

面向多源异构数据的云推送平台的研究与应用

陆佳炜 王辰昊 肖 刚 徐 俊

(浙江工业大学计算机科学与技术学院 杭州 310023)

摘 要 在传统推送中,多源异构数据的推送面临时效性不强、安全性不高、数据难以重用等问题。针对多源异构数据的特征,综合移动互联网的安全性和隐私性等特点,提出了一种多维决策云推送模型来计算分布式环境中多源异构数据的特征值和特征向量,以快速分离数据源中的同构数据和异构数据,并基于此模型设计了云推送平台,其利用云推送技术来实现同构数据和异构数据的自动分离和高效推送。根据云推送平台在实验环境中的运行情况及相关指标分析,说明该平台适用于多源异构数据的推送,是一种高效可行的推送方式。

关键词 云推送,多源异构数据,云计算,数据推送

中图法分类号 TP391 文献标识码 A

Research and Application of Cloud Push Platform Based on Multi-source and Heterogeneous Data

LU Jia-wei WANG Chen-hao XIAO Gang XU Jun

(College of Computer Science & Technology, Zhejiang University of Technology, Hangzhou 310023, China)

Abstract In the traditional push, the push of multi-source heterogeneous data is weak on timeliness, low security and difficult to reuse. According to the characteristics of heterogeneous data, characteristics of mobile Internet security and privacy, a multi dimensional decision cloud push model was proposed, to compute the eigenvalues and eigenvectors of multi-source and heterogeneous data in distributed environment, for the rapid separation of data source in homogeneous data and heterogeneous data. And based on this model, the cloud push platform was designed, and the automatic separation and high efficiency push of homogeneous data and heterogeneous data was realized by using cloud push technology. According to the running situation and related index analysis of the cloud platform in the experimental environment, the platform can be applied to the multi source and heterogeneous data.

Keywords Cloud push, Multi-source heterogeneous data, Cloud computing, Data push

近几年,随着移动通信技术的快速发展,移动互联网已成为互联网发展的主流,成为人们获取信息的主要渠道。通过移动互联网,大量多源异构数据整合在一起,创造了全球性的信息共享环境。为了满足用户的需求,各种移动应用呈爆炸性增长,需要推送的数据也向多源化发展。这些多源异构数据对移动终端设备上信息的捕获提出了较高要求,主要体现在移动性和时限性两个方面。移动性要求低功耗、低速率的消息传输;时限性要求信息在规定时间内发送到移动终端设备上。

如何快速有效地推送多源异构数据成为社会面临的新问题。这些多源异构数据难以共享,也没有统一的方法进行推送,传统的数据推送方法也已经不能适应现在用户对信息的实时需求。目前主要的推送方式是由服务器主动将变化的信息发送给用户,无需用户参与,减少了交互的次数和负担,缩短了反应时间,提高了效率,但是这些方法都不适用于多源异构数据的推送。IOS平台和Android平台虽然都有自己的推送系统,但在网络、操作系统和应用方面的限制,使其在使用上有一定的局限性;谷歌云消息服务^[13]在国内也无法使用;

iPhone上的APNS也仅适用于IOS^[14],无法跨平台推送。所以,国内外目前对于跨平台的多源异构数据还无法进行实时高效的推送。

为了克服现有的不足,提出了面向多源异构数据的MD-CP(Multiple Decision Cloud Push)模型,设计了云推送平台,并对多源异构数据进行推送实验。实验结果显示,相比于原有的服务器推送机制,面向多源异构数据的云推送平台不仅减少了数据推送更新过程中的同步数据量,缩短了时间,而且解决了多源异构数据的跨平台推送的问题,特别在数据量巨大且工作在低带宽、网络不可靠的情况下,工作效率有明显提升。

1 相关工作

国内外学者和研究机构从不同的视角对多源异构数据和推送方式进行了研究。

首先,从数据传输方向出发的代表性工作主要是,Yang Wang, Bharadwaj Veeravalli^[1]设计了一种分阶段获取云端分享数据的算法,其能够有效地控制数据的传输成本;国内的许

本文受浙江省自然科学基金(LQ12F02016),浙江省科技厅公益性技术应用研究项目(2014C33071, 2014C31078),浙江省重大科技专项(2014C01048)资助。

陆佳炜(1981—),男,硕士,讲师,主要研究方向为云计算、软件工程, E-mail: viivan@zjut.edu.cn; 王辰昊(1989—),男,硕士生,主要研究方向为云计算; 肖 刚(1965—),男,博士,教授,主要研究方向为图像识别、云制造; 徐 俊(1979—),男,硕士,高级实验师,主要研究方向为云计算。

福龙、刘明等人[2]进一步提出了一种基于相对距离感知的动态数据传输策略,其采用传感器节点到汇聚点的相对距离来计算节点传输概率的大小,并以此作为消息传输时选择下一跳的依据。

从决策方向出发的研究有,Yang Fang-Chun,Su Sen 提出一种基于模糊多属性决策理论的语义 Web 服务组合的优化选择算法[3],该算法能够评价以实数、区间数和语言型数据描述的信息,从而进行综合决策,同济大学的蒋乾悦、张亚英进一步提出了一种基于模糊综合决策的服务器推送方法[4],该方法将传统的长轮询技术和轮询相结合,得到了一种复合式轮询技术。

在利用推送技术实现系统的研究中,中国科学院软件研究所的刘鑫、陈伟[5]提出了一种基于 AJAX 和 Server Push 的 Web 树组件[6],其为用户提供了类似于在 windows 资源管理器中对目录树操作的基本功能和用户体验。

但以上方法均只是通过修改推送方式来实现对单一数据源进行推送,并没有过多考虑多源异构数据的推送问题。本文设计的云推送平台面向多源异构数据,通过计算数据的特征值和特征向量来快速分离同构数据和异构数据,以实现数据的高效推送。

2 云推送平台设计

2.1 云推送平台逻辑架构

本文提出的云推送是指云计算环境下[7]由云服务器端主动向客户端发送消息的技术,用户不需要主动检查更新,云服务器端将各个系统中的多源异构数据通过云推送平台主动[8]推送给用户。云推送平台逻辑架构如图 1 所示。

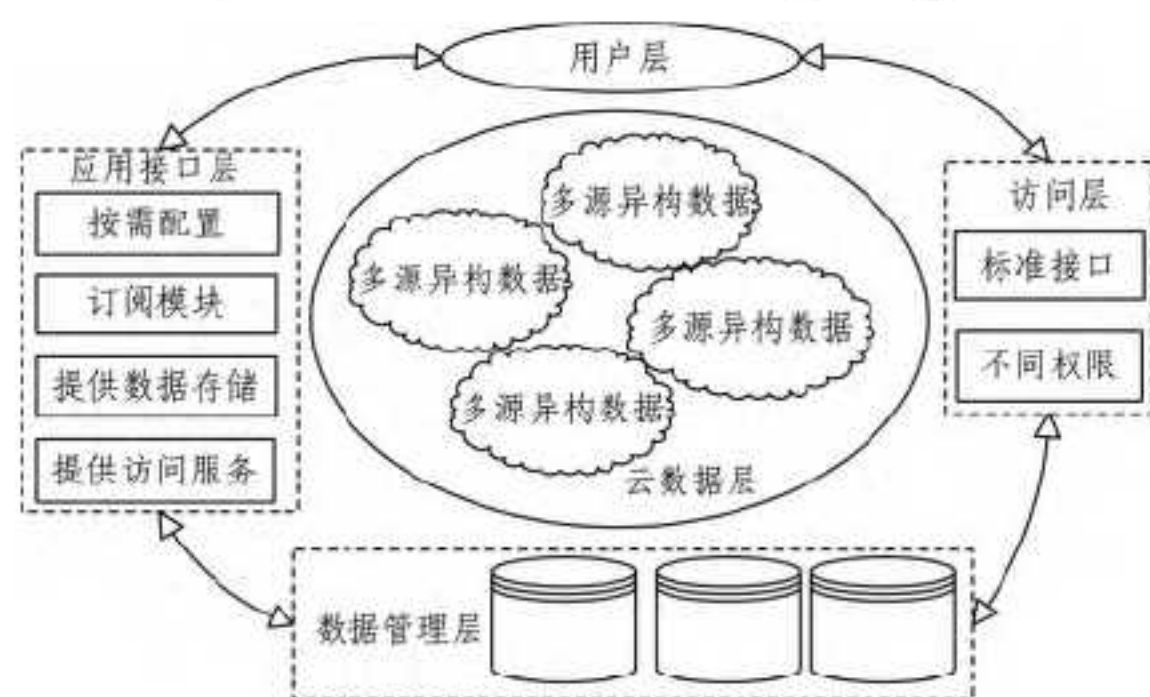


图 1 云推送平台逻辑架构

如图 1 所示,云推送平台逻辑架构一般可分为如下 5 层。

1) 云数据层

该层是云推送平台中的主要部分,各个系统中的多源异构数据形成了云数据,所有数据都属于该层,其实现了数据的逻辑虚拟化。

2) 数据管理层

该层是架构的核心,实现了云推送中数据与推送的协同工作,对外也提供具有相同数据的访问服务。

3) 应用接口层

该层可按需配置,根据用户订阅的不同应用或模块,提供不同数据存储和访问的应用服务,如工资推送服务、学习成绩服务、炒股信息服务、视频监控服务等。

4) 访问层

授权用户均可通过标准的接口来访问云推送系统,根据

不同的用户权限,各类用户访问的方式也不完全相同。

5) 用户层

用户层主要负责管理各类注册用户,同时也管理各类登录终端,如 PC,IOS,Android 等。

2.2 云推送平台功能设计

针对互联网传递消息类型的差异性,设计的云推送平台可以推送 3 种类型的消息:通知、透传消息及富媒体[9],同时应支持向所有用户或根据标签分类向特定用户群体推送消息,提供用户信息及通知消息统计信息,并且可以跨平台使用(PC,IOS,Android),用户还可以根据自己的需要添加自定义功能。云推送平台的功能设计如图 2 所示。

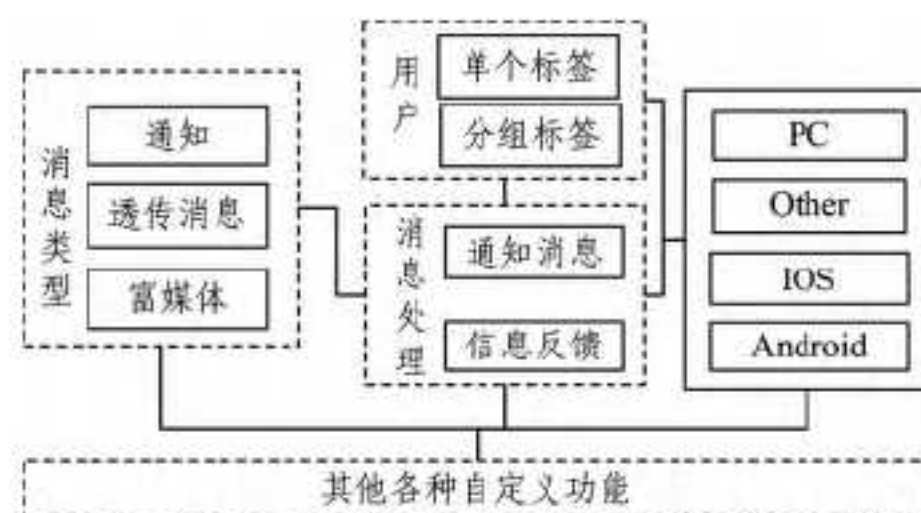


图 2 云推送平台的功能设计

2.3 云推送过程设计

云推送就像是一个信息快递服务,当用户成功订阅云推送平台的内容后,平台需要向自己的用户发送信息,通过云端和客户端之间建立的长连接,实时向客户端推送消息。文中提出的云推送过程在传统推送的基础上,围绕“订阅_收集_决策_推送”这一周期展开[10]。云推送过程要解决的问题主要体现在以下两点:1)如何收集多源异构数据;2)如何使多源异构数据快速有效地推送到用户手中。

针对这两个问题,提出了一种 MDCP 模型用于云推送平台,并基于此模型给出了云推送平台的设计过程,如图 3 所示。根据用户订阅的模块选择对应的多源异构数据,将数据收集[11]到信息库中,MDCP 模型将访问信息库的数据,通过确定权值和属性分离进行决策推送,即为整个“订阅_收集_决策_推送”周期。根据该周期的循环性,用 MDCP 模型筛选出重复数据和变化数据信息,再根据特征向量区分出多源同构数据还是多源异构数据,最后根据用户订阅的系统的数量、网络的带宽、推送的数据的数量以及保密性[12]等综合判断所使用的推送方式。

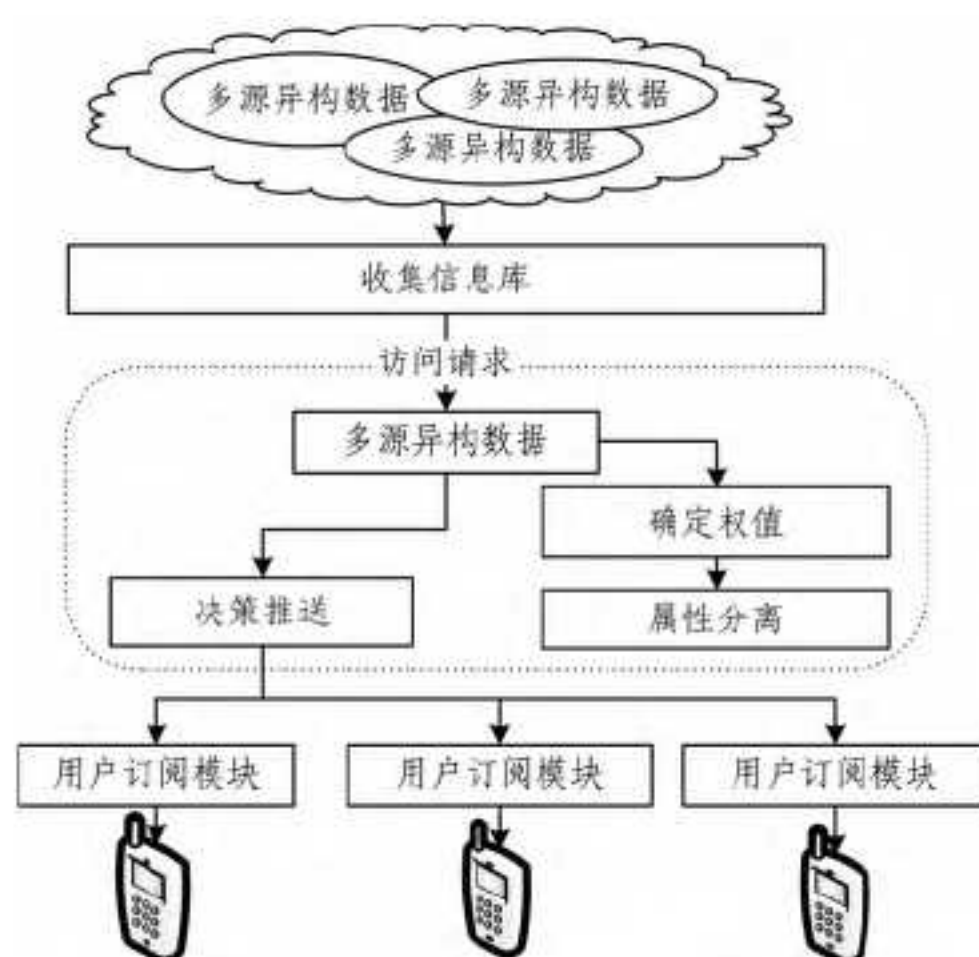


图 3 云推送过程设计

3 MDCP 模型研究

3.1 多源异构数据成分分析

由于对异构性问题欠考虑,导致在推送时效率降低^[13],甚至无法工作。所以在面对异构数据时,选择一个合适的推送方式是必要的。

本文需要解决的问题就是利用这些分布在不同云服务器上的云数据来得到数据的主成分,通过将每个主要数据成分值 X_i 与给出的 μ 相减即得到该成分在该方向上的偏移值^[14],将每个数据的主要成分值 $X_i (i=1,2,\dots,n)$ 填入矩阵 S 中。

$$S=[X_1, X_2, X_3, \dots, X_n] \quad (1)$$

再将矩阵 S 乘以它的转置矩阵 S^T 后求和再取平均。由于该矩阵乘以它的对角矩阵后得到的为确定的数值,因此该值可以用来表示整体的异构程度,用 V 表示。

$$V=\frac{1}{N}\sum_{i=1}^N\{|(X_i-\mu)|\cdot|(X_i-\mu)^T|\} \quad (2)$$

其中, μ 是根据评估得出的权重值,根据不同的系统得到其取值范围; N 为样本总量; 由于此时 V 为确定的数值,该样本的异构程度可用特征值 V 表示。

如果出现多源异构数据需要推送,此时每个数据的主成分不能分为 X_1-X_N , 需要 X_i 变为 X_{ij} , 所以矩阵 S 变为:

$$S=\begin{bmatrix} X_{11} & X_{12} & \dots & X_{1N} \\ X_{21} & X_{22} & \dots & X_{2N} \\ \dots & \dots & \dots & \dots \\ X_{i1} & X_{i2} & \dots & X_{iN} \end{bmatrix} \quad (3)$$

此时, V 为矩阵, 视为特征矩阵。所以通过 V 乘以它的对角阵 V^A 与 $V-kE$ 的绝对值相等, 即可解出远数据的特征值 k , E 为单位矩阵:

$$|V-kE|=|V \cdot V^A| \quad (4)$$

V 给出了特征向量, 通过化简求出了相应的特征值 k , 特征值 k 的大小和 V 共同度量了不同云数据的异构程度。

但是该方法中没有考虑到数据传输负荷的问题, 分布式环境中不同的云数据传输受到通信率、计算能力等因素的限制, 希望减少云服务器之间的数据通信量, 避免直接传输大量样本, 可将数据以矩阵方式进行排列组合, 再基于此数据结构进行云推送。

3.2 多源同构数据的推送

在云数据中, 虽然存在着大量的多源异构数据, 但是还存在着海量多源同构数据。当用户需要的推送内容属于同构数据时, 使用协方差矩阵将数据按特征向量分解, 分解同构数据就能使用更高效的推送方式, 不需要在各个点反复迭代^[15], 节约了数据通信率。

首先, 用第一个云数据站点上的样本估计局部方差和均值向量 μ_1 , 再取得第二个云数据站点上的样本估计局部方差和均值向量 μ_2 , 直到得到所有云数据站点的样本 N_K 。在局部方差和均值向量的规定范围内, 则认为是这些云数据站点提供的是同构数据, 求出平均均值向量 $\bar{\mu}$, 再通过 $\bar{\mu}$ 求出平均参数协方差:

$$\bar{\mu}=\frac{1}{N+N_K} \quad (5)$$

$$\sum_{i=1}^{N+N_K} X_i=\frac{1}{N+N_K}(N_{\mu}+N_{K\mu_k}) \quad (6)$$

因此在各个云数据样本充足的情况下, 计算同构数据的准确性是非常高的。由于是同构数据, 次增量更新方程为:

$$\Sigma^*=\frac{1}{N+N_K}\sum_{i=1}^{N+N_K}(x_i-\mu^*)(x_i-\mu^*)^T \quad (7)$$

$$\mu^*=\bar{\mu}-\mu_K \quad (8)$$

其中, $K=1,2,3,\dots$ 。在上述算法中, 可以注意到协方差矩阵是对称的, 因此在推送过程中, 同构数据每次更新只需要传输半个协方差矩阵, 降低了数据传输量。由于多源同构数据传输量较少、计算量有限、开销量小, 因此选择小型传输协议, MDCP 模型使用基于 WebSocket 的 MQTT^[16] 进行推送, 同时降低网络流量。若系统数目为 M , 参数个数为 d , 该算法的时间复杂度为:

$$O(M,d)=(M-1)(d(d+1)/2+d+1)$$

3.3 多源异构数据的推送

多源异构数据是对同构数据主要成分进行扩展。为了方便数学计算, 引入置换矩阵 P , 对云服务器上取得的样本进行置换映射, 结果记为 y , 即:

$$y=(y_a, y_b)^T=P_m X \quad (9)$$

其目的是将当前样本中与同构数据中相同的部分集中到向量的前面, 用 y_a 表示, 而将不同的部分放置在向量的后面, 用 y_b 表示。同样, 对均值矩阵和协方差矩阵进行置换, 结果分别记为 μ, Σ :

$$\mu=(\mu_a, \mu_b)^T=P_m \mu$$

$$\Sigma=\begin{pmatrix} \sum a \sum b \\ \sum b \sum c \end{pmatrix}=P_m \Sigma P_m^T \quad (10)$$

μ 和置换矩阵都是上一次计算得出的量, $\sum a$ 是在 m 样本中已经得到的协方差矩阵, $\sum c$ 是与同构数据中不同元素的协方差矩阵, 而 $\sum b$ 是它们之间协方差构成的矩阵。

该公式符合高维分布, 所以云数据样本就可以采用该高维分布计算, 其表达式如下:

$$N(y|\mu, \Sigma)=\frac{1}{(2\pi)^{1/2} \Sigma^{1/2}} \exp\left\{-\frac{1}{2} |(y-\mu)^T| \Sigma (y-\mu)^{-1}\right\} \quad (11)$$

y, μ, Σ 通过式(5)、式(7)、式(8)3个公式代入, 进行简化, 并取自然对数可得:

$$l(\mu, \Sigma y)=\sum_i \log P(\mu_a, \sum a | y_{aj}, \mu_b, \sum c, \sum b) \quad (12)$$

再求偏导, 化简后可得:

$$\sum a=\frac{1}{N_m} \sum_{i=1}^{N_m} (y_a - \mu_a)(y_a - \mu_a)^T - \Sigma_c^{-1} \Sigma_b^T \frac{1}{N_m} \sum_{i=1}^{N_m} (y_b - \mu_b)(y_b - \mu_b)^T \Sigma_c^{-1} \Sigma_b + \Sigma_b \Sigma_c^{-1} + \Sigma_b \Sigma_c^{-1} \Sigma_b^T \quad (13)$$

其中第一行是与同构数据相同的元素项, 已经在式(8)中计算出来, 第二行就是异构数据。所以信息源更新后推送异构数据的公式即为:

$$\Sigma d=\frac{1}{N} \sum_{i=1}^N (y_b - \mu_b)(y_b - \mu_b)^T \quad (14)$$

由于在异构数据中, 每种云数据的元素的个数是不一定的, 因此需要查找每个 N 。由于转置矩阵的存在, 仅需要传输一半的数据, 降低了数据传输量。因此传输复杂度为:

$$O(M,d)=(M-1)(d^2+2d)$$

这部分复杂的异构数据推送, 计算量比较大, 对网络要求也较高, MQTT 传输协议已经不再适合, 所以 MDCP 模型将

调用互联网上公开的第三方云推送服务来进行推送。第三方平台的消息推送服务完全免费,可以零成本使用,并拥有强大的服务器集群,具有极高的吞吐能力,用户订阅的消息能以更快的速度送达用户端。

3.4 MDCP 模型运行流程

通过上述分析,原有的推送系统确实难以实现对多源异构数据的实时高效推送。当得到需要推送的任务之后,调用 MDCP 模型,其运行过程如图 4 所示。

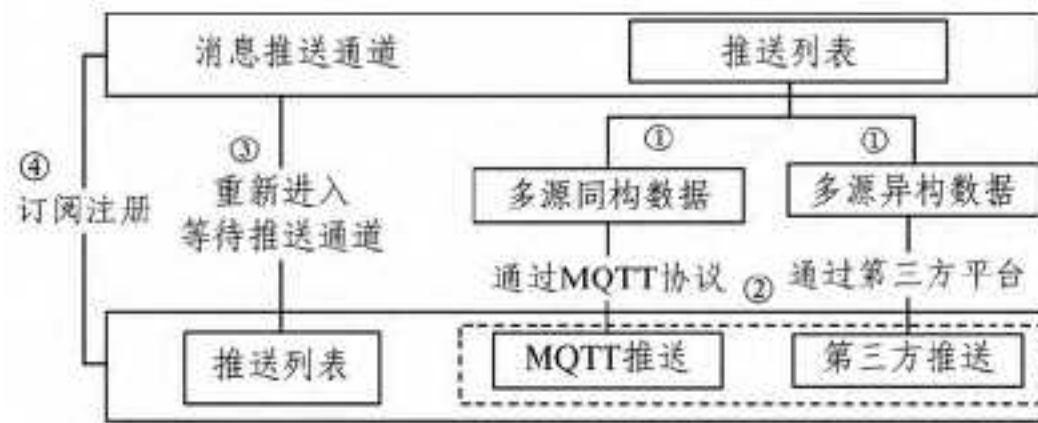


图 4 MDCP 模型运行过程

当消息进入推送列表时,MDCP 模型通过特征值计算分离出多源同构数据和多源异构数据,若是多源同构数据,则使用 MQTT 协议推送方式(图 4①);若是多源异构数据,则使用第三方平台推送(图 4②);若无法推送,则重新进入等待通道(图 4③)。

4 案例分析与实验

使用云推送平台在安卓和 IOS 中进行了测试,云数据来

$$V_1 = \begin{bmatrix} |0.085-0.3| & |0.122-0.2| & |0.231-0.2| & |0.081-0.1| & |0.098-0.1| & |0.95-0.1| \\ |0.127-0.3| & |0.128-0.2| & |0.246-0.2| & |0.200-0.1| & |0.310-0.1| & |0.810-0.1| \\ |0.891-0.3| & |0.129-0.2| & |0.871-0.2| & |0.764-0.1| & |0.990-0.1| & |0.769-0.1| \\ |0.361-0.3| & |0.765-0.2| & |0.560-0.2| & |0.809-0.1| & |0.591-0.1| & |0.760-0.1| \\ |0.819-0.3| & |0.581-0.2| & |0.129-0.2| & |0.794-0.1| & |0.899-0.1| & |0.870-0.1| \\ |0.270-0.3| & |0.800-0.2| & |0.531-0.2| & |0.750-0.1| & |0.999-0.1| & |0.969-0.1| \end{bmatrix}$$

$$V = \frac{1}{5} \sum_{i=1}^5 V_i \cdot V_i^T$$

$$= \begin{bmatrix} 0.012 & 0.167 & 0.231 & 0.431 & 0.981 & 0.993 \\ 0.125 & 0.145 & 0.246 & 0.265 & 0.230 & 0.231 \\ 0.801 & 0.120 & 0.871 & 0.564 & 0.923 & 0.928 \\ 0.323 & 0.765 & 0.523 & 0.857 & 0.391 & 0.125 \\ 0.832 & 0.581 & 0.127 & 0.756 & 0.867 & 0.849 \\ 0.246 & 0.800 & 0.531 & 0.750 & 0.012 & 0.980 \end{bmatrix}$$

得到特征矩阵之后,通过式(4)可以求出特征值,

$$\begin{bmatrix} 0.012 & 0.167 & 0.231 & 0.431 & 0.981 & 0.993 \\ 0.125 & 0.145 & 0.246 & 0.265 & 0.230 & 0.231 \\ 0.801 & 0.120 & 0.871 & 0.564 & 0.923 & 0.928 \\ 0.323 & 0.765 & 0.523 & 0.857 & 0.391 & 0.125 \\ 0.832 & 0.581 & 0.127 & 0.756 & 0.867 & 0.849 \\ 0.246 & 0.800 & 0.531 & 0.750 & 0.012 & 0.980 \end{bmatrix} -$$

$$k \begin{bmatrix} 100000 \\ 010000 \\ 001000 \\ 000100 \\ 000010 \\ 000001 \end{bmatrix}$$

化简后可得,

源于成绩系统、工资系统和微影视系统,采集到数据后将数据的主要成分值权重元素设为:用户权限、用户登录时间、用户发布/订阅的模式(一对多/一对一)、用户登录数量、传输数据量。权重 μ 是经过综合考虑而确定的,假设本系统中成分值权重为(0.3,0.2,0.2,0.1,0.1,0.1)。权重越高,该成分值的变化对最后 MDCP 推送结果的影响越大^[14,15]。由于数据来源于 3 个系统,并且微影视系统的数据类型不同(存在富媒体信息),因此采集到的数据属于多源异构数据,如表 1 所列。

表 1 云数据数据成分值

工资系统	成绩系统	影响系统	工资系统	成绩系统	影响系统
0.085	0.122	0.231	0.081	0.098	0.95
0.127	0.128	0.246	0.200	0.310	0.810
0.891	0.129	0.871	0.764	0.990	0.769
0.361	0.765	0.560	0.809	0.591	0.760
0.819	0.581	0.129	0.794	0.899	0.870
0.270	0.800	0.531	0.750	0.999	0.969

将表 1 中数据代入式(3)中,可得,

$$S = \begin{bmatrix} 0.085 & 0.122 & 0.231 & 0.081 & 0.098 & 0.95 \\ 0.127 & 0.128 & 0.246 & 0.200 & 0.310 & 0.810 \\ 0.891 & 0.129 & 0.871 & 0.764 & 0.990 & 0.769 \\ 0.361 & 0.765 & 0.560 & 0.809 & 0.591 & 0.760 \\ 0.819 & 0.581 & 0.129 & 0.794 & 0.899 & 0.870 \\ 0.270 & 0.800 & 0.531 & 0.750 & 0.999 & 0.969 \end{bmatrix}$$

将矩阵 S 和权重值 μ 代入式(2)中,可得,

$$k_1 = 1.2, k_2 = 1.6, k_3 = 4.9$$

$k_3 = 4.9$ 明显属于异构数据, k_1, k_2 可以视为多源同构数据。

$$y_a = \begin{bmatrix} 0.085 & 0.122 & 0.231 & 0.081 \\ 0.127 & 0.128 & 0.246 & 0.200 \\ 0.891 & 0.129 & 0.871 & 0.764 \\ 0.361 & 0.765 & 0.560 & 0.809 \\ 0.819 & 0.581 & 0.129 & 0.794 \\ 0.270 & 0.800 & 0.531 & 0.750 \end{bmatrix}$$

$$y_b = \begin{bmatrix} 0.098 & 0.95 \\ 0.310 & 0.810 \\ 0.990 & 0.769 \\ 0.591 & 0.760 \\ 0.899 & 0.870 \\ 0.999 & 0.969 \end{bmatrix}$$

$$y = (y_a, y_b)^T$$

$$= \begin{bmatrix} 0.085 & 0.127 & 0.891 & 0.361 & 0.819 & 0.270 \\ 0.122 & 0.128 & 0.129 & 0.765 & 0.581 & 0.800 \\ 0.231 & 0.246 & 0.871 & 0.560 & 0.129 & 0.531 \\ 0.081 & 0.200 & 0.764 & 0.809 & 0.794 & 0.750 \\ 0.098 & 0.310 & 0.990 & 0.591 & 0.899 & 0.999 \\ 0.950 & 0.810 & 0.769 & 0.760 & 0.870 & 0.969 \end{bmatrix}$$

表2 两个时间段的数据成分值均值

工资系统	成绩系统	影响系统	工资系统	成绩系统	影响系统
0.05	0.08	0.085	0.09	0.095	0.10
0.120	0.139	0.184	0.618	0.200	0.08
0.641	0.149	0.140	0.210	0.300	0.20
0.2003	0.09	0.051	0.230	0.201	0.90
0.230	0.12	0.232	0.560	0.640	0.80
0.917	0.60	0.090	0.150	0.150	0.78

由表2所列,6位不同用户分别在2个时间段内请求推送以上3个系统的信息,云推送平台记录下的数据成分均值,构成均值矩阵 μ :

$$\mu = \begin{bmatrix} 0.05 & 0.08 & 0.085 & 0.09 & 0.095 & 0.10 \\ 0.120 & 0.139 & 0.184 & 0.618 & 0.200 & 0.08 \\ 0.641 & 0.149 & 0.140 & 0.210 & 0.300 & 0.20 \\ 0.2003 & 0.09 & 0.051 & 0.230 & 0.201 & 0.90 \\ 0.230 & 0.12 & 0.232 & 0.560 & 0.640 & 0.80 \\ 0.917 & 0.60 & 0.090 & 0.150 & 0.150 & 0.78 \end{bmatrix}$$

MDCP最终确定综合推送方式,根据式(10)得到:

$$\Sigma = P_m \Sigma P_m^T = \begin{bmatrix} 0.05 & 0.08 & 0.08 & 0.09 & 0.09 & 0.10 \\ 0.10 & 0.10 & 0.10 & 0.60 & 0.20 & 0.08 \\ 0.60 & 0.10 & 0.10 & 0.20 & 0.30 & 0.20 \\ 0.20 & 0.09 & 0.05 & 0.20 & 0.20 & 0.90 \\ 0.20 & 0.10 & 0.20 & 0.50 & 0.60 & 0.80 \\ 0.90 & 0.60 & 0.09 & 0.10 & 0.15 & 0.70 \end{bmatrix}$$

将 y, μ, Σ 代入式(11)中,化简可得

$$N = \frac{1}{(2\pi)^{1/2} \cdot (0.35)^{1/2}} \exp\left\{-\frac{1}{2} \cdot 0.35 \cdot \frac{1}{0.15}\right\} = 1.06$$

将 N 带入式(12)中的 N_M, μ, y 同时带入式(12)中,通过偏导化简求和后可以得出,

$$\Sigma l = \begin{bmatrix} 0.038 & 0.050 & 0.500 & 0.020 & 0 & 0 \\ 0.050 & 0.040 & 0.010 & 0 & 0.750 & 0 \\ 0.500 & 0.010 & 0.028 & 0 & 0 & 0.980 \end{bmatrix}$$

由于每次推送部分为更新的数据,后面3列数据为上一次更新数据,因此去除矩阵中后面3列多余数据,得到更新数据 Σd :

$$\Sigma d = \begin{bmatrix} 0.038 & 0.050 & 0.500 \\ 0.050 & 0.040 & 0.010 \\ 0.500 & 0.010 & 0.028 \end{bmatrix}$$

从矩阵中可以看出, Σd 为对称矩阵,所以每次更新是传输的数据量为原来传统推送方式的 $1/2$,大大提高了推送效率。

5 算法有效性分析

为衡量提出的MDCP算法的能力,从平均传输率、通信率、静置时流量、推送留存率来对多源异构数据云推送平台中的算法有效性进行评估。

平均传输率是指数据传输平均的“倍速”数。单倍数传输时,即可记为1倍速,普通推送的平均传输率为10倍速,结果如表3所列。

表3 平均传输率数据表

	MDCP推送	普通推送
平均传输率	100	10
最好值(%)	0.01028	0.02489
最差值(%)	0.01523	0.02489
平均值(%)	0.01308	0.02489

通信率是指单位时间内用户与云推送平台的通信次数,测试用户是否愿意使用该平台进行数据推送,并同时测试了在通信次数多时会不会产生其他问题(结果如图6所示)。传统的推送方式通信率基本不变是由于在推送任务队列消息的整个过程中一直都会向服务器发送请求,而本文提出的云推送模型处于信息收集阶段,随着系统的运行,任务数量增多,优势就逐渐显示出来,即任务数越多,花费的通信量反而变少。

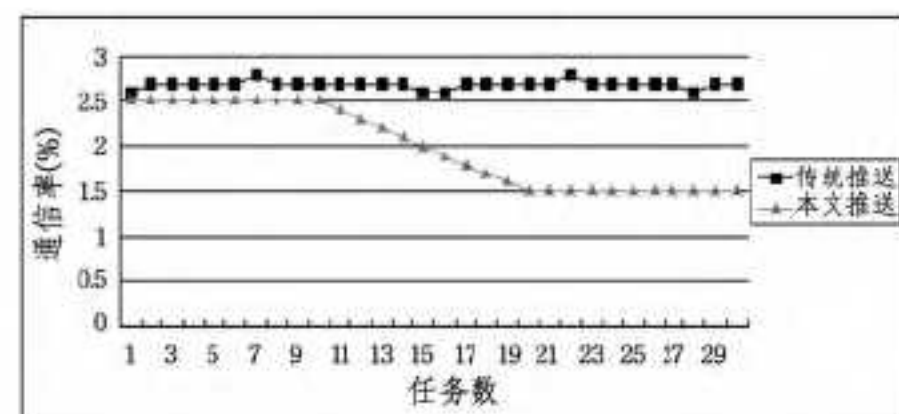


图6 通信率变化图

静置时流量是指手机静置时由于推送而产生的额外流量。测试云推送平台是否会因为通信率的改善而产生大量流量,分别使用云推送平台和传统推送平台进行测试(如表4所列)。实验结果表明,云推送平台在移动设备静置时间较长情况下流量消耗少于传统推送平台。

表4 静置时流量对比

场景	云推送平台	传统推送平台
流量(kB)	实测值	实测值
静置(8小时)	6.8	8.46
静置(12小时)	10.66	12.97
重连次数	0	0

留存率是指在一定时间内,如1~6个星期,用户还留有该推送消息的比例,也能从一定程度反映出该模型对用户的影响。传统推送平台与云推送平台留存率对比图如图7、图8所示。结果表明,几乎每星期云推送平台的留存率都大于传统推送平台留存率的一倍以上,6个星期后传统推送平台的留存率已经不到20%,而云推送平台几乎达到50%。

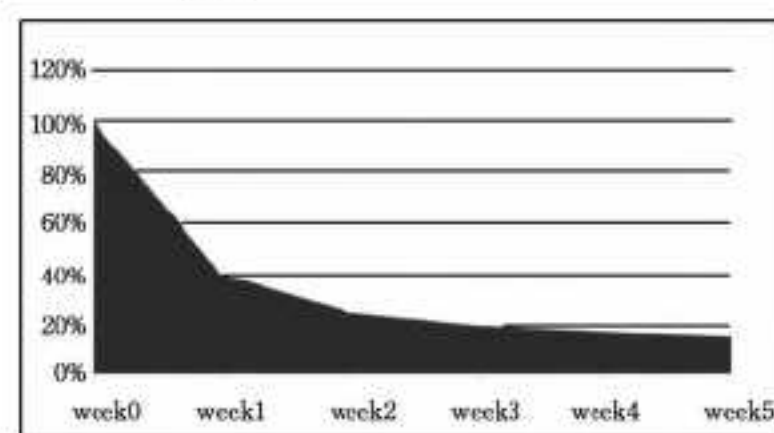


图7 传统推送平台留存率

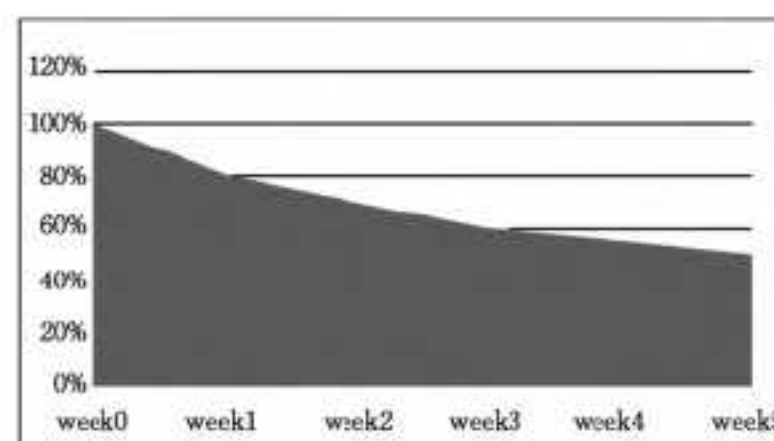


图8 云推送平台留存率

结束语 针对传统的推送方式在推送多源异构数据时遇到的效率低、实时性差等问题,设计了面向多源异构数据的云推送平台来适应云推送环境,并通过MDCP模型解决了多源异构数据推送问题,满足了用户需求。

(下转第567页)

- RSSI Technology [J]. Energy Procedia, 2011(11):1070-1073
- [9] Luo Xiao-wei, O'Brien W J, Julien C L. Comparative evaluation of Received Signal-Strength Index (RSSI) based indoor localization techniques for construction jobsites [J]. Advanced Engineering Informatics, 2010, 25(2):355-363
- [10] Chung Wan-young, Lee B G, Yang C S. 3D virtual viewer on mobile device for wireless sensor network-based RSSI indoor tracking system [J]. Sensors & Actuators: B. Chemical, 2009, 140(1):35-42
- [11] Li Xiao-feng, Chen Liang-feng, Wang Jian-ping, et al. A Novel Method to Improve the Accuracy of the RSSI Techniques Based on RSSI-D [J]. Journal of Networks, 2014, 9(12):3400-3406
- [12] Huang Bo-wei, Liu Hong-li, Zhan Jie. A New Algorithm of Mobile Node Localization Based on RSSI [J]. Wireless Engineering and Technology, 2011, 2(2):112-117
- [13] Wu Duan-po, Jin Xin-yu, Jiang Lu-rong. Analysis of handoff algorithm-based on Doppler Effect and RSSI measurements in GSM-R network [J]. Journal of the Chinese Institute of Engineers, 2014, 37(3):325-331
- [14] Abou-El-Hossein K, Olufayo O, Mkoko Z. Diamond tool wear during ultra-high precision machining of rapidly solidified aluminium RSA 905 [J]. Wear, 2013, 302(1/2):1105-1112
- [15] Toyoda N, Houzumi S, Mashita T, et al. Novel precision machining using gas cluster ion beams [J]. Surface & Coatings Technology, 2007, 201(19):8624-8627
- [16] Fang Su-ping, Liu Yong-sheng, Otsubo H, et al. An automatic leveling method for the stage of precision machining center [J]. the International Journal of Advanced Manufacturing Technology, 2012, 61(1):303-309
- [17] Saied Y B, Oliveureau A, Zeghlache D, et al. Trust management system design for the Internet of Things: A context-aware and multi-service approach [J]. Computers & Security, 2013(39):351-365
- [18] Ondemir O, Gupta S M. Quality management in product recovery using the Internet of Things: An optimization approach [J]. Computers in Industry, 2013, 63(3):491-504
- [19] Kyriazis D, Smart T V. Autonomous and Reliable Internet of Things [J]. Procedia Computer Science, 2013(21):442-448
- [20] Hernandez-Bravo A, Carretero J. Approach to Manage Complexity in Internet of Things [J]. Procedia Computer Science, 2014(36):210-217
- [21] Sheng Q Z, Zeadally S, Luo Zong-wei, et al. Ubiquitous RFID: Where are we? [J]. Information Systems Frontiers, 2010, 12(5):485-490
- [22] Wu Tin-yu, Liaw G H, Huang S W, et al. A GA-based mobile RFID localization scheme for internet of things [J]. Personal and Ubiquitous Computing, 2012, 16(3):245-258
- [23] King R. Planning and managing disruptive business models for RFID-startups-The disruptor's arrow of time [J]. Communications in Mobile Computing, 2012, 1(1):1-5
- [24] Chu He-qun, Wu Guang-min, Chen Jian-ming, et al. Design and simulation of self-powered radio frequency identification (RFID) tags for mobile temperature monitoring [J]. Science China Technological Sciences, 2013, 56(1):1-7
- [25] 袁巨龙, 张飞虎, 戴一帆. 超精密加工领域科学技术发展研究 [J]. 机械工程学报, 2010(15):161-177
- [26] 石欣, 印爱民, 陈曦. 基于 RSSI 的多维标度室内定位算法 [J]. 仪器仪表学报, 2014(2):261-268
- [27] 詹杰, 刘宏立, 刘述钢. 基于 RSSI 的动态权重定位算法研究 [J]. 电子学报, 2011(01):82-89
- [28] 曾孝平, 王峰, 熊东. 基于 UHF RFID 的三维位置感知算法设计 [J]. 世界科技研究与发展, 2013(02):196-200
- [29] 吴桂平, 虞慧群, 范贵生. 一种基于 RSSI 距离比的传感器节点定位算法 [J]. 华东理工大学学报(自然科学版), 2013(05):596-600
- [30] 姚锡凡, 练肇通, 杨屹. 智慧制造——面向未来互联网的人机物协同制造新模式 [J]. 计算机集成制造系统, 2014(06):1490-1498

(上接第 537 页)

然而,该平台能否满足所有的用户需求,能否使大量用户使用,还需要进行验证,云推送平台本身的性能提升以及各种演化方式也将是下一步的研究内容。相信随着这些关键问题的攻破,面向多源异构数据的云推送平台将为用户带来更好的推送体验。

参 考 文 献

- [1] Hu Meng-lan, Yang Wang, Bharadwaj V, et al. Practical Resource Provisioning and Caching with Dynamic Resilience for Cloud-Based Content Distribution Networks [J]. IEEE Transactions on Parallel & Distributed Systems, 2013, 25(8):2169-2179
- [2] 许富龙, 刘明, 龚海刚, 等. 延迟容忍传感器网络基于相对距离的数据传输 [J]. 软件学报, 2010, 21(3):490-504
- [3] Zhen L I, Yang F C, Sen S U. Fuzzy Multi-Attribute Decision Making-Based Algorithm for Semantic Web Service Composition [J]. Journal of Software, 2009, 20(3):583-596
- [4] 蒋乾悦, 张亚英. 基于模糊综合决策的服务器推送方法 [J]. 计算机科学, 2014, 41(5):86-90
- [5] 刘鑫, 陈伟. 基于 AJAX 和 Server Push 的 Web 树组件设计与实现 [J]. 计算机工程与设计, 2009, 30(3):679-682
- [6] Tan Zhi-peng, Zhou Wei, Sun Jian-liang, et al. An improvement of static subtree partitioning in metadata server cluster [J]. International Journal of Distributed Sensor Networks, 2012, 2012(1550-1329):1129-1132
- [7] Armbrust M, Fox A, Griffith R, et al. A view of cloud computing [J]. Commun. ACM(CACM), 2010, 53(4):50-58
- [8] 叶可江, 吴朝晖, 姜晓红, 等. 虚拟化云计算平台的能耗管理 [J]. 计算机学报, 2012, 35(6):1262-1285
- [9] 陈阳, 怀进鹏, 胡春明. 基于内存混合复制方式的虚拟机在线迁移机制 [J]. 计算机学报, 2011, 34(12):2278-2291
- [10] 冯朝胜, 秦志光, 袁丁. 云数据安全存储技术 [J]. 计算机学报, 2015, 38(1):150-163
- [11] 邵帅, 张宏科, 徐怀松. Sink 轨迹固定传感器网络的高效数据采集机制 [J]. 软件学报, 2010, 21(1):147-162
- [12] 丁滢, 王怀民, 史佩昌, 等. 可信云服务 [J]. 计算机学报, 2015, 38(1):133-149
- [13] 胡宁, 朱培栋, 邹鹏. 域间路由协同监测中的信息共享机制 [J]. 软件学报, 2011, 22(3):481-494
- [14] 谷峪, 于戈, 李晓静, 等. 基于动态概率路径事件模型的 RFID 数据填补算法 [J]. 软件学报, 2010, 21(3):438-451
- [15] 黎玲利, 王宏志, 高宏, 等. XML 数据流上 Top-K 关键字查询处理 [J]. 软件学报, 2012, 23(6):1561-1577
- [16] MQTT[EB/OL]. <http://mqtt.org>, 2015