

网站推荐方法和系统以及网络服务器

申请号: [201110288443.1](#)

申请日: 2011-09-26

申请(专利权)人 [中国科学院计算机网络信息中心](#)
地址 [100190 北京市海淀区中关村南四街四号](#)
发明(设计)人 [吴军 王欣 金键](#)
主分类号 [H04L29/08 \(2006.01\) I](#)
分类号 [H04L29/08 \(2006.01\) I](#)
公开(公告)号 [102316166A](#)
公开(公告)日 [2012-01-11](#)
专利代理机构 [北京同立钧成知识产权代理有限公司](#) [11205](#)
代理人 [丁琛](#)



(12) 发明专利申请

(10) 申请公布号 CN 102316166 A

(43) 申请公布日 2012. 01. 11

(21) 申请号 201110288443. 1

(22) 申请日 2011. 09. 26

(71) 申请人 中国科学院计算机网络信息中心
地址 100190 北京市海淀区中关村南四街四号

(72) 发明人 吴军 王欣 金键

(74) 专利代理机构 北京同立钧成知识产权代理有限公司 11205
代理人 丁琛

(51) Int. Cl.
H04L 29/08 (2006. 01)

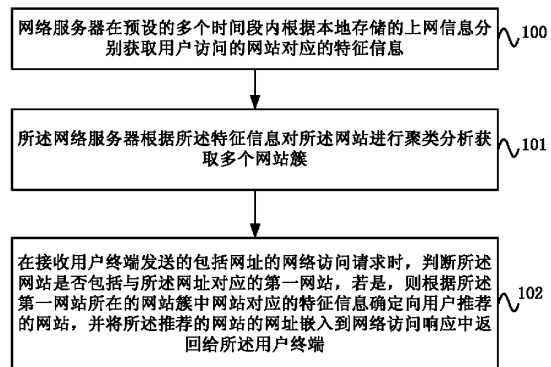
权利要求书 2 页 说明书 7 页 附图 2 页

(54) 发明名称

网站推荐方法和系统以及网络服务器

(57) 摘要

本发明提供一种网站推荐方法和系统以及网络服务器,其中,该方法包括:网络服务器根据本地存储的上网信息在预设的多个时间段内分别获取被用户访问过的网站对应的特征信息,根据特征信息对网站进行聚类分析获取多个网站簇,以便在接收用户终端发送的包括网址的网络访问请求时,若判断获知网站包括与网址对应的第一网站,则根据第一网站所在的网站簇中网站对应的特征信息确定向用户推荐的网站,并将推荐的网站的网址嵌入到网络访问响应中返回给用户终端,实现了网络服务器能够基于网站对应的用户网络访问行为向进行网络访问的用户推荐更多的网站,从而使用户获取更多感兴趣的资讯。



1. 一种网站推荐方法,其特征在于,包括:

网络服务器在预设的多个时间段内根据本地存储的上网信息分别获取用户访问的网站对应的特征信息;

所述网络服务器根据所述特征信息对所述网站进行聚类分析获取多个网站簇,以便在接收用户终端发送的包括网址的网络访问请求时,判断所述网站是否包括与所述网址对应的第一网站,若是,则根据所述第一网站所在的网站簇中网站对应的特征信息确定向用户推荐的网站,并将所述推荐的网站的网址嵌入到网络访问响应中返回给所述用户终端。

2. 根据权利要求1所述的网站推荐方法,其特征在于,若判断获知所述网站没有包括与所述网址对应的第一网站,所述方法还包括:

所述网络服务器向其余网络服务器广播包括所述网址和所述多个时间段的上网信息查询请求,若接收到所述其余网络服务器返回的所述第一网站在所述多个时间段内的上网信息,则根据所述上网信息获取所述第一网站对应的特征信息;

所述网络服务器根据每个网站簇中网站对应的特征信息获取对应的聚集轮廓信息,并根据所述第一网站对应的特征信息和所述聚集轮廓信息通过相似性度量确定所述第一网站所属的网站簇。

3. 根据权利要求1所述的网站推荐方法,其特征在于,所述根据所述特征信息对所述网站进行聚类分析包括:

根据所述特征信息通过分裂法、层次法、基于密度的方法、基于网格的方法和基于模型的方法对所述网站进行聚类分析。

4. 根据权利要求1所述的网站推荐方法,其特征在于,所述根据所述第一网站所在的网站簇中网站对应的特征信息确定向用户推荐的网站包括:

根据所述第一网站所在的网站簇中的其余网站对应的特征信息获取所述其余网站的推荐分数;

根据所述推荐分数按照预设的推荐准则确定向用户推荐的网站。

5. 根据权利要求1-4任一项所述的网站推荐方法,其特征在于,所述特征信息包括:各预设时间段内所述网站被用户访问的频率特征、方差特征和熵特征中的至少一种特征。

6. 根据权利要求1-4任一项所述的网站推荐方法,其特征在于,所述网址包括:域名和/或IP地址。

7. 一种网络服务器,其特征在于,包括:

第一获取模块,用于在预设的多个时间段内根据本地存储的上网信息分别获取用户访问的网站对应的特征信息;

第二获取模块,用于根据所述特征信息对所述网站进行聚类分析获取多个网站簇;

判断模块,用于在接收用户终端发送的包括网址的网络访问请求时,判断所述网站是否包括与所述网址对应的第一网站;

处理模块,用于若判断获知所述网站包括与所述网址对应的第一网站,则根据所述第一网站所在的网站簇中网站对应的特征信息确定向用户推荐的网站,并将所述推荐的网站的网址嵌入到网络访问响应中返回给所述用户终端。

8. 根据权利要求7所述的网络服务器,其特征在于,所述处理模块,还用于:

若判断获知所述网站没有包括与所述网址对应的第一网站,则向其余网络服务器广播

包括所述网址和所述多个时间段的上网信息查询请求,若接收到所述其余网络服务器返回的所述第一网站在所述多个时间段内的上网信息,则根据所述上网信息获取所述第一网站对应的特征信息;

根据每个网站簇中网站对应的特征信息获取对应的聚集轮廓信息,并根据所述第一网站对应的特征信息和所述聚集轮廓信息通过相似性度量确定所述第一网站所属的网站簇。

9. 根据权利要求 8 所述的网络服务器,其特征在于,所述第二获取模块具体用于:

根据所述特征信息通过分裂法、层次法、基于密度的方法、基于网格的方法和基于模型的方法对所述网站进行聚类分析。

10. 一种网站推荐系统,其特征在于,包括如权利要求 7 或 8 或 9 任一项所述的网络服务器,以及用户终端。

网站推荐方法和系统以及网络服务器

技术领域

[0001] 本发明涉及通信技术,尤其涉及一种网站推荐方法和系统以及网络服务器。

背景技术

[0002] 随着电子信息技术的发展,网络已经改变了人们的生活方式,举例来说,人们可以利用网络获取自己感兴趣的书籍、电影、音乐、甚至商品,因此,网络带给了我们高效便捷的生活,人们已经习惯利用计算机、手机等具有上网功能的设备,通过浏览自己感兴趣的网页进行学习、娱乐、购物来满足自身多方位的需求。

[0003] 人们利用网络可以更加高效的获取丰富的信息进行学习和娱乐,具体地,网络服务器会根据用户访问的网站类型向其推荐同一种类型的相关网站供用户参考,比如用户访问的是属于信息技术类型的网站,网络服务器会向用户推荐信息技术类型中的其他网站供用户参考;网络服务器会存储用户经常访问的网站类型并获取相关的网站推荐给用户,从而使用户可以获取更多感兴趣的资讯。

[0004] 但是,现有技术中的网络服务器只是根据用户访问的网站类型获取相关的网站推荐给用户供用户参考,使用户获得的信息有限,具有一定的局限性。

发明内容

[0005] 针对现有技术的上述缺陷,本发明实施例提供一种网站推荐方法和系统以及网络服务器。

[0006] 本发明实施例提供一种网站推荐方法,包括:

[0007] 网络服务器在预设的多个时间段内根据本地存储的上网信息分别获取用户访问的网站对应的特征信息;

[0008] 所述网络服务器根据所述特征信息对所述网站进行聚类分析获取多个网站簇,以便在接收用户终端发送的包括网址的网络访问请求时,判断所述网站是否包括与所述网址对应的第一网站,若是,则根据所述第一网站所在的网站簇中网站对应的特征信息确定向用户推荐的网站,并将所述推荐的网站的网址嵌入到网络访问响应中返回给所述用户终端。

[0009] 本发明实施例提供一种网络服务器,包括:

[0010] 第一获取模块,用于在预设的多个时间段内根据本地存储的上网信息分别获取用户访问的网站对应的特征信息;

[0011] 第二获取模块,用于根据所述特征信息对所述网站进行聚类分析获取多个网站簇;

[0012] 判断模块,用于在接收用户终端发送的包括网址的网络访问请求时,判断所述网站是否包括与所述网址对应的第一网站;

[0013] 处理模块,用于若判断获知所述网站包括与所述网址对应的第一网站,则根据所述第一网站所在的网站簇中网站对应的特征信息确定向用户推荐的网站,并将所述推荐的

网站的网址嵌入到网络访问响应中返回给所述用户终端。

[0014] 本发明实施例提供一种网站推荐系统,包括上述的网络服务器以及用户终端。

[0015] 本发明实施例提供的网站推荐方法和系统以及网络服务器,通过网络服务器根据本地存储的上网信息在预设的多个时间段内分别获取被用户访问过的网站对应的特征信息,根据特征信息对网站进行聚类分析获取多个网站簇,当接收到用户终端发送的包括网址的网络访问请求时,判断经过聚类分析的网站是否包括与该网址对应的第一网站。若包括,则根据第一网站所在的网站簇中网站对应的特征信息确定向用户推荐的网站,并将推荐的网站的网址嵌入到网络访问响应中返回给用户终端,实现了网络服务器能够基于网站对应的用户网络访问行为向进行网络访问的用户推荐更多的网站,从而使用户获取更多感兴趣的资讯。

附图说明

[0016] 图 1 为本发明网站推荐方法一个实施例的流程图;

[0017] 图 2 为本发明网站推荐方法另一实施例的流程图;

[0018] 图 3 为本发明网络服务器一个实施例的结构示意图;

[0019] 图 4 为本发明网站推荐系统一个实施例的结构示意图。

具体实施方式

[0020] 为使本发明实施例的目的、技术方案和优点更加清楚,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。

[0021] 图 1 为本发明网站推荐方法一个实施例的流程图,如图 1 所示,该方法包括:

[0022] 步骤 100,网络服务器在预设的多个时间段内根据本地存储的上网信息分别获取用户访问的网站对应的特征信息;

[0023] 用户可以通过手机、计算机等具有上网功能的用户终端向网络服务器发送网络访问请求进行网络访问,网络服务器能够按照预设的刷新时间存储一段时间内进行网络访问的用户的上网信息。可以理解的是,本实施例中网络服务器的刷新时间是根据具体的应用需要进行设置的比如三天或者一个星期。网络服务器存储的用户的上网信息具体包括:用户终端的 IP 地址、每次访问的网站和对应的开始时间和结束时间。

[0024] 网络服务器在预设的多个时间段内根据本地存储的上网信息分别获取被用户访问过的网站对应的特征信息,需要说明的是,本实施例中的特征信息反映了被用户访问过的网站在预设的不同的时间段内所对应的用户访问网站的行为特征,特征信息具体可以包括在每个预设时间段内网站被用户访问的频率特征、方差特征和熵特征中的至少一种特征,其中,频率特征反映了网站在各预设的时间段内被用户访问的频繁度;方差特征反映了网站在各预设的时间段内被用户访问的次数的方差,用来衡量网站在各预设时间段内被用户访问的次数变化的剧烈程度;熵特征反映了网站在各预设的时间段内被访问的用户的 IP 地址的熵,用来衡量网站的用户稳定性,举例来说,在上午八点至十点的时间段内,网站 A 一共被访问了 5 次,IP1 访问了 1 次,IP2 访问了 3 次,IP3 访问了 1 次,则用户的 IP 地址

的熵为： $-(1/5) \log(1/5) + (3/5) \log(3/5) + (1/5) \log(1/5)$ 。

[0025] 本实施例中预设的多个时间段可以根据具体的应用情况预先在网络服务器进行设置,举例来说,若预先设置的多个时间段为每天的 8:00 ~ 10:00、10:00 ~ 12:00、18:00 ~ 21:00 和 21:00 ~ 24:00,即网络服务器在设置的各时间段内根据存储的上网信息进行统计获取在每个时间段内被用户访问过的每个网站对应的特征信息。在具体的实现过程中,网络服务器会将获取的特征信息比如:频率特征、方差特征和熵特征中的至少一种特征进行模数转换,或对其中几种特征的数字量进行加权获取对应的特征信息。值得注意的是,本实施例中的特征信息并不局限于上述列举的几种特征,还可以根据获取的具体上网信息进行调整从而获取其他的特征信息,具体的处理过程如上,此处不再赘述。为了更清楚的说明特征信息含义,举例说明如表 1 所示,表 1 表示了预设的多个时间段内被用户访问过的网站对应的特征信息,特征信息是在每个预设的时间段内针对被用户访问过的每个网站的频率特征、方差特征和熵特征进行模数转换加权后获取的数值。

[0026] 表 1

网站 \ 时段	8:00 ~	10:00 ~	18:00 ~	21:00 ~
	10:00	12:00	21:00	24:00
A	15	5	0	0
B	0	0	32	4
C	12	0	0	56
D	9	47	0	0

[0028] 步骤 101,所述网络服务器根据所述特征信息对所述网站进行聚类分析获取多个网站簇;

[0029] 网络服务器根据在多个预设时间段内获取的被用户访问过的网站对应的特征信息,对所有网站进行聚类分析获取多个网站簇。聚类分析(Cluster Analysis)又称群分析,是将数据分类到不同的类或者簇这样的一个过程,所以同一个簇中的对象有很大的相似性,而不同簇间的对象有很大的相异性。聚类分析的计算方法主要包括分裂法(partitioning methods)、层次法(hierarchical methods)、基于密度的方法(density-based methods)、基于网格的方法(grid-based methods)和基于模型的方法(model-based methods)。每一种聚类方法的具体实施过程属于现有技术,为了更清楚的说明聚类分析的过程,以基于分裂法中的 K-均值和以基于模型的方法中的利用概率潜在语义模型(Probabilistic Latent Semantic Analysis, PLSA)为例进行具体说明,其余的聚类方法不再一一赘述。

[0030] 1:介绍 K-均值的算法如下:

[0031] 步骤(1):当网站簇 $k = 2$ 为例作说明,在网站 A 至网站 Z 中随机选择 2 个网站作为初始质心(类别的中心),假设选择网站 A 和网站 B;

[0032] 步骤(2):对于剩下的每一个网站根据特征信息计算其到每个质心的距离,比较每个网站到网站 A 和网站 B 的距离,选择距离大的划分到一个类中;对剩下的网站依次类推,最终可以将所有的网站划分到以网站 A 和网站 B 为质心的两个类中;

[0033] 步骤 (3) :对每一个类重新计算质心,计算方法为将各网站的权重求平均,计算出每个类的新质心后,对于所有的网站,计算其到每个质心的距离,如此反复,直到质心不再发生变化。

[0034] 步骤 (4) :对于每一个类计算类内均方误差,即类内所有网站到质心的距离,比较它们的均方误差,趋势应该为逐渐减小,当均方误差值由显著下降到不那么显著下降的 K 值就可以作为最终的 K,即网站簇的个数。

[0035] 2 :介绍 PLSA 模型的算法如下 :

[0036] 步骤 (1) :定义似然函数 $L(D, T) = \sum_{i=1}^m \sum_{j=1}^n w_{t_j}^{d_i} \log P(d_i, t_j)$;

[0037] 步骤 (2) :利用 EM 算法的迭代计算确定参数值,EM 算法其具体步骤如下所示 :

[0038] E 步利用旧参数计算潜在因素变量的后验概率,公式如下所示 :

$$[0039] \quad P(z_k | d_i, t_j) = \frac{P(z_k) \cdot P(d_i | z_k) \cdot P(t_j | z_k)}{\sum_{k'=1}^l P(z_{k'}) \cdot P(d_i | z_{k'}) \cdot P(t_j | z_{k'})}$$

[0040] M 步通过最大化似然函数的期望函数求得新参数,公式如下所示 :

$$[0041] \quad Q = E(L^C) = \sum_{i=1}^m \sum_{j=1}^n w_{d_i}^{t_j} \sum_{k=1}^l P(z_k | d_i, t_j) \log P(d_i, t_j)$$

[0042] 步骤 (3) :最大化过程中更新各参数的公式如下所示 :

$$[0043] \quad P(z_k) = \frac{\sum_{i=1}^m \sum_{j=1}^n w_{d_i}^{t_j} P(z_k | d_i, t_j)}{\sum_{i=1}^m \sum_{j=1}^n \sum_{k'=1}^l w_{d_i}^{t_j} P(z_{k'} | d_i, t_j)} = \frac{\sum_{i=1}^m \sum_{j=1}^n w_{d_i}^{t_j} P(z_k | d_i, t_j)}{\sum_{i=1}^m \sum_{j=1}^n w_{d_i}^{t_j}}$$

$$[0044] \quad P(d_i | z_k) = \frac{\sum_{j=1}^n w_{d_i}^{t_j} P(z_k | d_i, t_j)}{\sum_{i'=1}^m \sum_{j=1}^n w_{d_{i'}}^{t_j} P(z_k | d_{i'}, t_j)}$$

$$[0045] \quad P(t_j | z_k) = \frac{\sum_{i=1}^m w_{d_i}^{t_j} P(z_k | d_i, t_j)}{\sum_{i=1}^m \sum_{j'=1}^n w_{d_i}^{t_{j'}} P(z_k | d_i, t_{j'})}$$

[0046] 步骤 (4) :重复上述 E 步和 M 步的计算过程,在此过程中始终会单调递增,当达到最大值时,参数值确定,更新过程停止。

[0047] 步骤 102,在接收用户终端发送的包括网址的网络访问请求时,判断所述网站是否包括与所述网址对应的第一网站,若是,则根据所述第一网站所在的网站簇中网站对应的特征信息确定向用户推荐的网站,并将所述推荐的网站的网址嵌入到网络访问响应中返回给所述用户终端。

[0048] 网络服务器接收到用户终端发送的包括网址的网络访问请求时,根据网址查询经过聚类的网站判断是否包括与该网址对应的第一网站。若判断获知经过聚类的网站中包括该第一网站,说明该第一网站也经过了聚类分析,根据网址查询步骤 101 中获取的网站簇并确定第一网站所在的网站簇,由于上网信息中用户访问过的网站的表现形式网址,其中,网站的网址包括域名或 IP 地址,可以通过域名服务器对域名和 IP 地址进行转换确定用户

访问的网站。基于上述可以获知基于用户对网站的访问行为来看,该网站簇中的网站具有相似性。获取第一网站所在的网站簇中的网站对应的特征信息,并根据特征信息按照设置的推荐规则确定向用户推荐的网站,举例来说,可以从第一网站所在的网站簇中除第一网站之外随机选择一定数量的网站推荐给用户,因为用户访问的第一网站所对应的用户访问行为与该网站簇中其余网站对应的用户访问行为相似,本实施例可以基于网站对应的用户访问行为向用户推荐用户可能感兴趣的网站。需要说明的是,推荐规则根据具体的应用场景进行具体设置,本实施例不对具体的推荐规则作限制。

[0049] 网络服务器将向用户推荐的网站的网址嵌入到网络访问响应中返回给用户终端。其中,网站的网址包括域名和 / 或 IP 地址,网络服务器上的上网信息中存储的被用户访问的网站的网址是用域名或 IP 地址来表示的,若根据上网信息判断获知向用户推荐的网站的网址是 IP 地址,网络服务器可以直接将 IP 地址嵌入到网络访问响应中返回给用户终端,也可以向域名服务器发送包括 IP 地址的域名反查询请求,域名服务器通过 PTR 类型的域名解析向网络服务器返回与 IP 地址对应的域名,网络服务器将网站的 IP 地址和对应的域名都嵌入到网络访问响应中返回给用户终端供用户进行参考,向用户终端返回域名,方便用户记忆和书写,从而使用户更加方便的对推荐的网站进行检索和访问。若根据上网信息判断获知向用户推荐的网站的网址是域名,网络服务器可以直接将域名嵌入到网络访问响应中返回给用户终端,也可以向域名服务器发送包括域名的域名查询请求,域名服务器通过 A 类型的域名解析向网络服务器返回与域名对应的 IP 地址,网络服务器将网站的 IP 地址和对应的域名都嵌入到网络访问响应中返回给用户终端供用户进行参考,向用户终端返回 IP 地址,从而使用户更加直接的对推荐的网站进行检索和访问,不需要向域名服务器发起域名查询请求。

[0050] 本实施例提供的网站推荐方法,通过网络服务器根据本地存储的上网信息在预设的多个时间段内分别获取被用户访问过的网站对应的特征信息,根据特征信息对网站进行聚类分析获取多个网站簇,当接收到用户终端发送的包括网址的网络访问请求时,判断经过聚类分析的网站是否包括与该网址对应的第一网站。若包括,则根据第一网站所在的网站簇中网站对应的特征信息确定向用户推荐的网站,并将推荐的网站的网址嵌入到网络访问响应中返回给用户终端,实现了网络服务器能够基于网站对应的用户网络访问行为向进行网络访问的用户推荐更多的网站,从而使用户获取更多感兴趣的资讯。

[0051] 图 2 为本发明网站推荐方法另一实施例的流程图,如图 2 所示,该方法包括:

[0052] 步骤 200,网络服务器在预设的多个时间段内根据本地存储的上网信息分别获取用户访问的网站对应的特征信息;

[0053] 步骤 201,所述网络服务器根据所述特征信息对所述网站进行聚类分析获取多个网站簇;

[0054] 步骤 202,所述网络服务器在接收用户终端发送的包括网址的网络访问请求时,判断所述网站是否包括与所述网址对应的第一网站,若不是,则向其余网络服务器广播包括所述网址和所述多个时间段的上网信息查询请求,若接收到所述其余网络服务器返回的所述第一网站在所述多个时间段内的上网信息,则根据所述上网信息获取所述第一网站对应的特征信息;

[0055] 网络服务器接收到用户终端发送的包括网址的网络访问请求时,根据网址查询经

过聚类的网站判断是否包括与该网址对应的第一网站。若判断获知经过聚类的网站中不包括该第一网站,说明该第一网站在各预设的时间段内没有被用户通过该网络服务器访问过,也就是在各预设的时间段内用户通过该网络服务器访问的网站不包括该第一网站。

[0056] 网络服务器向互联网系统中的其余网络服务器广播包括第一网站的网址和各预设时间段的上网信息查询请求,其余的网络服务器根据接收到的上网信息查询请求,各网络服务器均根据第一网站的网址从本地存储的各预设时间段内的上网信息中查询是否包括该第一网站的上网信息,若该网络服务器能够接收到其余网络服务器返回的在各预设时间段内第一网站的上网信息,根据第一网站的上网信息获取该第一网站对应的特征信息,具体的特征信息获取过程参见上述实施例一中的步骤 100,此处不再赘述。

[0057] 步骤 203,所述网络服务器根据每个网站簇中网站对应的特征信息获取对应的聚集轮廓信息,并根据所述第一网站对应的特征信息和所述聚集轮廓信息通过相似性度量确定所述第一网站所属的网站簇;

[0058] 网络服务器根据上述步骤 201 中获取的每个网站簇中网站的特征信息获取对应的聚集轮廓信息,聚集轮廓信息即每一个网站簇中的网站对应的特征信息的平均权重;

[0059] 网络服务器根据第一网站的特征信息和获取的聚集轮廓信息进行相似性度量,值得注意的是,相似性度量的方法很多例如皮尔森相关系数或者余弦系数等,本实施例不作具体限制。通过相似性度量获取第一网站对应的特征信息与各个聚集轮廓信息的匹配分数以确定第一网站所属的网站簇,匹配分数越大,说明第一网站与该网站簇中的网站的相似度越高,选择最大匹配分数的网站簇确定为第一网站所属的网站簇。

[0060] 步骤 204,根据所述第一网站所在的网站簇中网站对应的特征信息确定向用户推荐的网站,并将所述推荐的网站的网址嵌入到网络访问响应中返回给所述用户终端。

[0061] 获取第一网站所在的网站簇中的网站的特征信息,并根据特征信息按照设置的推荐规则确定向用户推荐的网站,具体地,可以对第一网站所在的网站簇中的其余网站的特征信息进行加权平均获取其余网站的推荐分数,根据其余的每个网站的推荐分数按照预设的推荐准则确定向用户推荐的网站,比如根据每个网站的推荐分数从高往低进行排列直到预设的推荐网站的数量为止,将选出来的网站作为向用户推荐的网站。网络服务器将推荐的网站的网址嵌入到网络访问响应中返回给用户终端供用户进行参考,具体过程参见上述实施例,此处不再赘述。

[0062] 本实施例中的步骤 201 和步骤 202 的具体实施过程参见图 1 所示的实施例,此处不再赘述。

[0063] 本实施例提供的网站推荐方法,通过网络服务器根据本地存储的上网信息在预设的多个时间段内分别获取被用户访问过的网站对应的特征信息,根据特征信息对网站进行聚类分析获取多个网站簇,当接收到用户终端发送的包括网址的网络访问请求时,若判断获知经过聚类的网站不包括与网址对应的第一网站,则向其余网络服务器进行广播查询,若接收到其余网络服务器返回的第一网站的上网信息,则确定第一网站所在的网站簇,并根据第一网站所在的网站簇中网站的特征信息确定向用户推荐的网站,并将推荐的网站的网址嵌入到网络访问响应中返回给用户终端,实现了网络服务器能够基于网站对应的用户网络访问行为向进行网络访问的用户推荐更多的网站,从而使用户获取更多感兴趣的资讯。

[0064] 本领域普通技术人员可以理解：实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成，前述的程序可以存储于一计算机可读取存储介质中，该程序在执行时，执行包括上述方法实施例的步骤；而前述的存储介质包括：ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0065] 图3为本发明网络服务器一个实施例的结构示意图，如图3所示，该网络服务器包括：第一获取模块11、第二获取模块12、判断模块13和处理模块14，其中，第一获取模块11用于在预设的多个时间段内根据本地存储的上网信息分别获取用户访问的网站对应的特征信息；第二获取模块12用于根据特征信息对网站进行聚类分析获取多个网站簇；判断模块13用于在接收用户终端发送的包括网址的网络访问请求时，判断网站是否包括与网址对应的第一网站；处理模块14用于若判断获知网站包括与网址对应的第一网站，则根据第一网站所在的网站簇中网站对应的特征信息确定向用户推荐的网站，并将推荐的网站的网址嵌入到网络访问响应中返回给用户终端。

[0066] 针对图3所示的实施例，第二获取模块12可以根据特征信息通过分裂法、层次法、基于密度的方法、基于网格的方法和基于模型的方法对网站进行聚类分析。

[0067] 本实施例提供的网络服务器中各模块的功能和处理流程，可以参见上述图1所示的方法实施例，其实现原理和技术效果类似，此处不再赘述。

[0068] 基于图3所示的实施例，进一步地，处理模块14还用于若判断获知网站没有包括与网址对应的第一网站，则向其余网络服务器广播包括网址和多个时间段的上网信息查询请求，若接收到其余网络服务器返回的第一网站在多个时间段内的上网信息，则根据上网信息获取第一网站对应的特征信息；根据每个网站簇中网站对应的特征信息获取对应的聚集轮廓信息，并根据第一网站对应的特征信息和聚集轮廓信息通过相似性度量确定第一网站所属的网站簇。

[0069] 本实施例提供的网络服务器中各模块的功能和处理流程，可以参见上述图2所示的方法实施例，其实现原理和技术效果类似，此处不再赘述。

[0070] 图4为本发明网站推荐系统一个实施例的结构示意图，如图4所示，该系统包括：网络服务器1以及用户终端2，其中，网络服务器1可以为本发明实施例提供的网络服务器，用户终端2为本发明实施例涉及到的用户终端，本实施例提供的网站推荐系统中各装置的功能和处理流程，可以参见上述方法和装置实施例，其实现原理和技术效果类似，此处不再赘述。

[0071] 最后应说明的是：以上实施例仅用以说明本发明的技术方案，而非对其限制；尽管参照前述实施例对本发明进行了详细的说明，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发明各实施例技术方案的精神和范围。

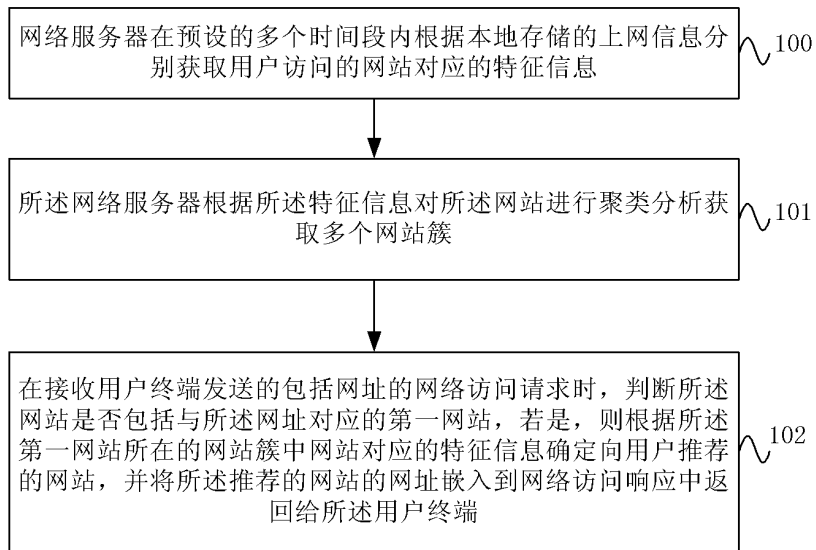


图 1

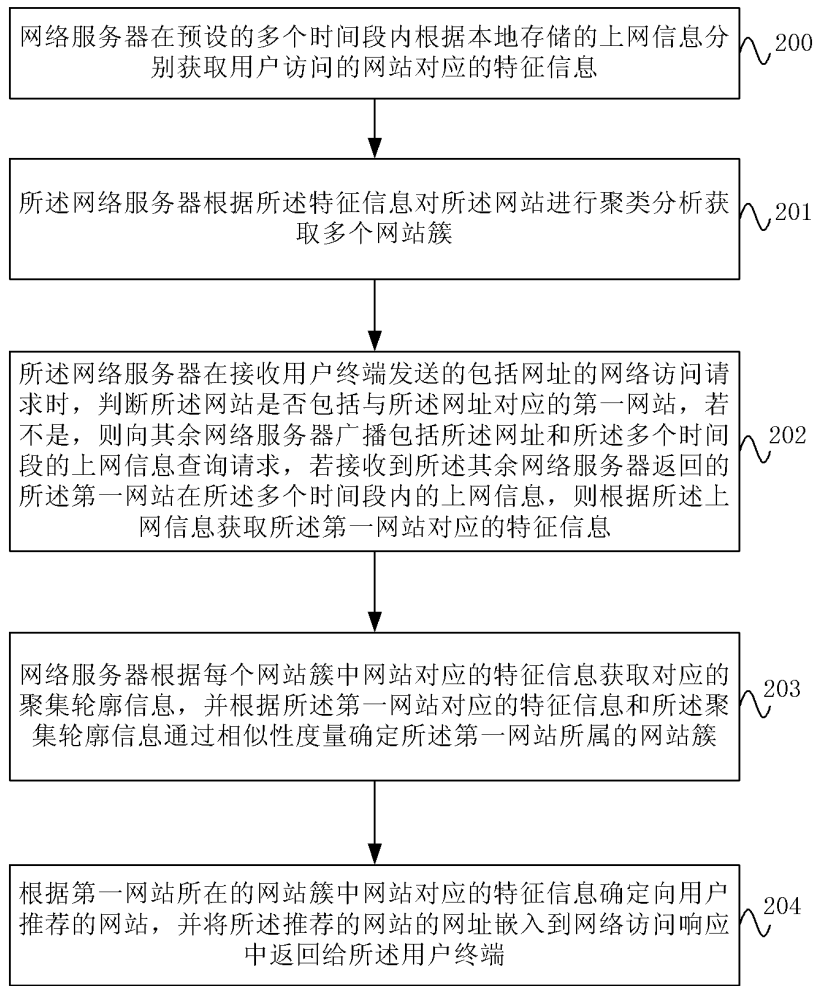


图 2

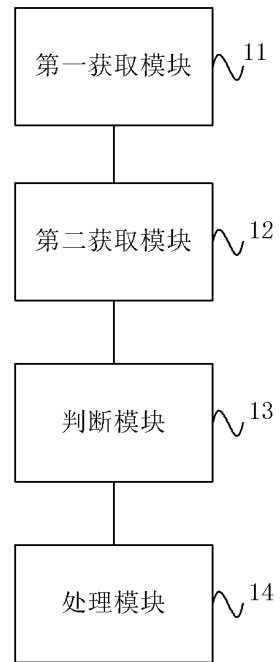


图 3

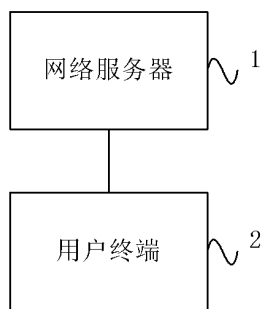


图 4