

134 | ACL 2018论文精读：问答系统场景下，如何提出好问题？

2018-08-10 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 07:49 大小 4.48M



今年 7 月 15 日~20 日，计算语言学协会年会 ACL 2018 (56th Annual Meeting of the Association for Computational Linguistics)，在澳大利亚的墨尔本举行，这是自然语言处理和计算语言学领域的顶级会议。

计算语言学协会 (ACL) 最早成立于 1962 年，每年都赞助举行各种学术交流和研讨大会。ACL 大会是 ACL 的旗舰会议，可以说这个会议是了解自然语言处理每年发展情况的重量级场所。

会议今年收到了 1018 篇长论文和 526 篇短论文的投稿。最终，大会接收了 256 篇长论文以及 125 篇短论文，综合录用率达到 24.7%。

今天，我们来看这次会议的一篇最佳论文，题目是《学习提出好问题：使用完美信息的神经期望价值对澄清问题进行排序》（[Learning to Ask Good Questions: Ranking Clarification Questions using Neural Expected Value of Perfect Information](#)）。

首先给你简单介绍下论文的作者。

第一作者萨德哈·饶（Sudha Rao）来自马里兰大学学院市分校（University of Maryland, College Park），是计算机系的博士生。她已经在 ACL、EMNLP、NAACL 等自然语言处理大会上发表了多篇论文，并且在微软研究院实习过。

第二作者是饶的导师哈尔·道姆三世（Hal Daume III），是马里兰大学学院市分校计算机系的一名教授，目前也在纽约的微软研究院工作。他是机器学习和自然语言处理领域的专家，在诸多领域都发表过不少研究成果，论文引用数达到 9 千多次。

论文的主要贡献

这篇论文关注的是“**问答系统**”（Question & Answering）。问答系统不仅在实用领域受到大量用户的青睐，产生了诸如 Quora、知乎、Stack Overflow 等知名的在线问答服务，也在人工智能系统开发领域受到研究者的关注。

我们曾经提到过“图灵测试”，用来衡量一个系统或者说是一个机器人是否具有真正的人工智能，这个测试其实就是建立在人机问答的交互场景下的。因此，建立有效的问答系统一直是人工智能研究，特别是自然语言处理研究的核心课题之一。

这篇论文的作者们认为，在问答系统的场景中，一个非常重要的手段是针对已经提出的问题进行“**澄清式**”（Clarification）提问，从而能够引导其他回答者更加有效地进行回答。也就是说，作者们研究的主题是，**如何找到这些具有桥梁作用的“澄清式问题”**，这是这篇论文的第一个重要贡献。

论文的第二个主要贡献是利用了“决策论”（Decision Theoretic）框架下的 **EVPI**（Expected Value of Perfect Information，完美信息的期望价值），来衡量一个澄清式问题会对原始的问题增加多少有用的信息。简而言之，**EVPI 就是这篇论文提出来的一个衡量有用信息的测度**（Measure）。

论文的第三个贡献是通过 Stack Exchange 平台（Stack Overflow 是其一个子站点），构造了一个 7 万 7 千条含有澄清式问题的数据集。作者们从这个数据集中选取了 500 个样本

进行了实验，并且发现，提出的模型要明显好于一些之前在问题系统中的类似算法。

论文的核心方法

既然这篇论文的一个核心贡献是提出了“澄清式提问”这么一个新的概念，用于帮助问答系统的开发。那么，**究竟什么是“澄清式提问”呢？**

实际上在这篇文章里，作者们并没有对“澄清式提问”给出一个清晰的定义，而是仅仅提供了一个实例来解释什么是“澄清式提问”。

例如，一个用户在 Ask Ubuntu 这个子论坛里，询问在安装 APE 程序包时遇到的问题。这个时候，如果我们需要问“澄清式问题”，究竟什么样的问题可以激发其他人或者提出澄清式问题的人来进一步解答原始的问题呢？

我们看下面几个从不同角度提出的问题：可以问这个用户使用的 Ubuntu 系统具体的版本号；也可以问用户的 WiFi 网卡信息，还可以问用户是不是在 X86 体系下运行 Ubuntu。

那么，在这一个场景下，后两个问题要么无法为原始的问题提供更多有价值的信息，要么就是彻底的不相关，而第一个问题关于具体的版本号，很明显是用户可以提供的，并且可以帮助回答问题的人来缩小问题的范围。

这也带出了这篇论文的第二个贡献点，**如何衡量一个帖子的价值呢？**

要回答这个问题，我们需要知道这里有两种帖子是模型需要处理的。第一种帖子集合是候选的澄清式问题集合。第二种帖子集合是候选的最终回答集合。我们最终的目的是得到最佳的最终回答。这里面起到“搭桥”作用的就是澄清式问题。

所以，作者们就构造了一个针对每一个最终问题的 EVPI 值，用于衡量这个问题的“期望价值”。为什么是期望价值呢？因为这里面有一个不确定的因素，那就是根据不同的澄清式问题，可能会产生不同的回答。因此，作者们在这里使用了概率化的表达。

也就是说，EVPI 的核心其实就是计算给定当前的原始问题以及某一个澄清式回答的情况下，某一个最终回答的概率，乘以这个回答所带来的“收益”。当我们针对候选最终回答集合中所有的回答都进行了计算以后，然后求平均，就得到了我们针对某一个澄清式回答的 EVPI。换句话说，**某一个澄清式回答的 EVPI 就是其所能产生的所有可能的最终回答的加权平均收益。**

从上面这个定义中，我们有两点不确定。第一，我们并不知道给定当前的原始问题以及某一个澄清式回答的情况下，某一个最终回答的条件概率；第二，我们并不知道问题的收益。因此，作者们利用了两个**神经网络模型**来对这两个未知量进行**联合学习**（Joint Learning）。这可以算是本文在建模上的一个创新之处。

具体来说，首先，作者们利用**LSTM**来针对原始问题、候选澄清问题、以及最后解答产生相应的**表达向量**。然后，原始问题和某一个候选澄清问题的表达向量，通过一个神经网络产生一个**综合的表达**。最后，作者们定义了一个**目标函数**来针对这些初始的表达向量进行优化。

这个目标是需要我们学习到的答案的表达靠近初始得到的答案的表达，同时，也要靠近最终答案的表达，如果这个最终答案所对应的问题也靠近原来的问题。换句话说，**如果两个问题的表达相近，答案的表达也需要相近**。

那什么样的问题是相近的问题呢？作者们利用了 Lucene 这个信息检索工具，根据一个原始的问题寻找相近的问题。这里，作者们并没有真实的标签信息，所以利用了一些方法来标注数据，从而能够让模型知道两个问题是否相关。

论文的实验结果

作者们利用了 Stack Exchange 来构建一个分析澄清式问题的数据集。具体的思路是，如果原始问题曾经被作者修改过，那么后面的某一个帖子中所提出的问题就会被当作是澄清式问题，而原始问题就被当作是因为澄清式问题而得以改进的帖子。很明显，这是一个非常粗略的数据收集条件。当原始问题被作者修改过以后，并且最后因为这个修改得到回复，就被认为是一个最终的答案。经过这么一番构建，作者们整理了 7 万 7 千多条帖子。

作者们利用论文提出的方式和其他的经典模型进行比较。最后的结论是，提出的模型能够更好地找到最佳的澄清式问题，效果要好于仅仅是简单利用神经网络，来匹配原始问题和相应的澄清式问题。

小结

今天我为你讲了 ACL 2018 的一篇最佳论文。

一起来回顾下要点：第一，这篇论文提出了“澄清式提问”这个概念，来帮助问答系统的开发；第二，文章提出了一系列方法，对澄清式问题进行描述和衡量；第三，文章构建了一个数据集，通过实验论证了所提出方法的有效性。

最后，给你留一个思考题，通过这篇文章关于澄清式问题的介绍，你能否给澄清式问题下一个定义呢？

欢迎你给我留言，和我一起讨论。

 极客时间

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 133 | ICML 2018论文精读：优化目标函数的时候，有可能放大了“不公平”？

下一篇 135 | ACL 2018论文精读：什么是对话中的前提触发？如何检测？

精选留言 (1)

 写留言



叶晓锋

2018-08-23



澄清式问题的目的是要获取更多的信息，帮助回答原始问题。澄清式问题的优劣在于问题所能获取的信息是否有助于为原始问题提供更多信息。具体说，作者认为某问题导致了原始问题的更新，这个就是原始问题，继续，如果更新原始问题导致了后续的回答，原始问题就被认为得到最终回答。

