

公共地图服务的群体用户访问行为 时序特征模型及预测

吴华意^{1,2} 李锐^{1,2} 周振¹ 蒋捷³ 桂志鹏⁴

1 武汉大学测绘遥感信息工程国家重点实验室,湖北 武汉,430079

2 地球空间信息技术协同创新中心,湖北 武汉,430079

3 国家基础地理信息中心,北京,100830

4 武汉大学遥感信息工程学院,湖北 武汉,430079

摘要:群体用户对公共地图服务的访问行为具有社会性,存在着一定的群体访问行为模式。该模式具有高强度的访问聚集性与突发性,且决定着公共地图服务对云计算资源的需求。如何有效地表达和捕捉群体用户访问聚集性及其访问强度的时序变化特征,进行准确的公共地图服务负载预测,是实现按需选择和调度云计算资源,应对海量用户并发服务挑战的关键。本文基于海量的公共地图服务用户访问日志和时间序列聚类方法,建立了群体用户访问到达行为的时序分布模型;考虑降低负载预测复杂度的同时,利用访问强度具有多峰值、变强度以及周期性的特点,分割访问到达率在一个周期内时间序列上的模式区间,实现访问强度时序聚类划分的最优;基于各访问模式区间不同的访问到达概率密度的分布,提出了基于累积概率分布的时间序列平滑预测服务负载方法,该预测方法的算法复杂度低,且所需的先验数据量小。实验证明,本文提出的基于时序的群体用户访问到达率最优分割方法及其预测方法可以以较高的准确率预测服务负载。该方法在应对海量用户并发访问挑战的同时,可提高云计算资源的利用效率,解决公共地图服务质量与服务成本的平衡性问题。

关键词:公共地图服务;访问到达率;时间序列聚类;Fisher 最优分割法;资源调度

中图法分类号:P208;TP311 文献标志码:A

地理空间信息系统向着网络化、规模化、虚拟化、服务化和时空化的方向发展,以 Google Maps/Google Earth、World Wind、Bing Map、天地图等为代表的公共地图服务平台为用户带来了全新的体验^[1]。随着 Internet 的快速发展和移动网络的不断普及,公共地图服务用户与日俱增,导致其实时在线操作达到了百万级。高强度的访问及访问的集中性带来了一系列的问题,如服务器过载、访问延迟、通信错误增多等。服务器有限的能力需通过扩展硬件来满足其数据量和用户量日益增长的需要,且其服务维护使用成本逐渐增高。云计算是一种新一代的计算模式,其计算资源的虚拟化可使计算资源按需提供服务,并且可以动态地扩展计算资源。将公共地图服务迁移到云中,可满足其对计算资源日益剧增的需求。但云

计算资源的分配方式是按需分配,特别是一些能提供较高服务性能的商业云,将收取一定的资源租赁费用。因此,如何有效地选择和利用云服务资源,控制公共地图服务成本,在投入和质量需求之间到达最佳的平衡点,对于拥有海量用户并对地理空间服务质量有较高要求的公共地图服务来说尤其重要^[2-5]。而此方面的研究多聚集在云服务计算能力和云服务可扩展性评估^[6]、云服务资源的地理位置部署^[7]、云服务的计费模式^[8]、云服务就绪度^[9]等方面,极少从云服务支撑的公共地图服务应用的自身特点与需求选择云服务资源。

用户的访问行为直接影响着公共地图服务对计算资源的需求。而海量用户对公共地图服务的访问行为具有社会性,存在着一定的群体模式。通过分析和挖掘群体用户的访问行为模式,可针

对性地解决服务质量与服务成本的相关性问题^[10]。且公共地图服务拥有大量的日志数据,提供了丰富的用户动态行为信息。近期,不断有研究人员利用不同的公共地图服务日志数据进行用户访问行为的统计分析与仿真,以期建立合适的用户访问模型。并且许多研究学者认为研究和挖掘大规模用户在地理信息服务系统中漫游时所表现出来的社会学规律和交互模式,为系统的性能优化、提高用户漫游流畅体验提供非常重要的指引作用^[11]。Fisher^[12] 和 Talagala^[13] 分别对地理信息的访问请求频率进行统计,提出数字地球中影像数据的访问请求服从社会学中的幂律分布。王浩等^[14]通过实验分析扩展了其结论,得出影像数据瓦片请求符合幂律分布中的 Zipf-like 分布。幂律分布全局地反映了地理信息访问中的一种时间长期累积访问特征,但不能实时反映用户访问模式的变化。Park 等^[15]认为,用户的访问模式取决于地理数据的空间位置性。Yang 等^[16]提出用户访问公共地图服务平台时带给服务器的负载具有潮汐特征。这些都说明了因地理数据存在时空属性,所以用户的访问请求同样存在时空聚集性。此访问特性对服务负载,如流量波动、服务队列长度等公共地图服务性能产生了重要的影响,同时也决定着公共地图服务对计算资源的需求。与此同时,公共地图服务中用户的访问到达在时间上是随机的。因此,如何有效地表达和捕捉用户访问的聚集性及其访问强度随时间的变化,进行准确、快速的服务负载预测,解决由于用户访问的集中性和突发性所带来的服务器性能问题,是其难点与根本。本文分析了大量的非结构化的用户访问公共地图服务历史记录,基于时间序列聚类方法发现群体用户访问强度随时间变化的规律;利用群体用户访问强度具有多峰值、变强度以及周期性的特点,将用户访问到达率的时间序列分割成若干个相对较短但不重叠的子序列访问模式区间,更精确地匹配群体用户访问在时间上的聚集性和突发性特征;并对各个时序区间分别预测访问负载,提高预测准确度的同时,实现一种简单有效且只需少量的先验历史数据的负载预测方法。此方法可用于指导公共地图服务资源的需求和规划,以提高资源的利用效率,并满足用户对公共地图服务质量的需求。

1 群体用户访问行为的时序建模

若将一个用户访问公共地图服务的请求到达

看成一个随机点,则这是一个源源不断出现的随机过程。在这过程中,任一时间段内达到的用户请求数也是随机的,为公共地图服务带来随机的访问强度与负载。单位时间内平均到达的用户请求数可称为公共地图服务的访问到达率 λ ,而群体用户访问公共地图服务的访问强度通常可用访问到达率来表达,且可用于公共地图服务中服务器负载的预测,它属于时间序列预测的范畴。对于访问量稳定的网络系统,呈现单峰值特征,可使用回归分析、灰度预测、拓扑预测、线性网络预测、BP 神经网络预测、Karkov 等方法预测用户访问行为的趋势。而对于类似公共地图服务的网络系统,群体用户的访问在时序上呈现非均质特征,用户到达率所构成的时间序列具有多峰值和周期性。因此,需针对群体用户访问特征,阶段性、周期性地建立准确的访问负载预测模型。同时也需考虑应对群体用户密集访问的预测方法对公共地图服务系统资源占用最小,保证预测精度高的同时,算法时间复杂度小。

1.1 样本数据及预处理

本文采用公共地理信息服务平台天地图服务器端采集的用户访问日志数据作为本文研究的基础数据。天地图是国家测绘地理信息局建设的地理信息综合服务网站。它是数字中国的重要组成部分,是国家地理信息公共服务平台的公众版。访问该网站的用户具有广域的地域性,日志数据以非结构化形式记录在不同的集群服务器和不同的文件中。选取的样本数据集为 2014 年 2 月海量用户访问的日志文件,数据存于上百个文件中,每日文件大小为 20 G。本文建立 MongoDB 分布式文件存储的数据库,基于 MapReduce 并行计算框架对数据进行格式化处理,建立高效的索引和存储,形成每日近 9 千万条数据,包括时间、用户 IP、请求的数据经纬度、层数、数据来源等。因天地图的访问中存在人类用户(普通的访问用户)和机器用户(程序下载数据资源用户),为了真实可靠地反映人类用户的群体行为,在数据预处理中根据机器用户的特征,如单个用户访问量巨大,或单个用户连续爬取一定经纬度范围的数据,或连续大量访问某个区域数据等,剔除了机器用户的访问记录。为了保证时序分析在时间上和访问特征聚类的精确度,以 10 min 为间隔,统计天地图用户访问的平均到达率用于样本分析。

1.2 群体用户访问的时序分布模型

对于公共地图服务平台而言,用户访问行为具有一定的模式性,即在不同的时间区间,用户到

达率 λ 的值是变化的。从时间序列的角度分析,若将一个周期的时间分为若干等长时段,每个时段的请求数可以构成一个时间序列 $L(S, t)$:

$$L(S, t) = \{S(t_1), S(t_2), \dots, S(t_n)\} \quad (1)$$

式中, S 和 t 分别是用户到达时间序列 $L(S, t)$ 的到达率因子和时间因子; $S(t_i)$ 表示 t_i 时间段内到达的用户请求数。那么对于时间序列 $L(S, t)$ 而言,其均值的期望值表达式为:

$$E(S_i) = \lambda(t) * T \quad (2)$$

式中, T 为每个时段的时间长度; $\lambda(t)$ 表示 t 时段的用户平均到达率。由于时间序列的均值与时间 t 相关,则该序列并非一个平稳的过程。从实际的群体用户访问记录分析来看,一天中用户的访问存在短时期的“黄金时段”,用户到达时间分布曲线在不同的时间段呈现出不同的分布模式,即用户到达率的概率分布具有多峰值的特性。同时,由于用户每天的访问行为模式具有相似性,则访问强度的变化具有一定的周期性规律。若对一

个周期内的用户访问行为在时序上进行模式划分,则在同一个访问模式下的时间区间内,可以认为用户的到达率 $\lambda(t)$ 变化很小。如此, $\lambda(t)$ 可以表示为具有 k 种模式的一个分段常值函数模型:

$$\lambda(t) = \begin{cases} \lambda_1, & 0 < t \leq t_1 \\ \vdots \\ \lambda_k, & t_{k-1} < t \leq t_k \end{cases} \quad (3)$$

依据上述群体用户访问的时序方法,为样本数据建立访问平均到达率 $\lambda(t)$ 的时序分析图。图 1 是统计样本数据后得到的 2014 年 2 月 1 日到 2 月 28 日的平均到达率变化图。从图 1 可以看出,公共地图服务平台每天的用户到达率 $\lambda(t)$ 的时序变化规律相似,且以天为周期单位。在这种周期分类上,用户工作日(暖色曲线)的访问行为模式与节假日(冷色曲线)的访问行为模式不同,体现在白日时间的高峰期访问到达率明显不同。因此,可先粗略地将用户的访问行为分为工作日模式与节假日模式。

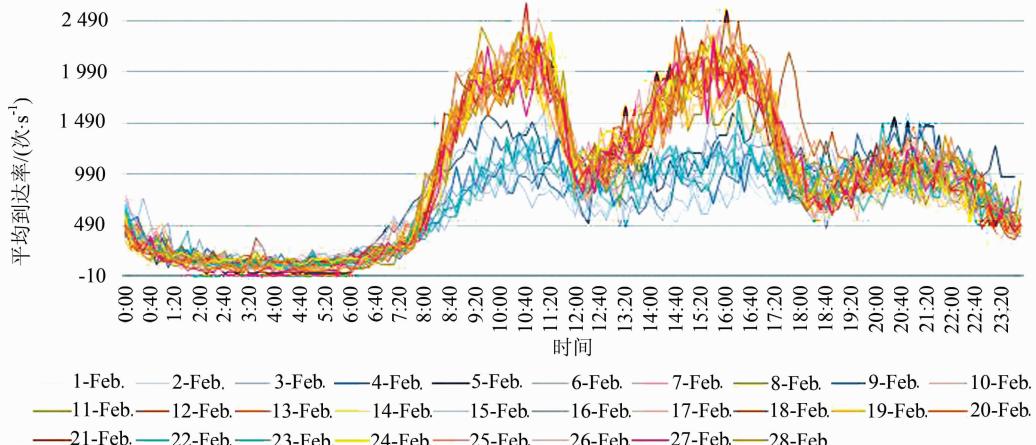


图 1 工作日与节假日用户到达率变化情况

Fig. 1 User Arrival Rate in Working Days and Holidays

图 1 显示出在一个周期内,群体用户访问到达率分布曲线在不同的时间段也呈现出不同的分布模式,其访问到达率的分布具有明显的区别。因此,若要实现群体用户访问行为的精确预测,则需对工作日模式与节假日模式进行进一步时序聚类划分,如式(3)所示。

1.3 群体用户访问到达率时间序列的最优分割

若要利用群体用户访问的模式性及周期性进行公共地图服务负载预测,首先需分割和挖掘群体用户到达率在时间序列中不同的访问模式,以获得式(3)中各访问模式的时间区间 t_i 。研究可知,对群体用户访问到达率聚类是模式分割的有效方法。图 1 显示出群体用户访问模式在时间上是连续的,因此本文基于 Fisher 最优分割方法进

行有序聚类,进而有效地发现用户访问到达率在时间序列上的模式区间,即对一个周期的时间序列 $L(S, t)$ 进行合理的划分,得到不同访问模式所在的时间区间集合 $T = \{T_1, T_2, \dots, T_k\}$ 。

Fisher 最优分割算法的思想是:当数据按一定的时间顺序排列,为了探索事物发展的周期性或旋回性,并合理地划分呈现不同模式的阶段,则需要进行最优分割。将一定数量的、规律性的、接近的若干相邻数据归为一个阶段;而相邻两个阶段的数据呈现的模式出现明显的差异,前一个阶段的最后一项数据与下一阶段的头条数据之间便为模式分割的界限。

本文研究需要从用户到达时间序列 $L(S, t)$ 中发现用户访问的模式 $T = \{T_1, T_2, \dots, T_k\}$,各模式

T_i ($1 \leq i \leq k$) 在时间上相邻接, 但用户平均到达率 $\lambda(t_i)$ 却不同, 且各模式适用于群体用户访问周期变化。图 1 显示, 公共地图服务用户在工作日与节假日的访问模式明显不同, 所以对此模式的分割与模式挖掘分为工作日模式与节假日模式。

本文以工作日模式为例, 建立多变量的、有序的表达方法实现群体用户访问到达率的最优分割。其中, 因短时间内的访问到达率相似, 以每 10 min 为时间间隔聚类得到一个周期内 144 个平均到达率, 建立有序的访问到达率时间序列表达, 如式(4)所示。将 2014 年 2 月 1 日至 2 月 21 日工作日的访问到达率作为训练样本集, 如表 1 所示。

$$L(S, t)_{d_m} = \{S(t_1), S(t_2), \dots, S(t_i), \dots, S(t_n)\} \quad (i = 1 \sim 144) \quad (4)$$

表 1 2014 年 2 月 1 日至 21 日工作日的访问到达率

Tab. 1 User Arrival Rate of Working Days
on 1-21, Feb. 2014

时间段	t	$d_{2,7}$	$d_{2,8}$...	$D_{2,20}$	$D_{2,21}$
0:00~0:10	t_1	204	268	...	521	424
0:10~0:20	t_2	207	192	...	454	304
0:20~0:30	t_3	162	258	...	263	234
...
15:20~15:30	t_{93}	2 187	1 925	...	1 840	2 288
15:30~15:40	t_{94}	2 072	1 598	...	1 993	2 183
...
23:40~23:50	t_{143}	306	358	...	600	420
23:50~24:00	t_{144}	440	419	...	505	492

将式(4)展开形成多个周期内访问到达率时间序列的平均到达率矩阵:

$$L(S, t) = \begin{bmatrix} S(t_1 d_1), S(t_1 d_2), \dots, S(t_1 d_m) \\ S(t_2 d_1), S(t_2 d_2), \dots, S(t_2 d_m) \\ \vdots & \vdots \\ S(t_n d_1), S(t_n d_2), \dots, S(t_n d_m) \end{bmatrix} \quad (5)$$

基于 Fisher 最优分割法对矩阵 $L(S, t)$ 进行最优分割^[17], 将值相似且相邻的行向量(访问到达率向量) $S(t_i), S(t_{i+1}), \dots, S(t_n)$ 聚为一类, 形成 K 类访问模式以及各模式所在的时间区间集合 $\{T_1, T_2, \dots, T_k\}$ 。方法与步骤如下。

1) 计算访问到达率向量的直径

设分割后访问到达率向量的某一访问模式包含时序 $\{S_{(i)}, S_{(i+1)}, \dots, S_{(j)}\} = \{S(t_i), S(t_{i+1}), \dots, S(t_j)\}$ ($j > i$), 记为 $G = \{i, i+1, \dots, j\}$, 则该访问模式的向量均值 \tilde{S}_G 为:

$$\tilde{S}_G = \frac{1}{j-i+1} \sum_{h=i}^j S_{(h)} \quad (6)$$

设模式内访问到达率向量的直径为访问到达

率向量集的离差平方和 $D(i, j)$, 则有:

$$D(i, j) = \sum_{h=i}^j (S_{(h)} - \tilde{S}_G)(S_{(h)} - \tilde{S}_G)^T \quad (7)$$

2) 计算访问到达率向量分割损失函数

设 $b(n, K)$ 为将有序的 n 个访问到达率向量被分割成 K 个访问模式的一种分类方法, 则 $b(n, K)$ 可表达为:

$$\begin{cases} G_1 = \{i_1, i_1 + 1, \dots, i_2 - 1\} \\ G_2 = \{i_2, i_2 + 1, \dots, i_3 - 1\} \\ \vdots \\ G_k = \{i_k, i_k + 1, \dots, n\} \end{cases} \quad (8)$$

式中, 分点 $1 = i_1 < i_2 < \dots < i_k < n = i_{k+1} - 1$ ($i_{k+1} = n+1$); $b(n, K)$ 损失函数为:

$$\Gamma[b(n, K)] = \sum_{e=1}^K D(i_e, i_{e+1} - 1) \quad (9)$$

当 n, K 固定时, $\Gamma[b(n, K)]$ 越小, 即各类模式的离差平方和越小, 则模式的分类是合理的。因此在此寻找一种合适的 $b(n, K)$, 使其损失函数值 Γ 最小。

记 $P(n, K)$ 为使 Γ 达到最小的分类方法。利用递推式(10)求出 K 取不同值时, 各分类方法的最小损失函数以及各模式之间的分割点:

$$\begin{cases} \Gamma[P(n, 2)] = \min_{2 \leq j \leq n} \{D(1, j-1) + D(j, n)\} \\ \Gamma[P(n, K)] = \min_{K \leq j \leq n} \{\Gamma[P(j-1, K-1)] + D(j, n)\} \end{cases} \quad (10)$$

3) 计算访问到达率向量分割的最优解

若访问到达率向量分割的次数 K ($1 < K < n$) 已知, 求使损失函数值最小的分割方法 $P(n, K)$ 。首先寻求分割点 j_k , 使其满足:

$$\Gamma[P(n, K)] = \Gamma[P(j_k - 1, K - 1)] + D(j_k, n) \quad (11)$$

计算得到第 k 类 $G_k = \{i_k, i_k + 1, \dots, n\}$ 。然后寻求 j_{k-1} , 使它满足式(12), 得到第 $k-1$ 类 $G_{k-1} = \{i_{k-1}, i_{k-1} + 1, \dots, j_k - 1\}$ 。

$$\Gamma[P(j_{k-1}, K - 1)] = \Gamma[P(j_{k-1} - 1, K - 2)] + D(j_{k-1}, j_k - 1) \quad (12)$$

根据递推式(10)可以得到所有类 G_1, G_2, \dots, G_k , 则所求的最优分割方法为:

$$P(n, K) = \{G_1, G_2, \dots, G_k\} \quad (13)$$

如此计算得出在 K 次分割下, 各个模式中时间序列的分割点 $\{j_1, j_2, \dots, j_k\}$ 。

训练实验中, 对训练样本中的访问到达率数据进行 2~15 次最优分割, 分别得到各次分割下的最小损失函数值, 如图 2 所示。

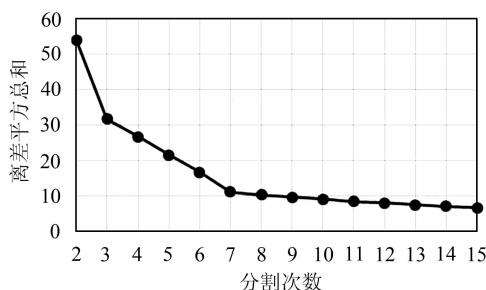


图 2 最小损失函数值变化图

Fig. 2 Variation of Minimum Loss Function Value

K 值的确定依据 Fisher 原理: 在 Fisher 最优分割过程中, 最小损失函数值 Γ 随分割数 K 的增加而减少。而当分割数增加到某一数值后, 最小损失函数值曲线将急剧变缓, 达到一定的平衡, 此时的 K 值为最佳分割值。本文基于非负斜率方法确定 K 值, 如式(14)所示:

$$\beta(K) = \left| \frac{P(n, K) - P(n, K-1)}{K - (K-1)} \right| \quad (14)$$

当 $\beta(K)$ 较大时, 表示分 K 类优于分 $K-1$ 类; 当 $\beta(K)$ 接近于 0 时, 此时 K 为合适的取值。由图 2 与式(14)可计算出 $K=7$ 时为最佳分割取值。且本文对访问到达规律的研究是为了准确且简单有效地预测公共地图服务的访问负载。若群体用户访问到达率在一个周期内的时序访问模式区间过多, 则预测所需的先验数据量越大, 预测的复杂度及所需的计算存储容量也将随之递增; 模式区间过少, 则影响预测精度。因此, 综合考虑最优分割原理与访问负载预测的可行性, 分割方法 $P(n, 7)$ 是合适的。本文在实验中通过训练样本集得到了各访问模式所在时间区间的集合 $\{T_1, T_2, \dots, T_7\}$, 如表 2 所示。表 2 显示, K 值为 7 时, 各访问模式的离差平方和值较小。

表 2 各访问模式所在时间区间的集合

Tab. 2 Time Sets of Every Access Mode

T_k	时间段	访问模式离差平方和
T_1	01:10~07:50	2.027 1
T_2	07:50~08:50	0.677 6
T_3	08:50~11:40	2.189 8
T_4	11:40~14:10	1.502 4
T_5	14:10~17:30	2.231 5
T_6	17:30~23:00	2.007 1
T_7	23:00~01:10	0.557 8

2 基于累积概率分布的时间序列平滑预测访问负载方法

2.1 访问到达率的累积概率分布及其时间序列分析

群体用户访问到达率变化的周期性及模式性

为访问到达率的预测提供了可靠的条件。§ 1.3 中对群体用户访问达到率时序的最优分割可以很好地对不同时间段内具有不同强度特征的访问达到率进行有效的划分, 是实现准确预测访问负载的基础。而在此之上挖掘不同时间序列模式下的访问到达率的概率分布是准确预测访问负载的关键。本节为不同时序下的访问到达模式建立了一种累积概率分布, 表达该模式下的访问到达概率密度的分布; 并基于不同访问模式下的累积概率分布和基于时间序列平滑预测方法, 对不同访问模式下的访问负载进行预测, 实现基于少量的先验历史数据即可快速准确地预测。

定义 1 访问到达率的累积概率分布函数为访问到达率概率密度函数的积分, 即为访问到达率小于或等于某个数值的概率。本文根据 § 1.3 得出的用户访问行为模式时序的划分方法, 对不同访问模式下的用户到达率分布进行累积概率分布分析。各模式下的访问到达率的累积概率分布函数为:

$$F(x) = P(\lambda(t) \leq x) = \int_{-\infty}^x f_k(x) dx, t \in T_k \quad (15)$$

根据式(15), 当 $t \in T_k (k=1, \dots, 7)$ 时, 用户到达率 $\lambda(t) \leq x$ 的概率为 $P(\lambda(t) \leq x)$ 。各模式访问到达率的累积概率分布如图 3 所示。

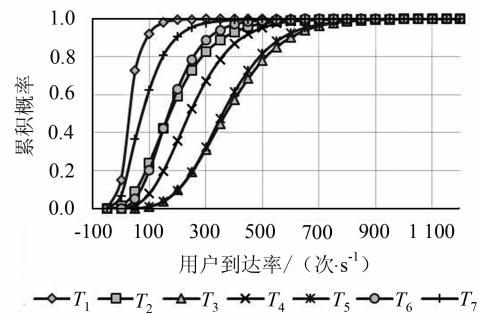


图 3 各模式访问到达率的累积概率分布

Fig. 3 Cumulative Probability Distribution of the Access Arrival Rate

访问到达率的预测是为了寻求不同时间区间的访问峰值, 更好地进行云计算资源的利用与分配。在此, 令 $F(x)=0.99$, 可以求得不同访问模式下的访问到达率峰值 $\lambda_k, t \in T_k (k=1, \dots, 7)$, 即当 $t \in T_k$ 时, $\lambda(t) \leq \lambda_k$ 的概率为 0.99。 λ_k 可以用于衡量 T_k 访问模式区间内群体用户的访问强度与负载。

定义一个周期 d_u 内不同访问模式下的用户访问到达率参数 λ_k 为:

$$\lambda_k = \lambda_{(d_u, k)} \quad (k=1, \dots, 7; u=1, \dots, N) \quad (16)$$

根据式(16),对样本数据工作日访问到达率进行分模式累积概率分布分析,得到 λ_k 的时序分布如图 4 所示。

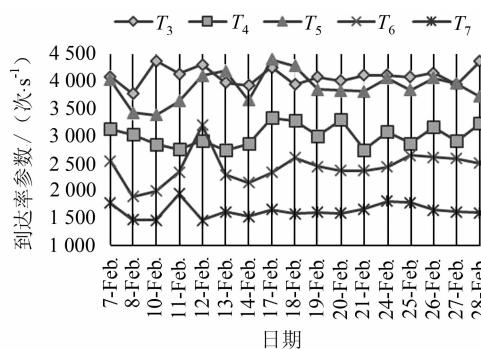


图 4 工作日分模式下的到达率参数时序分布

Fig. 4 Variation of Arrival Rate Parameters in Working Days

从图 4 可以看出,各周期同一访问模式下的到达率参数 λ_k 波动较小,有一定的相关性和稳定性。说明 § 1.3 中访问模式的分割适应并匹配周期的变化体现了群体用户访问的一种周期性的规律。基于 λ_k 的这种相对稳定的变化,可以对 λ_k 进行平滑预测。

2.2 群体用户访问负载的平滑预测

群体用户访问量变化的规律具有周期性,且同一访问模式下访问到达率参数的分布具有一定的稳定性。根据此群体用户的访问特性,本文基于一次移动平均法对用户的访问到达率参数 λ_k 进行预测。

预测方法的思想为:对同一周期内的不同访问模式 T_k 分别预测。收集各访问模式 N 个周期的观察值,即 N 个周期内 T_k 模式访问到达率参数 $\lambda_{(d_u, k)}$,计算该组观察值的均值,利用这一均值预测下一周期该访问模式下的访问到达率。

移动平均法有两种极端情况:① 在移动平均

值的计算中包括过去观察值的实际个数 $N=1$,此时利用最新的观察值作为下一期的预测值;② $N=N'$,这时利用全部 N' 个观察值的算术平均值作为预测值。

当数据的随机因素较大时,宜选用较大的 N ,较大限度地平滑由随机性所带来的严重偏差;反之,当数据的随机因素较小时,宜选用较小的 N ,这有利于跟踪数据的变化,并且预测值滞后的期数也少。

设到达率参数时间序列为 $\lambda_{(d_1, k)}, \lambda_{(d_2, k)}, \dots, \lambda_{(d_{N+1}, k)}$,则对周期 d_{N+1} 的 T_k 模式访问到达率参数 $\lambda_{(d_{N+1}, k)}$ 移动平均法预测可以表示为:

$$F_{(d_{N+1}, k)} = (\lambda_{(d_1, k)} + \lambda_{(d_2, k)} + \dots + \lambda_{(d_N, k)}) / N = \frac{1}{N} \sum_{u=1}^N \lambda_{(d_u, k)} \quad (17)$$

式中, $\lambda_{(d_N, k)}$ 为最新观察值; $F_{(d_{N+1}, k)}$ 为第 d_{N+1} 周期的预测值。

由移动平均法计算公式可以看出,每一新预测值是对前一移动平均预测值的修正, N 越大, 平滑效果越好。

该预测方法的优点是计算量小,其计算复杂度为 $O(1)$,且移动平均能较好地反映时间序列的趋势及其变化。其缺点是计算移动平均值必须具有 N 个过去观察值,当 N 较大或时序访问模式个数 K 值较大时,计算过程中需要使用大量的先验数据,这些都需要大量的计算资源和存储资源。本文提出的最优分割中, K 值为 7,使用的先验数据量相对很小。因此,群体用户访问负载的平滑预测方法是否会增加系统的消耗取决于 N 值。

2.3 预测结果分析

基于 2014 年 2 月所有工作的用户访问日志数据,对各访问模式下的访问到达率进行预测,取 $N=2, \dots, 10$ 。预测准确率和均方误差如表 3 所示。

表 3 预测准确率和均方误差

Tab. 3 Prediction Accuracy Rate and Mean Square Error

N	T_1		T_2		T_3		T_4		T_5		T_6		T_7	
	准确率/%	均方误差	准确率/%	均方误差	准确率/%	均方误差	准确率/%	均方误差	准确率/%	均方误差	准确率/%	均方误差	准确率/%	均方误差
2	89.33	4.62	92.26	9.92	96.37	7.81	93.77	9.21	92.80	13.13	88.79	14.97	93.07	6.61
3	90.55	4.63	93.14	10.00	96.74	6.58	93.32	9.73	93.42	12.13	89.47	14.50	93.16	5.80
4	92.31	3.64	93.88	9.95	96.71	6.52	93.90	9.73	93.44	12.03	92.13	12.42	95.15	4.16
5	92.70	3.46	92.95	9.85	97.36	5.46	92.76	10.51	94.08	11.31	96.24	4.59	95.60	3.54
6	92.40	3.60	93.66	9.46	96.99	6.03	93.12	10.09	94.13	11.30	95.57	5.46	95.33	3.72
7	93.39	3.25	94.01	9.72	97.34	5.43	92.73	10.15	94.72	11.09	95.26	5.89	95.70	3.73
8	94.21	3.11	93.27	10.32	97.48	5.54	93.78	8.57	96.17	7.68	94.79	6.20	95.77	3.89
9	94.22	3.12	92.77	10.96	98.03	4.90	94.62	7.39	97.32	5.40	96.35	4.62	95.93	3.84
10	93.56	3.32	92.94	11.21	97.78	5.29	94.04	7.72	97.24	5.21	96.65	4.40	95.2	4.02

基于本文提出的方法进行访问负载预测时,移动的项数 N 越大,对原数列波动的曲线修匀得越光滑,便越能显示出各个访问模式下访问负载的一种周期性发展趋势,但 N 越大,所需的先验数据越多,系统消耗也越大。综合预测准确率和均方误差值,当 $N=9$ 时,平滑预测效果较好,预测准确率较高,在 92.77~98.03 之间,即 $N=9$ 时即可体现出群体用户访问的周期性规律。这进一步说明了本文提出的基于时序的访问模式分割能准确地体现访问强度变化的规律性和区间性,是公共地图服务中访问负载预测准确率较高,且需要少量的预测先验数据的关键和基础。

2.4 群体用户访问负载模型应用的可行性验证

群体用户访问达到率决定着公共地图服务平台承受的负载,影响着其资源需求和云中资源分配,对云的弹性服务研究具有指导意义^[18,19]。基于上述建立的时间序列的访问负载时序模型与预测方法,在云仿真平台 Cloudsim 上初步验证了群体用户访问负载预测方法应用于云计算资源分配的可行性与有效性。实验中采用日志驱动仿真实验,用离散事件模拟机制实现用户的访问及其对访问请求的处理。云中单台虚拟机计算资源的配置如下:型号为 IBM System X, CPU 个数为 2 个,CPU 主频为 2.6 G,核心为 4,内存 64 G,硬盘 2 T,平均响应时间 0.02 s。其中云任务量为预测的 2014 年 2 月 28 日用户访问到达率的分段模式,公共地图服务平台负载均衡上采用轮询方式将计算资源赋予处理用户访问的请求,所有的用户共用 1 个数据中心。因公共地图服务用户在地图中浏览时,若需要保证较好的漫游体验,响应时间则不超过 0.02 s^[20]。因此实验中以平均响应时间 0.02 s 为性能目标,寻求在不同时间段访问强度下所需的计算资源的临界值,即虚拟机个数,如表 4 所示。由此可见,针对不同周期/时段预测的访问强度,云服务商可为公共地图服务平台进行有效的、合适的资源分配,比如计算资源、存储资源、网络带宽等分配,保证公共地图服务质量的同时,高效地按需使用云服务资源,实现云的弹性服务。

表 4 计算资源的临界值

Tab. 4 Critical Value of Computing Resources

T_k	时间段	到达率/s	服务器个数/个
T_1	01:10~07:50	914	2
T_2	07:50~08:50	2 825	6
T_3	08:50~11:40	4 100	9
T_4	11:40~14:10	3 007	7
T_5	14:10~17:30	3 897	8
T_6	17:30~23:00	2 413	5
T_7	23:00~01:10	1 625	4

3 结语

公共地图服务的群体用户访问具有周期性、时序性和聚集性,此特征有利于其服务性能和系统资源分配的优化。本文建立了公共地图服务中群体用户访问行为的时序分布模型,体现出群体用户并发访问的多峰值变强度特征;并基于此模型提出一种准确且简单有效的、阶段性的时序区间模式预测群体用户访问负载的方法。该预测方法依据群体用户访问的聚集性与周期性,并经过验证可较好地满足公共地图服务中群体用户并发访问的性能需求,从而用于云计算资源分配的优化。今后可结合群体用户访问的地域性和时序性,进一步探究群体用户访问中精确的时空规律,结合云资源的时空表达及公共地图服务性能的表达方法,建立精确、高效的云资源时空分配策略。

致谢:感谢国家基础地理信息中心和公共地理信息服务平台天地图提供的帮助与支持。

参 考 文 献

- [1] Li Deren, Sui Haigang, Shan Jie, et al. Discussion on Key Technologies of Geographic National Conditions Monitoring [J]. *Geomatics and Information Science of Wuhan University*, 2012, 37(5): 505-512 (李德仁, 眭海刚, 单杰, 等. 论地理国情监测的技术支撑 [J]. 武汉大学学报·信息科学版, 2012, 37(5): 505-512)
- [2] Wu Huayi, Zhang Hanwu. QoGIS: Concept and Research Framework [J]. *Geomatics and Information Science of Wuhan University*, 2007, 32(5): 385-388 (吴华意, 章汉武. 地理信息服务质量 (QoGIS): 概念和研究框架 [J]. 武汉大学学报·信息科学版, 2007, 32(5): 385-388)
- [3] Zhang Hanwu, Wu Huayi, Hu Yueming, et al. From Quality of Geospatial Data to Quality of Geospatial Information Services [J]. *Geomatics and Information Science of Wuhan University*, 2010, 35(9): 1 104-1 107 (章汉武, 吴华意, 胡月明, 等. 从地理空间数据质量到地理空间信息服务质量 [J]. 武汉大学学报·信息科学版, 2010, 35(9): 1 104-1 107)
- [4] Gong J, Wu H, Zhang T, et al. Geospatial Service Web: Towards Integrated Cyberinfrastructure for GIScience [J]. *Geo-spatial Information Science*, 2012, 15(2): 73-84
- [5] Li Deren. The Geo-spatial Information Science Mission [J]. *Geo-spatial Information Science*, 2012, 15(1): 1-2

[6] Gao J, Pattabhiram P, Bai X, et al. SaaS Performance and Scalability Evaluation in Clouds[C]. 2011 IEEE 6th International Symposium on Service Oriented System Engineering (SOSE), Irvine, CA, 2011

[7] Badger L, Grance T, Patt-Corner R, et al. Cloud Computing Synopsis and Recommendations[S]. U. S. Department of Commerce: NIST Special Publication 800-146, 2012

[8] Wu Jun, Xu Ming. A Comparative Analysis on the Billing Model of Public Cloud Service[J]. *Telecommunications Science*, 2012, 28(1): 127-132(吴俊, 徐溟. 公有云服务计费模式比较研究[J]. 电信科学, 2012, 28(1): 127-132)

[9] Furht B, Escalante A. Handbook of Cloud Computing[M]. New York: Springer, 2010

[10] Zhang Bo, Wu Lili, Zhou Min. The Analysis of User Behavior Based on Web Usage Mining[J]. *Computer Science*, 2006, 33(8): 213-214(张波, 巫莉莉, 周敏. 基于 Web 使用挖掘的用户行为分析[J]. 计算机科学, 2006, 33(8): 213-214)

[11] Li R, Guo R, Xu Z, et al. A Prefetching Model Based on Access Popularity for Geospatial Data in a Cluster-based Caching System [J]. *International Journal of Geographical Information Science*, 2012, 26 (10): 1 831-1 844

[12] Fisher D. Hotmap: Looking at Geographic Attention[J]. *IEEE Transactions on Visualization & Computer Graphics*, 2007, 13(6): 1 184-1 191

[13] Talagala N, Asami S, Patterson D, et al. The Art of Massive Storage: A Web Image Archive [J]. *Computer*, 2000, 33(11): 22-28

[14] Wang Hao, Pan Shaoming, Peng Min, et al. Zipf-like Distribution and Its Application Analysis for Image Data Tile Request in Digital Earth[J]. *Geomatics and Information Science of Wuhan University*, 2010, 35 (3): 356-359 (王浩, 潘少明, 彭敏, 等. 数字地球中影像数据的 Zipf-like 访问分布及应用分析[J]. 武汉大学学报·信息科学版, 2010, 35 (3): 356-359)

[15] Park D J, Kim H J. Prefetch Policies for Large Objects in a Web-enabled GIS Application[J]. *Data & Knowledge Engineering*, 2001, 37(1): 65-84

[16] Yang Chaowei, Wu Huayi, Huang Qunying, et al. Using Spatial Principles to Optimize Distributed Computing for Enabling the Physical Science Discoveries[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2011, 108 (14): 5 498-5 503

[17] Wang Binfei. Access Pattern Analysis and Performance Optimization of IP Network Video Service System[D]. Hefei: University of Science and Technology of China, 2010(王炳飞. IP 网络视频服务系统访问模式分析和性能优化[D]. 合肥: 中国科学技术大学, 2010)

[18] Wu H Y, Li Z L, Zhang H W, et al. Monitoring and Evaluating Web Map Service Resources for Optimizing Map Composition over the Internet to Support Decision Making[J]. *Computers and Geosciences*, 2011, 37(4): 485-494

[19] Li D, Zhang J, Wu H. Spatial Data Quality and Beyond [J]. *International Journal of Geographical Information Science*, 2012, 26(12): 2 277-2 290

[20] Zhang Hanwu, Zhu Xinyan, Zhang Songbo. Response Time in WebGIS Based on Vector Data[J]. *Journal of Geomatics*, 2005, 30(2): 25-27(章汉武, 朱欣焰, 张松波. 基于矢量的 WebGIS 用户响应时间问题的若干研究[J]. 测绘信息与工程, 2005, 30(2): 25-27)

Research and Prediction on Time-Sequence Characteristics of Group-User Access Behavior in Public Map Service

WU Huayi^{1,2} LI Rui^{1,2} ZHOU Zhen¹ JIANG Jie³ GUI Zhipeng⁴

1 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

2 Collaborative Innovation Center of Geospatial Technology, Wuhan 430079, China

3 National Geomatics Center of China, Beijing 100830, China

4 School of Remote Sensing and Information Science, Wuhan University, Wuhan 430079, China

Abstract: Group-user access behavior in public map service has a social nature and there is a certain group - user access pattern , which has a high access aggregative and outburst feature. However, the

(下转第 1316 页)

Analysis of Routing Algorithm for Space Delay/ Disruption Tolerant Network

WANG Sai¹ DENG Fuxing¹ CHENG Zijing² WANG Zhaojun²
ZHANG An'an² WU Jing¹

1 Electronic Information School, Wuhan University, Wuhan 430079, China

2 Space Star Technology Co., Ltd., Beijing 100086, China

Abstract: With the development of space communications and the increasing number of satellites, routing between satellites in the space DTN (delay-tolerant network) network becomes more and more important. Thus, a large number of complex satellite network routing strategies have been proposed. Considering the regularity and predictability of space DTN networks, this paper examines the contact graph routing (CGR) protocol based on satellite regularities; the simulation tool ONE is adopted to simulate real satellite network scenarios. Simulation results show that the delivery ratio, average delay, and the network overhead ratio of CGR are better than those of the Epidemic and Prophet routing strategies.

Key words: space networks; delay-tolerant network; DTN routing

First author: WANG Sai, PhD candidate, specializes in space network. E-mail: hellowangsai@whu.edu.cn

Foundation support: The National Natural Science Foundation of China, No. 60903195; the Natural Science Foundation of Hubei Province, No. 2011CDA042; the Innovation Fund of CAST Satellite Application Research Institute.

(上接第 1286 页)

feature has a great influence on the demands for cloud computing resources for public map service. Thus, how to effectively express and capture the access aggregative feature and the changes of access intensity over time, and predict the access load of public map service accurately, is the important key for selecting and scheduling cloud computing resources on demand, that can address the challenge of concurrent service for massive users. Based on the volume user access logs from public map service and the time-sequence clustering method, this paper first builds a time-sequence distribution model for group-user access arriving behavior; then using the features of multi-peak, variable and periodicity in access intensity, this paper optimally partitions the time-sequence of access arrival rate in a period into different temporal patterns; as there are different probability density distribution of access arrival rate in different temporal patterns, this paper proposes a method of service load forecasting method based on a smoothing time-sequence of cumulative probability distribution. This method has a low complexity and needs few priori data. Experimental results and method application prove that the optimal partition and prediction for the access arrival rate of group-user access based on a time-sequence have a good service response performance for massive users concurrent access, improve the utilization of cloud computing resource, and balance the service quality and cost in public map service.

Key words: public map service; access arrival rate; time series clustering; Fisher optimum split method; resource scheduling

First author: WU Huayi, professor, PhD. His research interests include geospatial data sharing and service, geospatial data mining, geographic information engineering and applications. E-mail: wuhuayi@whu.edu.cn

Corresponding author: LI Rui, PhD, associate professor. E-mail: ruili@whu.edu.cn

Foundation support: The National Natural Science Foundation of China, No. 41371370; the National Key Basic Research and Development Program (973 Program) of China, No. 2012CB719906.