

漫威电影中的智能问答

Intellectual dialogue in Marvel Cinematic Universe

王元卓 中国科学院计算技术研究所
仇韞琦 百度在线网络技术(北京)有限公司
陆 源 竞技世界(北京)网络技术有限公司

在漫威系列电影中,钢铁侠的人工智能助手贾维斯扮演了很重要的角色。比如在电影《钢铁侠2》中,当钢铁侠说“贾维斯,把我的战甲发射给我”时(如图1所示),贾维斯就会将Mark42战甲发射到钢铁侠所在的具体位置。



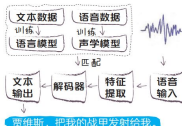
图1 《钢铁侠2》电影中的片段

目前市面上有很多类似的人工智能助手,包括各大品牌的智能音箱以及手机上自带的语音助手。如果它们是钢铁侠的助手,可能会问得更具体,比如:先生,发射哪件战甲?钢铁侠说:战力最强的战甲。助手回复:先生,没有找到“战力最强的战甲”。之所以贾维斯能够与钢铁侠实现智能对话,理解其意图,是因为贾维斯这样的助手采用了智能问答技术。如果让我们来实现一个类似贾维斯的人工智能助手,那该怎么做呢?可能的实现过程如图2所示。

07 《钢铁侠》中的科学

① 语音识别

智能问答的第一步就是将问题的语音转换成系统可以识别的文本，以便进行后续的自然语言处理。



贾维斯，把我的战甲发射给我。

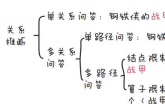
② 问题理解

· 可以把“把我的战甲发射给我”这个指令转换成另一个问题去理解，也就是“钢铁侠的常用战甲中战力最强的是哪个”。

· 通常智能问答中，系统首先要判断问题的主题，找到其中的主题实体。

· 然后通过问题的关系类型判断未进行问题理解，典型的关系类型包括以下几种。

主题实体推断：钢铁侠



11

电影中智能的问答

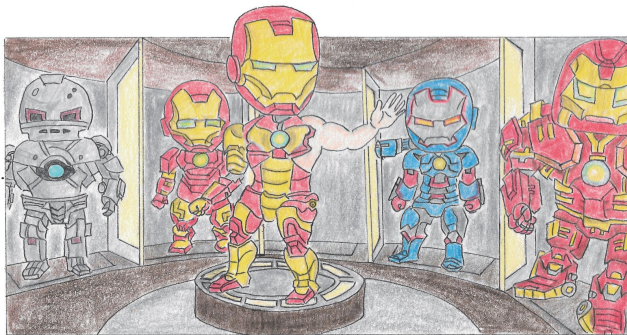
钢铁侠：贾维斯，把我的战甲发射给我。
贾维斯：好的，先生……
贾维斯：Mark42战甲已发射，预计15秒后到达。

科幻

现实中可能的问答

钢铁侠：贾维斯，把我的战甲发射给我。
贾维斯：先生，发射哪件战甲？
钢铁侠：战力最强的战甲……
贾维斯：先生，没有找到“战力最强的战甲”。
钢铁侠：快点，我快坚持不住了……

科学



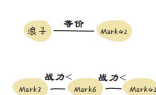
③ 答案搜索



· 确定问题的主题和类型之后，就是在知识图谱中搜索可能的答案。

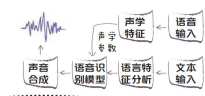
· 如果直接搜索到了确定的答案，就直接将相关的实体、关系、属性等知识输出去生成答案。

④ 答案推理



⑤ 语音合成

智能问答最后一步就是进行语音的回复，也就是将生成的文本信息转换成语音来实现类似于人与人之间的沟通方式。



⑥ 答案生成

根据对问题的理解和答案的搜索与推理，得到如下信息集合。

(Mark42战甲，发射时间，发射坐标，目标坐标)

飞行时间 = $\frac{\text{距离}(\text{发射坐标}, \text{目标坐标})}{\text{Mark42战甲飞行速度}}$

到达时间 = 飞行时间 - (当前时间 - 发射时间)

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

“Mark42战甲已发射，预计15秒后到达。”

生成自然语言回答语句

图2 智能问答讲解图

(选自《科幻电影中的科学：科学家奶爸的AI手绘》)

第一步是实现语音识别功能，我们要将问题的语音转换为系统可以识别的文本，以便进行后续的自然语言处理。通常，我们需要构建一个声学模型和一个语言模型，声学模型可以理解为对声音的建模。我们知道声音实际上是一种波。要想对声音进行分析，首先需要对声音进行分帧处理，也就是把声音切成多个小段，每小段称为一帧。分帧之后，通过特征提取将每一小段语音的声波转变为向量，这个向量包含了这帧语音的内容信息。我们把这些向量输入声学模型，它能通过解码器将向量输入转换成声学表示的输出，更准确地说是给出语音属于某个声学符号的概率。在英文中，这个声学符号可以是音节或者更小的颗粒度的音素；在中文中，这个声学符号可以是声母/韵母或者颗粒度同英文一样小的音素。语言模型的作用可以简单理解为消解多音字的问题，在声学模型给出发音序列之后，从候选的文字序列中找出概率最大的字符串序列。

得到了文本信息之后，第二步是问题理解，我们要让系统理解自然语言问题背后的语义。比如钢铁侠的指令“把我的战甲发射给我”就可以转变为另一个问题来理解，也就是“钢铁侠的常用战甲中战力最强的是哪个”。通常在智能问答中，系统首先要判断问题的主题，找到其中的主题实体，比如“钢铁侠的常用战甲中战力最强的是哪个”这个问题里的主题实体就是“钢铁侠”。得到主题实体之后，系统通过问题的关系推断进行问题理解，典型的关系推断包括单关系和多关系。单关系就是问题里涉及知识图谱中的单个关系，比如钢铁侠的战甲就是单关系；多关系是问题涉及了知识图谱中的多个关系，具体又分为单路径和多路径。单路径是指问题的主题实体到答案之间是一条不带分叉的关系路径，比如钢铁侠的战甲的战力；而多路径是指主题实体到答案之间的关系路径上有各种分叉作为约束条件，这些约束主要是结点约束（比如钢铁侠能发射小型飞弹的战甲是哪个）和算子约束（比如钢铁侠的战甲里战力最强的是哪个）。

在得到问题的主题实体和关系类型之后，我们可以形成一个结构化的查询，在知识图谱中搜索可

能的答案。如果直接搜索到了确定的答案,就直接将相关的实体、关系、属性等知识输出,生成答案。但很多时候,答案的内容是无法从知识图谱中直接获取的,需要进行各种推理,这个推理的过程就需要运用各种算法对知识图谱中知识的等价、比较、逻辑等进行运算,从而推理出可能的答案,比如在钢铁侠的战甲里,浪子是等价于Mark42的。而随着战甲的更新,其战力也在不断增强,因此Mark42的战力要比Mark6和Mark3高。

根据对问题的理解和答案的搜索与推理,系统得到了问题的答案,钢铁侠想要的战甲是Mark42。同时系统通过卫星定位到钢铁侠所在的位置,也就是战甲发射的目标坐标,根据Mark42战甲所在位置的发射坐标和战甲的飞行速度,可以计算出战甲到达钢铁侠那里的具体时间,最后生成相应的自然语言回答问句“Mark42战甲已发射,预计15秒后到达。”

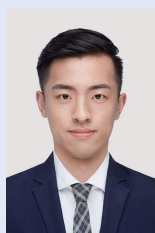
但生成的这个问句依旧是文本形式,智能问答的最后一步就是将这个文本转化为语音。这个过程是第一步语音识别的逆过程:首先抽取出文本输入的语言特征,然后将其输入语音识别模型,再结合已经设置好的声学参数合成声音,从而实现类似于人与人之间的沟通方式。

日常生活中我们经常接触到的在线客服问答和智能语音助手等,正是结合了智能问答技术来模拟的在线客服人员,其理解用户的提问,以智能客服的形式发挥客服作用,甚至比人工客服回答得更加快速、准确。

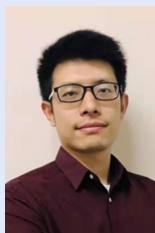
作者简介



王元卓 (1978-), 男, 博士, 中国科学院计算技术研究所研究员、博士生导师, 中科大数据研究院院长, 中国科普作家协会副理事长, 中国计算机学会科学普及工作委员会主任, 主要研究方向为大数据与人工智能。



仇毓琦 (1994-), 男, 博士, 百度在线网络技术(北京)有限公司高级工程师, 主要研究方向为自然语言处理、信息检索、智能问答。在WWW、WSDM、CIKM、COLING等国际学术会议上发表过多篇论文。



陆源 (1990-), 男, 现就职于竞技世界(北京)网络技术有限公司, 从事数据产品工作, 主要研究方向为大数据与社交网络。热心科普创作, 科普畅销书《科幻电影中的科学: 科学家奶爸的宇宙手绘》《科幻电影中的科学: 科学家奶爸的AI手绘》作者之一。