

用户体验质量(QoE)的模型与评价方法综述

林 闯 胡 杰 孔祥震

(清华大学计算机科学与技术系 北京 100084)

摘 要 随着信息时代的到来,新的多样性的信息服务大量涌现,人们对信息服务的要求也越来越高.为了获得用户对服务的认可,服务提供商必须确保良好的用户体验质量.因此,用户体验质量不仅成为学术界重要的研究课题之一,而且也成为工业界关键的服务评价指标之一.文中综述了用户体验质量的模型与评价方法等方面的工作.介绍了该领域的研究现状与进展,总结归纳了用户体验质量的影响因素、量化方法及三类评价方法学,重点讨论了几种具体的基于不同学科的用户体验质量的模型和评价方法.基于这些讨论,该文以视频流媒体服务为背景,提出了基于隐马尔可夫的用户体验质量模型,建立了基于随机模型的评价方法,对未来的研究方向进行了展望.

关键词 用户体验质量;统计学模型;心理学模型;人工智能模型;随机模型

中图法分类号 TP393

DOI号: 10.3724/SP.J.1016.2012.00001

Survey on Models and Evaluation of Quality of Experience

LIN Chuang HU Jie KONG Xiang-Zhen

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084)

Abstract With the advent of the information age, new services of high diversity emerge continuously. The demand of people for services is growing higher. The service providers should improve the Quality of Experience of the users, in order to keep users for their services. Therefore Quality of Experience has become one of the important topic in the academic world as well as an ultimate business metrics in the industry. In this paper, we provide a complete overview on evaluation models and methods of Quality of Experience. We summarize the influence factors, qualification methods and three evaluation methodologies, and focus on the concrete models and evaluation methods based on several disciplines. Then we provide a model of Quality of Experience for video streaming based on hidden Markov model, propose an evaluation method of QoE based on stochastic model and indicate the directions of future research on Quality of Experience.

Keywords QoE; statistical models; psychological models; artificial intelligence models; stochastic models

1 引 言

随着计算机技术与网络技术的飞速发展,大量多样性的服务涌现出来,服务提供商之间的竞争日

益激烈.为了能在激烈的竞争中胜出,服务提供商必须确保自己提供的服务得到用户的认可.因此,他们迫切需要一种以用户认可程度为标准的服务的评价方法.目前,服务质量(Quality of Service, QoS)是最广泛采用的服务的度量标准.QoS评价指标主要

包括网络的吞吐率、时延、丢包率、抖动、误码率等。显然这些指标仅仅反映了服务技术层面的性能,甚至仅仅是网络传输层面的性能,忽略用户主观因素,因此它不能直接反映用户对服务的认可程度。用户体验质量(Quality of Experience, QoE)是一种以用户认可程度为标准的服务的评价方法。它综合了服务层面、用户层面、环境层面的影响因素,直接反映了用户对服务的认可程度。

目前关于 QoE 的研究主要分为以下几个部分: QoE 的定义及影响因素、QoE 量化方法、QoE 评价指标、QoE 评价方法与模型以及基于 QoE 的网络管理及服务的管理与优化。由于 QoE 的影响因素包含服务、用户、环境 3 个层面,而且不同用户对同一服务的要求也不相同。因此,这就给 QoE 的研究带来了非常大的挑战。首先, QoE 的研究需要涉及计算机科学、心理学、社会学、统计学等多学科的相关知识和理论;其次, QoE 的部分影响因素难以量化,如用户的主观因素及环境的影响因素;最后,对不同的服务难以统一建模处理。鉴于以上几个难点,目前关于 QoE 的研究呈现出以下特点:(1) 由于用户的主观因素(如期望、体验经历)和环境因素对 QoE 的影响难以量化评价,目前的研究方法大多都集中在对服务的研究上;(2) QoE 的评价指标分布在不同层次上,而且多数文献更多地从网络层、应用层去讨论评价指标,而没有从用户或者服务的角度出发来研究;(3) 对 QoE 评价指标的讨论还是基于具体的服务,不同服务的评价指标往往差异较大;(4) QoE 的理论还不成熟,不同的研究人员试图从不同的角度对 QoE 做出量化和评价,因此涌现出了许多不同的量化方法和评价方法;(5) 大多数关于 QoE 评价的研究集中在视频和音频业务中。

鉴于以上现状,本文试图在现有文献的基础上,给出一个较为全面的 QoE 定义,分析概括 QoE 的影响因素,总结现有的 QoE 的量化方法和评价方法,并且分析比较其优缺点。最后在这些评价方法的基础上,以视频流媒体为背景,提出了随机模型的 QoE 评价方法。

本文第 2 节对 QoE 的定义和影响因素进行论述;第 3 节总结和讨论现有的 QoE 的量化方法,并且简要介绍 QoE 评价方法学的分类;第 4 节讨论 4 种基于不同学科的 QoE 评价方法,并在对这些评价方法与模型进行分析和比较的基础上,讨论如何建立基于随机模型的 QoE 评价方法;最后对全文进行总结,并提出下一阶段的研究方向。

2 QoE 的定义及影响因素

现有的文献给出了许多关于 QoE 定义和影响因素的讨论,但是没有形成一个统一的论述。下面我们将在这些文献的基础上,给出一个更为全面的 QoE 定义,并且对 QoE 的影响因素进行概括和分析。

2.1 QoE 的理解与定义

QoE 最初被理解为用户对提供给 OSI 模型不同层次的 QoS 机制整体感知的度量^[1]。例如在文献[2]中, QoE 被定义为:确保通过 IP 网络平稳传输音频和视频质量的 QoS 机制的整体性能。因此,可以将 QoE 理解为在用户角度对 QoS 机制的评价。之后,国际电信联盟等对 QoE 的定义进行了扩展,它不再是对 QoS 机制进行评价,而是对整个服务进行评价。国际电信联盟对 QoE 的定义为:终端用户对应用或者服务整体的主观可接受程度^[3]。文献[4]中, QoE 被定义为:从用户角度体验到的一个系统的整体性能。

分析上述几个定义,可以总结出 QoE 是一种主观感受,是用户在与服务或者应用交互的过程中,由用户产生的对服务的一种主观感受。因此 QoE 有两个直接的影响因素:用户和服务。但是用户和服务交互的过程肯定会处在一定的客观环境中,而且客观环境对用户和服务的交互过程具有较大的影响。因此我们有必要在 QoE 定义中阐明这一因素。

综上所述,我们将 QoE 定义为用户在一定的客观环境中对所使用的服务或者业务的整体认可程度。

2.2 QoE 的影响因素

QoE 的影响因素的研究对于 QoE 的评价至关重要。因为 QoE 评价的基本目标就是要从已知或者容易测量的影响因素来预测难以直接测量的 QoE。下面,我们先对 QoE 的影响因素做一个较为全面的讨论。文献[5-7]详细讨论了 QoE 的影响因素,我们在这些文献的基础上,结合 2.1 小节中的讨论,将 QoE 影响因素分为 3 个层面:服务层面、环境层面及用户层面,如图 1 所示。

服务层面的影响因素又包括网络(或者传输)层、应用层及服务层的参数^[5]。传输层的参数反映网络传输的状况,如延迟、带宽、丢包率、误码率、抖动等。应用层的参数反映没有经过传输的服务的性能,包含了 OSI 模型中会话层、表示层、应用层对服务的影响,如 IPTV 服务中的内容分辨率、编解码类型等。服务层的参数确定了通信的语义、内容、优先级、

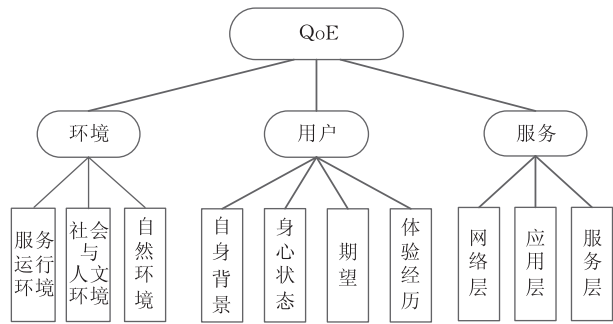


图 1 QoE 的影响因素

重要性以及定价,如服务层的配置(内容类型、服务的应用级别)及质量保证。环境层面的影响因素包括自然环境(如光照条件、噪声的大小、环境的固定或移动),人文与社会环境(如社会观念、文化规范)以及服务运行环境(包括软硬件环境)等。用户层面的因素包括用户的期望、体验经历、用户体验时所处的身心状态和自身背景(如年龄、性别、受教育程度、价值观念等)。文献[8-9]对这些影响因素如何相互作用进行了分析和讨论,这里不再赘述。

3 QoE 的量化与评价方法学分类

为了更好地研究 QoE 与其影响因素的关系,我们需要将 QoE 进行量化,也就是要建立 QoE 到实数的映射关系。这样就可以研究 QoE 关于其影响因素的函数模型,从而更准确地反映 QoE 与其影响因素的关系。在 3.1 节中,我们将介绍几种常用的量化方法。在将 QoE 量化之后,我们就需要研究如何获得 QoE 关于其影响因素的函数模型。因此在 3.2 节我们将介绍几类基本的思路,也就是几类 QoE 的评价方法学。

3.1 QoE 的量化

在文献[10]中,Stevens 将测量(或者量化)的水平分为 4 个等级,每一等级的测量水平都与一量表对应,即类别量表、顺序量表、等距量表及等比量表。这 4 种量表的测量水平依次增强,等比量表是最高级别的量表,它具有相等的单位和绝对零点,而且测量值之间的比值也是有意义的,因此它并不符合 QoE 的特征,一般不用等比量表来对 QoE 进行量化。通常采用类别量表、顺序量表、等距量表对 QoE 进行量化。

文献[11-12]采用类型量表的方法,将 QoE 分为可接受(Acceptable)和不可接受(Unacceptable)两个级别。这种量化方法(以下称为两类别法)不够

精细,但是方便用户进行评价,因为用户对这两种级别具有较大的辨识度。

文献[13]提出了用成对比较法(paired comparison)对 QoE 进行量化,属于等距量表法。成对比较法的大致思想为:首先准备 N 种样本,让用户对这 N 种样本进行两两比较,即共进行 $M = \binom{N}{2}$ 次比较,记录每次比较的结果。运用 Bradley-Terry-Luce (BTL) 模型^[14]处理 M 次比较得出的数据,从而给出每种样本的评分。

目前较广泛采用的是国际电信联盟(ITU)建议的“平均评估分值^[15]”(MOS,即 Mean Opinion Score)。它将 QoE 的主观感受分为 5 个层次,如表 1 所示,此量化方法较为细致地描述了用户的主观感受,是一种顺序量表法。

表 1 平均评估分值

| MOS | QoE | 损害程度 |
|-----|-----|---------|
| 5 | 优 | 不能察觉 |
| 4 | 良 | 可察觉但不严重 |
| 3 | 中 | 轻微 |
| 2 | 次 | 严重 |
| 1 | 劣 | 非常严重 |

表 2 中给出了 3 种方法的比较。两类别法虽然简单而且准确性高,但是它存在一些明显的缺陷。将 QoE 仅仅分为可接受和不可接受两个类别,显然不能满足用户以及供应商对业务及服务评价的要求。虽然成对比较法比 MOS 方法更加精确,但是它要求的工作量较大。比如一个业务有 20 个测试样本,用 MOS 方法只需要做出 20 次判断,而成对比较法需要做出 190 次判断。综合上述一系列的原因,MOS 方法仍然是目前最广泛使用的方法。

表 2 3 种方法之间的比较

| 量化方法 | 优点 | 缺点 |
|-------|---|--|
| 两类别法 | 简单,方便用户打分,准确性高 | 对用户的主观感受的描述不够细致 不能对用户给出的评价结果进行一致性检验 |
| MOS | 对用户主观感受的描述较为细致 | 几个层次之间的界限较为模糊,不同的人对这几个层次的理解不同 不能对用户给出的评价结果进行一致性验证 |
| 成对比较法 | 质量等级的数量可以变化 方便用户给出判断 可以对用户给出的评价结果进行一致性验证,准确性高 | 用户或测试人员完成评价需要的工作量较大 |

3.2 QoE 的评价方法学分类

现有的关于 QoE 的研究一般集中在音频、视频业务中. 下面我们就以音频、视频业务为例, 介绍几类 QoE 评价方法学. 按照是否有用户直接参与评价及是否给出 QoE 与其影响因素之间的关联模型可以将现有的评价方法分为主观评价方法、客观评价方法及主客观结合的方法, 其中主客观结合的方法又可以称为伪主观评价方法(在一些文献中, 伪主观的评价方法和客观评价方法都被称为客观的评价方法). 这种评价方法分类和第 4 节中根据不同学科的评价方法分类相比, 层次更高, 也更加抽象. 因此, 为了避免混淆, 本文把这种角度的评价方法分类称为评价方法学分类.

主观评价方法是让用户直接对所使用的业务做出评价, 因此它最准确, 一般可以作为评价其它评价方法的标准. 但是由于主观评价方法所需的测试环境较为严格、不可以实时地应用、可移植性差, 而且消耗的代价较大, 因此并未得到广泛的应用. 客观评价方法通过计算输出序列相对于输入序列的失真程度, 来对业务做出评价. 因此客观评价方法并没有考虑用户的因素, 不能直接、准确地反映用户对业务的评价. 现有的一些客观评价方法做出了一些改进, 将人类质量感知模型运用到对业务的评价中, 评价结果有了很大的改善. 例如, 峰值信噪比(Peak Signal to Noise Ratio, PSNR)是一种传统的客观的视频质量的评价方法, 它没有考虑人的视觉掩蔽的现象, 也就是说, 每一个像素的错误都会影响最后的评价结果, 而实际上这样的错误是完全不能被人眼觉察到的^[16]. MPQM^[16-18]将人类视觉特征考虑在内, 更准确地反映了用户对业务的主观体验. 伪主观评价方法将主观评价方法和客观评价方法结合起来. 它的基本思路是分析给定业务的特性, 确定其重要的 QoS 参数, 通过改变这些 QoS 参数, 得到评价的样本. 然后选择一定的测试人员进行评价, 将最终的结果与这些 QoS 参数关联起来, 这样就得到了 QoE 与 QoS 的关联模型. 从而可以用这个关联模型对此业务进行评价. 下面详细介绍这几类评价方法.

3.2.1 主观评价方法

文献[15]以语音业务为背景, 最先提出了主观评价的方法. 主观评价的方法就是要直接从用户感知出发, 对业务进行评价. 为了获得用户的感知, 就需要选择一定数量的测试人员, 让他们来使用给定的业务, 并按照 MOS 方法给出分数. ITU 陆续提出了一些关于音频、视频的主观评价方法. 例如 ITU-T

Rec P. 830、P. 835 和 P. 840 给出了语音业务的主观评价方法, ITU-T Rec BT 500 给出了视频业务的主观评价方法.

下面以 ITU-T Rec P. 800 中建议的语音评价的 ACR 测试方法为例来简介主观评价方法. ACR 测试方法分为 4 个部分, 首先是记录测试源, 这一部分主要考虑记录的环境、记录的介质、发送系统、语音材料的要求、记录的过程、说话人的选择等; 其次是链路状态的选择, 这一部分主要考虑语音输入、参考条件等; 再次是实验的设计; 最后是听力测试的部分. 这一部分主要考虑听力测试的环境、听力测试的系统、听力测试人员的选择、评分的标准以及统计分析和汇报结果.

综上所述, 主观评价的方法虽然可以直接、准确地反映用户对业务的主观体验, 但是由于它需要考虑太多的因素、实施步骤较复杂、代价大, 而且没有建立 QoE 关于其影响因素的函数模型, 不可以实时地应用, 不能移植, 所以主观评价方法没有得到广泛的应用. 但是, 由于它具有准确、直接的特点, 所以可以用它来对其它的评价方法进行验证和矫正.

3.2.2 客观评价方法

客观评价方法是过度量输出序列相对于原始序列的失真程度来对业务进行评价. 一般来说, 视频和音频业务的 QoE 客观评价方法根据其特定的应用可以选择以下 3 种方法^[4]:

(1) 全参考. 通过对系统输出序列(测试序列)和输入序列(参考序列)的比较获得质量评价. 这种方法结果精确度高, 但计算复杂度较大. 这种方法不能应用在分组交换网络中. 国际电信联盟使用这种方法来获得有线数字电视业务中用户对视频质量的要求^[19]. 文献[20]对原序列和语音序列的比较是在每一帧之间进行的, 这样就需要将这两个序列在时间上进行对准. 如果系统中存在时延抖动的话, 这样的比较就会存在一些问题.

(2) 部分参考. 在这种方法中, 选择输入、输出序列的部分参数进行比较从而获得质量评价. 如文献[21], 作者仅仅使用参数丢包率和延迟来研究它们与质量评价的关系.

(3) 无参考. 仅仅输出序列用来进行质量评价.

目前, 一些研究者[4, 22]试图通过研究 QoS 与 QoE 之间的关联模型来对 QoE 进行评价. 文献[4]将 IPTV 的 QoE 分为两个组成部分: (1) 与网络相关的 QoE, 用 QoE_N 表示; (2) 与用户相关的 QoE, 用 QoE_U 表示. 与网络相关的 QoE 主要由网络延迟、丢

包、抖动这些因素决定,它的函数模型如式(1)所示.与用户相关的 QoE 主要由视音频质量、视音频的同步及转换频道所用的时间决定,它的函数模型如式(2)所示.但是这篇文献没有给出如何用 QoE_N 和 QoE_U 来计算最终的 QoE.

$$QoE_N = \frac{1}{(Delay + K_1 Jitter) e^{packetLoss}} \quad (1)$$

式中: $Delay$ 为网络延迟; $Jitter$ 为网络抖动; K_1 为调节网络延迟和抖动之间的相对重要程度; $packetLoss$ 为丢包率.

$$QoE_U = \frac{\log_{10}(v_q)}{t_{zap} + K_1 t_{sync}} \quad (2)$$

式中: v_q 为视频质量; t_{zap} 为转换频道所用的时间; K_1 为调节 t_{zap} 和 t_{sync} 之间的重要程度; t_{sync} 表示视音频之间的同步误差.

文献[22]中讨论了 IPTV 业务中视频的 QoE 和 QoS 之间的关系,并给出了视频质量的 QoE 与 QoS 之间的关联模型,如式(3)所示.值得注意的是,它给出的并不是 IPTV 业务的 QoE 评价模型,而是 IPTV 业务中视频的 QoE 评价模型.

$$QoE_v = Q_r \cdot (1 - QoS(X))^{\frac{QoS(X) \cdot A}{R}} \quad (3)$$

式中: Q_r 为终端分辨率; $QoS(X)$ 为标准化的 QoS 的值,如式(4)所示; A 为预订服务的等级; R 为代表帧的结构.

$$QoS(X) = K \{L \times Wl + U \times Wu + J \times Wj + D \times Wd + B \times Wb \dots\} \quad (4)$$

式中: K 为整个 QoS 决定因子,由 IPTV 业务的类型而定; L 为丢包率; Wl 为丢包率的权重; U 为突发性; Wu 突发性的权重; J 为抖动; Wj 为抖动的权重; D 为延迟; Wd 为延迟的权重; B 为带宽; Wb 为带宽的权重;

综上所述,客观评价具有简单、实用的特点.所以它是目前 QoE 评价方法研究中重要的一部分.但值得注意的是,为了更准确地反映用户对业务或者服务的主观感受,客观评价方法需要考虑人类的感知模型,这就给 QoE 的客观评价带来了一定的难度.

3.2.3 伪主观评价方法

伪主观的评价方法^[23-24]结合了主观评价方法和客观评价方法的优点.它既有主观评价方法与用户感知一致的特点,又具有客观评价方法简便、可实时应用、可移植的特点.伪主观评价方法的基本步骤如图 2 所示^[11].

(1) 分析业务类型,确定此业务的 QoE 相应指标.如一个 IPTV 业务的 QoE 的指标包括音频质

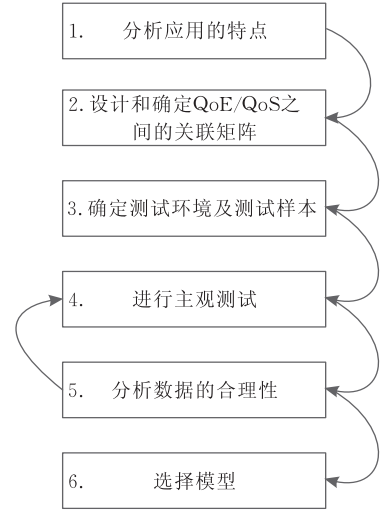


图 2 伪主观评价方法的步骤

量、视频质量以及音频视频之间的同步关系等.

(2) 确定决定这些 QoE 指标的相应 QoS 的参数. 确定 QoE 指标之后,需要找出影响每个评价指标的全部 QoS 参数或者关键的 QoS 参数. 这种关系可以通过 QoE 评价指标与 QoS 参数的关联矩阵表示出来.

(3) 准备测试的环境以及测试的样本. 测试环境可以有多种选择,测试环境可以选择在实验室的环境中^[11-12],也可以选择真实的生活环境中^[25]. 这两种选择各有利弊. 前者因为脱离了真实的生活环境,测试人员对服务的体验可能和真实环境中对服务的体验存在差距,不能准确地反映真实的生活环境中用户对服务的体验质量.但是由于在实验室有相关人员的监督,态度可能会相对认真.后者则会出现相反的情况.

(4) 挑选一定数量的测试人员进行业务评价,获取数据.测试人员应该从此业务的真实用户中选取,并且注意从不同年龄、不同性别、不同教育背景的用户群体中选择.文献[13]提出的一种获取数据的框架,可以很好地解决上面出项的问题.它将数据获取的任务放在互联网的环境中,并利用成对比较法,让广大的互联网用户参与测试,根据用户提供数据的一致性程度给予其相应的奖励.但是它要求的工作量较大.

(5) 分析数据的合理性,进行数据筛选.文献[26]给出了一个数据筛选的方法.在数据筛选之后,还要对筛选后的数据进行分析,判断其是否在统计上有意义.如果没有的话,还需要进一步进行数据获取的步骤.

(6) 利用得到数据,采用合适的数据模型将

QoS 与 QoE 关联起来. 数学模型可以有多种选择. 如采用基于统计学的判别分析方法^[11]、回归分析^[27-29]、人工智能中的支持向量机^[12,30]、决策树^[12,25,31]、粗糙集的方法^[32].

伪主观的评价方法由于有人工智能以及统计学等学科作为理论支撑,而且可以实时运用、准确性较高,目前受到研究人员越来越多的关注和运用.但是它也有一些不足之处,如需要获取足够的数据,进行较为复杂的模型训练等.

4 几种 QoE 的评价方法

在第 3 节中,我们介绍了 3 类 QoE 评价方法学.这 3 类评价方法学给出了 3 类基本的 QoE 评价思路.本章在这 3 类评价思路的基础上,讨论几种具体的基于不同学科的 QoE 评价方法,并且对这些评价方法进行优缺点分析.按照评价方法所采用的学科知识,可以将评价方法分为基于统计学的评价方法、基于心理学的评价方法及基于人工智能学的评价方法.本章将详细讨论这些评价方法.

4.1 基于统计学的评价方法

将统计学的方法应用于 QoE 评价,最主要的特点是可以进行 QoE 评价指标之间的相关性分析及 QoE 评价指标与 QoE 相关性的分析.这样就可以降低 QoE 评价问题的复杂度,简化 QoE 评价模型.另外,在 QoE 评价模型的选择时,可以采用统计学中的判别分析、回归分析方法.这一小节将详细讨论这些统计学的方法.

4.1.1 QoE 评价指标之间相关性的分析

在 QoE 评价问题的研究中,为了能对 QoE 有一个比较全面、完整的把握与认识,往往需要尽可能多地收集影响 QoE 的多个变量.但是这就带来了两个较大的问题:(1)收集数据的工作量变得十分庞大;(2)对研究 QoE 评价模型带来了很大的挑战.而且这些选取的变量往往存在着相关性,也就是不经过分析处理,直接使用这些变量就会做很多不必要的工作.因此可以利用统计学中因子分析^[33]的方法来对这些 QoE 的变量进行分析,从而简化问题和避免做一些不必要的工作.下面将介绍因子分析的基本步骤.

首先,对这些 QoE 的变量进行 KMO 检验. KMO 检验的目的是为了分析是否有必要对这些变量进行因子分析. KMO 统计量如式(5)所示,取值范围在 0 与 1 之间.一般认为, KMO 大于 0.6 以上,

就可以进行因子分析.

$$KMO = \frac{\sum \sum_{i \neq j} r_{ij}^2}{\sum \sum_{i \neq j} r_{ij}^2 + \sum \sum_{i \neq j} a_{ij}^2} \tag{5}$$

式中: r_{ij}^2 为两变量间的简单相关系数; a_{ij}^2 为两变量间的偏相关系数.

简单相关系数计算如(6)所示.偏相关系数可以通过求解所有变量的相关矩阵的逆矩阵来获得.

$$r = \frac{\sum_{i=1}^n (X_i - E(X))(Y_i - E(Y))}{\sqrt{\sum_{i=1}^n [X_i - E(X)]^2} \sqrt{\sum_{i=1}^n [Y_i - E(Y)]^2}} \tag{6}$$

其次,利用主成分分析法^[33]进行因子提取,并且进行旋转.主成分分析法(Principal Components Analysis, PCA)是一种对数据集进行降维的方法.它通过一个线性变换将变量映射为一组因子.然后依次取方差最大的前 m 个因子,并确保这 m 个因子累计贡献率达到 85%~95%.对因子载荷矩阵实行正交旋转的目的是使其结构简化,确保容易对每个因子进行恰当的解释.

最后,进行结果处理.首先要对得到的每个因子进行命名.因子的命名应能解释其代表的含义.其次还要计算每个因子的值,以便以后利用这些因子进行问题的研究.

文献[27]对 QoE 的主观参数进行了因子分析.它将主观参数分为质量层面和情绪层面.由于篇幅所限,本文仅介绍质量层面主观参数的分析.首先对质量层面的主观参数进行 KMO 检验,如表 3 所示.

表 3 QoE 质量层面的主观参数

| 质量层面 | KMO=0.75 |
|--------|----------|
| 内容 | 5 个级别 |
| 合适性 | 5 个级别 |
| 音频质量 | 5 个级别 |
| 图像质量 | 5 个级别 |
| 流畅性 | 5 个级别 |
| 视/音频同步 | 5 个级别 |

由于 KMO 值大于 0.6,所以可以用 PCA 方法进行数据降维.它的主要过程是,首先将这些数据进行标准化处理;其次利用 PCA 方法和方差极大旋转方法进行处理;最后对这些主成分进行可靠性分析,得出的最终结果如表 4 所示.

这样,通过 PCA 方法,就可以将质量层面的参数用两个不相关的参数,即空间质量(spatial quality)和时间质量(temporal quality)代替.由于经过变换之后,质量层面的参数从 6 个变成了 2 个,所以大大降低了原问题的复杂度.

表 4 质量层面的 PCA 分析

| 第 1 个主成分(空间质量) | | |
|----------------|-------------------|-------|
| 方差解释 | 可靠性 | |
| 43.26% | 克隆巴哈系数=0.73 | |
| 问题条目 | 因子载荷 | 公因子方差 |
| 内容 | 0.83 | 0.71 |
| 音频质量 | 0.72 | 0.57 |
| 合适性 | 0.72 | 0.52 |
| 图像质量 | 0.56 | 0.66 |
| 第 2 个主成分(时间质量) | | |
| 方差解释 | 可靠性 | |
| 18.91% | $r=0.28, p=0.000$ | |
| 问题条目 | 因子载荷 | 公因子方差 |
| 流畅性 | 0.85 | 0.75 |
| 视/音频同步 | 0.65 | 0.53 |

4.1.2 QoE 评价指标与 QoE 之间的相关性分析

由于 QoE 是一个多指标评价的问题,在 QoE 评价问题的研究中,势必会遇到判断提出的一个评价指标对 QoE 是否有影响或者影响多大的问题.解决这个问题,可以利用统计学中的相关性分析^[33]和方差分析^[33].

在研究 QoE 评价指标与 QoE 之间的相关性时可以用到 Pearson 相关模型.相关系数 r 定义如式(6)所示.从实践角度而言,相关系数 r 仅是根据样本数据而不是总体数据计算出来的,存在着抽样误差.因此,不能根据 r 值的大小来判断两变量间的相关关系,必须进行显著性检验.

文献[27]给出了 QoE 的 5 个评价指标与 QoE 之间的相关性分析,如表 5 所示.

表 5 QoE 的 5 个评价指标与 QoE 之间的相关性

| 主成分 | 皮尔逊相关系数 | 显著性 |
|------|----------|-----------|
| 空间质量 | $r=0.82$ | $p=0.000$ |
| 时间质量 | $r=0.43$ | $p=0.000$ |
| 满意程度 | $r=0.73$ | $p=0.000$ |
| 兴趣 | $r=0.37$ | $p=0.000$ |
| 关注级别 | $r=0.12$ | $p=0.000$ |

除此之外,文献[27-28]还利用单因素方差分析法(ANOVA)来分析 QoE 与位置之间的关系.在这里就不再做详细讨论.

4.1.3 QoE 统计学模型与模型验证

在进行了 QoE 评价指标之间的相关性分析和 QoE 评价指标与 QoE 之间相关性的分析之后,就需要给出 QoE 的评价模型,也就是给出将 QoE 的这些评价指标映射为最终 QoE 的函数模型.可以采用统计学中回归分析^[27-29]和判别分析^[11,33]的方法来解决这个问题.

若将 QoE 映射为一连续的值,则应该采用回归分析的方法.QoE 评价模型一般采用线性回归模

型、指数回归模型以及对数回归模型^[27-29].模型的选取与服务或业务的类型有关.文献[28]指出,实时交互的业务适合用指数回归模型进行分析,而传统的网页浏览的业务则适合用对数回归模型进行分析.回归分析的基本研究方法^[33]是:首先,做散点图以观察曲线的形状.若坐标轴相关点呈团状分布,则表示两变量没有任何关系.如果两变量有关系,则它们可能是最简单的直线线性关系或非线性关系.非线性关系又分两种:一种是本质线性关系或拟线性关系,可转换成线性关系,用最小二乘法求出相关系数;另一种是本质非线性关系,不可转换成线性关系,仅能用迭代方法或分段平均值方法求解.

若将 QoE 映射为离散的值,则 QoE 的评价问题就等价为一个分类问题,这时可以采用判别分析法^[11].判别分析的具体步骤为:(1)选择研究对象的指标,即 QoE 的评价指标;(2)收集数据,得到训练样本,并利用训练样本,给出判别函数;(3)对判别函数进行验证分析;(4)输出结果,根据研究对象的实际情况分析输出结果,得出结论.根据上述步骤就可以得到 QoE 的评价模型.

在得到 QoE 的函数模型以后,往往需要对其进行验证.常用的验证方法有自身验证、外部数据验证、样本二分法以及交互验证等.

4.2 基于心理学的评价方法

韦伯-费希纳定律^[34]是心理学领域的一个重要原理,它描述了物理刺激的程度和它被人感受的强度之间的关系,适用于中等强度的刺激.在许多场景,这种关系都呈现一种对数的特征.可以利用这一性质来研究 QoE 的评价^[35].

4.2.1 韦伯-费希纳定律

韦伯-费希纳定律是由德国物理学家、心理物理学创始人费希纳在韦伯研究的基础上,提出的一个描述连续意义上心理量与物理量关系的定律.

1834 年,德国的生理学家韦伯首次提出人类感觉系统中的“最小可觉差”的理论.最小可觉差是指当物理刺激程度的变化超过了它实际刺激程度的一定比例时,感觉系统能够区分出变化.这个比例是个常数,如式(7)所示,其中 ΔS 为物理刺激的变化量, S 为物理刺激, k 为常数.大量的实验表明,当我们手中的物体的重量增加接近 3% 时,我们可以感觉到重量的增加,和初始物体重量的绝对值无关.

$$\frac{\Delta S}{S} = k$$

(7)

1860 年,费希纳在韦伯研究的基础上提出了一

个假定:把最小可觉差作为感觉量的单位,即每物理刺激增加一个差别阈限,心理量增加一个单位.上述假定用微分的形式表示,如式(8)所示,其中 dP 表示感觉的变化,而用 dS/S 表示物理刺激的相对变化.

$$dP = k \frac{dS}{S} \quad (8)$$

由式(8)可以推出式(9),其中 S_0 指可被感觉到的最小的物理刺激的程度.

$$P = k \cdot \ln \frac{S}{S_0} \quad (9)$$

韦伯-费希纳定律是人类感觉系统的非常基础的一个原理,适用于生理和精神层面,比如人的视觉(恒心亮度在对数级别进行度量)、听觉(声音的强度用分贝进行度量)、味觉、嗅觉、触觉、时间感知等^[35].因此,可以利用韦伯-费希纳定律来进行 QoE 的相关研究.

4.2.2 韦伯-费希纳定律与 QoE 评价

文献[35]通过两个例子描述了韦伯-费希纳定律在 QoE 评价问题中的意义.

第 1 个例子讨论了语音业务的 QoE 与比特率的关系,如图 3 所示.它将比特率作为唯一变化的因素,变化范围从 2.4 kbps 到 24.8 kbps,并且利用 PSQA(Pseudo-Subjective Quality Assessment)方法计算出了语音业务的 MOS 分数.由于图 3 中横坐标表示比特率,并且经过了对数处理,因此可以看出语音业务的 QoE 与比特率成对数的关系.如果将比特率理解为物理刺激,而语音业务的 QoE 理解为人的感觉,则运用韦伯-费希纳定律,可以得出式(10).可以看出利用韦伯-费希纳定律得出的结论和实验数据相符,也就是说利用韦伯-费希纳定律可以直接解释语音业务的 QoE 与比特率之间的关系.

$$dQoS \propto QoS \cdot dQoE \quad (10)$$

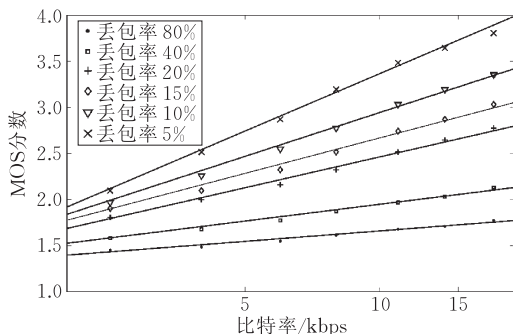


图 3 语音业务的 QoE 与比特率之间的对数关系

第 2 个例子讨论了 QoE 评价的 IQX 假设^[35-36]. IQX 假设将 QoE 描述为一个网络级别的 QoS 参数(如丢包率、延迟等)的函数,如式(11)所示.在这里,

同样以语音业务的 QoE 评价为例,控制其它因素不变,而只考虑丢包率,并利用 PESQ 算法,给出其 QoE 的 MOS 分数,结果如图 4 所示.图中纵坐标 QoE 进行了对数处理.可以用指数回归来描述图中给出的结果,如式(12)所示.

$$\frac{dQoE}{dQoS} = -\beta(QoE - \gamma) \quad (11)$$

$$QoE = 3.0819 \cdot e^{-4.6446 p_{loss}} + 1.07 \quad (12)$$

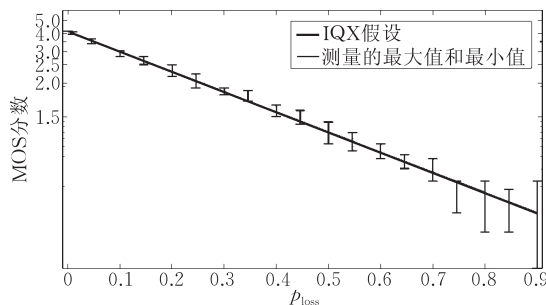


图 4 语音业务的 QoE(MOS 分数)与丢包率(p_{loss})的关系

很遗憾,在这里就不能将丢包率作为物理刺激, QoE 作为人的感觉,利用韦伯-费希纳定律来直接解释这一现象,但是这同韦伯-费希纳定律有着一定的联系.

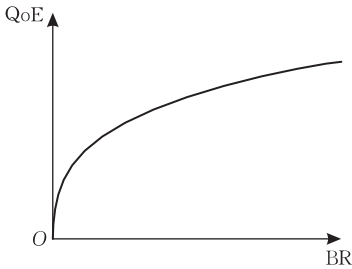
IQX 假设的前提是用户对 QoS 的敏感度跟当前 QoE 有关.也就是说,一定程度的 QoS 的变化对最终 QoE 的改变是不同的,与当前用户的 QoE 有关,如式(11)所示.文献[35]通过令 $QoE' = QoE - \gamma$,将式(11)变形为式(13),从而发现了一个规律:即和韦伯-费希纳定律相比, QoE 与 QoS 互换了位置.这是因为韦伯-费希纳定律是基于这样一个假设,物理刺激的改变量需要是当前物理刺激程度的一定比例.而 IQX 的假设是对一定的物理刺激的改变量,人感觉的变化是与当前人感觉所处的级别有关.

$$dQoE' \propto QoE' \cdot dQoS \quad (13)$$

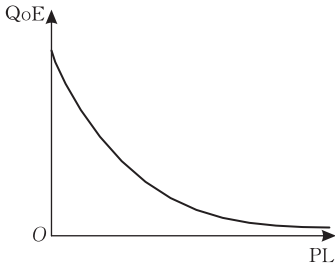
韦伯-费希纳定律和 IQX 假设存在以下关系,如图 5 所示.首先,韦伯-费希纳定律和 IQX 假设分别适用于不同的 QoS 与 QoE 之间的关系.其次,当技术层面的参数较小时,用户感觉随着技术层面的参数变化较快,当用户体验的技术层面的参数较大时,用户感觉随着技术层面的参数变化较慢;最后, QoE 与 QoS 之间总存在对数的关系,或者 QoE 是 QoS 的对数函数,或者 QoS 是 QoE 的对数函数.但是如何从心理学的角度对 IQX 假设进行解释,它和韦伯-费希纳定律有什么更深层次的关系,现有文献中并没有给出,还需要进一步的研究和探索.

4.3 基于人工智能学的评价方法

本文将基于人工智能的评价方法分为两类,即



(a) 语音业务的QoE与比特率(BR)之间对数的关系



(b) 语音业务QoE与丢包率(PL)之间指数的关系

图 5 语音业务 QoE 与 QoS 参数之间的关系

基于模糊层次分析法^[37] (Fuzzy Analytic Hierarchy Process, FAHP) 的评价方法和基于机器学习^[38] 的评价方法, 下面分别对这两种评价方法进行详细的讨论.

4.3.1 基于 FAHP 的 QoE 评价

层次分析法 (Analytic Hierarchy Process, AHP) 是由美国运筹学家 Thomas 首次提出, 它是用来解决多指标综合问题的一种方法. 它结合定性和定量的方法, 同时减少了主观因素的影响, 确保系统和模型的合理性. 它提供了一个科学和合理的评价分析方法. 模糊层次分析法结合 AHP 和模糊逻辑, 在一定程度减少了主观因素对评价的影响. 下面介绍 FAHP 中的一些主要的定义和定律^[32,37].

定义 1. 假设矩阵 $A = (a_{ij})_{n \times n}$ 中 $0 \leq a_{ij} \leq 1$, 则 A 是一个模糊矩阵.

定义 2. 如果模糊矩阵 $A = (a_{ij})_{n \times n}$ 中 $a_{ij} + a_{ji} = 1$, 则 A 称为模糊互补矩阵.

定义 3. 如果模糊互补矩阵 $A = (a_{ij})_{n \times n}$ 和任意的整数 k 满足 $a_{ij} = a_{ik} - a_{jk} + 0.5$, 则 A 称为模糊一致性矩阵.

假设模糊互补矩阵 A , 计算矩阵每行的和, 如式(14)所示:

$$r_i = \sum_{k=1}^n a_{ik} \tag{14}$$

进行如式(15)所示的数学变换:

$$b_{ij} = \frac{r_i - r_j}{2(n-1)} + 0.5 \tag{15}$$

这样一个模糊矩阵 $B = (b_{ij})_{n \times n}$ 就建立了. 计算

矩阵 B 每行的和, 并进行标准化, 就获得了权重向量 $w = (w_1, w_2, \dots, w_n)$. 计算公式如式(16)所示.

$$w_i = \frac{\sum_{j=1}^n b_{ij} + \frac{n}{2} - 1}{n(n-1)}, i = 1, 2, \dots, n \tag{16}$$

利用模糊层次分析法解决问题的基本步骤如下所示.

(1) 根据决策的条件建立评价问题的层次模型. 一般来说, 一个层次模型包括目标层, 指标层以及方案层. 如果有必要的话, 还可以在指标层下面加入子指标层.

(2) 建立模糊互补的判断矩阵. 评价专家和决策者给出指标之间的两两比较的值, 可以采用如表 6 所示的 0.1~0.9 标度.

(3) 计算权重向量. 通过上面所述的方法, 就可以将模糊互补矩阵转换成模糊一致矩阵, 每一层的权重向量就可以得到.

(4) 根据步骤 3 的结果, 计算出最终的值.

表 6 0.1~0.9 标度

| 标度 | 说明 |
|----------|---|
| 0.5 | 两元素同等重要 |
| 0.6 | 一元素比另一元素稍微重要 |
| 0.7 | 一元素比另一元素明显重要 |
| 0.8 | 一元素比另一元素重要得多 |
| 0.9 | 一元素比另一元素极端重要 |
| 0.1, 0.2 | 若元素 a_i 与元素 a_j 相比较得到判断 r_{ij} , 则元素 |
| 0.3, 0.4 | a_i 与元素 a_j 相比较得到的判断为 $r_{ji} = 1 - r_{ij}$ |

文献[32]根据 FAHP 方法, 建立了短信服务 QoE 的评价体系结构, 如图 6 所示. 它首先将短信服务的 QoE 评价指标归纳为可访问性、即时性、完整性、内容质量及可持续性. 其次, 找出了与每一个 QoE 评价指标关联的 QoS 参数. 这样就建立了短信服务 QoE 的评价体系结构.

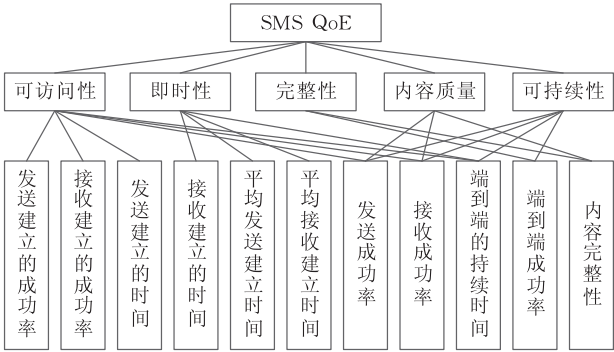


图 6 短信服务的 QoE 评价体系结构

在建立 QoE 评价体系结构之后, 还需要确定与短信服务 QoE 相关的各个 QoS 参数的权重. 这时, 需要首先确定 QoE 评价指标两两比较的值以及每

一个 QoE 评价指标下相关的 QoS 参数两两比较的值,然后按照上面所述的方法,就可以得出的最后结果,如图 7 所示.

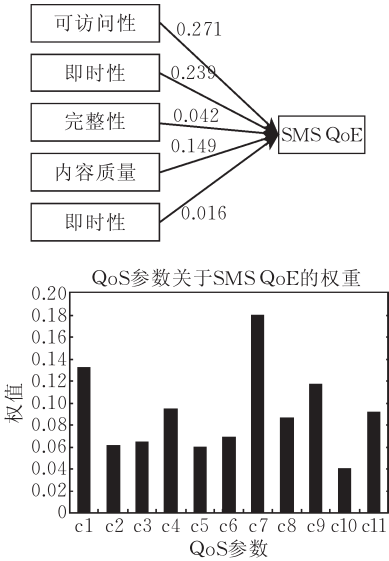


图 7 短信服务的 QoE 的评价模型

从上述讨论中,可以发现,FAHP 方法可以很好地解决多指标以及多层次的问题,但是它需要依赖专家的经验,并且无法描述同一层次指标之间的关系(需要保证同一层指标之间是相互独立的).

4. 3. 2 基于机器学习的 QoE 评价

QoE 的评价问题本质上是一个把众多的影响因素,包括客观环境、用户的主观因素以及业务或者服务的技术性能因素映射到用户感觉层面的一个问题.因此上 QoE 的评价问题必须要研究人的感觉与这些因素之间的映射关系.这就造成了 QoE 评价问题的两大难点:(1)需要研究人的感觉与这些因素之间的关系.虽然在 4. 2 小节中,给出了一些关于人的感觉与物理刺激的讨论,但这些讨论还远远不能解决 QoE 的评价问题;(2) QoE 的影响因素较多,难以给出一个具体的函数将这些影响因素映射为 QoE.运用机器学习的方法可以较好地避开这两个问题,但是机器学习的过程较为复杂,需要大量的计算,这也给采用机器学习进行 QoE 评价的工作带来了较大的难度.

用向量 x 表示影响 QoE 的因素或者 QoE 的评价指标, y 表示 QoE.那么机器学习的目的就是利用现有的一些机器学习的方法,生成一个将 x 映射为 y 的函数 f ,即 $y = f(x)$.下面将分别讨论决策树^[12,25,31]、支持向量机^[12,25,30-31]的方法.

1) 决策树

决策树^[38]是一个分层的树状结构的模型.每一

个中间节点要选定一个属性(即 x 中的一个分量),并根据这个属性部署一个测试(或者一个问题),而每一个叶子节点表示一个决策(一个类型或者一个标记).决策树中一个非常重要的算法是 ID3 算法,它是后来提出的其它算法的基础.ID3 算法是一个迭代的算法,对决策树中每一个节点按照选定的属性部署一个测试.属性的选择依据信息增益的度量.对每一个属性可能的值,都要产生一个分支.针对每个分支,使用剩下的训练数据重复进行上述过程.当没必要继续对训练数据进行分类时,这个算法完成了一个路径.这时候,叶子节点就被增加,并且将其和剩下的数据中数量最多的那一类型进行关联.C4. 5 算法是继 ID3 算法之后被提出的,它和 ID3 的原理大致相同,但是克服了许多 ID3 算法的弱点,比如它可以处理连续值的属性,训练的数据可以有丢失的值,有许多对待过度拟合的修剪算法.因此可以采用 C4. 5 算法进行 QoE 的评价.C4. 5 算法的步骤如下所示:

- (1)对数据进行预处理,将连续型的数据进行离散化处理,形成训练集;
- (2)计算每个属性的信息增益,从而求出其信息增益率,选择信息增益率最大的属性,作为当前属性节点,从而获得决策树的根节点;
- (3)根节点属性每一个可能的取值对应一个样本子集,对样本子集重复上述过程,直到每一个样本子集不需要再进行分类;
- (4)验证决策树的性能,如有出现过度拟合的现象,要对决策树进行修剪.

文献[12]给出了用决策树解决 QoE 评价问题的方法,并获得了较高的准确率.它将时间信息(temporal information)、空间信息(spatial information)、比特率(bitrate)以及帧频率(frame rate)作为视频业务的 QoE 评价指标,并通过 C4. 5 算法,得到了视频业务的 QoE 评价模型.

2) 支持向量机

支持向量机^[38](Support Vector Machine, SVM)是从线性可分情况下的最优分类面发展而来的,采用了保持经验风险值固定而最小化置信界限的策略.SVM 在分类问题上具有良好的性能.

支持向量机是线性的两类分类器中的一种.两类分类问题中的数据可以用两个标记中表示,简记为+1 和-1.记号 X 表示由一组分量 x_i 组成的向量.记号 X_i 表示在数据集集中的第 i 个向量, y_i 表示数据 X_i 的标记.向量 X 的集合称为特征空间.一个二

元组 (\mathbf{X}_i, y_i) 表示一个样本. 一个线性的分类器是基于一个线性的鉴别函数, 如式(17)所示. 其中 \mathbf{w} 称为权向量, b 称为偏移.

$$f(\mathbf{X}) = \mathbf{w}^T \mathbf{X} + b \quad (17)$$

支持向量机的基本思路就是要找到这样一个鉴别函数: 首先, 每个类别的样本点到鉴别函数所对应的超平面的最小距离相等并且最大; 其次, 令这些到超平面距离最小的样本点对应的鉴别函数值为 1 或 -1. 这样支持向量机解决分类问题就可以转换为带有约束条件的优化问题, 如式(18)所示.

$$\begin{aligned} & \text{Minimize } \frac{1}{2} \|\mathbf{w}\|^2 \\ & \text{Subject to: } y_i(\mathbf{w}^T \mathbf{X}_i + b) \geq 1, i = 1, \dots, n \end{aligned} \quad (18)$$

引入一组拉格朗日因子 λ_i , 可以上面的优化问题转换成下面的优化问题, 如式(19)所示.

$$\begin{aligned} & \max_{\lambda} \left\{ \sum_{i=1}^N \lambda_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \right\} \\ & \text{Subject to: } 0 \leq \lambda_i, i = 1, \dots, N, \sum_{i=1}^N \lambda_i y_i = 0 \end{aligned} \quad (19)$$

在这个优化问题中的限制条件保证了分类器的正确性. 但是数据不都是线性可分的, 如果遇到线性不可分的数据, 就不能采用上述的限制条件, 即使是线性可分的数据, 也可以通过错分一些数据, 来增大超平面到两类样本之间的距离. 考虑到这样, 通过引入松弛因子 $\xi_i > 0$ (其中 $i = 1, \dots, n$), 上面的问题就变为式(20)所示的优化问题.

$$\begin{aligned} & \text{Minimize } \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \\ & \text{Subject to: } y_i(\mathbf{w}^T \mathbf{X}_i + b) \geq 1 - \xi_i, i = 1, \dots, n \end{aligned} \quad (20)$$

其中, $C > 0$.

同样, 引入一组拉格朗日因子 λ_i , 上面的优化问题就转换为式(21)所示的优化问题.

$$\begin{aligned} & \max_{\lambda} \left\{ \sum_{i=1}^N \lambda_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \right\} \\ & \text{Subject to: } 0 \leq \lambda_i \leq C, i = 1, \dots, N \\ & \sum_{i=1}^N \lambda_i y_i = 0 \end{aligned} \quad (21)$$

利用 SMO 算法^[39]就可以解决式(19)和式(21)的优化问题, 从而得到最后的结果. 以上介绍的都是线性支持向量机. 当然也可以构造非线性的支持向量机, 它的基本思路就是利用一组非线性函数将原特征空间映射到新特征空间, 在新的特征空间设计

线性支持向量机.

文献[12]也给出了用支持向量机解决 QoE 评价问题的方法. 它同样将时间信息(temporal information)、空间信息(Spatial information)、比特率(Bitrate)以及帧频率(Frame rate)作为视频业务的 QoE 评价指标, 并通过 SMO 方法, 得到了视频业务的 QoE 评价模型.

由于支持向量机是一个两类的分类问题, 所以利用支持向量机解决 QoE 评价的问题有一个限制, 那就是它只能将 QoE 分成两个级别, 即可接受和不可接受. 如果将 QoE 分为 $n(n > 2)$ 个级别, 就需要在每两个类别之间构造一个超平面, 即共需构造 $\binom{n}{2}$ 个超平面. 这给支持向量机的使用带来了较大的不便.

值得注意的是, 在用机器学习解决实际问题时, 往往不是仅仅使用一种方法, 而是采用多种方法, 如上面提到的决策树、支持向量机, 然后对这几种方法得到的结果进行比较, 选择合适的模型.

4.4 基于随机模型的评价方法研究

在提出基于随机模型的评价方法之前, 我们先对前面介绍的 3 种基于不同学科的评价方法做一些分析和讨论. 基于统计学的评价方法的主要优点是可以利用统计学中因子分析等方法来研究 QoE 评价指标或者影响因素之间的相关性, 进而可以达到数据降维和问题简化的效果. 同时可以利用统计学中的回归分析和判别分析给出 QoE 的评价模型, 虽然精确度可能不如机器学习的方法^[12]. 基于 FAHP 的方法可以较好的解决 QoE 多指标的问题, 计算也较简单, 但是它需要依赖于专家知识, 这给 FAHP 方法的运用带来了较大的不便. 基于机器学习的方法有较高的精确度, 但是训练过程较为复杂. 另外, 本文也讨论了基于心理学的评价方法. 基于心理学的评价方法是基于心理学中的一些定律, 不需要进行复杂的训练以及计算的过程. 但它主要研究的是 QoE 与某些单一的 QoS 参数的函数模型, 无法解决 QoE 多影响因素的问题, 这给基于心理学的 QoE 评价方法的应用带来了一定的限制.

值得注意的是, 现有评价方法的模型都忽略一个这样一个事实: 用户之前的体验会对当前体验造成较大的影响. 因此我们提出了基于隐马尔可夫的 QoE 评价模型, 可以更合理地模型用户体验的过程, 从而建立一个更适用的 QoE 评价体系结构. 下面我们以 3G/4G 网络中的视频流媒体业务为例介

绍基于隐马尔可夫(Hidden Markov Model, HMM)的 QoE 评价模型.

4. 4. 1 评价参数体系

文献[40]提出了一个在 3G/4G 中通用的视频流媒体评价参数体系,如图 8 所示.

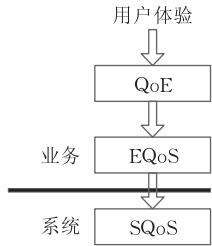


图 8 视频流媒体的 QoE 评价参数体系

QoE 参数是从用户角度定义的业务总体的性能参数,难以量化或者测量.它实际上只起一个从 QoE 到 QoS 参数的桥梁的作用,它是为了获得 QoE 对应的 QoS 参数.ESQoS 参数反映的是端到端的业务,它是从 QoE 参数映射得到,可以量化和测量.SQoS 参数反映的是网络为业务提供服务的能力.文献[40]详细地讨论了 ESQoS 和 SQoS 参数的选取、测量和计算的问题,但是它并没有给出如何通过这些已计算出的 ESQoS 参数得到最终 QoE.接下来我们将详细讨论如何利用隐马尔可夫模型将这些 ESQoS 参数映射为最终的 QoE.

4. 4. 2 评价模型的建立

下面首先简要介绍隐马尔可夫模型^[41-42].隐马尔可夫模型是一种统计概率模型.它是一个双重随机过程,即马尔可夫链^[43]和一般的随机过程.马尔可夫链用来描述状态的转移,用转移概率描述.一般随机过程用来描述状态和观察序列间的关系,用观察概率描述.隐马尔可夫模型用五元组 $\lambda = (N, M, \pi, \mathbf{A}, \mathbf{B})$ 来描述,或者可以简写为 $\lambda = (\pi, \mathbf{A}, \mathbf{B})$,其中 N 表示隐马尔可夫模型状态的数目,也就是隐马尔可夫模型中马尔可夫链的状态数目; M 表示隐马尔可夫模型中每个状态可能的观察值的数目; \mathbf{A} 表示隐马尔可夫模型中状态之间的转移概率矩阵; \mathbf{B} 表示每一个状态下每一个观测值对应概率矩阵; π 表示初始时,每个状态的概率.隐马尔可夫模型有 3 个基本的问题,评估问题、解码问题以及学习问题.评估问题是指如何计算给定的观察序列的概率.解码问题是指如何计算出最有可能产生给定观察序列的状态序列.学习问题是指如何根据观察序列求得模型参数或者调整模型参数.解码问题和学习问题分别可以采用 Viterbi 算法和 Baum-Welch 算法进行求解.

为了解决前面讨论的几种模型的缺陷,我们就必须要将用户之前体验对当前体验造成的影响考虑在内.为了简化问题,我们做出以下两个假设:

- (1) 用户对服务的主观体验能够以会话为单位进行讨论,也就是说每经过一次会话,用户就可以形成对服务的一个整体感知;
- (2) 仅仅前一次会话的用户体验会对当前的体验造成影响.

在上述两个假设的基础上,我们就可以建立基于隐马尔可夫的 QoE 评价模型.首先,以会话为单位,将用户的体验序列化.即有几次会话,就可以有几次完整的用户体验,而且这些体验存在先后顺序.其次,我们利用前面提到的 MOS 方法,将每次会话的用户体验进行量化,将 MOS 中的 5 个等级分别作为隐马尔可夫模型的 5 个状态,如图 9 所示.最后,将每次会话 ESQoS 参数作为隐马尔可夫模型的观测值.这样我们就确定了隐马尔可夫模型中的状态和观测值表示的含义.另外还需要确定 $\pi, \mathbf{A}, \mathbf{B}$ 等参数.由于 ESQoS 具有多个分量,而且是连续的.直接利用 ESQoS 参数进行训练可能会使问题变得非常复杂,因此可以利用前面提到的 PCA 方法对 ES-QoS 参数进行处理,然后将得到的主成分进行离散化.将离散化后的主成分作为隐马尔可夫模型的观测值,这样就可以用 Baum-Welch 算法进行求解. Baum-Welch 估计算法是一种迭代的算法.利用 Baum-Welch 算法时,需要由用户自己初始化各个参数的值,然后通过不断的迭代,当误差到了一定范围之内时,就得到了模型中各个参数的值.通过以上步骤,我们就建立了基于隐马尔可夫的 QoE 评价模型.

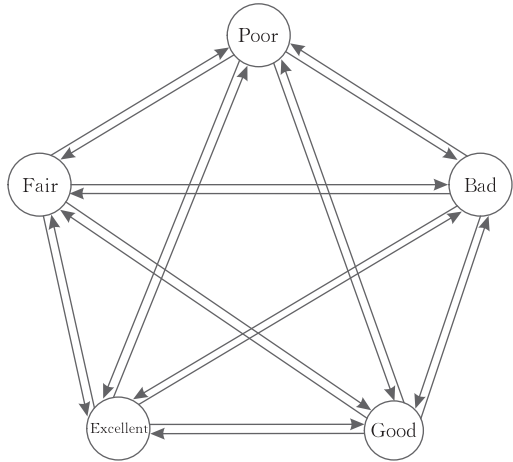


图 9 基于隐马尔可夫的 QoE 评价模型

4. 4. 3 评价模型的验证与应用

在得到 QoE 的评价模型之后,我们就可以利用

自身验证、外部数据验证以及交互验证等一些验证方法对模型的性能进行分析和讨论。

上述的模型将 MOS 方法中 QoE 的 5 个级别分别对应隐马尔可夫模型的 5 个状态,而经过处理的每次会话的 ESQoS 参数值则对应隐马尔可夫模型的观测向量。在获得 QoE 评价模型的具体参数值之后, QoE 的评价问题,就等价于如何利用一次或者连续的多次会话的 ESQoS 参数形成的观测序列对隐马尔可夫模型状态序列进行估计,这就等价于 HMM 的解码问题,可以利用 Viterbi 算法^[41-42]进行求解。Viterbi 算法是一种动态规划的算法,其复杂度为 $O(K^2L)$,其中 K 表示状态个数; L 表示序列长度。由于 K 为常数 5,所以在此算法可以在线性时间内得到问题的解。

基于隐马尔可夫的 QoE 评价模型可以克服现有模型的共有缺陷,可以较好地对用户体验的过程建模。但是它也存在一些缺点。首先,基于隐马尔可夫的 QoE 评价模型基于前面提出的两个假设,若这两个假设不符合实际,或和实际情况差别较大,那么基于隐马尔可夫的 QoE 评价就失去了合理性。其次,模型的训练也比较复杂。由于 Baum-Welch 算法是一种迭代的算法,所以需要较长时间来得到问题的解。最后,对某些服务,用户可能不会在较短的时间内连续体验多次,这时基于隐马尔可夫的 QoE 评价模型和前面提到的其它评价模型相比就不再具有优越性。

当然,基于隐马尔可夫的 QoE 评价模型还处于理论阶段,还没有应用于实践。对它的验证和完善还需要进一步的研究。

5 总 结

本文在现有文献的基础上,给出了 QoE 的定义以及影响因素,并且综述了 QoE 的量化方法和评价方法。其中详细介绍了基于统计学、心理学及人工智能的 QoE 评价方法,并且分析比较了这几种方法之间的优缺点。在分析和总结现有评价模型和方法的基础上,本文以视频流媒体业务为例,运用隐马尔可夫模型对用户体验的过程进行建模,提出了基于隐马尔可夫的 QoE 评价模型,能够更合理地描述用户体验的过程,从而可以为构建更准确的 QoE 评价方法奠定基础。

QoE 评价还存在很多的问题需要进一步解决:

(1) QoE 多属性融合的问题。QoE 是一个多属

性融合的问题,现有文献中解决 QoE 多属性的问题只有统计学和人工智能学的一些方法。但这些方法是基于大量的样本数据,计算较为复杂,而且并不能揭示 QoE 与其属性之间本质上的关系。

(2) 人的主观因素以及环境因素对 QoE 的影响仍需进一步的研究和探索。由于人的主观因素与环境因素难以测量以及量化,所以这些因素对 QoE 的影响难以评价。需要借助心理学以及社会学的相关理论对这些影响因素进行研究和探索。

(3) 对不同的服务建立一套通用的评价指标还需要进一步的探索。文献[32, 44-45]给出了一些讨论,但是没有对这些评价指标的合理性和完备性进行分析。

参 考 文 献

- [1] Siller M, Woods J. QoE improvement of multimedia transmission//Proceedings of the IADIS International Conference, 2003, II: 821-825
- [2] O'Neil Timothy M. Quality of experience and quality of service. For IP video conferencing. Director of Technical Marketing of Polycom Video Communications, White Paper by Polycom, 2005
- [3] ITU-T: Definition of quality of experience (QoE). International Telecommunication Union, Liaison Statement, Ref.: TD 109rev2 (PLEN/12), 2007
- [4] Garcia M, Canovas A, Edo M et al. A QoE management system for ubiquitous IPTV devices//Proceedings of the 2009 Third International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, 2009: 147-152
- [5] Volk M, Sterle J, Sedlar U et al. An approach to modeling and control of QoE in next generation networks. IEEE Communications Magazine, 2010, 48(8): 126-135
- [6] Ries M, Froehlich P, Schatz R. QoE evaluation of high-definition IPTV services//Proceedings of the Radioelektronika (Radioelektronika). 2011: 1-5
- [7] Baraković S, Baraković J, Bajrić H. QoE dimensions and QoE measurement of NGN services//Proceedings of the Telecommunications Forum (TELFOR 2010). 2010: 15-18
- [8] Moller S, Engelbrecht K P, Kuhnel C, Wechsung I, Weiss B. A taxonomy of quality of service and Quality of Experience of multimodal human-machine interaction//Proceedings of the International Workshop on Quality of Multimedia Experience (QoMEX 2009). 2009: 7-12
- [9] ur Rehman Laghari K, Crespi N, Molina B, Palau C E. QoE aware service delivery in distributed environment//Proceedings of the 2011 IEEE Workshops of International Conference on Advanced Information Networking and Applications (WAINA). 2011: 837-842

- [10] Stevens S S, Volkman J. The relation of pitch to frequency: A revised scale. *The American Journal of Psychology*, 1940, 53(3): 329-353
- [11] Agboma F, Liotta A. QoE-aware QoS management//Proceedings of the MoMM2008. 2008: 111-116
- [12] Menkovski V, Sánchez A C. Predicting quality of experience in multimedia streaming//Proceedings of the MoMM2009. 2009: 52-59
- [13] Chen K T, Wu C C, Chang Y C, Lei C L. A crowd sourceable QoE evaluation framework for multimedia content//Proceedings of the 17th ACM International Conference on Multimedia. 2009: 491-500
- [14] Augustin T. Bradley-terry-luce models to incorporate within-pair order effects: Representation and uniqueness theorems. *British Journal of Mathematical and Statistical Psychology*, 2004, 57(2): 281-294
- [15] International Telecommunication Union, Geneva. Methods for subjectivedetermination of transmission quality. Report: ITU-T-P. 800, 1996
- [16] Venkataraman M, Sengupta S, Chatterjee M et al. Towards a video QoE definition in converged networks//Proceedings of the 2nd International Conference on Digital Telecommunications (ICDT'07). 2007
- [17] Wang Y. Survey of objective video quality measurements. Worcester Polytechnic Institute; Technical Report, 2006
- [18] NetPredic, Inc. Performance Analysis for Video Streams across Networks. White Paper, 2003
- [19] ITU-T Rec J. 143. User requirements for objective perceptual video quality measurements in digital cable television. 2000
- [20] Winkler S. Quality of Experience (QoE): An important measure of success for IP-based video services. Symmetricom White Paper, 2007
- [21] Richards A, Rogers G, Antoniadis M et al. Mapping user level QoS from a single parameter//Proceedings of the 2nd IFIP/IEEE International Conference of Management of Multimedia Networks and Services, 1998
- [22] Hyun Jong Kim, Seong Gon Choi. A study on a QoS/QoE correlation model for QoE evaluation on IPTV service//Proceedings of the 12th International Conference on Advanced Communication Technology (ICACT). 2010, 2: 1377-1382
- [23] Koning T D D, Velthoven P, Knoche H et al. Of MOS and men: Bridging the gap between objective and subjective quality measurements in Mobile TV//Proceedings of the SPIE. 2007
- [24] Perreira F, Burnett I. Universal multimedia experiences for tomorrow. *IEEE Signal Processing Magazine*, 2003, 20(2): 63-73
- [25] Menkovski V, Exarchakos G, Liotta A. Online QoE Prediction//Proceedings of the QoMEX. 2010: 118-123
- [26] ITU-R Recommendation BT500. Methodology for the Subjective Assessment of the Quality of Television Pictures, 2002
- [27] Ketykó I, De Moor K, Joseph W et al. Performing QoE-Measurements in an Actual 3G Network//Proceedings of the 2010 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting. 2010: 1-6
- [28] De Moor K, Joseph W, Ketykó I et al. Linking users' subjective QoE evaluation to signal strength in an IEEE 802.11b/g wireless LAN environment. *Eurasip Journal on Wireless Communications and Networking*, 2010, 2010: 1-12
- [29] Chen Kuan-Ta, Huang Chun-Ying, Huang Polly et al. Quantifying skype user satisfaction//Proceedings of the ACM Conference on Applications, Technologies, Architectures and Protocols for Computer Communications (SIGCOMM'06). 2006: 399-410
- [30] Farrugia R A, Debono C J. A support vector machine approach for detection and localization of transmission errors within standard H. 263++ Decoders. *IEEE Transactions on Multimedia*, 2009, 11(7): 1323-1330
- [31] Menkovski V, Exarchakos G, Liotta A. Machine learning approach for quality of experience aware networks//Proceedings of the 2nd International Conference on Intelligent Networking and Collaborative Systems (INCOS). 2010: 461-466
- [32] Du Yu, Zhou Wen-An, Chen Bao-Fu et al. A QoE based evaluation of service quality in wireless communication network//Proceedings of the New Trends in Information and Service Science. 2009: 552-557
- [33] Huang R L, Guan Y H. Data Statistics Analysis. Beijing: Higher Education Press, 2010(in Chinese)
(黄润龙, 管于华. 数据统计分析. 北京: 高等教育出版社, 2010)
- [34] Urban F M. The weber-fechner law and mental measurement. *Journal of Experimental Psychology*, 1933, 16(2): 221-238
- [35] Reichl P, Egger S, Schatz R. The logarithmic nature of QoE and the role of the weber-fechner law in QoE assessment//Proceedings of the IEEE ICC 2010. 2010
- [36] Fiedler M, Hossfeld T, Phuoc Tran-Gia. A generic quantitative relationship between quality of experience and quality of service. *IEEE of Network*, 2010, 24(2): 36-41
- [37] Mikhailov L, Tsvetinov P. Evaluation of services using a fuzzy analytichierarchy process. *Appl. Soft Comput.*, 2004, 5(1): 23-33
- [38] Bao Jun-Peng, Zhang Xuan-Ping. Artificial Intelligence Introduction. Beijing: China Machine Press, 2011(in Chinese)
(鲍军鹏, 张选平. 人工智能导论. 北京: 机械工业出版社, 2011)
- [39] Platt J. Sequential minimal optimization: A fast algorithm for training support vector machines. Microsoft Research; Technical Report MST-TR-98-14, 1998
- [40] End-2-end multimedia services performance metrics. Technical Report; 3GPP TR 26.944, 2007
- [41] Schuller B, Rigol G I, Lang M. Hidden Markov model-based speech emotion recognition//Proceedings of ICASSP. 2003: 1-4

[42] Rabiner L R, Juang B. An introduction to Hidden Markov Models. IEEE ASSP Magazine, 1986, 3(1): 4-16

[43] Lin Chuang. Computer Network and Computer System Performance Evaluation. Beijing: Tsinghua University Press, 2001: 25-26(in Chinese)
(林闯. 计算机网络和计算机系统的性能评价. 北京: 清华大学出版社, 2001: 25-26)

[44] Gong Yan, Yang Fangchun, Huang Lin et al. Model-based approach to measuring quality of experience//Proceedings of the International Conference on Emerging Network Intelligence. 2009: 29-32

[45] Wu Wan-Min, Arefin A, Rivas R. Quality of experience in distributed interactive multimedia environments: Toward a theoretical framework//Proceedings of the MM'09. 2009: 481-490



LIN Chuang, born in 1948, Ph. D. , professor, Ph. D. supervisor. His research interests include computer networks, performance evaluation, network security analysis, and Petri net theory and its applications.

HU Jie, born in 1990, Ph. D. candidate. His research interest is Quality of Experience.

KONG Xiang-Zhen, born in 1985, Ph. D. candidate. His research interests include the technology of virtualization, the dependability and security of system.

Background

Quality of service (QoS) is the most popular service metrics, but it focuses on the technology performance of service, doesn't take the human perception or context into account. So it can't reflect the users' satisfaction directly. Quality of Experience (QoE) is the overall performance of a system from the point of view of the users. It takes all the influence factors of users' satisfaction for service into account and reflects the experience of the users directly. This paper provides a complete overview of current evaluation models and methods of Quality of Experience, and then proposes an evaluation method based on stochastic models. This work is partly supported by the National Basic Research Program

(973 Program) of China (Nos. 2010CB328105, 2009CB320504), National Natural Science Foundation of China (No. 60932003). These projects aim to provide better performance in computer networks and information systems. Our group has working on the performance evaluation of the computer networks and computer systems using the stochastic theoretical models. Many good papers have been published in respectable international conferences and transactions, such as INFOCOM, IWQoS and IEEE Transactions on Parallel and Distributed Systems. This paper summarizes the evaluation models and methods of Quality of Experience.