(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 101930438 B
(45)授权公告日 2016.08.31

(21)申请号 200910146331.5
(22)申请日 2009.06.19
(73)专利权人 阿里巴巴集团控股有限公司
地址 英属开曼群岛大开曼岛资本大厦一座四层847号邮箱
(72)发明人 郭宁 邢飞 谢宇恒 侯磊 张勤
(74)专利代理机构 北京集佳知识产权代理有限公司 11227
代理人 逯长明 王宝筠
(51)Int.Cl.
*G06F 17/30*(2006.01)

(56)对比文件
CN 101233513 A,2008.07.30,
CN 1335574 A,2002.02.13,
CN 1710560 A,2005.12.21,

审查员 李楠

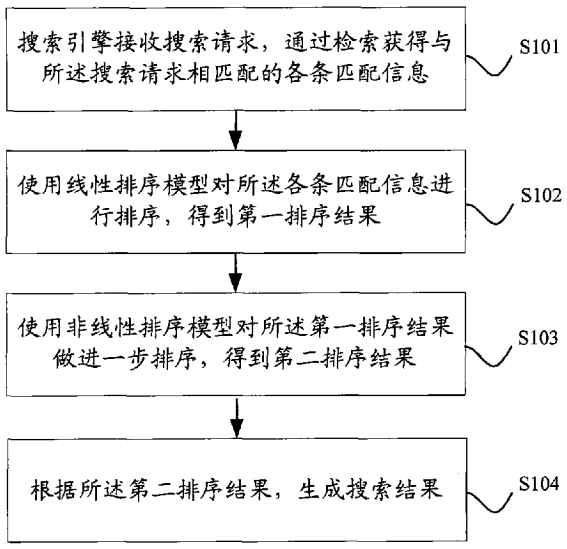
权利要求书2页 说明书8页 附图3页

(54)发明名称

一种搜索结果生成方法及信息搜索系统

(57)摘要

本申请公开了一种搜索结果生成方法及信息搜索系统。一种搜索结果生成方法，包括：信息搜索系统接收搜索请求，通过检索获得与所述搜索请求相匹配的各条匹配信息；使用线性排序模型对所述各条匹配信息中的N1条匹配信息进行排序，得到第一排序结果；使用非线性排序模型对所述第一排序结果中的前N2条匹配信息进行排序，得到第二排序结果，其中N2＜N1；根据所述第二排序结果，生成搜索结果。应用以上技术方案，可以有效减小使用非线性排序模型所处理的数据量，从而提高对匹配信息排序的整体处理速度。




CN 101930438 B

1.一种搜索结果生成方法，其特征在于，包括：

信息搜索系统接收搜索请求，通过检索获得与所述搜索请求相匹配的各条匹配信息；

使用线性排序模型对所述各条匹配信息中的N1条匹配信息进行排序，得到第一排序结果，其中，N1≤所检索到的匹配信息的总数目；

使用非线性排序模型对所述第一排序结果中的前N2条匹配信息进行排序，得到第二排序结果，其中N2＜N1；

根据所述第二排序结果，生成搜索结果以便展现；

其中，所述线性排序模型和所述非线性排序模型所使用的主要搜索特征相同。

2.根据权利要求1所述的方法，其特征在于，在通过检索获得与所述搜索请求相匹配的各条匹配信息之后，还包括：

对所述各条匹配信息进行排序预处理，由所述各条匹配信息中选取N1条匹配信息作为后续步骤排序的对象；其中，N1＜所检索到的匹配信息的总数目。

3.根据权利要求1或2所述的方法，其特征在于，

所述线性排序模型或非线性排序模型的输入为匹配信息的至少一个特征值，输出为匹配的信息的排序分值，所述排序分值，用于确定匹配信息的排列顺序。

4.根据权利要求3所述的方法，其特征在于，

所述线性排序模型所使用的特征，与所述非线性模型所使用的特征完全相同或部分相同。

5.根据权利要求4所述的方法，其特征在于，

所述匹配信息的特征值，由匹配信息自身所确定，或者由匹配信息与所述搜索请求共同确定。

6.根据权利要求1或2所述的方法，其特征在于，

所述线性排序模型或非线性排序模型，是通过训练所确定的模型。

7.一种信息搜索系统，其特征在于，包括：

信息检索单元，用于接收搜索请求，通过检索获得与所述搜索请求相匹配的各条匹配信息；

线性排序单元，用于使用线性排序模型对所述信息检索单元检索获得的各条匹配信息中的N1条匹配信息进行排序，得到第一排序结果，其中，N1≤所检索到的匹配信息的总数目；

非线性排序单元，用于使用非线性排序模型对所述线性排序单元排序得到的第一排序结果中的前N2条匹配信息进行排序，得到第二排序结果，其中N2＜N1；

结果生成单元，用于根据所述第二排序结果，生成搜索结果以便展现；

其中，所述线性排序模型和所述非线性排序模型所使用的主要搜索特征相同。

8.根据权利要求7所述的系统，其特征在于，还包括：

排序预处理单元，用于在所述信息检索单元获得所述各条匹配信息之后，对所述各条匹配信息进行排序预处理，由所述各条匹配信息中选取N1条匹配信息作为所述线性排序单元排序的对象；其中，N1＜所检索到的匹配信息的总数目。

9.根据权利要求7或8所述的系统，其特征在于，

所述线性排序模型或非线性排序模型的输入为匹配信息的至少一个特征值，输出为匹

配的信息的排序分值，所述排序分值，用于确定匹配信息的排列顺序。

10.根据权利要求7或8所述的系统，其特征在于，

所述线性排序模型或非线性排序模型，是通过训练所确定的模型。

# 一种搜索结果生成方法及信息搜索系统

## 技术领域

[0001] 本申请涉及计算机应用领域，特别是涉及一种搜索结果生成方法及信息搜索系统。

## 背景技术

[0002] 信息搜索系统是一种能够为用户提供信息检索服务的系统，以互联网中常用的搜索引擎为例，作为应用在互联网领域的搜索系统，搜索引擎目前已经成为用户上网必不可少的辅助工具之一。从用户的角度看，搜索引擎一般提供一个包含搜索框的页面，用户在搜索框输入关键词或其他搜索条件，通过浏览器提交给搜索引擎后，搜索引擎就会返回与用户输入的关键词内容相匹配的信息。

[0003] 针对同样的用户搜索请求(例如用户在搜索时所输入的搜索关键词)，搜索引擎往往能够检索到多条匹配信息，这个数量可能会达到数十至数万。而从用户的角度来讲，往往只会重点关注在搜索结果中排序比较靠前的信息。这样，在搜索引擎向用户提供搜索结果时，如何对这些信息进行排序就显得尤为重要，搜索结果的排序是否合理将直接影响着用户的体验。

[0004] 搜索引擎在对信息进行排序时，会综合考虑一种或多种因素(例如：搜索关键词在匹配信息中出现的次数、搜索关键词在匹配信息中所处的位置等等)，构建形如$y=f(x_1, x_2, \ldots, x_n)$的排序模型，根据该模型为每条匹配信息进行打分，最后依据分值高低对每条匹配信息进行排序。其中，上述模型的输入参量，即函数自变量$x_1, x_2, \ldots, x_n$，分别表示所考虑的各种因素，称为匹配信息的特征，模型的输出即应变量y表示匹配信息的得分值。

[0005] 根据$y=f(x_1, x_2, \ldots, x_n)$具体形式的不同，可以将排序模型分为线性排序模型和非线性排序模型两大类。一般而言，相对于线性排序模型，非线性排序模型的拟合能力更强，因此使用非线性排序模型可以实现更好的搜索效果(即匹配信息的排列顺序更符合用户的实际需求，或者与用户期待的顺序更趋于一致)。但是，由于非线性排序模型的复杂度高，因此，其处理速度较为缓慢。特别是在对大量匹配信息进行排序处理时，需要占用很长的时间来生成搜索结果，对用户体验造成了影响。

## 发明内容

[0006] 为解决上述技术问题，本申请提供一种搜索结果生成方法及信息搜索系统，以提高对匹配信息排序的处理速度，提升用户体验，技术方案如下：

[0007] 本申请提供一种搜索结果生成方法，包括：

[0008] 信息搜索系统接收搜索请求，通过检索获得与所述搜索请求相匹配的各条匹配信息；

[0009] 使用线性排序模型对所述各条匹配信息中的N1条匹配信息进行排序，得到第一排序结果，其中，$N1 \leq$所检索到的匹配信息的总数目；

[0010] 使用非线性排序模型对所述第一排序结果中的前N2条匹配信息进行排序，得到第

二排序结果，其中N2＜N1；

[0011] 根据所述第二排序结果，生成搜索结果。

[0012] 本申请还提供一种信息搜索系统，其特征在于，包括：

[0013] 信息检索单元，用于接收搜索请求，通过检索获得与所述搜索请求相匹配的各条匹配信息；

[0014] 线性排序单元，用于使用线性排序模型对所述信息检索单元检索获得的各条匹配信息中的N1条匹配信息进行排序，得到第一排序结果，其中，N1≤所检索到的匹配信息的总数目；

[0015] 非线性排序单元，用于使用非线性排序模型对所述线性排序单元排序得到的第一排序结果中的前N2条匹配信息进行排序，得到第二排序结果，其中N2＜N1；

[0016] 与现有技术相比，本申请实施例所提供的技术方案，首先使用线性排序模型对N1条匹配信息进行排序处理，然后对排序结果的前N2条再使用非线性排序模型进行排序处理。由于线性排序模型的处理速度是能够保证的，因此对于大量(N1条)的匹配信息，首先利用线性排序模型进行预处理，然后通过设置N2＜N1，可以有效减小使用非线性排序模型所处理的数据量，从而提高对匹配信息排序的整体处理速度。

## 附图说明

[0017] 为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术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中记载的一些实施例，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。

[0018] 图1为本申请实施例一的搜索结果生成方法的流程图；

[0019] 图2为本申请实施例二的搜索结果生成方法的流程图；

[0020] 图3为本申请实施例二的搜索效果示意图；

[0021] 图4为本申请实施例信息搜索系统的结构示意图；

[0022] 图5为本申请实施例信息搜索系统的另一种结构示意图。

## 具体实施方式

[0023] 首先对本申请实施例的一种搜索结果生成方法进行说明，包括：

[0024] 信息搜索系统接收搜索请求，通过检索获得与所述搜索请求相匹配的各条匹配信息；

[0025] 使用线性排序模型对所述各条匹配信息中的N1条匹配信息进行排序，得到第一排序结果；

[0026] 使用非线性排序模型对所述第一排序结果中的前N2条匹配信息进行排序，得到第二排序结果，其中N2＜N1；

[0027] 根据所述第二排序结果，生成搜索结果。

[0028] 为了使本技术领域的人员更好地理解本申请中的技术方案，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，所描述的实施例仅仅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

技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都应当属于本申请保护的范围。

[0029] 下面以网络搜索应用为例，对本申请所提供的技术方案进行详细说明，图1所示为本申请实施例的一种搜索结果生成方法的流程图，包括以下步骤：

[0030] S101，搜索引擎接收搜索请求，通过检索获得与所述搜索请求相匹配的各条匹配信息；

[0031] 当用户需要在网络上搜索信息时，会输入一个或多个搜索条件，一般最为常用的搜索条件是搜索关键词，根据具体搜索应用场景的不同，有些搜索引擎还可以支持更多类型的搜索条件，例如信息发布时间、信息属性等等，本申请实施例中，将各种搜索条件统称为搜索请求。搜索引擎接收到搜索请求之后，检索与搜索请求相匹配的信息。对应不同的搜索应用场景，检索到的信息类型也有所不同，例如：在网页搜索中，检索到的信息为网页；在电子商务搜索中，检索到的信息为商品；在文献搜索中，检索到的信息为期刊或论文等等。其中，根据搜索请求检索与之相匹配的信息，其实现方法与现有技术相同，本申请实施例对此不再进行详细说明。

[0032] S102，使用线性排序模型对所述各条匹配信息进行排序，得到第一排序结果；

[0033] 本步骤中，使用线性模型为每条匹配信息进行打分，然后依据分值高低对每条匹配信息进行排序。

[0034] 线性排序模型的数学表达式形式如下：

[0035] $y=f(x_1,x_2,...,x_n)$

[0036] $=a_1x_1+a_2x_2+...+a_nx_n$

[0037] 在上述模型中，应变量y与每个自变量分别构成一次函数关系，其中，模型的输入参量$x_1,x_2,...,x_n$，分别表示在排序时需要考虑的各种因素，称为匹配信息的特征；$a_1,a_2,...,a_n$分别为每个特征的加权系数，$a_n$的大小反映$x_n$对应特征对于排序的重要程度。模型的输出y表示匹配信息的排序分值。

[0038] 根据具体的搜索应用需求，系统会根据匹配信息的一个或多个特征，来计算每条匹配信息的排序分值大小。这些特征可能涉及多个方面，举例如下：

[0039] 1)搜索关键词在匹配信息中出现的次数。

[0040] 一般认为，搜索关键词在某条匹配信息中出现的次数越多，则该条匹配信息应该获得越高的排序分值。

[0041] 2)搜索关键词在匹配信息中所处的位置。

[0042] 一般认为，如果搜索关键词出现在某条匹配信息的标题、摘要等重要部分，则该条匹配信息应该获得较高的排序分值。

[0043] 3)匹配信息的用户反馈量。

[0044] 用户反馈量能够反映用户对某条信息关注程度，搜索引擎可以通过读取用户反馈日志，获得各条匹配信息所对应的用户反馈量，并根据用户反馈量为各条匹配信息打分，其基本原则是：用户对某条匹配信息的关注程度越高，则该条匹配信息应该获得越高的排序分值。

[0045] 4)匹配信息的来源。

[0046] 匹配信息的来源也可以作为确定其排列顺序的因素，例如，对于网页搜索来说，如

果匹配信息来源于大型门户网站或官方网站，则可获得较高的排序分值。

[0047] 以上仅仅列举了几种常用的匹配信息特征，匹配信息还具有很多可以用来计算排序分值的特征，这里不再一一说明。

[0048] 当一个排序模型确定以后，该模型所要使用的特征种类以及数量也就确定了。系统在对匹配信息进行排序时，首先要获取每条匹配信息的每个特征值，然后根据排序模型计算出每条匹配信息的排序分值，最后根据排序分值大小对每条匹配信息进行排序。

[0049] 举例说明，假设排序模型为$y=f(x_1,x_2,x_3)$，则其使用的特征数量为3，待排序的匹配信息数量为10，则系统需要分别获取10组$(x_1,x_2,x_3)$的特征值，然后分别计算出10个y值，最后根据10个y值的大小对这10条匹配信息进行排序。

[0050] S103，使用非线性排序模型对所述第一排序结果做进一步排序，得到第二排序结果；

[0051] 本步骤的执行方法与S102类似，不同之处在于，本步骤所依据的排序模型为非线性模型。

[0052] 对匹配信息进行排序的目的，是希望最终展现给用户的搜索结果能够更加符合用户的实际需求。可以想象的是，匹配信息的各个特征与其最终的排序得分在客观上是存在某种对应关系的。建立排序模型的目的，就是尽量去拟合这种对应关系。本领域技术人员所公知的是，线性函数的拟合能力是有限的，而非线性函数在理论上可以拟合任何形式的关系。因此，在多数情况下，使用非线性排序模型，可以实现更好的搜索效果，即匹配信息的排列顺序更符合用户的实际需求。

[0053] 由于非线性函数的计算复杂度高于线性函数，因此，在同等条件下，使用非线性模型进行排序，其处理速度一般会远远低于线性模型。这里所述的同等条件，包括：使用同样的特征值、处理相同数量的匹配信息。

[0054] 为了实现更高的排序速度，同时保证搜索效果，本实施例所采用的方案是：先使用线性模型对匹配信息进行第一次排序，得到第一排序结果，然后再使用非线性模型对第一排序结果进行第二次排序。其中，第二次排序所处理的匹配信息数量小于第一次排序所处理的匹配信息数量。

[0055] 假设第一次排序所处理的匹配信息数量为N1，可以理解的是，从整体上看，经过第一次排序处理后，排在前面的匹配信息基本上都是比较符合用户需求的，但是由于线性模型的局限性，其具体的排列顺序与用户的实际区别需求可能还有较大的差距。那么，对于这部分信息，可以进一步使用非线性排序模型进行排序处理，即：对于在第一排序结果中靠前的N2条匹配信息，使用非线性排序模型进行排序处理，得到第二排序结果。

[0056] 其中，N2的取值，可以根据具体的搜索需求确定，考虑到一般用户只会关注搜索结果的前几页，因此，可以根据每页可显示的匹配信息条数，为N2选取一个较小的值(相对于N1)，例如200、400等；或者，也可以根据N1来设定N2，例如，将N2的值取为N1的1/10、1/20等。

[0057] 本领域技术人员可以理解的是，相对于线性排序模型，在非线性排序模型中，可以适当减少一些细节特征以提高第二次排序的处理速度，或者适当增加一些细节特征以实现更好的搜索效果。但是，为了保证第一次排序和第二次排序的结果在整体上的一致性，线性排序模型和非线性排序模型所使用的主要特征应该是相同的，当然，线性排序模型和非线性排序模型也可以使用完全相同的特征。

[0058] S104，根据所述第二排序结果，生成搜索结果。

[0059] 搜索引擎根据使用第二次排序的结果，生成最终的搜索结果展现给用户。

[0060] 在本实施例中，首先使用线性排序模型对N1条匹配信息进行排序处理，然后对排序结果的前N2条再使用非线性排序模型进行排序处理。由于线性排序模型的处理速度是能够保证的，因此对于大量(N1条)的匹配信息，首先利用线性排序模型进行预处理，然后通过设置N2＜N1，可以有效减小使用非线性排序模型所处理的数据量，从而提高对匹配信息排序的整体处理速度。

[0061] 实施例二：

[0062] 传统的排序方法，是由人工设计排序模型，其局限性在于只能处理一些简单的特征组合。Learning to Rank(排序学习)是目前比较流行的一种排序方法，与传统的排序方法相比，Learning to Rank方法可以把更多的特征列入考虑。其原理是使用数据样本对排序模型进行训练，令模型学习用户的实际需求，从而使得排序结果更符合用户的实际需求。特别是对于非线性模型，通过训练，可以使排序结果与用户期待的排序结果基本趋于一致。

[0063] 在本申请的优选实施方案中，可以将经训练所确定的线性模型和非线性模型用于第一次排序和第二次排序，由于这类模型所涉及的特征往往比较多，计算复杂度高，因此，为了保证处理速度，可以在第一次排序之前，再增加一个排序预处理的步骤。参见图2所示，本实施例所提供的一种搜索结果生成方法包括以下步骤：

[0064] S201，搜索引擎接收搜索请求，通过检索获得与所述搜索请求相匹配的各条匹配信息；

[0065] S202，对各条匹配信息进行排序预处理。

[0066] S203，使用线性排序模型对经过排序预处理的匹配信息进行排序，得到第一排序结果；

[0067] S204，使用非线性排序模型对所述第一排序结果做进一步排序，得到第二排序结果；

[0068] S205，根据所述第二排序结果，生成搜索结果。

[0069] 本实施例与实施例一相比，主要的区别是增加了一个预处理的步骤S202，其目的是减小使用线性排序模型所处理的数据量。所述预处理，可以是过滤操作，例如滤掉一些过期的、链接无效的匹配信息；也可以是简单的排序操作，一般是采用一些简单传统排序算法，例如TF-IDF，BM25等，这些算法所使用的排序模型由人工设计，所涉及的特征也很少。其特点是速度快，但是相应的排序效果也比较差。

[0070] 可见，从原理上讲，S202对于S203的作用，相当于S203对于S204的作用。排序预处理的速度比线性模型要快很多，而效果也比较差。假设S201中共检索到N0条匹配信息，S202的作用是通过预处理，从N0条信息中选择出N1条匹配信息(或者将N1条匹配信息排在前面)，以供线性排序模型处理。从数量上来讲，N1一般是远小于N0的，因而可以显著提高第一次排序的处理速度。

[0071] 下面以一个简单的示意图，说明排序预处理、第一次排序，第二次排序的关系及效果。首先做一个假设：将所有的匹配信息按照用户的实际需求分为两类：真正相关的匹配信息和一般匹配信息。排序的目的，就是尽量将所有真正相关的匹配信息排在前面。如图3所示，实心圆代表真正相关的匹配信息，空心圆代表一般匹配信息。

[0072] 1)假设N0＝100，在100条匹配信息中共有5条真正相关的匹配信息，经过排序预处理之后，将5条匹配信息全部排在了前10位，如图3a所示。

[0073] 2)取N1＝10，经过第一次排序处理后，排序结果如图3b所示，可见，相对于图3a，5条匹配信息都排在了更为靠前的位置。

[0074] 3)取N2＝6，经过第二次排序处理后，排序结果如图3c所示，可见，5条匹配信息全部被排在了最前面。

[0075] 当然，以上例子仅用于示意性说明，在实际的应用中，N值可能会达到几十万、几百万或更多。而N1和N2的值可以结合排序的模型的复杂程度和实际需求(包括总数据量、用户习惯等)确定，例如，可以将N1设为2000-5000，N2设为100-1000，等等。

[0076] 实施例三：

[0077] 下面将结合一个具体的应用实例，对本申请的搜索结果生成方法进行说明。

[0078] S301，信息搜索系统接收搜索请求，通过检索获得与所述搜索请求相匹配的各条匹配信息；

[0079] S302，对各条匹配信息进行排序预处理。

[0080] S303，使用线性排序模型对经过排序预处理的匹配信息进行排序，得到第一排序结果；

[0081] 本实施例中，取N1＝3000，即预处理结果的前3000条，使用线性模型进行第一次排序，所采用的线性模型为：

[0082] $y_1=0.15x_1+0.1732x_2+0.873x_3+0.245x_4+0.042x_5$

[0083] 其中$x_1$至$x_5$为第一次排序时所考虑的匹配信息的特征，含义如下：

[0084] $x_1$：考虑的特征为：搜索关键词在匹配信息文本中出现的次数，将该次数做归一化处理后即为$x_1$的值。由模型可知，该值越高，则最终计算得到的排序分值越高。

[0085] $x_2$：考虑的特征为：搜索关键词在匹配信息标题中出现的次数，将该次数做归一化处理后即为$x_2$的值。由模型可知，该值越高，则最终计算得到的排序分值越高。

[0086] $x_3$：考虑的特征为：搜索关键词在匹配信息标题中的距离。有时，用户会采用多个关键词进行搜索，这种情况下认为，多个关键词在标题中的距离越小，则越符合用户的需求。$x_3$值的计算方法为：

[0087] $$1-\frac{\text{搜索关键词的距离}}{\text{标题中的总词数}}$$

[0088] 其中，搜索关键词的距离，是以“词”为单位计算的。根据一定的划词规则，可以将任意的词组或短句划分为若干个“词”。举例说明，如果某条匹配信息的标题为“电脑主机和显示器的选购方法”，则根据划词规则，可以将其划分为：电脑/主机/和/显示器/的/选购/方法，共7个词。如果用户搜索的关键词为“电脑”和“显示器”，则在上述标题中，这两个关键词之间隔了两个词，即距离为2，相应的$x_3$值为1-(2/7)＝5/7。

[0089] 可以理解的，如果搜索关键词与标题完全匹配，则关键词的距离为0，$x_3$值取1，如果搜索关键词在标题中没有出现，则$x_3$值取0。

[0090] $x_4$：考虑的特征为：搜索关键词和匹配信息标题的编辑距离。搜索关键词和匹配信息的标题的相似程度，也可以作为计算匹配信息排序得分的一个因素。该相似程度可以以“编辑距离”来衡量。该编辑距离也是以“词”为单位计算的。例如，用户搜索的关键词为：“显

示器”，则与标题“电脑主机和显示器的选购方法”的编辑距离为6，相应的$x_4$值为1-(6/7)=1/7

[0091] 可以理解的是，如果搜索关键词与标题完全匹配，则编辑距离为0，$x_4$值取1，如果搜索关键词在标题中没有出现，则编辑距离为∞，$x_4$值取0。

[0092] $x_5$：考虑的特征为，搜索关键词在匹配信息文本中的IDF(InverseDocument Frequency，反文档频率)值，将IDF做归一化处理后即为$x_5$的值。

[0093] 需要说明的是，为了模型计算方便，上述的$x_1$至$x_5$都是经过归一化处理后的值(即取值在[0,1]区间内)，对于归一化处理的具体方法，本申请实施例不做限定。

[0094] S304，使用非线性排序模型对所述第一排序结果做进一步排序，得到第二排序结果；

[0095] 本实施例中，取N2=600，即第一排序结果的前600条，使用非线性模型进行第二次排序，所采用的非线性模型为：

[0096] $$y_2=\frac{1}{1+e^{-(0.23x_1+0.122x_2+0.7653x_3+0.189x_4+0.156x_5)}}$$

[0097] 其中$x_1$至$x_5$为第二次排序时所考虑的匹配信息的特征，与第一次排序时所考虑的匹配信息的特征相同。

[0098] S305，根据所述第二排序结果，生成搜索结果。

[0099] 本实施例中，所采用的线性模型及非线性模型均为通过训练所确定的模型。本实施例是基于网页搜索或电子商务搜索等应用需求所提出。可以理解的是，这只是本申请技术方案的一种具体的实施方式。事实上，通过选择不同的排序模型，可以将本申请技术方案应用于各类搜索需求，例如图书数据库搜索、文献数据库搜索等。并且应用范围也不局限于互联网领域，其他如单机、局域网中的搜索，都可以应用本申请所提供的技术方案。

[0100] 相应于上面的方法实施例，本申请还提供一种信息搜索系统，参见图4所示，包括：

[0101] 信息检索单元410，用于接收搜索请求，通过检索获得与所述搜索请求相匹配的各条匹配信息；

[0102] 线性排序单元420，用于使用线性排序模型对所述信息检索单元410检索获得的各条匹配信息中的N1条匹配信息进行排序，得到第一排序结果，其中，N1≤所检索到的匹配信息的总数目；

[0103] 非线性排序单元430，用于使用非线性排序模型对所述线性排序单元420排序得到的第一排序结果中的前N2条匹配信息进行排序，得到第二排序结果，其中N2<N1；

[0104] 结果生成单元440，用于根据所述第二排序结果，生成搜索结果。

[0105] 本申请所提供的信息搜索，首先由线性排序单元420使用线性排序模型对N1条匹配信息进行排序处理，然后由非线性排序单元430对排序结果的前N2条再使用非线性排序模型进行排序处理。由于线性排序模型的处理速度是能够保证的，因此对于大量(N1条)的匹配信息，首先利用线性排序模型进行预处理，然后通过设置N2<N1，可以有效减小使用非线性排序模型所处理的数据量，从而提高对匹配信息排序的整体处理速度。

[0106] 参见图5所示，上述的信息搜索系统，还可以包括：

[0107] 排序预处理单元411，用于在所述信息检索单元410获得所述各条匹配信息之后，对所述各条匹配信息进行排序预处理，由所述各条匹配信息中选取N1条匹配信息作为所述

线性排序单元420排序的对象；其中，N1小于所检索到的匹配信息的总数目。

[0108] 使用排序预处理单元411，可以使线性排序单元420减少数据处理量，在不影响最终搜索效果的情况下，进一步提高整个系统的搜索处理速度。

[0109] 以上所提供的信息搜索系统，可以是应用于互联网搜索的搜索引擎，也可以是应用于单机、局域网络的搜索的信息搜索系统。

[0110] 为了描述的方便，描述以上装置时以功能分为各种单元分别描述。当然，在实施本申请时可以把各单元的功能在同一个或多个软件和/或硬件中实现。

[0111] 通过以上的实施方式的描述可知，本领域的技术人员可以清楚地了解到本申请可借助软件加必需的通用硬件平台的方式来实现。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品可以存储在存储介质中，如ROM/RAM、磁碟、光盘等，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例或者实施例的某些部分所述的方法。

[0112] 本说明书中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于系统实施例而言，由于其基本相似于方法实施例，所以描述得比较简单，相关之处参见方法实施例的部分说明即可。以上所描述的系统实施例仅仅是示意性的，其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性劳动的情况下，即可以理解并实施。

[0113] 本申请可用于众多通用或专用的计算系统环境或配置中。例如：个人计算机、服务器计算机、手持设备或便携式设备、平板型设备、多处理器系统、基 于微处理器的系统、置顶盒、可编程的消费电子设备、网络PC、小型计算机、大型计算机、包括以上任何系统或设备的分布式计算环境等等。

[0114] 本申请可以在由计算机执行的计算机可执行指令的一般上下文中描述，例如程序模块。一般地，程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。也可以在分布式计算环境中实践本申请，在这些分布式计算环境中，由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中，程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

[0115] 以上所述仅是本申请的具体实施方式，应当指出，对于本技术领域的普通技术人员来说，在不脱离本申请原理的前提下，还可以做出若干改进和润饰，这些改进和润饰也应视为本申请的保护范围。

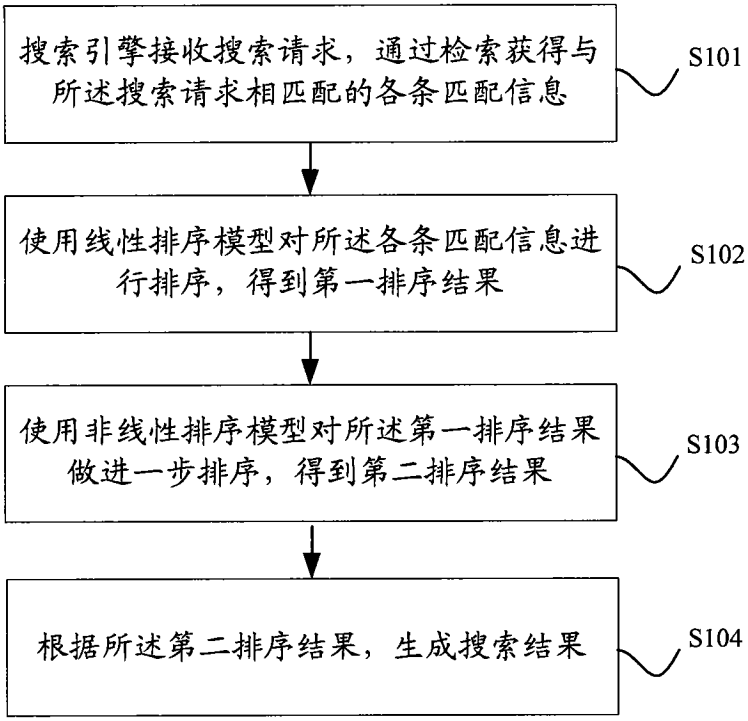

搜索引擎接收搜索请求，通过检索获得与所述搜索请求相匹配的各条匹配信息
S101
使用线性排序模型对所述各条匹配信息进行排序，得到第一排序结果
S102
使用非线性排序模型对所述第一排序结果做进一步排序，得到第二排序结果
S103
根据所述第二排序结果，生成搜索结果
S104

图1

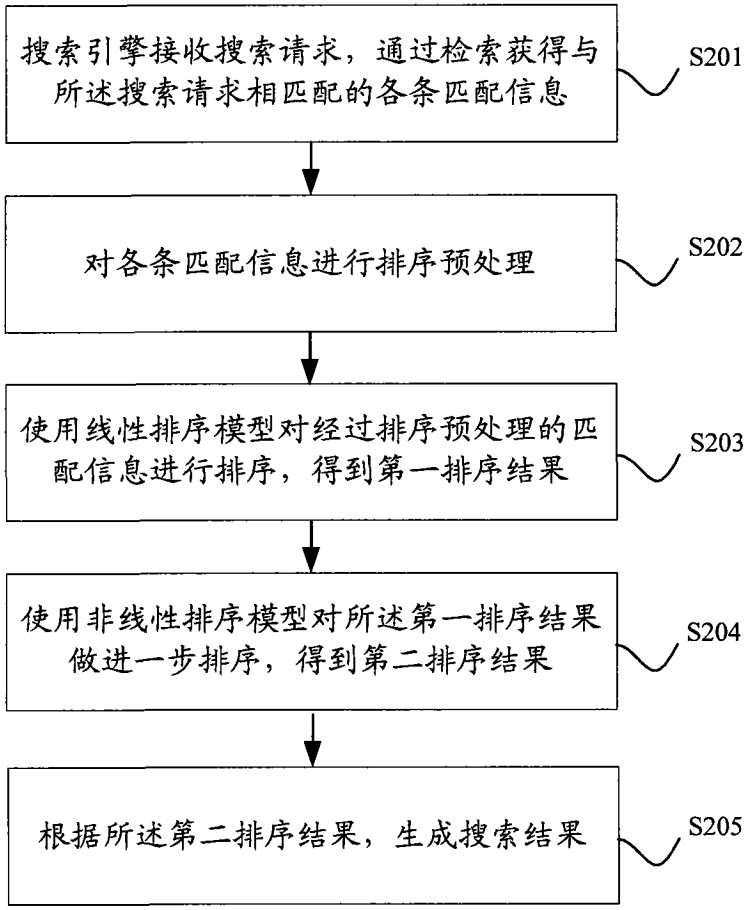
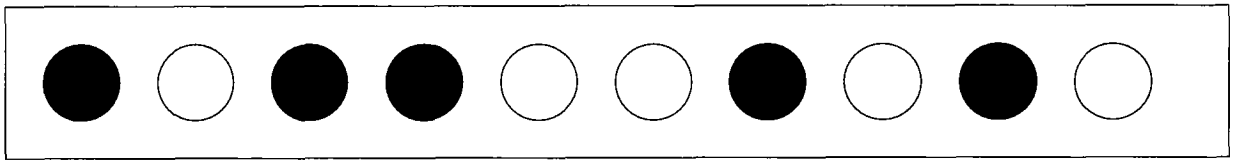
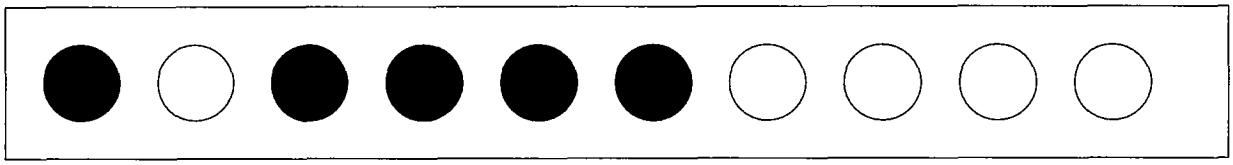

搜索引擎接收搜索请求，通过检索获得与所述搜索请求相匹配的各条匹配信息
S201
对各条匹配信息进行排序预处理
S202
使用线性排序模型对经过排序预处理的匹配信息进行排序，得到第一排序结果
S203
使用非线性排序模型对所述第一排序结果做进一步排序，得到第二排序结果
S204
根据所述第二排序结果，生成搜索结果
S205

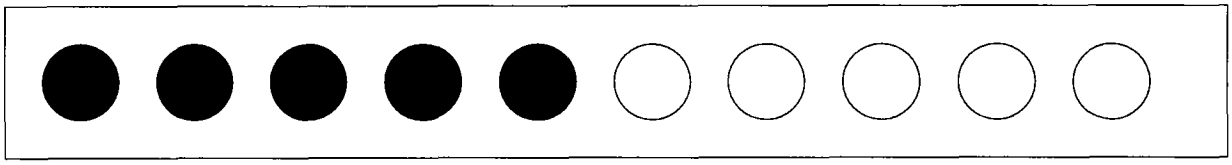
图2



a



b



c

图3

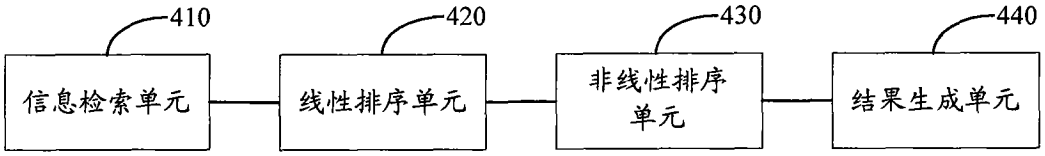

410
420
430
440
信息检索单元
线性排序单元
非线性排序单元
结果生成单元

图4

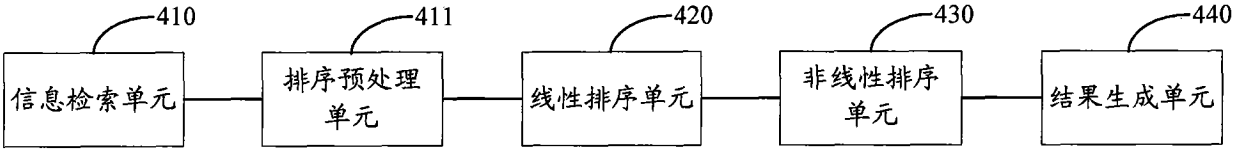

410
411
420
430
440
信息检索单元
排序预处理单元
线性排序单元
非线性排序单元
结果生成单元

图5