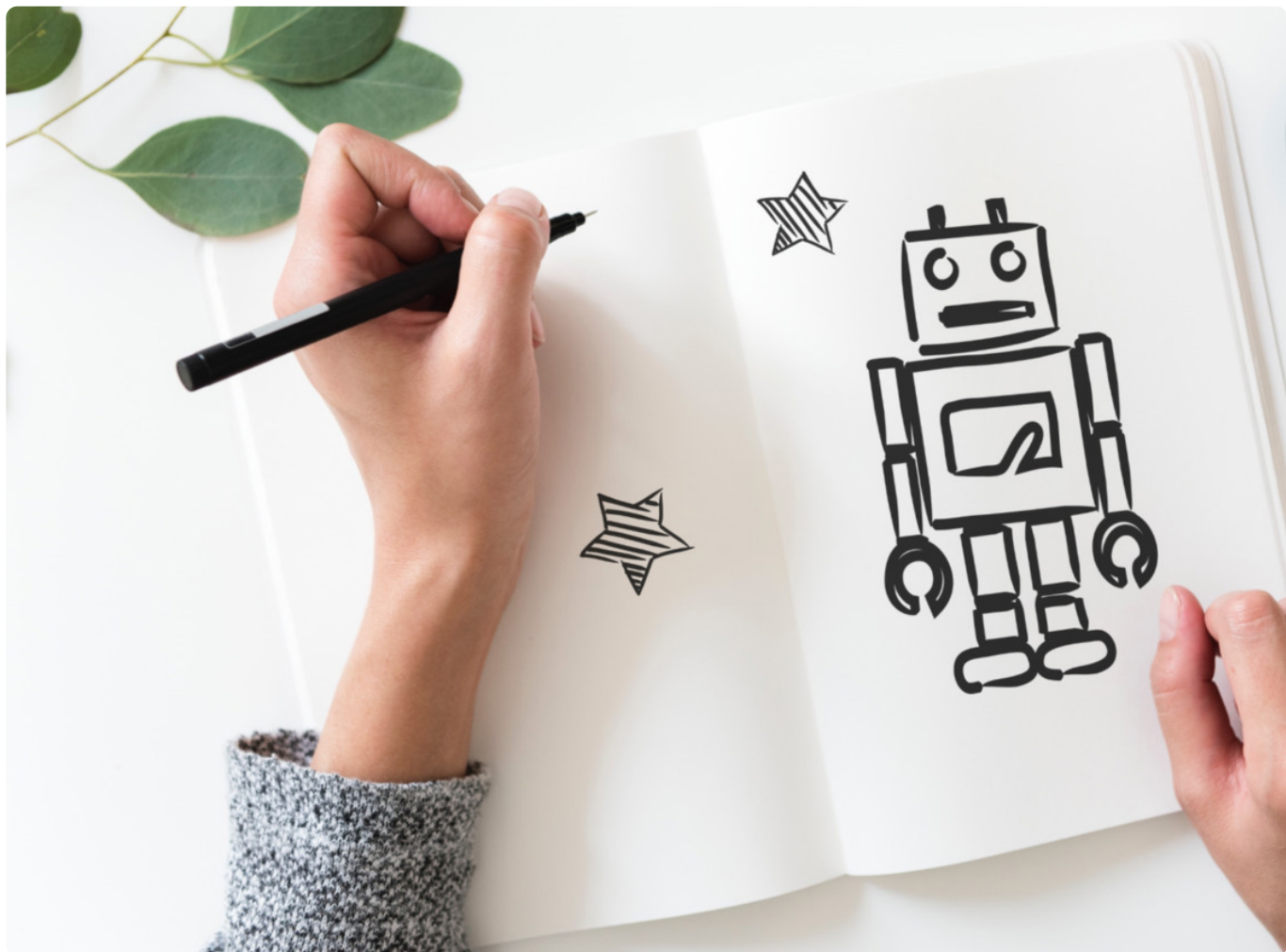


097 | 聊天机器人有哪些核心技术要点？

2018-05-16 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 05:35 大小 2.56M



对话系统分为“任务型”和“非任务型”两种基本类型。周一的分享里，我们讨论了任务型对话系统的一些技术要点，重点介绍了任务型对话系统的各个组件及其背后的模型支撑。

今天，我们就来看一看**非任务型对话系统**的主要技术要点。非任务型的对话系统有时候又会被称作是“**聊天机器人**”。

基于信息检索的对话系统

我们前面讲过，对话系统，特别是非任务型对话系统，也就是聊天机器人，有一个很重要的功能，就是在一个知识库的基础上和用户进行对话。这个知识库可以是海量的已经存在的人机对话，也可以是某种形式的知识信息。

比如，一个关于篮球的聊天机器人，那就需要这个系统能够访问有关篮球球队、运动员、比赛、新闻等有关篮球信息的知识库。同时，在这个对话系统运行了一段时间之后，我们就会慢慢积累很多有关篮球的对话。这些对话就成为了系统针对当前输入进行反应的基础。

针对当前的输入，利用之前已经有过的对话进行回馈，这就是基于信息检索技术的对话系统的核心假设。一种最基本的做法就是，找到和当前输入最相近的已有对话中的某一个语句，然后回复之前已经回复过的内容。

比如，当前的问话是“迈克尔·乔丹在职业生涯中一共得过多少分？”如果在过去的对话中，已经有人问过“迈克尔·乔丹为芝加哥公牛队一共得过多少分？”。那么，我们就可以根据这两句话在词组上的相似性，返回已经回答过的答案来应对当前的输入。

当然，上面这种对话系统可能会显得比较原始。但是，一旦我们把整个问题抽象成广义的搜索问题，其实就可以建立非常复杂的检索系统，来对我们究竟需要回复什么样的内容进行建模。

比如，我们可以把输入当作查询关键词，只不过从性质上来说，当前的输入语句往往要长于传统的查询关键词，但是在技术上依然可以使用各种搜索技术，例如通常的排序学习等方法都适用于这样的场景。

从理论上讲，基于检索的对话系统有很多先天的问题。比如，从根本上，搜索系统就是一个“无状态”（Stateless）的系统。特别是传统意义上的搜索系统，一般没有办法对上下文进行跟踪，其实从整个流程上讲，这并不是真正意义上的对话，当然也就谈不上是“智能”系统。

基于深度学习的对话系统

那么，如何能够让对话系统真正**对状态进行管理**，从而能够对上下文的信息进行回馈呢？

最近一段时间以来，**基于深度学习的对话系统逐渐成为了对话系统建模的主流**，就是因为这些模型都能够比较有效地对状态进行管理。

那么，在这么多的深度对话系统中，首当其冲的一个经典模型就是“**序列到序列**”（Sequence To Sequence）模型，简称**S2S**模型。S2S 模型认为，从本质上对话系统是某种程度上的“翻译”问题，也就是说，我们需要把回应输入的句子这个问题看作是某种语言的语句翻译成目标语言语句的一个过程。S2S 模型也广泛应用在机器翻译的场景中。

具体来说，S2S 把一串输入语句的字符，通过学习转换成为一个中间的状态。这其实就是一个“编码”（Encode）的过程。这个中间状态，可以结合之前字句的中间状态，从而实现对上下文进行跟踪的目的。这个部分，其实就成为很多具体模型各具特色的地方。总的来说，中间状态需要随着对话的演变而产生变化。然后，我们需要一个“解码”（Decode）的过程，把中间的状态转换成为最后输出的字句。

从整个流程上来说，S2S 其实非常像我们已经介绍过的深度序列模型，例如 RNN 和 LSTM。从实现上来说，很多 S2S 模型，其实都是直接利用 RNN 或者 LSTM 而得以实现的。因此，很多深度序列模型的技术都可以直接应用到对话系统中来。

另外，我们可以看到，相比于基于信息检索的系统来说，S2S 模型并没有一个“显式”的搜索过去信息的步骤，因此可以更加灵活地处理语言上的多样性，以及不是完全匹配的问题。因此，从实际的效果中来看，S2S 模型在对话系统中取得了不小的成功。

实际系统的一些问题

在实际的开发中，非任务型对话系统会有一系列的实际问题需要解决。

首先，因为是开放性的对话系统，其实并没有一个标准来衡量这些聊天机器人式的系统的好坏。究竟什么样的系统是一个好的聊天系统，依旧是一个非常具有争议的领域。

其次，人们在实际的应用中发现，基于深度学习的序列模型，虽然常常能够给出比较“人性化”的语句回答，但是很多回答都没有过多的“意义”，更像是已经出现过的语句的“深层次”的“翻译”。因此在最近的一些系统中，人们又开始尝试把信息检索系统和 S2S 模型结合起来使用。

最后，我们需要提出的就是，非任务型对话系统和任务型对话系统有时候常常需要**混合使用**。比如，在一个订票系统中，可能也需要掺杂一般性的对话。如何能够有效地进行两种系统的混合，肯定又是一种新的挑战。

总结

今天我为你介绍了非任务型对话系统的基本技术要点。

一起来回顾下要点：第一，我们讲了基于信息检索，也就是搜索技术的对话系统；第二，我们聊了聊如何利用 RNN 或者是序列模型对对话系统进行建模。

最后，给你留一个思考题，你觉得在什么样的条件下，非任务型聊天机器人可以展现出真正的“人工智能”呢？

欢迎你给我留言，和我一起讨论。



AI 技术内参

你的360度人工智能信息助理

洪亮劼
Etsy 数据科学主管
前雅虎研究院资深科学家



新版升级：点击「 请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 096 | 任务型对话系统有哪些技术要点？

下一篇 098 | 什么是文档情感分类？

精选留言 (3)

 写留言



夏洛克的救...

2018-05-17

👍 1

还有，输入的中间状态和反馈给输入方的中间状态指的是同一个？

展开 ▾



夏洛克的救...

2018-05-17

👍 1

这个中间状态，可以结合之前字句的中间状态，从而实现对上下文进行跟踪的目的。

这句话太笼统，不是很理解其过程。上下文的中间状态是如何关联起来的？

展开 ▾



散人

2018-05-16

👍 1

不谈技术层面，非任务聊天机器人的智能我认为主要表现在一下几个方面：

一是情感的表达，可以根据具体情景表现出不同的情绪，高级一点的话，有一定的共情能力。

二是具有个性，具有鲜明的特征，聊天过程不机械，套路少。

三是说人话，产生的回答自然无语病。...

展开 ▾