



(12) 发明专利

(10) 授权公告号 CN 107025228 B

(45) 授权公告日 2021.01.26

(21) 申请号 201610065638.2

(56) 对比文件

(22) 申请日 2016.01.29

CN 104572734 A, 2015.04.29

(65) 同一申请的已公布的文献号

US 2015088850 A1, 2015.03.26

申请公布号 CN 107025228 A

CN 105279288 A, 2016.01.27

(43) 申请公布日 2017.08.08

赵美雪.基于用户兴趣模型的个性化推荐算法研究.《中国优秀硕士学位论文全文数据库 信息科技辑》.2016,(第01期),第2章第2.2节.

(73) 专利权人 阿里巴巴集团控股有限公司

审查员 杨琦

地址 英属开曼群岛大开曼资本大厦一座四层847号邮箱

(72) 发明人 姜晓燕 代斌 杨旭 褚嵒 赵耀

(74) 专利代理机构 北京睿博行远知识产权代理有限公司 11297

代理人 龚家骅

(51) Int.Cl.

G06F 16/9535 (2019.01)

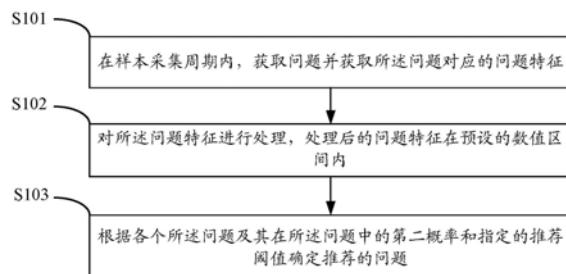
权利要求书2页 说明书6页 附图2页

(54) 发明名称

一种问题推荐方法及设备

(57) 摘要

本发明公开了一种问题推荐方法及设备,该方法包括:在样本采集周期内,获取问题并获取所述问题对应的问题特征;对所述问题特征进行处理,处理后的问题特征在预设的数值区间内;根据各个所述问题及其在所述问题中的第二概率和指定的推荐阈值确定推荐的问题;其中,各个所述问题及其在所述问题中的第二概率是通过所述处理后的问题特征和第一概率得到的;所述第一概率是通过所述问题特征得到的。应用本申请的方法,通过对以往问题特征计算得到需要向用户推荐的问题,提高了向用户推荐问题的准确率。



1. 一种问题推荐方法,其特征在于,该方法包括:

在样本采集周期内,获取问题并获取所述问题对应的问题特征;

对所述问题特征进行处理,处理后的问题特征在预设的数值区间内;

根据各个所述问题及其在所述问题中的第二概率和指定的推荐阈值确定推荐的问题;

其中,各个所述问题及其在所述问题中的第二概率是通过所述处理后的问题特征和第一概率得到的;所述第一概率是通过分类模型对所述问题特征进行分析得到的;所述问题特征包括数值型特征和文本型特征,所述数值型特征为连续的,所述文本型特征为不连续的;所述第二概率是通过对所述处理后的问题特征和所述第一概率进行深度神经网络DNN计算得到的;

在所述深度神经网络DNN训练前,通过第一概率控制深度神经网络DNN中网络节点所占的权重。

2. 如权利要求1所述的方法,其特征在于,所述获取问题,具体包括:

在特征获取周期内获取所述问题;

若存在特征获取周期内未获取的问题,则所述未获取的问题的值为空;

若不存在特征获取周期内未获取的问题,则以获取的问题作为所述问题。

3. 如权利要求1所述的方法,其特征在于,所述获取所述问题对应的问题特征,具体包括:

在特征获取周期内获取问题特征;

若存在特征获取周期内未获取的问题特征,且该问题特征为所述数值型特征时,则以获取的与问题对应的问题特征的数值的均值作为问题特征;

若存在特征获取周期内未获取的问题特征,且该问题特征为所述文本型特征时,则以获取的与问题对应的问题特征中出现频率最高的作为问题特征;

若不存在特征获取周期内未获取的问题特征,则以获取的问题特征作为所述问题特征。

4. 如权利要求1所述的方法,其特征在于,对所述问题特征进行处理,具体包括:

若所述问题特征为数值型特征,则对问题特征进行归一化处理;

若所述问题特征为文本型特征,则对问题特征进行向量化处理,向量化处理后的问题特征为数值型特征。

5. 一种问题推荐设备,其特征在于,所述设备包括:

获取模块:在样本采集周期内,获取问题并获取所述问题对应的问题特征;

处理模块:对所述问题特征进行处理,处理后的问题特征在指定的数值区间内;

确定模块:根据各个所述问题及其在所述问题中的第二概率和指定的推荐阈值确定推荐的问题;

其中,各个所述问题及其在所述问题中的第二概率是通过所述处理后的问题特征和第一概率得到的;所述第一概率是通过分类模型对所述问题特征进行分析得到的;所述问题特征包括数值型特征和文本型特征,所述数值型特征为连续的,所述文本型特征为不连续的;所述第二概率是通过对所述处理后的问题特征和所述第一概率进行深度神经网络DNN计算得到的;

在所述深度神经网络DNN训练前,通过第一概率控制深度神经网络DNN中网络节点所占

的权重。

6. 如权利要求5所述的问题推荐设备,其特征在于,所述获取模块具体用于:

在特征获取周期内获取所述问题;

若存在特征获取周期内未获取的问题,则所述未获取的问题的值为空;

若不存在特征获取周期内未获取的问题,则以获取的问题作为所述问题。

7. 如权利要求5所述的问题推荐设备,其特征在于,所述获取模块具体用于:

在特征获取周期内获取问题特征;

若存在特征获取周期内未获取的问题特征,且该问题特征为所述数值型特征时,则以获取的与问题对应的问题特征的数值的均值作为问题特征;

若存在特征获取周期内未获取的问题特征,且该问题特征为所述文本型特征时,则以获取的与问题对应的问题特征中出现频率最高的作为问题特征;

若不存在特征获取周期内未获取的问题特征,则以获取的问题特征作为所述问题特征。

8. 如权利要求5所述的问题推荐设备,其特征在于,所述处理模块具体用于:

若所述问题特征为数值型特征,则对问题特征进行归一化处理;

若所述问题特征为文本型特征,则对问题特征进行向量化处理,向量化处理后的问题特征为数值型特征。

一种问题推荐方法及设备

技术领域

[0001] 本申请涉及通信技术领域,特别涉及一种问题推荐方法,本申请同时还涉及一种问题推荐设备。

背景技术

[0002] 随着双十一、双十二等电商节日的兴起,越来越多人开始网上购物,但“节日经济”对电商也造成了双重冲击即销量和客服的压力暴增。企业的客服一般分为人工客服和自助客服,客服压力的增加促使企业使用更多的人工客服,投入的成本也随之越来越多。

[0003] 因此,需要自助客服系统具有更大的处理能力,才能满足客服需求。自助客服系统,能够自动处理解决用户的问题。而自助客服的系统中待处理数据量的增加,使现有的方法已经不能够处理全量数据。现有的算法随着问题增加,其计算效率下降。并且大部分特征是稀疏的,而现有技术适用于处理稠密特征,这样,系统中问题特征量增加的同时,对用户问题的预测精度就会下降。另外现有技术中的模型单一,效果受限。所以,随着信息的不断爆炸,现在的机器学习模型已经不能满足需求。

[0004] 因此,本领域技术人员亟待解决的技术问题就是如何通过对以往问题特征计算得到需要向用户推荐的问题,提高向用户推荐问题的准确率,进而在自助客服节点解决用户问题,减少用户进入人工客服,降低人工客服成本。

发明内容

[0005] 本发明提供了一种问题推荐方法,用于提高向用户推荐问题的准确率。所述方法包括以下步骤:

[0006] 在样本采集周期内,获取问题并获取所述问题对应的问题特征;

[0007] 对所述问题特征进行处理,处理后的问题特征在预设的数值区间内;

[0008] 根据各个所述问题及其在所述问题中的第二概率和指定的推荐阈值确定推荐的问题;

[0009] 其中,各个所述问题及其在所述问题中的第二概率是通过所述处理后的问题特征和第一概率得到的;所述第一概率是通过所述问题特征得到的。

[0010] 优选地,所述问题特征包括数值型特征和文本型特征,所述数值型特征为连续的,所述文本型特征为不连续的。

[0011] 优选地,所述获取问题,具体包括:

[0012] 在特征获取周期内获取所述问题;

[0013] 若存在特征获取周期内未获取的问题,则所述未获取的问题的值为空;

[0014] 若不存在特征获取周期内未获取的问题,则以获取的问题作为所述问题。

[0015] 优选地,所述获取所述问题对应的问题特征,具体包括:

[0016] 在特征获取周期内获取问题特征;

[0017] 若存在特征获取周期内未获取的问题特征,且该问题特征为所述数值型特征时,

则以获取的与问题对应的问题特征的数值的均值作为问题特征；

[0018] 若存在特征获取周期内未获取的问题特征,且该问题特征为所述文本型特征时,则以获取的与问题对应的问题特征中出现频率最高的作为问题特征;

[0019] 若不存在特征获取周期内未获取的问题特征,则以获取的问题特征作为所述问题特征。

[0020] 优选地,对所述问题特征进行处理,具体包括:

[0021] 若所述问题特征为数值型特征,则对问题特征进行归一化处理;

[0022] 若所述问题特征为文本型特征,则对问题特征进行向量化处理,向量化处理后的问题特征为数值型特征。

[0023] 优选地,所述第二概率是通过对所述处理后的问题特征和所述第一概率进行深度神经网络DNN计算得到的。

[0024] 相应的本申请还提出了一种问题推荐设备,所述设备包括:

[0025] 获取模块:在样本采集周期内,获取问题并获取所述问题对应的问题特征;

[0026] 处理模块:对所述问题特征进行处理,处理后的问题特征在指定的数值区间内;

[0027] 确定模块:根据各个所述问题及其在所述问题中的第二概率和指定的推荐阈值确定推荐的问题;

[0028] 其中,各个所述问题及其在所述问题中的第二概率是通过所述处理后的问题特征和第一概率得到的;所述第一概率是通过所述问题特征得到的。

[0029] 优选地,所述问题特征包括数值型特征和文本型特征,所述数值型特征为连续的,所述文本型特征为不连续的。

[0030] 优选地,所述获取模块具体用于:

[0031] 在特征获取周期内获取所述问题;

[0032] 若存在特征获取周期内未获取的问题,则所述未获取的问题的值为空;

[0033] 若不存在特征获取周期内未获取的问题,则以获取的问题作为所述问题。

[0034] 优选地,所述获取模块具体用于:

[0035] 在特征获取周期内获取问题特征;

[0036] 若存在特征获取周期内未获取的问题特征,且该问题特征为所述数值型特征时,则以获取的与问题对应的问题特征的数值的均值作为问题特征;

[0037] 若存在特征获取周期内未获取的问题特征,且该问题特征为所述文本型特征时,则以获取的与问题对应的问题特征中出现频率最高的作为问题特征;

[0038] 若不存在特征获取周期内未获取的问题特征,则以获取的问题特征作为所述问题特征。

[0039] 优选地,所述处理模块具体用于:

[0040] 若所述问题特征为数值型特征,则对问题特征进行归一化处理;

[0041] 若所述问题特征为文本型特征,则对问题特征进行向量化处理,向量化处理后的问题特征为数值型特征。

[0042] 优选地,所述第二概率是通过对所述处理后的问题特征和所述第一概率进行深度神经网络DNN计算得到的。

[0043] 由此可见,通过应用本申请的技术方案,对问题特征进行处理和分类计算,然后对

得到的结果进行进一步的深度计算,即可准确地输出各个所述问题和第二概率。本申请提高了向用户推荐问题的准确率。本申请可处理稠密性的问题特征值,同时适合处理大规模的稀疏数据,提高了预测精度。

附图说明

- [0044] 图1为本申请提出的一种问题推荐方法的流程示意图;
- [0045] 图2为本申请具体实施例所提出的一种DNN模型示意图;
- [0046] 图3为本申请提出的一种问题推荐设备的结构示意图。

具体实施方式

[0047] 有鉴于现有技术中的问题,本发明提出了问题推荐方法,该方法应用于问题推荐系统中,结合机器学习模型和深度神经网络模型DNN进行模型训练。该系统能够根据历史记录为用户推荐其所需的问题,并且擅长处理稀疏型和稠密性的问题特征,可用于提高向用户推荐问题的准确率。

[0048] 如图1所示,为本申请提出了的一种问题推荐方法的流程示意图,包括以下步骤:

[0049] S101,在样本采集周期内,获取问题并获取所述问题对应的问题特征。

[0050] 本申请旨在向需要问题推荐服务的用户推荐问题。用户可以是该需要问题推荐的用户和其他用户。本发明的系统中存在用户的历史记录,在历史记录中包含各个问题以及对应的问题特征。问题推荐系统通常包括收集层、处理层、存储层和输出层。收集层负责收集其他设备发送的问题及问题特征。处理层利用收集到的问题及问题特征进行模型训练。存储层负责数据存储,其中存储了用户的历史记录。输出层进行问题及问题特征的输出。本申请中的问题推荐系统可以在服务器上实现,优选采用分布式服务器。并且本申请可以使用一个服务器,也可以使用多个服务器组成的集群。

[0051] 所述问题特征包括数值型特征和文本型特征,所述数值型特征为连续的,比如,数值型特征为使用过某个应用软件的次数,数值9代表使用过9次,所述文本型特征为不连续的,比如,文本型特征为发票状态,对应未开发票和已开发票。因为历史记录中的问题和问题特征都具有一定的时效性,所以设置样本采集周期,以采集一段时期之内问题及问题特征,比如一周内、一个月内。其他设备向系统发送问题及问题特征时,由于不同设备的IP不同,一些问题、问题特征的获取时间较长,可能导致系统在很长时间内不能获取完毕。

[0052] 为了更高效地对数据进行处理,在本申请的优选实施例中,设定特征获取周期,在特征获取周期内获取所述问题。若存在特征获取周期内未获取的问题,则所述未获取的问题的值为空,若不存在特征获取周期内未获取的问题,则以获取的问题作为所述问题。在特征获取周期内获取问题特征,若存在特征获取周期内未获取的问题特征,且该问题特征为所述数值型特征时,则以获取的与问题对应的问题特征的数值的均值作为问题特征,若存在特征获取周期内未获取的问题特征,且该问题特征为所述文本型特征时,则以获取的与问题对应的问题特征中出现频率最高的作为问题特征,若不存在特征获取周期内未获取的问题特征,则以获取的问题特征作为所述问题特征。

[0053] 在获取问题及其对应的问题特征之后,推荐系统通过对问题特征进行筛选,以删除一些特征,比如删除所有用户都相同的问题特征、容易超出特征获取周期的问题特征和

与经营业务无关的问题特征。经过筛选得到的特征可以为后续建立分类模型作准备。

[0054] S102,对所述问题特征进行处理,处理后的问题特征在预设的数值区间内。

[0055] 在获取问题和对应的问题特征后,问题推荐系统对问题特征进行处理。若所述问题特征为数值型特征,则对问题特征进行归一化处理以使处理后的问题特征在指定的数值区间内;若所述问题特征为文本型特征,则对问题特征进行向量化处理,以使处理后的问题特征为数值型特征并在指定的数值区间内。

[0056] 在本申请的优选实施例中,若问题特征是数值型特征,可以采用百分位分箱算法进行归一化处理,使所有问题特征在经过处理后都处在指定的数值区间内。在进行百分位分箱算法处理时,将原始数值归纳到100个箱中,然后对箱进行编码,比如0.01,0.02……1.00。经过处理的数值型特征处于0至1的数值区间内。

[0057] 由于文本型特征是以文字的形式呈现的,无法参与计算。所以,需要对文本型特征进行向量化处理,将问题特征由文本型特征转化为数值型特征。可以采用one hot编码处理文本型特征,计算各个特征的频率,从而按频率给出one hot编码。举例来说,文本型特征为发票状态,对应未开发票和已开发票,经过向量化处理后,得到数值型特征0和1,处于0至1的数值区间内。

[0058] 问题特征经过处理后,处于指定的数值区间内,以便于参与后续计算。需要说明的是,本申请需要得到处于指定数值区域内的问题特征,故进行上述百分位分箱算法和向量化处理方法仅为本申请优选实施例提出的示例,本申请的保护范围并不限于此,以上仅为本申请优选实施例提出的示例,在此基础上还可以选择其他方式来进行计算,以使本申请适用于更多的应用领域,这些改进都属于本发明的保护范围。

[0059] S103,根据各个所述问题及其在所述问题中的第二概率和指定的推荐阈值确定推荐的问题。

[0060] 在获取问题和对应的问题特征后,本申请还需要对问题特征进行简单分类模型计算,可以采用决策树算法,从而得到第一概率。在进行决策树计算时,首先需要有两轮采样过程。第一轮对问题特征进行随机采样,得到决策树可处理的问题特征。第二轮进行重要特征采样,根据所述可处理的问题特征计算权重。

[0061] 在当今数据接口越来越多的情况下,数据集的原始变量、衍生变量会越来越多,因此信息值IV(InformationValue)在实际数据应用中十分重要。信息值IV用来表示每一个变量对目标变量来说有多少“信息”的量,从而使得特征选择变得简单快速。

[0062] 在进行特征选择时,总是在将特征的重要程度量化之后再进行选择,而如何量化特征,就成了各种方法间最大的不同。在信息增益中,重要性的衡量标准就是看特征能够为分类系统带来多少信息,带来的信息越多,该特征越重要。因此对于一个特征而言,信息增益为系统存在该特征和不存在该特征时候的信息量各是多少,两者的差值就是这个特征给系统带来的信息量,即信息增益IG(Information Gain)。

[0063] 因为信息值IV和信息增益IG都可以表示问题特征对应的权重,所以,所述权重为信息值IV和/或信息增益IG,根据权重选出重要特征,然后根据所述重要特征建立分类模型。然后,通过分类模型对筛选得到的问题特征进行分析得到第一概率。将各个问题特征经过决策树计算后得到的相应概率作为所述第一概率。

[0064] 需要说明的是,本申请进行以上基于决策树算法得到的各个问题特征对应概率值

的过程仅为本申请优选实施例提出的示例,除此之外还可以选择比如逻辑回归计算等的其他计算方法,以使本申请适用于更多的应用领域,这些改进都属于本发明的保护范围。

[0065] 在得到处理后的问题特征和第一概率后,对这两者进行计算,得到各个所述问题及其在所述问题中的第二概率。在优选的实施例中,所述各个所述问题及其在所述问题中的第二概率通过DNN计算得到。问题推荐系统中的深度神经网络DNN (Deep Neural Network),包括输入节点和计算节点。DNN计算包括以下步骤:(1)输入节点获取所述处理后的问题特征和第一概率。(2)计算节点通过全连接层、激活函数ReLU和多分类损失函数softmax loss对处理后的问题特征和第一概率进行计算,得到第二概率。

[0066] 在具体的应用场景中,操作步骤如图2所示:

[0067] a.由输入层获取所述处理后的问题特征和第一概率。

[0068] 在DNN训练前,可以使用决策树对数据进行初步分类,同时可以通过第一概率控制深度神经网络DNN中网络节点所占的权重。

[0069] b.由中间层即计算层进行问题推荐,计算层通过全连接层、激活函数ReLU和多分类损失函数softmax loss对所述处理后的问题特征和第一概率进行计算,得到各个问题特征对应的问题及第二概率。

[0070] 利用激活函数ReLU进行计算,使一部分网络中神经元的输出为0,因而造就了网络的稀疏性,并且减少了参数的相互依存关系,缓解了过拟合问题的发生。同时,使计算节点的计算量较小,有利于提高系统推荐问题的效率。另外,DNN训练可以使用GPU,能够针对矩阵计算进行加速,进一步提高计算速度。除激活函数ReLU之外,还可以使用sigmoid层进行计算。

[0071] c.输出层输出各个所述问题及其对应的第二概率。

[0072] 需要说明的是,本申请是由第一概率和处理后得到的数值型特征,得到第二概率,本申请提出的计算方式是DNN计算,本申请的保护范围并不限于此,以上仅为优选实施例提出的示例,在此基础上还可以选择其他方式来进行计算,以使本申请适用于更多的应用领域,这些改进都属于本发明的保护范围。

[0073] 在本申请的优选实施例中,问题推荐系统根据各个所述问题及其在所述问题中的第二概率和指定的推荐阈值,确定推荐的问题。然后根据阈值得到阈值内的问题特征,进而以该问题特征对应的问题作为推荐的问题。比如在阈值内得到六个问题的问题特征,则系统推荐这六个问题。本发明在针对各个用户对应的历史记录中的问题和问题特征进行计算,进而确定待推荐的问题之后,当用户访问问题推荐系统的时候直接调用相应的结果。通过本申请中的问题推荐系统,用户能够直接获取与之相关性非常高的问题。

[0074] 为达到以上技术目的,本申请还提出了一种问题推荐设备,如图3所示,所述设备包括:

[0075] 获取模块310:在样本采集周期内,获取问题并获取所述问题对应的问题特征;

[0076] 处理模块320:对所述问题特征进行处理,处理后的问题特征在指定的数值区间内;

[0077] 确定模块330:根据各个所述问题及其在所述问题中的第二概率和指定的推荐阈值确定推荐的问题;

[0078] 其中,各个所述问题及其在所述问题中的第二概率是通过所述处理后的问题特征

和第一概率得到的；所述第一概率是通过所述问题特征得到的。

[0079] 在优选的实施例中，所述问题特征包括数值型特征和文本型特征，所述数值型特征为连续的，所述文本型特征为不连续的。

[0080] 在优选的实施例中，所述获取模块具体用于：

[0081] 在特征获取周期内获取所述问题；

[0082] 若存在特征获取周期内未获取的问题，则所述未获取的问题的值为空；

[0083] 若不存在特征获取周期内未获取的问题，则以获取的问题作为所述问题。

[0084] 在优选的实施例中，所述获取模块具体用于：

[0085] 在特征获取周期内获取问题特征；

[0086] 若存在特征获取周期内未获取的问题特征，且该问题特征为所述数值型特征时，则以获取的与问题对应的问题特征的数值的均值作为问题特征；

[0087] 若存在特征获取周期内未获取的问题特征，且该问题特征为所述文本型特征时，则以获取的与问题对应的问题特征中出现频率最高的作为问题特征；

[0088] 若不存在特征获取周期内未获取的问题特征，则以获取的问题特征作为所述问题特征。

[0089] 在优选的实施例中，所述处理模块具体用于：

[0090] 若所述问题特征为数值型特征，则对问题特征进行归一化处理；

[0091] 若所述问题特征为文本型特征，则对问题特征进行向量化处理，向量化处理后的问题特征为数值型特征。

[0092] 在优选的实施例中，所述第二概率是通过对所述处理后的问题特征和所述第一概率进行深度神经网络DNN计算得到的。

[0093] 通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到本申请可以通过硬件实现，也可以借助软件加必要的通用硬件平台的方式来实现。基于这样的理解，本申请的技术方案可以以软件产品的形式体现出来，该软件产品可以存储在一个非易失性存储介质(可以是CD-ROM, U盘, 移动硬盘等)中，包括若干指令用以使得一台计算机设备(可以是个人计算机, 服务器, 或者网络设备等)执行本申请各个实施场景所述的方法。

[0094] 本领域技术人员可以理解附图只是一个优选实施场景的示意图，附图中的模块或流程并不一定是实施本申请所必须的。

[0095] 本领域技术人员可以理解实施场景中的装置中的模块可以按照实施场景描述进行分布于实施场景的装置中，也可以进行相应变化位于不同于本实施场景的一个或多个装置中。上述实施场景的模块可以合并为一个模块，也可以进一步拆分成多个子模块。

[0096] 上述本申请序号仅仅为了描述，不代表实施场景的优劣。

[0097] 以上公开的仅为本申请的几个具体实施场景，但是，本申请并非局限于此，任何本领域的技术人员能思之的变化都应落入本申请的保护范围。

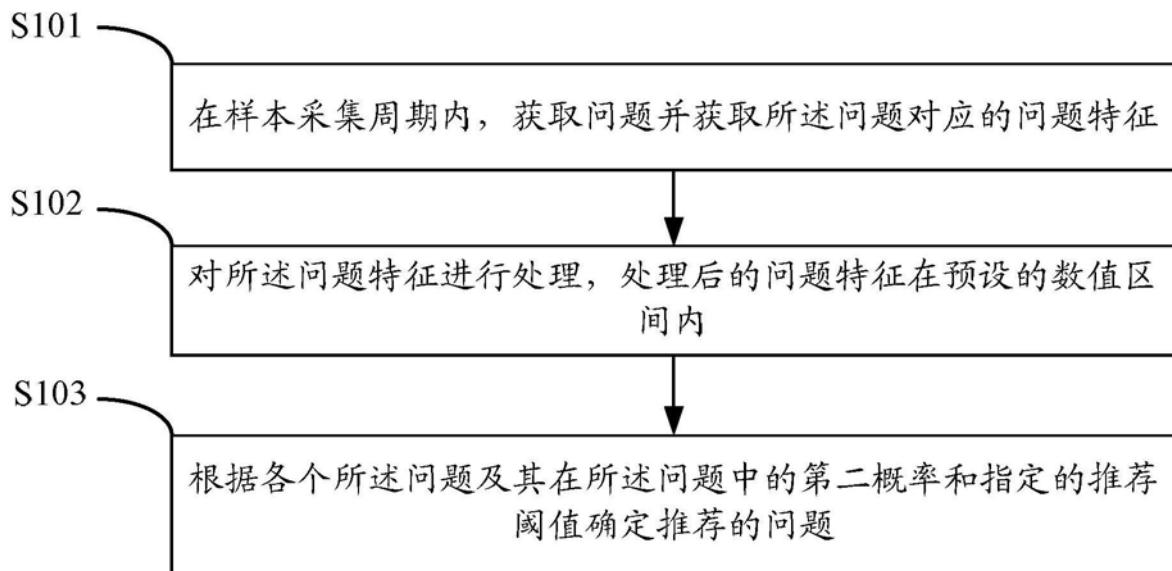
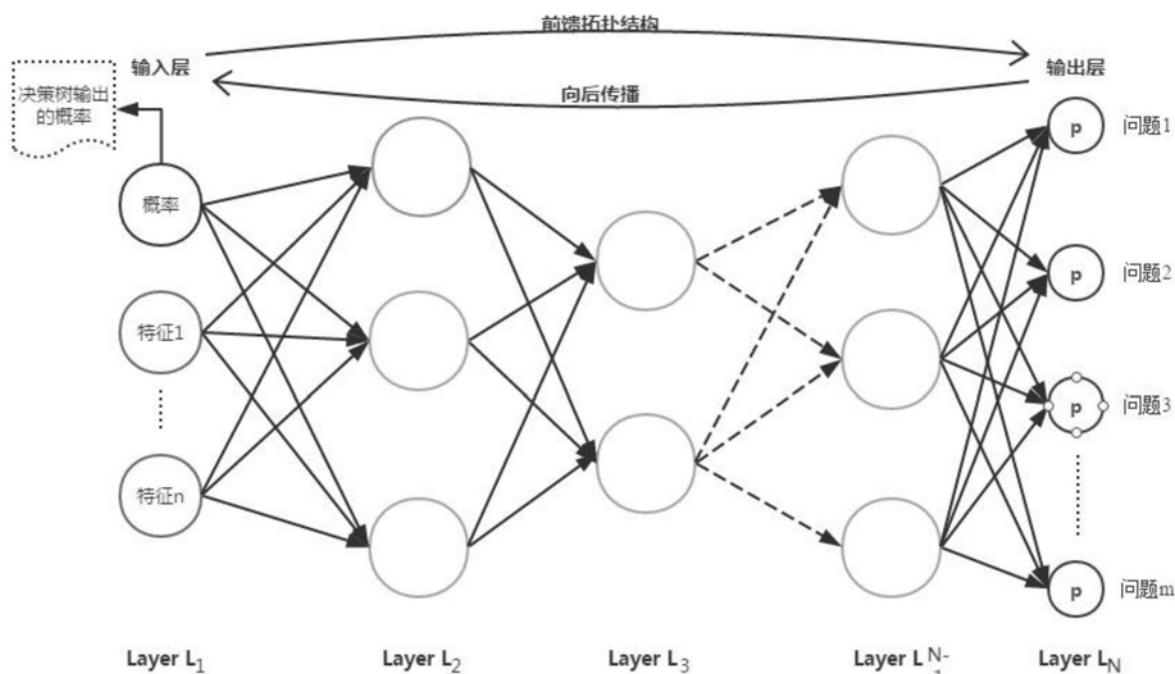


图1



智能客服推荐问题DNN模型

图2



图3