

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 11 月 26 日 (26.11.2020)



(10) 国际公布号
WO 2020/232877 A1

(51) 国际专利分类号:
G06F 16/332 (2019.01) **G06F 16/35** (2019.01)

(21) 国际申请号: PCT/CN2019/102923

(22) 国际申请日: 2019 年 8 月 28 日 (28.08.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910421574.9 2019 年 5 月 21 日 (21.05.2019) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(72) 发明人: 王健宗(WANG, Jianzong); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。
程宁(CHENG, Ning); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳众鼎专利商标代理事务所(普通合伙)(SHENZHEN ZHONGDING INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市罗湖区笋岗东路 3012 号中民时代广场 B 座 7 楼 701 室, Guangdong 518000 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,

(54) **Title:** QUESTION ANSWER SELECTION METHOD AND APPARATUS, COMPUTER DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 一种问题答案选取方法、装置、计算机设备及存储介质

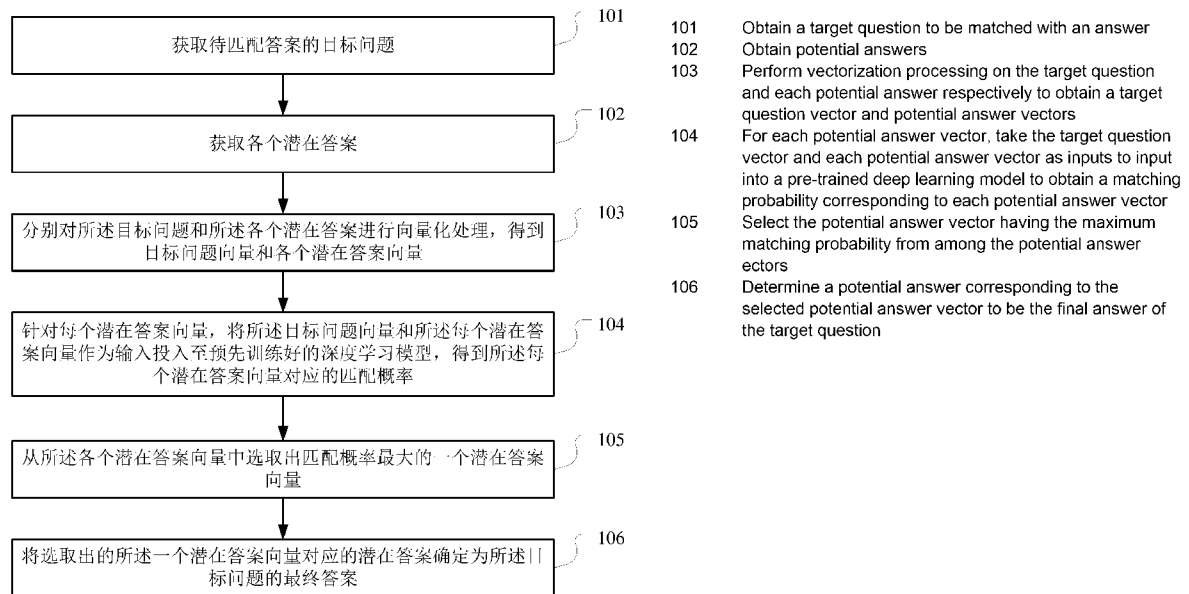


图 2

(57) **Abstract:** Disclosed by the present application are a question answer selection method and apparatus, a computer device, and a storage medium, which are applied to the technical field of deep learning, and are used to solve the problem of how to find the best answer from among a plurality of potential answers. The method provided in the present application comprises: obtaining a target question to be matched with an answer; obtaining potential answers; performing vectorization processing on the target question and each potential answer respectively to obtain a target question vector and potential answer vectors; for each potential answer vector, taking the target question vector and each potential answer vector as inputs to input into a pre-trained deep learning model to obtain

JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

a matching probability corresponding to each potential answer vector, wherein the matching probability represents the probability of a potential answer corresponding to each potential answer vector matching the target question; selecting the potential answer vector having the maximum matching probability from among the potential answer vectors; and determining a potential answer corresponding to the selected potential answer vector to be the final answer of the target question.

(57) 摘要: 本申请公开了一种问题答案选取方法、装置、计算机设备及存储介质, 应用于深度学习技术领域, 用于解决如何从多个潜在答案中寻找最佳答案的问题。本申请提供的方法包括: 获取待匹配答案的目标问题; 获取各个潜在答案; 分别对目标问题和各个潜在答案进行向量化处理, 得到目标问题向量和各个潜在答案向量; 针对每个潜在答案向量, 将目标问题向量和每个潜在答案向量作为输入投入至预先训练好的深度学习模型, 得到每个潜在答案向量对应的匹配概率, 匹配概率表征了每个潜在答案向量对应的潜在答案与目标问题匹配的概率; 从各个潜在答案向量中选取匹配概率最大的一个潜在答案向量; 将选取出的一个潜在答案向量对应的潜在答案确定为目标问题的最终答案。

一种问题答案选取方法、装置、计算机设备及存储介质

本申请以 2019 年 05 月 21 日提交的申请号为 201910421574.9，名称为“一种问题答案选取方法、装置、计算机设备及存储介质”的中国发明专利申请为基础，并要求其优先权。

技术领域

本申请涉及深度学习技术领域，尤其涉及一种问题答案选取方法、装置、计算机设备及存储介质。

背景技术

智能问答是人工智能领域中一个长期被研究的课题。在实际应用中，当用户提问后，经过语音识别、语义识别等技术手段确定用户的问题之后，往往可以从预设的答案数据库中搜索到与问题相关的多个潜在答案，然而，若将这些潜在答案均作为问题的最终答案提供给用户显然难以满足用户的需求。

因此，寻找一种从多个潜在答案中寻找最佳答案的方法成为本领域技术人员亟需解决的问题。

发明内容

本申请实施例提供一种问题答案选取方法、装置、计算机设备及存储介质，以解决如何从多个潜在答案中寻找最佳答案的问题。

一种问题答案选取方法，包括：

获取待匹配答案的目标问题；

获取各个潜在答案；

分别对所述目标问题和所述各个潜在答案进行向量化处理，得到目标问题向量和各个潜在答案向量；

针对每个潜在答案向量，将所述目标问题向量和所述每个潜在答案向量作为输入投入至预先训练好的深度学习模型，得到所述每个潜在答案向量对应的匹配概率，所述匹配概率表征了所述每个潜在答案向量对应的潜在答案与所述目标问题匹配的概率；

从所述各个潜在答案向量中选取匹配概率最大的一个潜在答案向量；

将选取出的所述一个潜在答案向量对应的潜在答案确定为所述目标问题的最终答案。

一种问题答案选取装置，包括：

目标问题获取模块，用于获取待匹配答案的目标问题；

潜在答案获取模块，用于获取各个潜在答案；

向量化模块，用于分别对所述目标问题和所述各个潜在答案进行向量化处理，得到目标问题向量和各个潜在答案向量；

模型识别模块，用于针对每个潜在答案向量，将所述目标问题向量和所述每个潜在答案向量作为输入投入至预先训练好的深度学习模型，得到所述每个潜在答案向量对应的匹配概率，所述匹配概率表征了所述每个潜在答案向量对应的潜在答案与所述目标问题匹配的概率；

答案向量选取模块，用于从所述各个潜在答案向量中选取匹配概率最大的一个潜在答案向量；

最终答案确定模块，用于将选取出的所述一个潜在答案向量对应的潜在答案确定为所述目标问题的最终答案。

一种计算机设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令，所述处理器执行所述计算机可读指令时实现上述问题答案选取方法的步骤。

一个或多个存储有计算机可读指令的可读存储介质，所述计算机可读存储介质存储有计算机可读指令，使得所述一个或多个处理器执行上述问题答案选取方法的步骤。

本申请的一个或多个实施例的细节在下面的附图和描述中提出，本申请的其他特征和优点将从说明书、附图以及权利要求变得明显。

附图说明

为了更清楚地说明本申请实施例的技术方案，下面将对本申请实施例的描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图 1 是本申请一实施例中问题答案选取方法的一应用环境示意图；

图 2 是本申请一实施例中问题答案选取方法的一流程图；

图 3 是本申请一实施例中问题答案选取方法步骤 103 在一个应用场景下的流程示意图；

图 4 是本申请一实施例中问题答案选取方法在一个应用场景下预先训练深度学习模型的流程示意图；

图 5 是本申请一实施例中问题答案选取方法步骤 304 在一个应用场景下的流程示意图；

图 6 是本申请一实施例中问题答案选取方法步骤 104 在一个应用场景下的流程示意图；

图 7 是本申请一实施例中问题答案选取装置一个应用场景下的结构示意图；

图 8 是本申请一实施例中问题答案选取装置另一个应用场景下的结构示意图；

图 9 是本申请一实施例中特征聚类模块的结构示意图；

图 10 是本申请一实施例中计算机设备的一示意图。

具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

本申请提供的问题答案选取方法，可应用在如图 1 的应用环境中，其中，客户端通过网络与服务器进行通信。其中，该客户端可以但不限于各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备。服务器可以用独立的服务器或者是多个服务器组成的服务器集群来实现。

在一实施例中，如图 2 所示，提供一种问题答案选取方法，以该方法应用在图 1 中的服务器为例进行说明，包括如下步骤：

101、获取待匹配答案的目标问题；

本实施例中，服务器可以根据实际使用的需要或者应用场景的需要获取到待匹配答案的目标问题。比如，可以是用户通过客户端提问，服务器经过对问题的分析和确认，确定出该用户想要询问的真实问题，该真实问题就是本实施例中所说的目标问题；或

者，可以通过第三方系统与服务器对接，该第三方系统需要为问题寻找最佳答案时，将该问题和与之对应的多个潜在答案提供给服务器，从而服务器可以获取到该问题作为待匹配答案的目标问题。

102、获取各个潜在答案；

服务器获取到待匹配答案的目标问题之后，可以一并获取到该目标问题对应的各个潜在答案，这些潜在答案可以是来自于服务器自身，比如上述服务器通过分析用户提问得出真实问题，服务器可以从对接的答案数据库中检索出与该真实问题相关的各个答案作为潜在答案；或者，这些潜在答案可以是第三方系统向服务器提供问题的同时一并提供的各个答案。

需要说明的是，本实施例中对服务器获取到目标问题和各个潜在答案的方式不作具体限定，但应当理解的是，所述各个潜在答案应当是与目标问题高度相关的答案。

103、分别对所述目标问题和所述各个潜在答案进行向量化处理，得到目标问题向量和各个潜在答案向量；

在得到所述目标问题和所述各个潜在答案之后，为了便于后续深度学习模型的识别，服务器需要分别对所述目标问题和所述各个潜在答案进行向量化处理，得到目标问题向量和各个潜在答案向量。具体地，服务器可以将所述目标问题和所述各个潜在答案以矩阵的形式记载，对于每个目标问题或每个潜在答案对应的矩阵，文本中的每个字词可以映射为矩阵中的一个行向量。

为便于理解，如图3所示，进一步地，步骤103可以包括：

201、针对每个目标文本，将所述每个目标文本中的各个目标字词分别转换为 GloVe (Global Vectors for Word Representation) 词向量，得到初始向量，所述目标文本为所述目标问题或所述潜在答案；

202、判断所述各个目标字词是否均被 GloVe 词向量覆盖，若是，则执行步骤 203，若否，则执行步骤 204；

203、确定所述初始向量为所述每个目标文本对应目标向量，所述目标向量为所述目标问题向量或所述潜在答案向量；

204、将未被 GloVe 词向量覆盖的目标字词转换为 TransE 词向量，得到补充向量；

205、将所述补充向量添加至所述初始向量，得到所述每个目标文本对应目标向量。

对于上述步骤 201，GloVe 的全称叫 Global Vectors for Word Representation，它是现有一个基于全局词频统计 (count-based & overall statistics) 的词表征 (word representation) 工具，它可以把一个单词表达成一个由实数组成的向量。本实施例中，服务器使用 GloVe 将所述目标文本中的各个目标字词分别转换成词向量，从而得到初始问题向量。

其中，这里所说的目标文本是指目标问题或潜在答案的文本。

对于上述步骤 202，考虑到目标文本中可能包含有专有名词，例如姓名、地点等，这些专有名词难以被 GloVe 全覆盖。因此，服务器可以判断所述各个目标字词是否均被 GloVe 词向量覆盖，如果该目标文本中的各个目标字词均已被覆盖，则可以执行步骤 203，直接确定所述初始向量为目标向量；反之，如果该目标文本中的各个目标字词未被全覆盖，则需要执行后续步骤 204 和步骤 205。

对于上述步骤 203，由上述内容可知，若所述各个目标字词均被 GloVe 词向量覆盖，则服务器可以确定所述初始向量为所述每个目标文本对应目标向量。其中，目标向量为所述目标问题向量或所述潜在答案向量。

对于上述步骤 204，若所述各个目标字词中任一个目标字词未被 GloVe 词向量覆盖，则可知该目标文本中存在无法被 GloVe 词向量覆盖的目标字词，为了补充这一部分的缺失，服务器可以将未被 GloVe 词向量覆盖的目标字词转换为 TransE 词向量，得到补充向量。

需要说明的，TransE，又称知识库方法，是一种现有的有效学习专有名词的算法模型，可以将学习到的字词转换为分布式向量表示。本实施例中，服务器可以采用 TransE 将未被 GloVe 词向量覆盖的目标字词进行向量转换，得到补充向量。

对于步骤 205，可知，服务器在得到补充向量之后，可以使用该补充向量添加至该初始向量中，以填补初始向量的缺失，从而得到该目标文本对应的目标向量。举例说明，假设该目标文本为“小明吃饭吗”，该句子中包括“小明”、“吃饭”和“吗”三个目标字词。服务器使用 GloVe 将“吃饭”和“吗”转换为词向量，分别为[1234]和[1235]，对于“小明”一词，服务器使用 TransE 将其转换为词向量，得到[1236]，再将[1236]添加至[1234]和[1235]中，得到该目标向量为[1236]、[1234]和[1235]，其中，该目标向量可以以一维向量的形式表达，即[1236]、[1234]和[1235]合并为[123612341235]作为该目标向量，也可以以二维向量的形式表达，即[1236]、[1234]和[1235]分别作为一个

二维向量的行向量，得到目标向量为：

$$\begin{bmatrix} 1 & 2 & 3 & 6 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 5 \end{bmatrix}$$

104、针对每个潜在答案向量，将所述目标问题向量和所述每个潜在答案向量作为输入投入至预先训练好的深度学习模型，得到所述每个潜在答案向量对应的匹配概率，所述匹配概率表征了所述每个潜在答案向量对应的潜在答案与所述目标问题匹配的概率；

可以理解的是，服务器在得到所述目标问题向量和各个潜在答案向量之后，针对每个潜在答案向量，可以将所述目标问题向量和所述每个潜在答案向量作为输入投入至预先训练好的深度学习模型，得到所述每个潜在答案向量对应的匹配概率。其中，所述匹配概率表征了所述每个潜在答案向量对应的潜在答案与所述目标问题匹配的概率，某个潜在答案向量对应的匹配概率越大，则该潜在答案向量对应的潜在答案与该目标问题越匹配，对于用户感受来说，可以认为该潜在答案向量对应的潜在答案越能满足用户的需求。

为便于理解，下面将对深度学习模型的训练过程进行详细描述。如图 4 所示，进一步地，所述深度学习模型包括分层循环编码器、特征聚类网络和相似度计算网络三部分，所述深度学习模型可以通过以下步骤预先训练好：

301、收集多对问答样本，每对所述问答样本包括一个样本问题和与所述样本问题对应的标准答案；

302、针对每对问答样本，对所述每对问答样本中的样本问题和标准答案进行向量化处理，得到所述每对问答样本的样本问题向量和标准答案向量；

303、将所述每对问答样本的样本问题向量和标准答案向量分别输入分层循环编码器进行编码，得到问题特征向量和答案特征向量，所述分层循环编码器由两个单层 GRU 神经网络组成；

304、将所述每对问答样本的答案特征向量输入特征聚类网络进行特征聚类计算，得到聚类更新后的答案特征向量；

305、针对每对问答样本，通过相似度计算网络计算所述每对问答样本的问题特征向量与所述聚类更新后的答案特征向量之间的相似度，得到所述每对问答样本中样本问题与标准答案的相似度值，作为样本匹配概率；

306、以所述每对问答样本对应的样本匹配概率为调整目标，调整所述分层循环编码器、所述特征聚类网络和所述相似度计算网络的网络参数，使得所述每对问答样本对应的样本匹配概率与 1 之间的误差最小化；

307、若每对问答样本对应的样本匹配概率与 1 之间的误差满足预设的训练终止条件，则确定所述深度学习模型已训练好。

对于步骤 301, 为了更好地训练深度学习模型, 可以预先收集大量的问答样本, 每对所述问答样本包括一个样本问题和与所述样本问题对应的标准答案。这里说的标准答案是指经过确认和审核, 可以解答该样本问题的答案, 在本实施例中, 可以认为标准答案就是样本问题的最正确的答案。

对于步骤 302, 与上述步骤 103 同理, 服务器可以针对每对问答样本, 对所述每对问答样本中的样本问题和标准答案进行向量化处理, 得到所述每对问答样本的样本问题向量和标准答案向量, 此处不再对向量化处理的内容进行描述, 具体可以参照步骤 103。

对于步骤 303, 本实施例中, 该所述分层循环编码器由两个单层 GRU (Gated Recurrent Unit) 神经网络组成。这里说的 GRU 神经网络是循环神经网络 RNN 的变种, 相比 RNN 网络存在容易遗忘长期数据的问题, GRU 神经网络在处理长文本、长向量的问题上有更加突出的表现, 可以在一定程度上解决长文本中较远状态被遗忘的问题。

对于所述每对问答样本的样本问题向量, 服务器首先将该样本问题向量输入到第一层单层 GRU 神经网络中进行对向量状态信息的更新, 得到第一层单层 GRU 神经网络输出的向量, 这里记为第一向量。在具体实现中, 可以将一个样本问题的文本段落划分为 c 个文本块, 每个文本块假设共 t 个单词, 每个单词对应的向量记为一个向量块, 该样本问题对应的样本问题向量由 $c*t$ 个向量块组成。服务器可以将该样本问题向量的各个向量块逐个输入到第一层单层 GRU 神经网络, 得到对各个向量块更新状态信息后的各个第一向量块, 记为 $h_{c,t}^Q$, c 为文本块的编号, t 为单词在当前文本块中的位置。在得到 $c*t$ 个第一向量

块 $h_{c,t}^Q$ 之后, 服务器再将每个文本块中最后一个单词对应的第一向量块 $h_{c,T}^Q$ 逐个输入到第二层单层 GRU 神经网络, 得到各个第二向量块, 可记为 $\mu_{c,T}^Q$, T 表示每个向量块中处于第 T 个位置的编号。这些第二向量块组成该样本问题向量对应的问题特征向量。

同理, 对于每对问答样本的标准答案向量与上述处理过程基本一致, 具体地, 服务器可以将该标准答案向量的各个向量块逐个输入到第一层单层 GRU 神经网络, 得到对各个向量块更新状态信息后的各个第三向量块, 记为 $h_{c,t}^A$, 服务器再将每个文本块中最后一个单词对应的第一向量块 $h_{c,T}^A$ 逐个输入到第二层单层 GRU 神经网络, 得到各个第四向量块, 可记为 $\mu_{c,T}^A$, 这些第四向量块组成该标准答案向量对应的答案特征向量。

具体地, 本实施例中的第一层单层 GRU 神经网络和第二层单层 GRU 神经网络可以是含有 200 个隐藏单元的单层 GRU 神经网络。

优选地, 为了优化该分层循环编码器的编码能力, 合理设置其更新步长, 可以采用梯度剪切值为 1Adam 优化器计算该分层循环编码器在编码时的更新补偿, 相较于常用的梯度下降算法具有更好的性能。

对于步骤 304, 针对步骤 303 中计算出的答案特征向量, 考虑到一般情况下答案的文本长度比问题的要长, 可以认为答案的文本中具有足够多的信息进行特征的聚类计算。因此, 本实施例中针对答案特征向量还需执行步骤 304, 而对步骤 303 计算得到问题特征向量则无需执行步骤 304。

具体地, 服务器在得到答案特征向量之后, 可以将所述每对问答样本的答案特征向量输入特征聚类网络进行特征聚类计算, 得到聚类更新后的答案特征向量。为便于理解, 如图 5 所示, 更进一步地, 步骤 304 可以包括:

401、针对各个特征向量块中的每个特征向量块, 计算所述每个特征向量块与各个预

设特征类别向量之间的相似度，得到各个相似度值，所述各个特征向量块是指在所述每对问答样本中、标准答案的各个字词在所述答案特征向量中的对应部分向量；

402、将所述各个相似度值分别与各自对应的预设特征类别向量进行加权求和运算，并将运算得到的向量与所述每个特征向量块并联，得到所述每个特征向量块对应的、聚类更新后的特征向量块；

403、将所述各个特征向量块各自对应的、聚类更新后的特征向量块组合得到所述每对问答样本的、聚类更新后的答案特征向量。

对于上述步骤 401，承接上述步骤 303 的内容，可知，这里所说的各个特征向量块即为上述的各个第四向量块 $\mu_{c,T}^A$ 。本实施例还预先设置各个预设特征类别向量，具体可以使用特征类别矩阵 m 表示，比如 m 的大小可以为 64×4 ，64 表示每一特征类别对应 64 个元素的一维向量，4 为预设的特征类别的数量。服务器计算所述每个特征向量块与各个预设特征类别向量之间的相似度时，可以将每个特征向量块分别与各个预设特征类别向量点乘，公式表示为 $\mu_{c,T}^A * m_k$ ，其中 k 为特征类别的编号，点乘后得到的结果经过 softmax 函数计算即可得到每个特征向量块分别与各个预设特征类别向量的相似度值，公式表示为 $p_k = \text{softmax}((\mu_{c,T}^A)^T m_k)$ 。

对于步骤 402，具体地，所述每个特征向量块对应的、聚类更新后的特征向量块 $e^{u,A}$ 可以通过下述公式表示：

$$e^{u,A} = \left\{ \sum_{k=1}^K p_k m_k, \mu_{c,T}^A \right\}$$

其中， K 为预设的特征类别的数量。

对于步骤 403，可以理解的是，服务器在计算得到每个特征向量块对应的、聚类更新后的特征向量块之后，由于各个特征向量块仅为标准答案对应的的向量的一部分，因此服务器可以将所述各个特征向量块各自对应的、聚类更新后的特征向量块组合得到所述每对问答样本的、聚类更新后的答案特征向量，也可以认为将这些各个特征向量块整理看作一个标准答案对应的、聚类更新后的答案特征向量。

对于步骤 305，服务器在得到所述每对问答样本的问题特征向量与所述聚类更新后的答案特征向量之后，即可通过相似度计算网络计算这两者的相似度作为两者之间的匹配概率，可知，相似度越高，表示两者越匹配；反之，相似度越低，表示两者越不匹配。具体地，所述相似度计算网络可以通过以下公式表示：

$$P = \sigma((\mu_{c,T}^Q)^T M e^{u,A} + b)$$

其中， P 为相似度值， M 和 b 为该相似度计算网络的网络参数， σ 是 sigmoid 函数。

如上所述， $\mu_{c,T}^Q$ 和 $e^{u,A}$ 分别为分层循环编码器输出的问题特征向量和步骤 304 得到的聚类更新后的答案特征向量。

对于步骤 306，可以理解的是，可以认为每对问答样本中的样本问题和标准答案之间是百分百匹配的，两者理论上的匹配概率应当为 100%，也就是 1。因此，在训练过程中，服务器可以以所述每对问答样本对应的样本匹配概率为调整目标，调整所述分层循环编码器、所述特征聚类网络和所述相似度计算网络的网络参数，使得所述每对问答样本对应的样本匹配概率与 1 之间的误差最小化，也即，尽可能让所述每对问答样本对应的

样本匹配概率逼近或者等于 1。

对于步骤 307，在针对每对问答样本均执行完成上述步骤 302-306 之后，可以判断每对问答样本对应的样本匹配概率与 1 之间的误差满足预设的训练终止条件，若满足，则说明该深度学习模型中的各个网路参数已经调整到位，可以确定所述深度学习模型已训练好；反之，若不满足，则说明该深度学习模型中的各个网路参数尚未调整到位，仍需继续训练。其中，该训练终止条件可以根据实际使用情况预先设定，具体地，可以将该训练终止条件设定为：若每对问答样本对应的样本匹配概率与 1 之间的误差均小于指定误差值，则认为其满足该预设的训练终止条件。或者，也可以将其设为：使用验证集中的问答样本执行上述步骤 302-306，若深度学习模型输出的样本匹配概率与 1 之间的误差在一定范围内，则认为其满足该预设的训练终止条件。其中，该验证集中的问答样本的收集与上述步骤 301 类似，具体地，可以执行上述步骤 301 收集得到多对问答样本后，将收集得到的问答样本中的一部分划分为训练集，剩余的问答样本划分为验证集。比如，可以将收集得到的问答样本中随机划分 80% 作为后续训练深度学习模型的训练集的样本，将其它的 20% 划分为后续验证深度学习模型是否训练完成，也即是否满足预设训练终止条件的验证集的样本。

优选地，在实际应用中，为了测试已训练完成的深度学习模型的效果，可以提取一对问答样本，并为该问答样本的样本问题创造 9 个错误的答案，由这 1 个样本问题和 10 个答案（1 个标准答案和 9 个错误答案）组成测试集。在深度学习模型训练完成之后，将这 1 个样本问题和 10 个答案输入到深度学习模型，得到深度学习模型输出的各个匹配概率，检验其中标准答案对应的匹配概率是否为最大，若是，则可以认为该深度学习模型训练后的效果合格；若否，则该深度学习模型训练后的效果不合格，可能需要重新调整深度学习模型的网络参数。

上面描述了该深度学习模型的训练过程，为便于理解，下面承接上述训练过程的内容，对步骤 104 使用该深度学习模型来得到目标问题向量与每个潜在答案向量之间匹配概率的过程进行详细描述。如图 6 所示，更进一步地，所述深度学习模型包括分层循环编码器、特征聚类网络和相似度计算网络三部分，步骤 104 可以包括：

501、针对每个潜在答案向量，将所述目标问题向量和所述每个潜在答案向量分别输入分层循环编码器进行编码，得到目标问题特征向量和目标答案特征向量；

502、将所述每个潜在答案向量对应的目标答案特征向量输入特征聚类网络进行特征聚类计算，得到聚类更新后的目标答案特征向量；

503、通过相似度计算网络计算所述目标问题向量与所述每个潜在答案向量对应的、聚类更新后的目标答案特征向量之间的相似度，得到所述目标问题与所述每个潜在答案的相似度值，作为所述每个潜在答案向量对应的匹配概率。

对于步骤 501，与上述步骤 303 同理，服务器针对每个潜在答案向量，将该目标问题向量和所述每个潜在答案向量分别输入分层循环编码器进行编码，得到目标问题特征向量和目标答案特征向量。对于分层循环编码器的编码过程请参阅上述步骤 303 的内容，此处不再赘述。

对于步骤 502，对于步骤 501 中得到目标答案特征向量，服务器还可以将所述每个潜在答案向量对应的目标答案特征向量输入特征聚类网络进行特征聚类计算，得到聚类更新后的目标答案特征向量，本步骤与上述步骤 304 同理，特征聚类计算的过程参阅步骤 304，此处不再赘述。

可以理解的是，由于答案的文本通常比问题的文本要长，加上现有的神经网络的表现会随着文本长度增长而迅速下降，考虑到这点，该深度学习模型中的特征聚类网络特地针对具有足够多信息的答案进行特征的聚类，从答案文本对应的向量数据中深度挖掘出带有特征的信息，进而再加入到后续步骤中参与匹配概率的计算，可以克服和改善现

有神经网络表现较差的问题。

对于步骤 503，与上述步骤 305 同理，服务器可以通过相似度计算网络计算所述目标问题向量与所述每个潜在答案向量对应的、聚类更新后的目标答案特征向量之间的相似度，得到所述目标问题与所述每个潜在答案的相似度值，作为所述每个潜在答案向量对应的匹配概率。可知，假设存在 100 个潜在答案，1 个目标问题，则步骤 503 最后计算得到 100 个匹配概率分别与 100 个潜在答案一一对应，每个匹配概率反映了与之对应的潜在答案与该目标问题之间的匹配程度。

105、从所述各个潜在答案向量中选取匹配概率最大的一个潜在答案向量；

由上述内容可知，服务器在得到所述每个潜在答案向量对应的匹配概率，为了筛选出各个潜在答案中的最优答案，先从所述各个潜在答案向量中选取匹配概率最大的一个潜在答案向量。具体，服务器也可以按照匹配概率从大到小为所述各个潜在答案向量进行排序，选取排序第一的潜在答案向量，这里选取出的排序第一的潜在答案向量也就是匹配概率最大的一个潜在答案向量。

106、将选取出的所述一个潜在答案向量对应的潜在答案确定为所述目标问题的最终答案。

在选取匹配概率最大的一个潜在答案向量之后，服务器可以将选取出的所述一个潜在答案向量对应的潜在答案确定为所述目标问题的最终答案。可以理解的是，选取出的该潜在答案相比其它潜在答案来说，与该目标问题更加匹配，因此可以认为该潜在答案为最优答案。

本申请实施例中，首先，获取待匹配答案的目标问题；并获取各个潜在答案；然后，分别对所述目标问题和所述各个潜在答案进行向量化处理，得到目标问题向量和各个潜在答案向量；接着，针对每个潜在答案向量，将所述目标问题向量和所述每个潜在答案向量作为输入投入至预先训练好的深度学习模型，得到所述每个潜在答案向量对应的匹配概率，所述匹配概率表征了所述每个潜在答案向量对应的潜在答案与所述目标问题匹配的概率；再之，从所述各个潜在答案向量中选取匹配概率最大的一个潜在答案向量；最后，将选取出的所述一个潜在答案向量对应的潜在答案确定为所述目标问题的最终答案。可见，本申请通过将目标问题和各个潜在答案向量化后投入到深度学习模型中，利用深度学习模型计算出目标问题与各个潜在答案的匹配概率，再从中取出匹配概率最大的潜在答案作为该目标问题的最终答案，实现了最佳答案的筛选，能够在面对多个潜在答案的情况下满足用户的问答需求。

应理解，上述实施例中各步骤的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不对本申请实施例的实施过程构成任何限定。

在一实施例中，提供一种问题答案选取装置，该问题答案选取装置与上述实施例中问题答案选取方法一一对应。如图 7 所示，该问题答案选取装置包括目标问题获取模块 601、潜在答案获取模块 602、向量化模块 603、模型识别模块 604、答案向量选取模块 605 和最终答案确定模块 606。各功能模块详细说明如下：

目标问题获取模块 601，用于获取待匹配答案的目标问题；

潜在答案获取模块 602，用于获取各个潜在答案；

向量化模块 603，用于分别对所述目标问题和所述各个潜在答案进行向量化处理，得到目标问题向量和各个潜在答案向量；

模型识别模块 604，用于针对每个潜在答案向量，将所述目标问题向量和所述每个潜在答案向量作为输入投入至预先训练好的深度学习模型，得到所述每个潜在答案向量对应的匹配概率，所述匹配概率表征了所述每个潜在答案向量对应的潜在答案与所述目标问题匹配的概率；

答案向量选取模块 605, 用于从所述各个潜在答案向量中选取匹配概率最大的一个潜在答案向量;

最终答案确定模块 606, 用于将选取出的所述一个潜在答案向量对应的潜在答案确定为所述目标问题的最终答案。

如图 8 所示, 进一步地, 所述深度学习模型包括分层循环编码器、特征聚类网络和相似度计算网络三部分, 所述深度学习模型可以通过以下模块预先训练好:

问答样本收集模块 607, 用于收集多对问答样本, 每对所述问答样本包括一个样本问题和与所述样本问题对应的标准答案;

样本向量化模块 608, 用于针对每对问答样本, 对所述每对问答样本中的样本问题和标准答案进行向量化处理, 得到所述每对问答样本的样本问题向量和标准答案向量;

样本编码模块 609, 用于将所述每对问答样本的样本问题向量和标准答案向量分别输入分层循环编码器进行编码, 得到问题特征向量和答案特征向量, 所述分层循环编码器由两个单层 GRU 神经网络组成;

特征聚类模块 610, 用于将所述每对问答样本的答案特征向量输入特征聚类网络进行特征聚类计算, 得到聚类更新后的答案特征向量;

相似度值计算模块 611, 用于针对每对问答样本, 通过相似度计算网络计算所述每对问答样本的问题特征向量与所述聚类更新后的答案特征向量之间的相似度, 得到所述每对问答样本中样本问题与标准答案的相似度值, 作为样本匹配概率;

网络参数调整模块 612, 用于以所述每对问答样本对应的样本匹配概率为调整目标, 调整所述分层循环编码器、所述特征聚类网络和所述相似度计算网络的网络参数, 使得所述每对问答样本对应的样本匹配概率与 1 之间的误差最小化;

训练完成确定模块 613, 用于若每对问答样本对应的样本匹配概率与 1 之间的误差满足预设的训练终止条件, 则确定所述深度学习模型已训练好。

如图 9 所示, 进一步地, 所述特征聚类模块 610 可以包括:

相似度计算单元 6101, 用于针对各个特征向量块中的每个特征向量块, 计算所述每个特征向量块与各个预设特征类别向量之间的相似度, 得到各个相似度值, 所述各个特征向量块是指在所述每对问答样本中、标准答案的各个字词在所述答案特征向量中的对应部分向量;

加权求和单元 6102, 用于将所述各个相似度值分别与各自对应的预设特征类别向量进行加权求和运算, 并将运算得到的向量与所述每个特征向量块并联, 得到所述每个特征向量块对应的、聚类更新后的特征向量块;

特征向量组合单元 6103, 用于将所述各个特征向量块各自对应的、聚类更新后的特征向量块组合得到所述每对问答样本的、聚类更新后的答案特征向量。

进一步地, 所述模型识别模块可以包括:

向量编码单元, 用于针对每个潜在答案向量, 将所述目标问题向量和所述每个潜在答案向量分别输入分层循环编码器进行编码, 得到目标问题特征向量和目标答案特征向量;

聚类计算单元, 用于将所述每个潜在答案向量对应的目标答案特征向量输入特征聚类网络进行特征聚类计算, 得到聚类更新后的目标答案特征向量;

匹配概率计算单元, 用于通过相似度计算网络计算所述目标问题向量与所述每个潜在答案向量对应的、聚类更新后的目标答案特征向量之间的相似度, 得到所述目标问题与所述每个潜在答案的相似度值, 作为所述每个潜在答案向量对应的匹配概率。

进一步地, 所述向量化模块可以包括:

第一转换单元, 用于针对每个目标文本, 将所述每个目标文本中的各个目标字词分别转换为 GloVe 词向量, 得到初始向量, 所述目标文本为所述目标问题或所述潜在答案;

字词判断单元, 用于判断所述各个目标字词是否均被 GloVe 词向量覆盖;

向量确定单元,用于若所述字词判断单元的判断结果为是,则确定所述初始向量为所述每个目标文本对应目标向量,所述目标向量为所述目标问题向量或所述潜在答案向量;

第二转换单元,用于若所述字词判断单元的判断结果是否为,则将未被 GloVe 词向量覆盖的目标字词转换为 TransE 词向量,得到补充向量;

向量添加单元,用于将所述补充向量添加至所述初始向量,得到所述每个目标文本对应目标向量。

关于问题答案选取装置的具体限定可以参见上文中对于问题答案选取方法的限定,在此不再赘述。上述问题答案选取装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中,也可以以软件形式存储于计算机设备中的存储器中,以便于处理器调用执行以上各个模块对应的操作。

在一个实施例中,提供了一种计算机设备,该计算机设备可以是服务器,其内部结构图可以如图 10 所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中,该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括可读存储介质、内存储器。该可读存储介质存储有操作系统、计算机可读指令和数据库。该内存储器为可读存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的数据库用于存储问题答案选取方法中涉及到的数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机可读指令被处理器执行时以实现一种问题答案选取方法。本实施例所提供的可读存储介质包括非易失性可读存储介质和易失性可读存储介质。

在一个实施例中,提供了一种计算机设备,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机可读指令,处理器执行计算机可读指令时实现上述实施例中问题答案选取方法的步骤,例如图 2 所示的步骤 101 至步骤 106。或者,处理器执行计算机可读指令时实现上述实施例中问题答案选取装置各模块/单元的功能,例如图 7 所示模块 601 至模块 606 的功能。为避免重复,这里不再赘述。

在一个实施例中,提供了一种计算机可读存储介质,该一个或多个存储有计算机可读指令的可读存储介质,计算机可读指令被一个或多个处理器执行时,使得一个或多个处理器执行计算机可读指令时实现上述问题答案选取方法的步骤,或者,该一个或多个存储有计算机可读指令的可读存储介质,计算机可读指令被一个或多个处理器执行时,使得一个或多个处理器执行计算机可读指令时实现上述装置实施例中问题答案选取装置中各模块/单元的功能。为避免重复,这里不再赘述。本实施例所提供的可读存储介质包括非易失性可读存储介质和易失性可读存储介质。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机可读指令来指令相关的硬件来完成,所述的计算机可读指令可存储于一计算机可读存储介质中,该计算机可读指令在执行时,可包括如上述各方法的实施例的流程。其中,本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用,均可包括和/或易失性存储器。存储器可包括只读存储器 (ROM)、可编程 ROM (PROM)、电可编程 ROM (EPROM)、电可擦除可编程 ROM (EEPROM) 或闪存。易失性存储器可包括随机存取存储器 (RAM) 或者外部高速缓冲存储器。作为说明而非局限, RAM 以多种形式可得,诸如静态 RAM (SRAM)、动态 RAM (DRAM)、同步 DRAM (SDRAM)、双数据率 SDRAM (DDRSDRAM)、增强型 SDRAM (ESDRAM)、同步链路 (Synchlink) DRAM (SLDRAM)、存储器总线 (Rambus) 直接 RAM (RDRAM)、直接存储器总线动态 RAM (DRDRAM)、以及存储器总线动态 RAM (RDRAM) 等。

所属领域的技术人员可以清楚地了解到,为了描述的方便和简洁,仅以上述各功能单

元、模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能单元、模块完成，即将所述装置的内部结构划分成不同的功能单元或模块，以完成以上描述的全部或者部分功能。

以上所述实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围，均应包含在本申请的保护范围之内。

权 利 要 求 书

1. 一种问题答案选取方法，其特征在于，包括：

获取待匹配答案的目标问题；

获取各个潜在答案；

分别对所述目标问题和所述各个潜在答案进行向量化处理，得到目标问题向量和各个潜在答案向量；

针对每个潜在答案向量，将所述目标问题向量和所述每个潜在答案向量作为输入投入至预先训练好的深度学习模型，得到所述每个潜在答案向量对应的匹配概率，所述匹配概率表征了所述每个潜在答案向量对应的潜在答案与所述目标问题匹配的概率；

从所述各个潜在答案向量中选取匹配概率最大的一个潜在答案向量；

将选取出的所述一个潜在答案向量对应的潜在答案确定为所述目标问题的最终答案。

2. 根据权利要求 1 所述的问题答案选取方法，其特征在于，所述深度学习模型包括分层循环编码器、特征聚类网络和相似度计算网络三部分，所述深度学习模型通过以下步骤预先训练好：

收集多对问答样本，每对所述问答样本包括一个样本问题和与所述样本问题对应的标准答案；

针对每对问答样本，对所述每对问答样本中的样本问题和标准答案进行向量化处理，得到所述每对问答样本的样本问题向量和标准答案向量；

将所述每对问答样本的样本问题向量和标准答案向量分别输入分层循环编码器进行编码，得到问题特征向量和答案特征向量，所述分层循环编码器由两个单层 GRU 神经网络组成；

将所述每对问答样本的答案特征向量输入特征聚类网络进行特征聚类计算，得到聚类更新后的答案特征向量；

针对每对问答样本，通过相似度计算网络计算所述每对问答样本的问题特征向量与所述聚类更新后的答案特征向量之间的相似度，得到所述每对问答样本中样本问题与标准答案的相似度值，作为样本匹配概率；

以所述每对问答样本对应的样本匹配概率为调整目标，调整所述分层循环编码器、所述特征聚类网络和所述相似度计算网络的网络参数，使得所述每对问答样本对应的样本匹配概率与 1 之间的误差最小化；

若每对问答样本对应的样本匹配概率与 1 之间的误差满足预设的训练终止条件，则确定所述深度学习模型已训练好。

3. 根据权利要求 2 所述的问题答案选取方法，其特征在于，所述将所述每对问答样本的答案特征向量输入特征聚类网络进行特征聚类计算，得到聚类更新后的答案特征向量包括：

针对各个特征向量块中的每个特征向量块，计算所述每个特征向量块与各个预设特征类别向量之间的相似度，得到各个相似度值，所述各个特征向量块是指在所述每对问答样本中、标准答案的各个字词在所述答案特征向量中的对应部分向量；

将所述各个相似度值分别与各自对应的预设特征类别向量进行加权求和运算，并将运算得到的向量与所述每个特征向量块并联，得到所述每个特征向量块对应的、聚类更新后的特征向量块；

将所述各个特征向量块各自对应的、聚类更新后的特征向量块组合得到所述每对问答样本的、聚类更新后的答案特征向量。

4. 根据权利要求 2 所述的问题答案选取方法，其特征在于，所述针对每个潜在答案

向量,将所述目标问题向量和所述每个潜在答案向量作为输入投入至预先训练好的深度学习模型,得到所述每个潜在答案向量对应的匹配概率包括:

针对每个潜在答案向量,将所述目标问题向量和所述每个潜在答案向量分别输入分层循环编码器进行编码,得到目标问题特征向量和目标答案特征向量;

将所述每个潜在答案向量对应的目标答案特征向量输入特征聚类网络进行特征聚类计算,得到聚类更新后的目标答案特征向量;

通过相似度计算网络计算所述目标问题向量与所述每个潜在答案向量对应的、聚类更新后的目标答案特征向量之间的相似度,得到所述目标问题与所述每个潜在答案的相似度值,作为所述每个潜在答案向量对应的匹配概率。

5. 根据权利要求 1 至 4 中任一项所述的问题答案选取方法,其特征在于,所述分别对所述目标问题和所述各个潜在答案进行向量化处理,得到目标问题向量和各个潜在答案向量包括:

针对每个目标文本,将所述每个目标文本中的各个目标字词分别转换为 GloVe 词向量,得到初始向量,所述目标文本为所述目标问题或所述潜在答案;

判断所述各个目标字词是否均被 GloVe 词向量覆盖;

若所述各个目标字词均被 GloVe 词向量覆盖,则确定所述初始向量为所述每个目标文本对应目标向量,所述目标向量为所述目标问题向量或所述潜在答案向量;

若所述各个目标字词中任一个目标字词未被 GloVe 词向量覆盖,则将未被 GloVe 词向量覆盖的目标字词转换为 TransE 词向量,得到补充向量;

将所述补充向量添加至所述初始向量,得到所述每个目标文本对应目标向量。

6. 一种问题答案选取装置,其特征在于,包括:

目标问题获取模块,用于获取待匹配答案的目标问题;

潜在答案获取模块,用于获取各个潜在答案;

向量化模块,用于分别对所述目标问题和所述各个潜在答案进行向量化处理,得到目标问题向量和各个潜在答案向量;

模型识别模块,用于针对每个潜在答案向量,将所述目标问题向量和所述每个潜在答案向量作为输入投入至预先训练好的深度学习模型,得到所述每个潜在答案向量对应的匹配概率,所述匹配概率表征了所述每个潜在答案向量对应的潜在答案与所述目标问题匹配的概率;

答案向量选取模块,用于从所述各个潜在答案向量中选取匹配概率最大的一个潜在答案向量;

最终答案确定模块,用于将选取出的所述一个潜在答案向量对应的潜在答案确定为所述目标问题的最终答案。

7. 根据权利要求 6 所述的问题答案选取装置,其特征在于,所述深度学习模型包括分层循环编码器、特征聚类网络和相似度计算网络三部分,所述深度学习模型通过以下模块预先训练好:

问答样本收集模块,用于收集多对问答样本,每对所述问答样本包括一个样本问题与和所述样本问题对应的标准答案;

样本向量化模块,用于针对每对问答样本,对所述每对问答样本中的样本问题和标准答案进行向量化处理,得到所述每对问答样本的样本问题向量和标准答案向量;

样本编码模块,用于将所述每对问答样本的样本问题向量和标准答案向量分别输入分层循环编码器进行编码,得到问题特征向量和答案特征向量,所述分层循环编码器由两个单层 GRU 神经网络组成;

特征聚类模块,用于将所述每对问答样本的答案特征向量输入特征聚类网络进行特征聚类计算,得到聚类更新后的答案特征向量;

相似度值计算模块,用于针对每对问答样本,通过相似度计算网络计算所述每对问答样本的问题特征向量与所述聚类更新后的答案特征向量之间的相似度,得到所述每对问答样本中样本问题与标准答案的相似度值,作为样本匹配概率;

网络参数调整模块,用于以所述每对问答样本对应的样本匹配概率为调整目标,调整所述分层循环编码器、所述特征聚类网络和所述相似度计算网络的网络参数,使得所述每对问答样本对应的样本匹配概率与1之间的误差最小化;

训练完成确定模块,用于若每对问答样本对应的样本匹配概率与1之间的误差满足预设的训练终止条件,则确定所述深度学习模型已训练好。

8. 根据权利要求7所述的问题答案选取装置,其特征在于,所述特征聚类模块包括:

相似度计算单元,用于针对各个特征向量块中的每个特征向量块,计算所述每个特征向量块与各个预设特征类别向量之间的相似度,得到各个相似度值,所述各个特征向量块是指在所述每对问答样本中、标准答案的各个字词在所述答案特征向量中的对应部分向量;

加权求和单元,用于将所述各个相似度值分别与各自对应的预设特征类别向量进行加权求和运算,并将运算得到的向量与所述每个特征向量块并联,得到所述每个特征向量块对应的、聚类更新后的特征向量块;

特征向量组合单元,用于将所述各个特征向量块各自对应的、聚类更新后的特征向量块组合得到所述每对问答样本的、聚类更新后的答案特征向量。

9. 根据权利要求7所述的问题答案选取装置,其特征在于,所述模型识别模块包括:

向量编码单元,用于针对每个潜在答案向量,将所述目标问题向量和所述每个潜在答案向量分别输入分层循环编码器进行编码,得到目标问题特征向量和目标答案特征向量;

聚类计算单元,用于将所述每个潜在答案向量对应的目标答案特征向量输入特征聚类网络进行特征聚类计算,得到聚类更新后的目标答案特征向量;

匹配概率计算单元,用于通过相似度计算网络计算所述目标问题向量与所述每个潜在答案向量对应的、聚类更新后的目标答案特征向量之间的相似度,得到所述目标问题与所述每个潜在答案的相似度值,作为所述每个潜在答案向量对应的匹配概率。

10. 根据权利要求6至9中任一项所述的问题答案选取装置,其特征在于,所述向量化模块包括:

第一转换单元,用于针对每个目标文本,将所述每个目标文本中的各个目标字词分别转换为 GloVe 词向量,得到初始向量,所述目标文本为所述目标问题或所述潜在答案;

字词判断单元,用于判断所述各个目标字词是否均被 GloVe 词向量覆盖;

向量确定单元,用于若所述字词判断单元的判断结果为是,则确定所述初始向量为所述每个目标文本对应目标向量,所述目标向量为所述目标问题向量或所述潜在答案向量;

第二转换单元,用于若所述字词判断单元的判断结果为否,则将未被 GloVe 词向量覆盖的目标字词转换为 TransE 词向量,得到补充向量;

向量添加单元,用于将所述补充向量添加至所述初始向量,得到所述每个目标文本对应目标向量。

11. 一种计算机设备,包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令,其特征在于,所述处理器执行所述计算机可读指令时实现如下步骤:

获取待匹配答案的目标问题;

获取各个潜在答案;

分别对所述目标问题和所述各个潜在答案进行向量化处理,得到目标问题向量和各个潜在答案向量;

针对每个潜在答案向量,将所述目标问题向量和所述每个潜在答案向量作为输入投入

至预先训练好的深度学习模型，得到所述每个潜在答案向量对应的匹配概率，所述匹配概率表征了所述每个潜在答案向量对应的潜在答案与所述目标问题匹配的概率；

从所述各个潜在答案向量中选取匹配概率最大的一个潜在答案向量；

将选取出的所述一个潜在答案向量对应的潜在答案确定为所述目标问题的最终答案。

12. 根据权利要求 11 所述的计算机设备，其特征在于，所述深度学习模型包括分层循环编码器、特征聚类网络和相似度计算网络三部分，所述深度学习模型通过以下步骤预先训练好：

收集多对问答样本，每对所述问答样本包括一个样本问题和与所述样本问题对应的标准答案；

针对每对问答样本，对所述每对问答样本中的样本问题和标准答案进行向量化处理，得到所述每对问答样本的样本问题向量和标准答案向量；

将所述每对问答样本的样本问题向量和标准答案向量分别输入分层循环编码器进行编码，得到问题特征向量和答案特征向量，所述分层循环编码器由两个单层 GRU 神经网络组成；

将所述每对问答样本的答案特征向量输入特征聚类网络进行特征聚类计算，得到聚类更新后的答案特征向量；

针对每对问答样本，通过相似度计算网络计算所述每对问答样本的问题特征向量与所述聚类更新后的答案特征向量之间的相似度，得到所述每对问答样本中样本问题与标准答案的相似度值，作为样本匹配概率；

以所述每对问答样本对应的样本匹配概率为调整目标，调整所述分层循环编码器、所述特征聚类网络和所述相似度计算网络的网络参数，使得所述每对问答样本对应的样本匹配概率与 1 之间的误差最小化；

若每对问答样本对应的样本匹配概率与 1 之间的误差满足预设的训练终止条件，则确定所述深度学习模型已训练好。

13. 根据权利要求 12 所述的计算机设备，其特征在于，所述将所述每对问答样本的答案特征向量输入特征聚类网络进行特征聚类计算，得到聚类更新后的答案特征向量包括：

针对各个特征向量块中的每个特征向量块，计算所述每个特征向量块与各个预设特征类别向量之间的相似度，得到各个相似度值，所述各个特征向量块是指在所述每对问答样本中、标准答案的各个字词在所述答案特征向量中的对应部分向量；

将所述各个相似度值分别与各自对应的预设特征类别向量进行加权求和运算，并将运算得到的向量与所述每个特征向量块并联，得到所述每个特征向量块对应的、聚类更新后的特征向量块；

将所述各个特征向量块各自对应的、聚类更新后的特征向量块组合得到所述每对问答样本的、聚类更新后的答案特征向量。

14. 根据权利要求 12 所述的计算机设备，其特征在于，所述针对每个潜在答案向量，将所述目标问题向量和所述每个潜在答案向量作为输入投入至预先训练好的深度学习模型，得到所述每个潜在答案向量对应的匹配概率包括：

针对每个潜在答案向量，将所述目标问题向量和所述每个潜在答案向量分别输入分层循环编码器进行编码，得到目标问题特征向量和目标答案特征向量；

将所述每个潜在答案向量对应的目标答案特征向量输入特征聚类网络进行特征聚类计算，得到聚类更新后的目标答案特征向量；

通过相似度计算网络计算所述目标问题向量与所述每个潜在答案向量对应的、聚类更新后的目标答案特征向量之间的相似度，得到所述目标问题与所述每个潜在答案的相似度值，作为所述每个潜在答案向量对应的匹配概率。

15. 根据权利要求 11 至 14 中任一项所述的计算机设备, 其特征在于, 所述分别对所述目标问题和所述各个潜在答案进行向量化处理, 得到目标问题向量和各个潜在答案向量包括:

针对每个目标文本, 将所述每个目标文本中的各个目标字词分别转换为 GloVe 词向量, 得到初始向量, 所述目标文本为所述目标问题或所述潜在答案;

判断所述各个目标字词是否均被 GloVe 词向量覆盖;

若所述各个目标字词均被 GloVe 词向量覆盖, 则确定所述初始向量为所述每个目标文本对应目标向量, 所述目标向量为所述目标问题向量或所述潜在答案向量;

若所述各个目标字词中任一个目标字词未被 GloVe 词向量覆盖, 则将未被 GloVe 词向量覆盖的目标字词转换为 TransE 词向量, 得到补充向量;

将所述补充向量添加至所述初始向量, 得到所述每个目标文本对应目标向量。

16. 一个或多个存储有计算机可读指令的可读存储介质, 其特征在于, 所述计算机可读指令被一个或多个处理器执行时, 使得所述一个或多个处理器执行如下步骤:

获取待匹配答案的目标问题;

获取各个潜在答案;

分别对所述目标问题和所述各个潜在答案进行向量化处理, 得到目标问题向量和各个潜在答案向量;

针对每个潜在答案向量, 将所述目标问题向量和所述每个潜在答案向量作为输入投入至预先训练好的深度学习模型, 得到所述每个潜在答案向量对应的匹配概率, 所述匹配概率表征了所述每个潜在答案向量对应的潜在答案与所述目标问题匹配的概率;

从所述各个潜在答案向量中选取匹配概率最大的一个潜在答案向量;

将选取出的所述一个潜在答案向量对应的潜在答案确定为所述目标问题的最终答案。

17. 根据权利要求 16 所述的可读存储介质, 其特征在于, 所述深度学习模型包括分层循环编码器、特征聚类网络和相似度计算网络三部分, 所述深度学习模型通过以下步骤预先训练好:

收集多对问答样本, 每对所述问答样本包括一个样本问题和与所述样本问题对应的标准答案;

针对每对问答样本, 对所述每对问答样本中的样本问题和标准答案进行向量化处理, 得到所述每对问答样本的样本问题向量和标准答案向量;

将所述每对问答样本的样本问题向量和标准答案向量分别输入分层循环编码器进行编码, 得到问题特征向量和答案特征向量, 所述分层循环编码器由两个单层 GRU 神经网络组成;

将所述每对问答样本的答案特征向量输入特征聚类网络进行特征聚类计算, 得到聚类更新后的答案特征向量;

针对每对问答样本, 通过相似度计算网络计算所述每对问答样本的问题特征向量与所述聚类更新后的答案特征向量之间的相似度, 得到所述每对问答样本中样本问题与标准答案的相似度值, 作为样本匹配概率;

以所述每对问答样本对应的样本匹配概率为调整目标, 调整所述分层循环编码器、所述特征聚类网络和所述相似度计算网络的网络参数, 使得所述每对问答样本对应的样本匹配概率与 1 之间的误差最小化;

若每对问答样本对应的样本匹配概率与 1 之间的误差满足预设的训练终止条件, 则确定所述深度学习模型已训练好。

18. 根据权利要求 17 所述的可读存储介质, 其特征在于, 所述将所述每对问答样本的答案特征向量输入特征聚类网络进行特征聚类计算, 得到聚类更新后的答案特征向量包括:

针对各个特征向量块中的每个特征向量块，计算所述每个特征向量块与各个预设特征类别向量之间的相似度，得到各个相似度值，所述各个特征向量块是指在所述每对问答样本中、标准答案的各个字词在所述答案特征向量中的对应部分向量；

将所述各个相似度值分别与各自对应的预设特征类别向量进行加权求和运算，并将运算得到的向量与所述每个特征向量块并联，得到所述每个特征向量块对应的、聚类更新后的特征向量块；

将所述各个特征向量块各自对应的、聚类更新后的特征向量块组合得到所述每对问答样本的、聚类更新后的答案特征向量。

19. 根据权利要求 17 所述的可读存储介质，其特征在于，所述针对每个潜在答案向量，将所述目标问题向量和所述每个潜在答案向量作为输入投入至预先训练好的深度学习模型，得到所述每个潜在答案向量对应的匹配概率包括：

针对每个潜在答案向量，将所述目标问题向量和所述每个潜在答案向量分别输入分层循环编码器进行编码，得到目标问题特征向量和目标答案特征向量；

将所述每个潜在答案向量对应的目标答案特征向量输入特征聚类网络进行特征聚类计算，得到聚类更新后的目标答案特征向量；

通过相似度计算网络计算所述目标问题向量与所述每个潜在答案向量对应的、聚类更新后的目标答案特征向量之间的相似度，得到所述目标问题与所述每个潜在答案的相似度值，作为所述每个潜在答案向量对应的匹配概率。

20. 根据权利要求 16 至 19 中任一项所述的可读存储介质，其特征在于，所述分别对所述目标问题和所述各个潜在答案进行向量化处理，得到目标问题向量和各个潜在答案向量包括：

针对每个目标文本，将所述每个目标文本中的各个目标字词分别转换为 GloVe 词向量，得到初始向量，所述目标文本为所述目标问题或所述潜在答案；

判断所述各个目标字词是否均被 GloVe 词向量覆盖；

若所述各个目标字词均被 GloVe 词向量覆盖，则确定所述初始向量为所述每个目标文本对应目标向量，所述目标向量为所述目标问题向量或所述潜在答案向量；

若所述各个目标字词中任一个目标字词未被 GloVe 词向量覆盖，则将未被 GloVe 词向量覆盖的目标字词转换为 TransE 词向量，得到补充向量；

将所述补充向量添加至所述初始向量，得到所述每个目标文本对应目标向量。

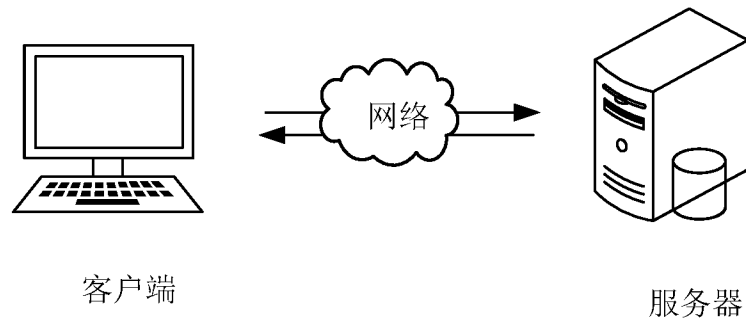


图 1

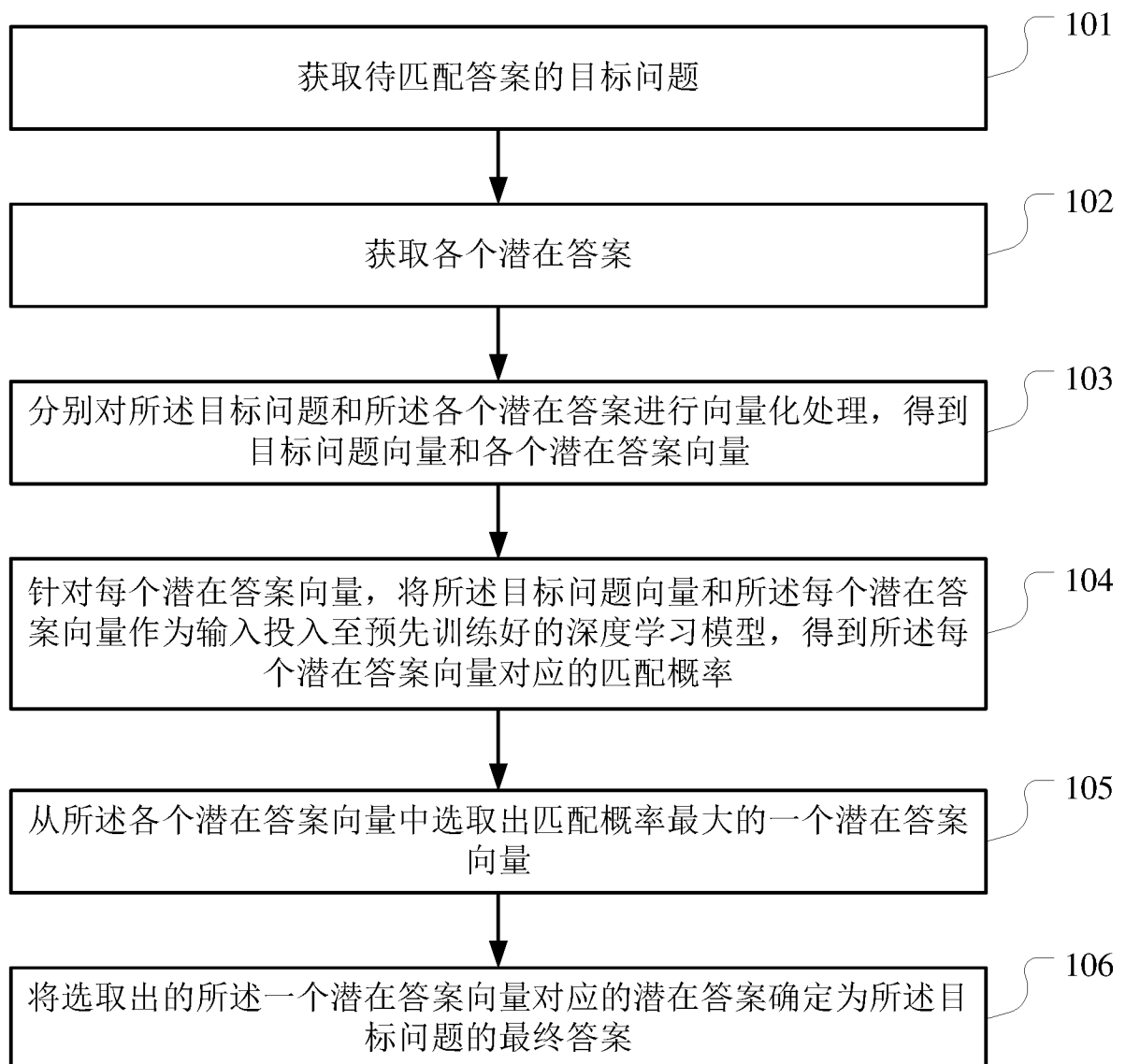


图 2

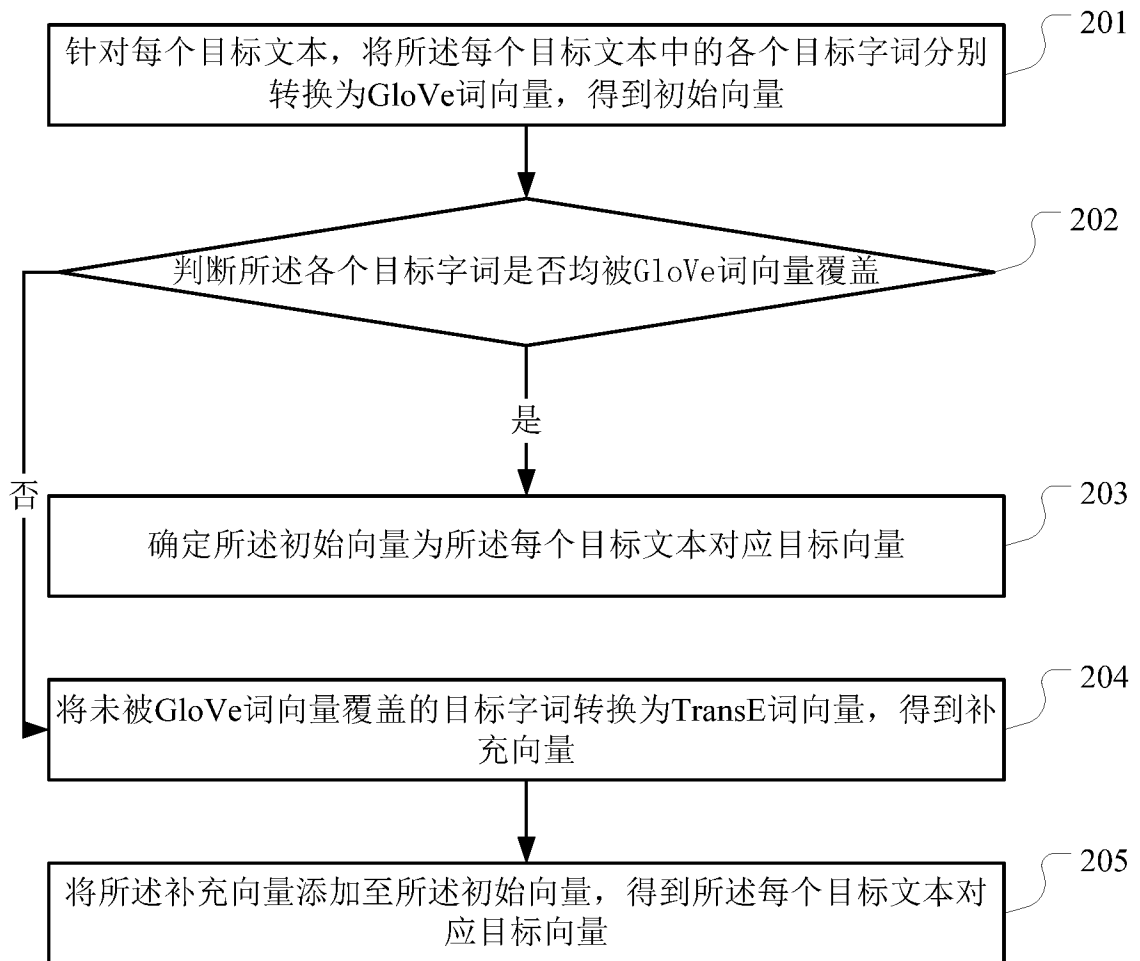


图 3

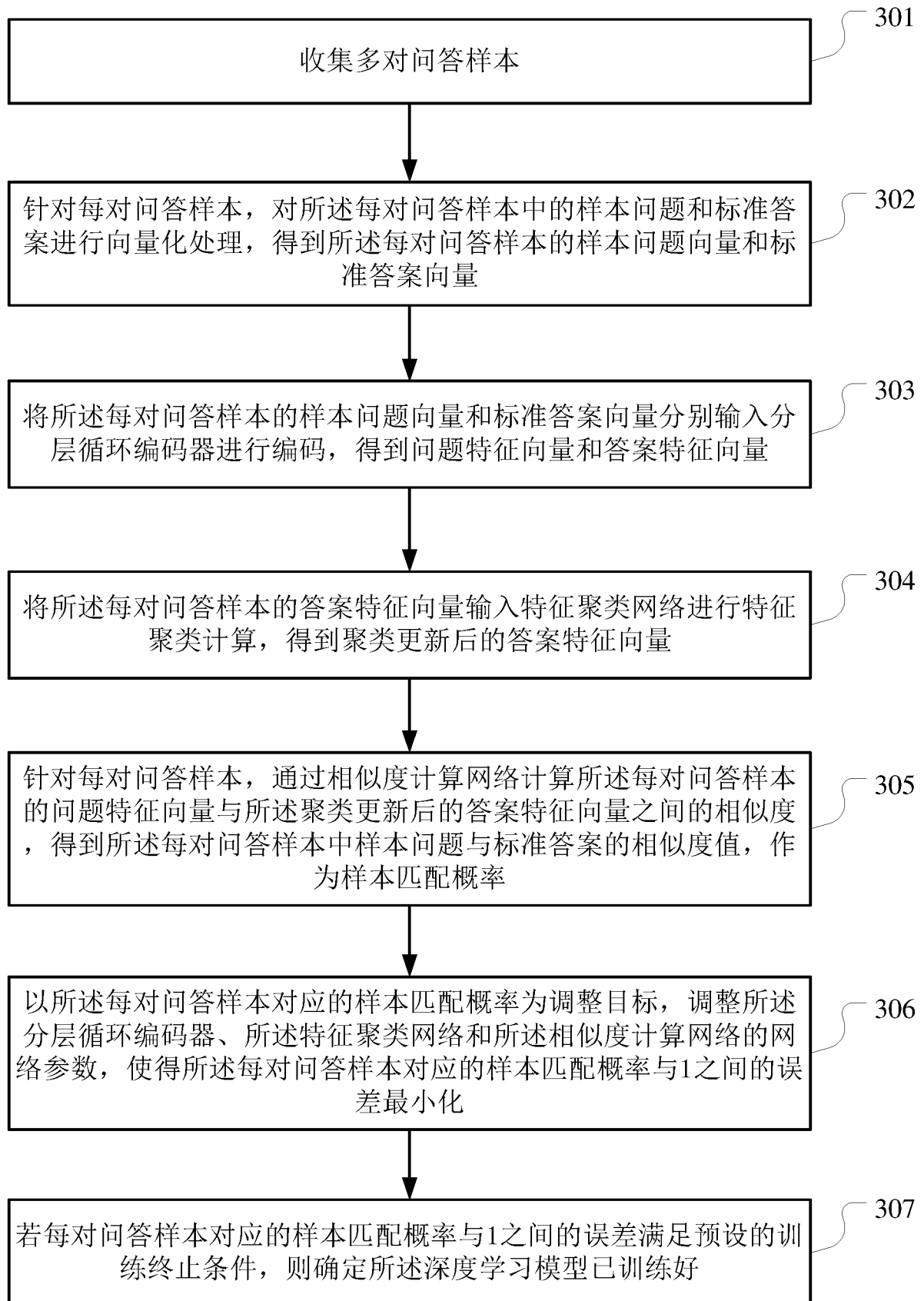


图 4

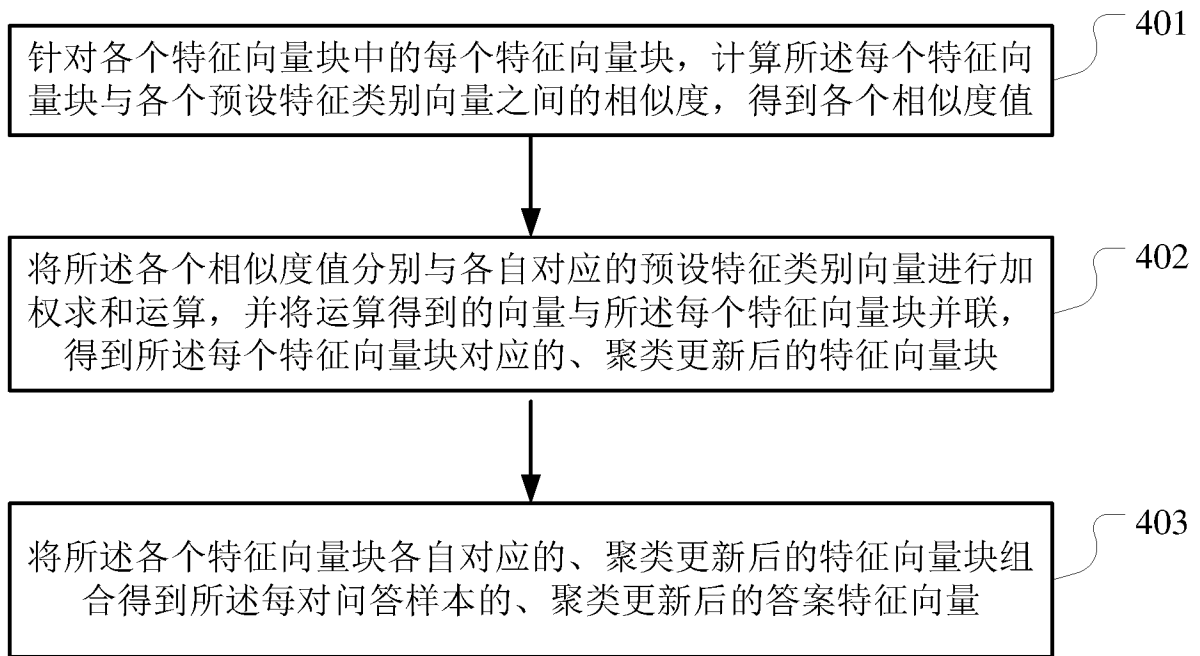


图 5

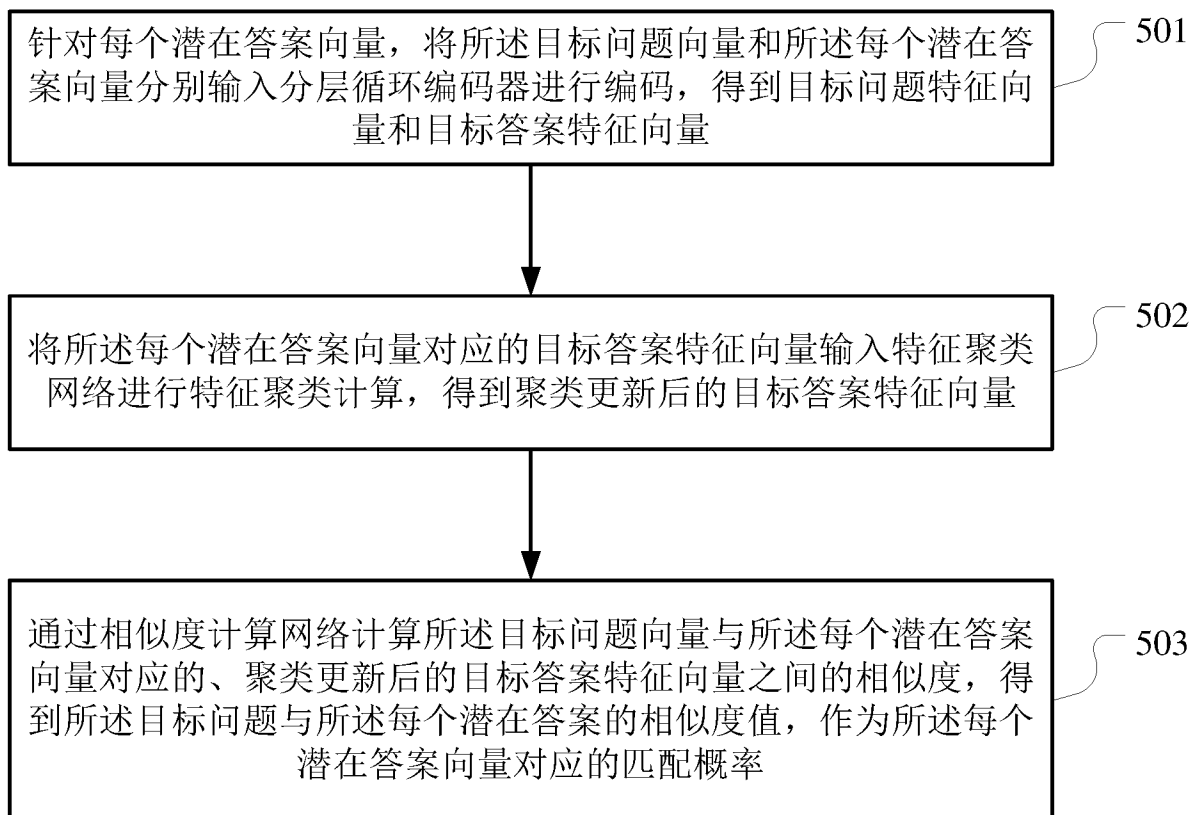


图 6

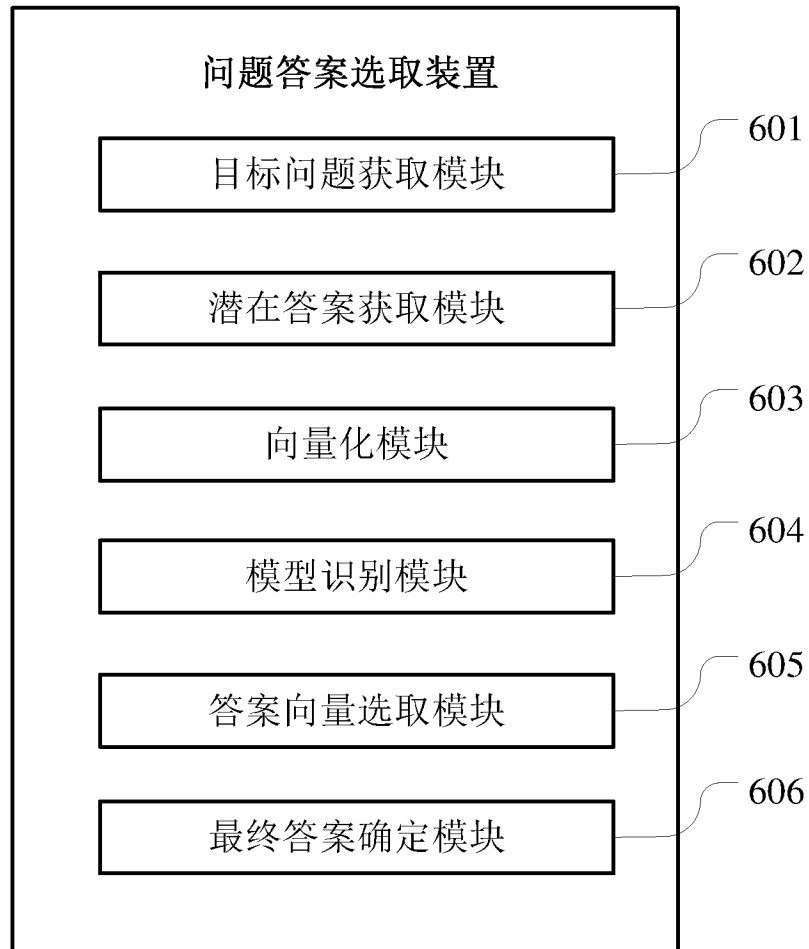


图 7

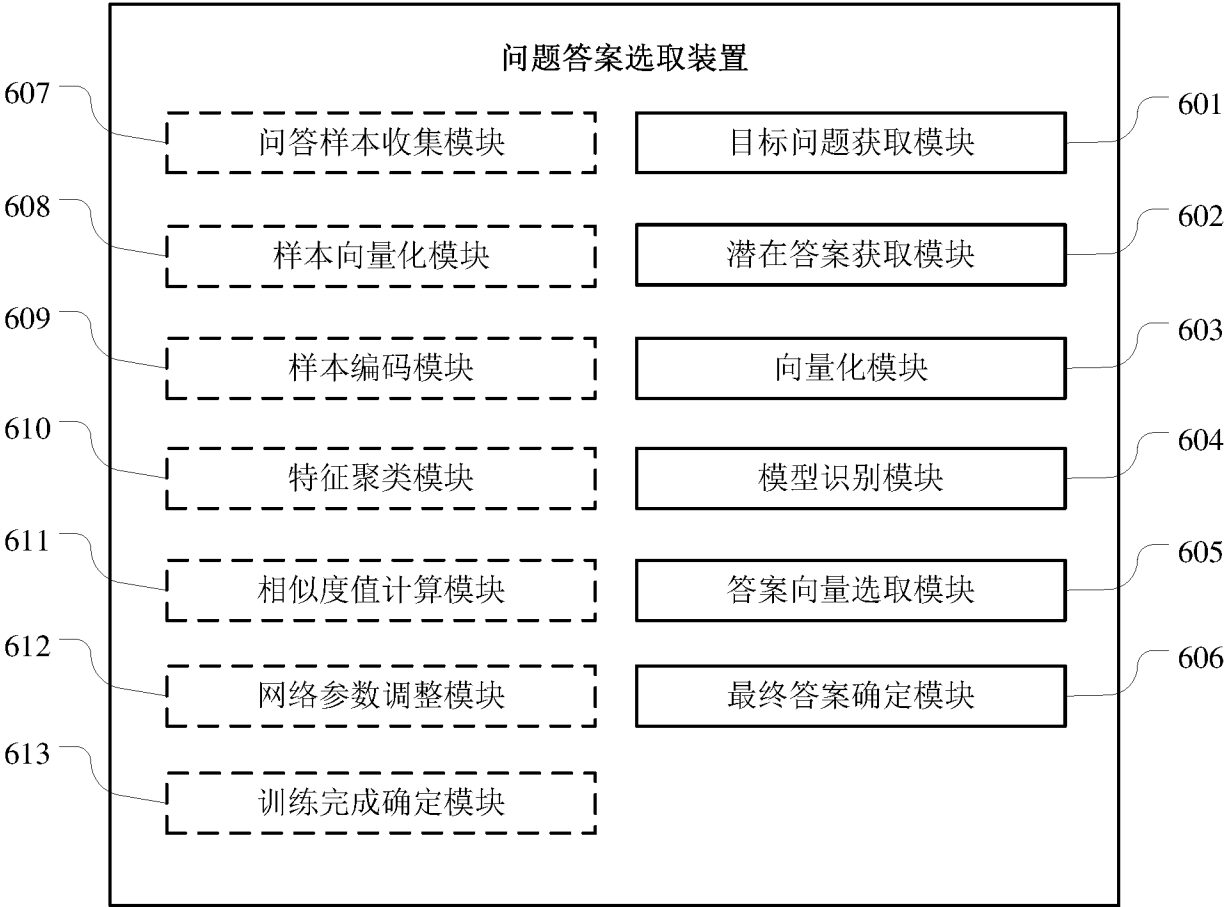


图 8

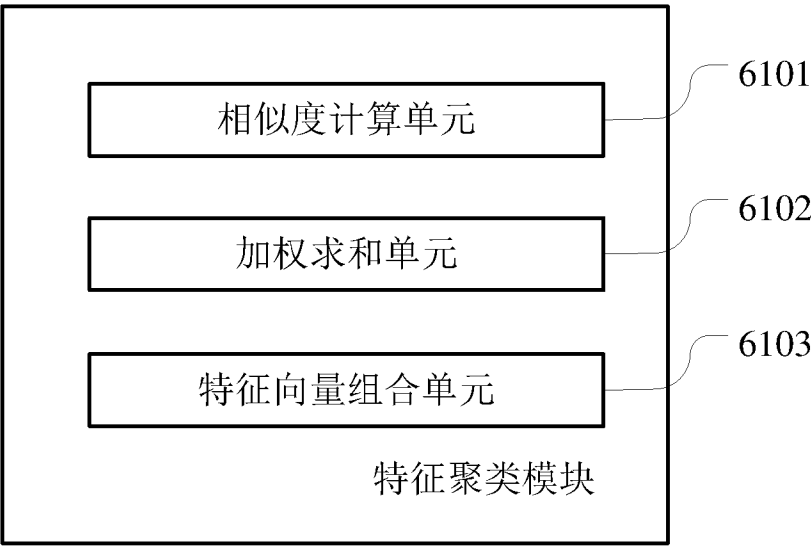


图 9

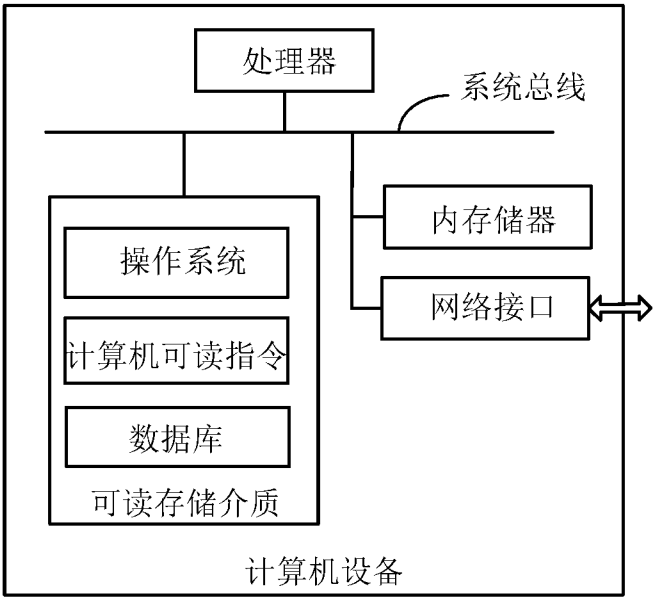


图 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/102923

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/332(2019.01)i; G06F 16/35(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F; H04L; H04W

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI, EPODOC, CNPAT, CNKI: 问题, 答案, 潜在, 候选, 备选, 目标, 向量, 匹配, 对应, 最佳, 答案, 解答, 模型, 学习, 深度, 概率, 编码, 聚类, 相似度, 神经网络, 特征, GloVe, question, answer, match, vector, probability, model, character

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 109740077 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 10 May 2019 (2019-05-10) description, paragraphs [0027]-[0076]	1-20
X	CN 109271505 A (SHENZHEN INTELLIGENT STRONG TECHNOLOGY CO., LTD.) 25 January 2019 (2019-01-25) description, paragraphs [0043]-[0078]	1-20
X	CN 108345585 A (ZHEJIANG UNIVERSITY) 31 July 2018 (2018-07-31) description, paragraphs [0033]-[0059]	1-20
A	US 2019065600 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 28 February 2019 (2019-02-28) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

06 January 2020

Date of mailing of the international search report

23 January 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/102923

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109740077	A	10 May 2019	None			
CN	109271505	A	25 January 2019	None			
CN	108345585	A	31 July 2018	None			
US	2019065600	A1	28 February 2019	US	2019065599	A1	28 February 2019

国际检索报告

国际申请号

PCT/CN2019/102923

A. 主题的分类 G06F 16/332(2019.01)i; G06F 16/35(2019.01)i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类																	
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) G06F; H04L; H04W 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) WPI, EPDOC, CNPAT, CNKI: 问题, 答案, 潜在, 候选, 备选, 目标, 向量, 匹配, 对应, 最佳, 答案, 解答, 模型, 学习, 深度, 概率, 编码, 聚类, 相似度, 神经网络, 特征, GloVe, question, answer, match, vector, probability, model, character																	
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 109740077 A (北京百度网讯科技有限公司) 2019年 5月 10日 (2019 - 05 - 10) 说明书第[0027]-[0076]段</td> <td>1-20</td> </tr> <tr> <td>X</td> <td>CN 109271505 A (深圳智能思创科技有限公司) 2019年 1月 25日 (2019 - 01 - 25) 说明书第[0043]-[0078]段</td> <td>1-20</td> </tr> <tr> <td>X</td> <td>CN 108345585 A (浙江大学) 2018年 7月 31日 (2018 - 07 - 31) 说明书第[0033]-[0059]段</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>US 2019065600 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2019年 2月 28日 (2019 - 02 - 28) 全文</td> <td>1-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 109740077 A (北京百度网讯科技有限公司) 2019年 5月 10日 (2019 - 05 - 10) 说明书第[0027]-[0076]段	1-20	X	CN 109271505 A (深圳智能思创科技有限公司) 2019年 1月 25日 (2019 - 01 - 25) 说明书第[0043]-[0078]段	1-20	X	CN 108345585 A (浙江大学) 2018年 7月 31日 (2018 - 07 - 31) 说明书第[0033]-[0059]段	1-20	A	US 2019065600 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2019年 2月 28日 (2019 - 02 - 28) 全文	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
X	CN 109740077 A (北京百度网讯科技有限公司) 2019年 5月 10日 (2019 - 05 - 10) 说明书第[0027]-[0076]段	1-20															
X	CN 109271505 A (深圳智能思创科技有限公司) 2019年 1月 25日 (2019 - 01 - 25) 说明书第[0043]-[0078]段	1-20															
X	CN 108345585 A (浙江大学) 2018年 7月 31日 (2018 - 07 - 31) 说明书第[0033]-[0059]段	1-20															
A	US 2019065600 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2019年 2月 28日 (2019 - 02 - 28) 全文	1-20															
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																	
<table border="0"> <tr> <td> * 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 </td> <td> “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件 </td> </tr> </table>			* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件													
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																
国际检索实际完成的日期 2020年 1月 6日		国际检索报告邮寄日期 2020年 1月 23日															
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		受权官员 杜少凤 电话号码 86-(10)-53961593															

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/102923

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	109740077	A	2019年 5月 10日	无	
CN	109271505	A	2019年 1月 25日	无	
CN	108345585	A	2018年 7月 31日	无	
US	2019065600	A1	2019年 2月 28日	US 2019065599 A1	2019年 2月 28日