

大尺度 IP 骨干网络流量矩阵估计方法研究

蒋定德^{1,2}, 王兴伟¹, 郭磊¹, 许争争¹, 陈振华¹

(1. 东北大学信息科学与工程学院, 辽宁沈阳 110004; 2. 电子科技大学宽带光纤传输与通信网技术重点实验室, 四川成都 610054)

摘要: 流量矩阵估计是当前的热点研究问题, 它被网络操作员用来进行负载均衡、路由最优化、流量侦测、网络规划等等。然而, 流量矩阵估计本身固有的高度病态特性, 使得精确地估计流量矩阵成为具有挑战性的研究课题。本文研究大尺度 IP 骨干网络的流量矩阵估计; 基于 RBF (Radial Basis Function) 神经网络, 提出一种新的估计方法 TMRI (Traffic Matrix Recurrence Inference)。TMRI 利用 RBF 神经网络强大的建模功能来建模流量矩阵估计问题, 将这一问题的病态特性克服于 RBF 神经网络的训练过程中, 从而避免复杂的数学建模过程。并在所建立的估计模型基础上, 将流量矩阵估计描述为约束条件下的最优化过程, 通过迭代寻优, TMRI 能进一步克服这一问题的病态特性。仿真结果表明 TMRI 能精确地估计流量矩阵和追踪它的动态变化, 与以前的方法相比, 具有更强的抗噪声性能和显著的性能改善。

关键词: 流量矩阵估计; 非平稳流量; RBF 神经网络; 病态特性; 最优化

中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112 (2011) 04-0763-09

Approach of Traffic Matrix Estimation in Large-scale IP Backbone Networks

JIANG Ding-de^{1,2}, WANG Xing-wei¹, GUO Lei¹, Xu Zheng-zheng¹, CHEN Zhen-hua¹

(1. College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110004, China;

2. Key Laboratory of Broadband Optical Fiber Transmission and Communication Networks, UESTC, Chengdu, Sichuan 610054, China)

Abstract: Traffic matrix estimation is an interesting research problem at present. Network operators use it to conduct load balancing, route optimization, traffic detecting, network dimensioning and so on. However, the highly ill-posed nature of traffic matrix estimation itself makes it being a challenging research subject to estimate accurately traffic matrix. This paper studies traffic matrix estimation in large-scale IP backbone networks. Based on RBF (radial basis function) neural network, a novel estimation method, namely TMRI (traffic matrix recurrence inference), is proposed. TMRI exploits the powerful modeling ability of RBF neural network to model traffic matrix estimation. The ill-posed nature of this problem will be overcome in the process of training the RBF neural network. Accordingly, the complex process of mathematic modeling can be avoided. Built on this estimation model, traffic matrix estimation is described into the optimal process under the constraints. By seeking the recurrent optimal solution, TMRI can further get rid of the ill-posed nature of this problem. Simulation results show that TMRI can accurately estimate traffic matrix and track its dynamics, and in contrast to previous methods, it holds the stronger robustness to noise and more evident performance improvement.

Key words: traffic matrix estimation; nonstationary traffic; radial basis function (RBF) neural network; ill-posed nature; optimization

1 引言

流量矩阵作为网络管理和流量工程的重要输入参数, 它表示网络中所有 OD (Origin-Destination) 节点对之间流动的流量。针对网络管理所作的许多重要决策都取决于网络中的流量在网络中怎样流动; 流量矩阵提供给网络操作员关于当前网络状态有价值的和全局的信息。

然而, 尽管流量矩阵非常重要, 但是直接测量流量矩阵, 特别是直接测量大尺度 IP 骨干网络上的流量矩阵非常困难, 甚至是不可能的^[1]。因此, 流量矩阵估计已成为当前的热点研究问题, 引起了研究人员的广泛关注^[2-9]。

然而, 如何有效地克服这一问题的病态特性是当前面临的主要挑战。Cao 等人^[10]将 OD 流建模为独立同分布的正态模型, 并修改传统的 EM 算法来获得局域网上

流量矩阵的估计值;Gunnar 等人^[1]对 IP 骨干网络流量矩阵进行了研究,提出流量矩阵具有时空分布特征,并分析了 OD 流的均方幂律关系;Zhang 等人^[3,4]研究了大尺度 IP 骨干网络流量矩阵,他们通过重力模型来获得流量矩阵的先验信息,提出了基于重力模型的流量矩阵反演方法;Tan 等人^[5]通过计算路由矩阵的 {1}-Inverse 来估计流量矩阵.然而,文献[11]分析认为,文献[10]所提出的估计方法是基于统计技术来获得流量矩阵估计值,这些方法对流量矩阵的先验信息非常敏感,而文献[3,4]提出的方法尽管降低了对先验信息的敏感性,但估计结果仍具有较大的误差,特别是当对网络流量的假设不满足重力模型条件时,估计误差更大.因此,由于流量矩阵估计的高度病态特性,要精确地估计流量矩阵非常困难.

本文研究了大尺度 IP 骨干网络流量矩阵的估计问题,基于 RBF 神经网络,提出一种新的估计方法 TMRI. RBF 神经网络^[12]是一种强大的建模工具,可进行线性和非线性、时变和非时变系统的建模,它被广泛用于信号处理、系统建模、生物医学等领域. TMRI 利用修改的 RBF 神经网络强大的建模功能,来捕获流量矩阵特征,建模流量矩阵估计问题.通过输入输出数据对训练 RBF 神经网络,流量矩阵估计的病态特性被逐步克服,并建立起流量矩阵的估计模型.基于建立的估计模型基础上,流量矩阵估计被描述为约束条件下的最优化过程,通过迭代寻优,流量矩阵估计的病态特性被 TMRI 进一步克服.仿真实验使用 Abilene^[13]和 GÉANT^[14]网络上的真实数据来验证 TMRI,仿真结果表明 TMRI 优于以前的估计方法,具有显著的性能改善.

2 问题阐述

图 1(a)和(b)分别表示了 Abilene 和 GÉANT 网络上的流量.从图 1 可以看到,Abilene 和 GÉANT 网络上的流量都是动态变化的,这是一种时变网络流量,要预测和估计这种流量非常困难.图 2(a)和(b)分别表示了 Abilene 和 GÉANT 网络上流量的 ACF (Autocorrelation Function).图 2 表明 Abilene 和 GÉANT 网络上流量的 ACF 都缓慢地趋近于零,这说明两个网络上的流量都是非平稳的.网络流量的这种非平稳性增加了流量矩阵预测的难度.如何准确捕获网络流量的时变非平稳特性是流量矩阵精确估计面对的主要困难之一.

流量矩阵表示在网络中所有 OD 节点间流动的流量,直接测量它非常的困难.在大尺度 IP 骨干网络中,链路负载是流量矩阵根据路由矩阵在链路上汇聚而成,它们之间具有某种线性约束关系.假设大尺度 IP 骨干网络有 p 个节点, Q 条链路,则有 $P = p^2$ 条 OD 流.假设时刻 t ,网络流量矩阵和链路负载分别为: $x(t) =$

$(x_1(t), x_2(t), \dots, x_p(t))^T$ 和 $y(t) = (y_1(t), y_2(t), \dots, y_Q(t))^T$,其中 $x_i(t)(i = 1, 2, \dots, P)$ 表示第 i 条 OD 流, $y_j(t)(j = 1, 2, \dots, Q)$ 表示第 j 条链路负载. IP 网络的路由矩阵表示为 $A = (A_{ij})_{Q \times P}$,其中,如果第 j 条 OD 流通过第 i 条链路,则 $A_{ij} = 1$,否则 $A_{ij} = 0$.流量矩阵、链路负载和路由矩阵之间的线性约束关系可以用如下的等式表示:

$$y(t) = Ax(t) \tag{1}$$

通过 SNMP (Simple Network Management Protocol) 测量可直接获得链路负载 $y(t)$, 根据网络配置信息可直接得到路由矩阵 A , 则流量矩阵估计就是在给定链路负载 $y(t)$ 和网络路由矩阵 A 的条件下, 如何精确获得流量矩阵的估计值. 如果流量矩阵 A 是一个满秩的方阵, 则等式(1)的解是唯一确定的. 然而, 在大尺度 IP 骨干网络中, 链路数目远远小于 OD 流的数目, 即 $Q \leq P$, 这表明等式(1)是一个高度欠定的线性问题, 满足等式(1)的解不唯一, 如何从无穷多个解中获得所需要的解是困难的. 因而, 流量矩阵估计是一个高度病态的问题, 如何有效克服这一问题的病态特性是目前面对的主要挑战.

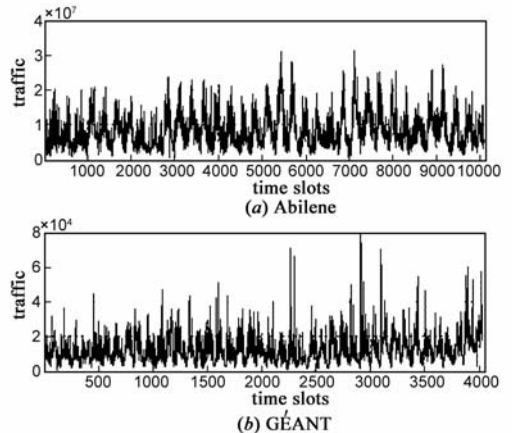


图1 Abilene和GÉANT网络的流量

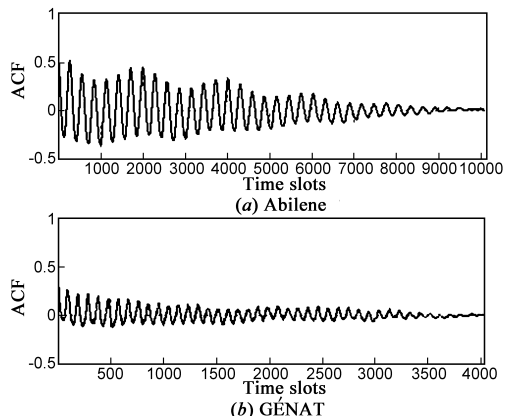


图2 Abilene和GÉANT网络的流量ACF

流量矩阵估计作为目前感兴趣和具有挑战性的研

究问题,研究人员采取各种方法来克服这一问题的病态特性,以获得尽可能精确的估计值. Soule 等人^[15]提出通过改变网络路由,然后在这种新的网络路由下进行新的链路负载测量来将欠定问题转变为满秩问题,以克服流量矩阵估计的病态特性; Bermolen 等人^[16]在二阶距统计下使用函数均方关系导出 Fisher 信息矩阵的解析表达式,然后获得流量矩阵估计器方差的 Cramer-Rao 下界,并通过 Cramer-Rao 下界,他们能获得流量矩阵估计值; Tang 等人^[17]使用分配模型来估计流量矩阵; Juva 等人^[18]基于链路负载协方差来快速估计流量矩阵; Erramill 等人^[19]提出用独立连接模型来获得流量矩阵的先验信息; Shioda 等人^[20]分析了 VPN 中的流量矩阵估计; Cao 等人^[21]提出一种伪似然方法来估计高速网络中的流量矩阵. 然而,所有这些方法都需要建立关于 OD 流的数学模型,并执行统计反演,或者结合部分 OD 流的直接测量来反演流量矩阵. 因此,需要复杂的数学计算.

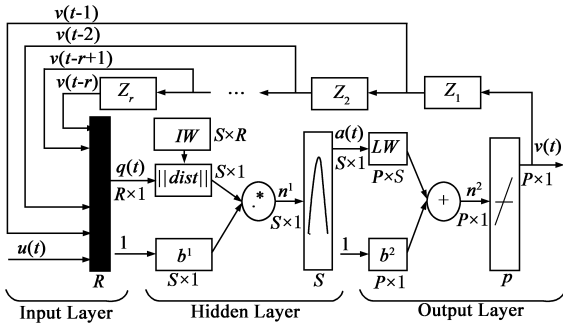


图3 用于流量矩阵估计的RBF神经网络结构

不同于上面的方法,本文提出的 TMRI 方法没有作关于 OD 流的任何假设,它直接使用 RBF 神经网络来建模大尺度 IP 流量矩阵估计问题,从而避免复杂的数学计算. 由于 RBF 神经网络具有学习和归纳的能力, TMRI 对 OD 流的假设并不敏感,它对噪声的干扰也足够健壮. 而且,由于 RBF 神经网络的并行结构, TMRI 是一种轻量级的方法,能快速的预测大尺度 IP 流量矩阵. 并且,估计模型的输出满足等式(1)的约束条件. 因此,根据估计模型获得的输出是满足等式(1)的一个解. 基于所获得的这个解,采用迭代寻优过程来进一步克服流量矩阵估计问题的病态特性,以获得流量矩阵的精确估计.

3 流量矩阵估计建模

由于流量矩阵具有时间的、空间的、时间空间的相关性,以及时变非平稳特性,因此,本节提出一种修改的 RBF 神经网络结构,以便更准确的捕获流量矩阵的这些特征. 并在修改的 RBF 神经网络基础上,提出大尺度 IP 骨干网络流量矩阵的估计模型.

3.1 用于流量矩阵估计的 RBF 神经网络模型

RBF 神经网络具有快速学习和归纳的能力,是一种强大的建模工具,能很好地用于流量矩阵估计问题. 为了精确地估计流量矩阵,本节提出一种适用于大尺度 IP 骨干网络流量矩阵估计问题的修改的 RBF 神经网络模型,如图 3 所示. 图 3 中, RBF 神经网络包括输入层、隐含层、输出层; $u(t) = (u_1(t), u_2(t), \dots, u_L(t))^T$ 为修改的 RBF 神经网络的输入; IW 和 LW 分别表示隐含层和输出层的网络权值矩阵; $\|dist\|$ 表示计算 R 维向量 $q(t)$ 与输入权值矩阵 IW 的行向量间的距离; S 和 P 分别表示隐含层和输出层的神经元数目; n^1 和 n^2 则分别表示隐含层和输出层的网络净输出; 第一和第二个圆圈中的“ \cdot ”和“ $+$ ”分别表示两个输入做向量点乘和向量加法运算; b^1 和 b^2 分别表示隐含层和输出层的网络阈值; $a(t) = (a_1(t), a_2(t), \dots, a_S(t))^T$ 为隐含层的输出; $v(t) = (v_1(t), v_2(t), \dots, v_N(t))^T$ 为修改的 RBF 神经网络的输出; $Z_d = z^{-d}I$ ($d = 1, 2, \dots, r$, 其中 $z^{-1}I$ 表示单位时间延迟, r 为网络引入的时间延迟数); $v(t-1), v(t-2), \dots, v(t-r+1)$, 和 $v(t-r)$ 分别表示 $v(t)$ 经延迟 Z_1, Z_2, \dots, Z_{r-1} , 和 Z_r 后的输出. 从图 3 可以看到,本节所提出的修改的 RBF 神经网络是一个多输入多输出系统,它完成了从输入 $u(t)$ 到输出 $v(t)$ 的映射,即可得到下面的等式:

$$\begin{cases} v_i(t) = \phi_i(\sum_{j=1}^M LW_{ij} a_j(t) + b_i^2), & i = 1, 2, \dots, P \\ a_j(t) = \varphi_j(\|IW - q\| * b_j^1), & j = 1, 2, \dots, S \\ q(t) = (u(t)^T, v(t-1)^T, \dots, v(t-r)^T)^T \end{cases} \quad (2)$$

其中, φ_j 表示隐含层的第 j 个径向基函数,一般是高斯函数,即 $\varphi_i(s) = e^{-s^2}$; ϕ_i 表示输出层的第 i 个激活函数,为线性函数. 等式(2)表示下面的映射关系:

$$\begin{cases} v_1(t) = f_1(u(t), v(t-1), \dots, v(t-r)) \\ v_2(t) = f_2(u(t), v(t-1), \dots, v(t-r)) \\ \dots \\ v_p(t) = f_p(u(t), v(t-1), \dots, v(t-r)) \end{cases} \quad (3)$$

其中, f_i 表示 $u(t), v(t-1), v(t-2), \dots, v(t-r)$ 到 $v_i(t)$ 的映射 ($i = 1, 2, \dots, P$). 则得到如下等式:

$$v(t) = f(u(t)^T, v(t-1)^T, \dots, v(t-r)^T) \quad (4)$$

其中, $f = (f_1, f_2, \dots, f_p)$. 等式(4)表明,图 3 所修改的 RBF 神经网络通过将网络的输出延迟反馈到网络的输入,使得修改后的网络输出 $v(t)$ 不仅和网络的输入 $u(t)$ 相关,而且和网络以前时刻的输出 $v(t-z)$ (其中 $z = 1, 2, \dots, r$) 相关. 这种映射关系有利于捕获流量矩阵的时间相关性. 另外,等式(4)也表明,图 3 所示的多输入多输出结构,以及 RBF 神经网络本身的并行结构有

利于捕获流量矩阵的空间相关性.更重要的是,RBF神经网络强大的学习和归纳的能力,有利于准确地建立起流量矩阵的估计模型.

3.2 流量矩阵估计模型

基于上面修改的 RBF 神经网络结构,图 4 表示了用于大尺度 IP 骨干网络流量矩阵的估计模型,其中,“Data Pretreating”和“Data Posttreating”分别表示数据预处理和后处理过程;“Training”和“Predicting”分别表示训练和预测过程; $Y(t)$ 和 $x(t)$ 分别为估计模型的输入和输出.为了更好地捕获流量矩阵的特征,链路负载的时间相关性被引入到估计模型中.估计模型的输入表示为:

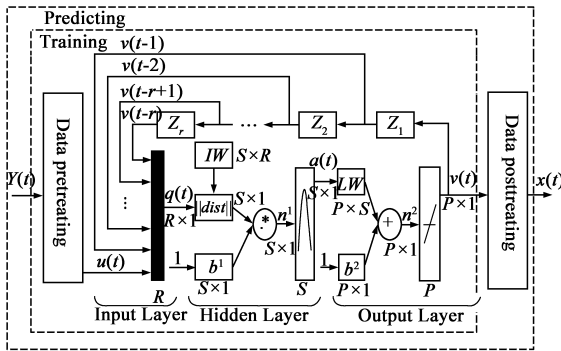


图4 流量矩阵估计的多输入多输出模型

$$Y(t) = (y(t)^T, y(t-1)^T, \dots, y(t-h)^T)^T \quad (5)$$

其中, h 为引入的最大时间延迟.从图 4, 得到如下等式:

$$u(t) = \xi(Y(t)) \quad (6)$$

$$\begin{cases} x(t) = \gamma(v(t)) \\ \text{s.t.} & y(t) = Ax(t) \\ & x_i(t) \geq 0, \quad i = 1, 2, \dots, P \end{cases} \quad (7)$$

其中, ξ 和 γ 分别表示数据预处理和后处理过程.等式(7)表明,数据后处理过程是在流量矩阵所满足的约束条件下进行的数据处理,这样有利于保证模型的估计值更接近真实值.根据等式(4)~(7),则得到如下训练过程:

$$v(t) = H(y(t)) \quad (8)$$

以及预测过程:

$$\begin{cases} x(t) = G(y(t)) \\ \text{s.t.} & y(t) = Ax(t) \\ & x_i(t) \geq 0, \quad i = 1, 2, \dots, P \end{cases} \quad (9)$$

其中, H 和 G 分别表示从 $y(t)$ 到 $v(t)$ 和到 $x(t)$ 的映射关系.等式(8)表明,根据图 4 所示的训练框图,经过 RBF 神经网络的学习和归纳,能建立准确捕获流量矩阵特征的模型,同时,结合流量矩阵满足的约束条件,就可建立等式(9)所表示的流量矩阵估计模型,完成图 4 预测框图表示的预测过程.从等式(8)和(9)可以看到,本节所提出的图 4 表示的估计模型,是一种多输入多输

出模型,这种模型结构有利于准确捕获流量矩阵特征.

4 流量矩阵递归反演

由于流量矩阵估计的病态特性.以及网络流量的时变非平稳和时间空间相关性特征,要精确估计流量矩阵非常困难.基于上面提出的估计模型,下面将流量矩阵描述为约束条件下的最优化过程,通过迭代寻优,来精确获得流量矩阵估计值,进一步克服流量矩阵估计本身的病态特性.

4.1 流量矩阵的样本协方差矩阵

以往研究表明,每日、每周、每月甚至每年的 IP 网络流量矩阵具有相似性,这种相似性说明流量矩阵具有明显的时间相关性,它具有一定的周期性.从图 1 也可以看到,Abilene 和 GÉANT 上的网络流量表示出明显的周期和时间相关性.可见,流量矩阵的历史数据能反映当前的某些特征,它能为当前流量矩阵预测提供某些附加信息,从而为获得准确的流量矩阵估计提供必要的约束条件.因此,本节分析流量矩阵的样本协方差矩阵.

假设 τ 个时刻的流量矩阵 $x(1), x(2), \dots, x(\tau)$ 已知,则流量矩阵的样本协方差矩阵 C_x 可表示为:

$$C_x = \frac{1}{\tau-1} \left(\sum_{i=1}^{\tau} (x(i) - \frac{1}{\tau} \sum_{i=1}^{\tau} x(i)) \right) \cdot \left(\sum_{i=1}^{\tau} (x(i) - \frac{1}{\tau} \sum_{i=1}^{\tau} x(i)) \right)^T \quad (10)$$

从等式(10)可以看到,样本协方差矩阵 C_x 的对角线元素表示了流量矩阵的时间相关性,非对角线元素表示了流量矩阵的时间空间相关性.因此,样本协方差矩阵 C_x 能捕获流量矩阵的时间空间相关性.

4.2 递归反演

在大尺度 IP 骨干网络中,流量矩阵满足等式(1)的线性约束.由于流量矩阵的每个元素都是实际可测的数据,则每条 OD 流也满足非负的约束条件.对于大尺度 IP 骨干网络流量矩阵,可得到如下最优化目标函数:

$$\begin{cases} \min & (x(t) - x_0(t))^T C_x^{-1} (x(t) - x_0(t)) \\ \text{s.t.} & y(t) = Ax(t) \\ & x_i(t) \geq 0, \quad i = 1, 2, \dots, P \end{cases} \quad (11)$$

其中, C_x^{-1} 为流量矩阵的样本协方差矩阵的逆矩阵, $x_0(t)$ 为一已知的流量矩阵.等式(11)表明,通过将样本协方差矩阵的逆矩阵 C_x^{-1} 引入到目标函数 $\min(x(t) - x_0(t))^T C_x^{-1} (x(t) - x_0(t))$ 中,在条件 $x_i(t) \geq 0 (i = 1, 2, \dots, P)$ 和 $y(t) = Ax(t)$ 的约束下,流量矩阵估计为满足样本协方差矩阵所表示的流量矩阵时间空间相关性的与已知流量矩阵 $x_0(t)$ 距离最小的取值.

根据拉格朗日乘法,上述条件最优化问题可以

变换为:

$$\min (y(t) - Ax(t))^T(y(t) - Ax(t)) + \lambda (x(t) - x_0(t))^T C_x^{-1}(x(t) - x_0(t)) \quad (12)$$

其中, λ 为某一常数, 一般取值为 0.01 ~ 1.

$$\text{令 } \Delta x(t) = x(t) - x_0(t) \quad (13)$$

将等式(13)带入等式(12), 得到:

$$\min (y(t) - Ax_0(t) + A\Delta x(t))^T(y(t) - Ax_0(t) + A\Delta x(t)) + \lambda (\Delta x(t))^T C_x^{-1}\Delta x(t) \quad (14)$$

根据矩阵分析和最优化理论, 得到等式(14)的最优解:

$$\Delta x(t) = (A^T A + \lambda C_x^{-1})^{-1} A^T (y(t) - Ax_0(t)) \quad (15)$$

根据等式(15), 得到如下迭代估计等式:

$$\begin{cases} x^{k+1}(t) = x^k(t) + \Delta x^{k+1}(t) \\ \Delta x^{k+1}(t) = (A^T A + \lambda C_x^{-1})^{-1} A^T (y(t) - Ax^k(t)) \\ x^0(t) = \hat{x}_r(t) \end{cases} \quad (16)$$

其中, k 为迭代变量, $\hat{x}_r(t)$ 为通过流量矩阵估计模型获得的初始值. 等式(11) ~ (16) 表明, 通过将流量矩阵估计表示为约束条件下的最优化过程, 经过不断迭代寻优, 流量矩阵的估计值将渐渐逼近真实值, 从而克服流量矩阵估计问题的病态特性, 并最终获得流量矩阵的精确估计.

4.3 反演算法

上面给出了整个 TMRI 方法的推理过程, 下面给出这一方法的完整步骤:

步骤 1 初始化图 4 所示估计模型;

步骤 2 用 $y(t)$ 、 $x(t)$ 作为输入输出数据对根据图 4 训练框图训练模型, 得到 RBF 神经网络权值矩阵, 构建等式(9)表示的流量矩阵估计模型;

步骤 3 根据等式(9)获得流量矩阵估计初始值 $\hat{x}_r(t)$;

步骤 4 根据等式(10)计算流量矩阵的样本协方差矩阵 C_x ;

步骤 5 给出误差限 δ 和 ϵ 、最大迭代步数 K , 并令 $k = 0$;

步骤 6 根据等式(16), 计算流量矩阵估计值 $x^{k+1}(t)$;

步骤 7 如果变量 $k > K$ 或者 $\|y(t) - Ax^{k+1}(t)\|_2 < \delta$ 或者 $\|x^{k+1}(t) - x^k(t)\|_2 < \epsilon$, 则获得流量矩阵的最终估计值 $\hat{x}(t)$ 并退出, 否则, 令 $k = k + 1$, 返回到步骤 6.

5 仿真分析

仿真实验使用来自 Abilene^[13] 和 GÉANT^[14] 网络上的真实数据来验证 TMRI 方法. 图 5 表示了用于仿真的

网络拓扑结构, 其中, 圆圈中数字表示路由器编号, 箭头旁边数字表示网络内部链路序号. Abilene 网络有 12 个节点, 144 条 OD 流; 而 GÉANT 网络有 23 个节点, 529 条 OD 流. 文献[3, 4]的 TomoGravity 方法和文献[5]的 $\{1\}$ -Inverse 方法被报道为目前流量矩阵估计的精确方法. 本文将 TMRI 和这两种方法进行比较, 分析三种方法的流量矩阵追踪、空间相对误差(Spatial relative errors: SREs)和时间相对误差(Temporal relative errors: TREs)、以及对噪声的健壮性. Abilene 网络上连续 5 周的数据被用来仿真三种方法, 其中前两周的数据用来训练图 4 所示模型, 其余 3 周数据用于性能分析; GÉANT 网络上连续 6 周的数据被用来仿真三种方法, 其中前三周的数据用于训练图 4 所示模型, 其余三周数据用于性能分析.

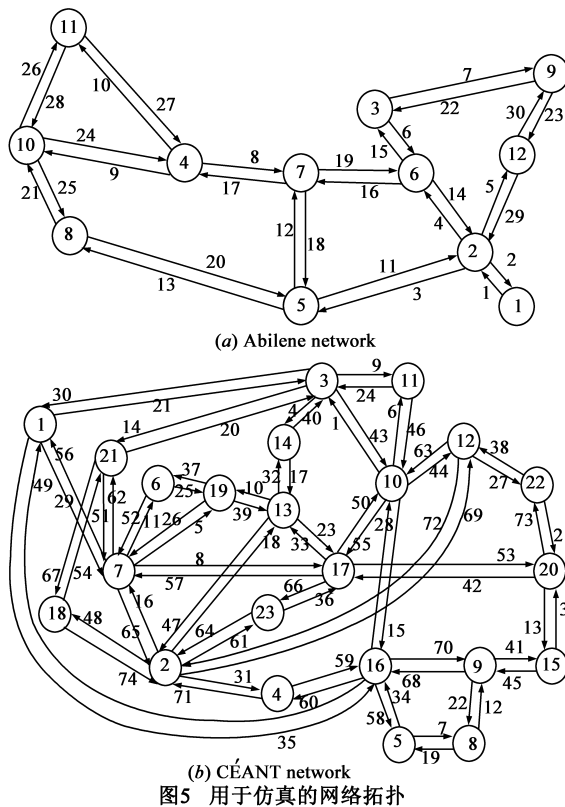


图5 用于仿真的网络拓扑

5.1 流量矩阵追踪

图 6 表示了三种方法在 Abilene 和 GÉANT 网络上的流量矩阵追踪, 其中, 真实值、TMRI、TomoGravity 和 $\{1\}$ -Inverse 方法分别用黑色、灰色、暗黑色和黑色点实线表示. 图 6 表明, 在 Abilene 和 GÉANT 网络上, OD 流都是动态变化的, 而且具有突变和周期性变化特征. 从图 6(a) 可以看到, 在 Abilene 网上, 三种方法都能准确跟踪流量矩阵的动态变化趋势, 但是, TomoGravity 和 $\{1\}$ -Inverse 产生过估计(如 OD 流 35 和 80)和欠估计(如 OD 流 119), 而 TMRI 能更精确地估计 OD 流, 估计值和真实值更接近. 同样地, 图 6(b) 表明, 在 GÉANT 网络上, 三

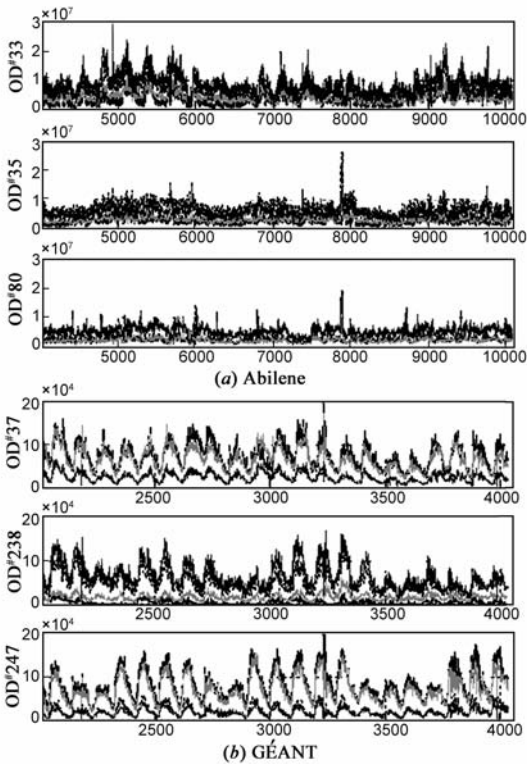


图6 Abilene和GÉANT网络上的OD流估计,真实值用黑色表示, TMRI、TomoGravity和{1}-Inverse方法分别用灰色、暗黑色和黑色实线表示

种方法也能准确跟踪流量的动态变化趋势,特别是具有明显周期性特征的OD流,三种方法的估计结果几乎能跟随OD流的周期性变化而变化.然而,TomoGravity和{1}-Inverse也产生过估计(如OD流238)和欠估计(如OD流37和247),而TMRI能精确地预测OD流的变化趋势,估计值更逼近真实值.更重要的是,图6表明,在Abilene和GÉANT网络上,分别只用两周和三周的数据来训练图4所示估计模型,却能精确预测连续三周的流量矩阵.这表明TMRI不仅能精确估计流量矩阵,而且能对流量矩阵作长期的精确预测.因此,相比之下,TMRI能精确估计流量矩阵,估计结果更接近真实值,而TomoGravity和{1}-Inverse的估计结果彼此接近,但产生较大估计误差;而且,估计模型一旦建立,TMRI就能用于流量矩阵的长期预测.

5.2 估计误差

流量矩阵估计的SREs表示估计误差随OD流变化的空间关系,反映了估计方法在空间上的估计精度;其TREs表示估计误差随时间变化而变化的情况,反映了估计方法在时间上的估计精度.流量矩阵的SREs和TREs是衡量估计方法好坏的一个重要指标,它们分别定义如下:

$$\begin{cases} err_{sp}(n) = \frac{\|\hat{x}_M(n) - x_M(n)\|_2}{\|x_M(n)\|_2}, t = 1, 2, \dots, M \\ err_{tm}(t) = \frac{\|\hat{x}_P(t) - x_P(t)\|_2}{\|x_P(t)\|_2}, n = 1, 2, \dots, P \end{cases} \quad (17)$$

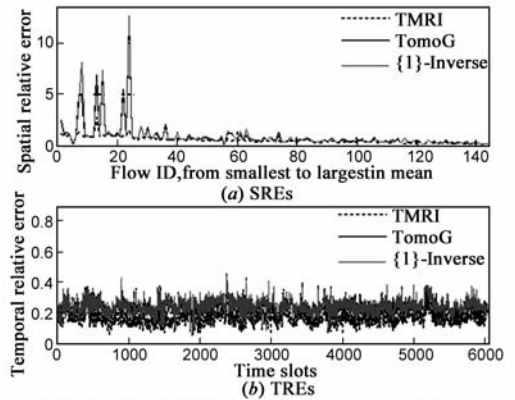


图7 在Abilene网络上, TMRI、TomoGravity和{1}-Inverse的SREs和TREs

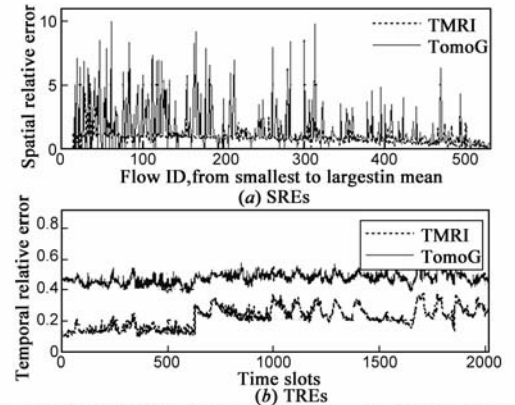


图8 在GÉANT网络上, TMRI和TomoGravity的SREs和TREs

其中, P 、 M 分别为大尺度IP骨干网络中OD流的总数和总的测量时刻; $err_{sp}(n)$ 表示相对于所有测量时刻,第 n 条OD流的SREs; $x_M(n)$ 和 $\hat{x}_M(n)$ 分别表示相对于所有测量时刻,第 n 条OD流的真实值和估计值; $err_{tm}(t)$ 表示时刻 t 时,所有OD流的TREs; $x_P(t)$ 和 $\hat{x}_P(t)$ 分别表示时刻 t 时,所有OD流的真实值和估计值.图7显示了三种方法在Abilene网络上的SREs和TREs;图8、9和10表示了三种方法在GÉANT网络上的SREs和TREs.图7表明,在Abilene网络上,TomoGravity和{1}-Inverse的SREs几乎一致,而TMRI比它们要小得多,特别是当OD流较小时,TMRI的SREs更小.TomoGravity和{1}-Inverse的TREs曲线和几乎完全重叠,即它们的TREs几乎相等,而TMRI的TREs要小得多.同样地,图8、9和10表示,在GÉANT网络上也有类似结论,即TomoGravity和{1}-Inverse的SREs和TREs几乎相等,而TMRI的SREs和TREs却要低得多.而且,在OD流较大和较小时,TomoGravity和{1}-Inverse都具有较大的SREs,而TMRI要小得多.并且TMRI的TREs明显低于TomoGravity和{1}-Inverse的TREs.这表明,与TomoGravity和{1}-Inverse相比,在Abilene和GÉANT网络上,TMRI的时间估计误差和空间估计误差都要小得多,而且TMRI不仅能精确估计较小的OD流,也能精确估计较大的

OD 流. 因此, TMRI 能更精确地进行流量矩阵估计.

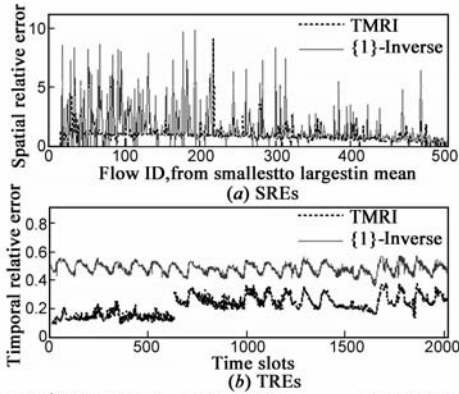


图9 在GÉANT网络上, TMRI和{1}-Inverse的SREs和TREs

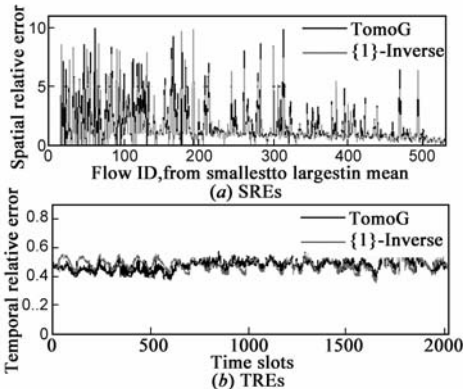


图10 在GÉANT网络上, TomoGravity和{1}-Inverse的SREs和TREs

为更准确评价三种方法的估计精度,下面分析 SREs 和 TREs 的 CDFs (Cumulative Distribution Functions), 即 CSRE (Cumulative distribution functions of Spatial Relative Errors) 和 CTRE (Cumulative distribution functions of Temporal Relative Errors). 图 11 和 12 分别表示 Abilene 和 GÉANT 网络上三种方法的 CSRE 和 CTRE. 图 11 显示在 Abilene 网络中, TomoGravity 和 {1}-Inverse 的 CSRE 和 CTRE 曲线远远低于 TMRI 的相应曲线. 在 SREs 为 0.9 时, TMRI 能精确估计 91.5% 的 OD 流, 而 TomoGravity 和 {1}-Inverse 仅能分别估计 82.0% 和 79.4% 的 OD 流; 而且, 在 TREs 为 0.25 时, TMRI 能精确估计 95% 以上的测量时刻, TomoGravity 和 {1}-Inverse 仅能分别精确估计 66% 和 65% 以下的测量时刻. 同样地, 图 12 显示, 在 GÉANT 网络中, TomoGravity 和 {1}-Inverse 的 CSRE 和 CTRE 曲线远远低于 TMRI 的相应曲线. 在 SREs 为 0.9 时, TMRI 能精确估计 63.0% 的 OD 流, 而 TomoGravity 和 {1}-Inverse 仅能分别精确估计 34.4% 和 42.3% 的 OD 流; 而且, 对于 95% 的测量时刻, TMRI 能以 0.332 以下的 TREs 精确估计, TomoGravity 和 {1}-Inverse 仅能分别以 0.531 和 0.536 以下的 TREs 精确估计. 因此, 从三种方法在两个仿真网络上的 CSRE 和 TSRE 进一步说明,

与 TomoGravity 和 {1}-Inverse 相比, TMRI 的时间估计误差和空间估计误差都要小得多, 它能更精确地估计流量矩阵.

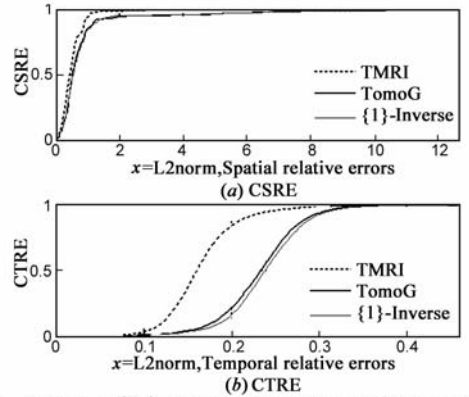


图11 Abilene网络上, TMRI、TomoGravity和{1}-Inverse的CSRE和CTRE

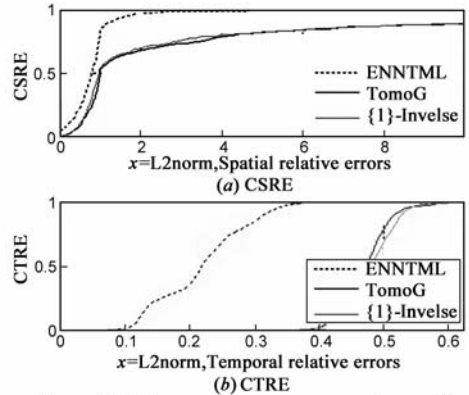


图12 GÉANT网络上, TMRI、TomoGravity和{1}-Inverse的CSRE和CTRE

图 13 表示 TMRI 相对 TomoGravity 和 {1}-Inverse 的平均性能改善. 其中, 在 Abilene 中, TMRI 相对 TomoGravity 和 {1}-Inverse 的平均性能改善分别达到 32.9% 和 34.8%, 而在 GÉANT 中, 分别达到 54.9% 和 53.4%. TMRI 的性能改善是非常明显的. 因此, TMRI 的估计性能要明显优于 TomoGravity 和 {1}-Inverse.

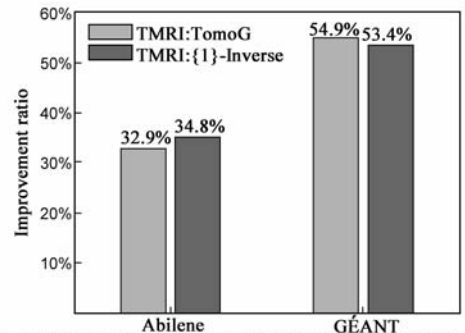


图13 TMRI相对TomoGravity和{1}-Inverse的平均性能改善

5.3 健壮性分析

下面讨论不同噪声干扰对三种方法估计结果的影响, 即分析三种方法的健壮性. 本文通过将噪声引入链

- cations. 2007, E90-B(12): 3673 – 3676.
- [8] Rahman MM, Saha S, Chengan U, Alfa AS. IP traffic matrix estimation methods: Comparisons and improvements[A]. Proc. of the IEEE International Conference on Communications [C]. 2006. 90 – 96.
- [9] Jiang D, Hu G. Large-scale IP traffic matrix estimation based on the recurrent multilayer perceptron network[A]. Proceedings of the IEEE International Conference on Communications [C]. Beijing, China, 19-23 May 2008. 366 – 370.
- [10] Cao J, Davis D, Weil SV, Yu B. Time-varying network tomography: router link data[J]. Journal of American Statistics Association, 2000, 95(452): 1063 – 1075.
- [11] Juva I. Sensitivity of traffic matrix estimation techniques to their underlying assumptions [A]. Proceedings of ICC' 07 [C]. 2007. 562 – 568.
- [12] Park, J, Sandberg JW. Universal approximation using radial basis functions network[J]. Neural Computation, 1991, 3(2): 246 – 257.
- [13] [OL] <http://www.cs.utexas.edu/~yzhang/research/Abilene-TM/>.
- [14] Uhlig S, Quoitin B, Lepropre J, Balon S. Providing public intradomain traffic matrices to the research community[J]. ACM SIGCOMM Computer Communication Review 2006, 36(1): 83 – 86.
- [15] Soule A, Nucci A, Cruz RL, Leonardi E, Taft N. Estimating dynamic traffic matrices by using viable routing changes[J]. IEEE/ACM Transactions on Networking, 2007, 15(3): 485 – 498.
- [16] Bermolen P, Vaton S, Juva I. Search for optimality in traffic matrix estimation: A rational approach by Cramér-Rao lower bounds[A]. Proc. of the 2nd Euro NGI Conf. on Next Generation Internet Design and Engineering [C]. 2006. 224 – 231.
- [17] Hong T, Tong LF, Guo GZ. An assignment model on traffic matrix estimation[A]. Proc. of the International Conference on Natural Computation [C]. 2006. 295 – 304.
- [18] Juva I, Vaton S, Virtamo J. Quick traffic matrix estimation based on link count covariances[A]. Proc of the IEEE International Conference on Communications [C]. Istanbul, 2006. 603 – 608.
- [19] Erramill V, Crovella M, Taft N. An independent-connection model for traffic matrices[A]. Proc. of the ACM SIGCOMM Internet Measurement Conference [C]. 2006. 251 – 256.
- [20] Shioda S, Ohtani K. Estimating the source-destination traffic matrix of a VPN from access-link loads[J]. Computer Communications, 2006, 29(18): 3663 – 3678.
- [21] Cao J, Chen A, Bu T. A quasi-likelihood approach for accurate traffic matrix estimation in a high speed network[A]. Proc. of the IEEE INFOCOM [C]. 2008. 13 – 18.

作者简介:



蒋定德 男, 1974年2月生于四川德阳, 博士, 副教授, 主要研究方向为网络测量、网络安全和认知网络等。

E-mail: jiangddqq@sina.com

王兴伟(通讯作者) (见本期第 756 页)

郭 磊 (见本期第 756 页)

许争争 女, 1980年1月生于湖北武汉, 博士研究生, 主要研究方向为数据建模和最优化。

陈振华 男, 1987年10月生于黑龙江嫩江, 保送硕士研究生, 主要研究方向为网络测量和认知网络。

