

基于 FSOM 的 VoIP QoS 反馈控制机制研究^①

张 武^② 王劲林

(中国科学院声学所国家网络新媒体工程技术研究中心 北京 100190)

摘 要 针对网络语音服务质量(VoIP QoS)问题展开了服务质量影响因素及相关评测方法的深入研究,提出了基于模糊自组织映射(FSOM)的 VoIP QoS 反馈控制算法。详细设计了反馈控制系统模型,其中引入了服务质量推理和会话再协商机制。搭建了研究所需的实验环境,据此进行实验数据的获取和仿真结果的验证。实验结果表明,该机制有效地改善了网络语音的服务质量,可将其应用于网络多媒体服务质量控制等其他领域。

关键词 网络语音(VoIP), 服务质量(QoS), 模糊自组织映射(FSOM)网络, 会话协商, 会话发起协议(SIP)

0 引 言

IP 网络已经成为支持数据、语音和视频业务的基础承载网络。由于其本质上是“尽力而为”的,服务质量(QoS)控制一直是研究的热点问题之一。在基于 IP 网络的数据、语音和视频业务中,网络语音(VoIP)业务本身的实时性和交互性对服务质量控制提出了较高的要求^[1,2]。VoIP QoS 控制方法可分为发送端 QoS 控制方法、网络层面 QoS 控制方法和接收端 QoS 控制方法^[3,4]。发送端 QoS 控制方法的基本原理是利用实时传输控制协议(real-time transport control protocol, RTCP)反馈回来的发送速率来确定端到端带宽,进而选择不同的编解码方式;网络传输层面的 VoIP QoS 控制方法以 InterServ/RSVP 模型、Diff-serv 模型和 MPLS 流量工程为主,辅以 P2P 重叠网络技术来保证 VoIP 的服务质量;接收端的 QoS 控制方法的基本原理是调整接收端缓存的尺寸,较好地平衡分组迟到所导致的丢包与总体延时的关系。综合考虑发送端、网络层面和接收端因素,本文提出了一种新的 VoIP QoS 反馈控制机制:在数据层面上进行 QoS 分析,确定影响 QoS 影响因素和需要调整会话终端的配置参数,通过信令层面协商需要调整的配置参数,实现 VoIP QoS 反馈控制。其中 VoIP QoS 评价分析采用了模糊自组织映射(fuzzy self-organization map, FSOM)网络方法。

1 VoIP QoS 影响因素与评价方法

网络语音服务质量控制的难点在于 VoIP QoS 具有很强的非线性。一方面,网络语音服务质量影响因素较多,并且影响因素与服务质量之间关系呈现出较强的非线性;另一方面,语音质量评测方式复杂,表现出很强的非线性。下面针对这两方面问题进行简要的分析。

1.1 影响因素

网络语音服务质量的主要影响因素包括编解码方式选择、打包周期、播放时间、带宽、乱序、丢包、延时和抖动等。这些 QoS 影响因素可分为两个层次——网络因素和终端因素。网络因素包括网络性能参数及其主要因素,其中网络性能参数包括数据包丢失率、乱序、延时及抖动等。影响网络性能的主要因素包括链路特性、网络拥塞及拥塞导致的丢包、路由选路等。终端因素包括编解码方式选择、数据包长、缓存长度与控制算法选择。网络性能参数、网络性能的主要因素和终端配置构成了影响因素的三个基本方面。其中网络性能导致原因是 QoS 最重要的影响因素,但网络运营商难于针对每个具体的 VoIP 会话调控网络性能。通过网络性能参数描述网络性能特征,将这些参数作为 QoS 反馈控制的输入条件,调整终端配置参数,可在一定程度上调控 QoS。

针对 VoIP 的具体应用,终端方面的影响因素主

^① 863 计划(2007AA01Z238)资助项目。

^② 男,1975 年生,博士,助研;研究方向:网络技术,多媒体技术,嵌入式系统;联系人, E-mail: zhangw@dsp.ac.cn (收稿日期:2009-06-24)

要包括:编解码选择(G.711、G.723.1、G.729等)、打包周期和播放时间;网络方面的影响因素主要包括:带宽、丢包、乱序、延时与抖动。如图1所示,带宽限制会导致丢包、延时及抖动;过大的延时和抖动会进一步导致丢包;较小的乱序可被播放缓存过滤,形成一定的延时和抖动;较大的乱序会导致分组失去时效性,直接造成丢包。带宽与编解码方式有直接的关系,语音数据所需速率超过网络提供的带宽,会造成严重的丢包;打包周期直接会影响网络的使用效率,如果与带宽配合得不好,会造成丢包、延时和抖动。播放时间与播放缓存是同一问题的两个方面,播放时间反映了播放缓存的时间尺度,播放缓存反映了播放时间的空间尺度。这些因素相互作用、相互转化,构成了影响因素与 VoIP QoS 之间的复杂非线性关系。

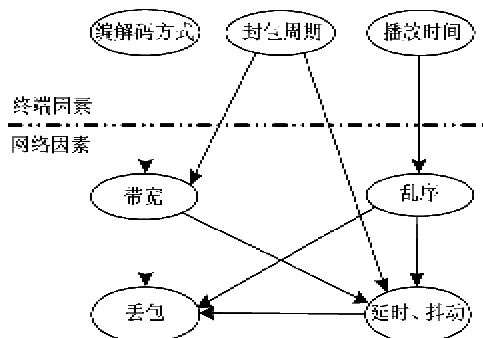


图1 语音 QoS 影响因素关系

1.2 评测方式的非线性分析

语音质量评测方式经历了由主观到客观、由人工方式到仪表方式的发展历程。主要评测的方法有 MOS(mean optimum score)模型、E-Model 模型、PSQM(perceptual speech quality measure)模型、PSQM+模型、PESQ(perceptual evaluation of speech quality)模型。下面对其进行简要的分析。

(1) MOS 模型^[5]

自发明电话之后,广泛采用的语音质量测量方法是利用人耳直接感知通话的语音质量。该方法是一种主观的评价方法,已经被广泛认同。经过标准化,该测量方法演变为平均主观值 MOS 模型,定义在 ITU-T P.800 中。利用该模型来进行网络的配置、基准确定和监听,并且要获得具有较好服务质量的网络配置,需要付出大量的人力成本,而且很难在线评价服务质量。

(2) E-Model 模型^[6]

ITU-TG.107 定义的 ETSI E-model 是一种规划工具。E-model 模型最终的结果是 R 值,用于全面地

描述网络传输等级,取值范围为 $[0,100]$ 。 R 值的基本计算公式如下:

$$R = R_0 - I_s - I_d - I_e + A \quad (1)$$

其中, R_0 是无网络延时和设备损伤因素的基本信号与收发噪声以及电流、背景噪声之比,即基本信噪比; I_s 是与语音信号传输同步的损伤; I_d 是语音信号传输延时后的损伤; I_e 是由设备引入的损伤,例如编码器损伤; A 是优势因素,致力于考虑呼叫者的期望因素,在大部分情况下,一般设置为 0。

在 VoIP 系统中,需要给每个设备确定设备损伤系数 I_e 。综合所有 I_e 值及其他参数来计算 R 因数。然后,进行从 R 参数到 MOS 值的非线性映射,实现语音质量的评测。此外,各个参数的选择均包含了诸多的非线性因素。尽管 E-Model 是优秀的规划工具,根据 R 因数部署 VoIP 系统,并获得语音质量的近似值。但模型本身做了大量的假设,所以它不能代替最终网络的真实测量,同样也难于在线 VoIP QoS 评估。

(3) PSQM 模型、PSQM+模型、PESQ 模型^[7,8]

ITU-T P.861 对 PSQM 模型进行了标准化,该标准主要面向移动传输的语音编解码估计,目前的 PSQM 仍旧没有考虑 VoIP 应用。为适应下一代网络业务——VoIP,测量模型需要面对更大的信号失真,ITU 成立了相关工作组开展 P.861 的修订。PSQM+ 针对上述问题,较好地处理了由突发错误带来的信号失真,但未有效地补偿变化的时延。ITU P.862 对 PESQ 模型进行了标准化,PESQ 把心理声学 and PSQM+ 的感知模型关联使用,没有面向流是 PESQ 的唯一缺陷,所以 PESQ 不能完全代替 PSQM+。利用 PSQM+ 和 PESQ 可以覆盖整个语音质量测试。

以上 VoIP QoS 评测方法都存在着一一定的非线性因素,各种模型实现的复杂度高,难于直接进行语音质量的在线检测。

2 VoIP QoS 反馈控制模型

如图2所示,VoIP QoS 反馈控制模型由待评估 VoIP 系统、VoIP QoS 推理系统和活动会话再协商机制三部分构成。模型的输入是配置参数,模型的输出是语音系统的特征参数。VoIP 系统包括网络 and 进行 VoIP 会话的终端;VoIP QoS 推理系统对 VoIP 系统的特征参数进行分析推理,确定出需要调整的配置参数;活动会话再协商机制将配置参数传送给终端,调整配置参数,实现了反馈控制。

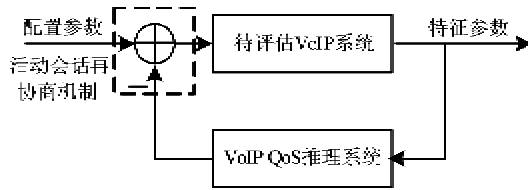


图2 服务质量反馈控制模型

由于 VoIP QoS 影响因素和测量方法的非线性,以及测量方法难以在线实现,VoIP QoS 推理系统成了反馈控制模型实现的关键部分。而模糊自组织映射(FSOM)网络可很好解决非线性问题,并且本文的思路与方法已在故障诊断领域得到了广泛应用,因此本文利用 FSOM 来实现 VoIP QoS 推理系统^[9]。

3 VoIP QoS 推理系统

3.1 模型描述

VoIP QoS 规划问题可以表达为一个四元式

$$Q_J = (X, Y, E^*, X^*) \quad (2)$$

其中: $X = (x_1, x_2, \dots, x_n)^T$ 为征兆集合,包括丢包率、延时及抖动参数等; $Y = (y_1, y_2, \dots, y_n)^T$ 为配置问题集合,包括编解码方式、数据包长度和缓存长度与管理等; $X^* \subseteq X$, 表示可观测到的征兆集合,是 X 的子集; $E^* \subseteq X \times Y$, 是征兆与配置问题的关系子集,表达了征兆与配置间的因果关系,具体通过模糊关系的隶属度函数来表达,最终表现为神经网络输入层与输出层的连接权值。

FSOM 非常适合线性数据的处理和分类,而语音质量又是非常典型的非线性问题,所以利用 FSOM 方法实现 VoIP QoS 推理系统。

基于 FSOM 的 VoIP QoS 推理系统模型如图 3 所示。VoIP 网关、传输网络组成的 VoIP 系统成为被判断系统——待评估 VoIP 系统,对其提取问题征兆,

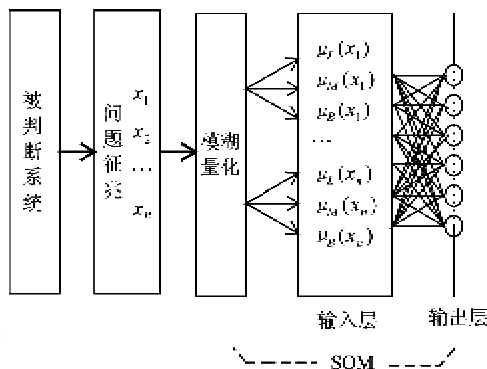


图3 基于 FSOM 的 VoIP QoS 推理系统模型

如网络性能参数、QoS 值(PESQ)。如上文所述,语音质量与影响因素之间具有复杂的非线性关系,难于用确定的数量关系来表达,而模糊理论能够较好地处理这些不确定量的关系,所以对系统“征兆”进行模糊化是非常必要的。然后,利用 SOM 网络的“聚类”机制,对话音质量的征兆进行分类,并且在低维空间内显示,实现推断机制。

3.2 FSOM 算法描述

模糊计算的实质是获得隶属度关系,而其关键问题在于确定隶属度函数。隶属度函数具体定义如下:

定义 设 $\tilde{A}$ 是论域 X 到 $[0, 1]$ 的一个映射,即 $\tilde{A}: X \rightarrow [0, 1], x \rightarrow \tilde{A}(x)$, 称 $\tilde{A}$ 是 X 上的模糊集,而函数 $\tilde{A}(\cdot)$ 称为模糊集 $\tilde{A}$ 的隶属度函数, $\tilde{A}(x)$ 称为 x 对模糊集的隶属度。其中 PESQ 值、带宽满足这种偏小型模糊分布;丢包率、单向延时、抖动、乱序满足偏大型模糊分布;发包周期满足中间型模糊分布-II。

利用征兆所对应的模糊分布对征兆进行模糊化处理,求解“故障”的隶属度:

$$\begin{aligned} \bar{X}_p &= [\pi(x_1, c, \lambda), \pi(x_2, c, \lambda), \dots, \pi(x_n, c, \lambda)]^T \\ &= [x_1^p, x_2^p, \dots, x_n^p]^T \end{aligned} \quad (3)$$

VoIP QoS 配置问题的任意输入样本为 $X_p = [x_1^p, x_2^p, \dots, x_n^p]^T$, 令 t 为学习次数, $\eta(t)$ 为学习速度,邻域为 $N_g(t)$, 连接权矢量 $W_j = [w_{j1}, w_{j2}, \dots, w_{jn}]^T (j = 1, 2, \dots, m \times n)$, 具体算法如下:

初始化将连接权 w_{ji} 赋予 $[0, 1]$ 区间的随机数, 确定学习速率 $\eta(0) = \eta_0, \eta_0 \in [0, 1]$, 确定邻域为 $N_g(0)$ 及总学习次数 T 。

利用公式(3)计算输入向量相对故障的隶属度值 $X_p = [x_1^p, x_2^p, \dots, x_n^p]^T$, 为了缩短网络的学习时间,将 X_p, W_j 进行归一化处理:

$$\bar{X}_p = [\bar{x}_1^p, \bar{x}_2^p, \dots, \bar{x}_n^p]^T = \frac{X_p}{\|X_p\|} \quad (4)$$

其中 $\|X_p\| = [(x_1^p)^2 + (x_2^p)^2 + \dots + (x_n^p)^2]^{\frac{1}{2}}$;

$$\bar{W}_j = [\bar{w}_{j1}, \bar{w}_{j2}, \dots, \bar{w}_{jn}]^T = \frac{W_j}{\|W_j\|} \quad (5)$$

其中 $\|W_j\| = [(w_{j1})^2 + (w_{j2})^2 + \dots + (w_{jn})^2]^{\frac{1}{2}}$ 。

计算 $\bar{W}_j$ 到输入矢量间的欧氏距离 d_j :

$$d_j = \|\bar{x}_p - \bar{W}_j\| = \left[\sum_{i=1}^n (\bar{x}_i^p - \bar{w}_{ji})^2 \right]^{\frac{1}{2}} \quad (6)$$

找出最小距离 d_g :

$$d_g = \min(d_j) \quad (7)$$

调整连接的权值 $\bar{w}_{ji}(t+1)$:

$$\overline{w_{ji}}(t+1) = \begin{cases} \overline{w_{ji}}(t) + \eta(t)[x_i^p - \overline{w_{ji}}(t)] \\ \overline{w_{ji}}(t) \end{cases} \quad (8)$$

连接权值归一化处理:

$$\overline{W_j}(t+1) = \frac{W_j(t+1)}{\|W_j(t+1)\|} \quad (9)$$

更新输入量,重复以上的运算。

更新学习速率 $\eta(t)$ 及邻域 $N_g(t)$:

$$\eta(t) = \eta_0(1 - \frac{t}{T}) \quad (10)$$

$$N_g(t) = \text{INT}[N_g(0)(1 - \frac{t}{T})] \quad (11)$$

式中, t 为学习次数; T 为总学习次数; $\text{INT}[x]$ 表示对 x 取整。令 $t = t + 1$, 重新进行运算, 直到 $t = T$ 为止。

4 活动会话再协商机制

对活动的会话修改配置可以通过实时传输协议 (real-time transport protocol, RTP) 带内控制方式、会话发起协议 (session initiation protocol, SIP) 带外控制方式和 SIP 带外加上端口变换控制方式来实现。

RTP 带内控制方式——将配置变化情况通过字段填充在 RTP 的负载中, 对端通过检查此字段来修改配置。优点是: 实现简单。缺点是: 增加了终端对每个 RTP 负载处理负担。

SIP 带外控制方式——通过 SIP 信令通知对端配置变化。优点是: 实现简单; 减少了终端处理的开

销。缺点是: 由于存在网络延时, SIP 控制信令很难与 RTP 媒体流保持同步, 可能会造成分组丢失或错误处理。

SIP 带外加上 RTP 端口变换控制方式——通过 SIP 信令协商通知对端会话媒体的传输格式以及新的 RTP 端口, 协商成功后, 会话双方关闭旧配置的编码方式, 利用新配置进行编码媒体处理 (停止旧的传输配置), 通过新端口发送; 接受旧端口上的数据, 继续按旧的配置进行解码媒体处理, 同时监听新端口, 如果收到新端口上的数据, 表示旧配置数据传输完成, 启动新的解码配置, 关闭旧的解码配置及端口。优点是: 克服了 RTP 带内控制方式、SIP 带外控制方式的问题。缺点是: 实现相对复杂。

对比以上三种方式, SIP 带外加上 RTP 端口变换控制方式具有明显的优势, 故选用该方式作为会话协商机制。如图 4 所示, 两终端设备 (RGW A、RGW B) 进行正常 VoIP 通信。它们分别扩展了服务质量推理机制和活动会话协商机制, 即 APP QoS (应用层 QoS) 和扩展 SDP 的 SIP 协议栈。APP QoS 在应用层进行了端到端的 QoS 协商, 确定新的终端配置参数, 由主叫终端发出 Invite 信令。该信令通过扩展的 SDP 字段传输修改配置参数, 不妨将它定义为 CC 字段。经过 3 次通讯 2 次握手后, 确认修改配置成功。两终端设备在新的 RTP 端口上进行正常的 VoIP 通信。

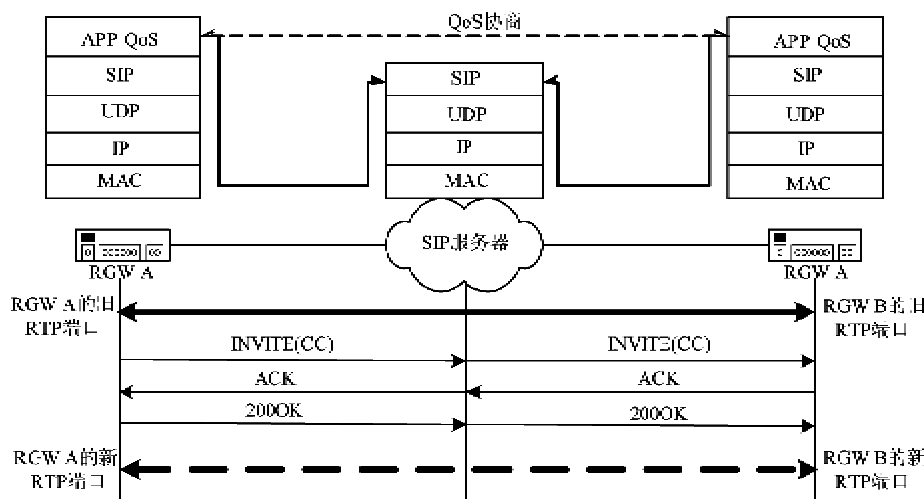


图 4 SIP 带外加 RTP 端口变换控制方式的实现流程

综上所述, 该服务质量反馈控制方法将应用层的服务质量推理和会话层的活动会话协商有机地结合起来, 实现了与传输网络无关的服务质量协商。

5 模型的实验与仿真

样本获取实验环境如图 5 所示, 包含的设备有两 VoIP 终端 (RGW A 和 RGW B)、语音质量测试

(Agilent VQT1981B)、网络环境模拟仪 (NetTest Line interEMULATOR)。

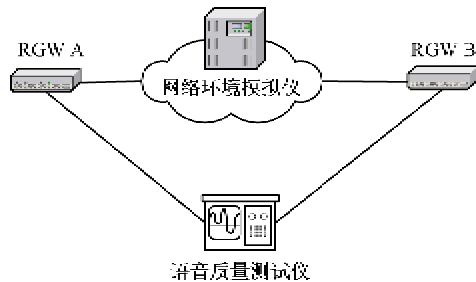


图 5 FSOM 数据采集验证系统

语音质量测试仪通过电话交换局端口连接 VoIP 终端的电话用户端接口,通过模拟摘机、呼叫、通话和挂机过程,可以客观测量 VoIP 系统的 DTMF、语音质量评分 (MOS、PSQM、PESQ)、网络延时及抖动。利用网络环境模拟仪来模拟 Internet 的丢包、延时、抖动、乱序等网络参数。利用该系统可以收集训练网络所需的数据,同时也可以将控制结果在该系统上进行验证。

5.1 典型配置“故障”样本

实验获得的典型配置“故障”样本如表 1 所示,对于每一个特征参数,0.8 为上限值,0.2 为下限值,0.5 为正常值。

表 1 VoIP 语音系统典型“故障”样本

	S_1	S_2	S_3	S_4	S_5	S_6
M_1	0.80	0.40	0.01	0.10	0.20	0.10
M_2	0.60	0.75	0.40	0.10	0.40	0.10
M_3	0.02	0.65	0.10	0.50	0.50	0.50
M_4	0.60	0.03	0.40	0.50	0.20	0.60
M_5	0.60	0.03	0.01	0.50	0.03	0.10
M_6	0.40	0.75	0.70	0.50	0.60	0.80

其中, M_1 、 M_2 、 M_3 、 M_4 、 M_5 、 M_6 是 6 组典型配置“故障”样本序号,代表 6 种故障类型。 S_1 、 S_2 、 S_3 、 S_4 、 S_5 、 S_6 为配置“故障”征兆的序号,代表着 6 种特征参数。它们的具体意义如下:

配置“故障”集合: $\{M_1$: G. 711 问题, M_2 : G. 729 问题, M_3 : 发送周期偏大问题, M_4 : 发送周期偏小问题, M_5 : 播放平均时间偏大问题, M_6 : 平均播放时间偏小问题 $\}$ 。

征兆集合: $\{S_1$: 带宽, S_2 : 丢包率, S_3 : 单向延时, S_4 : 发包周期, S_5 : 抖动, S_6 : 乱序 $\}$ 。

5.2 SOM 网络训练

建立网络模型时,首先应该确定输出层神经元

的个数。输出层神经元的个数 M 与总的故障类型有关。一般情况下,故障类型越多,输出层神经元也越多,需要多次整定才能确定最佳的输出层神经元个数。确定最佳神经元的标准是:既不要造成网络冗余,致使网络资源浪费,又不要过分简化网络,造成网络性能下降。针对本文语音质量推理问题,进行了多次仿真实验,仿真结果比较表明, 8×8 的输出具有故障辨识效果。典型配置“故障”样本的训练结果如图 6 所示。

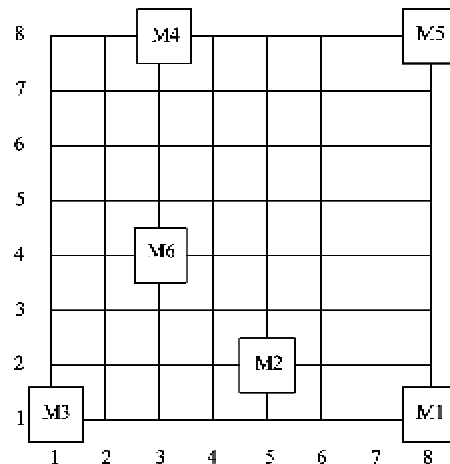


图 6 典型“故障”样本训练结果

5.3 试验验证

为了验证 5.3.7.3 训练的网络特性,通过实验平台采集了 4 组待诊断样本(如表 2 所示),其中 X_1 为带宽, X_2 为丢包率, X_3 为单向延时, X_4 为发包周期, X_5 为抖动, X_6 为乱序。

表 2 待诊断原始向量

	X_1	X_2	X_3	X_4	X_5	X_6
F_1	66.5	3.15	69.0	36.0	19.0	4.28
F_2	78.5	0.10	0.01	41.0	18.5	3.78
F_3	160	4.00	66.0	40.0	20.3	5.14
F_4	85.5	2.80	96.5	37.0	19.3	4.28

利用 5.3.7.1 的故障隶属度对原始样本进行了模糊化处理,结果如表 3 所示。

表 3 待诊断模糊化向量表

	S_1	S_2	S_3	S_4	S_5	S_6
F_1	0.80	0.31	0.11	0.14	0.14	0.18
F_2	0.58	0.04	0.01	0.60	0.10	0.13
F_3	0.06	0.75	0.10	0.60	0.60	0.58
F_4	0.40	0.20	0.40	0.20	0.20	0.20

将以上的模糊向量输入到训练好的 FSOM 网络,仿真结果如图 7 所示。对比图 6 和图 7 可知,样本 F₁ 的故障是 G.711 选择问题,将配置调整为 G.729 后,实测出 PESQ 由 3.40 变为 3.51;样本 F₂ 的故障是平均播放时间(65ms)问题,调整平均播放时间(55ms)后,实测出 PESQ 由 3.70 变为 3.76;样本 F₃ 的故障是发送周期(40ms)问题,调整发送周期(20ms)后,实测出 PESQ 由 3.50 变为 3.80;样本 F₄ 无故障。由此可见,该方法对于语音质量保证具有较为明显的效果。

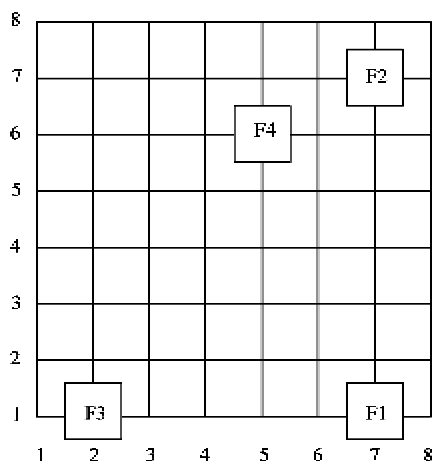


图7 “故障”网络训练结果输出

6 结 论

针对 VoIP QoS 控制问题,提出了一种基于 FSOM 的网络语音服务质量反馈控制机制。该方法包括基于模糊自组织神经网络的语音质量评价机制

和生活会话再协商机制,实现了面向业务的服务质量反馈控制,仿真实验证明了语音业务的动态服务质量控制的有效性。同时,该方法具有一定的通用性,可以将其推广到网络多媒体服务质量控制的其他领域。

参考文献

- [1] 李建东, 晋晓辉. IP 网络 QoS 研究现状. 西安电子科技大学学报, 2002, 29(2): 163-168
- [2] Borosa T, Marsic B, Pocuca S. QoS support in IP multimedia subsystem using DiffServ. In: Proceedings of the 7th International Conference on Telecommunications, Zagreb, Croatia, 2003. 669-672
- [3] Melvin H, Murphy L. Time synchronization for VoIP quality of service. *IEEE Internet Computing*, 2002, 6(3): 57-63
- [4] Mari M, Crespi N. New QoS control mechanism based on extension to SIP for access to UMTS core network via different kinds of access networks. In: Proceedings of the IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, Montreal, Canada, 2005. 150-157
- [5] P.800 ITU-T. Methods for Subjective Determination of Transmission Quality. ITU. 1993
- [6] G.107 ITU-T. The E-model, a Computational Model for Use in Transmission Planning. 1998
- [7] P.861 ITU-T. Objective Quality Measurement of Telephone-Band (300-3400 Hz) Speech Codecs. 1998
- [8] P. 862 ITU-T. Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs. 2001
- [9] 蒋东翔, 王风雨. 模糊自组织神经网络在航空发动机故障诊断中的应用. 航空动力学报, 2001, 16(1): 80-82

A feedback control algorithm of VoIP QoS based on FSOM

Zhang Wu, Wang Jinlin

(National Network New Media Engineering and Technology Research Center,
Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190)

Abstract

To improve the quality of service (QoS) of voice over IP (VoIP), the paper more thoroughly investigates the influence factors to QoS and its relative evaluation methods, and then presents a feedback control algorithm of VoIP QoS based on a fuzzy self-organization map (FSOM) network. A feedback control system model was designed in detail, which introduces a QoS reasoning scheme and an active session re-negotiation scheme. Through establishing the experimental environment, the experimental data were collected and the emulation results were validated. The experimental results demonstrate that the algorithm can effectively improve VoIP QoS. And the algorithm can be extended to QoS controlling scopes of network media.

Key words: voice over IP (VoIP), quality of service (QoS), fuzzy self-organization map (FSOM) network, session negotiation, session initiation protocol (SIP)