(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 108897685 A
(43)申请公布日 2018.11.27

(21)申请号 201810688272.3

(22)申请日 2018.06.28

(71)申请人 百度在线网络技术(北京)有限公司
地址 100085 北京市海淀区上地十街10号百度大厦三层

(72)发明人 李虓 王轶凡 吴迪 陈冠霖 贾晋康 金慈航

(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 11/36(2006.01)

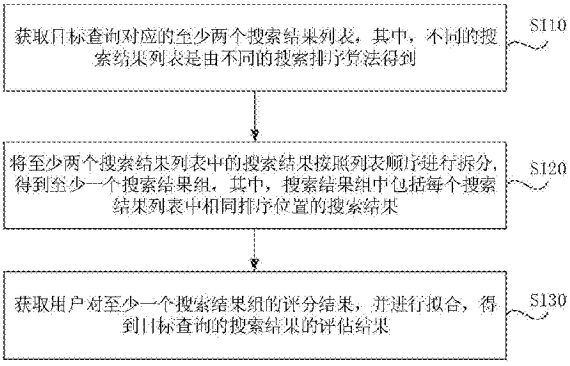
权利要求书2页 说明书10页 附图4页

(54)发明名称

搜索结果的质量评估方法、装置、服务器和介质

(57)摘要

本发明实施例公开了一种搜索结果的质量评估方法、装置、服务器和介质，其中，该方法包括：获取目标查询对应的至少两个搜索结果列表，其中，不同的搜索结果列表是由不同的搜索排序算法得到；将至少两个搜索结果列表中的搜索结果按照列表顺序进行拆分，得到至少一个搜索结果组，其中，搜索结果组中包括每个搜索结果列表中相同排序位置的搜索结果；获取用户对至少一个搜索结果组的评分结果，并进行拟合，得到目标查询的搜索结果的评估结果。本发明实施例解决了现有评估方法中评估成本较高与评估复杂性较大的问题，降低了评估成本和评估难度，提高了评估效率。



CN 108897685 A

1.一种搜索结果的质量评估方法，其特征在于，包括：

获取目标查询对应的至少两个搜索结果列表，其中，不同的搜索结果列表是由不同的搜索排序算法得到；

将所述至少两个搜索结果列表中的搜索结果按照列表顺序进行拆分，得到至少一个搜索结果组，其中，所述搜索结果组中包括每个搜索结果列表中相同排序位置的搜索结果；

获取用户对所述至少一个搜索结果组的评分结果，并进行拟合，得到目标查询的搜索结果的评估结果。

2.根据权利要求1所述的方法，其特征在于，每个搜索结果组中包括至少两个不同的搜索结果。

3.根据权利要求1所述的方法，其特征在于，如果每个搜索结果组中包括两个搜索结果，当任意两个搜索结果组互为对调搜索结果组时，其中，所述互为对调搜索结果组是指搜索结果相同且搜索结果的位置对调的搜索结果组，则相应的，获取用户对所述至少一个搜索结果组的评分结果，包括：

获取用户对所述互为对调搜索结果组中任一个搜索结果组的评分结果，并通过取相反数运算得到所述互为对调搜索结果组中另一个搜索结果组的评分结果。

4.根据权利要求1～3中任一所述的方法，其特征在于，获取用户对所述至少一个搜索结果组的评分结果，并进行拟合，得到目标查询的搜索结果的评估结果，包括：

获取用户对所述至少一个搜索结果组的评分结果；

根据预先确定的搜索结果列表中不同排序位置对应的权重，将所述至少一个搜索结果组的评分结果进行加权求和，得到目标查询的搜索结果的评估结果。

5.根据权利要求4所述的方法，其特征在于，在根据预先确定的搜索结果列表中不同排序位置对应的权重，将所述至少一个搜索结果组的评分结果进行加权求和之前，所述方法还包括：

获取样本查询对应的至少两个样本搜索结果列表；

利用并排评估方法对所述至少两个样本搜索结果列表中的搜索结果进行评估；

将所述至少两个样本搜索结果列表按照列表顺序进行拆分，得到至少一个样本搜索结果组；

获取用户对所述至少一个样本搜索结果组的评分结果；

将所述并排评估方法的评估结果和用户对所述至少一个样本搜索结果组的评分结果作为输入，通过回归训练得到所述权重。

6.一种搜索结果的质量评估装置，其特征在于，包括：

搜索结果列表获取模块，用于获取目标查询对应的至少两个搜索结果列表，其中，不同的搜索结果列表是由不同的搜索排序算法得到；

搜索结果列表拆分模块，用于将所述至少两个搜索结果列表中的搜索结果按照列表顺序进行拆分，得到至少一个搜索结果组，其中，所述搜索结果组中包括每个搜索结果列表中相同排序位置的搜索结果；

评估结果确定模块，用于获取用户对所述至少一个搜索结果组的评分结果，并进行拟合，得到目标查询的搜索结果的评估结果。

7.根据权利要求6所述的装置，其特征在于，所述搜索结果列表拆分模块中每个搜索结

果组中包括至少两个不同的搜索结果。

8.根据权利要求6所述的装置，其特征在于，如果所述搜索结果列表拆分模块中每个搜索结果组中包括两个搜索结果，当任意两个搜索结果组互为对调搜索结果组时，其中，所述互为对调搜索结果组是指搜索结果相同且搜索结果的位置对调的搜索结果组，则相应的，所述评估结果确定模块具体用于：

获取用户对所述互为对调搜索结果组中任一个搜索结果组的评分结果，并通过取相反数运算得到所述互为对调搜索结果组中另一个搜索结果组的评分结果，结合获取的除所述互为对调搜索结果组之外的搜索结果组的评分结果，进行拟合，得到目标查询的搜索结果的评估结果。

9.根据权利要求6～8中任一所述的装置，其特征在于，所述评估结果确定模块包括：

搜索结果组评分获取单元，用于获取用户对所述至少一个搜索结果组的评分结果；

加权求和单元，用于根据预先确定的搜索结果列表中不同排序位置对应的权重，将所述至少一个搜索结果组的评分结果进行加权求和，得到目标查询的搜索结果的评估结果。

10.根据权利要求9所述的装置，其特征在于，所述装置还包括：

样本搜索结果列表获取模块，用于获取样本查询对应的至少两个样本搜索结果列表；

样本搜索结果评估模块，用于利用并排评估方法对所述至少两个样本搜索结果列表中的搜索结果进行评估；

样本搜索结果列表拆分模块，用于将所述至少两个样本搜索结果列表按照列表顺序进行拆分，得到至少一个样本搜索结果组；

样本搜索结果组评分获取模块，用于获取用户对所述至少一个样本搜索结果组的评分结果；

权重确定模块，用于将所述并排评估方法的评估结果和用户对所述至少一个样本搜索结果组的评分结果作为输入，通过回归训练得到所述权重。

11.一种服务器，其特征在于，包括：

一个或多个处理器；

存储装置，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1～5中任一所述的搜索结果的质量评估方法。

12.一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1～5中任一所述的搜索结果的质量评估方法。

搜索结果的质量评估方法、装置、服务器和介质

技术领域

[0001] 本发明实施例涉及计算机技术领域，尤其涉及一种搜索结果的质量评估方法、装置、服务器和介质。

背景技术

[0002] 对于一个成熟的互联网搜索引擎而言，为了保证优秀的用户体验，需要持续不断地改进搜索结果排序算法。而判断一个新的搜索排序算法是否达到了预期效果，就需要科学地评估搜索结果质量。

[0003] 搜索引擎的根本在于找到符合用户需求的信息，这通常可以利用搜索结果与用户输入的关键词之间的相关程度，即结果相关性来反映，因此，结果相关性是评估搜索引擎最重要的指标。在对搜索引擎的搜索结果质量评估过程中，为了避免正反馈的问题，需要引入人力判断。现有技术中依赖人工评估的常用方法包括QU(query-url)评估方法和并排(Side-By-Side)评估方法。

[0004] 但是，QU评估方法的结果稳定性较低、结果冗余、结果缺乏多样性且不能很好满足查询者的需求；相比于QU评估方法，并排评估方法可以克服QU评估的缺陷，但随之产生了其他缺陷，例如评估成本较高，评估复杂性较大。

发明内容

[0005] 本发明实施例提供一种搜索结果的质量评估方法、装置、服务器和介质，以实现降低评估成本和评估难度的效果。

[0006] 第一方面，本发明实施例提供了一种搜索结果的质量评估方法，该方法包括：

[0007] 获取目标查询对应的至少两个搜索结果列表，其中，不同的搜索结果列表是由不同的搜索排序算法得到；

[0008] 将所述至少两个搜索结果列表中的搜索结果按照列表顺序进行拆分，得到至少一个搜索结果组，其中，所述搜索结果组中包括每个搜索结果列表中相同排序位置的搜索结果；

[0009] 获取用户对所述至少一个搜索结果组的评分结果，并进行拟合，得到目标查询的搜索结果的评估结果。

[0010] 第二方面，本发明实施例还提供了一种搜索结果的质量评估装置，该装置包括：

[0011] 搜索结果列表获取模块，用于获取目标查询对应的至少两个搜索结果列表，其中，不同的搜索结果列表是由不同的搜索排序算法得到；

[0012] 搜索结果列表拆分模块，用于将所述至少两个搜索结果列表中的搜索结果按照列表顺序进行拆分，得到至少一个搜索结果组，其中，所述搜索结果组中包括每个搜索结果列表中相同排序位置的搜索结果；

[0013] 评估结果确定模块，用于获取用户对所述至少一个搜索结果组的评分结果，并进行拟合，得到目标查询的搜索结果的评估结果。

[0014] 第三方面，本发明实施例还提供了一种服务器，包括：

[0015] 一个或多个处理器；

[0016] 存储装置，用于存储一个或多个程序，

[0017] 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一实施例所述的搜索结果的质量评估方法。

[0018] 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一实施例所述的搜索结果的质量评估方法。

[0019] 本发明实施例通过获取目标查询对应的至少两个搜索结果列表，其中，不同的搜索结果列表是由不同的搜索排序算法得到，按照列表中的排序将至少两个搜索结果列表中的搜索结果进行拆分，然后获取用户对拆分得到的至少一个搜索结果组的评分结果，进行拟合，最终得到目标查询的搜索结果的评估结果。本发明实施例结合对比与拆分的思想，解决了搜索引擎的现有评估方法中评估成本较高与评估复杂性较大的问题，降低了评估成本和评估难度，提高了评估效率。

附图说明

[0020] 图1是本发明实施例一提供的搜索结果的质量评估方法的流程图；

[0021] 图2是本发明实施例二提供的搜索结果的质量评估方法的流程图；

[0022] 图3是本发明实施例三提供的搜索结果的质量评估装置的结构示意图；

[0023] 图4是本发明实施例四提供的一种服务器的结构示意图。

具体实施方式

[0024] 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。

[0025] 实施例一

[0026] 图1是本发明实施例一提供的搜索结果的质量评估方法的流程图，本实施例可适用于对搜索引擎的搜索结果质量进行评估的情况，该方法可以由搜索结果的质量评估装置来执行，该装置可以采用软件和/或硬件的方式实现，并可集成在服务器中。如图1所示，该方法具体包括：

[0027] S110、获取目标查询对应的至少两个搜索结果列表，其中，不同的搜索结果列表是由不同的搜索排序算法得到。

[0028] 针对用户在一个或者多个搜索引擎中输入的同一目标查询，例如搜索关键词或图片等信息，获取分别基于不同搜索排序算法得到的搜索结果列表，可以实现对多种搜索排序算法的搜索结果质量的评估。具体的，可以同时选择两种搜索排序算法的搜索结果列表进行评估，也可以同时选择两种以上的搜索排序算法的搜索结果列表进行评估。示例性的，针对同一目标查询，分别获取一种搜索引擎的初始搜索排序算法F和改进后的搜索排序算法F1的搜索结果列表，进行评估，可判断算法改进后搜索引擎的搜索结果质量是否有所提高；或者分别获取多种搜索引擎对应的搜索排序算法的搜索结果列表，进行评估，可实现对该多种搜索引擎的搜索结果质量的比较。

[0029] S120、将至少两个搜索结果列表中的搜索结果按照列表顺序进行拆分，得到至少一个搜索结果组，其中，搜索结果组中包括每个搜索结果列表中相同排序位置的搜索结果。

[0030] 获取到同一目标查询的至少两个搜索结果列表之后，按照列表顺序进行拆分，得到至少一个搜索结果组，然后将目标查询与搜索结果组呈现给专家，进行搜索结果组的对比评分。经过列表拆分，专家评分过程中，无需关注整个搜索结果列表中不同排序位置的搜索结果之间的关联性，只需要关注单个搜索结果组中的搜索结果即可。相比于整体考虑整个搜索结果列表相关性进行对比评分，按照拆分后的搜索结果组进行比对评分，可以降低评估的难度和复杂性，提高评估效率，也可以提升评分稳定性，并且由于专家每次需要关注的搜索结果数量减少，专家的每次评分工作难度降低，因此评估成本也会降低。

[0031] S130、获取用户对至少一个搜索结果组的评分结果，并进行拟合，得到目标查询的搜索结果的评估结果。

[0032] 将每个搜索结果组的评分结果进行拟合，得到针对目标查询的多个搜索结果列表的综合对比评分，即完成对目标查询的搜索结果的评估。例如，如果最终得到的评估结果是一种搜索引擎的不同改进搜索排序算法的综合对比评分，评估结果可用于判断算法改进后搜索引擎的搜索结果质量的提高情况；如果最终得到的评估结果是多种搜索引擎对应的搜索排序算法的搜索结果列表之间的综合对比评分，评估结果可用于比较该多种搜索引擎的搜索结果质量。其中，具体的拟合过程，可以是按照以预先设定的规则将列表排序位置的影响因素加以考虑的算法进行拟合，也可以是按照预先训练的拟合模型进行拟合。

[0033] 对于搜索结果列表拆分与搜索结果组的评分过程，结合表1和表2进行示例性说明。以下以选取两种搜索排序算法的搜索结果列表为例，应知也可以选取三种及以上的搜索排序算法。针对用户输出的目标查询XX，基于搜索排序算法Fa和搜索排序算法Fb的搜索结果列表分别为列表Ra和列表Rb。

[0034] 如果直接采用并排(Side-By-Side)评估方法进行评分，即对搜索结果列表不进行拆分，如表1所示，专家整体考虑两个搜索结果列表Ra和Rb的搜索结果，给出综合对比评分。其中，既要考虑不同搜索结果列表中同一排序位置的搜索结果之间的对比，还需要考虑每个搜索结果与整体搜索结果列表中其他搜索结果之间的关联性。例如，以列表前三位的排序结果为例，根据并排评估方法，评分过程中专家需要同时考虑URL1、URL2、URL3、URL4、URL5、URL6这六个搜索结果的相关性，给出综合的对比评分。当需要同时考虑的结果数量增加，评估复杂度随之明显增加，评估效率也随结果列表的长度呈指数下降。

[0035] 如果对搜索结果列表进行拆分，如表2所示，得到多个搜索结果组，每组中包括两个搜索结果。专家评分过程中，只需要以搜索结果组为单位进行对比评分即可，相当于将对URL1、URL2、URL3、URL4、URL5和URL6的一次性综合对比评分，变为每次只考虑两个搜索结果的三次对比评分过程，每次只考虑搜索结果URL1与URL4或搜索结果URL2与URL5或搜索结果URL3与URL6。每次需要考虑的搜索结果数量减少，评估难度、复杂度和成本便得以降低。

[0036]

表 1 搜索结果列表拆分前

搜索排序算法 Fa	搜索排序算法 Fb
搜索结果列表 Ra	搜索结果列表 Rb
URL 1	URL 4
URL 2	URL 5
URL 3	URL 6
……	……
URL N	URL M

表 2 搜索结果列表拆分后

搜索排序算法 Fa	搜索排序算法 Fb
搜索结果列表 Ra	搜索结果列表 Rb
URL 1	URL 4
URL 2	URL 5
URL 3	URL 6
……	……
URL N	URL M

[0037] 需要说明的是，对于涉及人力判断的评估过程，互联网公司一般是将评估任务发放到外包公司，依赖外包公司的专家进行人工对比评分，根据每个专家的评分任务的复杂度，评估成本有所差异。例如，采用并排评估方法，由于专家评估过程需要考虑整个搜索结果列表的关联性，评估难度和复杂度较大，因此人均评估成本也较高；采用本方案，将搜索结果列表进行拆分，每个专家需要考虑的搜索结果数量减少，评估难度和复杂度下降，因此人均评估成本也较低。对于多人评估的情况，例如，并排评估方法采用平均6人/组对同一目标查询的同一组搜索结果进行对比评分，每个专家的评分成本是3X，则并排评估中该组专家对应的总共评估成本为18X；采用本方案将搜索结果列表进行拆分，利用平均8人/组对同一目标查询的同一组搜索结果进行对比评分，达到相同的评估质量，若每个专家的评分次数增加为2次，每个专家的人均评估成本较低，仅为X，最终该组专家对应的整体评估成本为16X，低于并排评估的成本。

[0038] 另外，本实施例中之所以考虑采用专家对搜索结果组进行对比评分，是因为依赖专家评分的搜索结果质量的评估方法不仅可以避免正反馈的现象，还可以解决小流量评估方法中存在的各种问题。小流量评估一般是通过随机采样，将搜索流量中的一小部分直接用于评估，评估依赖于搜索日志中记录的用户行为，该评估方法存在的问题在于：

[0039] 1)由于评估依赖用户的主观判断，用户间存在个体差异，对搜索结果的相关性标准并不一致。同样的搜索结果，在有些用户看来可能已经可以满足需求，而另一些用户却可能并不满意。

[0040] 2)由于评估依赖用户的被动反馈，用户行为中存在大量噪音，并非所有的用户行为都反应了用户对搜索结果的相关性判断。这种噪音虽然可以在统计意义下进行估计，但却难以提供指导搜索排序算法迭代改进的具体情形。

[0041] 专家评分过程中，首先针对要评估的搜索结果类型，制定一套严格的评估标准，并

基于评估标准，培训评估专家，然后由一组评估专家对经过随机抽样得到的目标查询的搜索结果进行评分，最终综合多个专家的意见给出每个搜索结果的评分，优点在于：1)具有客观性：具有严格制定的评估标准，2)具有主动性：由经过培训的专家进行评估。

[0042] 在上述技术方案的基础上，可选的，每个搜索结果组中包括至少两个不同的搜索结果。

[0043] 如果拆分得到的某个搜索结果组中的搜索结果完全相同，如下表3中所示，第一个搜索结果组包括URL1和URL1，则无需专家对该搜索结果组进行对比评分，系统可以统一赋以相同的对比评分，也可以将对比评分记为0。系统对具有相同搜索结果的搜索结果组可以进行标记，以便后续拟合过程中统一处理。如果某个搜索结果组中包括三个及以上的搜索结果，即使只有两个不同的搜索结果，其他搜索结果均相同，这时仍是需要专家对该搜索结果组进行对比评分，例如某个搜索结果组中包括的搜索结果为URL1、URL1、URL4，虽然前两个搜索结果相同，此时仍需要专家对该搜索结果组进行对比评分。在专家进行对比评分之前，通过将包括完全相同搜索结果的搜索结果组进行过滤，可以减少不必要的专家评分操作，进而降低评估成本。

[0044] 可选的，如果每个搜索结果组中包括两个搜索结果，当任意两个搜索结果组互为对调搜索结果组时，其中，互为对调搜索结果组是指搜索结果相同且搜索结果的位置对调的搜索结果组，则相应的，获取用户对至少一个搜索结果组的评分结果，包括：

[0045] 获取用户对互为对调搜索结果组中任一个搜索结果组的评分结果，并通过取相反数运算得到互为对调搜索结果组中另一个搜索结果组的评分结果。

[0046] 其中，互为对调搜索结果组如下表3中所示的第二个搜索结果组和第三个搜索结果组，当系统识别到拆分后得到的搜索结果组中存在互为对调搜索结果组时，只需要选取其中的一个呈现给专家进行对比评分即可，然后通过程序自动取相反数得到另一个搜索结果组的评分结果。例如，获取专家对包括URL2和URL5的搜索结果组的评分结果，通过取相反数，便可得到包括URL5和URL2的搜索结果组的评分结果，而无需对URL5和URL2进行再次单独评分，然后结合其他剩余的搜索结果组的评分结果，进行拟合，得到最终的评估结果。通过对互为对调搜索结果组的识别与评分的取反处理，可以减少专家评分的工作量，进而降低评估成本。

[0047] 表3搜索结果相同的搜索结果组和互为对调搜索结果组示例

[0048]

URL 1	URL 1
URL 2	URL 5
URL 5	URL 2

[0049] 本实施例技术方案通过获取目标查询对应的至少两个搜索结果列表，其中，不同的搜索结果列表是由不同的搜索排序算法得到，按照列表中的排序将至少两个搜索结果列表中的搜索结果进行拆分，然后获取用户对拆分得到的至少一个搜索结果组的评分结果，进行拟合，最终得到目标查询的搜索结果的评估结果。本发明实施例结合对比与拆分的思

想，使得专家评分过程中不需要考虑整个搜索结果列表不同排序位置的搜索结果之间的关联性，每次只需要关注单个搜索结果组中的搜索结果，解决了搜索引擎的现有评估方法中评估成本较高与评估复杂性较大的问题，降低了评估成本和评估难度，提高了评估效率，并且，通过对比取相对评分，相比于针对目标查询只考虑一个搜索结果且取绝对评分的QU评估方法，本实施例方案保证了评估的准确性与稳定性。

[0050] 实施例二

[0051] 图2是本发明实施例二提供的搜索结果的质量评估方法的流程图，本实施例是在上述实施例的基础上进一步进行优化。如图2所示，该方法具体包括：

[0052] S210、获取目标查询对应的至少两个搜索结果列表，其中，不同的搜索结果列表是由不同的搜索排序算法得到。

[0053] S220、将至少两个搜索结果列表中的搜索结果按照列表顺序进行拆分，得到至少一个搜索结果组，其中，搜索结果组中包括每个搜索结果列表中相同排序位置的搜索结果。

[0054] S230、获取用户对所述至少一个搜索结果组的评分结果。

[0055] S240、根据预先确定的搜索结果列表中不同排序位置对应的权重，将至少一个搜索结果组的评分结果进行加权求和，得到目标查询的搜索结果的评估结果。

[0056] 将目标查询的搜索结果列表进行拆分，专家对搜索结果组进行对比评分过程中，并不考虑搜索结果列表中排序位置的影响，而拟合过程中，将不同排序位置对应的权重进行考虑，通过加权求和，得到最终的评估结果，保证了评估的准确性。其中，评估结果的加权求和计算公式如下：

[0057] $$Top_n = \sum_{i=1}^{n} w_i Re_i$$

[0058] 其中，Re_i是第i个搜索结果组的对比评分结果，i也表示搜索结果列表的列表排序位置，n表示搜索结果组的数量，也表示搜索结果列表的总排序位数。

[0059] 可选的，在根据预先确定的搜索结果列表中不同排序位置对应的权重，将至少一个搜索结果组的评分结果进行加权求和之前，该方法还包括：

[0060] 获取样本查询对应的至少两个样本搜索结果列表；

[0061] 利用并排评估方法对至少两个样本搜索结果列表中的搜索结果进行评估；

[0062] 将至少两个样本搜索结果列表按照列表顺序进行拆分，得到至少一个样本搜索结果组；

[0063] 获取用户对至少一个样本搜索结果组的评分结果；

[0064] 将并排评估方法的评估结果和用户对至少一个样本搜索结果组的评分结果作为输入，通过回归训练得到搜索结果列表不同排序位置对应的权重。

[0065] 将对比与拆分的思想应用于样本查询对应的样本搜索结果列表，得到对应的样本搜索结果组的评分结果，然后结合利用并排评估方法对样本查询的搜索结果的评估结果，通过回归训练得到不同列表位置对应的权重，进而应用于本实施例的目标查询的搜索结果的评估过程，相当于本实施例评估方法是将并排评估方法的评估结果作为参考标准对不同的排序位置的重要性加以考虑，因此，可以有效地继承并排评估方法的稳定性与准确性的优势。其中，当样本搜索结果组中存在搜索结果相同或者互为对调搜索结果组的情况时，可

以同样采取前述实施例中的过滤与评分取反处理。考虑本实施例评估方法是以并排评估方法为基础，因此，并排评估方法相比于QU评估方法的优势，本实例评估方法均可以得以保留。

[0066] 本实施例技术方案通过获取目标查询对应的至少两个搜索结果列表，其中，不同的搜索结果列表是由不同的搜索排序算法得到，按照列表中的排序将至少两个搜索结果列表中的搜索结果进行拆分，然后获取用户对拆分得到的至少一个搜索结果组的评分结果，进行加权求和，最终得到目标查询的搜索结果的评估结果。本发明实施例结合对比与拆分的思想，解决了搜索引擎的现有评估方法中评估成本较高与评估复杂性较大的问题，降低了评估成本和评估难度，提高了评估效率，并且，由于评分拟合过程中的权重是基于并排评估方法得到，有效继承了并排评估方法的准确性。

[0067] 实施例三

[0068] 图3是本发明实施例三提供的搜索结果的质量评估装置的结构示意图，本实施例可适用于对搜索引擎的搜索结果质量进行评估的情况。本发明实施例所提供的搜索结果的质量评估装置可执行本发明任意实施例所提供的搜索结果的质量评估方法，具备执行方法相应的功能模块和有益效果。如图3所示，该装置包括搜索结果列表获取模块310、搜索结果列表拆分模块320和评估结果确定模块330，其中：

[0069] 搜索结果列表获取模块310，用于获取目标查询对应的至少两个搜索结果列表，其中，不同的搜索结果列表是由不同的搜索排序算法得到。

[0070] 搜索结果列表拆分模块320，用于将至少两个搜索结果列表中的搜索结果按照列表顺序进行拆分，得到至少一个搜索结果组，其中，搜索结果组中包括每个搜索结果列表中相同排序位置的搜索结果。

[0071] 评估结果确定模块330，用于获取用户对至少一个搜索结果组的评分结果，并进行拟合，得到目标查询的搜索结果的评估结果。

[0072] 可选的，搜索结果列表拆分模块320中每个搜索结果组中包括至少两个不同的搜索结果。

[0073] 可选的，如果搜索结果列表拆分模块320中每个搜索结果组中包括两个搜索结果，当任意两个搜索结果组互为对调搜索结果组时，其中，互为对调搜索结果组是指搜索结果相同且搜索结果的位置对调的搜索结果组，则相应的，评估结果确定模块330具体用于：

[0074] 获取用户对互为对调搜索结果组中任一个搜索结果组的评分结果，并通过取相反数运算得到互为对调搜索结果组中另一个搜索结果组的评分结果，结合获取的除互为对调搜索结果组之外的搜索结果组的评分结果，进行拟合，得到目标查询的搜索结果的评估结果。

[0075] 可选的，评估结果确定模块330包括：

[0076] 搜索结果组评分获取单元，用于获取用户对至少一个搜索结果组的评分结果；

[0077] 加权求和单元，用于根据预先确定的搜索结果列表中不同排序位置对应的权重，将至少一个搜索结果组的评分结果进行加权求和，得到目标查询的搜索结果的评估结果。

[0078] 可选的，该装置还包括：

[0079] 样本搜索结果列表获取模块，用于获取样本查询对应的至少两个样本搜索结果列表；

[0080] 样本搜索结果评估模块，用于利用并排评估方法对至少两个样本搜索结果列表中的搜索结果进行评估；

[0081] 样本搜索结果列表拆分模块，用于将至少两个样本搜索结果列表按照列表顺序进行拆分，得到至少一个样本搜索结果组；

[0082] 样本搜索结果组评分获取模块，用于获取用户对至少一个样本搜索结果组的评分结果；

[0083] 权重确定模块，用于将并排评估方法的评估结果和用户对至少一个样本搜索结果组的评分结果作为输入，通过回归训练得到搜索结果列表不同排序位置对应的权重。

[0084] 本实施例技术方案通过获取目标查询对应的至少两个搜索结果列表，其中，不同的搜索结果列表是由不同的搜索排序算法得到，按照列表中的排序将至少两个搜索结果列表中的搜索结果进行拆分，然后获取用户对拆分得到的至少一个搜索结果组的评分结果，进行拟合，最终得到目标查询的搜索结果的评估结果。本发明实施例结合对比与拆分的思想，解决了搜索引擎的现有评估方法中评估成本较高与评估复杂性较大的问题，降低了评估成本和评估难度，提高了评估效率。

[0085] 实施例四

[0086] 图4是本发明实施例四提供的一种服务器的结构示意图。图4示出了适于用来实现本发明实施方式的示例性服务器412的框图。图4显示的服务器412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。

[0087] 如图4所示，服务器412以通用服务器的形式表现。服务器412的组件可以包括但不限于：一个或者多个处理器416，存储装置428，连接不同系统组件(包括存储装置428和处理器416)的总线418。

[0088] 总线418表示几类总线结构中的一种或多种，包括存储装置总线或者存储装置控制器，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(Industry Subversive Alliance，ISA)总线，微通道体系结构(Micro Channel Architecture，MAC)总线，增强型ISA总线、视频电子标准协会(Video Electronics Standards Association，VESA)局域总线以及外围组件互连(Peripheral Component Interconnect，PCI)总线。

[0089] 服务器412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服务器412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。

[0090] 存储装置428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(Random Access Memory，RAM)430和/或高速缓存存储器432。服务器412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434可以用于读写不可移动的、非易失性磁介质(图4未显示，通常称为“硬盘驱动器”)。尽管图4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，例如只读光盘(Compact Disc Read-Only Memory，CD-ROM)，数字视盘(Digital Video Disc-Read Only Memory，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418相连。存储装置428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。

[0091] 具有一组(至少一个)程序模块442的程序/实用工具440,可以存储在例如存储装置428中,这样的程序模块442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442通常执行本发明所描述的实施例中的功能和/或方法。

[0092] 服务器412也可以与一个或多个外部设备414(例如键盘、指向终端、显示器424等)通信,还可与一个或者多个使得用户能与该服务器412交互的终端通信,和/或与使得该服务器412能与一个或多个其它计算终端进行通信的任何终端(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422进行。并且,服务器412还可以通过网络适配器420与一个或者多个网络(例如局域网(Local Area Network,LAN),广域网(Wide Area Network,WAN)和/或公共网络,例如因特网)通信。如图4所示,网络适配器420通过总线418与服务器412的其它模块通信。应当明白,尽管图中未示出,可以结合服务器412使用其它硬件和/或软件模块,包括但不限于:微代码、终端驱动器、冗余处理器、外部磁盘驱动阵列、磁盘阵列(Redundant Arrays of Independent Disks,RAID)系统、磁带驱动器以及数据备份存储系统等。

[0093] 处理器416通过运行存储在存储装置428中的程序,从而执行各种功能应用以及数据处理,例如实现本发明实施例所提供的搜索结果的质量评估方法,该方法包括:

[0094] 获取目标查询对应的至少两个搜索结果列表,其中,不同的搜索结果列表是由不同的搜索排序算法得到;

[0095] 将至少两个搜索结果列表中的搜索结果按照列表顺序进行拆分,得到至少一个搜索结果组,其中,搜索结果组中包括每个搜索结果列表中相同排序位置的搜索结果;

[0096] 获取用户对至少一个搜索结果组的评分结果,并进行拟合,得到目标查询的搜索结果的评估结果。

[0097] 实施例五

[0098] 本发明实施例五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实施例所提供的搜索结果的质量评估方法,该方法包括:

[0099] 获取目标查询对应的至少两个搜索结果列表,其中,不同的搜索结果列表是由不同的搜索排序算法得到;

[0100] 将至少两个搜索结果列表中的搜索结果按照列表顺序进行拆分,得到至少一个搜索结果组,其中,搜索结果组中包括每个搜索结果列表中相同排序位置的搜索结果;

[0101] 获取用户对至少一个搜索结果组的评分结果,并进行拟合,得到目标查询的搜索结果的评估结果。

[0102] 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。

[0103] 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。

[0104] 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。

[0105] 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终端上执行。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)。

[0106] 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。

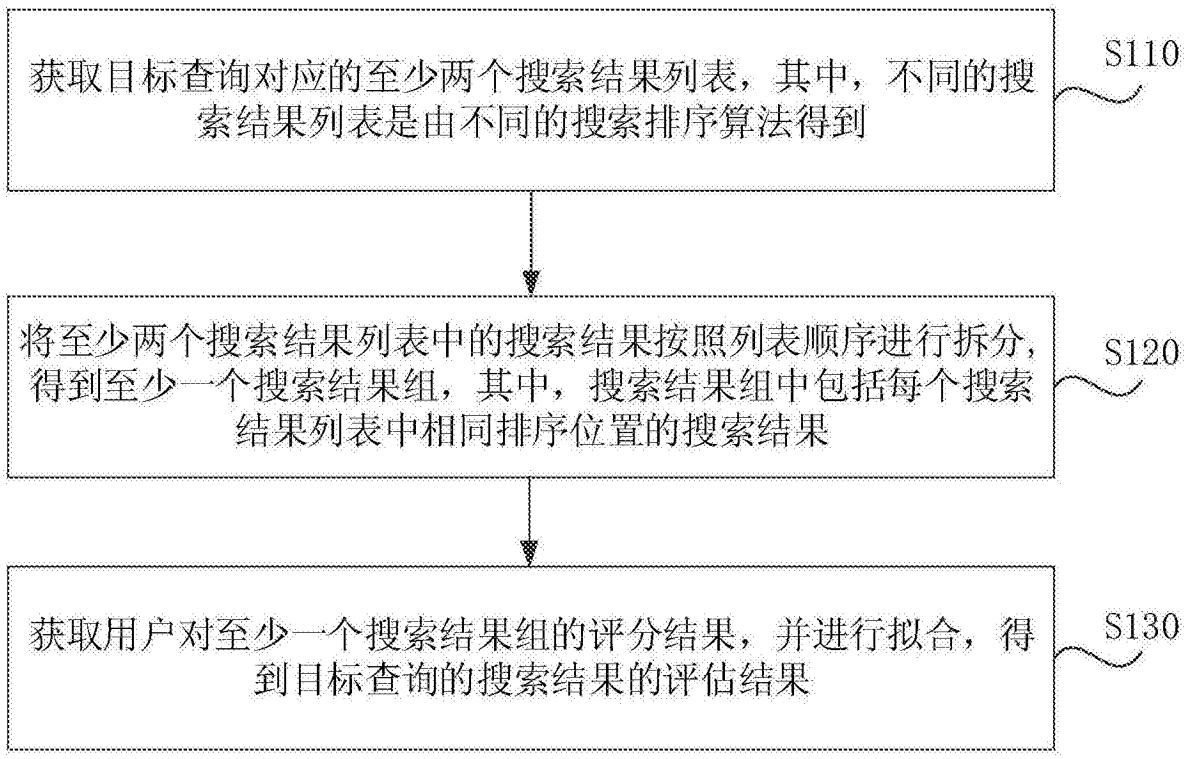

获取目标查询对应的至少两个搜索结果列表，其中，不同的搜索结果列表是由不同的搜索排序算法得到
S110
将至少两个搜索结果列表中的搜索结果按照列表顺序进行拆分，得到至少一个搜索结果组，其中，搜索结果组中包括每个搜索结果列表中相同排序位置的搜索结果
S120
获取用户对至少一个搜索结果组的评分结果，并进行拟合，得到目标查询的搜索结果的评估结果
S130

图1

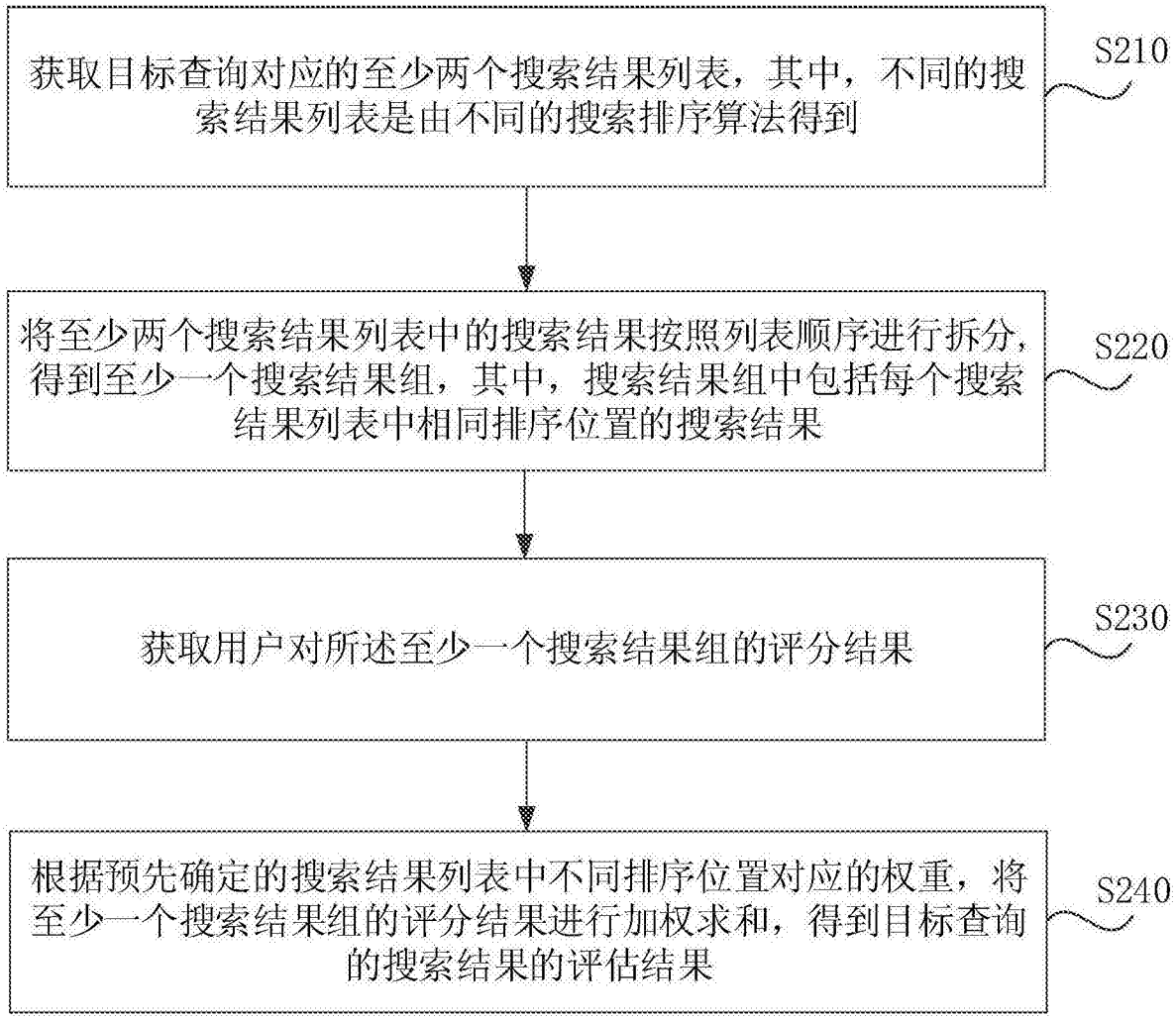

获取目标查询对应的至少两个搜索结果列表，其中，不同的搜索结果列表是由不同的搜索排序算法得到
S210
将至少两个搜索结果列表中的搜索结果按照列表顺序进行拆分，得到至少一个搜索结果组，其中，搜索结果组中包括每个搜索结果列表中相同排序位置的搜索结果
S220
获取用户对所述至少一个搜索结果组的评分结果
S230
根据预先确定的搜索结果列表中不同排序位置对应的权重，将至少一个搜索结果组的评分结果进行加权求和，得到目标查询的搜索结果的评估结果
S240

图2

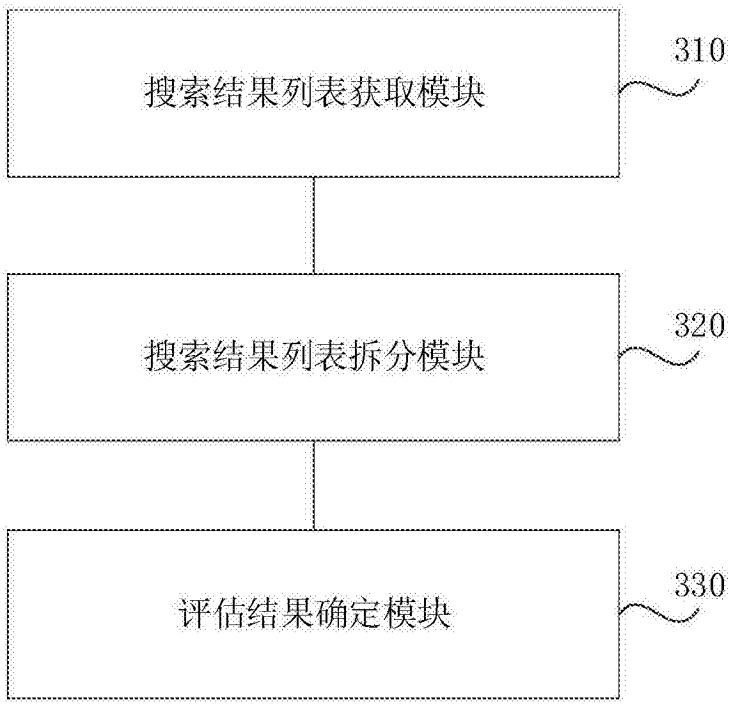


图3

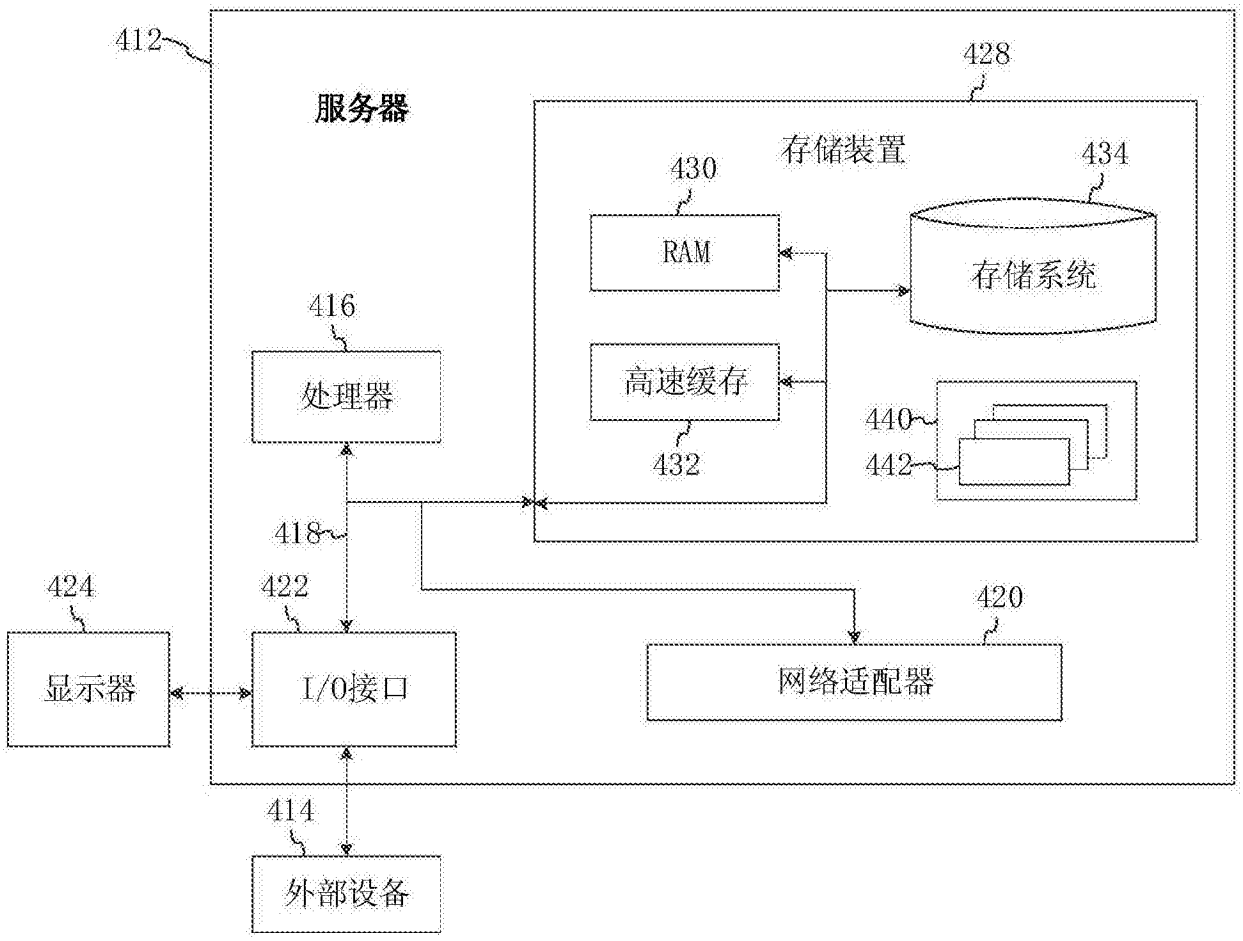

412
服务器
428
存储装置
430
RAM
434
存储系统
416
处理器
高速缓存
432
440
442
418
424
422
420
显示器
I/O接口
网络适配器
414
外部设备

图4