



(12) 发明专利申请

(10) 申请公布号 CN 116610853 A

(43) 申请公布日 2023. 08. 18

(21) 申请号 202210118904.9

G06N 3/084 (2023.01)

(22) 申请日 2022.02.08

(71) 申请人 中兴通讯股份有限公司

地址 518057 广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦

(72) 发明人 刘杨

(74) 专利代理机构 深圳市力道知识产权代理事

务所(普通合伙) 44507

专利代理师 张传义

(51) Int. Cl.

G06F 16/9535 (2019.01)

G06F 16/951 (2019.01)

G06F 16/906 (2019.01)

G06F 40/289 (2020.01)

G06N 3/045 (2023.01)

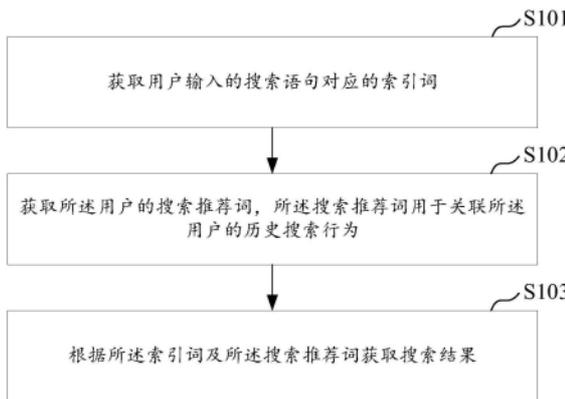
权利要求书2页 说明书10页 附图5页

(54) 发明名称

搜索推荐方法、搜索推荐系统、计算机设备及存储介质

(57) 摘要

本发明实施例提供一种搜索推荐方法、搜索推荐系统、计算机设备及存储介质,属于搜索技术领域。该方法包括:获取用户输入的搜索语句对应的索引词;获取所述用户的搜索推荐词,所述搜索推荐词用于关联所述用户的历史搜索行为;根据所述索引词及所述搜索推荐词获取搜索结果。本发明实施例的技术方案通过在搜索时关联用户的历史搜索偏好改进搜索结果,并对搜索结果的排序进行优化,从而能够更准确地判断用户的搜索意图,提升了用户的体验。



1. 一种搜索推荐方法,其特征在于,包括:
获取用户输入的搜索语句对应的索引词;
获取所述用户的搜索推荐词,所述搜索推荐词用于关联所述用户的历史搜索行为;
根据所述索引词及所述搜索推荐词获取搜索结果。
2. 根据权利要求1所述的搜索推荐方法,其特征在于,所述根据所述索引词及所述搜索推荐词获取搜索结果之后,所述方法还包括:
对所述搜索结果进行相关性评分,并依据得到的相关性分数对所述搜索结果进行排序得到推荐结果。
3. 根据权利要求2所述的搜索推荐方法,其特征在于,所述方法还包括:
获取所述用户的偏好分类,依据所述偏好分类对所述推荐结果进行优化,得到优化后的推荐结果;
其中,依据所述偏好分类对所述推荐结果进行优化包括:
对属于所述偏好分类的搜索结果的相关性分数进行加权,并依据得到的所述加权后的相关性分数对所述搜索结果进行排序得到优化后的推荐结果。
4. 根据权利要求3所述的搜索推荐方法,其特征在于,所述方法还包括:
收集所述用户的信息和搜索历史,基于预设的BP神经网络模型分析所述用户的信息和搜索历史得到所述用户的搜索偏好;
基于所述搜索偏好更新所述用户的搜索推荐词及偏好分类。
5. 根据权利要求1所述的搜索推荐方法,其特征在于,所述获取用户输入的搜索语句对应的索引词,包括:
将用户输入的搜索语句与预设的字典树进行分词匹配,得到与所述搜索语句对应的索引词。
6. 根据权利要求5所述的搜索推荐方法,其特征在于,所述方法还包括:
对预设的知识库中的文章进行分类,依据将所述文章的内容进行分词的结果与所述文章所属的类别构造所述预设的字典树。
7. 根据权利要求2所述的搜索推荐方法,其特征在于,所述根据所述索引词及所述搜索推荐词获取搜索结果,还包括:
从预设的关系数据库中查找与所述索引词关联的同义词;
根据所述索引词、所述同义词以及所述搜索推荐词获取搜索结果。
8. 根据权利要求7所述的搜索推荐方法,其特征在于,所述对所述搜索结果进行相关性评分包括:
为所述索引词、所述同义词及所述搜索推荐词设置不同权重对所述搜索结果进行相关性评分,得到所述搜索结果的相关性分数。
9. 一种搜索推荐系统,其特征在于,包括:
索引词获取模块,用于获取用户输入的搜索语句对应的索引词;
用户偏好模块,用于获取所述用户的搜索推荐词,所述搜索推荐词用于关联所述用户的历史搜索行为;
搜索模块,用于根据所述索引词及所述搜索推荐词获取搜索结果。
10. 一种计算机设备,其特征在于,所述计算机设备包括处理器、存储器、存储在所述存

存储器上并可被所述处理器执行的计算机程序以及用于实现所述处理器和所述存储器之间的连接通信的数据总线,其中所述计算机程序被所述处理器执行时,实现如权利要求1至8中任一项所述的搜索推荐方法的步骤。

11.一种存储介质,用于计算机可读存储,其特征在于,所述存储介质存储有一个或者多个程序,所述一个或者多个程序可被一个或者多个处理器执行,以实现权利要求1至8中任一项所述的搜索推荐方法的步骤。

搜索推荐方法、搜索推荐系统、计算机设备及存储介质

技术领域

[0001] 本发明涉及搜索技术领域,尤其涉及一种搜索推荐方法、搜索推荐系统、计算机设备及存储介质。

背景技术

[0002] 随着电信行业的不断发展,业务知识不断更新,电信客服人员为获取相关业务信息,经常通过统一的知识管理系统进行检索,且根据工作岗位、习惯等因素影响会经常搜索同类型知识。精准、高效的业务信息检索,将最符合搜索预期的目标文章推荐给搜索人员显得十分重要。

[0003] 当前搜索排序的方法如下:用户输入关键词作为搜索词后,系统后台服务器直接将待搜索词与知识库中的检索词进行匹配,获得多个搜索目标文章。然后,针对每篇文章,按照“相关性评分”计算出每篇文章的排序分值,根据分值将这些文章进行排序并推荐给用户。其中,相关性得分表征用户的搜索词与该文章的契合程度,由系统服务器利用特定的算法计算。然而对于一些用户的搜索语句与搜索关键词不匹配时,则无法准确判断用户的搜索意图,导致搜索结果不能够令用户满意。

发明内容

[0004] 本发明实施例的主要目的在于提供一种搜索推荐方法、搜索推荐系统、计算机设备及存储介质,旨在改善现有的知识搜索无法与用户的历史行为关联,无法准确判断用户搜索意图导致搜索结果正确率不足的问题,提升了用户的体验。

[0005] 第一方面,本发明实施例提供一种搜索推荐方法,包括:

[0006] 获取用户输入的搜索语句对应的索引词;获取所述用户的搜索推荐词,所述搜索推荐词用于关联所述用户的历史搜索行为;根据所述索引词及所述搜索推荐词获取搜索结果。

[0007] 第二方面,本发明实施例还提供一种搜索推荐系统,包括:

[0008] 索引词获取模块,用于获取用户输入的搜索语句对应的索引词;用户偏好模块,用于获取所述用户的搜索推荐词,所述搜索推荐词用于关联所述用户的历史搜索行为;搜索模块,用于根据所述索引词及所述搜索推荐词获取搜索结果。

[0009] 第三方面,本发明实施例还提供一种计算机设备,所述计算机设备包括处理器、存储器、存储在所述存储器上并可被所述处理器执行的计算机程序以及用于实现所述处理器和所述存储器之间的连接通信的数据总线,其中所述计算机程序被所述处理器执行时,实现如本发明说明书提供的任一项搜索推荐方法的步骤。

[0010] 第四方面,本发明实施例还提供一种存储介质,用于计算机可读存储,所述存储介质存储有一个或者多个程序,所述一个或者多个程序可被一个或者多个处理器执行,以实现如本发明说明书提供的任一项搜索推荐的方法的步骤。

[0011] 本发明实施例提供了一种搜索推荐方法、搜索推荐系统、计算机设备及存储介质,

本发明实施例提供的搜索推荐方法,通过本申请实施例提供的搜索推荐方法,通过对搜索语句基于知识库字典树提取所包含的索引词,并获取用户的搜索推荐词进行搜索,对搜索语句进行了扩展,获得了更多符合用户搜索意图的结果。另一方面,将代表用户搜索偏好的搜索推荐词以及偏好类别与搜索语句包含的索引词、分类信息一并用于对文章进行相关性评分并根据评分结果进行排序,从而使得更符合用户搜索意图的文章会显示在前。更进一步地,通过对用户搜索和浏览数据的深度学习,预测和分析用户的搜索偏好,并将分析结果用于修正用户的搜索偏好,使得用户的搜索偏好能根据搜索历史智能调整。获得了更符合用户意图的搜索推荐结果,提升了用户的满意度。

附图说明

[0012] 为了更清楚地说明本申请实施例技术方案,下面将对实施例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的前提下,还可以根据这些附图获得其他的附图。

[0013] 图1为本发明实施例提供的一种搜索推荐方法的流程示意图;

[0014] 图2为本发明实施例在电信行业中的搜索推荐方法的流程示意图;

[0015] 图3为本发明实施例一所构造的字典树的示意图;

[0016] 图4为本发明实施例提供的一种搜索推荐系统的流程示意图;

[0017] 图5为本发明实施例提供的一种计算机设备的结构示意图。

具体实施方式

[0018] 下面将结合本发明实施例中的附图,对本发明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发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0019] 附图中所示的流程图仅是示例说明,不是必须包括所有的内容和操作/步骤,也不是必须按所描述的顺序执行。例如,有的操作/步骤还可以分解、组合或部分合并,因此实际执行的顺序有可能根据实际情况改变。

[0020] 应当理解,在此本发明说明书中所使用的术语仅仅是出于描述特定实施例的目的而并不意在限制本发明。如在本发明说明书和所附权利要求书中所使用的那样,除非上下文清楚地指明其它情况,否则单数形式的“一”、“一个”及“该”意在包括复数形式。

[0021] 本发明实施例提供一种搜索推荐方法、系统、计算机设备及存储介质。其中,该搜索推荐方法可应用于移动终端中,该移动终端可以手机、平板电脑、笔记本电脑、台式电脑、个人数字助理和穿戴式设备等电子设备。

[0022] 下面结合附图,对本发明的一些实施例作详细说明。在不冲突的情况下,下述的实施例及实施例中的特征可以相互组合。

[0023] 请参照图1,图1为本发明实施例提供的一种搜索推荐方法的流程示意图。

[0024] 如图1所示,该搜索推荐方法包括步骤S101至步骤S103。

[0025] 步骤S101、获取用户输入的搜索语句对应的索引词;

[0026] 具体地,将用户输入的搜索语句与预设的字典树进行分词匹配,得到与搜索语句

对应的索引词。首先,构建预设的字典树,通过对预设的知识库中的文章进行分类,依据将文章的内容进行分词的结果与文章所属的类别构造预设的字典树。进一步地,还可以根据预设的字典树中的类别得到索引词对应的业务类别。

[0027] 为了便于解释,以电信行业具体场景下的分类知识索引建立为例进行介绍,需要说明的是,本发明也适用于其它业务场景。

[0028] 具体的,先收集电信行业专业术语、文章,并对文章按业务进行分类;将电信行业专业术语和常用术语加入词库,知识文章的内容分别进行粗粒度分词和细粒度分词并去重得到分词结果,其中,粗粒度分词是语句包含自定义电信行业常用词时,按自定义词进行分词。根据文章分类构建不同类型的知识库索引,构造字典树,将上述分词结果和文章分类信息保存到字典树中。将用户的搜索语句输入预设的字典树中进行匹配,即可得到与搜索语句匹配的索引词。

[0029] 示例性地,可以通过ANSJ分词器对知识文章进行分词。

[0030] 步骤S102、获取用户的搜索推荐词,搜索推荐词用于关联用户的历史搜索行为;

[0031] 具体地,可根据用户的用户ID从用户数据库中获取该用户预先设置的搜索推荐词,并将该搜索推荐词用于与搜索语句所匹配的索引词一并用于搜索。进而在每次搜索中能够关联用户的历史搜索行为,一方面对搜索的输入信息进行了拓展,另一方面,也更多地基于用户的搜索推荐词使得搜索结果更加接近用户的搜索意图。

[0032] 进一步地,为了更好地预测和分析用户的搜索偏好,可以基于神经网络对用户的历史搜索和对文章的浏览行为进行分析,进而根据分析结果对用户的搜索推荐词进行更新和维护。具体地,通过收集用户的信息和搜索和浏览的数据,基于预设的BP神经网络模型分析用户的信息和搜索历史得到用户的搜索偏好,并根据搜索偏好来更新用户的搜索推荐词。

[0033] 步骤S103、根据索引词及搜索推荐词获取搜索结果。

[0034] 根据该索引词及搜索推荐词对知识库中的文章进行搜索,具体的,通过过滤知识库中不包含索引词和搜索推荐词的文章得到所需的搜索结果。

[0035] 进一步地,为了对搜索用的关键词进行进一步的拓展,还可以从预设的关系数据库中查找与索引词关联的同义词,根据索引词、同义词以及搜索推荐词从预设的知识库中得到匹配的搜索结果。

[0036] 为了更好地提升搜索结果对用户的满意度,还可以对搜索结果进行排序,进而实现将与用户搜索意图更相关的搜索结果展示在前面,使用户能更快地定位到所需要的知识文章。

[0037] 具体地,对搜索结果进行相关性评分,并依据得到的相关性分数对搜索结果进行排序得到推荐结果。

[0038] 由于需要对搜索结果进行排序,以便将更满足用户搜索意图的文章排序在更前面的位置。需要对搜索结果进行相关性评分,根据评分结果,分数越高的文章越接近用户的搜索意图,因此按照从高到低的顺序对搜索结果进行排序得到排序后的搜索结果,也即推荐结果展示给用户。

[0039] 进一步地,为了更好地体现搜索词汇在每次搜索中的重要性,通过为索引词、同义词及搜索推荐词设置不同权重对搜索结果进行相关性评分。具体地,索引词是根据用户的

搜索语句匹配出来的搜索实体词,因此设置最高权重,索引词的同义词设置第二高的权重,而用户的搜索推荐词代表的是用户的历史搜索行为,可能与本次搜索的联系不大,因此设置第三高权重。按照索引词的权重>同义词的权重>搜索推荐词的权重的权重设置策略,进而根据三种词汇以及各自的权重对所有搜索结果中的文章进行相关性的评分,并根据相关性分数对这些文章进行排序。

[0040] 示例性地,可基于TF-IDF对搜索结果中的文章进行相关性评分的计算。

[0041] 进一步地,依据偏好分类对推荐结果进行优化,得到优化后的推荐结果;其中,依据偏好分类对推荐结果进行优化包括:对属于偏好分类的搜索结果的相关性分数进行加权,并依据得到的加权后的相关性分数对搜索结果进行排序得到优化后的推荐结果。

[0042] 用户的偏好分类可以由用户预先根据自己的业务需求进行设置,为了更加智能地维护用户也偏好分类,还可以基于神经网络对用户的历史行为和各维度信息进行智能分析,从而智能维护和更新用户的偏好分类情况。

[0043] 示例性地,收集用户的信息包括:采集用户各维度的信息,如用户岗位信息、面向客户类型、工作时段;用户的搜索历史数据包括:用户对文章的浏览顺序、阅读文章的次数、阅读文章的时间、文章是否收藏等,将用户的上述搜索历史数据作为神经网络输入层的神经元。

[0044] 为各个分类建立BP神经网络模型,将上述各维度信息数据作为BP神经网络的输入,在隐含层利用激励函数计算并输出最近时段用户最近阅读文章偏好分值时,计算公式为:

[0045] 其中, x_i 为第 i 个输入的神经元数据, w_i 为第 i 个输入的神经元数据的权重, b_i 为隐含层的偏移量。

$$f = \sum w_i x_i + b_i$$

[0046] 由此,根据BP神经网络结构获取用户最近阅读文章偏好分值,统计最近阅读文章偏好分值高的文章,根据文章的类别、偏好分值、分类文章数量按自定义算法计算出用户偏好分类分值,将分值高的类别作为用户的偏好分类。

[0047] 还可以统计最近阅读文章偏好分值高的文章,根据这部分文章分析出文中频率较高的粗粒度词汇,将此类词汇作为用户的搜索推荐词对用户数据库中的搜索推荐词进行修正。

[0048] 通过对用户搜索和浏览数据的深度学习,预测和分析用户的搜索偏好,并将分析结果用于修正用户的搜索偏好,使得用户的搜索偏好能根据搜索历史智能调整。

[0049] 进一步地,还可以根据用户的历史搜索,针对排名在前面的文章记录用户访问频次,如文章的用户访问频次很少,或者用户不访问,将认为此类文章不符合用户的搜索意图,在排序时对其进行减分操作。

[0050] 进一步地,还可以对用户访问频次很少的文章进行类别分析,识别出用户不喜欢的类别,并将该不喜欢的类别作为对搜索结果进行排序时减分的依据。

[0051] 示例性地,针对推荐给用户的前100篇文章进行访问频次的记录,文章通过文章在数据库中的ID与用户ID进行关联,对访问次数进行计数,在排序时获取该访问次数,访问次数很小或者为0时,对相应的文章进行减分。

[0052] 示例性地,针对这些访问次数很小或者为0的文章,获取这些低频访问文章所属的

类别,当某一个类别的低频访问文章超过一定数量时,即认为该类别为用户的非偏好类别,在对搜索结果进行排序时,对属于非偏好类别的搜索结果进行减分。

[0053] 根据对搜索结果中的文章的分数按照从高到低的顺序进行排序,并将排序后的搜索结果作为推荐结果返回给用户。通过将代表用户搜索偏好的搜索推荐词以及偏好类别与搜索索引词一并用于对文章进行相关性评分并根据评分结果进行排序,从而使得更符合用户搜索意图的文章会显示在前面,方便用户更加快捷地找到所需的文章。

[0054] 本申请实施例提供的搜索推荐方法,通过对搜索语句基于知识库字典树提取所包含的索引词,并获取用户的搜索偏好进行搜索,对搜索语句进行了扩展,获得了更多符合用户搜索意图的结果。另一方面,将代表用户搜索偏好的搜索推荐词以及偏好类别与搜索索引词一并用于对文章进行相关性评分并根据评分结果进行排序,从而使得更符合用户搜索意图的文章会显示在前。更进一步地,通过对用户搜索和浏览数据的深度学习,预测和分析用户的搜索偏好,并将分析结果用于修正用户的搜索偏好,使得用户的搜索偏好能根据搜索历史智能调整。获得了更符合用户意图的搜索推荐结果,提升了用户的满意度。

[0055] 此外,本申请的实施例还提供了关于搜索推荐方法的两个具体实施例,具体如下:

[0056] 实施例一

[0057] 请参照图2和图3,图2为本发明实施例在电信行业中的搜索推荐方法的流程示意图,图3为实施例一所构造的字典树的示意图。该实施例仅为了简单起见,指定了具体场景下的搜索推荐,需要说明的是,本发明也适用于其它场景下的搜索推荐。

[0058] 步骤1,收集电信行业知识文章,例如:“互联网国内数据传送”、“互联网数据中心”、“呼叫中心”。并对文章按内容进行分类:“互联网国内数据传送”属于基础电信业务,“互联网数据中心”、“呼叫中心”属于增值电信业务。

[0059] 步骤2,对知识文章进行分词,对“互联网国内数据传送”的分词结果包括:“互联网”、“国内”、“数据”、“传送”,对“互联网数据应用中心”的分词结果包括:“互联网”、“数据”、“中心”、“数据中心”,对“呼叫中心”的分词结果包括:“呼叫”、“中心”。针对以上知识文章分词后的结果如下表格所示:

索引词	分类 ID	分类
互联网	A1, B1	基础电信业务 增值电信业务
国内	A1	基础电信业务
数据	A1, B1	基础电信业务 增值电信业务
传送	A1	基础电信业务
中心	B1	增值电信业务
数据中心	B1	增值电信业务
呼叫	B1	增值电信业务
中心	B1	增值电信业务

[0061] 依据上述表格所示的分词结果及分类构建如图3所示的字典树。

[0062] 步骤3,获取用户ID(如Test1),并获取用户输入的搜索语句,例如“互联网数据”。

[0063] 步骤4,用户的搜索语句经过字典树节点进行匹配,得到索引词为“互联网”、“数

据”，和匹配的分类“基础电信业务”、“增值电信业务”。

[0064] 步骤5,通过用户ID获取数据库存储的用户搜索偏好中的搜索推荐词,如“国内”。

[0065] 步骤6,根据索引词、搜索推荐词所属分类在不同分类索引进行匹配搜索,得到匹配的文章包括:“互联网国内数据传送”、“互联网数据中心”。

[0066] 步骤7,为搜索词、推荐词汇设置不同的权重值,其中索引词的权重大于搜索推荐词的权重,并根据搜索词、搜索推荐词的权重利用TF-IDF评分算法对匹配文章进行评分,计算词w在文档d中的词频tf即词w在文档d中出现次数count(w,d)和文档d中总词数size(d)的比值: $tf(w,d) = count(w,d) / size(d)$ 。

[0067] “互联网国内数据传送”、“互联网数据中心”、“呼叫中心”统计如下:

tf(w,d)	互联网	数据	国内
互联网国内数据传送	1/4	1/4	1/4
互联网数据中心	1/4	1/4	1/4
呼叫中心	0/2	0/2	0/2

[0069] 词w在整个文档集合中的逆向文档频率idf,即文档总数n与词w所出现文件数docs(w,D)比值的对数 $idf = \log(n/docs(w,D))$ 。

[0070] 例如当文档总数为10000时,可统计出下面数据:

tf(w,d)	互联网	数据	国内
出现次数	200	400	2000
idf	$\log(10000/200)$	$\log(10000/400)$	$\log(10000/2000)$

[0072] Tf-idf模型根据tf和idf为每一个文档d和由关键词w[1]...w[k]组成的查询串q计算一个数值,用于表示查询词串q与文档d的匹配度:

[0073] $tf-idf(q,d) = \sum_{i=1..k} |tf-idf(w[i],d)| = \sum_{i=1..k} |tf(w[i],d) * idf(w[i])|$

[0074] 进一步地,加入了加权系数H,最终查询匹配度为:

[0075] $tf-idf(q,d,H) = \sum_{i=1..k} |tf-idf(w[i],d)| = \sum_{i=1..k} |H * tf(w[i],d) * idf(w[i])|$

[0076] 例如预定偏好分值比例如下:

	索引词	搜索推荐词	偏好分类	不偏好分类
加权系数H	1	0.2	1.2	0.8

	索引词	搜索推荐词	偏好分类	不偏好分类
加权系数H	1	0.2	1.2	0.8
[0078] 互联网国内数据传送	$1/4 * \text{Log}(10000/200) * 1.2$	$1/4 * \text{Log}(10000/400) * 1.2$	$1/4 * \text{Log}(10000/400) * 0.2$	1.0129494044227 391808819863273 81
互联网数据中心	$1/4 * \text{Log}(10000/200)$	$1/4 * \text{Log}(10000/400)$	$1/4 * \text{Log}(10000/400) * 0.2$	0.7742275032520 141035896958289 5662
呼叫中心	$0/2 * \text{Log}(10000/200)$	$0/2 * \text{Log}(10000/400)$	$0/2 * \text{Log}(10000/400) * 0.2$	0

[0079] 所计算的得分如下：

[0080] 根据上述的相关性计算，“互联网国内数据传送”分值大于“互联网数据中心”。

[0081] 进一步地，通过用户ID查询数据库，得到偏好分类如“基础电信业务”，再根据偏好分类“基础电信业务”对搜索结果中属于该偏好分类的文章进行加分处理，对“互联网国内数据传送”进一步进行加分。

[0082] 步骤8，最后将搜索结果按照分数从大到小的分值排序形成搜索时结果列表，最后将搜索结果列表返回给用户。

[0083] 实施例二

[0084] 本发明实施例二提供了用户偏好筛选方法在在电信客服系统中的应用。该实施例仅为了简单起见具体场景下的用户偏好筛选，需要说明的是，本发明也适用于其它应用场景下的偏好筛选。

[0085] 步骤1，采集用户各维度信息，如用户岗位信息(业务咨询、费用查询、业务办理、新业务推销等)、面向客户类型(家庭客户、政企客户、公用电话、无线市话)、工作时段。以及用户对文章的搜索和浏览数据，包括：浏览顺序、阅读文章次数、阅读文章时间、是否收藏等，将用户的各维度信息和搜索浏览数据作为输入层的神经元。

[0086] 步骤2，为各个分类建立BP神经网络模型，将上述各维度信息数据作为BP神经网络的输入，在隐含层利用激励函数计算并输出最近时段阅读文章偏好分值，根据BP神经网络结构获取用户最近阅读文章偏好分值。

[0087] 步骤3，统计最近阅读文章偏好分值高文章，根据文章所属分类、偏好分值、分类文章数量按自定义算法计算出用户偏好分类分值，将分值高的作为用户的偏好分类。

[0088] 步骤4，统计最近阅读文章偏好分值高的文章，根据这部分文章分析出文中相同、频率较高的粗粒度词汇，将此类词汇作为用户搜索推荐词，修正更新用户数据库中的搜索推荐词。

[0089] 步骤5，利用推送给用户的排序条文，继续采集用户阅读行为的数据，包括：匹配文章点击次数、阅读时长、是否收藏等作为步骤1的输入。浏览顺序靠前，多次点击、阅读时间长、收藏某文章则反映了用户喜欢该文章；反之，则说明用户不喜欢该文章。

[0090] 最后，通过基于BP神经网络对用户的历史行为以及的用户信息更新用户的搜索推荐词和偏好分类。

[0091] 请参照图4,图4为实施本实施例提供的搜索推荐系统的一场景示意图,如图4所示,上述实施例提供的搜索推荐系统包括:索引词获取模块201、用户偏好获取模块202,搜索模块203。

[0092] 其中,索引词获取模块201,包括:搜索请求获取模块2011、索引词匹配模块2012。搜索请求获取模块2011用于接收用户的搜索语句,索引词匹配模块2012用于提取输入搜索语句中的索引词与分类。索引词与分类生成的具体方法为:通过搜索语句与预设字典树的匹配,从字典树中提取搜索语句中包含的索引词以及索引词对应的分类信息。

[0093] 用户偏好获取模块202,包括,搜索推荐词获取模块2021和偏好分类获取模块2022,搜索推荐词获取模块2021用于根据用户的ID信息从数据库中获取用户的搜索推荐词;偏好分类获取模块2022用于根据用户的ID信息从数据库获取用户的偏好分类,并将该偏好分类作为后续对搜索结果排序的依据之一。

[0094] 搜索模块203,包括同义词获取模块2031、匹配搜索模块2032。同义词获取模块2031用于获取从搜索语句中提取的索引词的同义词;匹配搜索模块2032用于根据索引词、同义词以及用户的搜索推荐词从知识数据库中进行搜索匹配得到搜索结果。

[0095] 进一步地,搜索推荐系统还包括:排序模块204,具体可分为:权重标记模块2041、相关性评分模块2042、排序推荐模块2043。其中,权重标记模块2041用于对索引词、同义词以及用户的搜索推荐词赋予不同的权重;相关性评分模块2042用于根据索引词、同义词、搜索推荐词以及相应的不同的权重对搜索结果进行相关性评分,其评分结果将用于对搜索结果进行排序的依据。排序推荐模块2043用于根据搜索结果的相关性评分结果以及用户的偏好分类对搜索结果按照分数从高到低的顺序进行排序形成推荐结果,并将搜索的推荐结果返回给用户。

[0096] 进一步地,搜索推荐系统还包括:用户偏好筛选模块205,具体包括:用户信息获取模块2051、搜索历史获取模块2052,用户偏好分析模块2053、用户偏好修正模块2054。其中,用户信息获取模块2051用于收集用户的基础信息;浏览历史获取模块2052用于获取用户的文章搜索和查阅文章的历史数据(如浏览文章顺序、阅读次数、是否收藏);用户偏好分析模块2053用于根据用户的信息和文章的浏览历史结合BP神经网络预测用户对文章的偏好程度,提取偏好程度高的文章,将偏好文章数量和偏好程度汇总与融合提取偏好分类,根据偏好的文章按照预定筛选原则筛选出搜索推荐词;用户偏好修正模块2054用于根据BP神经网络的分析 and 预测结果来更新用户数据库中的偏好分类和搜索推荐词。

[0097] 进一步地,搜索推荐系统还包括:知识库分类索引模块206,具体可分为:知识库分类模块2061、知识库分词模块2062、字典树构造模块2063。其中,知识库分类模块2061用于将知识库的文章按照业务进行分类;知识库分词模块2062用于提取知识库文章中的专业词汇和其他常用自定义词汇作为索引词;字典树构造子模块2063用于根据索引词文章所属分类构造字典树。

[0098] 请参阅图5,图5为本发明实施例提供的一种计算机设备的结构示意图。

[0099] 如图5所示,计算机设备300包括处理器301和存储器302,处理器301和存储器302通过总线303连接,该总线比如为I²C(Integrated Circuit)总线。

[0100] 具体地,处理器301用于提供计算和控制能力,支撑整个计算机设备的运行。处理器301可以是中央处理单元(Central Processing Unit,CPU),该处理器301还可以是其他

通用处理器、数字信号处理器 (Digital Signal Processor, DSP)、专用集成电路 (Application Specific Integrated Circuit, ASIC)、现场可编程门阵列 (Field-Programmable Gate Array, FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。其中,通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

[0101] 具体地,存储器302可以是Flash芯片、只读存储器 (ROM, Read-Only Memory) 磁盘、光盘、U盘或移动硬盘等。

[0102] 本领域技术人员可以理解,图5中示出的结构,仅仅是与本发明实施例方案相关的部分结构的框图,并不构成对本发明实施例方案所应用于其上的计算机设备的限定,具体的服务器可以包括比图中所示更多或更少的部件,或者组合某些部件,或者具有不同的部件布置。

[0103] 其中,处理器用于运行存储在存储器中的计算机程序,并在执行计算机程序时实现本发明实施例提供的任意一种搜索推荐方法。

[0104] 在一实施例中,处理器用于运行存储在存储器中的计算机程序,并在执行计算机程序时实现如下步骤:

[0105] 获取用户输入的搜索语句对应的索引词;

[0106] 获取用户的搜索推荐词,搜索推荐词用于关联用户的历史搜索行为;

[0107] 根据索引词及搜索推荐词获取搜索结果。

[0108] 在一实施例中,处理器在实现搜索推荐方法时,用于实现:对搜索结果进行相关性评分,并依据得到的相关性分数对搜索结果进行排序得到推荐结果。

[0109] 在一实施例中,处理器在实现搜索推荐方法时,用于实现:依据偏好分类对推荐结果进行优化,得到优化后的推荐结果;其中,依据偏好分类对推荐结果进行优化包括:对属于偏好分类的搜索结果的相关性分数进行加权,并依据得到的加权后的相关性分数对搜索结果进行排序得到优化后的推荐结果。

[0110] 在一实施例中,处理器在实现搜索推荐方法时,用于实现:收集用户的信息和搜索历史,基于预设的BP神经网络模型分析用户的信息和搜索历史得到用户的搜索偏好;基于搜索偏好更新用户的搜索推荐词及偏好分类。

[0111] 在一实施例中,处理器在实现获取用户输入的搜索语句对应的索引词时,用于实现:将用户输入的搜索语句与预设的字典树进行分词匹配,得到与搜索语句对应的索引词。

[0112] 在一实施例中,处理器在实现搜索推荐方法时,用于实现:对预设的知识库中的文章进行分类,依据将文章的内容进行分词的结果与文章所属的类别构造预设的字典树。

[0113] 在一实施例中,处理器在实现根据索引词及搜索推荐词获取搜索结果时,用于实现:从预设的关系数据库中查找与索引词关联的同义词;根据索引词、同义词以及搜索推荐词获取搜索结果。

[0114] 在一实施例中,所述处理器在实现对所述搜索结果进行相关性评分时,用于实现:为索引词、同义词及搜索推荐词设置不同权重对搜索结果进行相关性评分,得到搜索结果的相关性分数。

[0115] 需要说明的是,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,上述描述的计算机设备的具体工作过程,可以参考前述搜索推荐方法实施例中的对应过

程,在此不再赘述。

[0116] 本发明实施例还提供一种存储介质,用于计算机可读存储,存储介质存储有一个或者多个程序,一个或者多个程序可被一个或者多个处理器执行,以实现如本发明实施例说明书提供的任一项搜索推荐方法的步骤。

[0117] 其中,存储介质可以是前述实施例的计算机设备的内部存储单元,例如计算机设备的硬盘或内存。存储介质也可以是计算机设备的外部存储设备,例如计算机设备上配备的插接式硬盘,智能存储卡(Smart Media Card,SMC),安全数字(Secure Digital,SD)卡,闪存卡(Flash Card)等。

[0118] 本领域普通技术人员可以理解,上文中所公开方法中的全部或某些步骤、系统、装置中的功能模块/单元可以被实施为软件、固件、硬件及其适当的组合。在硬件实施例中,在以上描述中提及的功能模块/单元之间的划分不一定对应于物理组件的划分;例如,一个物理组件可以具有多个功能,或者一个功能或步骤可以由若干物理组件合作执行。某些物理组件或所有物理组件可以被实施为由处理器,如中央处理器、数字信号处理器或微处理器执行的软件,或者被实施为硬件,或者被实施为集成电路,如专用集成电路。这样的软件可以分布在计算机可读介质上,计算机可读介质可以包括计算机存储介质(或非暂时性介质)和通信介质(或暂时性介质)。如本领域普通技术人员公知的,术语计算机存储介质包括在用于存储信息(诸如计算机可读指令、数据结构、程序模块或其他数据)的任何方法或技术中实施的易失性和非易失性、可移除和不可移除介质。计算机存储介质包括但不限于RAM、ROM、EEPROM、闪存或其他存储器技术、CD-ROM、数字多功能盘(DVD)或其他光盘存储、磁盒、磁带、磁盘存储或其他磁存储装置、或者可以用于存储期望的信息并且可以被计算机访问的任何其他的介质。此外,本领域普通技术人员公知的是,通信介质通常包含计算机可读指令、数据结构、程序模块或者诸如载波或其他传输机制之类的调制数据信号中的其他数据,并且可包括任何信息递送介质。

[0119] 应当理解,在本发明说明书和所附权利要求书中使用的术语“和/或”是指相关联列出的项中的一个或多个的任何组合以及所有可能组合,并且包括这些组合。需要说明的是,在本文中,术语“包括”、“包含”或者其他其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者系统不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者系统所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、方法、物品或者系统中还存在另外的相同要素。

[0120] 上述本发明实施例序号仅仅为了描述,不代表实施例的优劣。以上所述,仅为本发明的具体实施例,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到各种等效的修改或替换,这些修改或替换都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应以权利要求的保护范围为准。

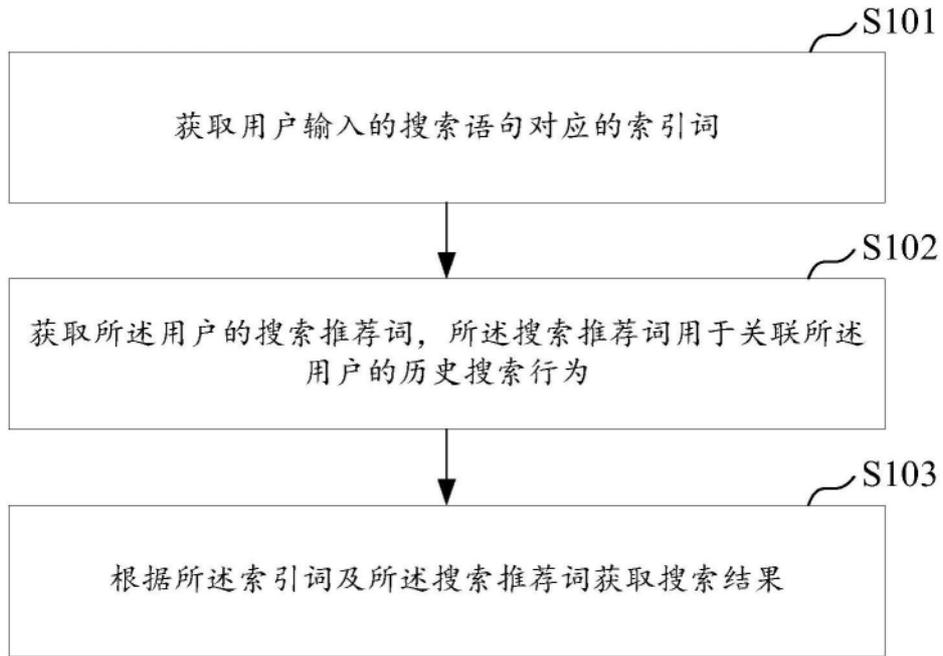


图1

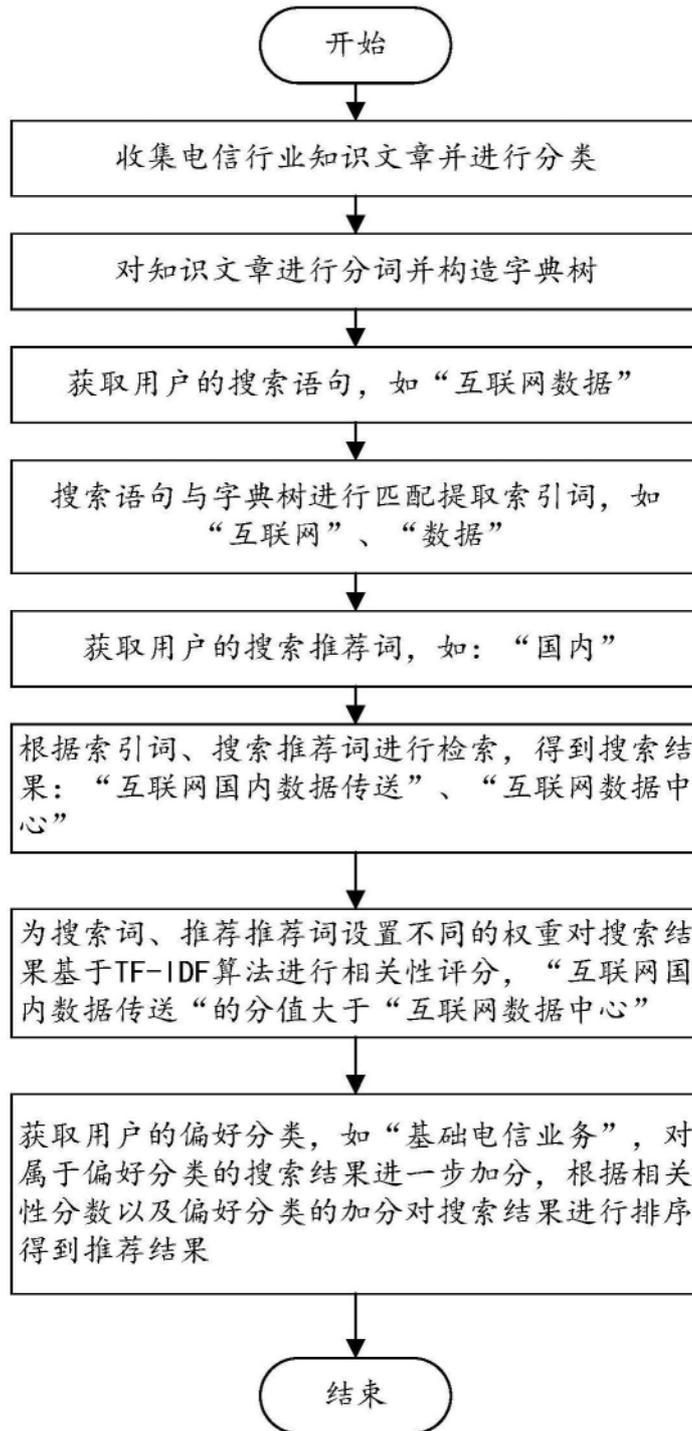


图2

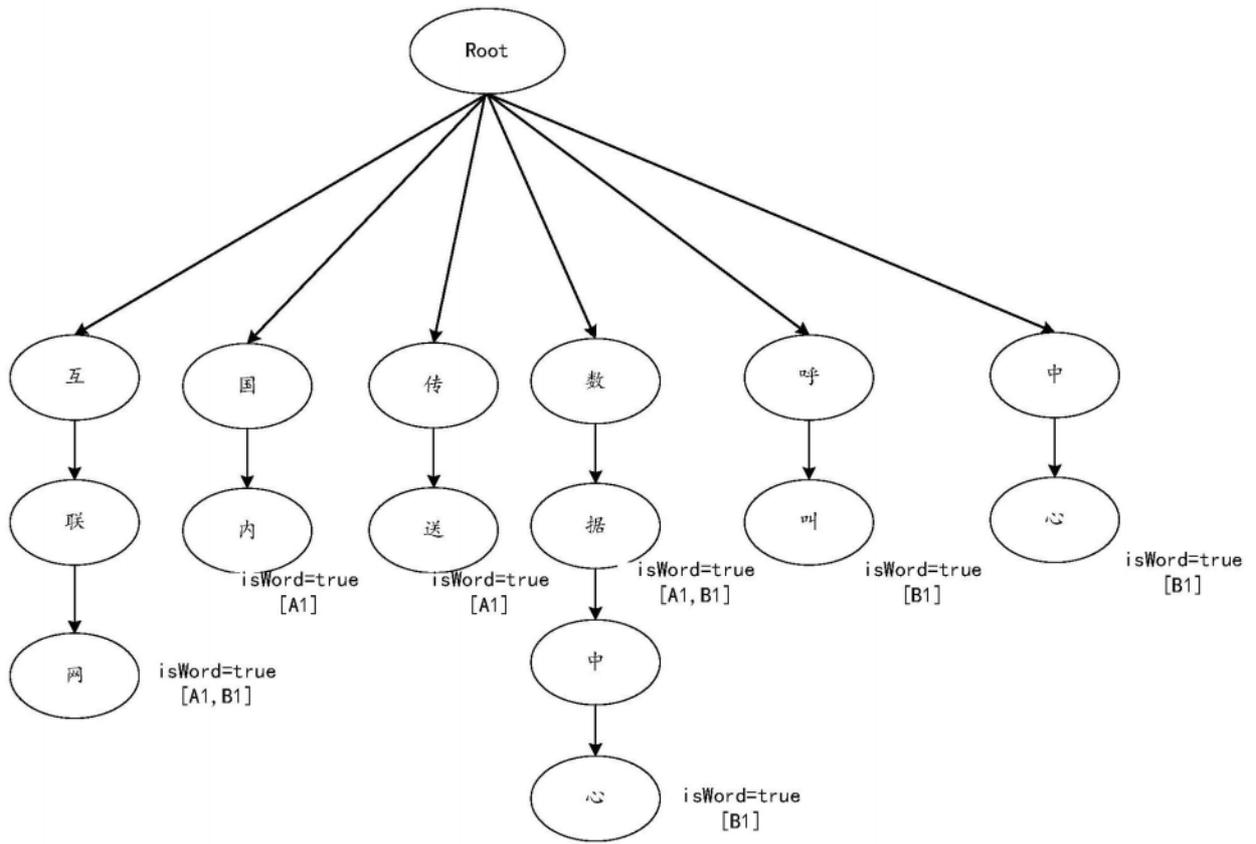


图3

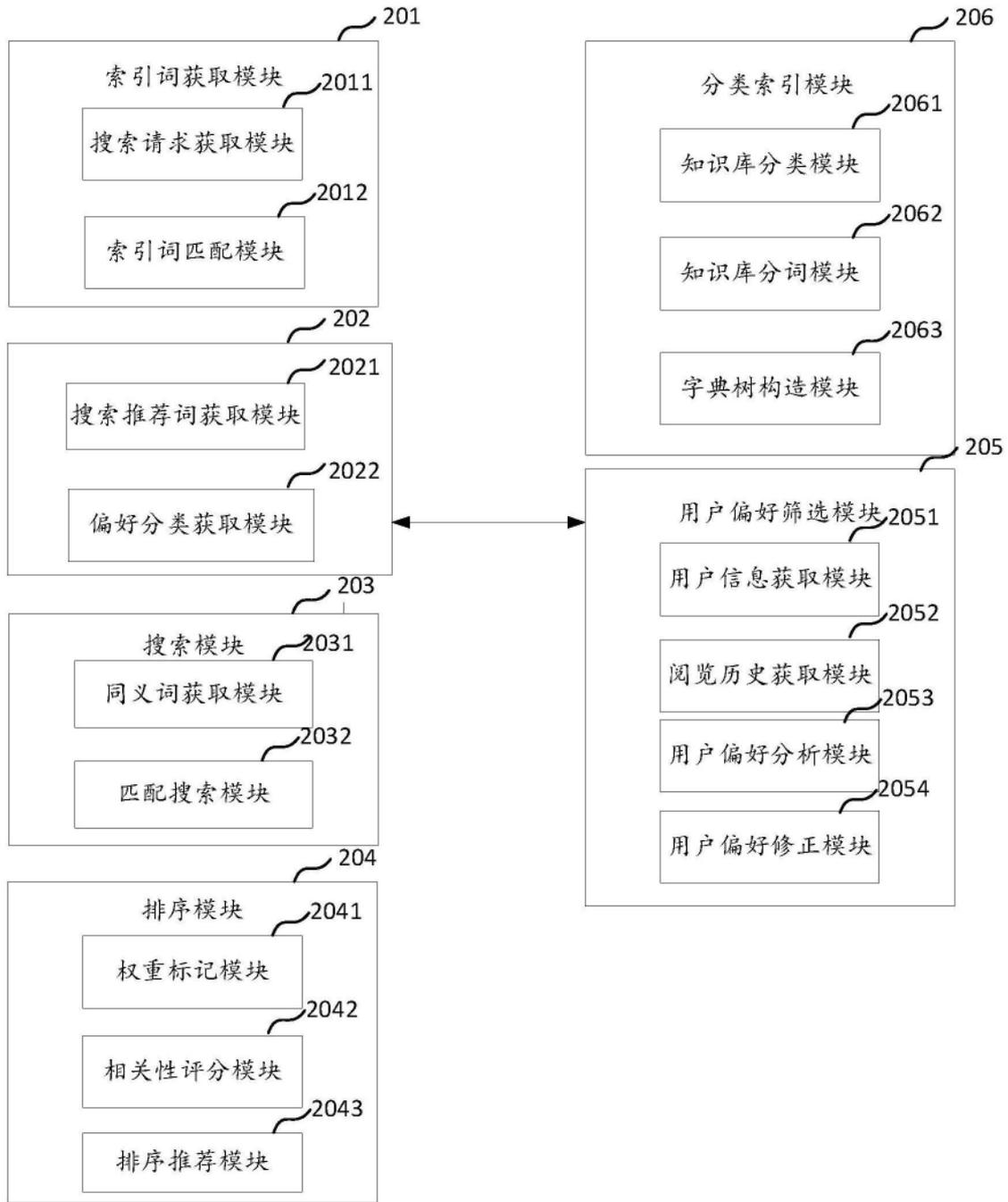


图4

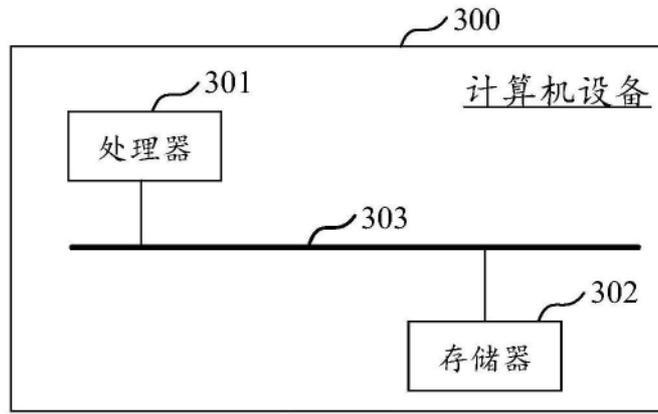


图5