情?

A：我们对交互这个事情还是研究了挺久，我们觉得不管是对什么样的群体，更便捷的交互方式它都是需要的，不管是国籍、男女、年龄，他们其实都是需要一个更便捷、更自然、更容易的交互方式，所以我们就坚信语音交互是一个值得做的事情。我们也相信有些观点，很多事情是我们做到了之后用户才发现“我需要它”，如果在没有实现之前，用户很多是抵触的心理，所以因为相信而干。

Q：我大概七八年前就去科大讯飞调研过，当时他们就在做语音识别，而且也做了很久，那个时候已经做了很久。在我印象中阿里做语音这块技术肯定是没有他们时间久，你们是理念上有什么不同，你们在这两块技术深度上的方向有什么不同?

A：科大讯飞是一个非常让人尊敬的高科技公司，我们在语音上面的时间可能没有他们久，但是我们阿里还是很有积累的，以前我们在云上也有很多的语音识别的需求，所以我们也是一直有团队在做这块，而且我们的团队也是不小的，整个阿里大概也是有很多人一直在执着的去做语音识别这

个事情，至于刚才讲到跟它的方向，因为现在深度学习是最有希望的方向，我觉得不管是科大讯飞还是我们，都是关注于深度学习这个大的框架，深度学习是一个非常大的框，里面有很多优化的点，可能有新的神经网络设计，一些新的优化的方法，我觉得可能各家都会根据自己的想法去做相应的优化，最后大家收敛到某一个点或者是会造成一个比较大的差异这个我不太清楚。

Q：我们现在能不能做得到辨别出每个人的声音作为唯一性的东西?

A：声纹识别在人数比较少，比如说六个人左右，我们是能够做到的，但是它毕竟没有人脸信息这么丰富，所以人一多的时候有可能会出问题，在家庭环境里面还是OK的，它遇到更大的规模的话就会有问题。

Q：我自己在用小鱼在家这个东西，我的体验会觉得他们也在不停的迭代，我来之前去创新工场那边，因为他们现在分拆了两个，一个是针对家庭这边，一个是针对企业级的，我想问"天猫精灵"这边有没有类似这种方向的分拆?

A：我们也跟行业有些垂直应用，比如一个是酒店，像跟万达酒店我们有很多合作，就是我们在每个酒店的房间里面放一个天猫精灵，然后用它来控制酒店里面的智能家居，比如叫服务员来送东西，我们跟南方航空提供一些特制性的特点，比如在他们VIP的休息室里面，所以我们也很重视跟不同的垂直产品有一些深度的融合，我们也是在做这个事情。

Q：像天猫精灵类似的产品，它的安全性怎么来保障？比如作为用户在家里怎么能够让用户相信，在我不想让它听到我说话的时候，它真的没有在听。

A：天猫精灵是有一个唤醒词的，就是"天猫精灵"，只有在叫了唤醒词之后，它才开始开麦，把这个声音进行处理，如果没有唤醒词的话，就不会去提取任何的声音，也不会做这个事情，因为如果你需要天猫精灵就表示你要跟它说话了。

Q（接上一提问）：这个怎么让用户相信？因为毕竟开麦的时候它是在监控我的声音?

A：这是一个很有意思的问题，我们只能够给用户解释，并且我们也有相

应的条款去跟他说明我们对数据的保护，我们内部对用户的数据保护是非常厉害的，基本上是用户说了“天猫精灵”，我们基本上很难能够去得到原声数据，基本上不可能，因为我们要做这些算法的，这些数据都是处理过的数据，不可能接触到原声数据的，所以我们数据保护是非常严格的。

Q：它的安全应该是从两方面，软件和硬件，这方面如果安全系数不够高一点，到了很多家庭里面如果被人黑掉，可以窃听很多家庭的隐私。

A：我们跟安全团队也是有合作的，不过这个事情可能跟其他所有的设备一样都存在这种风险，手机、汽车，都存在这个风险，我们能做的事情就是说魔高一尺道高一丈，能够有个更好的安全措施把这个问题解决掉。

高山大学思享课II期学员现场合影

高山大学（GASA）-

高山大学是一所以“科学复兴”为使命，以“没有受教，求知探索”为校训，致力于给创业者、企业家培养科学精神的新型大学。由长城会创始人文厨先生创办，由前金山软件CEO张宏江、创新工场创始人李开复、斯坦福大学物理系教授张首晟、清华大学教授鲁白、红杉资本中国创始合伙人沈南鹏、阿里巴巴集团技术委员会主席王坚、卡内基梅隆大学机器学习学院创始院长Tom Mitchell、加州大学伯克利分校教授杨培东等联合发起并作为

校董。

思享课-

思享课是高山大学基于“终身学习”的办学理念，“虚实结合”地探索科学。科学家务“虚”，谈理论和趋势；创业者落“实”，讲运用和实践。返回搜狐，查看更多

责任编辑：