

基于场景的移动应用服务 QoS 建模与评估

胡焰智^{1,2}, 章锋斌^{1,2}, 甘志春¹, 田田², 尹才华²

(1. 国防信息学院 二系, 湖北 武汉 430010; 2. 重庆通信学院 二系, 重庆 400035)

摘要: 针对目前静态服务 QoS 建模难以有效表征移动应用服务 QoS 的时变性和位置相关性, 构建基于场景的移动应用服务 QoS 模型。定义移动应用服务 QoS 指标体系并构建了 QoS 数值对矩阵, 提出服务 QoS 总体评估、特定区域或特定时段评估、热点区域或繁忙时段评估 3 类评估问题。采用遗传算法和数据聚类算法, 通过服务请求位置密度实现区域自动划分, 通过服务请求频次实现时段自动划分, 量化位置区域和时间这 2 个场景条件, 根据限定条件重构 QoS 数值对矩阵, 基于 QoS 数值对矩阵的统计分析给出了服务 QoS 评估方法。最后阐述 QoS 归一化评估和应用框架。

关键词: 移动应用; 服务质量; 时变性; 位置相关

中图分类号: TP393

文献标识码: A

QoS modeling and evaluation of mobile application service based on scene

HU Yan-zhi^{1,2}, ZHANG Feng-bin^{1,2}, GAN Zhi-chun¹, TIAN Tian², YIN Cai-hua²

(1. The Second Department, National Defense Information Academy, Wuhan 430010, China;

2. The Second Department, Chongqing Communication College, Chongqing 400035, China)

Abstract: The current static QoS model can not reflect the QoS time varying and location correlation of mobile application service. An QoS model was presented based on scene for mobile applications service. The QoS index system of mobile application service was defined and the QoS value matrix was constructed. Further, three types of assessment problems were proposed, which were QoS overall assessment, QoS assessment of specific region or specific time periods, and QoS assessment of hot region and busy time. By using genetic algorithm and data clustering algorithm, the location and time of the two scene conditions were quantified, which combine the region division of service request location density and the time division of service request frequency. According to the limiting conditions, QoS value matrix was reconstructed, and the service evaluation method was introduced based on the statistical analysis of the matrix. Finally, the normalized evaluation process and application framework of QoS were given.

Key words: mobile application; service quality; time varying; location correlation

1 引言

2015年《中国移动互联网发展报告》指出, 移动互联网开始渗透到人们生活和工作的各个领域, 形成移动互联网生态系统。目前, 国内外移动互联网应用规模已达百万数量级, 新闻资讯、移动音视频服务、手机游戏、移动支付、基于位置的服务(LBS)等丰富多彩的移动互联网应用和商业模式日趋成熟。

与此同时, 由于移动应用的普及和服务功能的同质化, 移动应用服务的非功能性需求, 特别是服务质量(QoS)正得到越来越多的关注^[1,2], 归纳起来, 移动应用服务 QoS 需求主要源于 3 点: 1) 人们更加注重服务的感受, 已不仅仅只满足于功能要求, 还对服务质量提出了更高的标准, 如服务响应低时延、高安全性等; 2) 相似服务区分的重要依据, 大量的移动应用为用户提供丰富的信息服务, 同时

收稿日期: 2015-11-06

基金项目: 国家科技重大专项子课题基金资助项目(2013ZX03006003-006)

Foundation Item: The Sub-Topic of National Science and Technology Major Project (2013ZX03006003-006)

也带来了服务选择问题，仅依据用户的功能需求难以区分服务，必须基于服务 QoS 来进一步支持服务的动态选择；3) 服务管控的重要参考，QoS 成为业务评价和业务调整的参考指标，也是服务提供商实施用户差分服务和业务差分服务的依据。

当前对服务 QoS 的研究较多，而 QoS 建模与评估是进行 QoS 相关研究的基础与前提。对服务 QoS 建模的研究可分为 3 类：1) 通过服务 QoS 属性的归纳和描述，建立静态指标体系模型，一致的观点是服务 QoS 指标确定不仅与网络传输性能相关，还要表征服务提供者的能力和服务使用者的感受^[3-6]；2) 利用概率模型描述 QoS 属性，包括离散随机变量、标准连续概率分布和任意连续概率分布，它在一定程度上反映了 QoS 属性的动态性^[7-9]；3) 考虑时间要素影响的概率分布建模方法，更加精确地描述服务 QoS，揭示 QoS 随时间而变化的规律^[10]。但是，移动应用服务是基于场景的服务，其服务体验会受到时间、空间位置等场景的影响而发生变化，服务 QoS 建模必须考虑这种特殊性，目前的研究仍有不足，移动应用服务 QoS 模型没有完整表征时间要素和位置相关性的影响。基于此，提出基于场景的移动应用服务 QoS 建模，并阐述评估方法。

2 移动应用服务 QoS 及特征分析

2.1 移动应用服务 QoS

服务 QoS 是非功能性需求的主要内容，具体包括哪些指标属性尚未形成统一的标准，文献[1~6]根据不同的应用给出了多种归纳描述。本文认为服务是面向用户的，考虑移动应用服务 QoS 属性的定义应基于移动用户的体验，即是否能成功地接入服

务，服务的时效性如何，服务运行中是否存在异常，服务实现的功能或提供的信息是否为所需服务，服务的安全性如何以及服务能否应对动态变化的状况？可以将这些问题落实到 6 个关键指标需求上来：服务的可用性、时效性、可靠性、准确性、安全性和适应性，具体如表 1 所示。

需要指出的是，定义可测试的 QoS 指标相当困难，有很多因素和参数需要考虑，有时难以明确有意义的数字化度量值。可用性、时效性和可靠性都从服务使用的角度量化了各个属性值，而准确性、安全性和适应性表达特征难以量化，需要用户给出主观评价，甚至需要借助第三方专家或参考用户群体的长期效果来评定。

2.2 基本假设和考虑

1) 终端是移动应用的前提和基础，本文不考虑终端性能的影响，假设用户终端具有支持所有移动应用的能力，那么影响服务 QoS 的因素可以归为 3 类：①用户行为因素，包括应用服务并发数量、移动用户位置变化状况等；②网络行为因素，包括无线信号稳定性、传输带宽等；③业务行为因素，包括业务流程、业务服务支撑等。

2) 移动应用服务是基于场景的服务，场景是其区别于传统互联网服务的新要素，场景构成包括空间与环境、用户实时状态、用户生活习惯等，但对于某个移动应用服务而言，影响其服务 QoS 的场景变化主要是时间和位置的不同，本文也将以时间和位置变化来表征不同场景。

3) 移动应用服务 QoS 的时变特征，这是网络应用服务的共性。首先网络应用服务的 QoS 是非静态的，各种影响移动应用服务的因素都具有一定的

表 1 移动应用服务 QoS 模型指标

QoS 属性	QoS 描述	计算或表征方法	计量单位
可用性	接入服务成功的概率	$\frac{n_s}{n_r}$ ，其中， n_r 为服务被请求的次数， n_s 为请求成功的次数	百分比
时效性	服务响应时间或信息更新时间	$T_s + T_d$ ，其中， T_s 是服务处理时间， T_d 为发送服务请求和接收服务响应的消息传输时延	时间单位，如 ms
可靠性	服务运行中保持正常的概率	$\frac{n_c}{n_s}$ ，其中， n_s 为成功接入服务次数， n_c 为能够完成既定服务功能的次数	百分比
准确性	功能吻合程度或信息可信程度	移动应用服务为用户提供的功能或信息与用户预期的吻合程度	等级
安全性	安全策略	移动应用服务具备的安全保障能力，包括认证、授权、机密性、完整性和不可否认性等	等级
适应性	应对动态环境的调整策略	移动应用服务在用户场景发生变化时维持服务的能力	等级

随机性，难以准确地用单值来对其进行表示；其次网络应用服务的 QoS 是随时间变化的，在不同时间段内呈现给用户不同的 QoS 性能表现，分时段表征 QoS 的特征是符合实际的。

4) 空间位置相关性是移动应用服务 QoS 所特有，体现了用户和终端“移动”而导致其接入服务位置不断变化的特点。例如在一些信号强度较弱的地方，无线承载带宽和稳定性下降，从而影响到应用服务的 QoS 性能。与时变特征类似，QoS 也需要区分不同空间区域来表征。

5) 在移动应用服务运行过程中，网络和服务提供商可以通过网管和监测设备，对网络状态信息和服务交互信息进行周期性收集和分析，包括消息传输和处理时延、用户位置等，并假设在使用服务时，用户能通过服务评价功能反馈其他主观 QoS 属性评价信息。

3 移动应用服务 QoS 建模

3.1 模型描述

对于一个正常运行的移动应用服务， T 和 S 分别表示服务运行对应的时间集合和空间位置集合，用户在位置 s 、时刻 t 向网络侧发起一次服务请求并接受该移动服务，其中， $t \in T, s \in S$ ，该服务提供方或网络根据交互记录可以得到 QoS 数值对，其格式为 (t, s, q) ，其中， $q = (q_1, q_2, \dots, q_n)$ ，表示 QoS 的 n 个属性对应的计量值。以表 1 所示的 6 个 QoS 属性为例，可用性和可靠性是累积成功次数的，即

$$q_1 = \begin{cases} 0, & \text{服务请求接入失败} \\ 1, & \text{服务请求接入成功} \end{cases}$$

$$q_3 = \begin{cases} 0, & \text{服务功能未成功完成} \\ 1, & \text{服务功能已成功完成} \end{cases}$$

q_2 是以时间为单位的响应时间表征值， q_4 、 q_5 和 q_6 是主观评定的等级。在这一过程中，用户发起服务请求的时间和位置都是随机的，并且可能以非均匀间隔多次接入服务，同时，会有多位用户访问该移动应用服务。因此，通过一定时间的记录，可以得到该服务的 QoS 数值对矩阵为

$$Q = \begin{bmatrix} t_1 & s_1 & q_{1,1} & q_{2,1} & \dots & q_{n,1} \\ t_2 & s_2 & q_{1,2} & q_{2,2} & \dots & q_{n,2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ t_m & s_m & q_{1,m} & q_{2,m} & \dots & q_{n,m} \end{bmatrix} \quad (1)$$

其中， $t_i, s_i, q_{1,i}, q_{2,i}, \dots, q_{n,i}$ 表示对该服务第 i 次请求对

应的 QoS 数值对， $i = 1, 2, \dots, m$ ，根据上述数据和信息，需要解决服务提供方或用户关注的以下几类问题。

- 1) 如何对该移动应用服务总的服务 QoS 进行评价。
- 2) 如何对特定位置区域或特定时段的服务 QoS 进行评价，例如针对服务运行的某一行政地理区域、某一移动通信服务小区和某个工作时段（08:00~12:00）等进行评估。
- 3) 热点区域或繁忙时段的服务 QoS 如何评价。

3.2 模型求解

1) 服务 QoS 总体评估

服务可用性基于 $q_{1,i}$ 形成的 0 和 1 序列进行计

算，即 $\varphi_1 = \frac{\sum_{i=1}^m q_{1,i}}{m}$ ，同理，服务可靠性 $\varphi_3 = \frac{\sum_{i=1}^m q_{3,i}}{m}$ ，

其结果均为百分比形式。服务准确性 φ_4 、安全性 φ_5 和适应性 φ_6 分别以 $q_{4,i}$ 、 $q_{5,i}$ 和 $q_{6,i}$ 形成的序列求均值得

$$\text{到，即有 } \varphi_4 = \frac{\sum_{i=1}^m q_{4,i}}{m}、\varphi_5 = \frac{\sum_{i=1}^m q_{5,i}}{m} \text{ 和 } \varphi_6 = \frac{\sum_{i=1}^m q_{6,i}}{m}，$$

其结果仍为等级值。

对于时效性 φ_2 ，同样也可以采取序列均值来表示。由于其需要考虑的因素多，不同场景下响应时间差异可能较为明显，当需要更准确的指标信息时，采用概率表示比均值更科学，例如：0.5 s 以下响应时间的概率为 85%（或 85% 情况下响应时间低于 0.5 s）、1 s 以下响应时间的概率为 98% 等。因此， φ_2 应基于 $q_{2,i}$ 形成的时间值随机序列进行概率统计分析来表征。

2) 特定位置区域或特定时段服务 QoS 评估

基于特定位置区域或特定时段的服务 QoS 评估本质是基于 QoS 数值对矩阵在限定条件下的统计分析，特定位置区域的评估过程如下。

①初始化。对特定位置区域边界进行抽样取点 p_1, p_2, \dots, p_k ，构成一个常规的多边形 D （一般考虑为凸多边形或凹多边形），将 QoS 数值对矩阵中用户服务请求位置在多边形内部的数值对才纳入统计评估分析范围。

②位置判断。判断服务请求 s_i 所表征的位置是否位于多边形 D 内，采用射线法进行，即构建直角坐标系，并确定服务请求点 s_i 、区域样点 p_1, p_2, \dots, p_k 及多边形 D 的位置；以位置 s_i 为起点，沿 X 正轴方向画出水平射线，并计算其与多边形 D

的交点数；对起点位置 s_i 进行微量移动到 s_i' ，该微量移动对 s_i 的 X 和 Y 坐标均要有很小的变化，再以 s_i' 的位置为起点，沿 X 正轴方向画出水平射线，计算其与多边形 D 的交点数；若 2 次的交点数均为奇数，则认为服务请求 s_i 位于该特定区域内，记为 $s_i \in D$ 。需要说明的是，针对极特殊形状的多边形 D ，该判断过程仍可能判断失误，但对于大数据量的统计而言，是可以接受的。

③统计分析。对 QoS 数值对矩阵 Q 遍历 s_i ， $i=1,2,\dots,m$ ，所有位于多边形 D 内的数值对构成新的 QoS 数值对矩阵 Q_1 ，具体过程如下。

```

 $Q_1 = []$ ,  $j = 1$ 
遍历  $s_i$ ,  $i = 1, 2, \dots, m$ 
    if ( $s_i \in D$ )
         $Q_1(j, :) = Q(i, :)$ 
         $j = j + 1$ 
    end
end
    
```

然后基于矩阵 Q_1 ，采用第一类问题同样的方法进行评。特定时段的评估算法过程类似，但评估范围的判断存在差异，假设此特定时段为 $[T_1, T_2]$ ，判断过程如下。

```

 $Q_2 = []$ ,  $j = 1$ 
遍历  $t_i$ ,  $i = 1, 2, \dots, m$ 
    if ( $T_1 \leq t_i \leq T_2$ )
         $Q_2(j, :) = Q(i, :)$ 
         $j = j + 1$ 
    end
end
    
```

然后基于新的 QoS 数值对矩阵 Q_2 进行统计分析与评估。

对于时间和位置区域均有限定，即特定区域特定时段内的服务 QoS 评估，方法类似，修改判断条件即可，过程如下。

```

 $Q_3 = []$ ,  $j = 1$ 
遍历  $(t_i, s_i)$ ,  $i = 1, 2, \dots, m$ 
    if ( $T_1 \leq t_i \leq T_2$ ) && ( $s_i \in D$ )
         $Q_3(j, :) = Q(i, :)$ 
         $j = j + 1$ 
    end
end
    
```

3) 热点区域或繁忙时段服务 QoS 评估

移动应用在热点区域或繁忙时段的服务提供对于该服务“黏住”用户和扩展应用意义重大，也是核心用户群体对服务效果体验的主要内容，它的

评估也是对 QoS 数值对矩阵在限定条件下的统计分析，但其限定条件并不明确，热点区域或繁忙时段均为模糊描述，这也是进行准确评估的难点。后续分析中，并不重复分析评估方法流程，仅针对热点区域或繁忙时段的限定条件量化问题提出方法。

4 移动应用服务 QoS 场景条件量化方法

4.1 基于服务请求位置密度的区域划分

服务热点地域也是服务交互多的区域，直观来看，就是服务请求位置集中、密度大的区域，下面分析单热点和多热点的服务区域自动划分。

1) 基于单中心点的搜索聚类

对于 QoS 数值对矩阵 Q ，遍历所有服务位置，找出离所有点距离之和最小的位置作为整个服务区域的中心点，然后逐步删除离中心最远的点，删除到一定比例后，剩下的服务请求位置聚合为热点区域，假设 2 个位置点之间的距离表示为 $|s_1 - s_j|$ ，则聚类过程如下。

```

 $S_m = \sum_{j=1}^m |s_1 - s_j|$ ,
for  $i=2:m$ 
     $S_i = \sum_{j=1}^m |s_i - s_j|$ 
    if ( $S_i < S_m$ )
         $S_m = S_i$ 
         $k_{min} = i$ 
    end
end
end
    
```

$S_i = \text{sort}(|s_{k_{min}} - s_j|)$, $j = 1, 2, \dots, m$

其中，sort 表示对距离大小按升序排列， S_i 表示序号值集合，删除一部分远距离位置点，剩余序号值集合为 $S'_i = S_i(1:[\mu \text{length}(S_i)])$ ，其中[*]表示取整， $0 < \mu < 1$ ，即序列 S'_i 元素个数为 S_i 的一部分，占比为 μ ，且服务位置相对集中，然后可针对该部分的 QoS 数值对矩阵进行统计分析。例如服务可用性可以表示为 $\varphi_i = \frac{\sum_{i \in S'_i} q_{1,i}}{\text{length}(S'_i)}$ 。

上述方法准确性较高，但当面临大数据量的服务记录时，计算量显著增加。对此，针对大数据记录划分服务热点也可以采用一种简化的迭代算法，将所有位置点的线性均值作为中心点，然后删除离

中心最远的点，再计算剩余所有位置点的线性均值为新的中心点，重复迭代此过程，当删除到一定比例后，剩下的服务请求位置聚合为热点区域，其流程如图 1 所示。

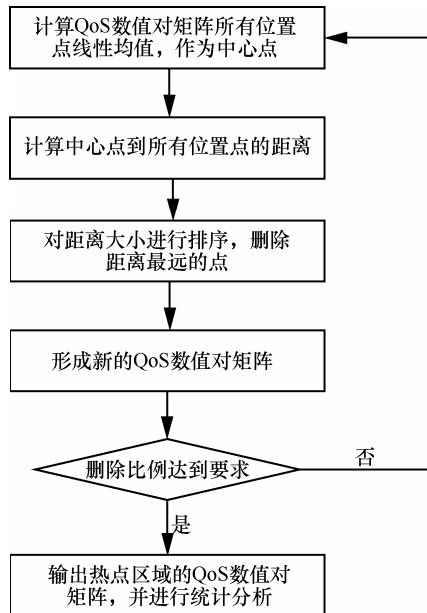


图 1 基于单中心点搜索聚类的迭代流程

2) 基于 k -means 算法的多中心聚类

对于多个服务热点的聚类，需采用数据聚类算法。 k -means 算法简单有效，适合对大规模数据进行聚类时应用。基于 k -means 算法，根据位置点的距离，将 m 个记录的服务位置点划分为 k 簇，使每一个簇内服务点位置集中，而簇间距离相对较远。其算法流程如下。

①任意选择 k 个位置点作为初始的簇中心。

②遍历所有位置点，将其分类给距离最近的簇中心，形成 k 个簇。

③计算每个簇内所有位置点的平均值，作为该簇新的簇中心， k 个簇对应形成 k 个新的簇中心。

④与第③步计算的簇中心相比，值是否发生变化，没有变化表明分簇处于稳定，否则转第④步。

在上述计算过程中，初始点选择对算法结果影响较大，可以进行多次初始选择和聚类，从中选择一组最优的结果，也可以根据移动应用服务热点的先验知识确定初始点，例如商业区和办公区所在位置点可能为热点。

4.2 基于服务请求频次的时段划分

1) 业务频次数据预处理

以一天的 24 h 作为完整的时间域，也是时间

划分的总集合，为了服务请求频次统计和时间分割的需要，首先将整个时间域均匀离散化，假设每个基本时间单元时长为 m min，则可以得到的基本时间单元数为 $l = \frac{24 \times 60}{m} = \frac{1440}{m}$ ， l 为整数，且考虑到

频次统计方便， m 取值可以选取为 10、15、20 等。例如，当 $m = 15$ 时，就可以形成 96 个均匀的基本时间单元，其对应的时间段分别为 [00:00~00:15]、[00:15~00:30]、[00:30~00:45]、...、[23:30~23:45]、[23:45~24:00]。基于上述时间段起止时间，对 QoS 数值对矩阵 Q 遍历 t_i ，对每次服务请求时间进行归类，从而统计出每个基本时间单元内服务请求频次，可以得到一个离散的服务请求频次数值序列

$$a_1, a_2, a_3, \dots, a_{l-1}, a_l$$

2) 业务时段自动划分模型

首先应明确业务时段等级划分，如繁忙、较忙、正常、较空闲、空闲等，并以此作为时段划分的标准。时段自动划分所要解决的问题是对离散均匀的基本时间单元进行聚类，即根据服务请求频次的多少将基本时间单元划分到不同业务时段中。需要强调的是，此处聚类问题与一般的数据聚类存在一个显著差异，基于服务请求频次的数值序列具有时序性，即归为一类的数据在时间上是连成一个时间片段的。因此，对数据序列值自身聚类难度较大，可以采取序列分割的方式进行求解，其聚类的关键是确定服务请求频次数值序列的分割点。具体模型构建如下。

假设划分 u 个业务时段，需要设置 $u-1$ 个序列分割点，即从业务频次数列 $a_1, a_2, a_3, \dots, a_{l-1}, a_l$ 的下标值序列 $2, 3, \dots, l-1$ 中选取 $u-1$ 个值，其对应的下标集合 $G = \{g_1, g_2, \dots, g_{u-1}\}$ ，且 $g_1, g_2, \dots, g_{u-1} \in \{2, 3, \dots, l-1\}$ ，对 G 中元素值按升序排列有 $G' = \text{sort}(G) = \{g'_1, g'_2, \dots, g'_{u-1}\}$ ，且 $1 < g'_1 < g'_2 < \dots < g'_{u-1} < l$ ，若存在下标相等的情况，则相邻 2 个时段合并。通过分割， u 个业务时段下标集合 G_1, G_2, \dots, G_u 与完整下标值序列的对应关系如下

$$\underbrace{1, 2, \dots, g'_1}_{G_1}, \underbrace{g'_1 + 1, \dots, g'_2}_{G_2}, \dots, \underbrace{g'_{u-1} + 1, \dots, l}_{G_u}$$

从上述可以看出，每个 G 值组合对应一种聚类方案，于是服务请求频次数列聚类转化为寻找一个

合适的分割点序列下标集合 G ，使 Δ 最小

$$\Delta = \sum_{j=1}^{u-1} \sum_{k \in G_j} \left| a_k - \frac{1}{h_j} \sum_{i \in G_j} a_i \right| \quad (2)$$

其中， h_j 是 G_j 集合中元素的个数。

3) 基于遗传算法的模型求解

由于遗传算法简单通用，具有优越的全局搜索优化性能，因此，下面采用遗传算法对上述模型进行求解，具体算法流程如下。

① 输入。服务请求频次数值序列 $a_1, a_2, \dots, a_{l-1}, a_l$ ，种群规模 M ，最大进化代数 W ，适应度计算控制参数 α 和 β ，选择运算调节参数 γ ，选择门限参数 $rand$ ，且考虑 $200 \leq M \leq 400$ ， $100 \leq W \leq 300$ ， α 为正数，接近于 0，一般取值为 10^{-6} ， β 取值为正整数 1, 2 或 3， $\gamma > 0$ 且接近于 1， $rand$ 是 $[0,1]$ 之间的随机数。

② 初始化。对集合 G 中的元素值采用二进制编码方式，每个元素编码长度为 $[lb \ l] + 1$ ，其中 $[*]$ 表示取整运算，则 G 编码长度为 $(u-1)([lb \ l] + 1)$ 。随机生成 M 组 G 值，其对应的 M 组二进制序列作为初始种群。

③ 遗传进化。具体求解流程如图 2 所示。群体进化主要是进行个体检测修正、选择、交叉、变异运算，到达最大迭代次数后，终止计算。

④ 结束。在最后群体中，选择具有最大适应度个体作为最优解输出，也即得到最优的分割点序列下标集合 G 。

4) 业务时段划分示例

假设在一个移动服务区内，采用随机双峰模型来模拟 24 h 内具有 2 个业务高峰的服务请求频次变化曲线，如图 3 所示。设置一个基本时间单元的时长为 15 min，则整个时间域共有 96 个基本时间单元，通过服务请求频次统计，基于上述业务时段自动划分方法，采用遗传算法将整个时间周期划分为 6 个时段，其结果如图 4 所示。从结果可以划分出时段 1 为 00:00~05:30；时段 2 为 05:30~10:45；时段 3 为 10:45~15:30；时段 4 为 15:30~19:30；时段 5 为 19:30~20:45；时段 6 为 20:45~24:00，其中较为直观地区分出了各个时段，也可以明确量化繁忙时段。

5 QoS 归一化评估和应用框架

通过上述模型求解，可以得到移动应用服务 QoS 的各个属性表征和计算方法，对于每一个移动

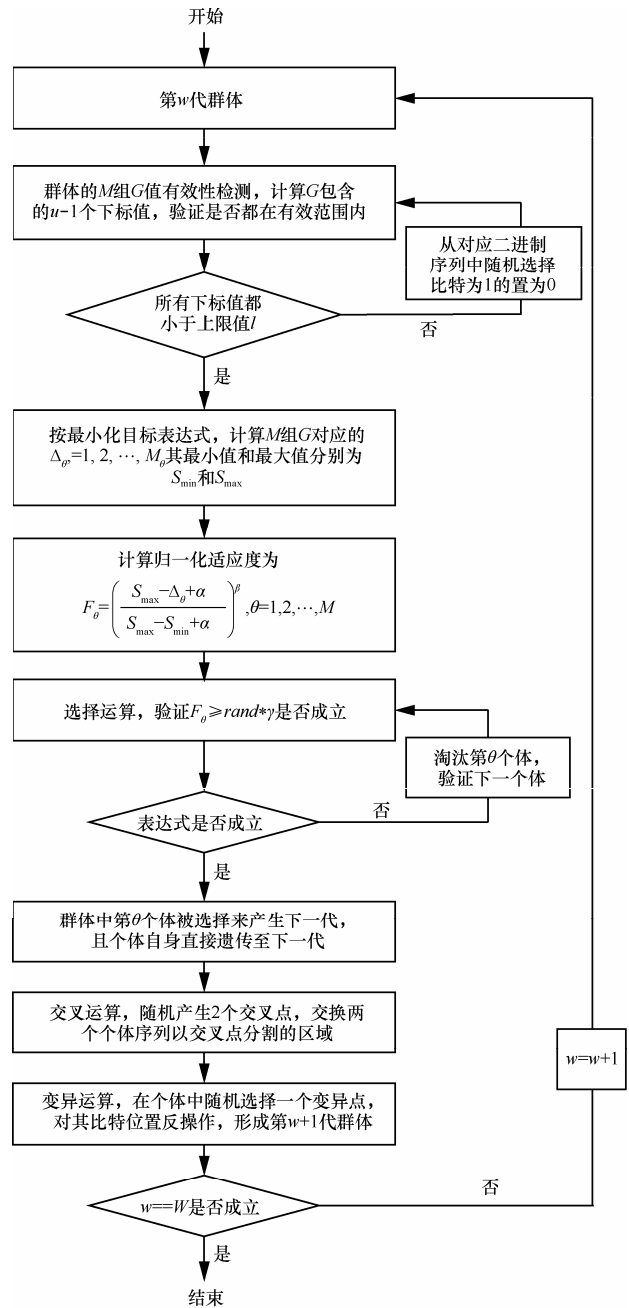


图 2 遗传进化计算流程

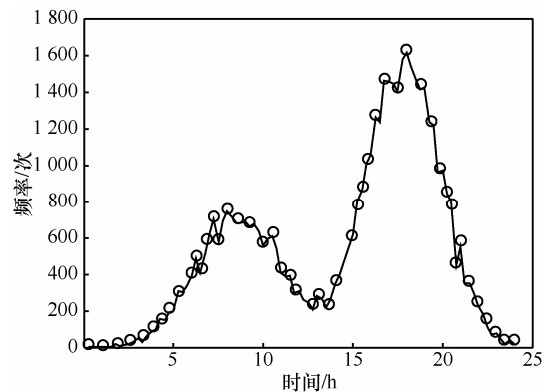


图 3 周期内业务频次变化曲线

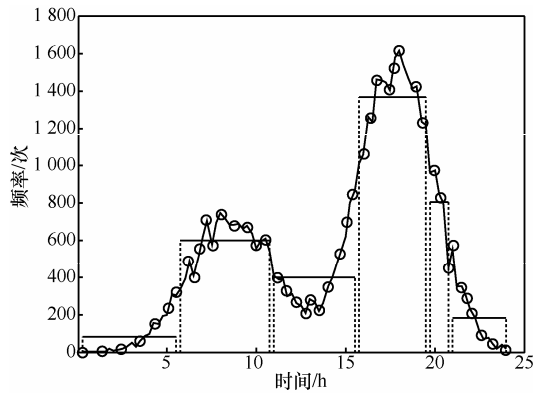


图 4 周期内时段自动划分结果

应用服务对象均可得到指标独立的评估结果。但有时也需要得到某个服务总的评价值，例如同质移动应用服务的比较和选择。基于此，在上述计算分析的基础上阐述归一化评估方法。归一化评估能获得服务的一个总的评价值，它的求解需要解决 3 个问题：数据去模糊化、数据规范化、QoS 属性权重量化。

1) 数据去模糊化

移动应用服务 QoS 的一些属性难以测算，无法直接给出量化值，例如服务适应性、正确性等。这类属性指标表征，往往基于用户的服务体验和领域专家打分，给出不同等级来评价，如（优秀，好，较好，一般，差），这些评价具有一定的模糊性，不易参与进一步的计算评估，为更好地在模糊评价的基础上体现出倾向性，采取三角模糊数评价这类指标较为可行，其基本形式为 $[c_l, c_m, c_u]$ ，其中， c_l 、 c_m 和 c_u 分别表示三角模糊数的下界、核和上界值。如某一指标采用三角模糊数评价为 $[0.7, 0.73, 0.8]$ ，表明其位于 0.7~0.8 之间，最有可能为 0.73。为了进行 QoS 的综合评估，对这三三角模糊数必须进行去模糊化^[11]，即形成一个实数形式的评价值，其计算如下

$$r = \frac{\int_{x \in c} x \mu(x) dx}{\int_{x \in c} \mu(x) dx} = \frac{\int_{c_L}^{c_M} x \mu_1(x) dx + \int_{c_M}^{c_U} x \mu_2(x) dx}{\int_{c_L}^{c_M} \mu_1(x) dx + \int_{c_M}^{c_U} \mu_2(x) dx} \quad (3)$$

其中， $\mu(x)$ 表示属性的隶属度函数，而 $\mu_1(x)$ 、 $\mu_2(x)$ 分别是其对应 $[c_L, c_M]$ 、 $[c_M, c_U]$ 2 个区间段对应的函数表达式。

2) 数据规范化

通过数据规范化，将具有不同量纲的 QoS 各属性值转化到统一的基准上，一般映射到 $[0, 1]$ 区间的无量纲数值。数据规范化的方法很多，可以采用常见的最大最小归一化计算方法，成本型的最大最小归一化计算公式

$$y = \frac{x_{\max} - x + \alpha}{x_{\max} - x_{\min} + \alpha} \quad (4)$$

其中， y 是 x 对应的归一化数值， α 为接近于 0 的正数，一般取值为 10^{-6} 。当某一移动应用服务平均时效性为 3 s，它的归一化需要结合历史记录和用户体验确立的时效性范围，如 $[2 \text{ s}, 7 \text{ s}]$ ，基于上述规范化公式可以得到平均时效性的归一化值为 0.8。也可以采用其对应的效益型归一化计算公式为

$$y = \frac{x - x_{\min} + \alpha}{x_{\max} - x_{\min} + \alpha} \quad (5)$$

需要指出的是，多个指标进行规范化，为了便于进行综合评估，应同时选择成本型或同时选择效益型。

3) QoS 属性权重量化

对于某一移动应用服务，虽然每个属性都要考虑，但是不同指标重要性不一样，将多个独立指标值进行聚合，需要确定在移动应用服务 QoS 的综合评估中各属性所占的权重，一般可通过层次分析法求解得到。

4) 移动应用服务 QoS 应用框架

综合上述 QoS 建模和评估过程，构建了移动应用服务 QoS 应用框架，如图 5 所示。首先确定 QoS

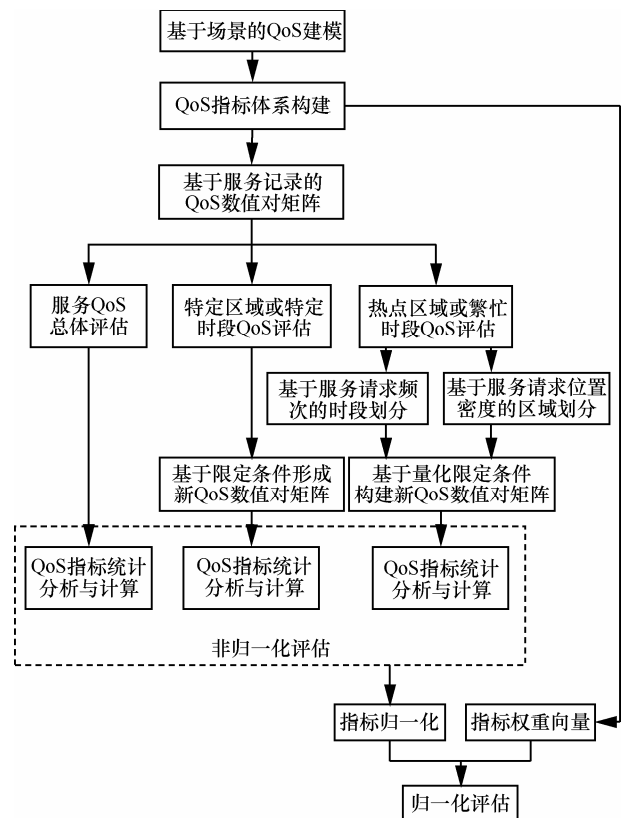


图 5 移动应用服务 QoS 应用框架

指标体系, 基于服务记录形成的 QoS 数值对矩阵, 然后开展 3 种类型的评估服务 QoS 总体评估、特定区域或特定时段 QoS 评估、热点区域或繁忙时段 QoS 评估。特别地, 通过运行时段和位置区域 2 项限定条件的量化, 能够对基于场景的移动应用服务 QoS 进行评估, 该评估结果为多属性多类型值构成的向量。同时, 在 QoS 指标统计分析计算的基础上, 进行指标归一化, 并结合不同指标属性的权重向量, 进行归一化评估, 能够表征不同时段、不同区域服务 QoS 的总评价值。

6 结束语

本文基于移动用户的服务体验, 定义了移动应用服务 QoS 指标体系, 构建了基于场景的移动应用服务 QoS 模型, 该模型能够表征移动应用服务 QoS 的时变性和位置相关性。在服务 QoS 数值对矩阵构建的基础上, 采用基遗传算法和数据聚类算法, 通过服务请求位置密度实现区域自动划分, 通过服务请求频次实现时段自动划分, 量化了位置区域和时间这 2 个场景条件, 根据限定条件重构 QoS 数值对矩阵, 基于 QoS 数值对矩阵的统计分析给出了服务 QoS 评估方法。最后阐述 QoS 归一化评估和应用框架。本文的模型和求解方法是针对服务记录进行统计分析和计算, 可操作性强, 下一步还需采集实际案例数据对建模与评估的方法开展进一步研究。

参考文献:

- [1] GUISELLE A, RAKESH N. Network and QoS-based selection of complementary services[J]. IEEE Transaction on Service Computing, 2015, 8(1): 79-91.
- [2] RAJESWARI M, SAMBASIVAM G, BALAJI N, *et al.* Appraisal and analysis on various web service composition approaches based on QoS factors[J]. Journal of King Saud University – Computer and Information Sciences, 2014, 26(1): 143-152.
- [3] 武蕾, 刘玉龙, 刘士军. 多策略 QoS 感知的服务选择与排序模型[J]. 通信学报, 2012, 33(Z1): 141-147.
WU L, LIU Y L, LIU S J. Multi-strategy QoS-aware service selection and ranking model[J]. Journal on Communications, 2012, 33(Z1): 141-147.
- [4] 马建威, 舒振, 罗雪山. 新型指控系统军事信息服务 QoS 技术研究[J]. 中国电子科学研究院学报, 2009, 4(5): 464-468.
- MA J W, SHU Z, LUO X S. Survey on QoS technology for military information service in next generation C2 system[J]. Journal of CAEIT, 2009, 4(5): 464-468.
- [5] WU B, WU X C. A QoS-aware method for web services discovery[J]. Journal of Geographic Information System, 2010, 2(1): 40-44.
- [6] JAMOUCSI Y, DRISS M, JEZEQUEL J M, *et al.* QoS assurance for service-based applications using discrete-event simulation[J]. IJCSI International Journal of Computer Science Issues, 2010, 7(6): 1-11.
- [7] HWANG S Y, WANG H J, TANG J, *et al.* A probabilistic approach to modeling and estimating the QoS of web-service-based workflows[J]. Information Science, 2007, 177(23): 5484-5503.
- [8] ROSARIO S, BENVENISTE A, HAAR S, *et al.* Probabilistic QoS and soft contracts for transaction-based web services orchestrations[J]. IEEE Transactions on Services Computing, 2008, 1(4): 187-200.
- [9] ZHENG H Y, YANG J, ZHAO W L, *et al.* QoS analysis for web service compositions based on probabilistic QoS[A]. Proc of The 9th International Conference on Service Oriented Computing[C]. 2011. 47-61.
- [10] CHEN L L, YANG J, ZHANG L. Time based QoS modeling and prediction for web services[A]. Proc of the 9th International Conference on Service-Oriented Computing[C]. 2011. 532-540.
- [11] 杨放春, 苏森, 李祯. 混合 QoS 模型感知的语义 Web 服务组合策略[J]. 中国科学 E 辑: 信息科学, 2008, 38(10): 1697-1716.
YANG F C, SU S, LI Z. Semantic Web service composition strategy based on mixed QoS model[J]. Science China Information Sciences, 2008, 38(10): 1697-1716.

作者简介:



胡焰智 (1982-), 男, 湖北黄冈人, 国防信息学院博士生, 重庆通信学院讲师, 主要研究方向为移动应用服务和移动互联网。

章锋斌 (1980-), 男, 安徽东至人, 博士, 重庆通信学院副教授, 主要研究方向为无线通信理论与技术。

甘志春 (1962-), 男, 江苏苏州人, 博士, 国防信息学院教授, 主要研究方向为无线通信理论与技术。

田田 (1982-), 女, 河南新乡人, 硕士, 重庆通信学院讲师, 主要研究方向为移动通信应用技术。

尹才华 (1978-), 男, 重庆人, 硕士, 重庆通信学院副教授, 主要研究方向为移动通信应用技术与服务质量。