



(12)发明专利

(10)授权公告号 CN 105701216 B

(45)授权公告日 2017.03.08

(21)申请号 201610022317.4

(51)Int.Cl.

(22)申请日 2016.01.13

G06F 17/30(2006.01)

(65)同一申请的已公布的文献号

(56)对比文件

申请公布号 CN 105701216 A

CN 104516887 A, 2015.04.15,

CN 104881484 A, 2015.08.02,

(43)申请公布日 2016.06.22

CN 104834698 A, 2015.08.12,

(73)专利权人 北京三快在线科技有限公司

US 8209724 B2, 2012.06.26,

地址 100190 北京市海淀区北四环西路9号

2106-030

审查员 徐波

(72)发明人 陈焯 张涛 王兴 郑志昊 李璟

叶树蕙 沈丹 王观海 张弓

赵晓蕾 管宏

(74)专利代理机构 北京润泽恒知识产权代理有

限公司 11319

代理人 苏培华

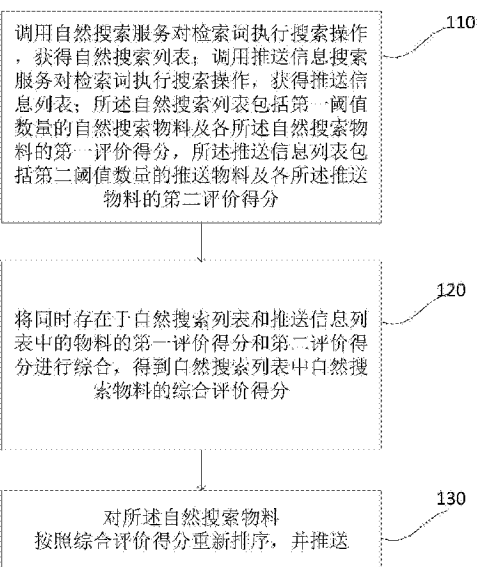
权利要求书2页 说明书12页 附图5页

(54)发明名称

一种信息推送方法及装置

(57)摘要

本申请提供了一种信息推送方法,属于计算机技术领域。本申请通过调用自然搜索服务对检索词执行搜索操作,获得包括第一阈值数量的自然搜索物料和各自然搜索物料第一评价得分的自然搜索列表,及调用推送信息搜索服务对检索词执行搜索操作,获得包括第二阈值数量的推送物料及各所述推送物料的第二评价得分的推送信息列表;并将同时存在于自然搜索列表和推送信息列表中的物料的第一评价得分和第二评价得分进行综合,最后按照综合评价得分将所述自然搜索物料重新排序,并推送,提前了自然搜索列表中具有推送属性的搜索结果的推送位置,即保证了搜索结果的准确性,又保证了具有推送属性的搜索结果的点击率。



1. 一种信息推送方法,其特征在于,包括:

调用自然搜索服务对检索词执行搜索操作,获得自然搜索列表;调用推送信息搜索服务对检索词执行搜索操作,获得推送信息列表;所述自然搜索列表包括第一阈值数量的自然搜索物料及各所述自然搜索物料的第一评价得分,所述推送信息列表包括第二阈值数量的推送物料及各所述推送物料的第二评价得分;

将同时存在于所述自然搜索列表和所述推送信息列表中的物料的第一评价得分和第二评价得分进行综合,得到所述自然搜索列表中自然搜索物料的综合评价得分;

对所述自然搜索物料按照综合评价得分重新排序,并推送。

2. 如权利要求1所述的方法,其特征在于,所述调用自然搜索服务对检索词执行搜索操作,获得自然搜索列表,进一步包括:

调用自然搜索服务,执行自然搜索物料召回,并按照文本相关性选择第一阈值数量的自然搜索物料;

利用预设的排序方法对所述第一阈值数量的自然搜索物料按照排序得分排列,并将该自然搜索物料的排序得分作为第一评价得分。

3. 如权利要求2所述的方法,其特征在于,所述预设的排序方法通过采集搜索日志中的与用户体验指标相关的数据训练得到。

4. 如权利要求1所述的方法,其特征在于,所述调用推送信息搜索服务对检索词执行搜索操作,获得推送信息列表,进一步包括:

调用推送信息搜索服务,执行推送物料召回,并按照文本相关性选择第二阈值数量的推送物料;

根据预设的推送评价参数和所述推送评价参数的得分权重,分别计算所述第二阈值数量的推送物料的第二评价得分。

5. 如权利要求4所述的方法,其特征在于,所述根据预设的推送评价参数和所述推送评价参数的得分权重,分别计算所述第二阈值数量的推送物料的第二评价得分,具体为:

获得当前检索词对应物料的推送评价参数的得分权重;

将所述推送物料的所述推送评价参数值与所述得分权重的乘积作为所述推送物料的第二评价得分。

6. 如权利要求5所述的方法,其特征在于,所述方法还包括:根据搜索日志,在预设的用户体验指标约束下,优化并更新检索词对应物料的推送评价参数的得分权重。

7. 一种信息推送装置,其特征在于,包括:

搜索模块,用于调用自然搜索服务对检索词执行搜索操作,获得自然搜索列表;调用推送信息搜索服务对检索词执行搜索操作,获得推送信息列表;所述自然搜索列表包括第一阈值数量的自然搜索物料及各所述自然搜索物料的第一评价得分,所述推送信息列表包括第二阈值数量的推送物料及各所述推送物料的第二评价得分;

得分综合模块,用于将同时存在于所述自然搜索列表和所述推送信息列表中的物料的第一评价得分和第二评价得分进行综合,得到所述自然搜索列表中自然搜索物料的综合评价得分;

推送模块,用于对所述自然搜索物料按照综合评价得分重新排序,并推送。

8. 如权利要求7所述的装置,其特征在于,所述搜索模块进一步包括:

自然搜索子模块,用于调用自然搜索服务,执行自然搜索物料召回,并按照文本相关性选择第一阈值数量的自然搜索物料;

自然搜索排序子模块,用于利用预设的排序方法对所述第一阈值数量的自然搜索物料按照排序得分排列,并将该自然搜索物料的排序得分作为第一评价得分。

9.如权利要求8所述的装置,其特征在于,所述预设的排序方法通过采集搜索日志中的与用户体验指标相关的数据训练得到。

10.如权利要求7所述的装置,其特征在于,所述搜索模块进一步包括:

推送搜索子模块,用于调用推送信息搜索服务,执行推送物料召回,并按照文本相关性选择第二阈值数量的推送物料;

推送得分计算子模块,用于根据预设的推送评价参数和所述推送评价参数的得分权重,分别计算所述第二阈值数量的推送物料的第二评价得分。

11.如权利要求10所述的装置,其特征在于,所述推送得分计算子模块具体用于:

获得当前检索词对应物料的推送评价参数的得分权重;

将所述推送物料的所述推送评价参数值与所述得分权重的乘积作为所述推送物料的第二评价得分。

12.如权利要求11所述的装置,其特征在于,所述装置还包括:

评价参数更新模块,用于根据搜索日志,在预设的用户体验指标约束下,优化并更新检索词对应物料的推送评价参数的得分权重。

一种信息推送方法及装置

技术领域

[0001] 本申请涉及计算机技术领域,特别是涉及一种信息推送方法及装置。

背景技术

[0002] 随着互联网技术的发展,使用搜索引擎的用户越来越多,互联网用户可以通过搜索引擎获取各种信息,搜索引擎几乎成为网民进入互联网的入口。因此,基于搜索引擎的大用户量的特点,越来越多的企业利用搜索引擎进行信息推送,例如:推送广告信息、推送网站链接等。比较常见的一种基于搜索的信息推送方式是:根据检索词推送与该检索词相关的具有推送标识的信息,提前搜索排名。

[0003] 现有技术中的信息推送方式通常为:搜索服务器接收到Web页面或者终端应用程序发送的检索词之后,以检索词作为输入参数分别调用自然搜索服务和推送信息搜索服务;自然搜索服务根据检索词召回搜索物料,并按照与检索词的相关性以及用户点击反馈率等排序参数得分由高到低的顺序排序,得到自然搜索结果列表;推送信息搜索服务根据检索词召回推送物料,并按照与检索词的相关性进行排序,筛选出相关度较高的预设数量的推送物料,并根据预设推送参数计算所述筛选出的预设数量的推送物料推送得分,并按照得分由高到低降序排列,得到推送信息列表;搜索服务器将推送信息列表中前N(N为正整数)个推送信息和自然搜索结果分别推送至Web页面或者终端应用程序的预设固定位置进行显示。

[0004] 现有技术中的信息推送方式至少存在以下缺陷:推送信息出现在搜索结果的固定位置,可能出现推送信息与检索词的相关度比其附近的自然搜索结果与检索词的相关度差很多的情况,直观表现为搜索结果不准确,导致对搜索结果的点击率下降。

发明内容

[0005] 本申请所要解决的技术问题是:提供一种信息推送方法,解决现有技术中推送信息出现在搜索结果的固定位置,可能出现推送信息与检索词的相关度相比其附近的自然搜索结果与检索词的相关度差很多的情况,从而导致对搜索结果的点击率下降的问题。

[0006] 为了解决上述问题,本申请实施例提供了一种信息推送方法,包括:调用自然搜索服务对检索词执行搜索操作,获得自然搜索结果列表;调用推送信息搜索服务对检索词执行搜索操作,获得推送信息列表;所述自然搜索结果列表包括第一阈值数量的自然搜索物料及各所述自然搜索物料的第一评价得分,所述推送信息列表包括第二阈值数量的推送物料及各所述推送物料的第二评价得分;将同时存在于所述自然搜索结果列表和所述推送信息列表中的物料的第一评价得分和第二评价得分进行综合,得到所述自然搜索结果列表中自然搜索物料的综合评价得分;对所述自然搜索物料按照综合评价得分重新排序,并推送。

[0007] 所述调用自然搜索服务对检索词执行搜索操作,获得自然搜索结果列表,进一步包括:调用自然搜索服务,执行自然搜索物料召回,并按照文本相关性选择第一阈值数量的自然搜索物料;利用预设的排序方法对所述第一阈值数量的自然搜索物料按照排序得分排列,

并将该自然搜索物料的排序得分作为第一评价得分。

[0008] 具体实施时,所述预设的排序方法通过采集搜索日志中的与用户体验指标相关的数据训练得到。

[0009] 所述调用推送信息搜索服务对检索词执行搜索操作,获得推送信息列表,进一步包括:调用推送信息搜索服务,执行推送物料召回,并按照文本相关性选择第二阈值数量的推送物料;根据预设的推送评价参数和所述推送评价参数的得分权重,分别计算所述第二阈值数量的推送物料的第二评价得分。

[0010] 具体实施时,所述根据预设的推送评价参数和所述推送评价参数的得分权重,分别计算所述第二阈值数量的推送物料的第二评价得分,具体为:

[0011] 获得当前检索词对应物料的推送评价参数的得分权重;将所述推送物料的所述推送评价参数值与所述得分权重的乘积作为所述推送物料的第二评价得分。

[0012] 在本申请的一个优选实施例中,所述方法还包括:根据搜索日志,在预设的用户体验指标约束下,优化并更新检索词对应物料的推送评价参数的得分权重。

[0013] 相应的,本申请实施例还提供了一种信息推送装置,包括:搜索模块,用于调用自然搜索服务对检索词执行搜索操作,获得自然搜索列表;调用推送信息搜索服务对检索词执行搜索操作,获得推送信息列表;所述自然搜索列表包括第一阈值数量的自然搜索物料及各所述自然搜索物料的第一评价得分,所述推送信息列表包括第二阈值数量的推送物料及各所述推送物料的第二评价得分;得分综合模块,用于将同时存在于所述自然搜索列表和所述推送信息列表中的物料的第一评价得分和第二评价得分进行综合,得到所述自然搜索列表中自然搜索物料的综合评价得分;推送模块,用于对所述自然搜索物料按照综合评价得分重新排序,并推送。

[0014] 所述搜索模块进一步包括:自然搜索子模块,用于调用自然搜索服务,执行自然搜索物料召回,并按照文本相关性选择第一阈值数量的自然搜索物料;自然搜索排序子模块,用于利用预设的排序方法对所述第一阈值数量的自然搜索物料按照排序得分排列,并将该自然搜索物料的排序得分作为第一评价得分。

[0015] 具体实施时,所述预设的排序方法通过采集搜索日志中的与用户体验指标相关的数据训练得到。

[0016] 所述搜索模块进一步包括:推送搜索子模块,用于调用推送信息搜索服务,执行推送物料召回,并按照文本相关性选择第二阈值数量的推送物料;

[0017] 推送得分计算子模块,用于根据预设的推送评价参数和所述推送评价参数的得分权重,分别计算所述第二阈值数量的推送物料的第二评价得分。

[0018] 具体实施时,所述推送得分计算子模块具体用于:获得当前检索词对应物料的推送评价参数的得分权重;将所述推送物料的所述推送评价参数值与所述得分权重的乘积作为所述推送物料的第二评价得分。

[0019] 在本申请的另一优选实施例中,所述装置还包括:评价参数更新模块,用于根据搜索日志,在预设的用户体验指标约束下,优化并更新检索词对应物料的推送评价参数的得分权重。

[0020] 本申请通过搜索服务器调用自然搜索服务对检索词执行搜索操作,获得包括第一阈值数量的自然搜索物料和各自然搜索物料第一评价得分的自然搜索列表,搜索服务器并

行调用推送信息搜索服务对检索词执行搜索操作,获得包括第二阈值数量的推送物料及各所述推送物料的第二评价得分的推送信息列表;然后,将同时存在于自然搜索列表和推送信息列表中的物料的第一评价得分和第二评价得分进行综合,得到自然搜索列表中该物料的综合评价得分,对于仅存在于自然搜索列表中的物料其综合评价得分等于第一评价得分;最后,按照综合评价得分将所述自然搜索物料重新排序,并推送。通过综合推送信息列表中的物料的第二评价得分,提前了自然搜索列表中具有推送属性的搜索结果的推送位置,即保证了搜索结果的准确性,又保证了具有推送属性的搜索结果的点击率,并且改善了用户体验。

附图说明

[0021] 为了更清楚地说明本申请实施例的技术方案,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。

[0022] 图1是本申请一个实施例的信息推送方法流程图;

[0023] 图2是本申请另一个实施例的信息推送方法流程图;

[0024] 图3是本申请一个实施例中计算推送评价参数的得分权重的流程图;

[0025] 图4是本申请又一个实施例的信息推送方法流程图;

[0026] 图5是本申请一个实施例的信息推送装置结构示意图;

[0027] 图6是本申请另一个实施例的信息推送装置结构示意图;

[0028] 图7是本申请又一个实施例的信息推送装置结构示意图。

具体实施方式

[0029] 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。

[0030] 信息推送有多种方式,包括:接收到用户输入的检索词后,将搜索到的相关信息推送给用户的情况;还包括没有接收到用户的任何触发,主动给用户发送信息的情况。本申请适用于接收到用户输入的检索词后推送相关信息的情况。根据属于本申请中推送的信息可以包括:网站链接、通知、广告等,本申请对此不做限定。对于推送的不同信息,对应有相应的推送信息搜索服务,各推送信息搜索服务根据前端页面发送的检索词在相应的推送数据库中进行检索,可以采用相同的检索方法,也可以根据推送策略的不同采用不同的检索方法,本申请对此不做限定。本申请的实施例中,为了便于理解,采用常用的推送广告的应用场景来描述本申请的信息推送方法。

[0031] 本申请实施例中的自然搜索服务是指搜索引擎中实现自然搜索(Natural Search)的程序模块或硬件,用于在搜索引擎里找到与搜索请求最相关的匹配页面;推送信息搜索服务是搜索引擎中实现非自然搜索的程序模块或硬件,用于在搜索引擎里找到与搜索请求关联的符合预设推送规则的推送信息的程序模块或硬件。对于被搜索的对象,在本

申请的实施例中统称为“物料”，包括：网址、各种网页内容等，例如：网站上的商家信息、产品介绍、文章、图片、网站地址等等。

[0032] 本申请公开的一种信息推送方法，如图1所示，该方法包括：

[0033] 步骤110，调用自然搜索服务对检索词执行搜索操作，获得自然搜索列表；调用推送信息搜索服务对检索词执行搜索操作，获得推送信息列表；所述自然搜索列表包括第一阈值数量的自然搜索物料及各所述自然搜索物料的第一评价得分，所述推送信息列表包括第二阈值数量的推送物料及各所述推送物料的第二评价得分；

[0034] 步骤120，将同时存在于所述自然搜索列表和所述推送信息列表中的物料的第一评价得分和第二评价得分进行综合，得到所述自然搜索列表中自然搜索物料的综合评价得分；

[0035] 步骤130，对所述自然搜索物料按照综合评价得分重新排序，并推送。

[0036] 本申请实施例中，可以由搜索服务器调用自然搜索服务和推送信息搜索服务。

[0037] 在步骤110之前，本申请还包括接收检索词的步骤。例如，用户通过Web页面或者应用程序(APP)的输入界面输入的检索词，由Web页面或者应用程序发送给搜索服务器，搜索服务器在接收到前端发送的检索词后，执行后续的步骤110至步骤130，完成该检索词相关的搜索结果和推送信息的推送操作。具体实施时，搜索服务器还可以通过其它方式接收检索词，例如，在导航系统中，当地理位置变化时，筛选功能根据当前的地理位置，将当前地理位置作为检索词，发送至搜索服务器，同样可以实现搜索功能。

[0038] 上述步骤110中，搜索服务器可以是一台物理机也可以是多台物理机，本申请对此不做限定，所述搜索服务器能够并行、独立地调用自然搜索服务和推送信息搜索服务。上述步骤110中，调用自然搜索服务对检索词执行搜索操作，和调用推送信息搜索服务对检索词执行搜索操作，可以同时执行，也可以先后执行，本申请对上述两个搜索操作的执行顺序不做限定。搜索服务器调用自然搜索服务对输入的检索词执行搜索操作，根据检索词的不同和搜索服务器中内置的数据库的不同可能会搜索到上万条搜索结果，并且，自然搜索服务会获得每个搜索结果的评价得分，用于指示搜索结果的匹配度。为了提高浏览的效率，搜索服务器会根据搜索结果的评价得分对搜索结果进行筛选，仅选择其中的一部分搜索结果进行推送。本申请实施例中的选择第一阈值数量的自然搜索结果进行推送，其中，所述第一阈值数量可以根据被推送端的接收能力或设置确定，如对于被推送端为移动终端的情形，可以设置第一阈值数量为200。同样，搜索服务器调用推送信息搜索服务对输入的检索词执行搜索操作，根据检索词的不同和搜索服务器中内置的推送信息数据库的不同可能会搜索到上百条推送信息，并且，信息推送搜索服务会获得每个推送信息的评价得分，用于指示推送信息的匹配度。根据被推送端的接收能力，搜索服务器会预设最大的推送信息数量，然后根据推送信息的评价得分进行筛选，仅选择其中的一部分推送信息进行推送。本申请实施例中的选择第二阈值数量的推送信息进行推送，例如，所述第二阈值数量为10。所述第二阈值数量通常小于第一阈值数量。

[0039] 待搜索物料可以存储在搜索服务器的数据库中，也可以存储在数据服务器中，包括自然搜索物料和推送物料，其中，推送物料也可以存在于自然搜索物料中，即部分自然搜索物料具有推送属性。每一个物料都设置有唯一的物料标识，例如网址的物料标识为网址的url。上述步骤120，通过对比自然搜索列表中的物料标识是否存在于推送信息列表中，可

以找出存在于自然搜索列表中具有推送属性的物料。具体实施时,遍历所述第一阈值数量的自然搜索物料,当所述推送信息列表中存在所述自然搜索物料对应的推送物料时,即对于所述自然搜索列表中的具有推送属性的物料,将其第一评价得分和推送信息列表中该物料标识的推送物料的第二评价得分之和作为该自然搜索物料的综合评价得分;对于所述自然搜索列表中不具有推送属性的物料的第一评价得分作为其综合评价得分。

[0040] 最后,在步骤130中,对所述自然搜索物料按照综合评价得分重新排序,并按照重新排序后的顺序进行推送。

[0041] 下面以在大众点评客户端搜索“火锅”为例,对本申请的具体实现方案进行说明。用户通过大众点评客户端输入检索词“火锅”,点击搜索后,客户端将检索词“火锅”发送至搜索服务器,搜索服务器分别调用自然搜索服务和推送信息搜索服务执行搜索操作。自然搜索服务在数据库中搜索到10万个物料,然后,按照物料的第一评价得分由高到低的顺序筛选出得分高的10000个物料组成自然搜索列表(如表1);推送信息搜索服务在推送信息库中搜索到1000个物料,然后,按照物料的第二评价得分由高到低的顺序筛选出得分高的100个物料组成推送信息搜索列表(如表2)。

[0042]

物料标识	第一评价得分	其他
老城一锅	85	---
李不管重庆老灶火锅	80	---
阳坊胜利涮肉	79	---
宴渝老灶达州铜锅	70	---

[0043]

.....
-------	-------	-------

[0044] 表1:自然搜索列表

[0045]

物料标识	第二评价得分	其他
宴渝老灶达州铜锅	75	---
小渔棠	60	---
.....

[0046] 表2:推送信息列表

[0047] 遍历自然搜索信息列表中物料,发现物料标识为“宴渝老灶达州铜锅”的物料既存在于自然搜索信息列表中,也存在于推送信息列表中,则标识为“宴渝老灶达州铜锅”的物料的综合评价得分为 $70+75=145$,其他不同时存在于自然搜索列表和推送信息列表中的物料的综合评价得分为该物料在自然搜索信息列表中的第一评价得分。经过综合第二评价得分,对自然搜索列表按照综合评价得分由高到低的顺序重新排序,如表3所示。

[0048]

物料标识	综合评价得分	其他
宴渝老灶达州铜锅	145	---

老城一锅	85	--
李不管重庆老灶火锅	80	--
阳坊胜利涮肉	79	--
.....

[0049] 表3:重新排序的自然搜索列表

[0050] 然后,按照重新排序后的顺序,将自然搜索列表推送至大众点评客户端。

[0051] 以上表格和数据仅作为用于帮助阅读者理解本申请的一个例子,不应作为对本申请的限定。本申请通过搜索服务器调用自然搜索服务对输入的检索词执行搜索操作,获得包括第一阈值数量的自然搜索物料和各自然搜索物料第一评价得分的自然搜索列表;搜索服务器并行调用推送信息搜索服务对输入的检索词执行搜索操作,获得包括第二阈值数量的推送物料及各所述推送物料的第二评价得分的推送信息列表;然后,将同时存在于所述自然搜索列表和所述推送信息列表中的物料的第一评价得分和第二评价得分进行综合,得到所述自然搜索列表中该物料的综合评价得分,对于仅存在于所述自然搜索列表中的物料其综合评价得分等于第一评价得分;最后,按照综合评价得分将所述自然搜索物料重新排序,并推送。通过综合推送信息列表中的物料的第二评价得分,提前了自然搜索列表中具有推送属性的搜索结果的推送位置,既保证了搜索结果的准确性,又保证了具有推送属性的搜索结果的点击率,并且改善了用户体验。

[0052] 在本申请的一个实施例中,如图2所示,所述搜索服务器调用自然搜索服务对检索词执行搜索操作,获得自然搜索列表,进一步包括:

[0053] 步骤1101,搜索服务器调用自然搜索服务,执行自然搜索物料召回,并按照文本相关性选择第一阈值数量的自然搜索物料;

[0054] 步骤1102,利用预设的排序方法对所述第一阈值数量的自然搜索物料按照排序得分排列,并将该自然搜索物料的排序得分作为第一评价得分。

[0055] 搜索服务器调用自然搜索服务,执行自然搜索物料召回,在数据库中搜索到海量物料后,利用简单排序方法,如按照文本相关性由高到低的顺序,或按照文本相关性由低到高的顺序对海量物料进行排序,然后对海量物料进行筛选,选择相关性较高的第一阈值数量的物料,作为待返回的自然搜索物料,加入自然搜索列表。文本相关性得分可以通过tf-idf算法计算检索词与物料标题的tf-idf值得到。如果按照文本相关性由高到低的顺序对搜索到海量物料进行排序,则选择前第一阈值数量的物料,作为待返回的自然搜索物料,加入自然搜索列表;如果按照文本相关性由低到高的顺序对搜索到海量物料进行排序,则选择后第一阈值数量的物料,作为待返回的自然搜索物料,加入自然搜索列表。

[0056] 然后,在步骤1102中,利用预设的排序方法对筛选出的所述第一阈值数量的自然搜索物料进行重新排序。所述预设的排序方法在执行过程中,按照排序得分由高到低的顺序输出重新排序后的所述第一阈值数量的自然搜索物料。将所述第一阈值数量的自然搜索物料的排序得分作为相应的物料的第一评价得分。

[0057] 具体实施时,所述预设的排序方法可以使用复杂排序算法,包括监督学习排序方法learningToRank、网页排名pagerank等,训练排序方法时,以用户体验为优化目标,因此,所述预设的排序方法通过采集搜索日志中的与用户体验指标相关的数据训练得到。其中,与用户体验指标相关的数据包括但不限于:点击次数、文本相关性。本申请通过预设的排序

方法得出能够反映用户体验的排序得分,作为自然搜索结果的第一评价得分,不限定使用那种排序方法,在本实施例中通过一种常用复杂排序方法:pair wise的ranking svm方法,来介绍排序方法的训练过程。

[0058] 首先,收集后端的自然搜索的查询日志。后端的查询日志通常保存在搜索服务器上。查询日志包括:查询唯一标识(queryid),城市,查询(query)类目,经纬度(geohash),检索词,自然搜索列表 d_i (物料id)、文本相关性得分(t_i)、类目相关性得分(c_i)等。其中,文本相关性得分通过tf-idf算法计算检索词与物料标题的tf-idf值得到;类目相关性得分可以利用余弦相似度算法计算query类目和物料类目之间的相似度,做为类目相关性。

[0059] 然后,收集用户在前端自然搜索列表的点击日志,包括:查询唯一标识(queryid),自然搜索列表 d_i (物料id)。

[0060] 再后,根据后端的自然搜索的查询日志和前端自然搜索列表的点击日志生成排序方法训练数据。每条训练数据格式为(label,queryid,城市,query类目,经纬度,检索词, t_i,c_i),其中queryid、城市、query类目、经纬度、检索词 t_i,c_i 来自后端的自然搜索的查询日志, t_i 和 c_i 是物料 d_i 与检索词的文本相关性和物料 d_i 与query类目的类目相似度,Label值根据每个后端的自然搜索的查询日志对应的用户点击次数Count设置。

[0061] 根据每个后端的自然搜索的查询日志对应的用户点击次数Count设置训练数据的Label值,具体如下:通过queryid关联后端的自然搜索的查询日志和前端自然搜索列表的点击日志,找出每个后端的自然搜索的查询日志对应的所有前端用户的自然搜索列表点击日志(queryid, d_i),计算对应用户点击次数Count。对每个后端的自然搜索的查询日志展开成N条训练数据,N等于自然搜索列表中的物料个数。设一条后端的自然搜索的查询日志有count条用户点击,count ≥ 0 ,如果count=0,则每条训练数据的label等于0;如果count > 0 ,则遍历后端的自然搜索的查询日志中的物料id列表,第1个被点击的物料label等于count,第2个被点击的物料label等于count-1,第i个被点击的物料label等于count+1-i,其中 $i \leq \text{count}$,其他未被点击的物料label等于0。

[0062] 最后,用pair wise的机器学习方法ranking svm,对上述训练数据进行学习,训练复杂排序模型,生成模型参数文件。至此排序方法训练完成。

[0063] 在自然搜索服务执行自然搜索物料召回后,并按照文本相关性筛选出第一阈值数量的自然搜索物料后,对每个物料生成对应的排序输入参数(queryid,城市,query类目,经纬度,检索词, t_i,c_i),用ranking svm加载训练得到的模型参数文件,计算每个物料对应的score,所有物料按照score由高到低的顺序排列,即完成复杂排序,每个物料对应的score作为该物料的第一评价得分。

[0064] 具体实施时,上述ranking svm方法使用的训练数据的特征还可以包括更多特征,也可以使用更复杂训练方法,本申请对具体的排序方法训练方法和训练数据不做限定。在本申请的实施例中,通过引入与用户体验指标相关的特征(如:文本相关性、点击次数)训练排序模型,使返回的自然搜索结果更符合用户体验。

[0065] 本实施例中,利用简单排序方法对自然搜索结果进行第一轮排序,并筛选,减少搜索结果的数量,可以提高推送的效率;利用预设的排序方法对筛选后的自然搜索结果进行再次排序,可以通过在预设的排序方法中引入搜索服务器关注的排序因子,提高自然搜索结果的针对性,从而进一步改善用户体验。

[0066] 在本申请的另一具体实施例中,如图2所示,所述搜索服务器调用推送信息搜索服务对输入的检索词执行搜索操作,获得推送信息列表,进一步包括:

[0067] 步骤1103,搜索服务器调用推送信息搜索服务,执行推送物料召回,并按照文本相关性选择第二阈值数量的推送物料;

[0068] 步骤1104,根据预设的推送评价参数和所述推送评价参数的得分权重,分别计算所述第二阈值数量的推送物料的第二评价得分。

[0069] 具体实施时,搜索服务器调用推送信息搜索服务,执行推送物料召回,在推送数据库中搜索到海量物料后,利用简单排序方法,如按照文本相关性由高到低的顺序或按照文本相关性由低到高的顺序,对海量物料进行筛选,选择相关性较高的第二阈值数量的物料,作为待推送物料,加入推送信息列表。具体实施时,文本相关性得分可以通过tf-idf算法计算检索词与物料标题的tf-idf值得到。如果按照文本相关性由高到低的顺序对搜索到海量物料进行排序,则选择前第一阈值数量的物料,作为待返回的自然搜索物料,加入自然搜索列表;如果按照文本相关性由低到高的顺序对搜索到海量物料进行排序,则选择后第一阈值数量的物料,作为待返回的自然搜索物料,加入自然搜索列表。

[0070] 然后,根据预设的推送评价参数和所述推送评价参数的得分权重,分别计算所述第二阈值数量的推送物料的第二评价得分。其中,推送评价参数的得分权重 λ 可以根据经验预先设定。具体实施时,可以为:首先,获得当前检索词对应物料的推送评价参数的得分权重 λ ;然后,将所述推送物料的所述推送评价参数值与所述得分权重 λ 的乘积作为所述推送物料的第二评价得分。物料的推送评价参数代表对该类物料的推送效果进行衡量的指标,不同类别的推送物料的推送评价参数略有不同,以推送广告物料为例,广告物料的推送参数为广告的有效成本(effective cost per mille)指的就是每一千次展示可以获得的广告收入,或者点击率。其中,推送物料的推送评价参数预先定义,所述推送评价参数的得分权重在搜索服务器中存储,根据搜索日志离线计算得到。

[0071] 下面以推送物料为广告物料为例,对根据预设的推送评价参数和所述推送评价参数的得分权重,分别计算所述第二阈值数量的推送物料的第二评价得分的步骤进行详细说明。假设预设的推送评价参数为广告物料的有效成本(eCPM),现有技术中,根据公式 $eCPM = CTR * Bid$ 可以计算得到广告物料的有效成本(eCPM),其中,CTR代表用户点击广告的概率,Bid表示广告出价。在推送数据库中,每个广告物料包括CTR和Bid信息,其中,Bid广告基本属性(广告主为一次点击的出价),为预先设置的数值;CTR根据动态计算获得。动态计算获得CTR可以采用现有技术中的计算模型,此处不再赘述。搜索服务器中读取推送评价参数的得分权重 λ ,根据公式 $AdScore = \lambda * CTR * Bid$,计算广告物料的第二评价得分AdScore。

[0072] 本实施例中仅以广告物料为例对根据预设的推送评价参数和所述推送评价参数的得分权重,分别计算所述第二阈值数量的推送物料的第二评价得分的步骤进行详细说明,本申请的推送物料不限于广告物料,推送物料的评价参数也限于广告物料的有效成本(eCPM),计算第二评价得分的方法也不限于上述公式。以上具体实施例不应作为对本申请的限定。

[0073] 本实施例中,利用简单排序方法对自然搜索结果进行第一轮排序,并筛选,减少搜索结果的数量,可以提高推送的效率;利用推送评价参数的得分权重计算推送物料的第二评价得分,可以动态平衡具有推送属性的物料在返回的自然搜索结果列表中的排列位置,从而进一步改善用户体验。

[0074] 若不考虑信息推送的效果, λ 可以设置为0; 若既要考虑信息推送的效果, 同时兼顾自然搜索结果的用户体验, 则需要结合第一评价得分和推送物料的评价参数设定 λ 的值。设定推送评价参数的得分权重 λ 的具体方法为: 收集查询日志; 按照业务规则对收集的查询日志分类; 以每一个分类中的日志作为样本, 采用批量梯度下降算法计算推送评价参数的得分权重 λ 。每一次搜索过程的都生成一条查询日志, 每一次搜索过程的日志采用一个列表来记录。例如, 你在东直门搜“火锅”的日志生成一个列表, 我在西直门搜“火锅”的日志生成另一个列表; 我在东直门搜“火锅”的日志生成又一个列表。一个列表对应一次搜索。每个列表中都包括: 唯一的查询标识、城市、查询类目、经纬度、检索词等查询条件, 还包括: 查询发生时间、返回给用户的所有结果 (即根据综合评价得分排序后的自然搜索列表)、推送信息列表、根据第一评价得分排序的自然搜索列表、排序因子等等各种信息。其中, 自然搜索列表中还包括: 物料标识、物料的第一评价得分; 推送信息列表中还包括推送物料的推送评价参数。然后, 按照业务规则对查询日志分类, 可以采用要素分类法, 如采用 (城市, 类目, 经纬度, 检索词) 的分类方法。最后, 以每一个分类中的日志作为样本, 采用批量梯度下降算法计算推送评价参数的得分权重 λ 。

[0075] 下面结合内容搜索中推送广告的例子详细说明设定推送评价参数的得分权重 λ 的具体方法。假设, 广告物料的推荐评价参数为用户点击广告的概率CTR和广告出价Bid; 自然搜索物料的排序因子为用户体验得分RelevanceScore。搜索系统的搜索服务器或日志服务器会记录查询日志, 即每一次搜索的日志, 每一次搜索的日志记录在一个列表中, 列表中包括查询发生时间、返回给用户的所有结果 (即根据综合评价得分排序后的自然搜索列表)、推送信息列表、根据第一评价得分排序的自然搜索列表、排序因子等等各种信息。其中, 根据第一评价得分排序的自然搜索列表中还包括: 物料标识id、物料的第一评价得分RelevanceScore; 推送信息列表中还包括推送物料的推送评价参数CTR和Bid。

[0076] 然后, 按照业务规则对查询日志分类, 如采用 (城市, 类目, 经纬度, 检索词) 的分类方法对收集的查询日志分类, 可以得到多个类别, 每个类别中又包含多个记录查询日志的列表。

[0077] 最后, 以每一个分类中的查询日志作为样本, 采用批量梯度下降算法计算推送评价参数的得分权重 λ 。如图3所示, 具体方法包括步骤310至步骤380。

[0078] 步骤310, 假设当前分类有K个列表, 计算当前分类的总用户体验得分Rel。首先, 设 $\lambda=0$, 对当前分类下每个列表中物料按照RelevanceScore降序排序, 计算每个列表前M个位置的物料的用户体验得分 $Rel_i = RelevanceScore_1 + RelevanceScore_2 + RelevanceScore_i + \dots + RelevanceScore_m$, 该分类的总用户体验得分为 $Rel = Rel_1 + Rel_2 + \dots + Rel_i + \dots + Rel_k$ 。

[0079] 步骤320, λ 取0和1之间的一个随机值, 计算当前分类的总混合排序后用户体验得分 Rel'_k 。然后, λ 取 (0, 1) 之间的一个随机值, 根据公式:

[0080] $RankScore = RelevanceScore + \lambda * CTR * Bid$;

[0081] 计算该分类中每个列表中物料的综合评价得分RankScore, 对每个列表按照RankScore由高到低的顺序排列。然后, 取每个列表前M个位置的用户体验得分 $Rel'_i = RelevanceScore_1 + RelevanceScore_2 + RelevanceScore_i + \dots + RelevanceScore_m$, 该分类的总用户体验得分为 $Rel' = Rel'_1 + Rel'_2 + \dots + Rel'_i + \dots + Rel'_k$ 。

[0082] 步骤330, 判断用户体验损失是否超过预设值, 若超过, 则返回步骤320, 重新设置 λ

值;否则,继续执行对 λ 值调优的步骤。用户体验损失等于 $(Re1-Re1')/Re1$,用户体验损失目标值预设 $X\%$,若 $(Re1-Re1')/Re1 > X\%$,则确定为用户体验损失超过预设目标值,需要重新估算 λ 值。否则,确定为用户体验损失低于预设目标值,在此条件下,进一步调优 λ 值。

[0083] 步骤340,根据广告收益与 λ 的增长对应关系,计算 λ 增长 $\Delta\lambda$ 时,广告收益的增长值。广告收益Revenue是RankScore的函数,表示为: $g(\text{RankScore})$,计算公式为:

[0084] $\text{Revenue} = g(\text{RankScore}) = g(\text{RelevanceScore} + \lambda * \text{CTR} * \text{Bid})$ 。

[0085] 针对每个列表,分别计算广告收益 $g(\lambda)$ 对 λ 的偏导数 $\frac{\partial g}{\partial \lambda}$ 。计算方法为 λ 增加一个

微小量 $\Delta\lambda$,计算广告收益的增长 ΔRev ,表示为 $\frac{\partial g}{\partial \lambda} = \frac{\Delta\text{Rev}}{\Delta\lambda}$ 。

[0086] 步骤350,根据当前分类下每个列表的 λ 的学习结果,更新 λ 值。更新 λ 值的公式为:

$\lambda = \lambda + \alpha \sum_i^K \frac{\partial g}{\partial \lambda_i}$,其中, α 是学习速率参数,可以根据学习效果和速度人工调优, K 为当前类别中列表的数量。

[0087] 步骤360,利用更新后的 λ 值重新计算混合排序后的用户体验得分。利用更新后的 λ 值重新计算当前分类中每个列表中物料的综合评价得分RankScore,对每个列表按照RankScore由高到低的顺序排列。然后,取每个列表前 M 个位置的用户体验得分进一步计算该分类的混合排序后的用户体验得分 $Re1'$,具体方式参见步骤320,此处不再赘述。

[0088] 步骤370,判断用户体验损失是否小于或等于预设值,若是,则重复执行步骤340至步骤360,继续优化 λ 值;否则,执行步骤380。利用更新后的混合排序后的用户体验得分 $Re1'$,判断是否满足 $(Re1-Re1')/Re1 \leq X\%$ 的约束,若是,则重复执行步骤340至步骤360,继续优化 λ 值。

[0089] 步骤380,将更新前的 λ 值作为最优值进行保存。当确定为当前 λ 值过大,将步骤350中更新前的 λ 值作为该类别的物料对应的推送评价参数的得分权重最优值进行保存,用于计算第二评价得分。

[0090] 通过在一定的用户体验损伤约束下,最大限度提高 λ 值,可以在设定的用户体验损伤约束下最大化推送评价参数的得分权重,从而最大化广告收益。本实施例仅以推送信息为广告为例进行阐述,但本申请的应用范围不限于推送广告。可以理解,本申请的方法还是用于推送其他信息,推送不同信息时,优化 λ 的思路不变,仅需针对 $g(\lambda)$ 的变化对计算公式做相应调整即可。

[0091] 基于前述实施例,本申请的另一优选实施中,如图4所示,所述方法还包括:

[0092] 步骤140,根据搜索日志,在预设的用户体验指标约束下,优化并更新检索词对应物料的推送评价参数的得分权重 λ 。每个类别的 λ 值可以通过离线批量梯度下降算法进行优化,然后上传至搜索系统,也就可以通过在线批量梯度下降算法进行优化,并实时更新。

[0093] 通过不断调优推送评价参数的得分权重 λ ,可以有效适应由于不同的查询时间、查询地点等的变化导致的搜索结果变化,使推送信息总是排列在搜索结果的恰当位置,不仅有效地改善了用户体验,也保证了具有推送属性的搜索结果的点击率。

[0094] 相应地,本申请还公开了一种信息推送装置,如图5所示,包括:

[0095] 搜索模块510,用于调用自然搜索服务对检索词执行搜索操作,获得自然搜索列

表;调用推送信息搜索服务对检索词执行搜索操作,获得推送信息列表;所述自然搜索列表包括第一阈值数量的自然搜索物料及各所述自然搜索物料的第一评价得分,所述推送信息列表包括第二阈值数量的推送物料及各所述推送物料的第二评价得分;

[0096] 得分综合模块520,用于同时存在于所述自然搜索列表和所述推送信息列表中的物料的第一评价得分和第二评价得分进行综合,得到所述自然搜索列表中自然搜索物料的综合评价得分;

[0097] 推送模块530,用于对所述自然搜索物料按照综合评价得分重新排序,并推送。

[0098] 本申请通过搜索服务器调用自然搜索服务对输入的检索词执行搜索操作,获得包括第一阈值数量的自然搜索物料和各自然搜索物料第一评价得分的自然搜索列表,同时,搜索服务器并行调用推送信息搜索服务对输入的检索词执行搜索操作,获得包括第二阈值数量的推送物料及各所述推送物料的第二评价得分的推送信息列表;然后,将同时存在于自然搜索列表和推送信息列表中的物料的第一评价得分和第二评价得分进行综合,得到自然搜索列表中该物料的综合评价得分,对于仅存在于自然搜索列表中的物料其综合评价得分等于第一评价得分;最后,按照综合评价得分将所述自然搜索物料重新排序,并推送。通过综合推送信息列表中的物料的第二评价得分,提前了自然搜索列表中具有推送属性的搜索结果的推送位置,即保证了搜索结果的准确性,又保证了具有推送属性的搜索结果的点击率,并且改善了用户体验。

[0099] 进一步地,如图6所示,所述搜索模块进一步包括:

[0100] 自然搜索子模块5101,用于调用自然搜索服务,执行自然搜索物料召回,并按照文本相关性选择第一阈值数量的自然搜索物料;

[0101] 自然搜索排序子模块5102,用于利用预设的排序方法对所述第一阈值数量的自然搜索物料按照排序得分排列,并将该自然搜索物料的排序得分作为第一评价得分。

[0102] 其中,所述预设的排序方法通过采集搜索日志中的与用户体验指标相关的数据训练得到。所述用户体验指标相关的数据包括但不限于:文本相关度、点击次数。预设的排序方法的训练过程参见方法实施例部分,此处不再赘述。

[0103] 本实施例中,利用简单排序方法对自然搜索结果进行第一轮排序,并筛选,减少搜索结果的数量,可以提高推送的效率;利用预设的排序方法对筛选后的自然搜索结果进行再次排序,可以通过在预设的排序方法中引入搜索服务器关注的排序因子,提高自然搜索结果的针对性,从而进一步改善用户体验。

[0104] 进一步地,如图6所示,所述搜索模块进一步包括:

[0105] 推送搜索子模块5103,用于调用推送信息搜索服务,执行推送物料召回,并按照文本相关性选择第二阈值数量的推送物料;

[0106] 推送得分计算子模块5104,用于根据预设的推送评价参数和所述推送评价参数的得分权重,分别计算所述第二阈值数量的推送物料的第二评价得分。

[0107] 具体实施时,所述推送得分计算子模块具体用于:获得当前检索词对应物料的推送评价参数的得分权重;将所述推送物料的所述推送评价参数值与所述得分权重的乘积作为所述推送物料的第二评价得分。当前检索词对应物料的推送评价参数的得分权重获取过程参见方法实施例部分和图3,此处不再赘述。

[0108] 本实施例中,利用简单排序方法对自然搜索结果进行第一轮排序,并筛选,减少搜

索结果的数量,可以提高推送的效率;利用推送评价参数的得分权重计算推送物料的第二评价得分,可以动态平衡具有推送属性的物料在返回的自然搜索列表中的排列位置,从而进一步改善用户体验。

[0109] 优选的,如图7所示,所述装置还包括:

[0110] 评价参数更新模块540,用于根据搜索日志,在预设的用户体验指标约束下,优化并更新检索词对应物料的推送评价参数的得分权重。每个类别的 λ 值可以通过离线批量梯度下降算法进行优化,然后上传至搜索系统,也就可以通过在线批量梯度下降算法进行优化,并实时更新。

[0111] 通过不断调优推送评价参数的得分权重 λ ,可以有效适应由于不同的查询时间、查询地点等的变化导致的搜索结果变化,使推送信息总是排列在搜索结果的恰当位置,不仅有效地改善了用户体验,也保证了具有推送属性的搜索结果的点击率。

[0112] 本说明书中的各个实施例均采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似的部分互相参见即可。对于装置实施例而言,由于其与方法实施例基本相似,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

[0113] 以上对本申请提供的一种信息推送方法、装置进行了详细介绍,本文中应用了具体个例对本申请的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本申请的方法及其核心思想;同时,对于本领域的一般技术人员,依据本申请的思想,在具体实施方式及应用范围上均会有改变之处,综上所述,本说明书内容不应理解为对本申请的限制。

[0114] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到各实施方式可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件实现。基于这样的理解,上述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计算机可读存储介质中,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)执行各个实施例或者实施例的某些部分所述的方法。

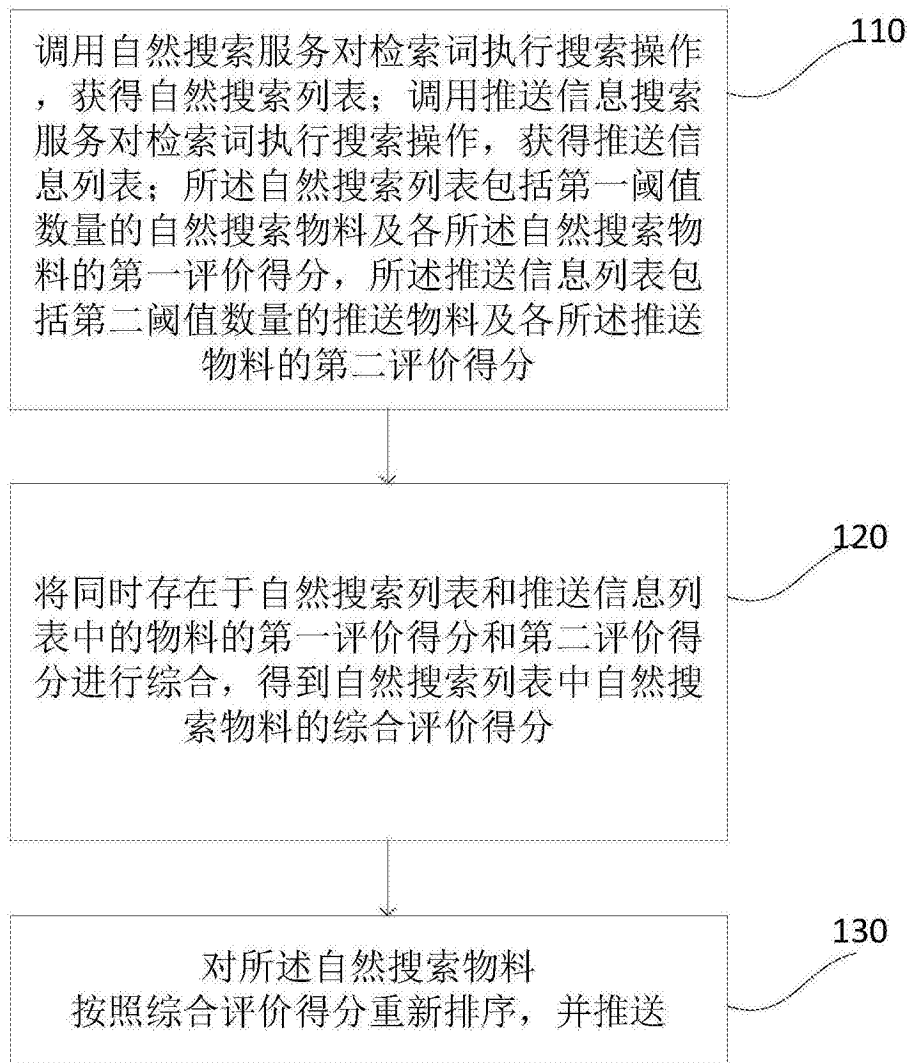


图1

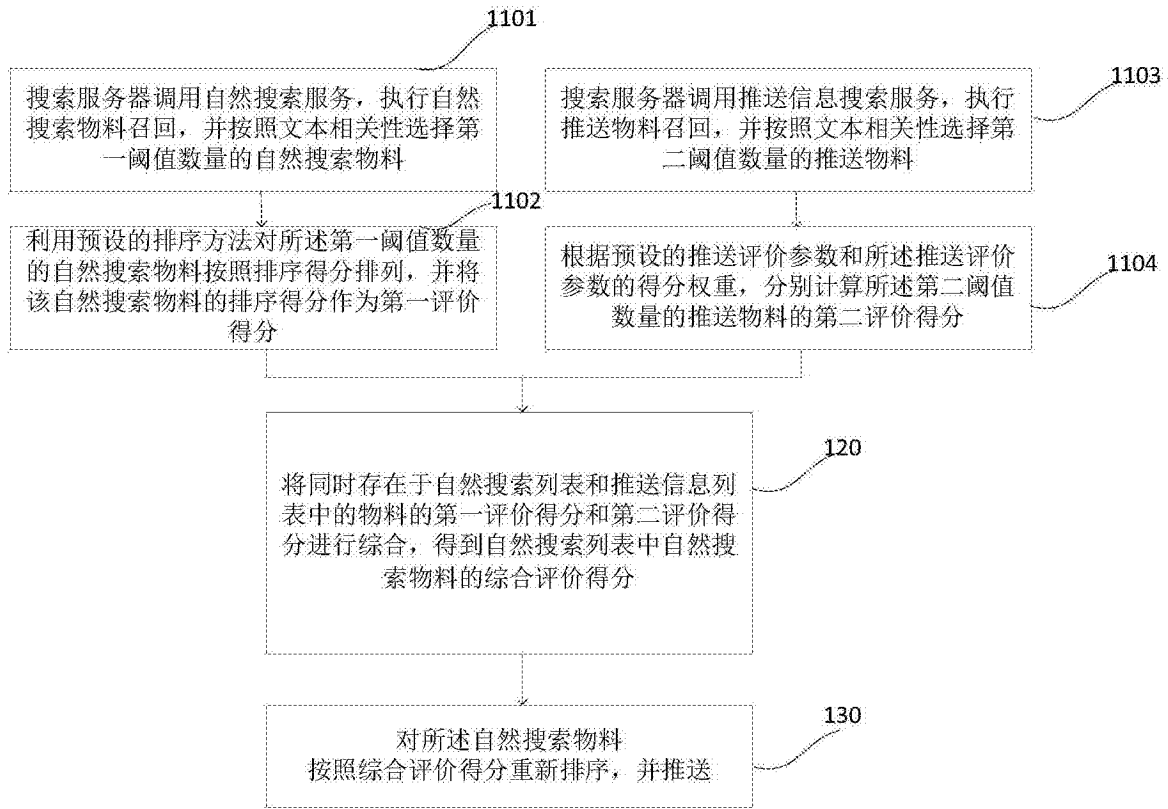


图2

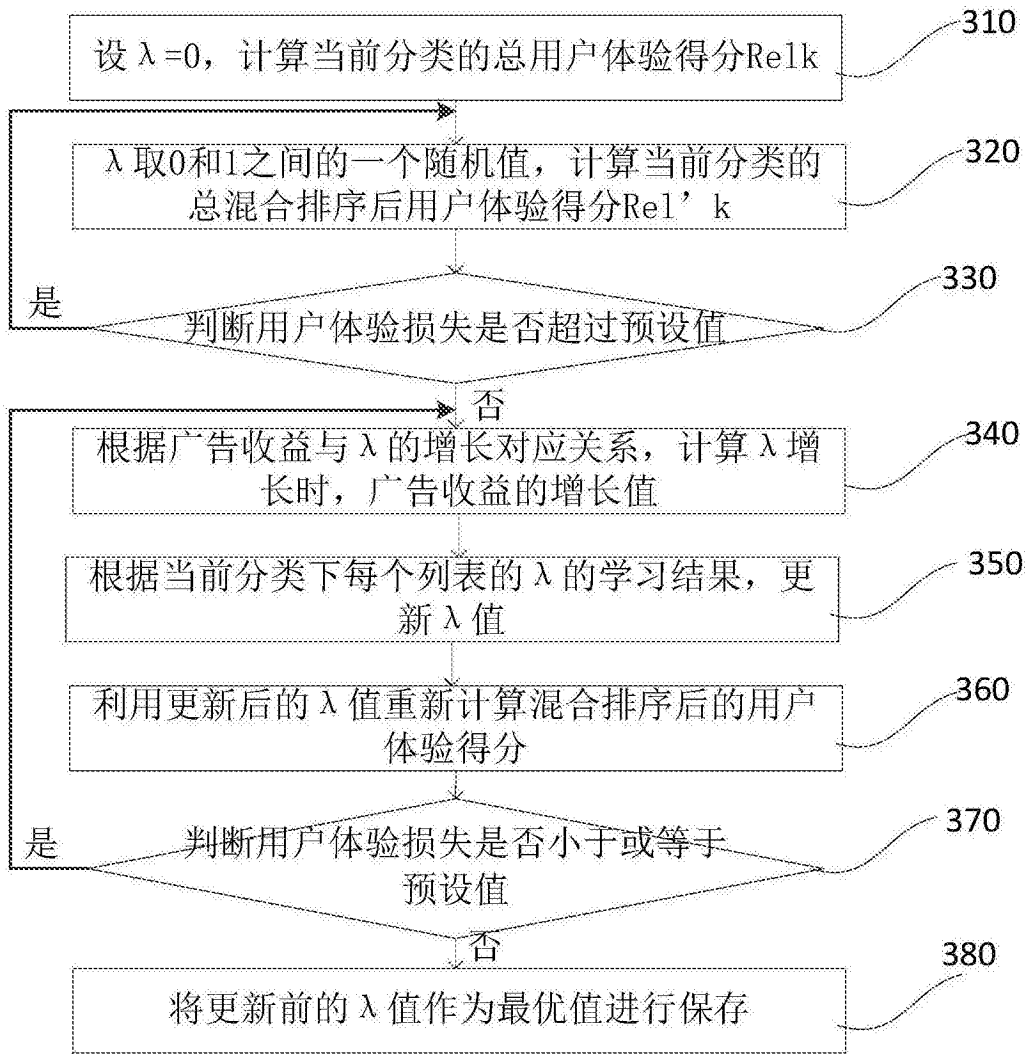


图3

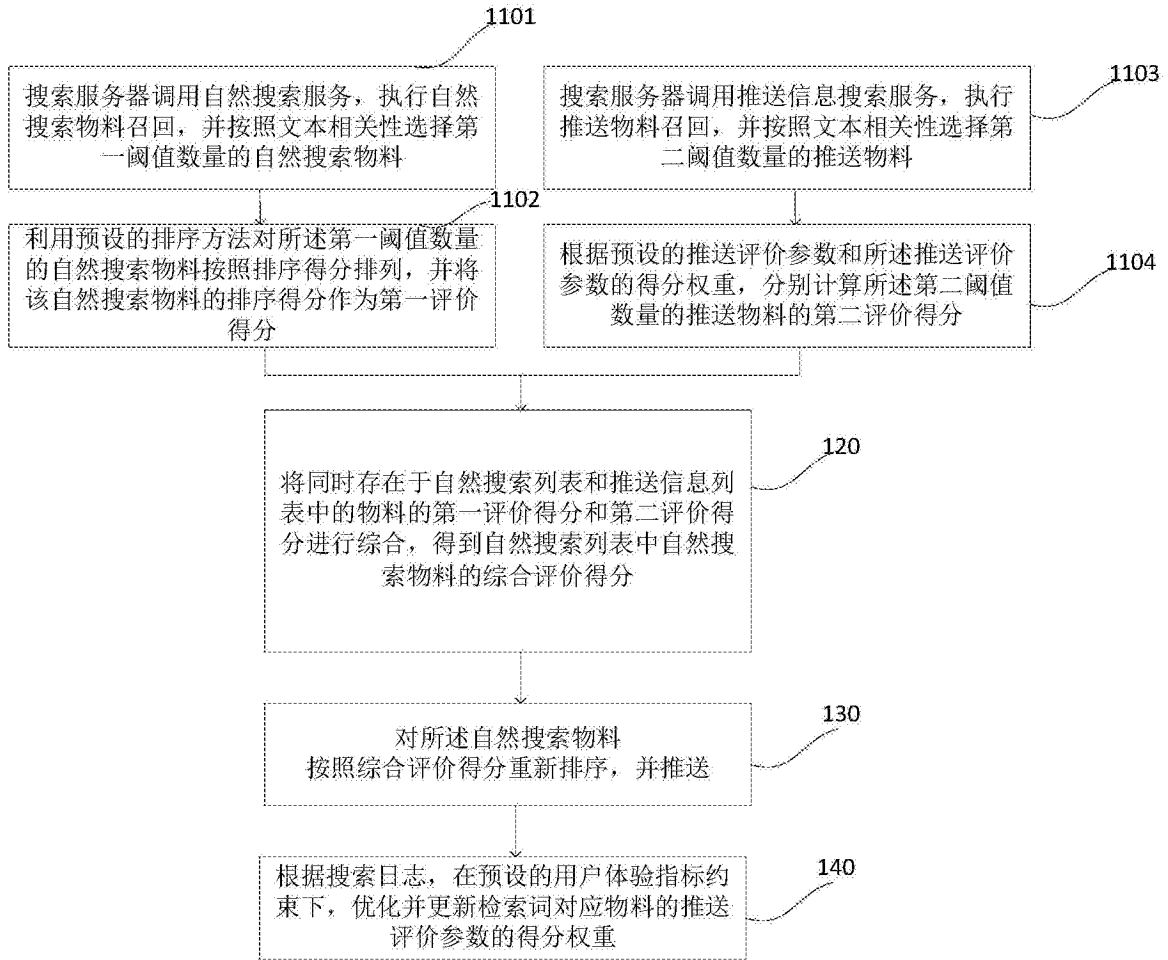


图4



图5



图6

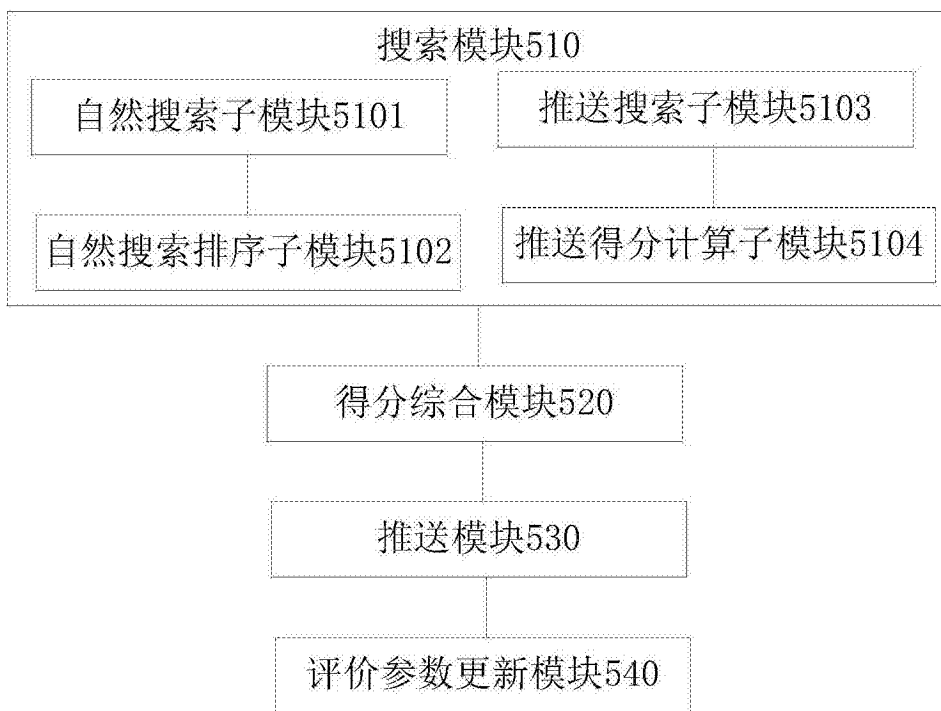


图7