

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2016 年 1 月 7 日 (07.01.2016)



(10) 国际公布号
WO 2016/000555 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2015/082282
- (22) 国际申请日: 2015 年 6 月 25 日 (25.06.2015)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201410307116.X 2014 年 6 月 30 日 (30.06.2014) CN
201410308039.X 2014 年 6 月 30 日 (30.06.2014) CN
- (71) 申请人: 北京奇虎科技有限公司 (BEIJING QIHOO TECHNOLOGY COMPANY LIMITED) [CN/CN]; 中国北京市西城区新街口外大街 28 号 D 座 112 室 (德胜园区), Beijing 100088 (CN)。奇智软件 (北京) 有限公司 (QIZHI SOFTWARE (BEIJING) COMPANY LIMITED) [CN/CN]; 中国北京市朝阳区酒仙桥路 6 号院 2 号楼 B 座 2 层、3 层 301-306 室, Beijing 100015 (CN)。
- (72) 发明人: 周楠 (ZHOU, Nan); 中国北京市朝阳区酒仙桥路 6 号院 2 号楼, Beijing 100015 (CN)。常富洋 (CHANG, Fuyang); 中国北京市朝阳区酒仙桥路 6

号院 2 号楼, Beijing 100015 (CN)。秦吉胜 (QIN, Jisheng); 中国北京市朝阳区酒仙桥路 6 号院 2 号楼, Beijing 100015 (CN)。

- (74) 代理人: 北京智汇东方知识产权代理事务所 (普通合伙) (WISEAST INTELLECTUAL PROPERTY LAW FIRM); 中国北京市海淀区花园路 13 号 5 幢 320 房间, Beijing 100088 (CN)。
- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE,

[见续页]

(54) Title: METHODS AND SYSTEMS FOR RECOMMENDING SOCIAL NETWORK-BASED CONTENT AND NEWS

(54) 发明名称: 基于社交网络的内容、新闻推荐方法和系统

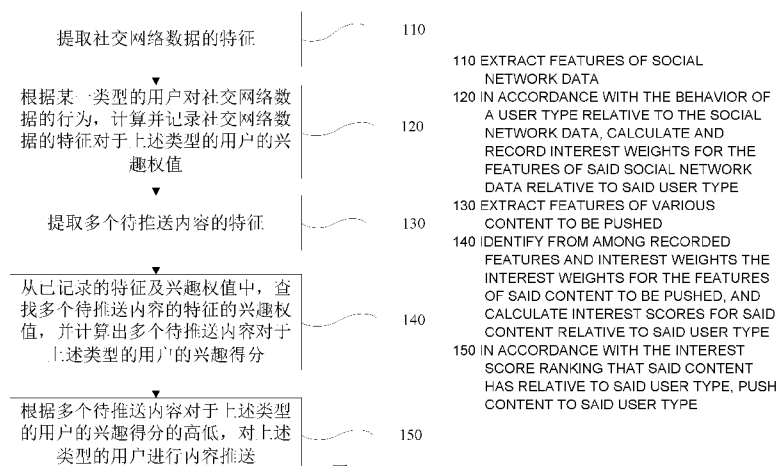


图 1 / Fig. 1

(57) Abstract: The present invention provides a method and a system for recommending social network-based content, and a method and a system for recommending news. The method for recommending social network-based content comprises: extracting features of social network data; in accordance with the behavior of a user type relative to the social network data, calculating and recording interest weights for the features of said social network data relative to said user type; extracting features of various content to be pushed; identifying from among recorded features and interest weights the interest weights for the features of said content to be pushed, and calculating interest scores for said content relative to said user type; in accordance with the interest score ranking that said content has relative to said user type, pushing content to said user type. In the present invention, the interests of various user types are analyzed and content matching user interests is pushed to the user.

(57) 摘要:

[见续页]



WO 2016/000555 A1



IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO,
RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD,
TG)。

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

本发明提供了一种基于社交网络的内容推荐方法和系统，以及一种新闻推荐方法和系统。其中，基于社交网络的内容推荐方法包括：提取社交网络数据的特征；根据某一类型的用户对所述社交网络数据的行为，计算并记录所述社交网络数据的特征对于所述类型的用户的兴趣权值；提取多个待推送内容的特征；从已记录的特征及兴趣权值中，查找所述多个待推送内容的特征的兴趣权值，并计算出所述多个待推送内容对于所述类型的用户的兴趣得分；根据所述多个待推送内容对于所述类型的用户的兴趣得分的高低，对所述类型的用户进行内容推送。通过本发明，可以分析不同类型用户的兴趣所在，并将符合用户兴趣的内容推送给用户。

基于社交网络的内容、新闻推荐方法和系统

技术领域

本发明涉及信息技术领域，特别是一种基于社交网络的内容推荐方法和系统，
5 以及一种新闻推荐方法和系统。

背景技术

获取新闻和资讯是现代社会中人们的生活习惯，随着计算机技术的发展和互联网用户规模的不断扩大，越来越多的人使用通过互联网获得各种各样所需的信息。
10 同时，通过互联网提供新闻和资讯服务的网站也越来越多，越来越多的突发新闻和事件是通过互联网得到迅速传播的，互联网信息呈现出爆发式的增长趋势。近些年来，移动互联网的快速发展使得用户的阅读时间变得越来越碎片化，在这种背景下，如何在海量的信息中筛选出最有价值的信息，向用户个性化推荐其最感兴趣的新闻和资讯，就变得极其重要。

15 现在的互联网新闻资讯阅读产品主要包括 web（网页）端和移动 app（应用程序）端，从新闻和资讯的整合方式来看，大部分依然是人工编辑和分类浏览的形式，这种方式的阅读会使用户浏览到大量不感兴趣的新闻和资讯，浪费用户的时间，同时产品本身也需要大量的编辑来进行新闻和资讯的更新和维护；以 google reader（谷歌阅读）为代表的订阅类新闻阅读产品是不同于上述产品的另外一种产品形式，
20 用户可以订阅自己感兴趣的网站的内容进行阅读和浏览，这种阅读形式减少了用户浏览到不感兴趣的内容的可能性，但是用户需要自己寻找自己感兴趣的内容和网站进行一系列设置，而对于大多数互联网用户来说，他们不喜欢这种繁琐的方式。

为了使得用户在最短的时间内以最便捷的方式获取最有价值和最感兴趣的新闻和资讯，必须采取一种更加智能的方式去提供给用户所需的信息，针对不同的用户
25 推荐用户最感兴趣和最有价值的新闻和资讯。

发明内容

鉴于上述问题，提出了本发明以便提供一种克服上述问题或者至少部分地解决上述问题的基于社交网络的内容推荐方法和系统，以及新闻推荐方法和系统。

30 根据本发明的一方面，提供了基于社交网络的内容推荐方法，其包括：提取社交网络数据的特征；根据某一类型的用户对所述社交网络数据的行为，计算并记录所述社交网络数据的特征对于所述类型的用户的兴趣权值；提取多个待推送内容的特征；从已记录的特征及兴趣权值中，查找所述多个待推送内容的特征的兴趣权值，并计算出所述多个待推送内容对于所述类型的用户的兴趣得分；根据所述多个待推
35 送内容对于所述类型的用户的兴趣得分的高低，对所述类型的用户进行内容推送。

根据本发明的另一方面，还提供了一种基于社交网络的内容推荐系统，其包括：第一特征提取模块，用于提取社交网络数据的特征；兴趣权值计算模块，用于根据

某一类型的用户对所述社交网络数据的行为，计算并记录所述社交网络数据的特征对于所述类型的用户的兴趣权值；第二特征提取模块，用于提取多个待推送内容的特征；兴趣得分计算模块，用于从已记录的特征及兴趣权值中，查找所述多个待推送内容的特征的兴趣权值，并计算出所述多个待推送内容对于所述类型的用户的兴趣得分；内容待推荐模块，用于根据所述多个待推送内容对于所述类型的用户的兴趣得分的高低，对所述类型的用户进行内容推送。

在本发明的基于社交网络的内容推荐方法和系统中，由于不同类型用户在网络上的社交行为，能够反映对该类型用户的兴趣所在，所以基于不同类型用户对社交网络数据的行为，分析得到社交网络数据特征对于不同类型用户的兴趣权值，以及计算出待推送内容对不同类型用户的兴趣得分，实际上是合理分出了不同类型用户对待推送内容的兴趣高低，按兴趣高低对不同类型的用户进行推荐。本发明的技术方案给用户展示推荐内容，大大减少了人工编辑的工作量，对用户而言，提升了推荐内容的可读性，减少了大量用户不喜欢的推荐内容，节约了用户的时间，推荐质量的提高也会带动更多的用户，提高了推荐内容的点击率，最终带来推送流量的稳步提升。

根据本发明的又一方面，提供了一种新闻推荐方法，其包括：提取搜索查询数据的特征；根据某一类型用户对所述搜索查询数据的行为，计算并记录所述搜索查询数据的特征对于所述类型用户的兴趣权值；提取多个待推送新闻的特征；从已记录的特征及兴趣权值中，查找所述多个待推送新闻的特征的兴趣权值，并计算出所述多个待推送新闻对于所述类型的用户的兴趣得分；根据所述多个待推送新闻对于所述类型的用户的兴趣得分的高低，对所述类型的用户进行新闻推送。

根据本发明的又一方面，提供了一种新闻推荐系统，其包括：第一特征提取模块，用于提取搜索查询数据的特征；兴趣权值计算模块，用于根据用户对所述搜索查询数据的行为，计算并记录所述搜索查询数据的特征对于所述类型的用户的兴趣权值；第二特征提取模块，用于提取多个待推送新闻的特征；兴趣得分计算模块，用于从已记录的特征及兴趣权值中，查找所述多个待推送新闻的特征的兴趣权值，并计算出所述多个待推送新闻对于所述类型的用户的兴趣得分；待推送新闻推荐模块，用于根据所述多个待推送新闻对于所述类型的用户的兴趣得分的高低，对所述类型的用户进行新闻推送。

在本发明的新闻推荐方法和系统中，由于不同类型用户对搜索查询数据的行为，能够反映对该类型用户的兴趣所在，所以基于不同类型用户对搜索查询数据行为，分析得到搜索查询数据特征对于不同类型用户的兴趣权值，以及计算出待推送新闻对不同类型用户的兴趣得分，实际上是合理分出了不同类型用户对待推送新闻的兴趣高低，按兴趣高低对不同类型的用户推荐新闻。本发明的技术方案给用户展示新闻，大大减少了人工编辑的工作量，对用户而言，提升了新闻的可读性，减少了大量用户不喜欢的新闻，节约了用户的时间，推荐质量的提高也会带动更多的用户，提高了每条新闻的点击率，最终带来新闻流量的稳步提升。

根据本发明的又一方面，提供了一种计算机程序，其包括计算机可读代码，当

所述计算机可读代码在计算设备上运行时，导致所述计算设备执行根据上文所述的基于社交网络的内容推荐方法和/或新闻推荐方法。

根据本发明的再一方面，提供了一种计算机可读介质，其中存储了上述的计算机程序。

- 5 上述说明仅是本发明技术方案的概述，为了能够更清楚了解本发明的技术手段，而可依照说明书的内容予以实施，并且为了让本发明的上述和其它目的、特征和优点能够更明显易懂，以下特举本发明的具体实施方式。

附图说明

- 10 通过阅读下文优选实施方式的详细描述，各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的，而并不认为是对本发明的限制。而且在整个附图中，用相同的参考符号表示相同的部件。在附图中：

15 图 1 示出了根据本发明的一个实施例的基于社交网络的内容推荐方法的流程图；

图 2 示出了根据本发明的另一个实施例的基于社交网络的内容推荐方法的流程图；

图 3 示出了根据本发明的一个实施例的基于社交网络的内容推荐方法的工作流程图；

20 图 4 示出了根据本发明的一个实施例的基于社交网络的内容推荐系统的框图；

图 5 示出了根据本发明的另一个实施例的基于社交网络的内容推荐系统的框图；

图 6 示出了根据本发明的一个实施例的新闻推荐方法的流程图；

图 7 示出了根据本发明的另一个实施例的新闻推荐方法的流程图；

25 图 8 示出了根据本发明的一个实施例的新闻推荐方法的工作流程图；

图 9 示出了根据本发明的一个实施例的新闻推荐系统的框图；

图 10 示出了根据本发明的另一个实施例的新闻推荐系统的框图；

图 11 示意性地示出了用于执行根据本发明的基于社交网络的内容推荐方法和/或新闻推荐方法的计算设备的框图；以及

30 图 12 示意性地示出了用于保持或者携带实现根据本发明的基于社交网络的内容推荐方法和/或新闻推荐方法的程序代码的存储单元。

具体实施方式

下面结合附图和具体的实施方式对本发明作进一步的描述。

- 35 如图 1 所示，本发明的一个实施例提供了一种基于社交网络的内容推荐方法，其包括：

步骤 110，提取社交网络数据的特征。本实施例中对于社交网络数据的类型不做限定，例如，可以是用户所使用的社交网站、社交工具，例如微博、博客等，例

如，可以是社交网站、社交工具的名称、分类、标签内容等。

步骤 120，根据某一类型的用户对社交网络数据的行为，计算并记录社交网络数据的特征对于上述类型的用户的兴趣权值。例如，用户在社交网络上频繁发出体育类消息，由此可见用户对体育类内容具有较高的兴趣。

5 步骤 130，提取多个待推送内容的特征。本实施例中的待推送内容包括但不限于新闻以及资讯，或其他形式的信息。

步骤 140，从已记录的特征及兴趣权值中，查找多个待推送内容的特征的兴趣权值，并计算出多个待推送内容对于上述类型的用户的兴趣得分。在本实施例的技术方案中，依据前述的社交网络数据的特征以及相应的兴趣权值可以建立用户的兴
10 趣模型，通过兴趣模型可以选择出需要推送给用户的候选推送内容。

步骤 150，根据多个待推送内容对于上述类型的用户的兴趣得分的高低，对上述类型的用户进行内容推送。本实施例中，基于兴趣得分对待推送内容进行排序，根据排序结果可以确定最终要推荐给用户的内容集合以及顺序。

在本实施例的技术方案中，基于兴趣得分的高低，也即不同类型用户对于待推
15 送内容的兴趣高低进行内容推送，大大减少了人工编辑的工作量，对用户而言，提升了推荐内容的可读性，减少了大量用户不喜欢的推荐内容，节约了用户的时间，推荐质量的提高也会带动更多的用户，提高了每条推荐内容的点击率，最终带来推送流量的稳步提升。

如图 2 所示，本发明的另一实施例还提供了一种基于社交网络的内容推荐方法，
20 其中，还包括：

步骤 160，根据上述类型的用户对多个待推送内容的点击行为，重新确定多个待推送内容的兴趣得分。

步骤 170，按重新确定的兴趣得分，计算多个待推送内容的特征的兴趣权值并进行记录。

25 在本实施例的技术方案中，如果用户点击阅读了推送内容，说明推送准确；但如用户对推送的某条内容点击了不感兴趣的按钮，表示用户对于该内容所对应的分类或主题等特征具有较低兴趣，此时根据用户的实际行为估算该内容的兴趣得分，并反向修正该内容中的特征的兴趣权值，以便于在以后使得计算的兴趣得分与用户的实际兴趣更符合。

30 本发明的另一实施例还提供了一种基于社交网络的内容推荐方法，其中，社交网络数据包括社交网络账号，社交网络数据的特征包括社交网络账号的类别和主题，上述类型的用户对社交网络数据的行为包括对相同类别或相同主题的社交网络账号的关注行为。

在本实施例的技术方案中，以目前流行的微博为例，如用户关注某一媒体账号
35 或者名人账号，则表示用户对该类型或主题的微博账号感兴趣，用户对同一标签下的微博账号关注的越多，则可设置较高的兴趣权值。目前微博账号均可设置对应的类别、主题或是其他形式的标签；也可以预先为不同微博账号制定至少一个标签，并在标签中记录微博账号的特征，微博账号的标签可以存储在数据库中，以在需要进行提取。

本发明的另一实施例还提供了一种基于社交网络的内容推荐方法，其中，社交网络数据包括社交网络账号发布的社交内容，社交网络数据的特征包括社交内容的类别和主题，上述类型的用户对社交网络数据的行为包括对相同类别或相同主题的社交内容的转发行为。

5 在本实施例的技术方案中，以目前流行的微博发出的正文为例，如用户对某一类别或主题的微博账号的正文的转发次数越多，则表示用户对该类别或主题的微博账号的正文具有较高的兴趣，则可以设置较高的兴趣权值。

10 本发明的另一实施例还提供了一种基于社交网络的内容推荐方法，其中，社交网络数据包括社交网络账号发布的 URL，社交网络数据的特征包括 URL 指向的推送内容的类别和主题，上述类型的用户对社交网络数据的行为包括对相同类别或相同主题的推送内容的 URL 的点击行为，或对相同类别或相同主题的推送内容上的页面标签的点击行为。

15 在本实施例的技术方案中，可以预先为不同推送内容设置类别标签，例如，如果推送内容为体育资讯，则其标签设置为体育标签。域名的类别标签可预先存储在数据库中。在本实施例的技术方案中，用户点击了某社交账号发布的 URL 指向的新闻，则表示用户对于该域名的类别和主题感兴趣，则可以设置较高的兴趣权值。

20 本发明的另一实施例还提供了一种基于社交网络的内容推荐方法，其中，社交网络数据包括社交网络账号发布的 URL，社交网络数据的特征包括 URL 中包含的域名的类别，上述类型的用户对社交网络数据的行为包括对相同类别的域名对应的 URL 的点击行为。

在本实施例的技术方案中，可以预先为不同域名设置类别标签，例如，一个域名的类别标签通常是这个域名下的网页所包含的网页的信息类别，比如 sports.abc.com，其下的网页可能包含了各个方面的体育信息，则可以把此域名的类别标签确定为“体育”。域名的类别标签可预先存储在数据库中。

25 在本实施例的技术方案中，用户点击了某社交账号发布的 URL 指向的新闻，则表示用户对于该域名的类别和主题感兴趣，则可以设置较高的兴趣权值。

本发明的另一实施例还提供了一种基于社交网络的内容推荐方法，其中，第 i 个待推送内容的兴趣得分为：

$$P = \frac{a}{b + e^{-g(V_i)}}$$

30 其中， $V_i = x_1 \times w_1 + x_2 \times w_2 + \dots + x_N \times w_N$ ，其中， $w_1 \dots w_N$ 为第 i 个待推送内容的 N 个特征， $x_1 \dots x_N$ 为对应 N 个特征的兴趣权值，a 为第一常数，b 为第二常数，e、g 均为固定常数。

35 在本实施例的技术方案中，基于上述的得分公式，可以实现一排序模型，该模型利用上述公式计算兴趣得分。排序模型实际上是一个逻辑回归分类器，该逻辑回归分类器的输入是一条推送内容的特征，输出是一条推送内容针对某一类型的用户的兴趣得分，得分越高表示该类型用户对这条推送内容可能越感兴趣。每条推送内容可以抽象为一个特征向量，向量的每个维度表示该条推送内容的主题、分类，甚

至关键词、热度等多个特征。

假设我们已经根据上述的兴趣权值得到模型系数向量为 $X = \{x_1, x_2, \dots, x_N\}$ ，则可将用来进行推送内容兴趣值计算的逻辑回归分类器表示为：

$$P(Y = 1 | news_i) = \frac{1}{1 + e^{-g(V)}}$$

- 5 其中， $V = XW$ ， X 表示上述类型的用户对应的模型系数向量， W 表示推送内容的特征向量，上述等式的左边的意义是当向用户推荐一条推送内容 $news_i$ 时，用户点击的可能性，所以计算得到的右边的兴趣得分可以作为对上述类型用户推送内容的依据。

- 10 结合前述的实施例，在用户对推送内容进行处理的情况下， W 已知， X 未知，求 X 。

根据用户的点击行为的反馈，可以得到用户点击过的推送内容集合和一批向用户推送过但是用户没有点击的内容集合，对于用户点击过的推送内容 $news_c$ ，可以得到：

$$P(Y = 1 | news_c) = \frac{1}{1 + e^{-g(V_c)}} = 1$$

对于用户没有点击过的推送内容 $news_d$ ，可以得到：

$$15 \quad P(Y = 1 | news_d) = \frac{1}{1 + e^{-g(V_d)}} = 0$$

这样根据一个用户对 m 条推送内容点击记录，我们就得到了 m 个形式如上所述两个表达式的式子，联立求解，即可得到该用户的排序模型系数向量 X ，也即修正了兴趣权值。

- 20 在兴趣权值修正之后，设模型系数向量为 $\{x_1, x_2, \dots, x_N\}$ ，将候选的推送内容集合中的每一条推送内容提取得到对应的特征向量 $W_i = \{w_1, w_2, \dots, w_N\}$ ，带入到模型中：

$$P(Y = 1 | news_i) = \frac{1}{1 + e^{-g(V_i)}}$$

- 25 其中， $V_i = x_1 \times w_1 + x_2 \times w_2 + \dots + x_N \times w_N$ ，计算则可得到 $P(Y = 1 | news_i)$ 。这个值就是该用户对此条内容的兴趣得分。根据候选推送内容兴趣得分的高低可以确定给该用户推荐内容的先后顺序，由此可见，本实施例的技术方案中根据用户对推送内容的实际点击行为，修正了兴趣权值，有利于再次更加准确地对用户进行内容推送，最终本实施例结合前述实施例得到的技术方案，其工作流程如图 3 所示。

需要说明的是，上述各个公式并不是实现本发明的唯一公式，仅作为实施例的一种实现方式。技术人员可以根据业务需要对公式做适当变形，依然落在本发明的范围之内，例如增添参数或倍数等。

如图 4 所示, 本发明的另一实施还提供了一种基于社交网络的内容推荐系统, 其包括:

第一特征提取模块 410, 用于提取社交网络数据的特征。本实施例中对于社交网络数据的类型不做限定, 例如, 可以是用户所使用的社交网站、社交工具, 例如 5 微博、博客等, 例如, 可以是社交网站、社交工具的名称、分类、标签内容等。

兴趣权值计算模块 420, 用于根据某一类型的用户对社交网络数据的行为, 计算并记录社交网络数据的特征对于上述类型的用户的兴趣权值。例如, 用户在社交网络上频繁发出体育类消息, 由此可见用户对体育类内容具有较高的兴趣。

第二特征提取模块 430, 用于提取多个待推送内容的特征。本实施例中的待推 10 送内容包括但不限于新闻以及资讯, 或其他形式的信息。

兴趣得分计算模块 440, 用于从已记录的特征及兴趣权值中, 查找多个待推送内容的特征的兴趣权值, 并计算出多个待推送内容对于上述类型的用户的兴趣得分。在本实施例的技术方案中, 依据前述的社交网络数据的特征以及相应的兴趣权值可以建立用户的兴趣模型, 通过兴趣模型可以选择出需要推送给用户的候选推送内容。

内容待推荐模块 450, 用于根据多个待推送内容对于上述类型的用户的兴趣得 15 分的高低, 对上述类型的用户进行内容推送。本实施例中, 基于兴趣得分对待推送内容进行排序, 根据排序结果可以确定最终要推荐给用户的内容集合以及顺序。

在本实施例的技术方案中, 基于兴趣得分的高低, 也即不同类型用户对于待推送内容的兴趣高低进行内容推送, 大大减少了人工编辑的工作量, 对用户而言, 提 20 升了推荐内容的可读性, 减少了大量用户不喜欢的推荐内容, 节约了用户的时间, 推荐质量的提高也会带动更多的用户, 提高了每条推荐内容的点击率, 最终带来推送流量的稳步提升。

如图 5 所示, 本发明的另一实施还提供了一种基于社交网络的内容推荐系统, 其中, 还包括:

第一重新确定模块 460, 用于根据上述类型的用户对多个待推送内容的点击行 25 为, 重新确定多个待推送内容的兴趣得分。

第二重新确定模块 470, 用于按重新确定的兴趣得分, 计算多个待推送内容的特征的兴趣权值并进行记录。

在本实施例的技术方案中, 如果用户点击阅读了推送内容, 说明推送准确; 但 30 如用户对推送的某条内容点击了不感兴趣的按钮, 表示用户对于该内容所对应的分类或主题等特征具有较低兴趣, 此时根据用户的实际行为估算该内容的兴趣得分, 并反向修正该内容中的特征的兴趣权值, 以便于在以后使得计算的兴趣得分与用户的实际兴趣更符合。

本发明的另一实施还提供了一种基于社交网络的内容推荐系统, 其中, 社交 35 网络数据包括社交网络账号, 社交网络数据的特征包括社交网络账号的类别和主题, 上述类型的用户对社交网络数据的行为包括对相同类别或相同主题的社交网络账号的关注行为。

在本实施例的技术方案中, 以目前流行的微博为例, 如用户关注某一媒体账号或者名人账号, 则表示用户对该类型或主题的微博账号感兴趣, 用户对同一标签下

的微博账号关注的越多，则可设置较高的兴趣权值。目前微博账号均可设置对应的类别、主题或是其他形式的标签；也可以预先为不同微博账号制定至少一个标签，并在标签中记录微博账号的特征，微博账号的标签可以存储在数据库中，以在需要时进行提取。

5 本发明的另一实施例还提供了一种基于社交网络的内容推荐系统，其中，社交网络数据包括社交网络账号发布的社交内容，社交网络数据的特征包括社交内容的类别和主题，上述类型的用户对社交网络数据的行为包括对相同类别或相同主题的社交内容的转发行为。

10 在本实施例的技术方案中，以目前流行的微博发出的正文为例，如用户对某一类别或主题的微博账号的正文的转发次数越多，则表示用户对该类别或主题的微博账号的正文具有较高的兴趣，则可以设置较高的兴趣权值。

15 本发明的另一实施例还提供了一种基于社交网络的内容推荐系统，其中，社交网络数据包括社交网络账号发布的 URL，社交网络数据的特征包括 URL 指向的推送内容的类别和主题，上述类型的用户对社交网络数据的行为包括对相同类别或相同主题的推送内容的 URL 的点击行为，或对相同类别或相同主题的推送内容上的页面标签的点击行为。

20 在本实施例的技术方案中，可以预先为不同推送内容设置类别标签，例如，如果推送内容为体育资讯，则其标签设置为体育标签。域名的类别标签可预先存储在数据库中。在本实施例的技术方案中，用户点击了某社交账号发布的 URL 指向的新闻，则表示用户对于该域名的类别和主题感兴趣，则可以设置较高的兴趣权值。

25 本发明的另一实施例还提供了一种基于社交网络的内容推荐系统，其中，社交网络数据包括社交网络账号发布的 URL，社交网络数据的特征包括 URL 中包含的域名的类别，上述类型的用户对社交网络数据的行为包括对相同类别的域名对应的 URL 的点击行为。

30 在本实施例的技术方案中，可以预先为不同域名设置类别标签，例如，一个域名的类别标签通常是这个域名下的网页所包含的网页的信息类别，比如 sports.abc.com，其下的网页可能包含了各个方面的体育信息，则可以把此域名的类别标签确定为“体育”。域名的类别标签可预先存储在数据库中。

35 在本实施例的技术方案中，用户点击了某社交账号发布的 URL 指向的新闻，则表示用户对于该域名的类别和主题感兴趣，则可以设置较高的兴趣权值。

本发明的另一实施例还提供了一种基于社交网络的内容推荐系统，其中，第 i 个待推送内容的兴趣得分为：

$$P = \frac{a}{b + e^{-g(V_i)}}$$

40 其中， $V_i = x_1 \times w_1 + x_2 \times w_2 + \dots + x_N \times w_N$ ，其中， $w_1, \dots, w_N$ 为第 i 个待推送内容的 N 个特征， $x_1, \dots, x_N$ 为对应 N 个特征的兴趣权值，a 为第一常数，b 为第二常数，e、g 均为固定常数。

在本实施例的技术方案中，基于上述的得分公式，可以实现一排序模型，该模

型利用上述公式计算兴趣得分。排序模型实际上是一个逻辑回归分类器，该逻辑回归分类器的输入是一条推送内容的特征，输出是一条推送内容针对某一类型的用户的兴趣得分，得分越高表示该类型用户对这条推送内容可能越感兴趣。每条推送内容可以抽象为一个特征向量，向量的每个维度表示该条推送内容的主题、分类，甚至关键词、热度等多个特征。

假设我们已经根据上述的兴趣权值得到模型系数向量为 $X = \{x_1, x_2, \dots, x_N\}$ ，则可将用来进行推送内容兴趣值计算的逻辑回归分类器表示为：

$$P(Y = 1 | news_i) = \frac{1}{1 + e^{-g(V)}}$$

其中， $V = XW$ ， X 表示上述类型的用户对应的模型系数向量， W 表示推送内容的特征向量，上述等式的左边的意义是当向用户推荐一条推送内容 $news_i$ 时，用户点击的可能性，所以计算得到的右边的兴趣得分可以作为对上述类型用户推送内容的依据。

结合前述的实施例，在用户对推送内容进行处理的情况下， W 已知， X 未知，求 X 。

根据用户的点击行为的反馈，可以得到用户点击过的推送内容集合和一批向用户推送过但是用户没有点击的内容集合，对于用户点击过的推送内容 $news_c$ ，可以得到：

$$P(Y = 1 | news_c) = \frac{1}{1 + e^{-g(V_c)}} = 1$$

对于用户没有点击过的推送内容 $news_d$ ，可以得到：

$$P(Y = 1 | news_d) = \frac{1}{1 + e^{-g(V_d)}} = 0$$

这样根据一个用户对 m 条推送内容点击记录，我们就得到了 m 个形式如上所述两个表达式的式子，联立求解，即可得到该用户的排序模型系数向量 X ，也即修正了兴趣权值。

在兴趣权值修正之后，设模型系数向量为 $\{x_1, x_2, \dots, x_N\}$ ，将候选的推送内容集合中的每一条推送内容提取得到对应的特征向量 $W_i = \{w_1, w_2, \dots, w_N\}$ ，带入到模型中：

$$P(Y = 1 | news_i) = \frac{1}{1 + e^{-g(V_i)}}$$

其中， $V_i = x_1 \times w_1 + x_2 \times w_2 + \dots + x_N \times w_N$ ，计算则可得到 $P(Y = 1 | news_i)$ 。

这个值就是该用户对此条内容的兴趣得分。根据候选推送内容兴趣得分的高低可以确定给该用户推荐内容的先后顺序，由此可见，本实施例的技术方案中根据用户对推送内容的实际点击行为，修正了兴趣权值，有利于再次更加准确地对用户进行内容推送，最终本实施例结合前述实施例得到的技术方案，其工作流程如图 3 所示。

根据本发明的实施例，下面将介绍本发明实施例提供的新闻推荐方法和系统。

如图 6 所示, 本发明的一个实施例提供了一种新闻推荐方法, 其包括:

步骤 610, 提取搜索查询数据的特征。本实施例中对于搜索查询数据的类型不做限定, 例如, 可以是用户对搜索得到的新闻的浏览情况等; 本实施例对搜索查询数据的特征也不进行限定, 例如, 可以是用户浏览的新闻的分类、标题、关键词、
5 新闻来源、网站来源、地域标签、点击率等。

步骤 620, 根据某一类型用户对搜索查询数据的行为, 计算并记录搜索查询数据的特征对于上述类型的用户的兴趣权值。例如, 对于浏览行为而言, 用户对首先浏览、重复浏览的新闻必然兴趣较高, 由此可分析用户的兴趣权值。

步骤 630, 提取多个待推送新闻的特征。

10 步骤 640, 从已记录的特征及兴趣权值中, 查找多个待推送新闻的特征的兴趣权值, 并计算出多个待推送新闻对于上述类型的用户的兴趣得分。在本实施例的技术方案中, 依据前述的搜索查询数据的特征以及相应的兴趣权值可以建立用户的兴趣模型, 通过兴趣模型可以选择出需要推送给用户的候选新闻。

步骤 650, 根据多个待推送新闻对于上述类型的用户的兴趣得分的高低, 对用
15 户进行推送。本实施例中, 基于兴趣得分对待推送新闻进行排序, 根据排序结果可以确定最终要推荐给用户的新闻集合以及顺序。

在本实施例的技术方案中, 基于兴趣得分的高低, 也即不同类型用户对于待推送新闻的兴趣高低进行新闻推送, 大大减少了人工编辑的工作量, 对用户而言, 提升了新闻的可读性, 减少了大量用户不喜欢的新闻, 节约了用户的时间, 推荐质量的
20 的提高也会带动更多的用户, 提高了每条新闻的点击率, 最终带来新闻流量的稳步提升。

如图 7 所示, 本发明的另一个实施例还提供了一种新闻推荐方法, 还包括:

步骤 660, 根据上述类型的用户对多个待推送新闻的点击行为, 重新确定多个待推送新闻的兴趣得分。

25 步骤 670, 按重新确定的兴趣得分, 计算多个待推送新闻的特征的兴趣权值并进行记录。

在本实施例的技术方案中, 用户点击并阅读了推送新闻的话, 则说明推送准确; 但如用户对推送的某条新闻点击了不感兴趣的按钮, 表示用户对于该新闻所对应的分类或主题等特征具有较低兴趣, 此时根据用户的实际行为估算该新闻的兴趣得分, 并反向修正该新闻的特征的兴趣权值, 以便于在以后使得计算的兴趣得分与用户的
30 实际兴趣更符合。

本发明的另一个实施例还提供了一种新闻推荐方法, 其中, 搜索查询数据包括查询词, 搜索查询数据的特征包括查询词的类别和主题, 上述类型的用户对搜索查询数据的行为包括对相同类别或相同主题的查询词的查询行为。

35 在本实施例的技术方案中, 可以预先根据查询词对应的新闻集合中新闻的类别标签和主题标签来确定该查询词的类别标签和主题标签, 并建立数据库进行存储, 则查询词的类别和主题可以从数据库中的类别标签和主题标签中进行提取。比如搜索查询词 abc, 获取到新闻中最多的主题标签是 t1, 则该查询词对应的主题标签是 t1, 获取到新闻最多的分类标签是 c1, 则该查询词对应的类别标签是 c1, 则可以提取 t1

和 $c1$ 作为查询词的类别和主题的特征。

本实施例的技术方案中，用户对查询词查询行为的不同主要包括：搜索频次的不同以及搜索时间的不同。搜索某一查询词的频次越高，说明用户的兴趣越高，则可为查询词的类别和主题设置较高的兴趣权值；同时，用户每次搜索该查询词的时间与当前时间越接近，也说明用户的兴趣越高，则可为查询词的类别和主题设置较高的兴趣权值。

本发明的另一个实施例还提供了一种新闻推荐方法，其中，搜索查询数据包括查询结果页上的 URL，搜索查询数据的特征包括 URL 指向的新闻的类别和主题，上述类型的用户对搜索查询数据的行为包括对相同类别或相同主题的新闻的 URL 的点击行为，或对相同类别或相同主题的新闻上的页面标签的点击行为。

在本实施例的技术方案中，可以预先为每个新闻设置一个类别标签和至少一个主题标签，并在其中记载该新闻的一个类别和至少一个主题。

在本实施例的技术方案中，用户点击阅读了搜索到的某条 URL 指向的新闻，则表示用户对于该新闻的类别和主题感兴趣，则可以设置较高的兴趣权值；或者，用户点击了某个 URL 指向的新闻分类频道，且该分类频道的新闻具有同一类别标签，则表示用户对该新闻的类别感兴趣，则可以设置较高的兴趣权值。

本发明的另一个实施例还提供一种新闻推荐方法，其中，搜索查询数据包括社交网络账号发布的 URL，搜索查询数据的特征包括 URL 中包含的域名的类别，上述类型的用户对搜索查询数据的行为包括对相同类别的域名对应的 URL 的点击行为。

在本实施例的技术方案中，可以预先为不同域名设置类别标签，例如，一个域名的类别标签通常是这个域名下的网页所包含的网页的信息类别，比如 sports.abc.com，其下的网页可能包含了各个方面的体育信息，则可以把此域名的类别标签确定为“体育”。域名的类别标签可预先存储在数据库中。

在本实施例的技术方案中，用户搜索到某社交账号发布的 URL，并点击阅读该 URL 指向的新闻，则表示用户对于该域名的类别和主题感兴趣，则可以设置较高的兴趣权值。

本发明的另一个实施例还提供了一种新闻推荐方法，其中，第 i 个待推送新闻的兴趣得分为：

$$P = \frac{a}{b + e^{-g(V_i)}}$$

其中， $V_i = x_1 \times w_1 + x_2 \times w_2 + \dots + x_N \times w_N$ ，其中， $w_1, \dots, w_N$ 为第 i 个待推送新闻的 N 个特征， $x_1, \dots, x_N$ 为对应 N 个特征的兴趣权值， a 为第一常数， b 为第二常数， e 、 g 均为固定常数。

在本实施例的技术方案中，基于上述的得分公式，可以实现一排序模型，该模型利用上述公式计算兴趣得分。排序模型实际上是一个逻辑回归分类器，该逻辑回归分类器的输入是一条新闻的特征，输出是一条新闻针对某一类型的用户的兴趣得分，得分越高表示该类型用户对这条新闻可能越感兴趣。每条新闻可以抽象为一个

特征向量，向量的每个维度表示该条新闻的主题、分类，甚至关键词、热度等多个特征。

假设我们已经根据上述的兴趣权值得到模型系数向量为 $\mathbf{X} = \{x_1, x_2, \dots, x_N\}$ ，则可将用来进行新闻兴趣值计算的逻辑回归分类器表示为：

$$5 \quad P(Y = 1 | news_i) = \frac{1}{1 + e^{-g(V)}}$$

其中， $V = \mathbf{XW}$ ， $\mathbf{X}$ 表示上述类型的用户对应的模型系数向量， $\mathbf{W}$ 表示新闻的特征向量，上述等式的左边的意义是当向用户推荐一条新闻 $news_i$ 时，用户点击的可能性，所以计算得到的右边的兴趣得分可以作为对上述类型用户推送新闻的依据。

10 结合前述的实施例，在用户对推送新闻进行处理的情况下， $\mathbf{W}$ 已知， $\mathbf{X}$ 未知，求 $\mathbf{X}$ 。

根据用户的点击行为的反馈，可以得到用户点击过的新闻集合和一批向用户推送过但是用户没有点击的新闻集合，对于用户点击过的新闻 $news_c$ ，可以得到：

$$P(Y = 1 | news_c) = \frac{1}{1 + e^{-g(V_c)}} = 1$$

对于用户没有点击过的新闻 $news_d$ ，可以得到：

$$15 \quad P(Y = 1 | news_d) = \frac{1}{1 + e^{-g(V_d)}} = 0$$

这样根据一个用户对 m 条推送新闻点击记录，我们就得到了 m 个形式如上所述两个表达式的式子，联立求解，即可得到该用户的排序模型系数向量 $\mathbf{X}$ ，也即修正了兴趣权值。

20 在兴趣权值修正之后，设模型系数向量为 $\{x_1, x_2, \dots, x_N\}$ ，将候选的新闻集合中的每一条新闻提取得到对应的特征向量 $\mathbf{W}_i = \{w_1, w_2, \dots, w_N\}$ ，带入到模型中：

$$P(Y = 1 | news_i) = \frac{1}{1 + e^{-g(V_i)}}$$

25 其中， $V_i = \mathbf{x}_1 \times \mathbf{w}_1 + \mathbf{x}_2 \times \mathbf{w}_2 + \dots + \mathbf{x}_N \times \mathbf{w}_N$ ，计算则可得到 $P(Y = 1 | news_i)$ 。这个值就是该用户对此条新闻的兴趣得分。根据候选新闻兴趣得分的高低可以确定给该用户推荐新闻的先后顺序，由此可见，本实施例的技术方案中根据用户对推送新闻的实际点击行为，修正了兴趣权值，有利于再次更加准确地对用户进行新闻推送，最终本实施例结合前述实施例得到的技术方案，其工作流程如图 8 所示。

需要说明的是，上述各个公式并不是实现本发明的唯一公式，仅作为实施例的一种实现方式。技术人员可以根据业务需要对公式做适当变形，依然落在本发明的

范围之内，例如增添参数或倍数值等。

如图 9 所示，本发明的另一实施例还提供了一种新闻推荐系统，其包括：

第一特征提取模块 910，用于提取搜索查询数据的特征。本实施例中对于搜索查询数据的类型不做限定，例如，可以是用户对搜索得到的新闻的浏览情况等；本
5 实施例对搜索查询数据的特征也不进行限定，例如，可以是用户浏览的新闻的分类、标题、关键词、新闻来源、网站来源、地域标签、点击率等。

兴趣权值计算模块 920，用于根据用户对搜索查询数据的行为，计算并记录搜索查询数据的特征对于上述类型的用户的兴趣权值。例如，对于浏览行为而言，用户对首先浏览、重复浏览的新闻必然兴趣较高，由此可分析用户的兴趣权值。

10 第二特征提取模块 930，用于提取多个待推送新闻的特征。

兴趣得分计算模块 940，用于从已记录的特征及兴趣权值中，查找多个待推送新闻的特征的兴趣权值，并计算出多个待推送新闻对于上述类型的用户的兴趣得分。在本实施例的技术方案中，依据前述的搜索查询数据的特征以及相应的兴趣权值可以建立用户的兴趣模型，通过兴趣模型可以选择出需要推送给用户的候选新闻。

15 待推送新闻推荐模块 950，用于根据多个待推送新闻对于上述类型的用户的兴趣得分的高低，按顺序将多个待推送新闻推送给上述类型的用户。本实施例中，基于兴趣得分对待推送新闻进行排序，根据排序结果可以确定最终要推荐给用户的新闻集合以及顺序。

在本实施例的技术方案中，基于兴趣得分的高低，也即不同类型用户对于待推送新闻的兴趣高低进行新闻推送，大大减少了人工编辑的工作量，对用户而言，提升了新闻的可读性，减少了大量用户不喜欢的新闻，节约了用户的时间，推荐质量的提高也会带动更多的用户，提高了每条新闻的点击率，最终带来新闻流量的稳步提升。

如图 10 所示，本发明的另一个实施例还提供了一种新闻推荐系统，还包括：

25 第一重新确定模块 960，用于根据上述类型的用户对多个待推送新闻的点击行为，重新确定多个待推送新闻的兴趣得分。

第二重新确定模块 970，用于按重新确定的兴趣得分，计算多个待推送新闻的特征的兴趣权值并进行记录。

本实施例的技术方案中，用户点击并阅读了推送新闻的话，则说明推送准确；
30 但如用户对推送的某条新闻点击了不感兴趣的按钮，表示用户对于该新闻所对应的分类或主题等特征具有较低兴趣，此时根据用户的实际行为估算该新闻的兴趣得分，并反向修正该新闻的特征的兴趣权值，以便于在以后使得计算的兴趣得分与用户的实际兴趣更符合。

本发明的另一个实施例还提供了一种新闻推荐系统，其中，搜索查询数据包括
35 查询词，搜索查询数据的特征包括查询词的类别和主题，上述类型的用户对搜索查询数据的行为包括对相同类别或相同主题的查询词的查询行为。

在本实施例的技术方案中，可以预先根据查询词对应的新闻集合中新闻的类别标签和主题标签来确定该查询词的类别标签和主题标签，并建立数据库进行存储，则查询词的类别和主题可以从数据库中的类别标签和主题标签中进行提取。比如搜

索查询词 abc, 获取到新闻中最多的主题标签是 t1, 则该查询词对应的主题标签是 t1, 获取到新闻最多的分类标签是 c1, 则该查询词对应的类别标签是 c1, 则可以提取 t1 和 c1 作为查询词的类别和主题的特征。

5 本实施例的技术方案中, 用户对查询词查询行为的不同主要包括: 搜索频次的不同以及搜索时间的不同。搜索某一查询词的频次越高, 说明用户的兴趣越高, 则可为查询词的类别和主题设置较高的兴趣权值; 同时, 用户每次搜索该查询词的时间与当前时间越接近, 也说明用户的兴趣越高, 则可为查询词的类别和主题设置较高的兴趣权值。

10 本发明的另一个实施例还提供了一种新闻推荐系统, 其中, 搜索查询数据包括查询结果页上的 URL, 搜索查询数据的特征包括 URL 指向的新闻的类别, 上述类型的用户对搜索查询数据的行为包括对相同类别的新闻的 URL 的点击行为, 或对相同类别或相同主题的新闻上的页面标签的点击行为。

在本实施例的技术方案中, 可以预先为每个新闻设置一个类别标签和至少一个主题标签, 并在其中记载该新闻的一个类别和至少一个主题。

15 在本实施例的技术方案中, 用户点击阅读了搜索到的某条 URL 指向的新闻, 则表示用户对于该新闻的类别和主题感兴趣, 则可以设置较高的兴趣权值; 或者, 用户点击了某个 URL 指向的新闻分类频道, 且该分类频道的新闻具有同一类别标签, 则表示用户对该新闻的类别感兴趣, 则可以设置较高的兴趣权值。

20 本发明的另一个实施例还提供了一种新闻推荐系统, 其中, 搜索查询数据包括社交网络账号发布的 URL, 搜索查询数据的特征包括 URL 中包含的域名的类别和主题, 上述类型的用户对搜索查询数据的行为包括对相同类别或相同主题的域名对应的 URL 的点击行为。

25 在本实施例的技术方案中, 可以预先为不同域名设置类别标签, 例如, 一个域名的类别标签通常是这个域名下的网页所包含的网页的信息类别, 比如 sports.abc.com, 其下的网页可能包含了各个方面的体育信息, 则可以把此域名的类别标签确定为“体育”。域名的类别标签可预先存储在数据库中。

在本实施例的技术方案中, 用户搜索到某社交账号发布的 URL, 并点击阅读该 URL 指向的新闻, 则表示用户对于该域名的类别和主题感兴趣, 则可以设置较高的兴趣权值。

30 本发明的另一个实施例还提供了一种新闻推荐系统, 其中, 第 i 个待推送新闻的兴趣得分为:

$$P = \frac{a}{b + e^{-g(V_i)}}$$

其中, $V_i = x_1 \times w_1 + x_2 \times w_2 + \dots + x_N \times w_N$, 其中, $w_1, \dots, w_N$ 为第 i 个待推送新闻的 N 个特征, $x_1, \dots, x_N$ 为对应 N 个特征的兴趣权值, a 为第一常数, b 为第二常数,

35 e、g 均为固定常数。

在本实施例的技术方案中, 基于上述的得分公式, 可以实现一排序模型, 该模型利

用上述公式计算兴趣得分。排序模型实际上是一个逻辑回归分类器，该逻辑回归分类器的输入是一条新闻的特征，输出是一条新闻针对某一类型的用户的兴趣得分，得分越高表示该类型用户对这条新闻可能越感兴趣。每条新闻可以抽象为一个特征向量，向量的每个维度表示该条新闻的主题、分类，甚至关键词、热度等多个特征。

- 5 假设我们已经根据上述的兴趣权值得到模型系数向量为 $X = \{x_1, x_2, \dots, x_N\}$ ，则可将用来进行新闻兴趣值计算的逻辑回归分类器表示为：

$$P(Y = 1 | news_i) = \frac{1}{1 + e^{-g(V)}}$$

- 其中， $V = XW$ ， X 表示上述类型的用户对应的模型系数向量， W 表示新闻的特征向量，上述等式的左边的意义是当向用户推荐一条新闻 $news_i$ 时，用户点击的可能性，所以计算得到的右边的兴趣得分可以作为对上述类型用户推送新闻的依据。

结合前述的实施例，在用户对推送新闻进行处理的情况下， W 已知， X 未知，求 X 。

根据用户的点击行为的反馈，可以得到用户点击过的新闻集合和一批向用户推送过但是用户没有点击的新闻集合，对于用户点击过的新闻 $news_c$ ，可以得到：

$$P(Y = 1 | news_c) = \frac{1}{1 + e^{-g(V_c)}} = 1$$

- 15 对于用户没有点击过的新闻 $news_d$ ，可以得到：

$$P(Y = 1 | news_d) = \frac{1}{1 + e^{-g(V_d)}} = 0$$

这样根据一个用户对 m 条推送新闻点击记录，我们就得到了 m 个形式如上所述两个表达式的式子，联立求解，即可得到该用户的排序模型系数向量 X ，也即修正了兴趣权值。

- 20 在兴趣权值修正之后，设模型系数向量为 $\{x_1, x_2, \dots, x_N\}$ ，将候选的新闻集合中的每一条新闻提取得到对应的特征向量 $W_i = \{w_1, w_2, \dots, w_N\}$ ，带入到模型中：

$$P(Y = 1 | news_i) = \frac{1}{1 + e^{-g(V_i)}}$$

其中， $V_i = x_1 \times w_1 + x_2 \times w_2 + \dots + x_N \times w_N$ ，计算则可得到 $P(Y = 1 | news_i)$ 。

- 25 这个值就是该用户对此条新闻的兴趣得分。根据候选新闻兴趣得分的高低可以确定给该用户推荐新闻的先后顺序，由此可见，本实施例的技术方案中根据用户对推送新闻的实际点击行为，修正了兴趣权值，有利于再次更加准确地对用户进行新闻推送，最终本实施例结合前述实施例得到的技术方案，其工作流程如图 8 所示。

在此处所提供的说明书中，说明了大量具体细节。然而，能够理解，本发明的实施例可以在没有这些具体细节的情况下实践。在一些实例中，并未详细示出公知

的方法、结构和技术，以便不模糊对本说明书的理解。

类似地，应当理解，为了精简本公开并帮助理解各个发明方面中的一个或多个，在上面对本发明的示例性实施例的描述中，本发明的各个特征有时被一起分组到单个实施例、图、或者对其的描述中。然而，并不应将该公开的方法解释成反映如下意图：即所要求保护的本发明要求比在每个权利要求中所明确记载的特征更多的特征。更确切地说，如下面的权利要求书所反映的那样，发明方面在于少于前面公开的单个实施例的所有特征。因此，遵循具体实施方式的权利要求书由此明确地并入该具体实施方式，其中每个权利要求本身都作为本发明的单独实施例。

本领域那些技术人员可以理解，可以对实施例中的设备中的模块进行自适应性地改变并且把它们设置在与该实施例不同的一个或多个设备中。可以把实施例中的模块或单元或组件组合成一个模块或单元或组件，以及此外可以把它分成多个子模块或子单元或子组件。除了这样的特征和/或过程或者单元中的至少一些是相互排斥之外，可以采用任何组合对本说明书（包括伴随的权利要求、摘要和附图）中公开的所有特征以及如此公开的任何方法或者设备的所有过程或单元进行组合。除非另外明确陈述，本说明书（包括伴随的权利要求、摘要和附图）中公开的每个特征可以由提供相同、等同或相似目的的替代特征来代替。

此外，本领域的技术人员能够理解，尽管在此所述的一些实施例包括其它实施例中有所包括的某些特征而不是其它特征，但是不同实施例的特征的组合意味着处于本发明的范围之内并且形成不同的实施例。例如，在下面的权利要求书中，所要求保护的实施例的任意之一都可以以任意的组合方式来使用。

本发明的各个部件实施例可以以硬件实现，或者以在一个或者多个处理器上运行的软件模块实现，或者以它们的组合实现。本领域的技术人员应当理解，可以在实践中使用微处理器或者数字信号处理器（DSP）来实现根据本发明实施例的基于社交网络的内容推荐系统，以及新闻推荐系统中的一些或者全部部件的一些或者全部功能。本发明还可以实现为用于执行这里所描述的方法的一部分或者全部的设备或者装置程序（例如，计算机程序和计算机程序产品）。这样的实现本发明的程序可以存储在计算机可读介质上，或者可以具有一个或者多个信号的形式。这样的信号可以从因特网网站上下载得到，或者在载体信号上提供，或者以任何其他形式提供。

例如，图 11 示出了可以实现在智能终端之间传输数据的方法的计算设备。该计算设备传统上包括处理器 1110 和以存储器 1120 形式的计算机程序产品或者计算机可读介质。存储器 1120 可以是诸如闪存、EEPROM（电可擦除可编程只读存储器）、EPROM、硬盘或者 ROM 之类的电子存储器。存储器 1120 具有用于执行上述方法中的任何方法步骤的程序代码 1131 的存储空间 1130。例如，用于程序代码的存储空间 1130 可以包括分别用于实现上面的方法中的各种步骤的各个程序代码 1131。这些程序代码可以从一个或者多个计算机程序产品中读出或者写入到这一个或者多个计算机程序产品中。这些计算机程序产品包括诸如硬盘，紧致盘（CD）、存储卡或者软盘之类的程序代码载体。这样的计算机程序产品通常为如参考图 12 所述的便携式或者固定存储单元。该存储单元可以具有与图 11 的计算设备中的存储器 1120 类似布置的存储段、存储空间等。程序代码可以例如以适当形式进行压缩。通常，存储单

元包括计算机可读代码 1131'，即可以由例如诸如 1110 之类的处理器读取的代码，这些代码当由计算设备运行时，导致该计算设备执行上面所描述的方法中的各个步骤。

5 本文中所称的“一个实施例”、“实施例”或者“一个或者多个实施例”意味着，结合实施例描述的特定特征、结构或者特性包括在本发明的至少一个实施例中。此外，请注意，这里“在一个实施例中”的词语例子不一定全指同一个实施例。

10 应该注意的是上述实施例对本发明进行说明而不是对本发明进行限制，并且本领域技术人员在不脱离所附权利要求的范围的情况下可设计出替换实施例。在权利要求中，不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包含”不排除存在未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本发明可以借助于包括有若干不同元件的硬件以及借助于适当编程的计算机来实现。在列举了若干装置的单元权利要求中，这些装置中的若干个可以是通过同一个硬件项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

15 此外，还应当注意，本说明书中使用的语言主要是为了可读性和教导的目的而选择的，而不是为了解释或者限定本发明的主题而选择的。因此，在不偏离所附权利要求书的范围和精神的情况下，对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。对于本发明的范围，对本发明所做的公开是说明性的，而非限制性的，本发明的范围由所附权利要求书限定。

权 利 要 求

1、一种基于社交网络的内容推荐方法，其包括：

提取社交网络数据的特征；

5 根据某一类型的用户对所述社交网络数据的行为，计算并记录所述社交网络数据的特征对于所述类型的用户的兴趣权值；

提取多个待推送内容的特征；

从已记录的特征及兴趣权值中，查找所述多个待推送内容的特征的兴趣权值，并计算出所述多个待推送内容对于所述类型的用户的兴趣得分；

10 根据所述多个待推送内容对于所述类型的用户的兴趣得分的高低，对所述类型的用户进行内容推送。

2、根据权利要求1所述的基于社交网络的内容推荐方法，其中，还包括：

根据所述类型的用户对所述多个待推送内容的点击行为，重新确定所述多个待推送内容的兴趣得分；

15 按重新确定的兴趣得分，计算所述多个待推送内容的特征的兴趣权值并进行记录。

3、根据权利要求1-2任一项所述的基于社交网络的内容推荐方法，其中，所述社交网络数据包括社交网络账号，所述社交网络数据的特征包括所述社交网络账号的类别和主题，所述类型的用户对所述社交网络数据的行为包括对相同类别或相同主题

20 主题的所述社交网络账号的关注行为。

4、根据权利要求1-3任一项所述的基于社交网络的内容推荐方法，其中，所述社交网络数据包括社交网络账号发布的社交内容，所述社交网络数据的特征包括所述社交内容的类别和主题，所述类型的用户对所述社交网络数据的行为包括对相同类别或相同主题的所述社交内容的转发行为。

25 5、根据权利要求1-4任一项所述的基于社交网络的内容推荐方法，其中，所述社交网络数据包括社交网络账号发布的 URL，所述社交网络数据的特征包括所述 URL 指向的推送内容的类别和主题，所述类型的用户对所述社交网络数据的行为包括对相同类别或相同主题的推送内容的 URL 的点击行为，或对相同类别或相同主题的推送内容上的页面标签的点击行为。

30 6、根据权利要求1-5任一项所述的基于社交网络的内容推荐方法，其中，所述社交网络数据包括社交网络账号发布的 URL，所述社交网络数据的特征包括所述 URL 中包含的域名的类别，所述类型的用户对所述社交网络数据的行为包括对相同类别的域名对应的 URL 的点击行为。

7、一种新闻推荐方法，其包括：

35 提取搜索查询数据的特征；

根据某一类型用户对所述搜索查询数据的行为，计算并记录所述搜索查询数据的特征对于所述类型用户的兴趣权值；

提取多个待推送新闻的特征；

从已记录的特征及兴趣权值中，查找所述多个待推送新闻的特征的兴趣权值，并计算出所述多个待推送新闻对于所述类型的用户的兴趣得分；

根据所述多个待推送新闻对于所述类型的用户的兴趣得分的高低，对所述类型的用户进行新闻推送。

5 8、根据权利要求 7 所述的新闻推荐方法，其中，还包括：

根据所述类型的用户对所述多个待推送新闻的点击行为，重新确定所述多个待推送新闻的兴趣得分；

按重新确定的兴趣得分，计算所述多个待推送新闻的特征的兴趣权值并进行记录。

10 9、根据权利要求 7-8 任一项所述的新闻推荐方法，其中，所述搜索查询数据包括查询词，所述搜索查询数据的特征包括所述查询词的类别和主题，所述类型的用户对所述搜索查询数据的行为包括对相同类别或相同主题的所述查询词的查询行为。

15 10、根据权利要求 7-9 任一项所述的新闻推荐方法，其中，所述搜索查询数据包括查询结果页上的 URL，所述搜索查询数据的特征包括所述 URL 指向的新闻的类别和主题，所述类型的用户对所述搜索查询数据的行为包括对相同类别或相同主题的新闻的 URL 的点击行为，或对相同类别或相同主题的新闻上的页面标签的点击行为。

20 11、根据权利要求 7-10 任一项所述的新闻推荐方法，其中，所述搜索查询数据包括社交网络账号发布的 URL，所述搜索查询数据的特征包括所述 URL 中包含的域名的类别，所述类型的用户对所述搜索查询数据的行为包括对相同类别的域名对应的 URL 的点击行为。

12、一种基于社交网络的内容推荐系统，其包括：

第一特征提取模块，用于提取社交网络数据的特征；

25 兴趣权值计算模块，用于根据某一类型的用户对所述社交网络数据的行为，计算并记录所述社交网络数据的特征对于所述类型的用户的兴趣权值；

第二特征提取模块，用于提取多个待推送内容的特征；

30 兴趣得分计算模块，用于从已记录的特征及兴趣权值中，查找所述多个待推送内容的特征的兴趣权值，并计算出所述多个待推送内容对于所述类型的用户的兴趣得分；

内容待推荐模块，用于根据所述多个待推送内容对于所述类型的用户的兴趣得分的高低，对所述类型的用户进行内容推送。

13、根据权利要求 12 所述的基于社交网络的内容推荐系统，其中，还包括：

35 第一重新确定模块，用于根据所述类型的用户对所述多个待推送内容的点击行为，重新确定所述多个待推送内容的兴趣得分；

第二重新确定模块，用于按重新确定的兴趣得分，计算所述多个待推送内容的特征的兴趣权值并进行记录。

14、根据权利要求 12-13 任一项所述的基于社交网络的内容推荐系统，其中，所述社交网络数据包括社交网络账号，所述社交网络数据的特征包括所述社交网络

账号的类别和主题，所述类型的用户对所述社交网络数据的行为包括对相同类别或相同主题的所述社交网络账号的关注行为。

15、根据权利要求 12-14 任一项所述的基于社交网络的内容推荐系统，其中，所述社交网络数据包括社交网络账号发布的社交内容，所述社交网络数据的特征包
5 括所述社交内容的类别和主题，所述类型的用户对所述社交网络数据的行为包括对相同类别或相同主题的所述社交内容的转发行为。

16、一种新闻推荐系统，其包括：

第一特征提取模块，用于提取搜索查询数据的特征；

10 兴趣权值计算模块，用于根据用户对所述搜索查询数据的行为，计算并记录所述搜索查询数据的特征对于所述类型的用户的兴趣权值；

第二特征提取模块，用于提取多个待推送新闻的特征；

兴趣得分计算模块，用于从已记录的特征及兴趣权值中，查找所述多个待推送新闻的特征的兴趣权值，并计算出所述多个待推送新闻对于所述类型的用户的兴趣得分；

15 待推送新闻推荐模块，用于根据所述多个待推送新闻对于所述类型的用户的兴趣得分的高低，对所述类型的用户进行新闻推送。

17、根据权利要求 16 所述的新闻推荐系统，其中，还包括：

第一重新确定模块，用于根据所述类型的用户对所述多个待推送新闻的点击行为，重新确定所述多个待推送新闻的兴趣得分；

20 第二重新确定模块，用于按重新确定的兴趣得分，计算所述多个待推送新闻的特征的兴趣权值并进行记录。

18、根据权利要求 16-17 任一项所述的新闻推荐系统，其中，所述搜索查询数据包括查询词，所述搜索查询数据的特征包括所述查询词的类别和主题，所述类型的用户对所述搜索查询数据的行为包括对相同类别或相同主题的所述查询词的查询
25 行为。

19、根据权利要求 16-18 任一项所述的新闻推荐系统，其中，所述搜索查询数据包括查询结果页上的 URL，所述搜索查询数据的特征包括所述 URL 指向的新闻的类别和主题，所述类型的用户对所述搜索查询数据的行为包括对相同类别或相同主题的新闻的 URL 的点击行为，或对相同类别或相同主题的新闻上的页面标签的点击
30 行为。

20、根据权利要求 16-19 任一项所述的新闻推荐系统，其中，所述搜索查询数据包括社交网络账号发布的 URL，所述搜索查询数据的特征包括所述 URL 中包含的域名的类别，所述类型的用户对所述搜索查询数据的行为包括对相同类别的域名对应的 URL 的点击行为。

35 21、一种计算机程序，包括计算机可读代码，当所述计算机可读代码在计算设备上运行时，导致所述计算设备执行根据权利要求 1-11 任一项所述的方法。

22、一种计算机可读介质，其中存储了如权利要求 21 所述的计算机程序。

1/8

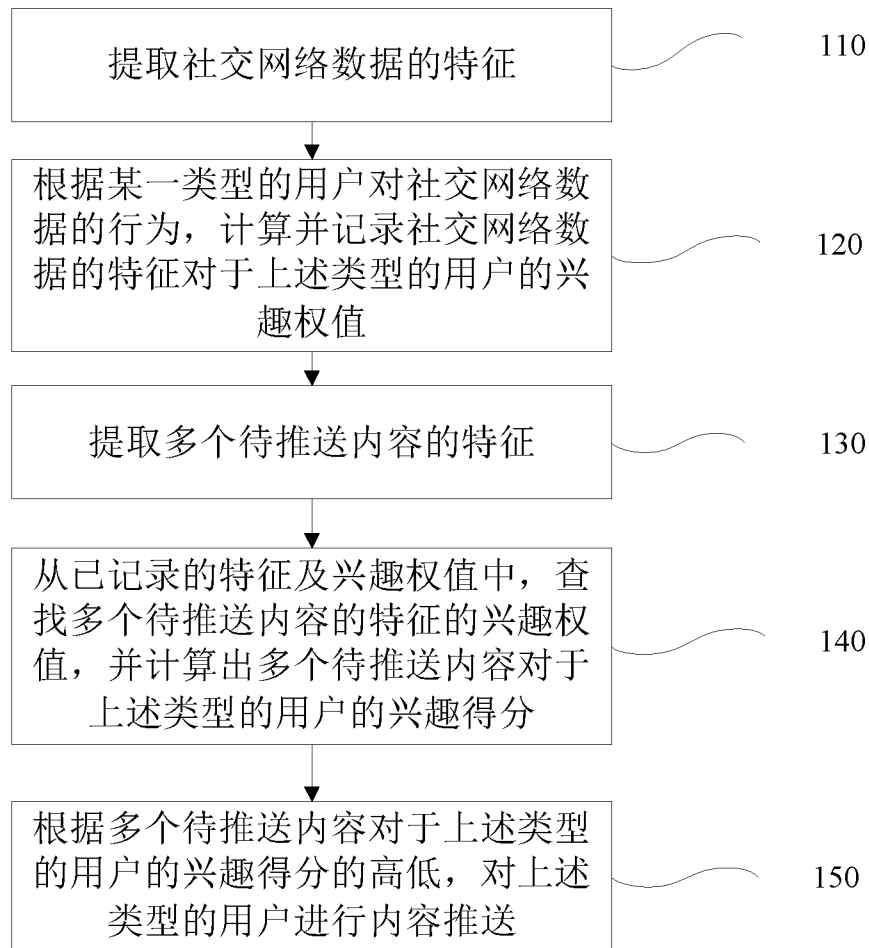


图 1

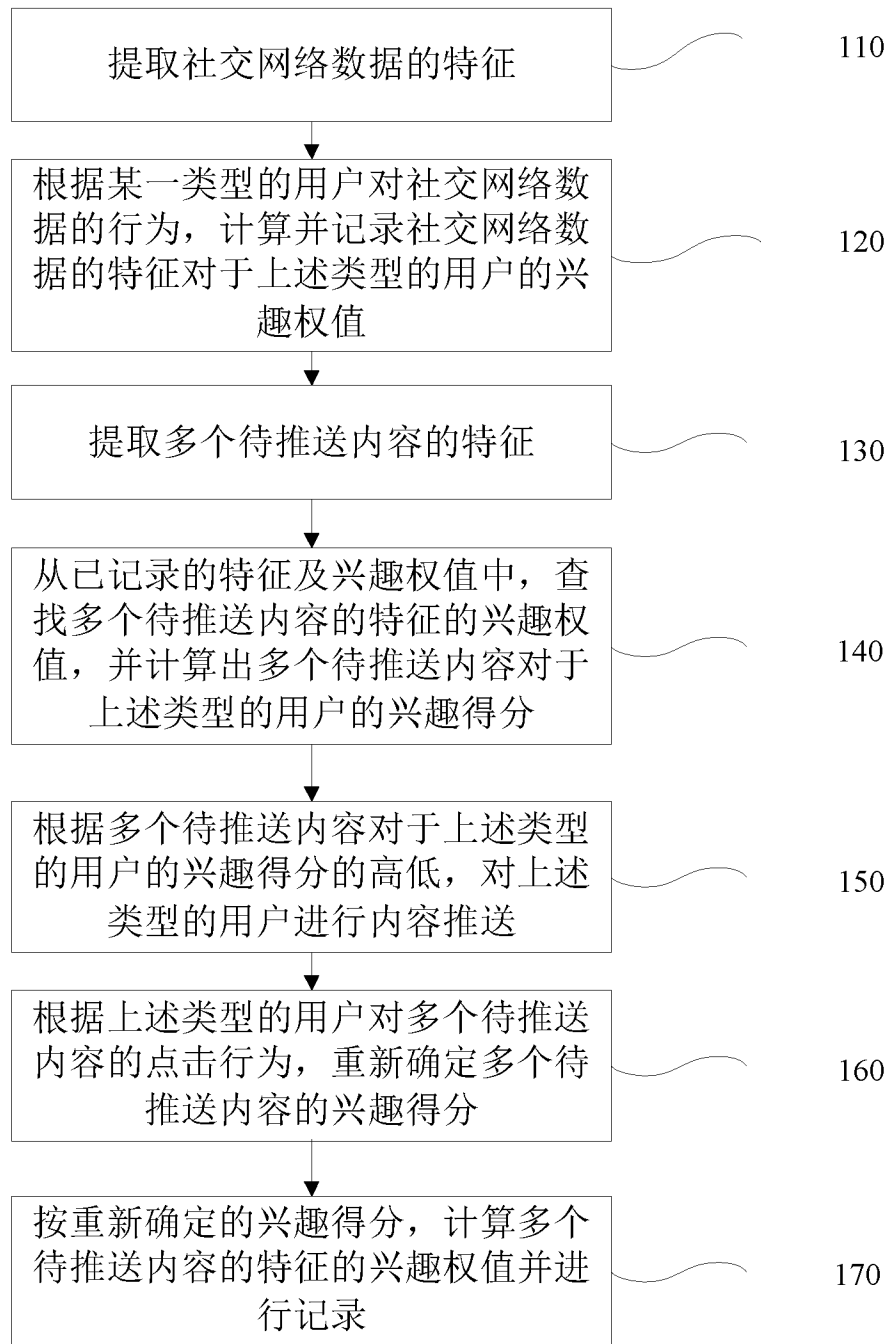


图 2

3/8

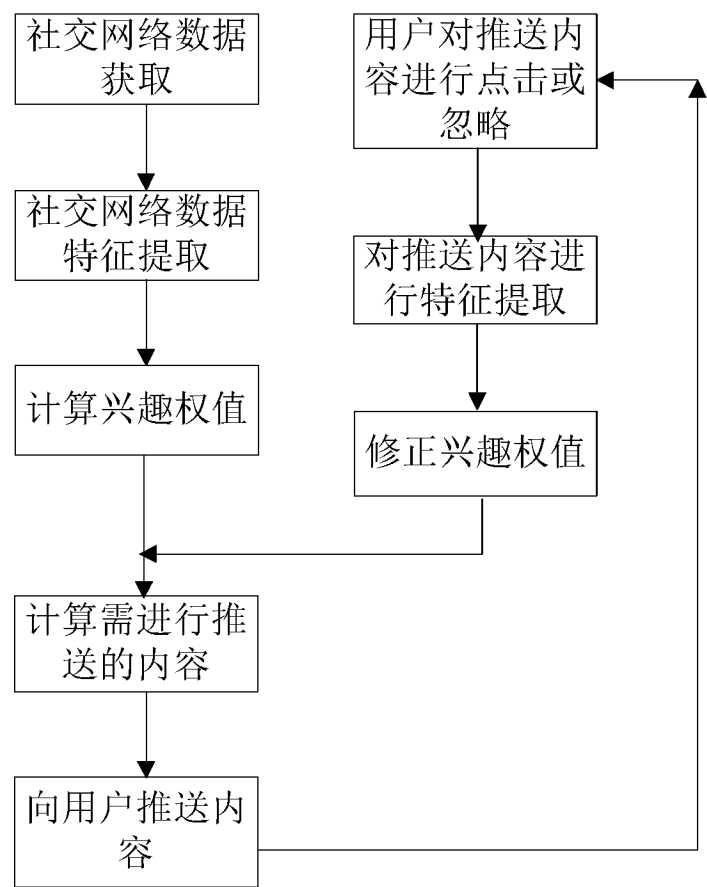


图 3



图 4

4/8



图 5

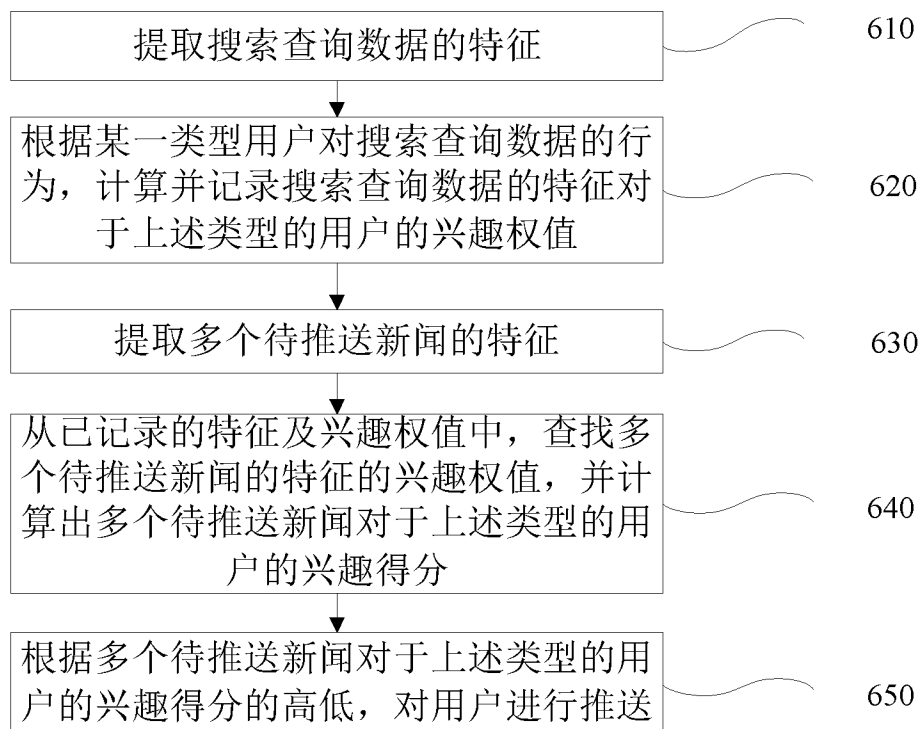


图 6

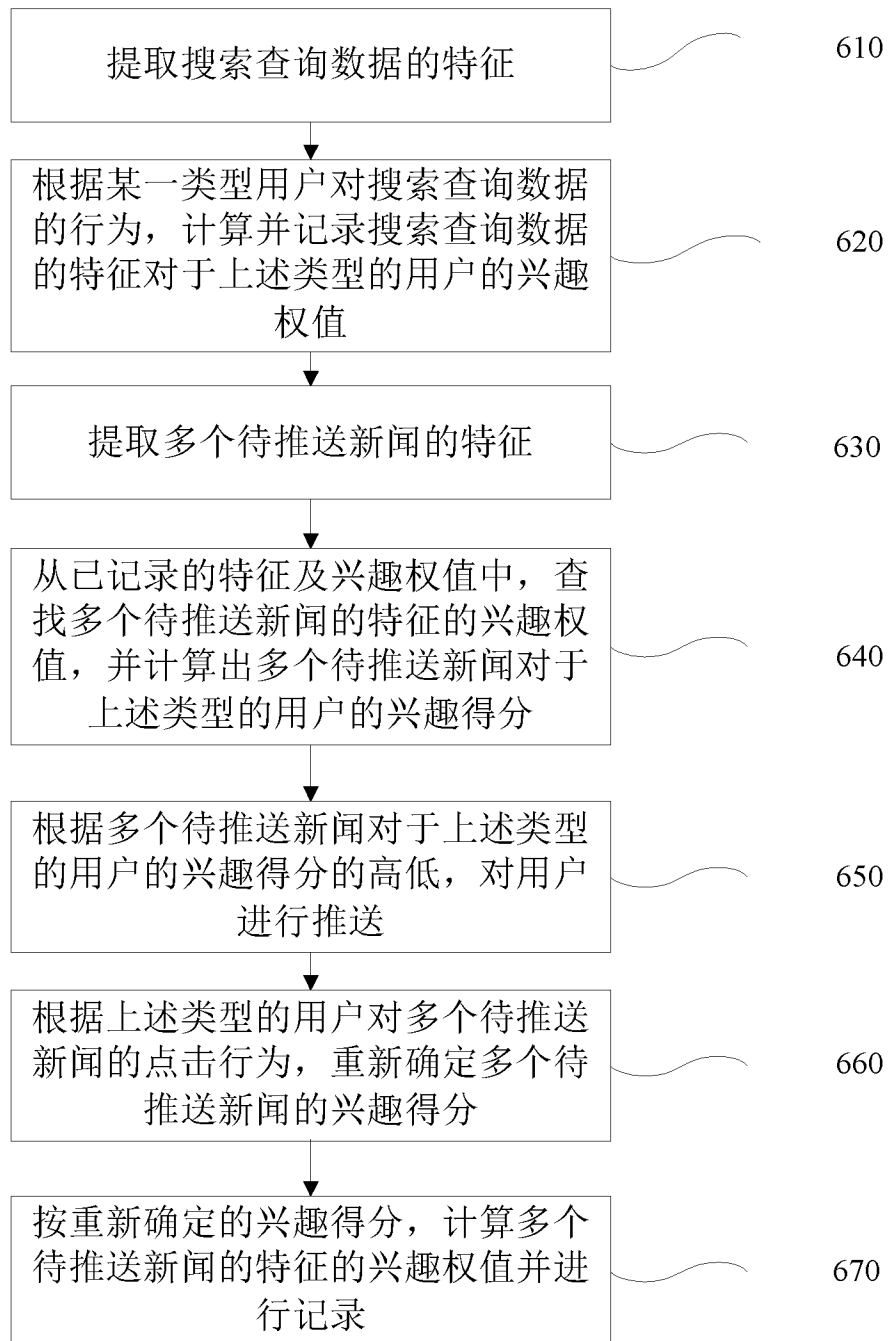


图 7

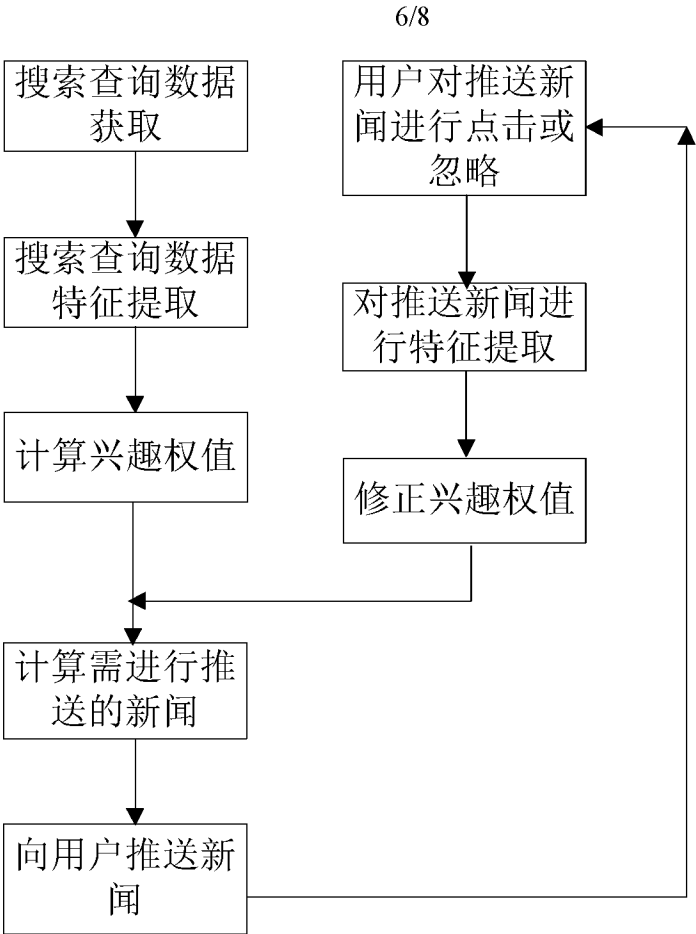


图 8



图 9



图 10

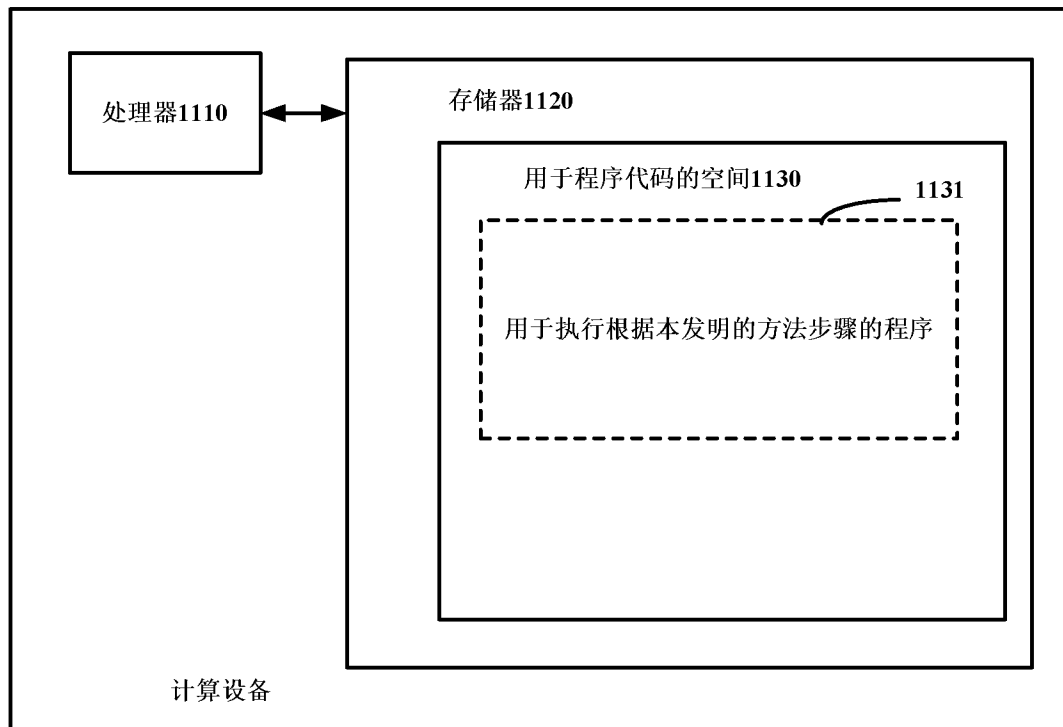


图 11

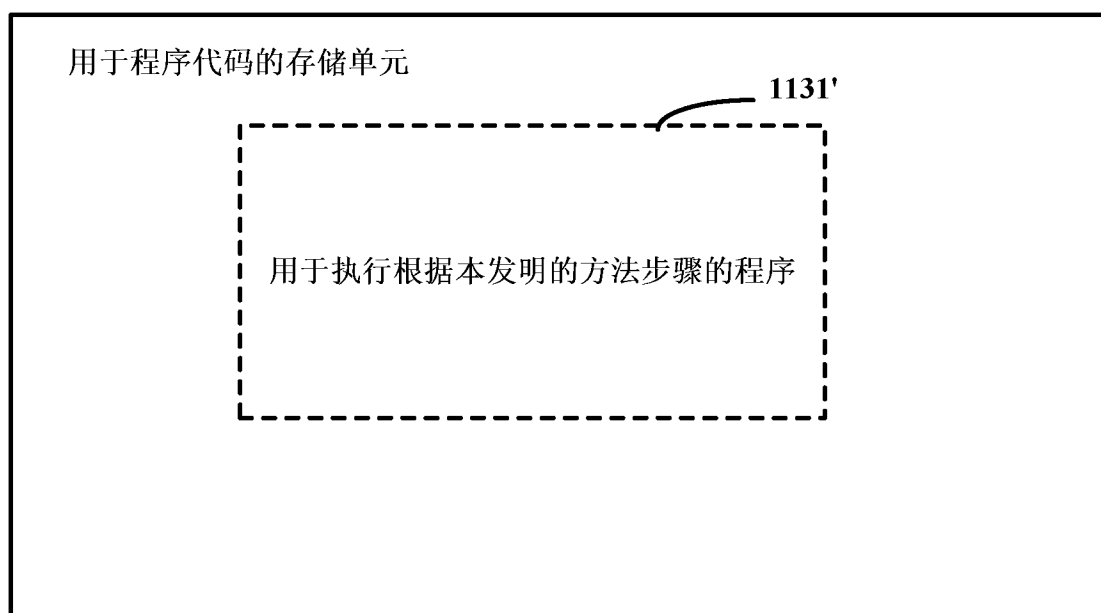


图 12

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/082282

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, DWPI, SIPOABS, CNKI: sociality, community, commend, deliver, push, interest, search, query, news

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 104063476 A (BEIJING QIHOO TECHNOLOGY CO., LTD.), 24 September 2014 (24.09.2014), the whole document	1-6, 12-15, 21, 22
PX	CN 104036038 A (BEIJING QIHOO TECHNOLOGY CO., LTD.), 10 September 2014 (10.09.2014), the whole document	7-11, 16-22
Y	CN 102867016 A (BEIJING KAXINREN INFORMATION TECHNOLOGY CO., LTD.), 09 January 2013 (09.01.2013), description, paragraphs 96-126	1-22
Y	CN 102611785 A (BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS), 25 July 2012 (25.07.2012), description, paragraphs 71-131	1-22

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“A” document defining the general state of the art which is not considered to be of particular relevance	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“E” earlier application or patent but published on or after the international filing date	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	“&” document member of the same patent family
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 23 September 2015 (23.09.2015)	Date of mailing of the international search report 30 September 2015 (30.09.2015)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer ZHANG, Yan Telephone No.: (86-10) 62089421

International application No.
PCT/CN2015/082282

Form PCT/ISA/210 (patent family annex) (July 2009)

A. 主题的分类 G06F 17/30 (2006.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类		
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G06F 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CNABS, DWPI, SIPOABS, CNKI: 社交, 社区, 推荐, 推送, 兴趣, 搜索, 查询, 查找, 新闻, sociality, community, commend, deliver, push, interest, serach, query, news		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 104063476 A (北京奇虎科技有限公司) 2014年 9月 24日 (2014 - 09 - 24) 全文	1-6, 12-15, 21, 22
PX	CN 104036038 A (北京奇虎科技有限公司) 2014年 9月 10日 (2014 - 09 - 10) 全文	7-11, 16-22
Y	CN 102867016 A (北京开心人信息技术有限公司) 2013年 1月 9日 (2013 - 01 - 09) 说明书第96-126段	1-22
Y	CN 102611785 A (北京邮电大学) 2012年 7月 25日 (2012 - 07 - 25) 说明书第71-131段	1-22
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件		
国际检索实际完成的日期 2015年 9月 23日		国际检索报告邮寄日期 2015年 9月 30日
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局 (ISA/CN) 北京市海淀区蓟门桥西土城路6号 100088 中国 传真号 (86-10)62019451		授权官员 张妍 电话号码 (86-10)62089421

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2015/082282

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	104063476	A	2014年 9月 24日	无	
CN	104036038	A	2014年 9月 10日	无	
CN	102867016	A	2013年 1月 9日	无	
CN	102611785	A	2012年 7月 25日	CN 102611785 B	2014年 4月 2日