

基于广义线性模型的 QoE 评估方法研究*

吴吉祥, 李凡, 夏靖波, 王恺

(空军工程大学 电讯工程学院, 西安 710077)

摘要: 基于简化的广义线性模型,建立了 QoE 等级概率分布回归模型。该模型能够通过网络 QoS 参数映射得到用户 QoE 等级累积概率分布。采用最大似然估计完成了对模型参数的估计,并使用皮尔逊方法对模型的拟合优度进行了检验。Web 业务的 QoE 评估实例验证了方法的有效性。

关键词: 用户感知质量; 服务质量; 广义线性模型; 最大似然估计; 皮尔逊检验

中图分类号: TP393 **文献标志码:** A **文章编号:** 1001-3695(2012)10-3911-03

doi:10.3969/j.issn.1001-3695.2012.10.082

QoE evaluation method through generalized linear model

WU Ji-xiang, LI Fan, XIA Jing-bo, WANG Kai

(Telecommunication Engineering Institute, Air Force Engineering University, Xi'an 710077, China)

Abstract: As the evaluation of QoE was an ordinal multiple categories problem, this paper established the regression model of QoE categories probability distribution based on the GLM. This model related cumulative probabilities of particular QoE categories to QoS parameters of network. Used the MLE and Pearson test to estimate the model parameters and validate the model respectively. Estimated the QoE of Web service utilize the proposed method, and the result proves that the method is effective.

Key words: quality of experience(QoE); quality of service(QoS); generalized linear model(GLM); maximum likelihood estimation(MLE); Pearson test

0 引言

用户的满意程度已经成为当前网络服务提供商最关注的问题。业务的服务质量可以采用 QoE 来进行评估。直接进行用户满意度测量和由可测量的服务质量(QoS)参数映射是评估用户 QoE 的两种主要方法。MOS(mean opinion score)是用于测量语音 QoE 的模型,ITU 把它写入了规范^[1]。MOS 能够直接、真实地反映用户对业务的满意情况。它通过计算所有用户的平均分来给出用户对业务的满意情况。然而,用户感受等级之间的差距并不像它们对应的数值那样是等差分布的,因此通过量化的平均值计算也不能够完全反映用户的满意情况。E-model 是一种用于评估语音 QoE 的重要模型,它建立了多个因子与用户感受之间的回归方程^[2]。另外,ITU 提出的 PSQM、PESQ^[3,4]等模型都是用于特定的语音业务,但是应用范围窄,代价大。通过研究 QoS 与 QoE 之间的关系完成了从可测量的 QoS 到定量的 QoE 的映射。文献[5]采用主成分分析和多元线性回归的方法建立了 QoE 与 QoS 之间的多元线性回归方程。然而,QoE 与 QoS 之间并不是简单的线性关系。文献[6]建立了 QoS 与 QoE 之间的非线性关系方程,却没有给出方程中参数获取的方法。文献[7]针对视频业务通过 BP 神经网络的方法,建立了带宽、时延等 QoS 参数与 QoE 之间的关系模型,遗憾的是它并未建立在人的真实感受之上。文献[8]提出采用广义线性模型(GLM)来建立 QoE 与 QoS 之间的映射关系

的一种新思路,而文中所给出的模型考虑了 QoS 之间的相互影响项,增多了模型需要估计的参数,从而使模型的复杂度变大。同时,它仅仅分析了链路能力这一个参数与 MOS 之间的关系。本文将对如何采用简化的 GLM 模型建立多个 QoS 参数与 QoE 等级概率分布之间的映射关系进行研究。

1 基本原理

1972 年 Nelder 等人引入广义线性模型一词,此后对其研究工作逐渐增加。广义线性模型是常见的正态线性模型的直接推广,其基本形式^[9]表示为

$$g(\mu_i) = g(E[Y_i]) = \sum_j X_{ij}\beta_j + \xi_i \quad (1)$$

$$\text{var}[Y_i] = \varphi V(\mu_i)/\omega_i \quad (2)$$

其中: Y_i 为因变量向量; $g(\cdot)$ 为联系函数; X_{ij} 为已知自变量矩阵; β_j 为需要估计的模型参数向量; ξ_i 为干扰项; φ 为方差函数 $V(x)$ 的散布参数; ω_i 为每一观察值的信度或权重; var 为方差。广义线性模型由随机部分、系统部分和联系函数三部分构成。随机成分用于明确因变量的概率分布,包含自然指数分布族中某一个分布的若干独立观察值 $Y = (y_1, y_2, \dots, y_n)$; 系统成分确定预测解释变量的线性函数,通过一个线性模型 $\eta = X\beta$ 将预测向量 η 与一组解释变量 X 联系起来;联系函数 $g(\cdot)$ 用来建立系统成分与随机成分的期望值之间的函数关系,即 $\eta_i = g(\mu_i)$, $g(\cdot)$ 为任意单调可导函数。式(1)未考虑自变量之间的相互影响,是简化的广义线性模型。

收稿日期: 2012-02-12; 修回日期: 2012-04-12 基金项目: 全军军事类研究生资助课题(2010XXX-X88, 2011XXX-X23)

作者简介: 吴吉祥(1985-), 男, 山东莱芜人, 博士研究生, 主要研究方向为军事通信网络管理、网络评估等(jixiang19850126@163.com); 夏靖波(1963-), 男, 教授, 博导, 主要研究方向为军事通信网络管理、安全与评估。

2 基于 GLM 的 QoE 评估方法

2.1 QoE 等级概率分布回归模型

参照 MOS 模型对用户满意情况的划分,根据用户对业务的满意情况,从高到低把 QoE 等级 C_i 划分为五个等级:

$$C_i \in [\text{bad, poor, fair, good, excellent}] \quad i = 1, 2, \dots, 5 \quad (3)$$

按照序数关系, C_1 为 Bad, 然后依次对应。便于说明,把 QoE 等级映射到连续的数字 i , 则由式(4)来表示 QoE 等级概率分布:

$$u_i = P(C \leq i | X) \quad 0 \leq u_i \leq 1 \quad (4)$$

其中: C 为需要预测的 QoE 等级; $X = (x_1, x_2, \dots, x_n)$ 为与 QoE 有关的 QoS 参数向量; i 是序数。

QoE 等级的累计概率分布问题作为一个有序多分类的问题,有多种联系函数可供选择。采用 logit 函数^[10]对 u_i 进行处理可得

$$Q_i = \ln\left(\frac{u_i}{1 - u_i}\right) \quad Q_i \in (-\infty, +\infty) \quad (5)$$

需要注明的是,由于 $0 \leq u_i \leq 1$, 当 u_i 为 1 或 0 的时候,式(5)的写法是不正确的,但是这时,中间变量 Q_i 趋向于无穷,在意义上是可以理解的,因而,在推导的过程中仍然用式(5)来完成。假设 Q_i 与自变量之间 X 具有线性关系,基于简化的广义线性模型的基本原理,根据式(1)建立如下线性关系:

$$Q_i = \sum_{l=1}^m \beta_l x_l + \beta_0 \quad (6)$$

其中:对于不同的 Q_i , 系数 β_l 是相同的,而 β_0 则是不同的。

把式(4)(5)代入式(6)后可得

$$\ln\left(\frac{P(C \leq i | X)}{1 - P(C \leq i | X)}\right) = \sum_{l=1}^m \beta_l x_l + \beta_0 \quad (7)$$

对式(7)进行变换得到

$$u_i = P(C \leq i | X) = \frac{\exp(\sum_{l=1}^m \beta_l x_l + \beta_0)}{1 + \exp(\sum_{l=1}^m \beta_l x_l + \beta_0)} \quad (8)$$

这样就建立了 QoE 等级的累计概率分布 u_i 与 QoS 参数 X 之间的回归模型。特定 QoS 参数预测的 QoE 等级 C 属于某一等级 i 的概率 P_i :

$$P_i = P(C = i | X) = \begin{cases} u_1 & i = 1 \\ u_i - u_{i-1} & i = 2, 3, 4, 5 \end{cases} \quad (9)$$

2.2 参数估计与拟合优度检验

普通线性回归假设响应变量为正态分布,采用最小二乘法进行参数估计。广义线性模型对响应变量的正态性不作要求,故不同于普通线性回归,对其参数可以采用极大似然估计法进行估计。

通过参数估计获得了对先验数据支持最好的回归模型,但是这只是建立在先验数据之上的。建立回归模型的目的是要根据所获得的 QoS 参数数据对 QoE 等级概率分布进行预测,因而需要对回归模型的拟合优度进行检验。利用 test 集对模型的拟合优度进行检验,从而确定回归模型是否满足要求。如果由模型预测出的 QoE 等级与实际 test 集中的用户等级在统计意义上一致,则说明所建立的回归模型具有较好的预测功能。

基于离差的方差检验方法并不适用于像 QoE 等级这样响应变量为离散变量的情况,可以采用皮尔逊 χ^2 检验法^[11]。皮

尔逊 χ^2 检验根据 t 实际频数与由模型预测出的期望频数相对平方偏差的总和进行检验。利用皮尔逊 χ^2 检验对模型的拟合优度进行检验:

$$\chi^2 = \sum_{i=1}^5 \frac{(\nu_i - NP_i)^2}{NP_i} \quad (10)$$

其中: χ^2 为皮尔逊检验统计量; N 为 test 集中样本总数; ν_i 为 test 集样本数据中 QoE 属于等级 i 的频数; P_i 为由模型根据 QoS 参数值计算出的 QoE 属于等级 i 概率。根据皮尔逊 χ^2 检验规则,如果 χ^2 充分大,则认为回归模型不够理想;否则认为回归模型具有较高的拟合优度,符合需求。

2.3 评估结果计算

根据前面建立的基于 GLM 的 QoE 等级概率分布回归模型,可以由实际测量的 QoS 参数 $X = (x_1, x_2, \dots, x_n)$ 来预测用户对当前网络的 QoE 等级累计概率分布 $u = (u_1, u_2, u_3, u_4, u_5)$ 。根据式(9)可以计算具体 QoE 等级的概率分布,即得到 $P = (P_1, P_2, P_3, P_4, P_5)$ 。这样不仅可以获得定量的 QoE 等级情况,还能清楚地了解用户对业务的感受情况是如何分布的,使得对网络业务的 QoE 情况一目了然。对网络业务 QoE 等级的确定可以通过求解概率最大的 QoE 等级来实现:

$$\bar{P}_i = \max(P_1, P_2, P_3, P_4, P_5) \quad (11)$$

若 \bar{P}_i 为 QoE 等级里最大的,则认为当前网络业务的 QoE 评估等级为 i ,进而根据式(3)确定其 QoE 最终评估为对应等级。

3 实例分析

Web 业务是网络中最重要也是使用最频繁的应用,人们通过浏览网页获取信息、进行购物等,Web 业务的好坏直接影响用户使用网络的感受。时延、丢包率和带宽是网络性能的三个基本参数,它们直接作用于网络上运行业务的质量,进而影响用户使用网络的感受。下面基于文中给出的基于 GLM 的 QoE 等级概率分布回归模型,通过分析时延、丢包率和带宽与用户 QoE 之间的关系,对 Web 业务的 QoE 进行评估。本文采用用户 QoE 调查与世纪前线(<http://benchmark.avl.com.cn/intro.html>)的测试直通车相结合的方式获取数据样本。

3.1 模型建立

利用世纪前线测试直通车的测试过程如下,在不同的时间段利用校园网络对测试网站世纪前线进行测试,首先控制校园网出口带宽为某一值,记录下 20 个用户当前访问测试网站的主观感觉,用户感知度按照 QoE 等级进行划分,同时利用世纪前线提供的工具得到对应的网络延迟、丢包率和网络带宽信息。然后,调整校园网出口带宽,进行多次测试获取数据。样本分 training 集和 test 集。基于 GLM 的 QoE 等级概率分布回归模型,所建立的 QoE 等级概率分布与网络时延、丢包率和网络带宽之间的回归模型如下:

$$u_i = P(C \leq i | X) = \frac{\exp(\beta_i + \beta_D D + \beta_E E + \beta_B B)}{1 + \exp(\beta_i + \beta_D D + \beta_E E + \beta_B B)} \quad (12)$$

其中: X 为某一次测试的网络 QoS 参数值,分别为时延(D)、丢包率(E)和网络带宽(B); β_D 、 β_E 和 β_B 分别为对应参数的回归系数, β_i 是截止系数,对于不同的 i , β_D 、 β_E 和 β_B 是相同的,而 β_i 却是不同的; u_i 是指用户的 QoE 不大于等级 i 的概率,反映了用户 QoE 等级的概率分布。根据 training 集样本数据,首先

采用最大似然估计的方法对模型参数进行估计,在这里置信区间选取 0.95,得到的结果如表 1 所示。

所有估计的 QoS 参数的系数都具有统计意义,因而全部列出。在 D 、 E 和 B 已知的情况下,通过式(12)计算用户 QoE 等级的累计概率分布。

3.2 模型检验

采用 test 集的样本数据,对上面建立的回归模型利用皮尔逊 χ^2 检验方法检验拟合优度。样本中共有 50 组数据,即 $N = 50$ 。根据式(12)和(9),计算用户 QoE 各个等级概率 P_i 。 P_i 与样本数据中 QoE 属于等级 i 的频数 v_i 如表 2 所示。

表 1 参数估计值

参数	估计值	参数	估计值
β_D	0.124	β_2	-2.708
β_E	124.154	β_3	1.100
β_B	-1.475	β_4	5.255
β_1	-6.848	β_5	9.103

表 2 频数分布

QoE 等级	实际频数	P_i
1	0	0.015 2
2	11	0.171 3
3	29	0.654 4
4	10	0.148 4
5	0	0.010 7

由式(10)计算 χ^2 为 3.248 3,而自由度为 4,则 $\chi_{0.05}^2(4) = 9.488$ 。于是得到:

$$\chi^2 < \chi_{0.05}^2(4)$$

因而,在水平 0.05 条件下认为用回归模型进行预测的结果和实际情况是相吻合的,即所建立的回归模型是符合要求的。为了分析在不同样本数情况下的拟合情况,还计算了 N 分别为 25、75、100、125 时的 χ^2 ,具体结果如图 1 所示。

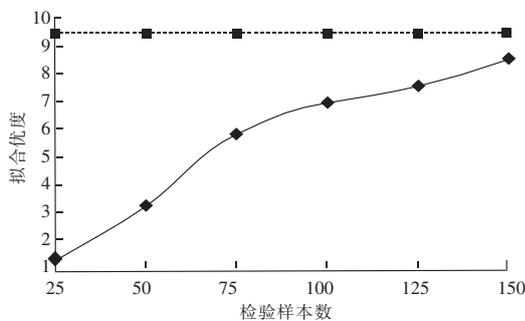


图 1 不同样本数的拟合情况

图 1 中,实线表示不同样本数情况下的 χ^2 ,虚线表示 $\chi_{0.05}^2(4)$ 。建立回归模型所使用的 training 集的样本数为 120,在进行模型检验的时候所使用 test 集的样本数最大为 150。在从图 1 中可以看出,在所进行的检验中,回归模型的拟合优度均符合要求。但是随着样本数 N 的不断增大, χ^2 的值也不断增大,这是因为样本数增大会带来更多回归模型预测准确度较低的样本点,这样就会导致模型的拟合优度降低。

3.3 QoE 评估

根据所建立的 QoE 等级概率回归模型,可以对 Web 业务的 QoE 进行评估。例如,已知 D 、 E 和 B 分别为 55、0.022 和 4,结合表 1 参数值和式(13)可以获得用户 QoE 从低到高的累计概率分布分别为 0.0393、0.7198、0.9914、0.9999、1.0000。按照式(9)可以计算用户 QoE 等级概率分别为 0.0393、0.6805、0.2716、0.0085、0.000。经过计算,就可以得出在 D 、 E 和 B 分别为 55、0.022 和 4 的 QoS 条件下,用户质量感受是如何分布的。根据式(12)求解概率最大的 QoE 等级,得到 $\bar{P}_2 = 0.6805$,即将会有 68.05% 比例的用户认为当前 Web 业务 QoE 属于等级 2。因而,对当前业务的用户指令评估为 poor。

为了对文中方法的有效性进行说明,把 MOS 线性回归方法和直接采用 GLM 的方法在评估的准确度上进行了对比。

MOS 线性回归方法通过建立 MOS 得分与 QoS 参数之间的线性回归方程对 MOS 值进行预测,进而确定 QoE 等级。QoE 评估准确度是由评估结果中正确的数量与测试样本中的实际数量相除所求的。从表 3 中可以看出采用 MOS 线性回归方法的评估准确度均低于采用 GLM 的评估方法,这充分说明了 GLM 模型用于 QoE 评估的有效性。

表 3 与线性回归方法的比较

方法	bad	poor	fair	good	excellent
MOS 线性回归方法	80.2	81.6	86.5	87.4	85.3
GLM 方法	85.9	87.5	91.8	93.4	92.5

4 结束语

用户的 QoE 评估为有序多分类问题,其响应变量为序数,而 GLM 模型属于指数非线性回归,对响应变量的要求很低,专门用于响应变量为离散属性变量、自变量为连续变量或离散属性变量的回归问题。文中给出的基于 GLM 基本原理构建的网络业务 QoE 评估方法能够有效地处理 QoE 等级为序数的问题。与 MOS 等方法仅仅给出一个 QoE 分数相比,该方法不仅对业务的 QoE 进行评估,还能够给出详细的 QoE 等级概率分布情况,能够为运营商或者用户提供更加丰富的网络业务质量信息。由于基于 GLM 模型的有序回归分析是一种统计方法,模型的优劣不可避免地取决于实测样本数据的准确性及代表性。在数据的处理过程中还存在着如何剔除样本中可信度较低或者非典型的样本数据等。在下一步的工作中将就如何获取高可信度的样本数据进行研究。

参考文献:

- [1] ITU-T P 800.1, Mean opinion score (MOS) terminology [S]. Geneva: ITU, 2001.
- [2] ITU-T G 107, The E-Model, a computational model for use in transmission planning [S]. Geneva: ITU, 2005.
- [3] ITU-T P 862, Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs [S]. Geneva: ITU, 2001.
- [4] ITU-T Rec P861, Objective quality measurement of telephone-band (300-3400 Hz) speech codecs [S]. Geneva: ITU, 1996.
- [5] ITO Y, TASAKA S. Quantitative assessment of user-level QoS and its mapping [J]. IEEE Trans on Multimedia, 2005, 7(3): 573-584.
- [6] KIM H J, LEE D H, LEE J M, et al. The QoE evaluation method through the QoS-QoE correlation model [C] // Proc of the 4th International Conference on Networked Computing and Advanced Information Management. 2008: 719-725.
- [7] DU Hai-qing, GUO Chang, LIU Yi-xi, et al. Research on relationship between QoE and QoS based on BP neural network [C] // Proc of IC-NIDC. 2009: 312-315.
- [8] JANOWSKI L, PAPIR Z. Modeling subjective tests of quality of experience with a generalized linear model [C] // Proc of QoMEx. 2009: 35-40.
- [9] 陈希孺. 广义线性模型(一) [J]. 数理统计与管理, 2002, 21(5): 54-61.
- [10] McCULLAGH P, NELDER J. Generalized linear models [M]. London: Chapman and Hall, 1983: 101-105.
- [11] TORRES V A, RIOS-CURIL A. The effect and adjustment of complex surveys on chi-squared goodness of fit tests; some monte carlo evidence [C] // Proc of the Survey Research Methods Section. [S. l.]: American Statistical Association, 1994: 602-607.