

GBS 业务流量的隐马尔可夫模型^①

程静文, 单 洪

(电子工程学院, 合肥 230037)

摘 要: 简单介绍了 GBS(全球广播业务)系统, 对其承载的业务进行分析, 提出了基于隐马尔可夫模型的业务流量建模方法. 在 Qualnet 中搭建了 GBS 仿真场景, 生成并测量了业务流量. 并在 Matlab 中实现了对 GBS 系统中四种业务流量的建模. 结果表明, 该模型只选用数据包大小和数据包间隔等受加密技术影响不大的流量特征进行建模, 更能真实地反映 GBS 系统的业务流量特征. 采用离散型的随机变量, 降低了建模的计算复杂度, 减少了建模时间, 提高了实时性.

关键词: GBS 系统; 隐马尔可夫模型; 业务建模

GBS Service Modeling Based on Hidden Markov Model

CHENG Jing-Wen, SHAN Hong

(Electronic Engineering Institute, Hefei 230037, China)

Abstract: This paper briefly analyses the type of service of the GBS(Global Broadcast Service), put forward an approach of the service modeling based on Hidden Markov Model(HMM). By simulating the GBS in Qualnet, the service traffic trace is captured. The proposed modeling approach is realized in Matlab. The HMM based modeling approach only used two variables: packet size and inter packet time, which are hardly influenced by encryption techniques. The experiment results show that by the use of discrete variables, the proposed modeling approach can decrease the calculating complexity and model constructing time, improve the time cost.

Key words: GBS; HMM(hidden markov model); service modeling

1 系统概述

全球广播业务(GBS)是在商用直播卫星系统(DBS)的基础上发展起来的军用信息传输业务. 它利用商用卫星大容量、高带宽的数据传输能力为终端用户提供战术视频、音频及文件等数据广播.

2 GBS广播系统

2.1 GBS 系统结构

GBS 系统由转发卫星、主注入点(PIP)、战术注入点(TIP)、卫星广播管理站(SMB)、后方信息资源中心和用户接收终端等几部分组成. 其部署结构如图 1 所示. 卫星系统主要搭载在 ka 波段的专用卫星和 ku 波段的商用卫星上. PIP 是经过卫星广播的所有地面信息固定注入端口. TIP 与 PIP 相比, 具有移动能力.

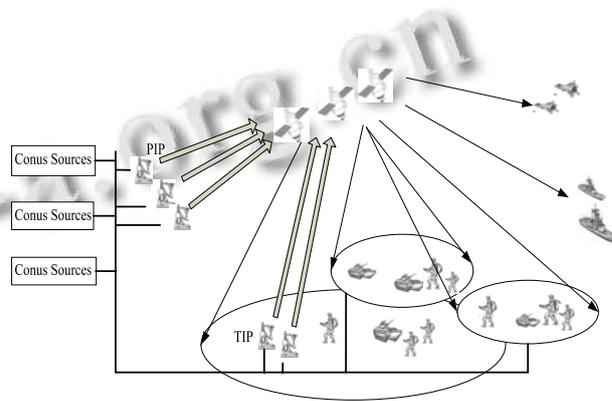


图 1 GBS 部署结构

2.2 GBS 系统业务分析

GBS 系统使用“智能发送”(smart push)和“用户牵引”(user pull)两种方式向终端节点分发业务信息. 对

^① 收稿时间:2012-12-10;收到修改稿时间:2013-02-04

“智能发送”方式,经 GBS 系统向各终端用户广播的业务如气象预报、教育训练、综合性新闻视频和图片数据等信息。

对“用户牵引”方式,经用户请求后 GBS 系统可以向用户终端提供端到端的 Web 业务,文件高速下载业务, Telnet 业务和流媒体业务等互联网应用。

为有效的对 GBS 系统的业务进行识别,将 GBS 系统传输的信息分为以下几种业务类型:

(1) 流媒体业务:由 IP 数据包来承载的各种实时音、视频业务。

(2) Web 业务:用户可通过卫星广播管理站(SBM)经 DVB-RCS 回传信道向后方资源中心请求 Web 业务,并通过浏览器进行浏览。

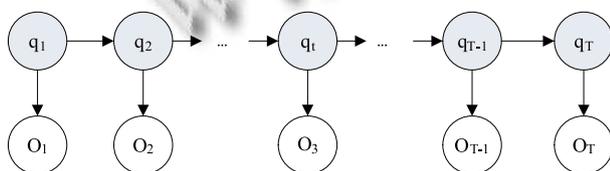
(3) 文件业务:以文件形式传输数据的业务,其数据内容包括指挥命令、天气预报、地图等。

(4) Telnet 业务:用户终端可以使用 Telnet 服务远程登录目标主机,实现对目标网络的管理和配置。

以上这几种业务构成了 GBS 系统最基本的业务载体,通过卫星向前沿终端用户广播的各种业务都可以通过这些业务载体承载。

3 隐马尔可夫模型

隐马尔可夫模型(HMM)^[1]是在马尔可夫模型的基础上发展起来的.它的理论基础是由 Baum 等人最先提出,在语音信号识别、故障诊断和生物信息学研究等领域得到了广泛的应用.与马尔可夫模型不同的是,隐马尔可夫模型的观测事件并不是与状态一一对应,而是通过一组观测概率分布相联系.它是一个双重随机过程,如图 2 所示.其中之一是马尔可夫链,这是一个基本的随机过程,描述状态之间的转移.另一个随机过程描述状态和观测变量之间的统计对应关系。



$$Q_i \in S = \{s_1, s_2, \dots, s_N\}$$

$$O_i \in V = \{v_1, v_2, \dots, v_M\}$$

图 2 HMM 示意图

3.1 HMM 的定义

一个 HMM 可以由以下参数描述:

(1) N : 模型中马尔可夫链的状态数,状态集合表示为 $S = \{S_1, S_2, \dots, S_N\}$,并用 q_t 表示 t 时刻的状态。

(2) M : 每个状态可以观察到的观察值数.标记 M 个观察值为 $V = \{V_1, V_2, \dots, V_M\}$, O_t 为 t 时刻观察到的观察值。

(3) 状态转移概率分布 $A = [a_{ij}]$, 这里 $a_{ij} = P\{q_{t+1} = s_j \mid q_t = s_i\}$, $(1 \leq i, j \leq N)$. 其满足 $\sum_{j=1}^N a_{ij} = 1$ 。

(4) 观察序列的概率分布矩阵 $B = [b_j(k)]$, 表示状态 S_j 输出相应观察值的概率,其中 $b_j(k) = P\{o_t = v_k \mid q_t = s_j\}$ $(1 \leq j \leq N, 1 \leq k \leq M)$. 其满足 $\sum_{j=1}^N b_j = 1$ 。

(5) 初始状态分布 $\pi = [\pi_i]$, $\pi_i = P\{q_1 = s_i\}$, $(1 \leq i \leq N)$. 用于表示观察序列 O 在 $t=1$ 时刻所处的状态 q_1 属于模型中各状态的概率分布,其满足 $\sum_{i=1}^N \pi_i = 1$ 。

3.2 HMM 建模的基本方法

定义模型 λ 的似然函数 $P(O \mid \lambda)$ 为由模型 λ 产生观察值序列 O 的概率. HMM 建模所需解决的问题是: 给定一个观察值序列 $O = \{O_1, O_2, \dots, O_T\}$, 训练初始模型参数至参数收敛且 $P(O \mid \lambda)$ 最大。

可使用 Baum-Welch 算法^[1]解决 HMM 参数重估计问题,它具有严格的数学证明,收敛速度快,能保证每一步迭代都朝着正确的方向进行. Baum-Welch 算法的基本思想是: 按照参数重估公式从现有的模型 λ' 估计出新的模型 λ , 可以证明, λ 在表示观察值序列方面由于 λ' , 即 $P(O \mid \lambda') \leq P(O \mid \lambda)$, 用 λ 代替 λ' , 重复上述过程直到模型参数处于收敛状态,即得到最大似然模型。

3.3 基于 HMM 的 GBS 系统业务流量建模

HMM 适用于动态过程时间序列建模,并具有强大的时序模式分类能力.其只需要少量的数据就能建立完整的特征库。

GBS 系统的业务分发流程本身就是一个双重随机过程: 用户终端收到的数据包外部特征是可观察的序列,但其业务请求分发过程是按照隐含的交互规则(不可观察的状态)进行通信.因此, HMM 可以准确对 GBS 系统中的业务进行建模。

3.3.1 业务流量特征参数的选取

GBS 系统所传输的数据经过了信源加密,因此,我们在选取流量特征时,选用对加密技术不敏感的包长和包到达时间间隔作为业务的流量特征.为了保证

参数特征数量级的一致性,我们对包到达时间间隔取对数处理^[2].

这样,我们可以得到一个二维业务特征矢量序列,记为: $O=\{O[1], O[2], \dots, O[k], \dots\}$. 其中 k 表示第 k 个数据包, $O[k]=(d[k], b[k])^T$. $d[k]=10\log_{10}(IPT/1)$ 表示第 k 个数据包到达时间间隔, $b[k]$ 表示第 k 个数据包的长度. 使用矢量量化技术对二维矢量序列进行降维处理, 将降维后的一维序列作为建模时输入的观测序列.

HMM 链具有多种不同的逻辑结构, 在实际建模时, 要根据实际情况进行选择. 在本文中, 由于我们使用的特征参数是按照时间序列进行统计的, 并且在无线卫星信道中, 数据包可能丢失, 状态无法按序转移. 因此, 我们选择从左到右有跨越的 HMM 模型对 GBS 业务进行建模, 如图 3 所示.

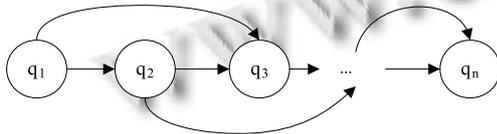


图 3 从左到右有跨越的 HMM 模型

理论上证明, Baum-Welch 算法只能收敛到 $P(O|\lambda)$ 的局部最大值, 因此, 要获得最优的模型必须选好初始模型参数. 一般认为, π 和 A 初值的选取对结果影响不大, 只需要满足约束条件即可^[3], 本文中我们对其均匀取值. 参数 B 的取值对结果影响较大, 本文采用 Segmental K-Means 算法^[4]对参数 B 的初值进行选取. 确定初始参数后, 使用 Baum-Welch 算法对 GBS 中不同业务进行建模, 建模过程如图 4 所示.

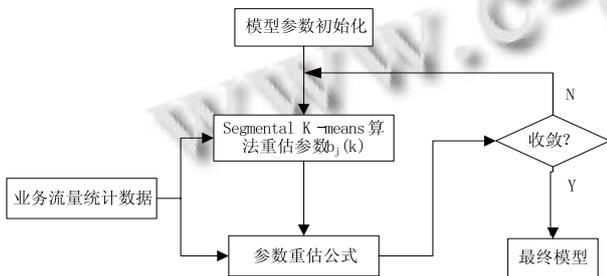


图 4 业务流量建模过程

4 GBS系统业务流量仿真建模

4.1 仿真环境

QualNet 是一个针对无线网络进行仿真的网络模拟器, 可以提供模型设计、仿真、数据收集和数据分析

析等各种研究工具. 本文中使用了 QualNet 搭建 GBS 系统的网络仿真场景, 构造并测量业务仿真流量, 在 Matlab 平台上对 GBS 系统业务流量进行建模.

4.2 业务仿真流量的构造与测量

在对实际网络业务流量特征研究的基础上, QualNet 网络模拟器建立了一些业务流量生成器, 用于产生不同类型的业务流量. 本文根据搭建的 GBS 系统网络仿真场景, 针对不同的终端节点生成了以下几种业务仿真流量:

- (1) 文件业务: 由 FTP 业务流量生成器生成, 主要用于传输图片、文档等文本文件.
- (2) 流媒体业务: 由 CBR 业务流量生成器生成. CBR 业务流量生成器用于模拟固定比特率的应用层业务流量, 常被用于模拟音频或视频业务的通信流量.
- (3) Web 业务: 由 HTTP 业务流量生成器生成. 用户可以通过回传链路请求 Web 服务, 获取互联网信息.
- (4) Telnet 业务: 由 Telnet 业务流量生成器生成.

设定 QualNet 仿真时间为 30 分钟, 记录以上业务生成器生成的流量, 对 trace 文件中业务流量样本统计, 如表 1 所示. 按数据包到达时间对各种业务流的数据包长度和包到达时间间隔数据进行统计处理, 得到所需业务特征序列.

表 1 业务仿真流量统计

GBS 业务	数据包总数	样本大小
流媒体业务	6807	975M
文件业务	3173	247M
Web 业务	10349	842M
Telnet 业务	5926	694M

4.3 实验结果与分析

将处理后的特征序列作观测值序列, 在 Matlab 平台上用 Baum-Welch 算法分别对以上几种业务流量建立 HMM 模型, 结果如下:

- (1) 流媒体业务: 建模时取 $N=4$, 如图 5 所示, 迭代 19 次后, 当 $Loglik=-21.927766$ 时, 模型收敛.
- (2) 文件业务: 建模时取 $N=5$, 如图 6 所示, 迭代 20 次后, 当 $Loglik=-16.442470$ 时, 模型收敛.
- (3) Telnet 业务: 与文件业务一样, 建模时取 $N=4$, 如图 7 所示, 迭代 16 次后, 当 $Loglik=-27.978559$ 时, 模型收敛.
- (4) Web 业务: 由于 Web 业务的交互过程更加复杂, 建模时取 $N=7$, 如图 8 所示, 迭代 28 次后, 当 $Loglik$

=-15.824286 时, 模型收敛.

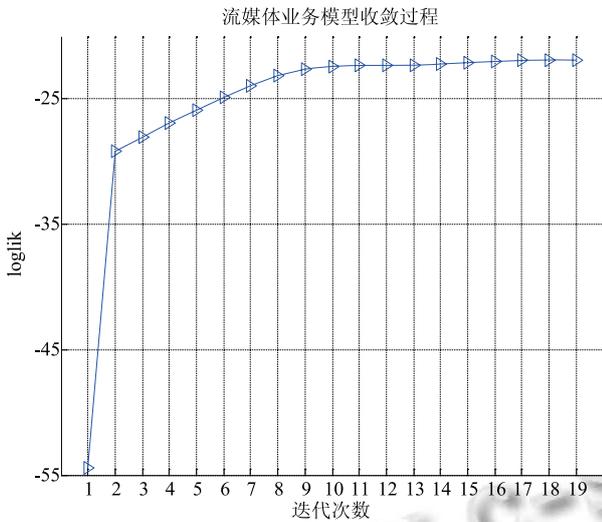


图 5 流媒体业务模型收敛过程

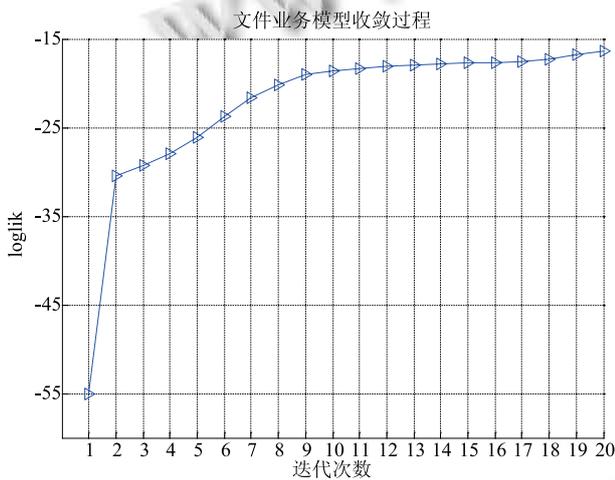


图 6 文件业务模型收敛过程

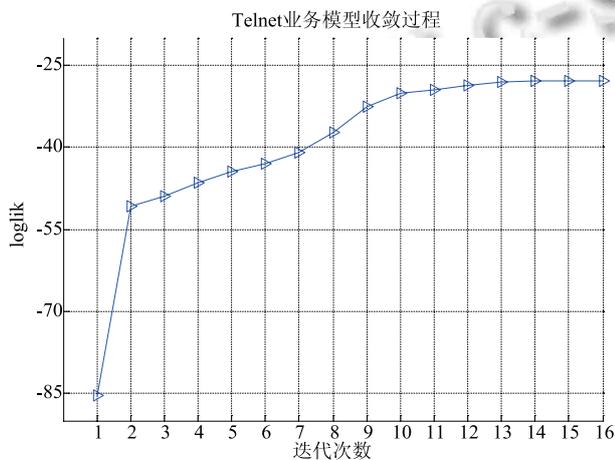


图 7 Telnet 业务模型收敛过程

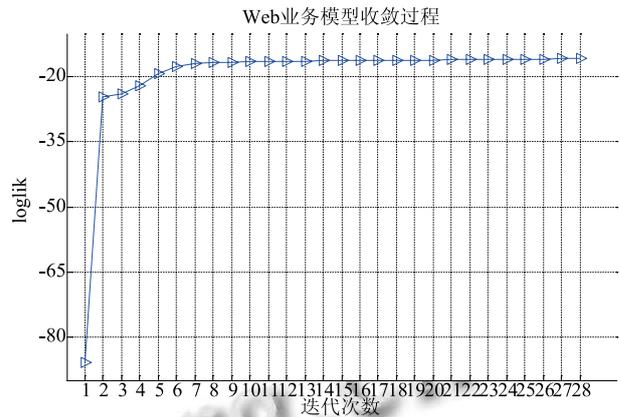


图 8 Web 业务模型收敛过程

文献[5]使用基于连续变量的 HMM 对互联网中的流量进行建模, 计算复杂度高, 当参数较多, 训练样本不够充足时, 模型参数训练的精度不高. 本文所使用离散型 HMM 和矢量降维后的特征序列对模型参数进行训练, 计算量较少, 实时性较高.

5 结束语

本文在分析 GBS 广播系统业务类型的基础上, 提出了基于隐马尔可夫模型的业务建模方法, 采用了对加密技术不敏感的数据包长度和包到达时间间隔作为业务特征, 详细介绍了建模算法和建模过程. 并且在 Qualnet 中搭建了 GBS 仿真场景, 构造并测量了业务流量. 在 Matlab 平台上实现了建模过程. 结果表明, 采用离散型随机变量的基于隐马尔可夫模型的业务建模降低了建模的计算复杂度, 减少了建模时间, 提高了实时性.

参考文献

- 1 Rabiner A, Lawrence R. A tutorial on hidden markov models and selected application in speech recognition. Proc. of the IEEE, 1989,(77)2:257-285.
- 2 何中阳, 杨白薇, 李鸥. 基于隐马尔可夫模型的协议识别技术. 解放军信息工程大学学报, 2011, 12(5): 596-600.
- 3 赵力. 语音信号处理. 北京: 机械工业出版社, 2003. 31-118.
- 4 Barlas Y, Kanar K. A dynamic pattern-oriented test for model validation. Proc. of 4th System Science European Congress. Spain: Valencia, 1999: 269-286.
- 5 Dainotti A, Pescapé A, Salvo Rossi P. An HMM approach to internet traffic modeling. Proc. of the IEEE GLOBECOM. 2006: 1-6.