

# 基于SLA的用户效用最大化建模

万彩云<sup>1</sup>, 张登银<sup>2</sup>

(1. 南京邮电大学理学院, 江苏南京 210003;

2. 南京邮电大学科技园管理办公室, 江苏南京 210003)

**摘要:**服务等级协议是用户与网络服务提供商签订的关于ISP如何向用户提供服务的协议。为了衡量用户对ISP所提供服务的满意程度,采用微观经济学中的效用函数,建立了一个多用户单资源的效用模型,并设计了基于用户效用最大化和基于资源使用最大化两个计费算法。理论分析和MATLAB数值仿真表明,基于用户效用最大化的算法更优,按此算法用户能获得更大的满意度。研究结果对于量化用户的满意度、提升对服务的性能监测和优化资源分配很有参考价值。

**关键词:**服务等级协议;效用模型;资源分配

**中图分类号:**TP31

**文献标识码:**A

**文章编号:**1673-629X(2011)12-0045-03

## SLA-Based User Effectiveness Maximization Modeling

WAN Cai-yun<sup>1</sup>, ZHANG Deng-yin<sup>2</sup>

(1. College of Science, Nanjing Univ. of Posts & Telecom., Nanjing 210003, China;

2. Executive Office, Sci-Tech Park, Nanjing Univ. of Posts and Telecom., Nanjing 210003, China)

**Abstract:** Service level agreement (SLA) is an agreement about how Internet service providers (ISP) provide service signed by the users and ISP. In this paper, to measure the user's satisfaction to the service provided by ISP, the utility function in the microeconomics is applied, a multiuser single resources utility model is established, and two cost algorithms are designed to maximize the effectiveness of users and the use of resources respectively. The theoretical analysis and the numerical results by MATLAB show that the algorithm based on the user effectiveness maximization is more effective. According to this algorithm, users can achieve greater satisfaction. The research results are valuable in measuring the satisfaction of users, upgrading the monitor of service performance and optimizing resource distribution.

**Key words:** service level agreement; utility model; resource distribution

## 0 引言

服务等级协议(SLA, service level agreement)是用户与ISP(网络服务提供商)签订的关于ISP如何向用户提供服务的协议<sup>[1,2]</sup>。SLA包括技术和商业法律两部分,其中技术部分被称为服务等级规范(SLS, service level specification)。SLS指定了与特定服务相关的网络性能参数及参数的阈值,这些参数用于描述ISP提供的网络服务,也是ISP进行网络计费的重要依据<sup>[3,4]</sup>。

为了衡量用户对ISP所提供服务的满意程度,文中引入效用函数<sup>[5]</sup>,以微观经济学的方法量化衡量用户主观的满意程度,以便于ISP对服务进行性能监测和优化资源分配,为用户提供最高满意度的服务。

## 1 基于SLA的用户效用最大化建模

所谓效用是指消费者通过消费一定数量的商品而获得的满足程度,效用函数则用来刻画满足水平(满意度)与所消费商品数量之间的关系,可表示成 $u = u(x_1, \dots, x_n) = u(X)$ ,  $u$ 为一个实函数,是由 $R_+^n \rightarrow R$ 的一个映射,  $x_i \geq 0$ 是第 $i$ 种商品的选择数。

把网络传输服务看作经济领域中的商品,用户所需服务量看作需求因子,ISP提供的服务量看作供应因子,通过对供求双方的交互和竞争的数学建模,就可以建立起网络服务的效用模型<sup>[6,7]</sup>。

由于不同的资源已由服务提供商预分配,可以理解为不同的资源的分配在相互之间没有任何关系,本质上只考虑一个资源<sup>[8]</sup>。假设 $n$ 个用户共享1种资源,每个用户与ISP签订的SLA协议级别从高到低依次为 $1, 2, \dots, n$ 。每个用户的协议资源数为 $x_j^{(0)}$  ( $j = 1, 2, \dots, n$ ),在第 $k$ 个周期,用户申请的资源数为 $x_j^{(k)}$ ,实际获得的资源数为 $y_j^{(k)}$ 。

设每个用户的效用与其所获得的资源数成正比,

收稿日期:2011-05-16;修回日期:2011-08-21

基金项目:华为公司高校科技基金(YJCB20070401N)

作者简介:万彩云(1969-),女,讲师,研究方向为计算机网络技术;张登银,博士,研究员,博士生导师,研究方向为信号与信息处理、IP网络技术。

而比例系数体现用户的级别高低,用户 1,2,⋯,n 对应的权系数分别为  $\lambda^{n-1}, \lambda^{n-2}, \dots, \lambda, 1$ , 每个级别的用户效用系数是下一个级别用户的  $\lambda$  倍,为使总效用最大,高级别的用户将优先获得资源<sup>[9,10]</sup>。

若一个用户所申请的资源数不超过其协议资源数,则分配给该用户的资源数就是其申请的数目,若申请的资源数超过协议资源数,则视情况许可或者分配其申请的资源数,或者分配协议资源数。数学模型为:

$$\begin{aligned} \max \quad & \sum_{j=1}^n \sum_{k=1}^m \lambda^{n-j} y_j^{(k)} \\ \text{s. t.} \quad & \sum_{j=1}^n y_j^{(k)} \leq R, k=1, 2, \dots, m \\ & y_j^{(k)} = \begin{cases} x_j^{(k)} & x_j^{(k)} \leq x_j^{(0)} \\ \in \{x_j^{(0)}, x_j^{(k)}\} & x_j^{(k)} > x_j^{(0)} \end{cases}, j=1, 2, \dots, n, k=1, 2, \dots, m \end{aligned} \quad (1)$$

其中  $\lambda$  为权系数,是一个充分大的正数,理论上为正无穷。该数学模型中,  $y_j^{(k)}$  是待求的变量,其余符号均表示已知数。

该数学模型有一个前提条件,  $\sum_{j=1}^n x_j^{(0)} \leq R$ , 即总资源数不小于用户协议资源数的总和。在不同的周期,资源的分配是没有关联的,因此对每个周期分别求解数学模型,在第  $k$  个周期,数学模型为:

$$\begin{aligned} \max \quad & \sum_{j=1}^n \lambda^{n-j} y_j^{(k)} \\ \text{s. t.} \quad & \sum_{j=1}^n y_j^{(k)} \leq R \\ & y_j^{(k)} = \begin{cases} x_j^{(k)} & x_j^{(k)} \leq x_j^{(0)} \\ \in \{x_j^{(0)}, x_j^{(k)}\} & x_j^{(k)} > x_j^{(0)} \end{cases}, j=1, 2, \dots, n \end{aligned} \quad (2)$$

设  $S = \{j_1, j_2, \dots, j_r\}$  表示申请资源数超过协议资源数的用户标号集合,  $j_1 < j_2 < \dots < j_r$ , 再设  $T = \{1, 2, \dots, n\} \setminus S$ 。对于  $j \in T$ , 有  $y_j^{(k)} = x_j^{(k)}$ , 对于  $j \in S$ , 有  $y_j^{(k)} \geq x_j^{(0)}$ , 令  $z_j^{(k)} = y_j^{(k)} - x_j^{(0)}$ , 数学模型可以转化为:

$$\begin{aligned} \max \quad & \sum_{j \in S} \lambda^{n-j} z_j^{(k)} \\ \text{s. t.} \quad & \sum_{j \in S} z_j^{(k)} \leq R - \sum_{j \in T} x_j^{(k)} \\ & z_j^{(k)} \in \{0, x_j^{(k)} - x_j^{(0)}\}, j \in S \\ \text{即} \quad & \sum_{i=1}^r \lambda^{n-j_i} z_{j_i}^{(k)} \\ \text{s. t.} \quad & \sum_{i=1}^r z_{j_i}^{(k)} \leq R - \sum_{j \in T} x_j^{(k)} \\ & z_{j_i}^{(k)} \in \{0, x_{j_i}^{(k)} - x_{j_i}^{(0)}\}, i=1, 2, \dots, r \end{aligned} \quad (3)$$

性质 1: 如果  $z_1 = \{0, \dots, 0, z_{j_p}^{(k)}, \dots, z_{j_r}^{(k)}\}$  与  $z_2 = \{0, \dots, 0, z_{j_p}^{(k)}, \dots, z_{j_r}^{(k)}\}$  是上述模型的两个可行解, 其中  $z_{j_p}^{(k)}(z_{j_i}^{(k)})$  是  $z_1(z_2)$  的第一个非零分量。若  $p < q$ , 则  $z_2$  必不是最优解。

证明: 只要证明  $z_1$  的目标函数值比  $z_2$  的目标函数值大。

$z_1$  的目标函数值为  $\lambda^{n-j_p} z_{j_p}^{(k)} + \dots + \lambda^{n-j_r} z_{j_r}^{(k)}$ ,  $z_2$  的目标函数值为  $\lambda^{n-j_q} z_{j_q}^{(k)} + \dots + \lambda^{n-j_r} z_{j_r}^{(k)}$ 。由于  $p < q$ , 可得  $n - j_p > n - j_q$ , 而  $\lambda$  是一个充分大的正数, 因此有

$$\lambda^{n-j_p} z_{j_p}^{(k)} + \dots + \lambda^{n-j_r} z_{j_r}^{(k)} > \lambda^{n-j_q} z_{j_q}^{(k)} + \dots + \lambda^{n-j_r} z_{j_r}^{(k)}$$

## 2 资源分配算法设计

由性质 1, 在分配过程中, 优先分配级别高的用户, 这和实际情况是相符的。进一步对每个周期进行分配, 可以得到如下用户效用最大化算法<sup>[11]</sup>。

算法 A (用户效用最大化):

Step0 输入总资源数  $R$ , 用户协议资源数  $x_j^{(0)}$  与申请资源数  $x_j^{(k)}(j=1, 2, \dots, n)$ , 并对申请资源  $j$  的用户按 SLA 协议从高等级到低等级排序  $1, 2, \dots, n$ ;

Step1 对所有用户进行预分配 (循环): 若  $x_j^{(k)} \leq x_j^{(0)}$ , 则  $y_j^{(k)} = x_j^{(k)}$ , 否则  $y_j^{(k)} = x_j^{(0)}$ , 并记录用户标号, 得到集合  $S = \{j_1, j_2, \dots, j_r\}$ ,  $j_1 < j_2 < \dots < j_r$ ;

Step2 对有标号记录的用户集合  $S$  进行再次分配, 即根据标号从小到大进行循环: 若  $x_j^{(k)} - y_j^{(k)}$  不超过剩余资源数, 则  $y_j^{(k)} = x_j^{(k)}$ , 否则不改变  $y_j^{(k)}$  的值;

Step3 求出这个周期的效用函数  $u = \sum_{j=1}^n \lambda^{n-j} y_j^{(k)}$ 。

显然, 算法 A 的思想就是在满足各个用户基本要求的基础上, 尽量满足高等级用户的要求, 高等级用户在效用函数中被定义为较高值的权系数  $\lambda^{n-j}$ , 从而获得较大的效用函数值。

由性质 1, 可以推测算法 A 是用户效用最大化的数学模型。为便于数值仿真分析, 得到直观的对比数据, 再给出一个算法 B, 主要思想就是在总资源满足的情况下, 谁超额要求多就给谁, 所谓尽量分配, 不浪费, 保证资源使用最大化。

算法 B (资源使用最大化):

Step0 输入总资源数  $R$ , 用户协议资源数  $x_j^{(0)}$  与申请资源数  $x_j^{(k)}(j=1, 2, \dots, n)$ , 并对申请资源  $j$  的用户按 SLA 协议从高等级到低等级排序  $1, 2, \dots, n$ ;

Step1 对所有用户进行预分配 (循环): 若  $x_j^{(k)} \leq x_j^{(0)}$ , 则  $y_j^{(k)} = x_j^{(k)}$ , 否则  $y_j^{(k)} = x_j^{(0)}$ , 并记录用户标号, 得到集合  $S = \{j_1, j_2, \dots, j_r\}$ ,  $j_1 < j_2 < \dots < j_r$ ;

Step2 对有标号记录的用户集合  $S$  进行再次分配, 剩余需再次分配资源的用户所要求的资源数为  $z_p^{(k)} = x_{j_p}^{(k)} - x_{j_p}^{(0)}(j_p \in S)$ , 得到资源集合  $M = \{z_{j_p}^{(k)}, z_{j_q}^{(k)}, \dots, z_{j_r}^{(k)}\}$ , 且  $z_{j_p}^{(k)} \geq z_{j_q}^{(k)} \geq \dots \geq z_{j_r}^{(k)}$ ; 依据  $z_p^{(k)}$  的值从大到小检查集合  $M$  (循环), 若  $z_p^{(k)}$  不超过剩余资源数, 则  $y_j^{(k)} = x_j^{(k)}$ , 否则不改变  $y_j^{(k)}$  的值;

Step3 求出这个周期的效用函数  $u = \sum_{i=1}^n \lambda^{-j} y_j^{(k)}$ 。

3 数值仿真分析

用计算机随机产生了 5 个用户的合同资源数<sup>[12]</sup>、需求资源数,仿真了 15 个周期。总资源数等于 5 个用户合同资源数的和。为了更接近于实际情况,按照一定的概率保持用户在相邻的周期的需求不发生变化。假设用户合同值为 [30,40,25,50,30],需求资源如表 1 所示。按照算法 A 和算法 B 得到的用户实际分配资源,分别如表 2 和表 3 所示。

表 1 用户需求资源

周期 用户	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	35	35	0	0	5	5	45	40	40	40	40	40	40	60	60
2	45	45	75	75	75	75	75	75	75	75	75	75	75	75	75
3	5	60	60	60	35	35	35	35	80	0	30	30	30	10	10
4	5	10	10	10	5	5	5	5	5	5	50	95	95	5	5
5	45	45	70	60	20	20	25	25	25	55	85	85	70	70	70

表 2 算法 A 得到的用户实际分配资源

周期 用户	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	35	35	0	0	5	5	45	40	40	40	30	30	30	60	60
2	45	45	75	75	75	75	75	75	75	40	40	40	40	40	40
3	5	25	60	60	35	35	25	25	25	0	25	25	25	10	10
4	5	10	10	10	5	5	5	5	5	5	50	50	50	5	5
5	45	45	30	30	20	25	25	25	55	30	30	30	30	30	30

表 3 算法 B 得到的用户实际分配资源

周期 用户	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	35	30	0	0	5	5	45	30	40	40	30	30	30	30	30
2	45	45	40	75	75	75	75	75	75	40	40	40	40	40	40
3	5	60	25	60	35	35	25	35	25	0	25	25	25	10	10
4	5	10	10	10	5	5	5	5	5	5	50	50	50	5	5
5	45	30	70	30	20	20	25	25	25	55	30	30	30	70	70

将每一个周期用户实际分配的资源  $y_j$  代入效用函数

$$u = \sum_{j=1}^5 10^{5-j} y_j^{(k)} \tag{4}$$

其中权系数为  $10^{5-j}, j = 1, 2, 3, 4, 5$

由公式(4)得到算法 A 与 B 在 15 个周期的效用函数值,如表 4 所示。

为直观起见,将表 4 两种算法的效用函数值作折线图对比,如图 1 所示,横坐标表示周期,纵坐标表示效用函数的值,小方框表示算法 B 按照超过额度由多到少进行分配(资源使用最大化)得到的值,小圆圈表示算法 A 按照级别由高到低进行分配(用户效用最大化)得到的值。

表 4 算法 A 与 B 的效用函数值

周期 算法	1(9)	2(10)	3(11)	4(12)	5(13)	6(14)	7(15)	8
A	395595	397645	81130	81130	128570	128570	527575	477575
B	395595	351130	42670	81130	128570	128570	527575	378575
A	477575	475105	343030	343030	343030	641080	641080	
B	477575	475105	343030	343030	343030	341120	341120	

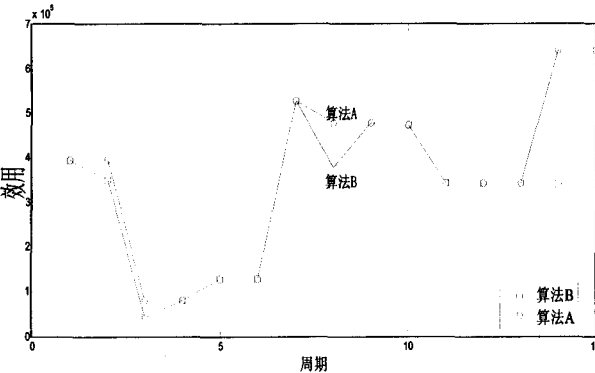


图 1 两种算法的效用函数对比

比较表 4 及图 1 所示的效用函数值可见,算法 A 的效用函数值总是大于等于算法 B 的效用函数值,算法 A 的折线图在算法 B 的折线图的上方或与之平行重合,说明算法 A 更优,按此算法能获得更大的用户满意度。

4 结束语

文中将 SLA 管理的问题转变成成为每个用户选择合适的服务水平,在满足资源限制的前提下实现效用函数值的最大化。对用户而言,服务水平越高,用户越满意;但对 ISP 来说,必须要保证所有服务至少都要达到 SLA 规定的最低要求。

在网络可用资源、服务质量都不断变化的动态环境下,SLA 管理的关键问题就是要保证所有服务的 QoS 水平都达到既定要求。效用函数为我们进行 SLA 管理提供了有效的依据:首先根据 SLA 合约和当前可用资源情况为每个用户进行资源分配并计算效用函数;利用效用函数进行服务性能监测的同时判断是否出现违反 SLA 的情况,如果有则重新计算服务质量并进行资源调整,以确保所有服务都达到 SLA 的规定,使用户得到满意的服务。

参考文献:

[1] 张登银,王雪强.下一代网络计费技术分析[J].电子工程师,2004,30(12):55-58.  
[2] 张登银,王雪强.基于服务级别和流量控制的网络计费[J].重庆邮电学院学报(自然科学版),2005,17(3):328-331.  
[3] 张登银,卢栋梁,王雪强.基于数据包的下一代网络计费策

线与数字的交叉次数; $G$ 为标准样本库的特征向量<sup>[10]</sup>。

该模板匹配算法核心思想就是通过度量待识别手写体数字和样本库中标准样本库的相似度,来分类确定待识别数字。设待识别数字特征向量 $X$ 和标准样本库中第 $i$ 个样本的特征向量 $G$ 相似度为 $R(X, G)$ 。 $R(X, G)$ 定义如下:

$$R(X, G) = \frac{\sum_{i=1}^m x_i \times g_{ik}}{\sqrt{\left[ \sum_{i=1}^m x_i^2 \times \sum_{k=1}^m g_{ik}^2 \right]}} = \cos \alpha$$

其中在上面的公式里<sup>[11]</sup>: $x_i$ 为待识别手写体数字特征向量 $X$ 的第 $i$ 个( $i=1, 2, \dots, 19$ )分量; $g_{ik}$ 为数字标准样本库中第 $i$ 个样本中的第 $k$ 个分量; $m$ 为数字标准样本类别数。公式中的分子为数字特征向量 $X, G$ 之间的积,分母为数字特征向量 $X, G$ 的模, $\alpha$ 是数字特征向量 $X, G$ 在 $m$ 维空间中的夹角。当 $X, G$ 两个数字特征向量完全相同时,它们之间的夹角大小为0,那么 $R(X, G)=1$ ,可以得出它们间的距离 $D(X, G)=0$ ,表明它们的相似度最大,最后计算最大相似度 $R(X, G)$ ,假如 $R(X, G) \geq$  给定阈值,这样可以找到与待识别数字最接近的样本类别<sup>[12]</sup>,最终确认出待识别的手写数字。

### 3 实验的结论与分析

实验采用 CCD 摄像头采集汇款单的手写体图像,图像预处理、关键特征提取、数字识别等过程,判定汇款单表格中有无数字、该数字是多少,最后识别出这些数字。本实验采用 100 张汇款单图像,对其进行试验,实验结果发现识别错误的手写体数字主要是 1, 7, 0, 6, 4, 原因在于 1 和 7, 0 和 6, 6 和 4 在书写结构方面的相似性非常高,而 2, 3, 5, 8, 9 的识别率达到 99.8%。

(上接第 47 页)

- 略研究[J]. 电子与信息学报, 2006, 28(12): 2382-2385.
- [4] 张登银, 李正, 程春玲. 基于 SLA 的下一代网络计费方法[J]. 计算机应用研究, 2009, 26(7): 2713-2716.
- [5] 肖红叶. 高级微观经济学[M]. 北京: 中国金融出版社, 2003.
- [6] Zhang Yong, Feng Wenying, Hurley R. Integration of QoS Queuing Schedules to QoS Caching Schemes[C]// Proceedings of the 5th IEEE/ACIS International Conference on Computer and Information Science. [s. l.]: IEEE/ACIS, 2006: 1-6.
- [7] Ibrahim H A, Nossier B M, Darwish M G. Billing System for Internet Service Provider (ISP) [C]// MELECON' 2002. Mediterranean: [s. n.], 2002: 260-268.

### 4 结束语

文中基于图像识别技术对手写体数字进行识别,相对于传统的光学的光电变换原理对数字进行识别,在识别率和实时性方面得到了有效的提高。但该系统还需要完善的是手写体数字特征提取,这样更进一步提高手写体数字识别率。

#### 参考文献:

- [1] 龚声蓉. 数字图像处理与分析[M]. 北京: 清华大学出版社, 2006.
- [2] 田村秀行. 计算机图像处理(日)[M]. 北京: 科学出版社, 2004.
- [3] 张婷. 基于图像识别技术的光学标记阅读机的研究与应用[D]. 合肥: 安徽大学, 2007.
- [4] 王虎. 基于图像识别的标记阅读机及选举计票系统研究[D]. 合肥: 安徽大学, 2006.
- [5] 刘克. 基于轮廓分段特征的手写阿拉伯数字识别[J]. 计算机学报, 1996(10): 1066-1073.
- [6] 孙家广. 计算机图形学[M]. 北京: 清华大学出版社, 1996.
- [7] 张站. 基于图像识别的阅卷系统的设计与实现[D]. 合肥: 安徽大学, 2008.
- [8] 于振华. 基于骨架结构特征的手写数字识别方法[J]. 模式识别, 2010(6): 95-98.
- [9] 杜敏. 基于混合特征提取的手写体数字识别方法的研究[J]. 西安交通大学学报, 1996(9): 94-98.
- [10] Mahmoud S A. Arabic Character Recognition Using Fourier Descriptors and Character Contour Encoding [J]. Pattern Recognition, 1994, 27: 815-824.
- [11] Pedrycz. Fuzzy Sets in Pattern Recognition: Methodology and Methods[J]. Pattern Recognition, 1990, 23: 121-146.
- [12] Burel D G, Potteier I, Catros D J Y. Recognition of Hand-Written Numerals[J]. IEEE Trans. Syst. Man Cybernet, 1990, 20: 835-847.
- [8] Davoyan R, Altmann J. Real-time market model for pricing differentiated services [C]// Fourth International Conference on Networking and Services. [s. l.]: [s. n.], 2008: 134-140.
- [9] 王平, 李信满, 赵宏, 等. 基于服务质量的网络计费的研究[J]. 小型微型计算机系统, 2002, 23(1): 74-77.
- [10] 李晓花, 李小军. IP 网络 QoS 几种排队算法的对比分析[J]. 四川理工学院院报(自然科学版), 2006, 19(3): 81-84.
- [11] 何芳, 邱雪松, 成璐. 面向业务的下一代互联网计费系统体系模型[J]. 北京邮电大学学报, 2004, 27: 54-58.
- [12] 马莉. MATLAB 语言实用教程[M]. 北京: 清华大学出版社, 2010.