(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 111737543 A
(43)申请公布日 2020.10.02

(21)申请号 201910447437.2

(22)申请日 2019.05.27

(71)申请人 北京京东尚科信息技术有限公司
地址 100086 北京市海淀区知春路76号8层
申请人 北京京东世纪贸易有限公司

(72)发明人 王培英 陈蒙

(74)专利代理机构 北京品源专利代理有限公司 11332
代理人 孟金喆

(51)Int.Cl.
*G06F 16/9032*(2019.01)
*G06K 9/62*(2006.01)
*G06Q 30/02*(2012.01)

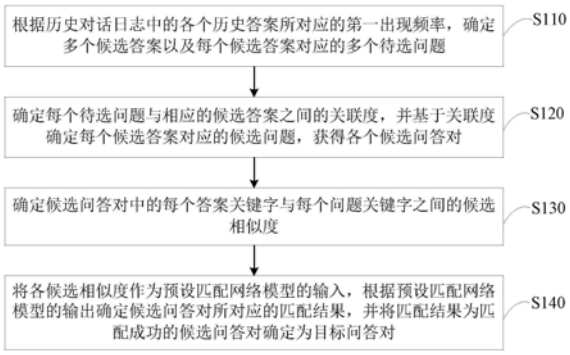
权利要求书3页 说明书13页 附图5页

(54)发明名称

一种问答对的提取方法、装置、设备和存储介质

(57)摘要

本发明实施例公开了一种问答对的提取方法、装置、设备和存储介质，该方法包括：根据历史对话日志中的各个历史答案所对应的第一出现频率，确定多个候选答案以及每个候选答案对应的多个待选问题；确定每个待选问题与相应的候选答案之间的关联度，并基于关联度确定每个候选答案对应的候选问题，获得各个候选问答对；确定候选问答对中的每个答案关键字与每个问题关键字之间的候选相似度；将各候选相似度作为预设匹配网络模型的输入，根据预设匹配网络模型的输出确定候选问答对所对应的匹配结果，并将匹配结果为匹配成功的候选问答对确定为目标问答对。通过本发明实施例的技术方案，可以提取出较多高质量的问答对，提高提取的准确性。




CN 111737543 A

1.一种问答对的提取方法,其特征在于,包括:

根据历史对话日志中的各个历史答案所对应的第一出现频率,确定多个候选答案以及每个所述候选答案对应的多个待选问题;

确定每个所述待选问题与相应的候选答案之间的关联度,并基于所述关联度确定每个所述候选答案对应的候选问题,获得各个候选问答对;

确定所述候选问答对中的每个答案关键字与每个问题关键字之间的候选相似度;

将各所述候选相似度作为预设匹配网络模型的输入,根据所述预设匹配网络模型的输出确定所述候选问答对所对应的匹配结果,并将所述匹配结果为匹配成功的候选问答对确定为目标问答对。

2.根据权利要求1所述的方法,其特征在于,根据历史对话日志中的各个历史答案所对应的第一出现频率,确定多个候选答案以及每个所述候选答案对应的多个待选问题,包括:

基于预设过滤规则,对历史对话日志中的各个历史答案进行过滤;

统计过滤出的历史答案所对应的第一出现频率,将所述第一出现频率大于或等于预设频率的历史答案作为候选答案,小于预设频率的历史答案作为第一答案;

确定每个所述第一答案与每个所述候选答案之间的答案相似度,将所述答案相似度大于或等于预设相似度的第一答案也作为候选答案;

根据所述历史对话日志中的各个对话记录,提取在每个所述对话记录中位于所述候选答案之前的用户问题,并根据从各所述对话记录中提取出的用户问题确定所述候选答案对应的多个待选问题。

3.根据权利要求2所述的方法,其特征在于,根据从各所述对话记录中提取出的用户问题确定所述候选答案对应的多个待选问题,包括:

若从所述对话记录中提取出的用户问题为多个,则获得距离所述候选答案最近的第一用户问题;

检测是否存在与所述第一用户问题连续提问的第二用户问题;

若是,则将所述第一用户问题和所述第二用户问题均作为所述候选答案对应的待选问题;

若否,则将所述第一用户问题作为所述候选答案对应的待选问题。

4.根据权利要求1所述的方法,其特征在于,确定每个所述待选问题与相应的候选答案之间的关联度,并基于所述关联度确定每个所述候选答案对应的候选问题,获得各个候选问答对,包括:

对各所述候选答案进行聚类,确定各个候选答案集合,并根据所述候选答案与所述待选问题之间的对应关系,确定每个所述候选答案集合对应的待选问题集合;

对所述待选问题集合中的各个待选问题进行问题关键字提取,确定每个所述待选问题集合对应的问题关键字集合;

根据各所述问题关键字集合,确定所述问题关键字集合中的每个所述问题关键字对应的关键字重要程度,并基于各所述关键字重要程度确定每个待选问题与相应的候选答案之间关联度;

将所述关联度最高的待选问题确定为相应的候选答案所对应的候选问题,获得候选问答对。

5.根据权利要求4所述的方法，其特征在于，根据各所述问题关键字集合，确定所述问题关键字集合中的每个所述问题关键字对应的关键字重要程度，包括：

根据每个所述问题关键字对应的第二出现频率、包含有同一所述问题关键字的问题关键字集合数量，以及问题关键字集合总数量，确定所述问题关键字集合中的每个所述问题关键字对应的关键字重要程度，其中，所述第二出现频率是指出现在同一问题关键字集合的出现频率。

6.根据权利要求5所述的方法，其特征在于，根据如下公式，确定所述问题关键字集合中的每个所述问题关键字对应的关键字重要程度：

$$S_{iw}=\frac{F_{iw}\times\log(N/n_w)}{\sqrt{\sum_{m\in D_i}(F_{im}\times\log(N/n_{\mathrm{m}}))^2}}$$

其中，$S_{iw}$是第i个问题关键字集合$D_i$中的问题关键字w对应的关键字重要程度；$F_{iw}$是问题关键字w出现在问题关键字集合$D_i$的第二出现频率；N为问题关键字集合总数量；$n_w$为包含有问题关键字w的问题关键字集合数量；m为问题关键字集合$D_i$中的任一问题关键字；$F_{im}$为问题关键字m出现在问题关键字集合$D_i$的第二出现频率；$n_m$为包含有问题关键字m的问题关键字集合数量。

7.根据权利要求4所述的方法，其特征在于，基于各所述关键字重要程度确定每个待选问题与相应的候选答案之间关联度，包括：

将所述待选问题中的各个问题关键字所对应的关键字重要程度进行相加，并将相加结果确定为所述待选问题与相应的候选答案之间的关联度。

8.根据权利要求1所述的方法，其特征在于，确定所述候选问答对中的每个答案关键字与每个问题关键字之间的候选相似度，包括：

确定所述候选问答对中的每个答案关键字所对应的答案词向量以及每个问题关键字所对应的问题词向量；

基于余弦距离公式，根据每个所述答案词向量和每个所述问题词向量，确定每个所述答案关键字与每个所述问题关键字之间的候选相似度。

9.根据权利要求1所述的方法，其特征在于，所述预设匹配网络模型为文本匹配MatchPyramid模型；

相应地，将各所述候选相似度作为预设匹配网络模型的输入，根据所述预设匹配网络模型的输出确定所述候选问答对所对应的匹配结果，包括：

将各所述候选相似度组成所述候选问答对所对应的候选相似度矩阵；

将所述候选相似度矩阵作为所述MatchPyramid模型的输入，根据所述MatchPyramid模型的输出确定所述候选问答对所对应的匹配结果。

10.根据权利要求1-9任一所述的方法，其特征在于，在将所述匹配结果为匹配成功的候选问答对确定为目标问答对之后，还包括：

获取与历史时间段所对应的当前时间段产生的当前对话日志；

根据所述当前对话日志中的各个当前答案所对应的第三出现频率，确定多个当前候选答案，以及每个所述当前候选答案对应的多个当前待选问题；

根据各所述当前候选答案、各所述当前待选问题、以及所述历史时间段对应的历史候

选答案和历史待选问题，从各所述当前待选问题中确定与所述当前候选答案对应的当前候选问题，获得各个当前候选问答对；

根据所述预设匹配网络模型，确定每个所述当前候选问答对所对应的匹配结果，并将所述匹配结果为匹配成功的当前候选问答对确定为目标问答对。

11.一种问答对的提取装置，其特征在于，包括：

候选答案确定模块，用于根据历史对话日志中的各个历史答案所对应的第一出现频率，确定多个候选答案以及每个所述候选答案对应的多个待选问题；

候选问答对获得模块，用于确定每个所述待选问题与相应的候选答案之间的关联度，并基于所述关联度确定每个所述候选答案对应的候选问题，获得各个候选问答对；

候选相似度确定模块，用于确定所述候选问答对中的每个答案关键字与每个问题关键字之间的候选相似度；

目标问答对确定模块，用于将各所述候选相似度作为预设匹配网络模型的输入，根据所述预设匹配网络模型的输出确定所述候选问答对所对应的匹配结果，并将所述匹配结果为匹配成功的候选问答对确定为目标问答对。

12.一种设备，其特征在于，所述设备包括：

一个或多个处理器；

存储器，用于存储一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求1-10中任一所述的问答对的提取方法步骤。

13.一种计算机可读存储介质，其上存储有计算机程序，其特征在于，该程序被处理器执行时实现如权利要求1-11中任一所述的问答对的提取方法步骤。

# 一种问答对的提取方法、装置、设备和存储介质

## 技术领域

[0001] 本发明实施例涉及计算机技术，尤其涉及一种问答对的提取方法、装置、设备和存储介质。

## 背景技术

[0002] 在电商等服务行业中，客服需要对用户频繁提出的业务知识类问题进行解答。针对于此，通常可以构建一个问答数据库，基于该问答数据库进行自动解答，从而降低人工客服数量和成本。

[0003] 在构建问答数据库时，可以是由运营人员根据线上用户的提问总结出高质量的问答对；也可以是从历史对话数据中自动提取出一些高频问答对，以提高提取效率。

[0004] 现有技术中，提取问答对的过程为：从历史对话数据中提取出比较高频且紧密出现在一起的问答对，并通过检测提取出的问答对中是否包含相同的关键字，来确定将该问答对是否匹配，从而将匹配的问答对添加至问答数据库中，比如：问题Q：“这个免费安装吗”，答案A：“您好安装首次安装免人工费”，可见，问题和答案中均包含关键字“安”、“装”、“免”和“费”，此时可以确定该问答对相匹配，从而可以将该问答对添加至问答数据库中。

[0005] 然而，在实现本发明过程中，发明人发现现有技术中至少存在如下问题：

[0006] 由于用户的问题与相应的答案往往是不相邻的，从而利用现有的提取方式无法提取出较多的有效问答对，并且容易混入高频噪声数据，比如欢迎语。而且在真实对话场景中，具有相同语义且匹配的问题和答案并不会总出现相同的关键字，比如问题Q：“明天到货吗”，答案A：“下单会显示送达时间的”。可见，现有的问答对匹配方式无法有效合理提取出高质量的问答对，降低了提取准确性。

## 发明内容

[0007] 本发明实施例提供了一种问答对的提取方法、装置、设备和存储介质，以提取出较多高质量的问答对，提高提取的准确性。

[0008] 第一方面，本发明实施例提供了一种问答对的提取方法，包括：

[0009] 根据历史对话日志中的各个历史答案所对应的第一出现频率，确定多个候选答案以及每个所述候选答案对应的多个待选问题；

[0010] 确定每个所述待选问题与相应的候选答案之间的关联度，并基于所述关联度确定每个所述候选答案对应的候选问题，获得各个候选问答对；

[0011] 确定所述候选问答对中的每个答案关键字与每个问题关键字之间的候选相似度；

[0012] 将各所述候选相似度作为预设匹配网络模型的输入，根据所述预设匹配网络模型的输出确定所述候选问答对所对应的匹配结果，并将所述匹配结果为匹配成功的候选问答对确定为目标问答对。

[0013] 第二方面，本发明实施例还提供了一种问答对的提取装置，包括：

[0014] 候选答案确定模块，用于根据历史对话日志中的各个历史答案所对应的第一出现

频率，确定多个候选答案以及每个所述候选答案对应的多个待选问题；

[0015] 候选问答对获得模块，用于确定每个所述待选问题与相应的候选答案之间的关联度，并基于所述关联度确定每个所述候选答案对应的候选问题，获得各个候选问答对；

[0016] 候选相似度确定模块，用于确定所述候选问答对中的每个答案关键字与每个问题关键字之间的候选相似度；

[0017] 目标问答对确定模块，用于将各所述候选相似度作为预设匹配网络模型的输入，根据所述预设匹配网络模型的输出确定所述候选问答对所对应的匹配结果，并将所述匹配结果为匹配成功的候选问答对确定为目标问答对。

[0018] 第三方面，本发明实施例还提供了一种设备，所述设备包括：

[0019] 一个或多个处理器；

[0020] 存储器，用于存储一个或多个程序；

[0021] 当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如本发明任意实施例所提供的问答对的提取方法步骤。

[0022] 第四方面，本发明实施例还提供了一种计算机可读存储介质，其上存储有计算机程序该程序被处理器执行时实现如本发明任意实施例所提供的问答对的提取方法步骤。

[0023] 上述发明中的实施例具有如下优点或有益效果：

[0024] 通过根据历史对话日志中的各个历史答案所对应的第一出现频率，从历史对话日志中先提取出多个候选答案，再提取出每个候选答案对应的多个待选问题，此时提取出的待选问题可以是与候选答案相邻的问题，也可以是非相邻的问题。通过根据每个待选问题与相应的候选答案之间的关联度，从多个待选问题中筛选出每个候选答案对应的候选问题，从而可以获得较多且有效的候选问答对。通过预设匹配网络模型，对每个候选问答对进行深度语义匹配，从而可以提取出较多高质量的问答对，提高提取的准确性。

## 附图说明

[0025] 图1是本发明实施例一提供的一种问答对的提取方法的流程图；

[0026] 图2是本发明实施例一所涉及的一种候选相似度矩阵的示例；

[0027] 图3是本发明实施例一所涉及的一种MatchPyramid模型的示例；

[0028] 图4是本发明实施例二提供的一种问答对的提取方法的流程图；

[0029] 图5是本发明实施例三提供的一种问答对的提取装置的结构示意图；

[0030] 图6是本发明实施例四所涉及的一种设备的结构示意图。

## 具体实施方式

[0031] 下面结合附图和实施例对本发明作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释本发明，而非对本发明的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与本发明相关的部分而非全部结构。

[0032] 实施例一

[0033] 图1为本发明实施例一提供的一种问答对的提取方法的流程图，本实施例可适用于提取高质量的问答对的情况。该方法可以由问答对的提取装置来执行，该装置可以由软件和/或硬件的方式来实现，集成于具有数据处理功能的终端中。该方法具体包括以下步

骤：

[0034] S110、根据历史对话日志中的各个历史答案所对应的第一出现频率，确定多个候选答案以及每个候选答案对应的多个待选问题。

[0035] 其中，历史对话日志可以是指在预设历史时间段内，用户与客服之间产生的各个对话记录。历史答案可以是指历史对话日志中客服回复的答案。待选问题可以是指历史对话日志中用户提出的问题。第一出现频率可以是指历史答案在历史对话日志中出现的次数。

[0036] 具体地，通过统计历史对话日志中的每个历史答案的第一出现频率，可以将第一出现频率大于或等于预设频率的历史答案确定为候选答案，或者也可以将各个第一出现频率进行降序排列，将排列后的前预设数量的第一出现频率所对应的历史答案确定为候选答案，从而可以确定出多个高频回复的候选答案。由于候选答案是高频出现的，并且用户的提问是千变万化的，使得候选答案可以对应多个不同的待选问题，从而可以根据每个候选答案在历史对话日志中反向寻找相应的多个待选问题。本实施例中的候选答案和待选问题并不是一起提取的，使得提取出的候选答案与待选问题不一定是紧密相邻的，从而有利于提取出较多的有效问答对。

[0037] S120、确定每个待选问题与相应的候选答案之间的关联度，并基于关联度确定每个候选答案对应的候选问题，获得各个候选问答对。

[0038] 其中，关联度可以是指候选答案与待选问题之间的匹配度，即匹配度越高，关联度越高。候选问答对可以是指由一个候选答案与相应的一个候选问题组成的问答对。

[0039] 具体地，本实施例可以根据每个待选问题中的关键字与相应的候选答案中的关键字之间的匹配度，确定每个待选问题与相应的候选答案之间的关联度。基于候选答案与每个待选问题之间的关联度，可以将关联度最高的待选问题确定为该候选答案对应的候选问题，生成一个候选问答对，从而可以获得多个有效的候选问答对。

[0040] S130、确定候选问答对中的每个答案关键字与每个问题关键字之间的候选相似度。

[0041] 其中，答案关键字可以是指候选答案中的一个有效的字或者词。问题关键字可以是指候选问题中的一个有效的字或者词。候选相似度可以是指在一个候选问答对中，每个答案关键字与每个问题关键字之间的相似程度。

[0042] 示例性地，S130可以包括：确定候选问答对中的每个答案关键字所对应的答案词向量以及每个问题关键字所对应的问题词向量；基于余弦距离公式，根据每个答案词向量和每个问题词向量，确定每个答案关键字与每个问题关键字之间的候选相似度。

[0043] 其中，词向量可以基于预设词库进行确定。词向量中元素的维度与预设词库中的词汇数量相同，并且词向量中的每位元素均可以按照预定顺序对应于预设词库中的每个词。答案词向量中的每位元素对应的元素值可以用于表征该元素对应的预设词库中的词作为答案关键字的上下文词的概率，从而可以根据预设词库中的每个词作为答案关键字的上下文词的概率来表示答案关键字所对应的答案词向量。同理，问题词向量中的每位元素对应的元素值可以用于表征该元素对应的预设词库中的词作为问题关键字的上下文词的概率，从而可以根据预设词库中的每个词作为问题关键字的上下文词的概率来表示问题关键字所对应的问题词向量。

[0044] 具体地，本实施例可以对候选问答对中的候选问题和候选答案分别进行分词处理，确定出各个问题关键字和各个答案关键字。分词处理可以包括但不限于中文分词、识别专有名词以及去停用词。通过对候选问题和候选答案进行分词处理，并剔除没有语义的停用词，比如："啊"、"吧"等，从而可以快速获取各个问题词向量和各个答案词向量，提高了词向量的生成效率。示例性地，本实施例可以基于如下的余弦距离公式，确定每个答案关键字与每个问题关键字之间的候选相似度：

[0045] $$\mathrm{M_{ij}} = \frac{A_i^T Q_j}{\|A_i\| . \|Q_j\|}$$

[0046] 其中，$M_{ij}$是指第i个答案关键字与第j个问题关键字之间的候选相似度；$A_i$是指第i个答案关键字对应的答案词向量；$Q_j$是第j个问题关键字对应的问题词向量。通过利用答案词向量与问题词向量的方式，确定答案关键字与问题关键字之间的候选相似度，可以衡量词的上下文重要性，避免发生语义偏移，以便提高语义匹配的准确性。

[0047] S140、将各候选相似度作为预设匹配网络模型的输入，根据预设匹配网络模型的输出确定候选问答对所对应的匹配结果，并将匹配结果为匹配成功的候选问答对确定为目标问答对。

[0048] 其中，预设匹配网络模型可以是指用于确定候选问答对是否匹配的神经网络模型。预设匹配网络模型可以为但不限于文本匹配MatchPyramid模型。匹配结果可以包括两种，分别为匹配成功和匹配失败。

[0049] 具体地，本实施例在使用预设匹配网络模型之前，还包括对预设匹配网络模型的训练过程。将多个匹配的问答对样本所对应的候选相似度以及多个不匹配的问答对样本(可以是指随机产生的问答对)所对应的候选相似度作为训练样本，对预设匹配网络模型进行训练。基于损失函数，根据预设匹配网络模型的训练匹配结果与标准匹配结果，计算每个问答对样本对应的训练误差，当训练误差大于或等于预设误差时，调整预设匹配网络模型的网络参数；当训练误差小于预设误差时，或者迭代次数等于预设次数时，可以确定预设匹配网络模型训练结束。本实施例可以将每个候选问答对对应的各候选相似度输入至训练结束后的预设回归网络模型中，根据预设回归网络模型的输出确定每个候选问答对所对应的匹配结果，从而可以基于预设回归网络模型实现对候选问答对是否匹配的自动判断，提高匹配效率。通过将匹配结果为匹配成功的候选问答对确定为目标问答对，以便可以将高质量的目标问答对添加至问答数据库中。

[0050] 本实施例的技术方案，通过根据历史对话日志中的各个历史答案所对应的第一出现频率，从历史对话日志中先提取出多个候选答案，再提取出每个候选答案对应的多个待选问题，此时提取出的待选问题可以是与候选答案相邻的问题，也可以是非相邻的问题。通过根据每个待选问题与相应的候选答案之间的关联度，从多个待选问题中筛选出每个候选答案对应的候选问题，从而可以获得较多且有效的候选问答对。通过预设匹配网络模型，对每个候选问答对进行深度语义匹配，从而可以提取出较多高质量的问答对，提高提取的准确性。

[0051] 在上述技术方案的基础上，S110可以包括：基于预设过滤规则，对历史对话日志中的各个历史答案进行过滤；统计过滤出的历史答案所对应的第一出现频率，将第一出现频

率大于或等于预设频率的历史答案作为候选答案，小于预设频率的历史答案作为第一答案；确定每个第一答案与每个候选答案之间的答案相似度，将答案相似度大于或等于预设相似度的第一答案也作为候选答案；根据历史对话日志中的各个对话记录，提取在每个对话记录中位于候选答案之前的用户问题，并根据从各对话记录中提取出的用户问题确定候选答案对应的多个待选问题。

[0052] 其中，预设过滤规则可以是用于过滤历史答案中的无意义答案的规则。无意义答案可以是指但不限于欢迎语、客套话和活动宣传语。本实施例中的候选答案为高频答案，使得候选答案所在的对话记录可以为一个或多个。

[0053] 具体地，通过基于预设过滤规则，将历史对话日志中的无意义的历史答案进行过滤，使得过滤后获得的历史答案均为针对于用户问题所回复的答案，从而可以避免混入高频噪声。通过在过滤后获得的所有历史答案中，将第一出现频率大于或等于预设频率的历史答案直接作为候选答案，以便获得高频回复的历史答案。同时将第一出现频率小于预设频率的历史答案作为第一答案，并可以基于编辑距离或者最长公共字串的占比来确定每个第一答案与每个候选答案之间的答案相似度。通过将答案相似度大于或等于预设相似度的第一答案也作为候选答案，从而也可以将第一出现频率较低，但与高频答案相似的第一答案也作为候选答案，从而可以避免长尾问答对的遗漏，并且也可以满足客服的多样性需求，比如某个高频的候选答案为：“包安装”；某个低频且与该候选答案相似的第一答案为：“亲，包安装哦”，通过将该第一答案也作为候选答案使得客服可以选取自身喜好的答案回复用户。在确定出每个候选答案之后，可以从历史对话日志中的各个对话记录中，提取出每个候选答案所在的对话记录，并根据对话记录中的位于候选答案之前的用户问题，确定出每个候选答案对应的多个待选问题，比如可以将对话记录中的位于候选答案之前的用户问题均作为该候选答案对应的待选问题。

[0054] 示例性地，根据从各对话记录中提取出的用户问题确定候选答案对应的多个待选问题，可以包括：若从对话记录中提取出的用户问题为多个，则获得距离候选答案最近的第一用户问题；检测是否存在与第一用户问题连续提问的第二用户问题；若是，则将第一用户问题和第二用户问题均作为候选答案对应的待选问题；若否，则将第一用户问题作为候选答案对应的待选问题。

[0055] 具体地，对于每个候选答案所在的每个对话记录而言，若在对话记录中位于该候选答案之前的用户问题仅存在一个，则可以直接将该用户问题确定为该候选答案对应的待选问题。若在对话记录中位于该候选答案之前的用户问题为多个，则将位于该候选答案之前且距离该候选答案最近的用户问题作为第一用户问题，并通过检测在对话记录中是否存在与第一用户问题连续提问的第二用户问题，若存在，则可以将第一用户问题和第二用户问题均作为该候选答案对应的待选问题，其中第二用户问题可以为一个或多个；若不存在，则可以将第一用户问题作为该候选答案对应的待选问题。

[0056] 示例性地，某个对话记录如下：

[0057] 问题$Q_1$：你好

[0058] 答案$A_1$：在的哦

[0059] 问题$Q_2$：保修多长时间

[0060] 问题$Q_3$：使用一段时间有问题，我找谁解决呀

[0061] 答案$A_4$:亲爱哒

[0062] 答案$A_5$:本产品全国联保,享受三包服务,保修期为:整机一年质保,主要零部件三年质保

[0063] 在上述的对话记录中,若答案$A_5$为候选答案,则距离候选答案$A_5$最近的第一用户问题为问题$Q_3$,并且存在与问题$Q_3$连续提问的第二用户问题$Q_2$,从而可以将问题$Q_2$和问题$Q_3$均作为候选答案$A_5$的待选问题。

[0064] 在上述技术方案的基础上,当预设匹配网络模型为文本匹配MatchPyramid模型时,S140中的“将各候选相似度作为预设匹配网络模型的输入,根据预设匹配网络模型的输出确定候选问答对所对应的匹配结果”,可以包括:将各候选相似度组成候选问答对所对应的候选相似度矩阵;将候选相似度矩阵作为MatchPyramid模型的输入,根据MatchPyramid模型的输出确定候选问答对所对应的匹配结果。

[0065] 其中,候选相似度矩阵可以是由候选问答对中的每个答案关键字与每个问题关键字之间的候选相似度组成的矩阵。例如,在候选问答对中的候选问题为:“请问安装收费吗”,候选答案为:“首次安装免人工费”时,图2给出了一种候选相似度矩阵的示例,如图2所示,候选相似度矩阵中的每个行向量表示一个问题关键字与每个答案关键字之间的候选相似度;每个列向量标识一个答案关键字与每个问题关键字之间的候选相似度。

[0066] 具体地,通过将候选问答对中的各个候选相似度矩阵组成一个候选相似度矩阵后,可以将该候选相似度矩阵作为一个二维的灰度图像,输入至MatchPyramid模型中,从而可以将问答对匹配的问题转换为一个图像分类的问题,并且根据MatchPyramid模型的输出可以确定出候选问答对所对应的匹配结果。图3给出了一种MatchPyramid模型的示例。如图3所示,MatchPyramid模型可以包含一层CNN(Convolutional Neural Network,卷积神经网络)、全连接层和一个输出层,其中CNN网络包含卷积层和池化层;全连接层用于对CNN层输出的低维向量x进行非线性转换,并且通过softmax归一化指数函数获得候选问答对的匹配程度。输出层中的$S_0$和$S_1$可以分别表示正负类得分,其中正类表示候选问答对匹配;负类表示候选问答对不匹配。通过将候选相似度矩阵输入至MatchPyramid模型中,可以根据MatchPyramid模型输出的正类得分确定候选问答对所对应的匹配结果,比如当正类得分大于0.5时,确定候选问答对所对应的匹配结果为匹配成功,否则确定匹配结果为匹配失败,从而可以对候选问答对是否匹配进行自动判断,大大提高了判断效率。

[0067] 实施例二

[0068] 图4为本发明实施例二提供的一种问答对的提取方法的流程图,本实施例在上述实施例的基础上,对“确定每个待选问题与相应的候选答案之间的关联度,并基于关联度确定每个候选答案对应的候选问题,获得各个候选问答对”进行优化。其中与上述各实施例相同或相应的术语的解释在此不再赘述。

[0069] 参见图4,本实施例提供的问答对的提取方法包括以下步骤:

[0070] S210、根据历史对话日志中的各个历史答案所对应的第一出现频率,确定多个候选答案以及每个候选答案对应的多个待选问题。

[0071] S220、对各候选答案进行聚类,确定各个候选答案集合,并根据候选答案与待选问题之间的对应关系,确定每个候选答案集合对应的待选问题集合。

[0072] 其中,候选答案集合可以由两个或两个以上的相似度大于预设相似度阈值的候选

答案组成。待选问题集合可以是由候选答案集合中的每个候选答案对应的多个待选问题组成。

[0073] 具体地，本实施例可以基于现有的聚类方式，对各候选答案进行聚类，以将相似的候选答案聚类到一起，获得多个候选答案集合，以便后续可以获取较多的问题关键字，提高关联度确定的准确性。针对每个候选答案集合，获得候选答案集合中的每个候选答案所对应的多个待选问题，并将获得的这些待选问题进行组合，确定该候选答案集合对应的待选问题集合。

[0074] S230、对待选问题集合中的各个待选问题进行问题关键字提取，确定每个待选问题集合对应的问题关键字集合。

[0075] 其中，问题关键字可以是待选问题中的一个字或者一个词。

[0076] 具体地，对于每个待选问题集合而言，可以将待选问题集合中的所有待选问题进行分词处理，获得各个待选问题中的每个问题关键字，便将所有的问题关键字进行组合，确定该待选问题集合对应的问题关键字集合。

[0077] S240、根据各问题关键字集合，确定问题关键字集合中的每个问题关键字对应的关键字重要程度，并基于各关键字重要程度确定每个待选问题与相应的候选答案之间关联度。

[0078] 其中，问题关键字对应的关键字重要程度可以是指问题关键字在问题关键字集合中的重要程度，其可以反映出问题关键字对于候选答案的重要程度。若问题关键字重要程度越高，则表示该问题关键字是候选答案的标准问题中的关键字的可能性也越高。

[0079] 具体地，本实施例可以基于TF-IDF(Term Frequency-Inverse Document Frequency，词频-逆文本频率指数)方式，根据每个问题关键字集合中的各个问题关键字，可以确定出每个问题关键字对应的关键字重要程度。根据每个待选问题中的各个问题关键字对应的关键字重要程度，可以计算出该待选问题对应的问题重要程度，并可以将该问题重要程度作为该待选问题与相应的候选答案之间的关联度。同理，可以确定出每个待选问题与相应的候选答案之间的关联度。

[0080] 示例性地，S240中的“根据各问题关键字集合，确定问题关键字集合中的每个问题关键字对应的关键字重要程度”，可以包括：根据每个问题关键字对应的第二出现频率、包含有同一问题关键字的问题关键字集合数量，以及问题关键字集合总数量，确定问题关键字集合中的每个问题关键字对应的关键字重要程度。

[0081] 其中，第二出现频率可以是指出现在同一问题关键字集合中的出现频率。本实施例可以通过统计同一问题关键字集合中的每个问题关键字的出现频率，作为第二出现频率。示例性地，可以根据如下公式，确定问题关键字集合中的每个问题关键字对应的关键字重要程度：

[0082] $$S_{iw}=\frac{F_{iw}\times\log(N/n_w)}{\sqrt{\sum_{m\in D_i}(F_{im}\times\log(N/n_{\mathrm{m}}))^2}}$$

[0083] 其中，$S_{iw}$是第i个问题关键字集合$D_i$中的问题关键字w对应的关键字重要程度；$F_{iw}$是问题关键字w出现在问题关键字集合$D_i$的第二出现频率；N为问题关键字集合总数量；$n_w$为包含有问题关键字w的问题关键字集合数量；m为问题关键字集合$D_i$中的任一问题关键字；

$F_{im}$为问题关键字m出现在问题关键字集合$D_i$的第二出现频率；$n_m$为包含有问题关键字m的问题关键字集合数量。具体地，本实施例可以对基于上述公式，对每个问题关键字对应的关键字重要程度进行归一化处理，以便可以准确地反映出每个问题关键词对应的关键词重要程度。

[0084] 示例性地，S240中的“基于各关键字重要程度确定每个待选问题与相应的候选答案之间关联度”，可以包括：将待选问题中的各个问题关键字所对应的关键字重要程度进行相加，并将相加结果确定为待选问题与相应的候选答案之间的关联度。

[0085] 具体地，对于每个待选问题而言，可以将该待选问题中的各个问题关键字对应的关键字重要程度进行相加，获得的相加结果可以作为该待选问题与相应的候选答案之间的关联度。

[0086] S250、将关联度最高的待选问题确定为相应的候选答案所对应的候选问题，获得候选问答对。

[0087] 具体地，对于每个候选答案对应的多个待选问题，可以基于每个待选问题对应的关联度，将关联度最高的待选问题作为该候选答案对应的候选问题，从而可以获得多个有效的候选问答对。

[0088] S260、确定候选问答对中的每个答案关键字与每个问题关键字之间的候选相似度。

[0089] S270、将各候选相似度作为预设匹配网络模型的输入，根据预设匹配网络模型的输出确定候选问答对所对应的匹配结果，并将匹配结果为匹配成功的候选问答对确定为目标问答对。

[0090] 本实施例的技术方案，通过对候选答案进行聚类确定出各个候选答案集合以及每个候选答案集合对应的待选问题集合，并对待选问题集合中的各个待选问题进行问题关键字提取，确定出每个问题关键字对应的关键字重要程度；基于各关键字重要程度确定每个待选问题与相应的候选答案之间关联度，并将关联度最高的待选问题确定为相应的候选答案所对应的候选问题，从而可以获得更加有效准确的候选问答对，进一步提高问答对提取的准确性。

[0091] 在上述技术方案的基础上，在S270之后，还包括：获取与历史时间段所对应的当前时间段产生的当前对话日志；根据当前对话日志中的各个当前答案所对应的第三出现频率，确定多个当前候选答案，以及每个当前候选答案对应的多个当前待选问题；根据各当前候选答案、各当前待选问题、以及历史时间段对应的历史候选答案和历史待选问题，从各当前待选问题中确定与当前候选答案对应的当前候选问题，获得各个当前候选问答对；根据预设匹配网络模型，确定每个当前候选问答对所对应的匹配结果，并将匹配结果为匹配成功的当前候选问答对确定为目标问答对。

[0092] 其中，当前对话日志可以是指当前时间段产生的各个对话记录，比如当天时间段内产生的当前对话日志。第三出现频率可以是指当前对话日志中的每个当前答案在当前对话日志中出现的次数。

[0093] 具体地，本实施例可以基于预设过滤规则，过滤掉当前对话日志中的无意义的当前答案，比如欢迎语、礼貌性回复等。根据过滤后获得的各个当前答案所对应的第三出现频率，可以将第三出现频率高于预设频率的当前答案作为当前候选答案。并可以基于与提取

历史待选问题相似的方式，从当前对话日志中提取出每个当前候选答案对应的多个当前待选问题。本实施例可以对当前候选答案和历史候选答案进行聚类，确定出各个第一候选答案集合，并根据历史候选答案与历史待选问题之间的对应关系以及当前候选答案与当前待选问题之间的对应关系，确定出每个第一候选答案集合对应的第一待选问题集合；对第一待选问题集合中的各个第一待选问题进行问题关键字提取，确定每个第一待选问题集合对应的第一问题关键字集合，其中第一待选问题可以为历史待选问题或者当前待选问题。基于与上述步骤S240的实现方式，可以根据各第一问题关键字集合，确定出第一问题关键字集合中的每个问题关键字对应的关键字重要程度，并可以将第一待选问题为当前待选问题中的各个当前问题关键字对应的关键字重要程度进行相加，获得的相加结果作为当前待选问题与相应的当前候选答案之间的关联度，并将关联度最高的当前待选问题作为相应的当前候选答案对应的当前候选问题，从而可以确定出多个当前候选问答对。通过将每个当前候选问答对中的每个当前答案关键字与每个当前问题关键字之间的候选相似度作为预设匹配网络模型的输入，根据预设匹配网络模型的输出可以确定出每个当前候选问答对所对应的匹配结果，并将匹配结果为匹配成功的当前候选问答对确定为目标问答对，从而可以将匹配成功的高质量的当前候选问答对实时添加至问答数据库中，实现了对短时间内的对话日志进行问答对的挖掘，以实时更新问答库，保证了问答对的时效性。

[0094] 以下是本发明实施例提供的问答对的提取装置的实施例，该装置与上述各实施例的问答对的提取方法属于同一个发明构思，在问答对的提取装置的实施例中未详尽描述的细节内容，可以参考上述问答对的提取方法的实施例。

[0095] 实施例三

[0096] 图5为本发明实施例三提供的一种问答对的提取装置的结构示意图，本实施例可适用于提取高质量的问答对的情况，该装置具体包括：候选答案确定模块310、候选问答对获得模块320、候选相似度确定模块330和目标问答对确定模块340。

[0097] 其中，候选答案确定模块310，用于根据历史对话日志中的各个历史答案所对应的第一出现频率，确定多个候选答案以及每个候选答案对应的多个待选问题；候选问答对获得模块320，用于确定每个待选问题与相应的候选答案之间的关联度，并基于关联度确定每个候选答案对应的候选问题，获得各个候选问答对；候选相似度确定模块330，用于确定候选问答对中的每个答案关键字与每个问题关键字之间的候选相似度；目标问答对确定模块340，用于将各候选相似度作为预设匹配网络模型的输入，根据预设匹配网络模型的输出确定候选问答对所对应的匹配结果，并将匹配结果为匹配成功的候选问答对确定为目标问答对。

[0098] 可选地，候选答案确定模块310，包括：

[0099] 历史答案过滤单元，用于基于预设过滤规则，对历史对话日志中的各个历史答案进行过滤；

[0100] 第一候选答案确定单元，用于统计过滤出的历史答案所对应的第一出现频率，将第一出现频率大于或等于预设频率的历史答案作为候选答案，小于预设频率的历史答案作为第一答案；

[0101] 第二候选答案确定单元，用于确定每个第一答案与每个候选答案之间的答案相似度，将答案相似度大于或等于预设相似度的第一答案也作为候选答案；

[0102] 待选问题确定单元，用于根据历史对话日志中的各个对话记录，提取在每个对话记录中位于候选答案之前的用户问题，并根据从各对话记录中提取出的用户问题确定候选答案对应的多个待选问题。

[0103] 可选地，待选问题确定单元，具体用于：若从对话记录中提取出的用户问题为多个，则获得距离候选答案最近的第一用户问题；检测是否存在与第一用户问题连续提问的第二用户问题；若是，则将第一用户问题和第二用户问题均作为候选答案对应的待选问题；若否，则将第一用户问题作为候选答案对应的待选问题。

[0104] 可选地，候选问答对获得模块320，包括：

[0105] 待选问题集合确定单元，用于对各候选答案进行聚类，确定各个候选答案集合，并根据候选答案与待选问题之间的对应关系，确定每个候选答案集合对应的待选问题集合；

[0106] 问题关键字集合确定单元，用于对待选问题集合中的各个待选问题进行问题关键字提取，确定每个待选问题集合对应的问题关键字集合；

[0107] 关联度确定单元，用于根据各问题关键字集合，确定问题关键字集合中的每个问题关键字对应的关键字重要程度，并基于各关键字重要程度确定每个待选问题与相应的候选答案之间关联度；

[0108] 候选问答对确定单元，用于将关联度最高的待选问题确定为相应的候选答案所对应的候选问题，获得候选问答对。

[0109] 可选地，关联度确定单元，包括：关键字重要程度确定子单元，用于根据每个问题关键字对应的第二出现频率、包含有同一问题关键字的问题关键字集合数量，以及问题关键字集合总数量，确定问题关键字集合中的每个问题关键字对应的关键字重要程度，其中，第二出现频率是指出现在同一问题关键字集合的出现频率。

[0110] 可选地，根据如下公式，确定问题关键字集合中的每个问题关键字对应的关键字重要程度：

[0111] $$S_{iw}=\frac{F_{iw}\times\log(N/n_w)}{\sqrt{\sum_{m\in D_i}(F_{im}\times\log(N/n_m))^2}}$$

[0112] 其中，$S_{iw}$是第i个问题关键字集合$D_i$中的问题关键字w对应的关键字重要程度；$F_{iw}$是问题关键字w出现在问题关键字集合$D_i$的第二出现频率；N为问题关键字集合总数量；$n_w$为包含有问题关键字w的问题关键字集合数量；m为问题关键字集合$D_i$中的任一问题关键字；$F_{im}$为问题关键字m出现在问题关键字集合$D_i$的第二出现频率；$n_m$为包含有问题关键字m的问题关键字集合数量。

[0113] 可选地，关联度确定单元，还包括：关联度确定子单元，用于将待选问题中的各个问题关键字所对应的关键字重要程度进行相加，并将相加结果确定为待选问题与相应的候选答案之间的关联度。

[0114] 可选地，候选相似度确定模块330，具体用于：确定候选问答对中的每个答案关键字所对应的答案词向量以及每个问题关键字所对应的问题词向量；基于余弦距离公式，根据每个答案词向量和每个问题词向量，确定每个答案关键字与每个问题关键字之间的候选相似度。

[0115] 可选地，预设匹配网络模型为文本匹配MatchPyramid模型；相应地，目标问答对确

定模块340还包括:匹配结果确定单元,用于将各候选相似度组成候选问答对所对应的候选相似度矩阵;将候选相似度矩阵作为MatchPyramid模型的输入,根据MatchPyramid模型的输出确定候选问答对所对应的匹配结果。

[0116] 可选地,该装置还包括:

[0117] 当前对话日志获取模块,用于在将匹配结果为匹配成功的候选问答对确定为目标问答对之后,获取与历史时间段所对应的当前时间段产生的当前对话日志;

[0118] 当前候选答案确定模块,用于根据当前对话日志中的各个当前答案所对应的第三出现频率,确定多个当前候选答案,以及每个当前候选答案对应的多个当前待选问题;

[0119] 当前候选问答对确定模块,用于根据各当前候选答案、各当前待选问题、以及历史时间段对应的历史候选答案和历史待选问题,从各当前待选问题中确定与当前候选答案对应的当前候选问题,获得各个当前候选问答对;

[0120] 目标问答对确定模块340,还用于根据预设匹配网络模型,确定每个当前候选问答对所对应的匹配结果,并将匹配结果为匹配成功的当前候选问答对确定为目标问答对。

[0121] 本发明实施例所提供的问答对的提取装置可执行本发明任意实施例所提供的问答对的提取方法,具备执行问答对的提取方法相应的功能模块和有益效果。

[0122] 实施例四

[0123] 图6为本发明实施例四提供的一种设备的结构示意图。图6示出了适于用来实现本发明实施方式的示例性设备12的框图。图6显示的设备12仅仅是一个示例,不应对本发明实施例的功能和使用范围带来任何限制。

[0124] 如图6所示,设备12以通用计算设备的形式表现。设备12的组件可以包括但不限于:一个或者多个处理器或者处理单元16,系统存储器28,连接不同系统组件(包括系统存储器28和处理单元16)的总线18。

[0125] 总线18表示几类总线结构中的一种或多种,包括存储器总线或者存储器控制器,外围总线,图形加速端口,处理器或者使用多种总线结构中的任意总线结构的局域总线。举例来说,这些体系结构包括但不限于工业标准体系结构(ISA)总线,微通道体系结构(MAC)总线,增强型ISA总线、视频电子标准协会(VESA)局域总线以及外围组件互连(PCI)总线。

[0126] 设备12典型地包括多种计算机系统可读介质。这些介质可以是任何能够被设备12访问的可用介质,包括易失性和非易失性介质,可移动的和不可移动的介质。

[0127] 系统存储器28可以包括易失性存储器形式的计算机系统可读介质,例如随机存取存储器(RAM)30和/或高速缓存存储器32。设备12可以进一步包括其它可移动/不可移动的、易失性/非易失性计算机系统存储介质。仅作为举例,存储系统34可以用于读写不可移动的、非易失性磁介质(图6未显示,通常称为“硬盘驱动器”)。尽管图6中未示出,可以提供用于对可移动非易失性磁盘(例如“软盘”)读写的磁盘驱动器,以及对可移动非易失性光盘(例如CD-ROM,DVD-ROM或者其它光介质)读写的光盘驱动器。在这些情况下,每个驱动器可以通过一个或者多个数据介质接口与总线18相连。系统存储器28可以包括至少一个程序产品,该程序产品具有一组(例如至少一个)程序模块,这些程序模块被配置以执行本发明各实施例的功能。

[0128] 具有一组(至少一个)程序模块42的程序/实用工具40,可以存储在例如系统存储器28中,这样的程序模块42包括但不限于操作系统、一个或者多个应用程序、其它程序模块

以及程序数据，这些示例中的每一个或某种组合中可能包括网络环境的实现。程序模块42通常执行本发明所描述的实施例中的功能和/或方法。

[0129] 设备12也可以与一个或多个外部设备14(例如键盘、指向设备、显示器24等)通信，还可与一个或者多个使得用户能与该设备12交互的设备通信，和/或与使得该设备12能与一个或多个其它计算设备进行通信的任何设备(例如网卡，调制解调器等等)通信。这种通信可以通过输入/输出(I/O)接口22进行。并且，设备12还可以通过网络适配器20与一个或者多个网络(例如局域网(LAN)，广域网(WAN)和/或公共网络，例如因特网)通信。如图所示，网络适配器20通过总线18与设备12的其它模块通信。应当明白，尽管图中未示出，可以结合设备12使用其它硬件和/或软件模块，包括但不限于：微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、RAID系统、磁带驱动器以及数据备份存储系统等。

[0130] 处理单元16通过运行存储在系统存储器28中的程序，从而执行各种功能应用以及数据处理，例如实现本发实施例所提供的一种问答对的提取方法步骤，该方法包括：

[0131] 根据历史对话日志中的各个历史答案所对应的第一出现频率，确定多个候选答案以及每个候选答案对应的多个待选问题；

[0132] 确定每个待选问题与相应的候选答案之间的关联度，并基于关联度确定每个候选答案对应的候选问题，获得各个候选问答对；

[0133] 确定候选问答对中的每个答案关键字与每个问题关键字之间的候选相似度；

[0134] 将各候选相似度作为预设匹配网络模型的输入，根据预设匹配网络模型的输出确定候选问答对所对应的匹配结果，并将匹配结果为匹配成功的候选问答对确定为目标问答对。

[0135] 当然，本领域技术人员可以理解，处理器还可以实现本发明任意实施例所提供的问答对的提取方法的技术方案。

[0136] 实施例五

[0137] 本实施例五提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如本发明任意实施例所提供的问答对的提取方法步骤，该方法包括：

[0138] 根据历史对话日志中的各个历史答案所对应的第一出现频率，确定多个候选答案以及每个候选答案对应的多个待选问题；

[0139] 确定每个待选问题与相应的候选答案之间的关联度，并基于关联度确定每个候选答案对应的候选问题，获得各个候选问答对；

[0140] 确定候选问答对中的每个答案关键字与每个问题关键字之间的候选相似度；

[0141] 将各候选相似度作为预设匹配网络模型的输入，根据预设匹配网络模型的输出确定候选问答对所对应的匹配结果，并将匹配结果为匹配成功的候选问答对确定为目标问答对。

[0142] 本发明实施例的计算机存储介质，可以采用一个或多个计算机可读的介质的任意组合。计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质。计算机可读存储介质例如可以是但不限于：电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读存储介质的更具体的例子(非穷举的列表)包括：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光

存储器件、磁存储器件、或者上述的任意合适的组合。在本文件中，计算机可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

[0143] 计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质，该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。

[0144] 计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括但不限于：无线、电线、光缆、RF等等，或者上述的任意合适的组合。

[0145] 可以以一种或多种程序设计语言或其组合来编写用于执行本发明操作的计算机程序代码，所述程序设计语言包括面向对象的程序设计语言，诸如Java、Smalltalk、C++，还包括常规的过程式程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络，包括局域网(LAN)或广域网(WAN)，连接到用户计算机，或者，可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。

[0146] 本领域普通技术人员应该明白，上述的本发明的各模块或各步骤可以用通用的计算装置来实现，它们可以集中在单个计算装置上，或者分布在多个计算装置所组成的网络上，可选地，他们可以用计算机装置可执行的程序代码来实现，从而可以将它们存储在存储装置中由计算装置来执行，或者将它们分别制作成各个集成电路模块，或者将它们中的多个模块或步骤制作成单个集成电路模块来实现。这样，本发明不限制于任何特定的硬件和软件的结合。

[0147] 注意，上述仅为本发明的较佳实施例及所运用技术原理。本领域技术人员会理解，本发明不限于这里所述的特定实施例，对本领域技术人员来说能够进行各种明显的变化、重新调整和替代而不会脱离本发明的保护范围。因此，虽然通过以上实施例对本发明进行了较为详细的说明，但是本发明不仅仅限于以上实施例，在不脱离本发明构思的情况下，还可以包括更多其他等效实施例，而本发明的范围由所附的权利要求范围决定。

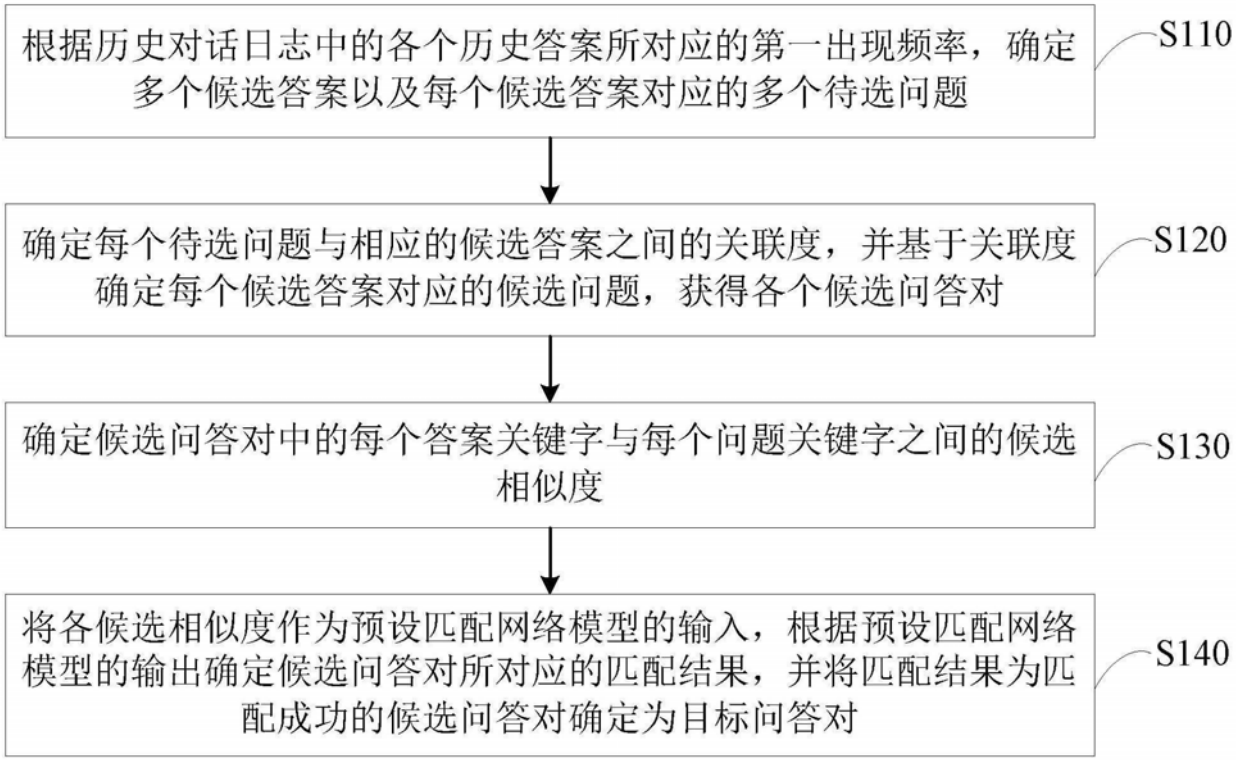



图1

| | 首次 | 安装 | 免收 | 人工 | 费 |
|---|---|---|---|---|---|
| 请问 | 0.1 | 0.23 | 0.16 | 0.24 | 0.31 |
| 安装 | 0.34 | 1.0 | 0.72 | 0.65 | 0.82 |
| 收费 | 0.12 | 0.87 | 0.62 | 0.71 | 0.95 |

图2

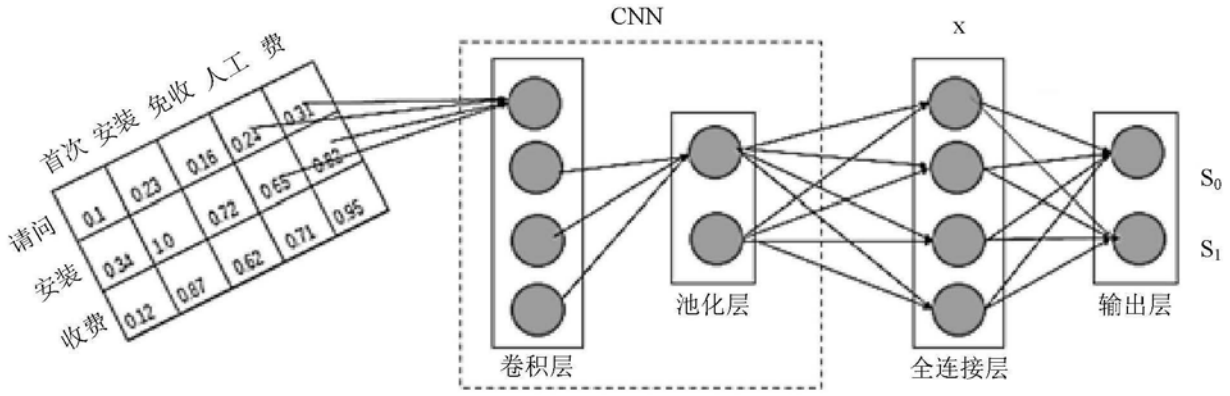

首次 安装 免收 人工 费
请问
安装
收费
01
0.23
0.16
0.23
0.31
0.34
1.0
0.72
0.65
0.82
0.12
0.87
0.62
0.71
0.95
CNN
卷积层
池化层
x
全连接层
输出层
S0
S1

图3

根据历史对话日志中的各个历史答案所对应的第一出现频率，确定多个候选答案以及每个候选答案对应的多个待选问题
S210
对各候选答案进行聚类，确定各个候选答案集合，并根据候选答案与待选问题之间的对应关系，确定每个候选答案集合对应的待选问题集合
S220
对待选问题集合中的各个待选问题进行问题关键字提取，确定每个待选问题集合对应的问题关键字集合
S230
根据各问题关键字集合，确定问题关键字集合中的每个问题关键字对应的关键字重要程度，并基于各关键字重要程度确定每个待选问题与相应的候选答案之间关联度
S240
将关联度最高的待选问题确定为相应的候选答案所对应的候选问题，获得候选问答对
S250
确定候选问答对中的每个答案关键字与每个问题关键字之间的候选相似度
S260
将各候选相似度作为预设匹配网络模型的输入，根据预设匹配网络模型的输出确定候选问答对所对应的匹配结果，并将匹配结果为匹配成功的候选问答对确定为目标问答对
S270

图4

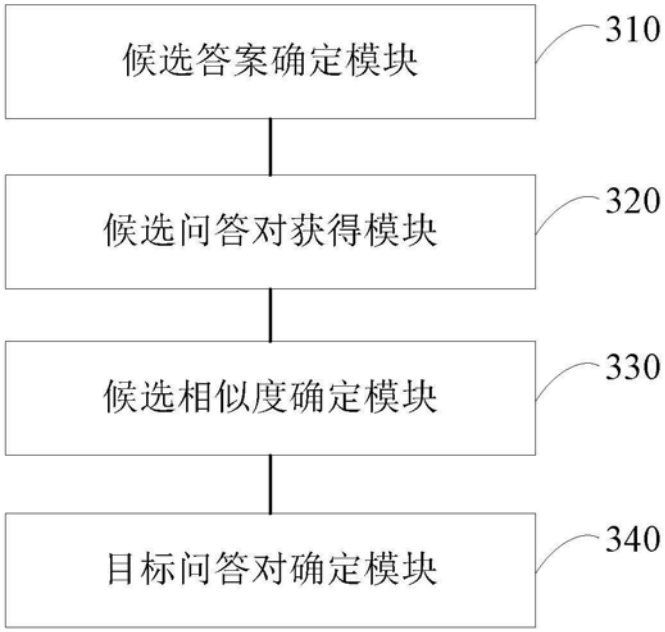



图5

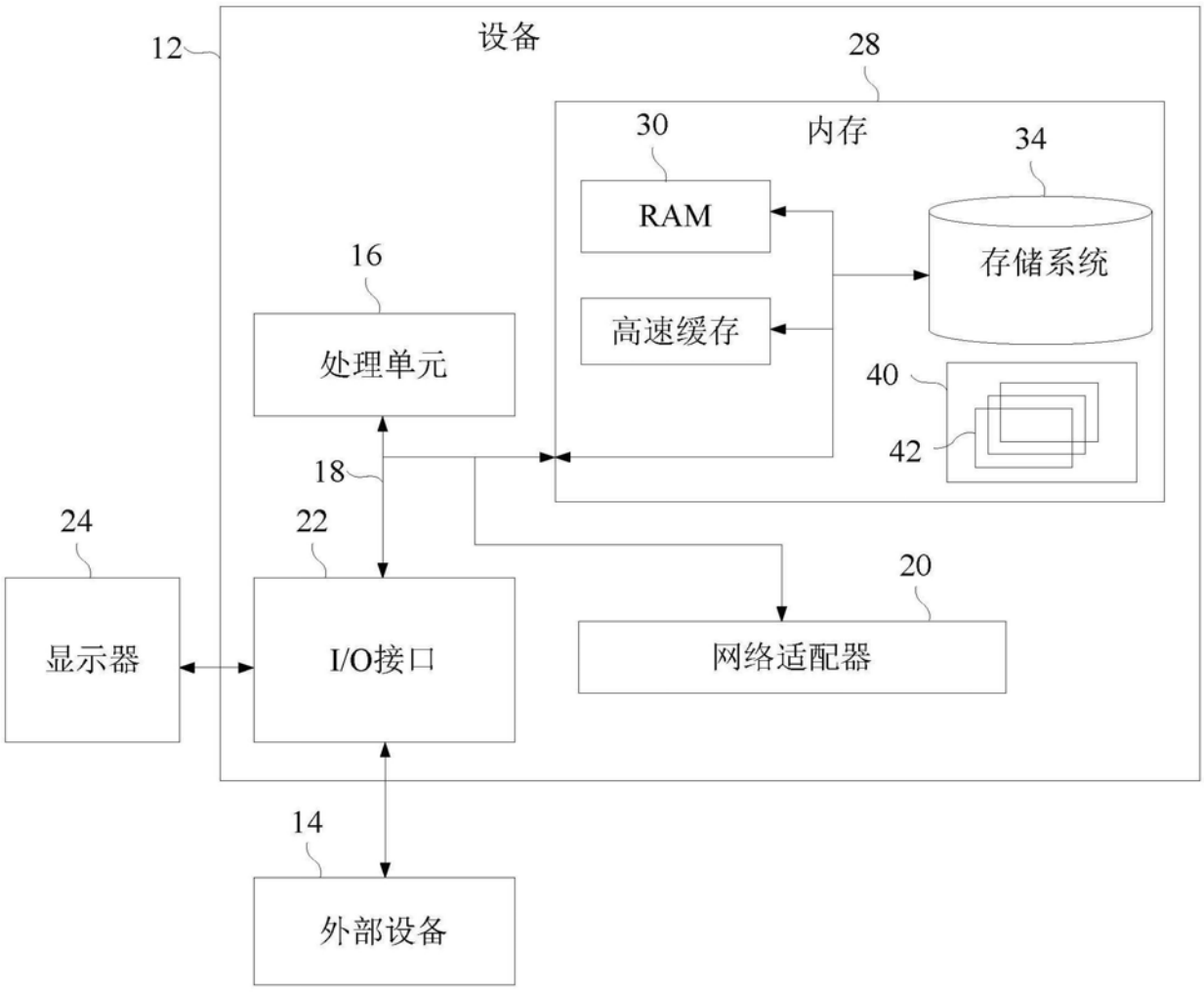

12
设备
28
30
内存
34
RAM
存储系统
16
高速缓存
处理单元
40
42
18
24
22
20
显示器
I/O接口
网络适配器
14
外部设备

图6