



(12) 发明专利申请

(10) 申请公布号 CN 104036002 A

(43) 申请公布日 2014. 09. 10

(21) 申请号 201410266630. 3

(22) 申请日 2014. 06. 16

(71) 申请人 深圳市英威诺科技有限公司

地址 518000 广东省深圳市南山区高新南四
道创维半导体大厦 6 楼 605-610

(72) 发明人 不公告发明人

(51) Int. Cl.

G06F 17/30 (2006. 01)

权利要求书1页 说明书2页

(54) 发明名称

一种智能推荐数据的技术方法

(57) 摘要

本发明主要提供了一种终端用户能够快速获取信息的一种解决方法,以解决现有技术中用户在运用互联网的过程中需要从众多的信息中去搜索、寻找自己喜欢的或需要的资讯所带来的不便。此项技术解决方法可以改变用户被动获取信息和花时间搜索信息的模式,让互联网真正改变我们人类的生活习惯。首先要使用爬虫获得大量的信息,然后使用提取关键字算法和主题算法,建立海量的信息源,其次需要构建用户数据模型,并根据数据模型采集用户的浏览信息,点击信息,智能设备的基本情况和使用行为,以及位置变化的轨迹等行为数据,然后使用模型算法不停的完善用户的特征画像。再次当一个用户在使用智能设备时,服务器会根据该用户的特征画像,判断该用户在该时刻需要什么信息,并给用户推荐什么信息。

1. 一种智能推荐信息的一种方法,应用于在使用互联网软件时分析用户喜好和需求并进行智能推荐信息的一种方法,其特征在于:

要使用爬虫获得大量的信息,然后使用提取关键字算法和主题算法,建立海量的信息源,构建用户数据模型,并根据数据模型采集用户的浏览信息,点击信息,智能设备的基本情况和使用行为,以及位置变化的轨迹等行为数据,然后使用模型算法不停的完善用户的特征画像,当一个用户在使用智能设备时,服务器会根据该用户的特征画像,判断该用户在该时刻需要什么信息,并给用户什么信息。

2. 根据权利要求 1 所述的方法,其特征在于,还包括:

搭建数据处理系统,抓取用户数据,根据用户的兴趣点,推荐服务器做更细分的多层级计算,如:体育 -- 篮球 -- NBA -- 乔丹等,并推荐信息给用户。

3. 根据权利要求 1 所述的方法,其特征在于,还包括:

构建用户模型和采集机制,根据用户的属性特征,唯一性,喜好,活动轨迹等结合场景算法,形成一个特定的场景,引导用户,为用户提供智能服务。

4. 根据权利要求 3 所述的方法,其特征在于,还包括:

根据关键字、主题、多级类别、pv、点击、分享、收藏、评论以及场景变化得到该用户的推荐信息列表,并根据特征总权重分数排序,根据智能终端请求信息的条数给用户推荐信息。

一种智能推荐数据的技术方法

[0001] 技术领域：

本发明涉及移动终端智能推送的技术方法，特别地涉及一种移动互联网大数据量下根据用户喜好和需求推送数据的技术实现方法。

[0002] 背景技术：

随着互联网的发展，特别是智能手机的普及和移动互联网到来，使越来越多的人能够随时随地的接触到互联网，从多人共用一台手机，到人手一台智能手机，使连接互联网的终端，越来越能体现一个人的特征，但是人类获取信息的方式却没有太多变化，还是遵循在输入框中输入关键字，然后点击查询，然后从大量的查询结果中找到自己想要的信息。在信息爆炸的今天这是一件很繁琐和费时的事情。而目前市场上根据用户喜好和需求推送信息(以下简称智能推荐)还未有先例。

[0003] 智能推荐是根据用户的喜好和需求给用户推送信息，不是简单的根据某一类别推送，也不是简单的相关类别的推荐，而是根据人的喜好和需求，如喜欢美食、爱打篮球、关注时政、爱玩游戏、经常出差、出去旅游等等。智能推荐将达到一种效果：当用户想去吃饭的时候，就会得到他附近的符合他饮食习惯的餐厅；当用户出差时，可实时提供当地的酒店、餐饮等相关信息给客户；同时还能够根据用户喜好变化而变化，如：用户浏览汽车方面的信息时，用户可以看到更多汽车方面的信息，当用户的想看体育方面的信息时，随着用户点击的增多，体育方面的信息就会增多，而汽车方面的信息就会相应减少，除此之外还会有场景识别等，通过场景识别来为客户推送相应的信息。

[0004] 目前实现智能推荐的方法是：根据用户在智能终端使用习惯和浏览并点击信息的习惯，再通过长期属性算法和实时熟悉算法，算出用户的长期属性和短期属性，并根据这些属性给每个用户绘制特征画像，根据特征画像给用户推送数据，并且能够根据用户的实时属性实时的递增或衰减相应的信息数据。

[0005] 发明内容：

本发明的主要目的是提供一种用户能够快速获取信息的解决方法，以解决现有技术中用户在运用互联网的过程中需要从众多的信息中去搜索、寻找自己喜欢的或需要的资讯所带来的不便。此项技术解决方法可以改变用户被动获取信息和花时间搜索信息的模式，让互联网正在改变我们人类的生活习惯。

[0006] 为解决上述问题本发明提供如下技术方案：

1、要使用爬虫获得大量的信息，然后使用提取关键字算法和主题算法，建立海量的信息源。

[0007] 2、构建用户数据模型，并根据数据模型采集用户的浏览信息，点击信息，智能设备的基本情况和行为，以及位置变化的轨迹等行为数据，然后使用模型算法不停的完善用户的特征画像。

[0008] 3、当一个用户在使用智能设备时，服务器会根据该用户的特征画像，判断该用户在该时刻需要什么信息，并给用户什么信息。

[0009] 具体实施方式：

一、首先要搭建爬虫系统,该系统要实时抓取大量的各种丰富的类型的数据,以保证服务器有足够丰富的数据,这样服务器才能满足更多的用户群体,还要抓取数据的原始热度,以保证数据的起初热度。

[0010] 二、要搭建数据处理系统,关键字提取算法很重要,要能通过自主学习不停的丰富自己的关键字库,类型的划分很重要,不能完全根据互联网的分类来分类,必须要有人工干预,因为互联网的分类杂乱无章,往往词不同意同,同时分的类别层级高低不一,如有的网站是”体育 -- 篮球”有的网站是”体育 --NBA “。在类型划分很重要,对实时算法会有影响,因此类型的划分要足够的细致与合理,如:“可以是体育 -- 篮球 --NBA-- 乔丹”。主题算法,要对信息进行主题的提取,这更能得到用户喜欢什么类型的主题。数据的处理结束后的灌库很重要,海量数据的存储,查找,更新是处理大数据量的能力,很好的处理掉这些数据,是能否快速响应用户的信息需求的根本,本系统使用了 google 的一个分布式多层级的海量数据存储的框架技术 leveldb。这个框架能高效的完成海量数据的存储,查找,计算的工作。

[0011] 三、构建用户模型和采集机制,一个真实的用户,包含着用户的基本属性,如性别、年龄的区分等,用户的数据模型,其中还包含用户的唯一标识符,就像我们的身份证号码一样,具有唯一性。根据真实用户的特性,用户数据模型按变化性分为长期属性和实时属性,长期属性如:性别、居住地、工作地、爱好、主题、关键字。短期属性如:pv、点击、多级类别、关键字。长期属性每隔一段时间(可配置)绘制用户特征画像,用于当用户没有实时属性可根据长期属性给用户推荐信息。短期属性会随着用户在智能终端的操作实时发生变化,并根据变化实时给用户推送信息

四、当用户使用智能终端时,会将该用户的唯一标识发送到推荐服务器,推荐服务器会根据用户的标识,得到该用户的特征画像,然后依次根据关键字、主题、多级类别、pv、点击、分享、收藏、评论以及场景变化得到该用户的推荐信息列表,并根据特征总权重分数排序,根据智能终端请求信息的条数给用户推荐信息。