



专栏：人工智能技术与应用

人工智能在客服系统的应用及关键技术

王铮, 任华, 路绪海

(中国电信股份有限公司上海研究院, 上海 200122)

摘 要: 客服体系架构中, 人工坐席、工单处理等环节需要大量的人工操作, 成本较高, 效率相对较低, 而且会有一定比例的问题及故障无法找到责任方甚至无法解决。在客服体系中, 引入人工智能技术后, 在一定程度上可以解决这些问题, 降低成本并提高客服的整体效率和效果。分析了人工智能应用于客服体系的关键技术, 包括语音识别、语音合成、自然语言处理等。就新技术在运营商客服体系中的应用进行了探讨, 并通过简单实例阐述了在实际应用时的效果。

关键词: 客服系统; 人工智能; 关键技术

中图分类号: TP391

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2018302

Key technologies of AI in customer service system

WANG Zheng, REN Hua, LU Xuhai

Shanghai Research Institute of China Telecom Co., Ltd., Shanghai 200122, China

Abstract: In the customer service system structure, manual seats, work order processing and other links require a lot of manual operations and participation, the cost is high, the efficiency is relatively low, and there will still be a certain percentage of problems and failures that can't be found and resolved. In the customer service system, the introduction of artificial intelligence technology can solve these problems to a certain extent, reduce costs and improve the overall efficiency and effectiveness of customer service. The key technologies which could be applied to the customer service system in AI were analyzed, including speech recognition, speech synthesis, and natural language processing. The application of new technology in the customer service system of the operator was discussed. The effect in practical application was illustrated by an example.

Key words: customer service system, AI, key technology

1 引言

在市场白热化竞争的当下, 企业之间不仅靠产品争夺客户, 往往服务更被用户看重, 客服成为企业生态链中不可或缺的一环, 且重要性不断

攀升。

伴随着技术的推进和提高, 近年来客服系统本身经历了至少 3 次变革, 有了 4 种形式及效率效果的提升。第一次是随着程控交换机的大量上线, 话路不再是瓶颈, 电话客服也随之被推广,

收稿日期: 2018-10-05; 修回日期: 2018-12-08

用户可以直接以语音的方式向企业咨询或反馈问题,最明显的就是 400 电话的大幅增多。第二次是在互联网特别是移动互联网方式的应用和用户大幅度增长后,以“在线客服+电话”的方式开展服务的客服也应运而生,这种模式打破了单一的电话形式,应用了互联网及移动互联网的使用习惯和应用模式,为用户提供了在线客服的新型交互方式。第三次是基于云计算的云客服阶段,通过云计算系统的技术特点和优势提供整合后的整体性客服,为用户提供更全面、更准确的客户服务,当然这是后台技术的升级,客服人员大部分都知道,感受更好。而这两年随着数据质量的不断提高、数据计算能力的不断提升,人工智能技术也突飞猛进,逐步在多个行业和多个领域中形成规模应用。基于语音识别及语义理解人工智能应用,在当前尤其火热。Siri 自诞生时起,就是用户们的“调戏对象”;科大讯飞、百度的语音识别已经从标准普通话往各种方言细分;小冰、小翼、小 I 等机器人如雨后春笋冒出;基于 AI 的客服系统也随之而来。这就是第四代的新型客服系统,系统运用人工智能技术在后台了解用户诉求、分析问题因果、给出答复或提出建议,并通过模拟人声或者文字与用户交流。当然这时与用户沟通的不再是活生生的人,而是一个个“机器人”了。

从企业而言,这是客服体系的质的飞越,以前恼人的周期性业务培训、人员流动及在岗缺位都不再存在了,成本下降了,效率却有了大幅度的提高。

本文主要介绍在客服体系中引入人工智能技术后的架构和关键技术,并重点描述如何在电信客服系统中使用符合电信要求的人工智能技术。

2 基于人工智能的客服系统通用架构

基于人工智能的客服系统,是在原有客服系统中搭建人工智能处理模块来替代原有的人工处理功能。或者说将人工智能处理模块引入客服系统整体架构中。基于人工智能的客服系统架构示意图如图 1 所示。

在传统的客服系统架构中引入了人工智能技术。在客服整体架构中用相应的技术模块,替代了传统架构中原有的功能,且套入了传统流程。

在新的系统架构中,使用语音识别和语音合成模块代替了原来人工与用户交流沟通的模式。使用自然语言处理相应的话单或工单,替代了之前由业务处理人员承担的人工操作;使用知识管理模块取代原有的知识库构建方式。用人工智能技术模块替代了原来依靠人工处理的、具有一定重复性和规律性的工作。

这样的目的主要是在不影响客服质量的前提

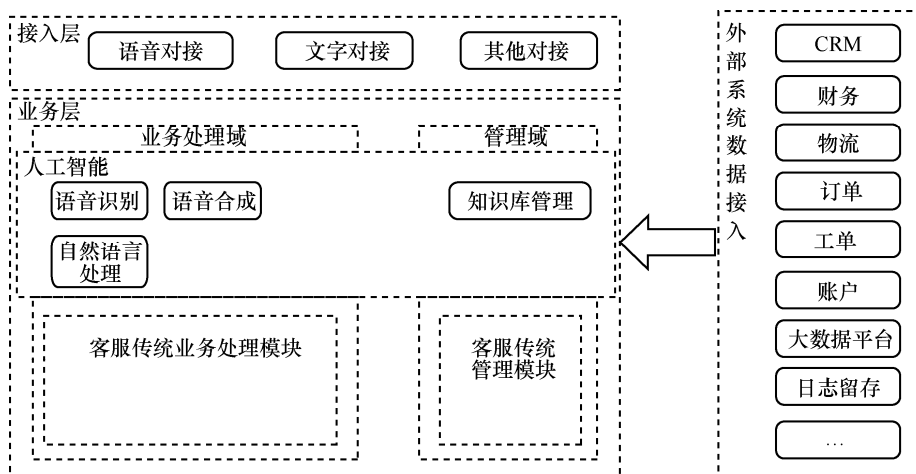


图 1 基于人工智能的客服系统架构示意图



下，降低原来的人工成本，提升效率和效果。

3 基于人工智能的客服系统的关键技术

基于人工智能的客服系统能够从技术上利用人工智能技术，主要体现在如下几大部分，即语音识别、自然语言处理、后台知识库管理以及语音合成。另外在整个端到端流程中也将在线用户情绪分析作为一个关键辅助技术。

3.1 语音识别技术

语音识别(automatic speech recognition, ASR)技术，是让机器通过识别和理解过程使之听懂人类语言的技术，语音识别技术是信息技术中人机交互的关键技术。随着语音识别在语音搜索、语音控制等领域的深入应用，语音识别技术被视为人工智能领域的关键技术。

3.1.1 前端语音处理

前端语音处理指利用信号处理的方法对说话人的语音进行检测、降噪等预处理，以便得到最适合识别引擎处理的语音。主要功能如下。

(1) 端点检测

端点检测是对输入的音频流进行分析，确定用户说话的起始和终止的处理过程。一旦检测到用户开始说话，语音开始流向识别引擎，直到检测到用户说话结束。这种方式使识别引擎在用户说话时即开始进行识别处理。

(2) 智能打断

智能打断功能是双向的，既可以打断播放的语音服务，也可以打断用户说话。该功能使人机交互更加高效、快捷、自然，有助于增强客户体验。

打断播放时指用户在自助语音服务的播放过程中可能会随时说话，而系统不必等待播放结束，能够自动对语义进行判断，并立即停止自动语音的播放，对用户的语音指示做出响应。

打断用户说话，则是在用户表达或者声音不清晰的时候，使用一定话术打断或告知用户。当然这个功能使用会非常谨慎，使用不当会造成用

户的反感甚至投诉。

3.1.2 后端识别处理

后端识别处理对说话人语音进行识别，得到最适合的结果，主要特性如下。

(1) 顽健识别

满足大词汇量、与说话人无关的识别要求，可以支持数万条语法规模的词汇量；并能适应不同年龄、不同地域、不同人群、不同信道、不同终端和不同噪声环境的应用环境。

(2) 置信度输出

置信度反映了识别结果的可信程度。语音识别引擎可以在返回识别结果时携带该识别结果的置信度，应用程序可以通过置信度的值进行分析和后续处理。

(3) 多识别结果

又称多候选技术，在某些识别过程中，识别引擎可以通过置信度判决的结果向应用程序返回满足条件的多个识别结果，而不是唯一的结果。识别系统提供了可能的识别结果列表，并按置信度结果从高到低进行排列。在业务设计中，可以根据应用要求向用户提供这些结果，供用户进行二次选择。通过置信度判决和多识别结果输出技术开发更加灵活、更加人性化的业务流程。

3.2 自然语言处理

自然语言处理(natural language processing, NLP)是使用自然语言与计算机进行通信的技术^[1]，是实现计算机与人间用自然语言进行通信的技术和方法。自然语言处理主要包括自然语言理解(NLU)和自然语言生成(NLG)两部分^[2]。

NLU会深刻理解每个单词对于中文就是字和词，试图理解是一个名词还是动词以及时态等特征。NLP系统有一个词库和一套语法规则，然后使用机器学习和深度学习，学习将这些规则应用于自然语言，并最终理解该段语言的意义。

NLG就是进行翻译转换工作，即将计算机的文本语言转换成人可以理解的文本，相对NLU而言简单的多。后继还可以使用语音合成技术将这

些文本转化成语音。

NLP 的基本操作如下。

(1) 分词^[3]

通常处理的自由文本分为中文、英文等。词为文本最基本的单位，分词是进行自然语言处理中最基本的步骤。分词算法分为词典方法和统计方法。中文的分词工具有很多，近年来常用的是 Jieba 和 stanford corenlp 等。

(2) 词性标注

在进行词性标注时，需先定义词性的类别：动词、名词、标点符号等。词性标注是语义识别、句法分析、信息抽取技术的基础技术。

(3) 句法分析

句法分析的目的是确定句子的句法结构，如主谓宾、动宾、定中、动补等。在客服系统中进行智能问答时有重要作用。

(4) 命名实体识别

命名实体识别是定位句子中出现的人名、地名、机构名、专有名词等。命名实体属于标注问题，因此可以采用 HMM/CRF 等进行模型的训练。

(5) 实体关系抽取

实体关系抽取是自动识别非结构化文档中两个实体之间的关联关系，属于信息抽取领域的基础知识之一。实体关系抽取包括规则匹配、有监督学习、无监督学习。

3.3 知识库管理

知识库是客服系统不可或缺的部分，不管是老版本客服还是人工智能客服，在管理流程中都需要在知识库中找到合适的标准知识和答案，再反馈给用户。

知识库的类型有很多，当前在用的相对普遍的是框架型、规则型、逻辑型 3 类知识库。

(1) 框架型知识库

框架型知识库主要由框架结构、框架行为、框架推理 3 部分组成。框架结构主要说明框架的属性；框架的行为主要是框架属性的方法（如平

移、旋转）及取值；根据框架与框架间的关系，框架型知识库能进行自动推理。

(2) 规则型知识库

规则型知识库主要包括规则库以及按照规则推理的方法，另外还包括规则优化及优选的逻辑和流程。

(3) 逻辑型知识库

逻辑型知识库的设计理念就是通过对问题的逻辑关系描述来告知用户寻求解释或解决的问题是什么，由知识库通过一系列的逻辑推理来找到解决问题的答案。

以上 3 种形式的知识库各有优缺点：框架型知识库描述应用领域对象相对自然，但是推理受限于框架结构本身；逻辑型知识库在逻辑推理方法和结果方面较好，但系统效率却较低；规则型知识库的可扩充性和可视性相当好，但对对象的描述能力却又较弱。所以，当前一般采用混合型的知识库构建方式，以满足不同的需求。

谷歌公司在 2012 年 5 月 17 日提出了知识图谱的概念，并以此为基础构建下一代智能化搜索引擎。本质上，知识图谱是一种揭示实体之间关系的语义网络，可以对现实世界的事物及其相互关系进行形式化地描述。当前知识图谱已被用来构建各种大规模的知识库^[4]。

知识图谱通常使用三元组的形式来表示，即 $G=(E,R,S)$ ，其中 $E=\{e_1,e_2,e_3,\dots,e_n\}$ 是知识库中的实体集合，共包含 $|E|$ 种不同的实体； $R=\{r_1,r_2,\dots,r_n\}$ 是知识库中的关系集合，共包含 $|R|$ 种不同关系； $S\subseteq E\times R\times E$ 代表知识库中的三元组集合。三元组的基本形式主要包括实体 1、关系、实体 2 和概念、属性、属性值等，实体是知识图谱中的最基本元素，不同的实体间存在不同的关系。概念主要指集合、类别、对象类型、事物的种类，例如人物、地理等；属性主要指对象可能具有的属性、特征、特性、特点以及参数；属性值主要指对象指定属性的值。每个实体（概念的外延）可用



一个全局唯一确定的 ID 来标识，每个属性—属性值对可用来刻画实体的内在特性，而关系可用来连接两个实体，刻画它们之间的关联。

知识图谱的架构主要包括自身的逻辑结构以及体系架构，目前，大多数知识图谱都采用自底向上的方式进行构建。知识图谱的构建包括 3 个步骤。

(1) 信息抽取

从各种类型的数据源中提取出实体（概念）、属性以及实体间的相互关系，在此基础上形成本体化的知识表达。

(2) 知识融合

在获得新知识之后，需要对其进行整合，以消除矛盾和歧义，比如某些实体可能有多种表达，某个特定称谓也许对应于多个不同的实体等。

(3) 知识加工

对于经过融合的新知识，需要经过质量评估之后（部分需要人工参与甄别），才能将合格的部分加入知识库中，以确保知识库的质量。新增数据之后，可以进行知识推理、拓展现有知识、得到新知识。

3.4 语音合成技术

语音合成（text to speech, TTS），又称为文本转化，其实质就是将文本数据转化为语音数据并播放语音的技术，是涉及声音学、语言学、心理学、数字信号处理、多媒体合成等多种学科的前沿高新技术。其表现就是计算机系统能说话，其目的就是用机器代替人来组织语音并说话。

语音合成技术的原理如图 2 所示。

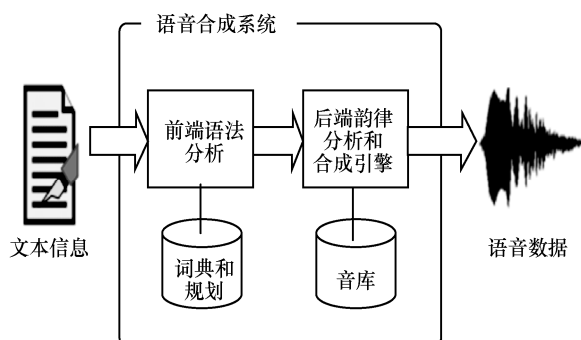


图 2 语音合成技术原理示意图

首先，语音合成系统从输入端口读入文本信息，由前端语法分析系统完成语法分析。主要包括：规范化文本，过滤不规范和无法发音的字符；通过自动分词和语句划分，分析文本中词、短语及语句，确定读音并对多音字进行处理；确定发音时的轻重音及语气变化。结合词典和规则进行处理后，得到格式规范并携带该文本的语法层次信息，向后端系统发送。

后端系统一般又称为韵律分析和语音合成系统，是在前端系统提供的语法分析的基础上，由韵律分析子系统根据韵律合成规则得到将输出语音的几个要素，包括音高、音强、音长、语调、停顿等。这些要素被输入语音合成引擎子系统，该子系统按照这些语音要素对照文本数据，运用语音合成算法合成出满足要求的音节波形数据，再将其拼接成语音流，输出到语音播放系统播放。

4 人工智能技术在电信客服系统中的应用

在电信客服系统中引入人工智能技术，其架构基本与基于人工智能的客服系统的通用架构一致，在现网客服系统中引入这些主要功能模块，并嵌入工作流程中。鉴于人工智能应用还不成熟，在技术上有待进一步细化、提升，同时在应用中也有待进一步细分场景和用途，在电信客服系统中人工智能客服的应用基本聚焦在某些特殊场景，如某类业务的主动外呼推广或者某类业务的投诉受理。

就技术的应用而言，由于行业的特殊性及本身企业的地位，当前电信客服系统主要在自然语言处理方面做了一些适应自身特点的技术应用优选和完善。

电信客服系统因为主要服务于企业本身，所以在自然语言处理中其最基本的语术一定是基于电信业务范畴的专业语言，但是用户的语言是日常语术，需要做专门的词向量处理及相应的短句匹配，才能更精确地分析问题和回答问题，满足

客户的要求。

4.1 词向量处理

以电信客服体系内所有的话单、工单作为客服系统专业语料库；再利用网页爬虫程序获取互联网网站的通用文本作为通用语料库。针对专业语料库和通用语料库应用的不同权重，采用“词库+词频统计”的分词方法，从这些文本数据中无监督地学习出词向量^[3]，采用词向量处理框架（如 Word2Vec）^[6]，建立一种适用于客服系统的词库。该词库包含基本词汇和词向量数据，通过计算词向量，可以输出和某个词相关性最高的几个词，或者从几个词中选出同类词，排除异类词等。在进行用户语言进行语义分析时，这些词向量可以协助联想、选择、提醒等，最终提高处理的速度和效率。词向量处理示意图如图 3 所示。

流程说明如下。

(1) 分词及向量处理

- 采用最大匹配法（正向逆向均可）与词库比对，当遇到无法匹配的字时，将其整体切割为一个词，并更新到基本词汇表；
- 输入下一篇文本后，采用更新后的基本词汇表继续做分词；
- 分词后生成词向量，并迭代计算，采用相应模型优化词向量。

(2) 更新词库^[7]

- 训练结束后，更新基础词汇表和词向量；
- 最后在词库中去掉小于某个预定阈值的词

向量和词。

4.2 短文本相似度比较

在文本处理后，需要就文本的问题在知识库中快速地找到最佳答案。电信的客服系统对于用户满意度要求非常高，为了解答用户问题，务必要找到答案，或者是最佳答案。在这里系统使用了短文本相似度^[8]比较的技术方案。

(1) 逐条比较短文本 1 与短文本 2~短文本 N 的相似度。

(2) 短文本 1 (记为 sentence1) 与短文本 2 (记为 sentence2) 的相似度 (记为 S) 为：

$$S = D / L_{\text{long}}, D = d_1 + d_2 + \dots + d_{L_{\text{max}}}, L_{\text{long}} = \text{Max}(\text{len}(\text{sentence1}), \text{len}(\text{sentence2})).$$

$d_1, d_2, \dots, d_{L_{\text{max}}}$ 的计算方法为：以较短文本为基础，计算 word1 与长文本每个 word 的欧氏距离，取最小值为 d_1 ；计算 word2 与长文本每个 word 的欧氏距离，取最小值为 d_2 ；直至短文本内所有 word 比较完毕。

将长文本内剩余 word 与短文本内全部 word 进行比较，计算长文本最后一个 word 与短文本所有 word 的欧氏距离后，取最小值为 $d_{L_{\text{max}}}$ 。

(3) 引入同义词与近义词表，同义词与近义词距离为 0。 S 越接近 0，代表越相似；越接近 1，代表不相似。

(4) 逐步递归，最后找到最接近的问题和答案。

4.3 应用案例及效果

电信运营商每月都会有一定比例的投诉或故

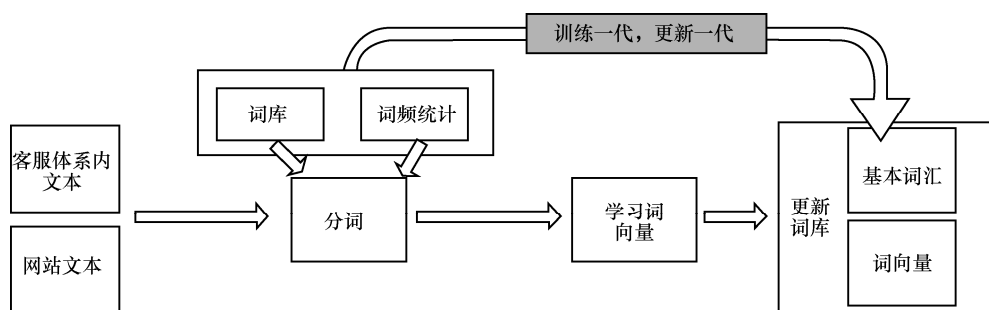


图 3 词向量处理示意图



障难以找到故障原因或者第一责任方，给排障和定则带来极大的困扰，造成了较大的资源浪费。另外，在客服与客户交互层面也有相当多的问题无法及时答复或解决，效率不高甚至经常引起更高层级的投诉或更多的问题，对于工作的顺利开展有极大的影响。在后台引入自然语言处理的新技术，能够在一定程度上减少类似问题，同时对于事件或故障的响应效率、问题的解决效率得到了较大的提升。

如下，就一个一般性问题以短文本的方式举例说明。

当前有工单短文本 A：“赠送您的免费宽带即将到期”，与以下工单短文本：B：“您的本月账单已出账，请通过邮箱查收”、C：“您报修的 IPTV 机顶盒故障已检修完毕”、D：“您的家庭套餐可以升级为十全十美套餐，续费享九折优惠”、E：“您已欠费三个月，您的手机将于本月底停机”，在之前现网系统或人工服务中，短文本 A 与已有的工单短文本均无法匹配，在没有更新知识库或有新的范例生成前，工单有可能无法解决。但是使用了相应的技术，就会有一定程度的匹配关系供参考。

短文本处理时首先将以上 5 条短文本转换成对应的词向量，每个词向量为 30 维，如图 4 所示。

短文本 A 中每个词向量与短文本 B、C、D、E 中的每个词向量相比较，取欧氏距离最小的值作为两个词向量的距离，再取均值作为两个短文本之间的相似度，那么取值较小则为更相似，则短文本 A 与短文本 B、C、D、E 之间的相似度为： $S_{AB} = 0.40$ ， $S_{AC} = 0.42$ ， $S_{AD} = 0.27$ ， $S_{AE} = 0.37$ 。

按以上结果，相似度排序为（近到远）：D：

您的家庭套餐可以升级为十全十美套餐，续费享九折优惠→E：您已欠费三个月，您的手机将于本月底停机→B：您的本月账单已出账，请通过邮箱查收→C：您报修的 IPTV 机顶盒故障已检修完毕。

虽然采用自然语言处理技术后，结果不一定完全准确，但是对于那些无解的问题至少提供了一定的判决及解决的可能。当然，在实际操作时，对于这些相似度低的问题，一定会有人工介入，而这些结果也为操作员提供了一定的解决思路 and 方向。

事实上，在应用到累积的无法定位问题后，有至少 30% 的问题最后及时定位到责任方并找到了根源。随着自然语言处理技术的大量应用、数

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	赠送	-2.13167	-0.12078	-0.67824	-0.3161	-1.31163	-2.50861	-1.24456	-1.79806	-0.98577	-1.74522	-1.00938	1.992646	-0.84667	-2.09141	-0.76516	1.174665	-0.09838	-1.04394	-0.25009
2	您	-0.57116	-0.44843	1.776688	0.960239	0.221712	-1.9963	1.110934	-1.21843	-0.0156	-1.04074	1.335017	-2.01046	-1.9917	-2.11612	-0.3059	0.188369	-0.5161	-0.23435	1.542313
3	免费	-1.18899	0.005642	-1.3071	-0.31661	-2.59882	-0.03954	-0.51719	-0.82785	-0.72444	-1.69822	-1.66173	1.682736	-0.1803	-2.58376	0.906559	-2.14045	0.514931	-0.57344	-0.3756
4	宽带	-0.83719	-0.80864	0.251608	-0.63303	0.106974	1.349145	1.490485	-2.84994	-0.56776	-0.98191	0.009623	-1.27426	-0.3685	-0.04919	-0.09639	-0.80785	-0.09941	0.978413	0.244119
5	即将	-4.49415	2.618621	-2.38246	-0.85351	1.428316	-2.91372	-1.52131	0.330393	-0.56051	-2.27576	0.008949	-0.37971	0.855873	-1.40433	-2.91091	0.253515	2.846673	-0.43426	-0.18285
6	到期	-1.31172	1.5695	-2.95109	-0.15815	0.293742	-0.57261	-1.66503	-2.71398	0.703909	0.004898	-0.15518	0.365578	0.198744	-0.96101	-2.39904	0.033427	0.589928	1.735399	0.666903
7																				
8	您	-0.57116	-0.44843	1.776688	0.960239	0.221712	-1.9963	1.110934	-1.21843	-0.0156	-1.04074	1.335017	-2.01046	-1.9917	-2.11612	-0.3059	0.188369	-0.5161	-0.23435	1.542313
9	本月	1.603484	0.118516	0.183163	2.851616	-1.38984	0.323749	1.240395	-2.61057	1.33414	1.463613	-1.21394	1.33061	-0.35891	1.615306	0.031166	1.979082	1.057611	0.983915	0.776918
10	账单	-2.39342	0.330807	-1.73321	-1.82339	0.733524	-1.64356	-0.9317	-0.09692	1.818382	-0.2104	2.722718	-2.27251	1.124705	2.152744	-2.51852	1.647695	1.346197	-1.08921	-0.97211
11	已	-0.67449	1.685305	1.523252	-0.75089	-2.61939	0.014362	-1.79199	-0.19949	0.656705	0.982653	1.720584	1.135383	-3.34757	-2.3993	1.741591	0.808956	-1.61125	1.750372	1.352754
12	出版	-0.12962	1.436871	1.446123	0.475699	-0.71278	-1.15219	2.122848	-1.33432	0.742563	0.446563	1.327928	1.370389	-0.32748	0.715708	0.310914	-0.22803	1.112952	2.647179	1.597395
13	请	0.54443	2.441972	0.325594	1.368916	-2.81839	2.902426	-0.72779	-0.65075	0.51623	-4.27629	-0.76272	3.680532	1.154752	-3.17973	0.197764	-3.69582	0.513955	1.945533	-0.26168
14	通过	-1.3136	2.901302	-0.04697	0.595871	3.641457	2.725097	-0.61125	-3.4539	-1.11138	0.609843	1.093309	-0.93886	-0.61142	0.738558	-0.82251	0.58865	0.732277	-4.11568	2.191067
15	邮箱	1.432163	1.164355	0.117632	-0.73584	-2.54733	-1.18121	0.709356	0.035095	1.367174	0.284783	1.29594	2.392256	-2.25468	-3.00869	-0.2347	1.008064	-1.26118	0.477622	0.707587
16	查收	-1.91454	-2.20849	2.138646	1.686555	1.636795	1.426687	2.967984	-0.0959	1.957289	1.04789	0.312155	2.028705	-1.03382	-2.85142	-1.35905	-2.23187	-1.9708	0.596261	-0.97706
17																				
18	您	-0.57116	-0.44843	1.776688	0.960239	0.221712	-1.9963	1.110934	-1.21843	-0.0156	-1.04074	1.335017	-2.01046	-1.9917	-2.11612	-0.3059	0.188369	-0.5161	-0.23435	1.542313
19	报修	-1.8626	-1.24639	0.992404	-0.35505	-1.21995	-0.98212	1.047398	0.667836	-2.2656	-2.98406	2.610678	-0.0389	-0.86086	-0.86604	-2.50582	-1.0134	1.334214	0.817273	-0.95435
20	IPTV	-3.44359	1.99903	1.316633	-0.97779	2.329243	-1.33106	2.017172	0.739509	4.001949	1.079552	1.24645	0.526406	-2.02119	0.082836	-1.13678	0.507477	2.017748	-1.86214	-0.97519
21	机顶盒	-0.53306	0.41232	0.763027	-0.90172	0.89879	0.403129	0.767148	0.918182	-0.45178	-0.38987	0.159712	-0.2442	0.490896	-0.27283	-0.30499	-0.52805	-0.46037	0.681224	-0.06275
22	故障	0.675676	0.756013	-0.58446	-0.22066	0.47788	-0.39628	0.060072	-2.8091	0.178231	1.101646	0.226292	1.037147	-1.16378	1.35074	-1.66676	0.308686	0.188782	0.852395	2.166865
23	已	-0.67449	1.685305	1.523252	-0.75089	-2.61939	0.014362	-1.79199	-0.19949	0.656705	0.982653	1.720584	1.135383	-3.34757	-2.3993	1.741591	0.808956	-1.61125	1.750372	1.352754
24	检修	-3.30466	-1.19405	-2.0395	-1.06668	0.801044	-1.17926	-0.28319	-1.17987	-0.31629	-1.1171	2.034387	-1.00045	1.275788	2.500851	-2.66672	1.586769	2.543636	0.035226	-3.53455
25	完毕	-0.54572	-0.73049	-1.71888	-1.26958	0.063359	-0.54254	-0.80259	-1.78783	-0.17042	0.744752	1.104963	-0.57018	-1.238	0.429593	-1.36495	1.203467	-0.50042	-0.07347	1.67067
26																				
27	您	-0.57116	-0.44843	1.776688	0.960239	0.221712	-1.9963	1.110934	-1.21843	-0.0156	-1.04074	1.335017	-2.01046	-1.9917	-2.11612	-0.3059	0.188369	-0.5161	-0.23435	1.542313
28	家庭	0.622689	-0.46697	-2.04399	1.216907	-1.7959	-1.09396	-0.35444	-0.52122	0.698918	1.589509	-1.2094	-1.7515	1.221143	0.493279	1.124272	0.072398	-1.8451	1.697494	1.498112
29	套餐	0.931374	-1.93147	1.694631	2.961726	-2.8581	-0.23946	-0.7248	0.674699	0.44198	-0.86743	1.943589	1.532229	-2.56993	-0.26742	1.827539	1.142195	1.039567	-0.36246	0.756889
30	可以	-3.7876	-1.91401	3.393902	-0.39834	0.39506	-0.39584	1.106389	0.727866	0.482851	0.088991	1.284662	1.394226	-3.55609	-0.99288	-0.68175	-0.7899	1.892808	-1.97753	-1.36404
31	升级	0.655194	1.348051	0.63117	-0.27801	-0.64124	0.500378	0.023276	-0.33775	1.409169	0.954984	0.934074	-1.49671	-1.77326	-1.18141	-2.07434	-0.41807	1.502526	1.156038	0.154842
32	为	-2.49394	0.635291	-1.1378	0.029719	2.069812	-2.14475	1.565098	-1.46544	0.864486	-0.63104	1.086972	-0.92419	-0.00647	1.311671	-1.93355	1.55135	1.322315	0.09612	-0.88405
	Sheet1	⊕																		

图 4 词向量处理结果示例

据量的增长以及知识库的完善和细化,在理论上无法解决的问题会逐步减少直至消失。

5 结束语

虽然基于人工智能的客服系统方兴未艾,伴随着大量的投入在线使用,系统的问题也开始逐步显现。表现在用户侧,则是系统的答复比较呆板、客服语气生硬(没有感情、缺乏人气),形式比较单一、基本采用一问一答的模式、主动关怀用户感受的能力不够,对于顾客的诉求或者反馈的问题答案相对有限、跳出一定范围容易出现答非所问的情况。表现在企业侧,则是虽然自动化程度高了,但是成功率未见显著提升,而用户满意度在某些业务方面反而有些下降。

电信作为提供普遍通信服务的企业,有大量的业务服务及客户服务的需求,而且对服务的质量也有相当高的要求。伴随着人工智能技术的演进及应用的成熟,对于服务的质量、效率、场景的要求也会越来越高、越来越多。有时在面对某些客户或事件,需要应用更为特殊及适配的话术及技巧;有时需要在整个客服活动中,跟踪和预测用户的情绪变化;有时在与人工客服自动呼转,需要挑选合适的话务员。这些都需要在处理的实时性以及技术的细化、深化上进行更多的研究,下更大的功夫。

总之,以后在客服体系内,人工智能的应用会越来越多,人工智能坐席会替代人工坐席,而机器也会越来越像人。

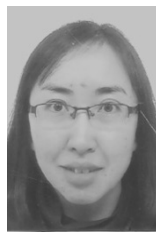
参考文献:

- [1] ZHENG X, CHEN H, XU T, et al. Deep learning for Chinese word segmentation and POS tagging[J]. Empirical Methods in Natural Language Processing, 2003: 647-657.
- [2] 王灿辉, 张敏, 马少平. 自然语言处理在信息检索中的应用综述[J]. 中文信息学报, 2007, 21(2): 35-45.
WANG C H, ZHANG M, MA S P. A survey of natural language processing in information retrieval [J]. Journey of Chinese Information Processing, 2007, 21(2): 34-45.
- [3] BIRD S, KLEIN E, LOPER E. Python 自然语言处理[M]. 陈涛, 张旭, 崔扬, 等译. 北京: 人民邮电出版社, 2014.
- [4] BIRD S, KLEIN E, LOPER E. Python natural language processing[M]. Translated by CHENG T, ZHANG X, CUI Y, et al. Beijing: Posts&Telecom Press, 2014.
- [4] 李涛, 王次臣, 李华康. 知识图谱的发展与构建[J]. 南京理工大学学报, 2017, 41(1): 22-34.
LI TAO, WANG C C, LI H K. Development and construction of knowledge graph[J]. Journey of Nanjing University of Science and Technology, 2017, 41(1): 22-34.
- [5] 蔡鑫, 娄京生. 基于 LSTM 深度学习模型的中国电信官方微博用户情绪分析[J]. 电信科学, 2017, 33(12): 136-141.
CAI X, LOU J S. Sentiment analysis of telecom official micro-blog users based on LSTM deep learning model [J]. Telecommunications Science, 2017, 33(12): 136-141.
- [6] ANOTONELLIS I, BOURAS C, POULOPOULOS V. Personalized news categorization through scalable text classification[C]//The 8th Asia-Pacific Web Conference on Frontiers of WWW Research and Development, January 16-18, 2006, Harbin, China. New York: ACM Press, 2006: 391-401.
- [7] HU M, LIU B. Mining and summarizing customer review[C]//The Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 22-25, 2004, Seattle, WA, USA. New York: ACM Press, 2004: 168-177.
- [8] JOULIN A, GRAVE E, BOJANOWSKI P. Bag of tricks for efficient text classification[J]. Computer Science, arXiv: 1607.01759v3.

[作者简介]



王铮(1973-), 男, 中国电信股份有限公司上海研究院高级工程师, 主要研究方向为大数据架构、数据挖掘分析和人工智能技术。



任华(1977-), 女, 中国电信股份有限公司上海研究院高级工程师, 主要研究方向为数据挖掘分析、大数据和人工智能技术及应用。



路绪海(1983-), 男, 中国电信股份有限公司上海研究院高级工程师, 主要研究方向为数据挖掘分析和人工智能技术。