

改进 PSO-LSSVM 算法的 SDN 网络流量预测模型^①



龙 霏, 余 铮, 刘 芬, 冯 浩, 代 荡 荡

(国网湖北省电力有限公司 信息通信公司, 武汉 430077)
通讯作者: 龙 霏, E-mail: longfei3@hb.sgcc.com.cn

摘 要: SDN 技术解决了 IP 网络布设困难、更新繁琐等突出问题, 近年来发展迅速. 本文针对 SDN 网络流量预测问题, 提出首先采用混沌理论对时间序列样本群进行相空间重构, 随后引入最小二乘支持向量机 (LSSVM) 构建 SDN 网络流量预测模型, 并结合改进的粒子群算法 (PSO) 对其关键参数进行优化. 实验结果证明, 该模型有效提高了 SDN 网络流量预测精度与误差控制水平, 具有良好的实际应用价值.

关键词: SDN; 网络流量; 预测模型; 相空间重构; LSSVM; PSO

引用格式: 龙霏,余铮,刘芬,冯浩,代荡荡.改进 PSO-LSSVM 算法的 SDN 网络流量预测模型.计算机系统应用,2021,30(7):283-289. <http://www.c-s-a.org.cn/1003-3254/8039.html>

SDN-Based Traffic Prediction Model Based on Improved PSO-LSSVM Algorithm

LONG Fei, YU Zheng, LIU Fen, FENG Hao, DAI Dang-Dang

(Information Communication Company, State Grid Hubei Electric Power Co. Ltd., Wuhan 430077, China)

Abstract: The Software Defined Networking (SDN) technology, which has been booming in recent years, solves the prominent problems of IP networks such as layout difficulty and complex updates. In response to SDN-based traffic prediction, the chaos theory is used to reconstruct the phase space of the time series sample group. Then, the Least Squares Support Vector Machine (LSSVM) is introduced to build the SDN-based traffic prediction model, and the key parameters are optimized by the improved Particle Swarm Optimization (PSO) algorithm. The experimental results show that the model effectively improves the accuracy and error control level of SDN-based traffic prediction and is valuable in practical application.

Key words: SDN; network traffic; prediction model; phase space reconstruction; LSSVM; PSO

1 引言

随着互联网规模的不断扩大, 各种接入设备和相关的数据类型也日益繁杂, 传统 IP 网络的管理方法已经无法满足大量新型应用服务对设备快速部署与技术实时更新的需求, 同时网络管理者在面对网络资源的合理化配置等任务时也遇到了较大的困难, 这些问题都极大地增加了新技术的应用成本. 软件定义网络 SDN (Software Defined Network) 技术改变了传统 IP 网络的

四层协议体系, 将数据传输与网络控制两个主要环节相互解耦^[1], 从而使得用户可以通过控制层提供的统一且便捷的管理工具, 以全局化的视角实现对分布式网络的集中调控, 并依据自身需求灵活的配置各类网络资源, 显著地提高了大规模网络的运维管理水平^[2]. 作为全类型网络管理的关键依据, 网络流量的特征及变化规律同样也成为了 SDN 技术研究的热点, 因此针对 IP 网络流量预测技术进行改进, 设计一款符合 SDN 网

^① 收稿时间: 2020-11-19; 修改时间: 2020-12-21; 采用时间: 2021-01-07; csa 在线出版时间: 2021-06-30

络流量特征的、精确度高的网络流量预测模型,以帮助网络应用开发人员更准确地分析 SDN 网络性能波动状态和优化网络资源配置,具有重要的理论意义和应用价值。

目前针对网络流量进行预测的方法很多,基本可分为两个主要类型,第 1 类是以时间序列法为核心,即根据网络流量在历史时间序列中的变化规律预测下一个有限时间间隔内的变化趋势,如王逸兮等人提出采用同点时间序列构建检测模型,提高了网络性能异常的检出率^[3],Auld 等人则设计了一套网络流量信息的规范化预处理算法,有效地控制了预测模型的计算规模^[4];叶德忠等人提出构建马尔可夫时变模型来完成针对网络流量信息挖掘工作,降低了预测误差^[5]等。这一类模型对于传统 IP 网络流量的监测效果较好,但对于具备长相关性特征的 SDN 网络数据流的预测精度则不太理想^[6]。第 2 类方法是以神经网络等智能算法为核心构建网络流量预测模型,如李晓方等人提出对源-目的流(OD 流)数据进行降维,并以 Elman 神经网络对低维度样本进行训练,从而发挥了 SDN 网络可实现精确测量数据流的优势^[7];NAREJOS 等人提出通过构建深层神经网络的方式提高对样本群的训练质量,提高了预测精度,但算法的空间复杂度也大幅上升^[8]。神经网络存在一些固有的缺陷,如对训练样本要求过高,对待有限的样本群时,缺乏足够的泛化能力;网络结构无法动态调整,训练时间过长;神经元规模可控性差,极易陷入“维数灾”难题等^[9]。

支持向量机(SVM)作为一种基于统计学习理论的人工智能算法,具有良好的泛化能力与非线性处理能力^[10],近年来在样本群分型与模式识别领域得到广泛的应用。如王鸣提出采用小波技术结合 SVM 构建流量预测模型,良好地控制了计算规模,提升了预测模型的执行效率^[11]。但根据目前的研究成果来看,与传统 IP 网络相比,SDN 网络流量的特征呈现出更为突出的混沌性、时变性和自相似性,这使得 SVM 技术必须在非线性混沌理论的辅助下,方可通过数学建模来描述复杂的网络流量变化规律。因此本文提出采用混沌理论中的多维相空间重构技术对 SDN 网络流量数据进行预处理,构建相空间中的样本矩阵,随后采用改进后的粒子群算法(PSO)对最小二乘支持向量机(LSSVM)中的关键参数进行优化,有效的提高了 LSSVM 的运算效率,并构建针对网络流量变化趋势的预测模型,对 SDN 网络流量样本群进行训练。通过仿真实验证明,该

模型与其他同类模型相比,在合理控制计算复杂度的条件下,表现出了更高的预测精度和误差控制能力。

2 基于混沌理论的相空间重构

相空间重构理论由 Takens 等人提出,该理论将多维状态空间中任一点的变化规律视为与之产生关联的其它点变化的综合影响,因此可构建相对时间序列。该序列中的任一点在起始时间点上的值设为基值,而在空间中其他延迟坐标点上的值则作为新维来处理。随着起始时间点的推移,不断构建出新的子序列,就可通过监测空间中各个时间点变化规律的方式,完整的构建出系统的动态模型^[12]。

假设混沌时间序列为 $x = \{x_i | x_i, i = 1, 2, 3, \dots, N\}$, N 为时间序列长度,采用时延 τ 和嵌入维 m 将其重构,可得到:

$$\begin{cases} X = \{X_i | X_i = [x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}]^T, i = 1, 2, \dots, M\} \\ M = N - (m-1)\tau \end{cases} \quad (1)$$

设系统动力学维数为 d , Takens 同样证明了若嵌入维 m 能够满足 $m \geq 2d + 1$, 则该模型与原系统可视为在拓扑意义上等价。虽然 Takens 在其证明过程中给出了嵌入维的计算方法,但其前提是时间序列无限长,且样本空间处于无干扰的理想状态,这显然与网络流量的特征不符^[13],因此本文提出采用 C-C 方法对 m 和 τ 这两个关键参数进行选取,按照以上设定,可将嵌入序列的关联积分通过式(2)表达:

$$C(m, N, r, \tau) = \frac{2}{M(M-1)} \sum_{1 \leq i < j \leq M} \theta(r - d_{ij}), r > 0 \quad (2)$$

式中, $d_{ij} = \|X_i - X_j\|_{\infty}$, 为相空间内任意两点间距,采用矢量差的无穷范数表示;且有 $x < 0$ 时, $\theta(x) = 0$, $x \geq 0$ 时, $\theta(x) = 1$ 。按重构时延将 $x = \{x_i | x_i, i = 1, 2, 3, \dots, N\}$ 等分为 t 个子序列,即得到如下分量:

$$\begin{cases} x^1 = \{x_i | i = 1, \tau + 1, \dots, N - \tau + 1\} \\ x^2 = \{x_i | i = 2, \tau + 2, \dots, N - \tau + 2\} \\ \vdots \\ x^t = \{x_i | i = \tau, 2\tau, \dots, N\} \end{cases} \quad (3)$$

由此可将检验统计量表达为:

$$S_1(m, N, r, \tau) = C(m, N, r, \tau) - C^m(1, N, r, \tau) \quad (4)$$

对上式进行分块平均,可得到:

$$S_2(m, N, r, \tau) = \frac{1}{t} \sum_{s=1}^t [C_s(m, N/\tau, r, \tau) - C_s^m(1, N/\tau, r, \tau)] \quad (5)$$

当 $N \rightarrow \infty$ 时, 可将上式简化为:

$$S_2(m, r, \tau) = \frac{1}{t} \sum_{s=1}^t [C_s(m, r, \tau) - C_s^m(1, r, \tau)] \quad (6)$$

当处于时间序列无限长且各元素之间不相关的理想条件下时, 若 m 和 τ 为定值, 则对于 $\forall r$, 均有 $S_2(m, r, \tau) = 0$ 成立. 但对于网络流量的样本空间而言, 其时间序列长度必然是有限的, 且元素之间存在复杂的关联, 因此 $S_2(m, r, \tau) \sim \tau$ 明显呈现出了时延的自相关性, 按照混沌理论求解相空间中最佳时延 τ_d 的方法, 可知:

$$\begin{aligned} \tau_d &= \min(\Delta S_2(m, \tau)) \\ &= \min(\max(S_2(m, r_j, \tau)) - \min(S_2(m, r_j, \tau))) \end{aligned} \quad (7)$$

确定了最佳重构时延 τ 后, 即可逆向推导得到嵌入维 m 的最佳值, 并完成时间序列在多维相空间中的重构.

3 基于改进 PSO 算法的 LSSVM 的优化

3.1 最小二乘支持向量机 LSSVM

支持向量机 (SVM) 具有良好的泛化能力, 但对于一个新的样本群进行训练时, 其计算复杂度高成为了一个很难克服的问题, 这就导致了当面对以时间序列构建的大规模 SDN 网络流量样本群时, SVM 的执行效率会大幅降低, 从而严重地影响了预测模型的实时性能^[14]. Este A 等人提出采用最小二乘法对 SVM 进行改进, 引入误差平方和替代 SVM 中的损失函数, 通过非线性映射函数 $\varphi(x)$ 将低维空间中的平面分类问题转化为高维空间的线性方程组求解问题, 从而有效地降低了原算法的复杂度^[15].

对于一个记录了输入量 x_i 和输出量 y_i 的网络流量样本群 $D = \{(x_i, y_i) | i = 1, 2, \dots, n\}$, 使用 $\varphi(x)$ 构建最小二乘支持向量机 (LSSVM) 的回归函数:

$$y(x) = w^T \varphi(x) + b \quad (8)$$

式中, w 表示高维特征空间采用的权值向量, b 为预设的偏置常数.

参照支持向量机理论, 可采用式 (9) 中列出的回归模型求解上述回归函数, 有:

$$\begin{cases} \min J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^n e_i^2 \\ \text{s.t. } y_i = w^T \phi(x_i) + b + e_i, i = 1, 2, \dots, n \end{cases} \quad (9)$$

该模型使用的正则化方法非常有助于提高模型的

泛化能力, γ 为正则化参数, e_i 为输出量与回归函数结果间的误差. 再引入 Lagrange 函数将式 (9) 转化更易求解的对偶空间优化模型:

$$L(w, b, e, \alpha) = J(w, e) - \sum_{i=1}^n \alpha_i (w^T \varphi(x) - b + e_i - y_i) \quad (10)$$

式中, $\alpha_i \in R$ 为 Lagrange 乘子. 经过多类函数的对比分析后, 选取 RBF 函数为 LSSVM 的核函数, 设其函数宽度为 σ , 可将其表达为:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (11)$$

对式 (10) 中的 $L(w, b, \eta, \alpha)$ 求偏导, 消除其它待求变量, 并引入式 (11) 核函数, 可最终将其简化为针对 α_i 和 b 的求解函数:

$$\hat{y}(x) = \sum_{i=1}^N \alpha_i K(x_i, x_j) + b \quad (12)$$

3.2 基于改进 PSO 算法的参数寻优

LSSVM 模型的拟合能力主要依赖正则化参数 γ 和核函数宽度值 σ 的选取质量, 其中 γ 影响了模型的拟合精度与泛化能力, σ 则直接决定了模型的计算量与执行效率^[16,17]. 目前通常采用的选取方法有网络搜索算法和遗传算法等, 前者的计算规模庞大, 算法实时性差, 后者则容易陷入局部极值陷阱问题, 本文提出采用改进粒子群算法实现对 LSSVM 模型参数 (γ, σ) 的组合优化, 取得了较好的参数寻优质量.

粒子群算法 (PSO) 针对群体内的每个粒子的位置与速度进行跟踪和调整, 从而实现对整个群体的优化效果^[18]. 假设采用 d 维向量对以上两种状态分别进行描述, 则有 $z_i = (z_{i1}, z_{i2}, \dots, z_{id})$, $v_i = (v_{i1}, v_{i2}, \dots, v_{id})$, 而经过调整后得到该粒子最佳位置记为 $p_i = (p_{i1}, p_{i2}, \dots, p_{id})$, 最终整个群体的最佳位置记为 $p_s = (p_{s1}, p_{s2}, \dots, p_{sd})$. 在 PSO 算法的 k 轮迭代过程中, 对群体中的每个粒子均采用式 (13) 进行状态参数的调整.

$$\begin{cases} v_{id}^{k+1} = \omega v_{id} + c_1 r_1 (p_{id} - z_{id}^k) + c_2 r_2 (p_{sd} - z_{id}^k) \\ z_{id}^{k+1} = z_{id}^k + v_{id}^{k+1} \end{cases} \quad (13)$$

式中, ω 为惯性权重, c_1 和 c_2 为 PSO 算法采用的学习因子, 用以调节个体粒子的变化量对群体状态造成的影响, r_1 和 r_2 是为了避免算法陷入局部最优而设定的随机数. 传统方法设定 ω 、 c_1 和 c_2 多是根据历史数据验算求

得,之后定期调整更新,这导致寻得的参数往往滞后于网络流量的变化现状,因此本文提出采用自适应方法,在算法执行过程中,动态的调整粒子搜索性能,从而提高算法的拟合度,首先 ω 的计算方法为:

$$\omega = \begin{cases} \omega_{\min} - \frac{(\omega_{\max} - \omega_{\min})(f - f_{\min})}{f_{\text{avg}} - f_{\min}}, & f < f_{\text{avg}} \\ \omega_{\max}, & \text{otherwise} \end{cases} \quad (14)$$

式中, f 为当前粒子个体的适应度, f_{\min} 和 f_{avg} 分别为整个群体中适应度的最小值和平均值。 c_1 和 c_2 的最佳取值则与迭代次数 k 密切相关,并随着 k 的增加而分别呈现出递增和递减趋势,因此可采用针对 k 的单调函数来完成每一轮次 c_1 和 c_2 的动态调整。

$$\begin{cases} c_1 = -2 \times \frac{k}{K_{\max}} + 2.5 \\ c_2 = 2 \times \frac{k}{K_{\max}} + 0.5 \end{cases} \quad (15)$$

式中, K_{\max} 为设定的最大迭代次数。在改进 PSO 算法执行的每个轮次,均自适应的采用以上方法确定 ω 、 c_1 和 c_2 ,从而可完成对 LSSVM 参数组合 (γ, σ) 的寻优工作。

对参数寻优效果进行分析,首先,根据式(13)可以看出, c_1 和 c_2 两个学习因子属于加速参数类型,决定了个体粒子分别向个体最佳位置 p_i 与群体最佳位置 p_s 的移动速度,显然,过高的 c_1 会使得粒子个体快速掠过 p_i 所在区域,继而导致粒子适应度的大幅波动,并最终使寻得的参数偏离最优解^[19];当 c_1 不断减小时,算法漏掉最佳位置的概率也随之降低,但与此同时,算法的收敛过程也会相应的延长,导致优化效率的下降。 c_2 则相反,由于群体最佳位置 p_s 的相对固定特征,使得较大的 c_2 会导致算法在确定粒子移动方向时耗费过多资源,而较小的 c_2 又会降低粒子抵达 p_s 区域的速度,增加算法收敛时间。改进后的算法在引入了动态调节机制之后, c_1 随着迭代次数的增加而减小,确保粒子个体能够捕捉到个体最佳位置,而 c_2 在算法执行的初期取较小值,在快速确定了 p_s 的方向后, c_2 随迭代次数不断增大,有效提高算法的收敛性能。

惯性权重参数 ω 的作用是帮助粒子个体构建全局搜索能力,对 ω 进行合理的调节,将有效的影响算法在全局搜索和局部搜索两方面侧重倾向^[20],若式(13)中缺少了第一项 ωv_{id} ,则算法将只能搜索有限范围,其结果往往会出现局部最优问题;若式(13)中去除与 c_1 和 c_2 相关的后两项,则此时粒子以初速度 v_0 开始移动。对

ω 的作用进行独立分析,可分为以下3种情况:

- 1) $\omega = 1$, 此时粒子将保持初速度,以原方向做匀速移动;
- 2) $0 < \omega < 1$, 此时粒子速度会从初速度开始,随着 ω 的减小而迅速降低至0;
- 3) $\omega > 1$, 则粒子的速度不断增大,直至达到最大速度 v_{\max} 。

以上3种情况均会对算法造成极大的干扰,粒子个体将会出现加速度恒定和移动方向固定等问题,因此当失去了加速参数 c_1 和 c_2 带来的调节效果后,单纯依赖惯性权重 ω 是无法实现对参数的可靠寻优的。若单独保留 c_1 项,则算法过程等价于多个粒子个体的独立搜索,参数寻优性能将大幅度下降;若单独保留 c_2 项,则算法仅具备群体搜索模式,虽然可开辟新的解空间,但对复杂度稍高的问题,算法很容易收敛于局部极值,而造成较大的偏差。针对以上分析,本文提出在 ω 的设定范围内,采用自适应算法实现对 ω 的动态调节目标,既在算法执行的初期有效地促进了全局搜索能力,提高了收敛速度,又在后期保障了局部搜索性能,改善了算法收敛精度,对于 LSSVM 关键参数 γ, σ 的寻优工作起到了显著的改进效果。

4 SDN 网络流量预测模型的构建与仿真

4.1 预测模型的构建与算法流程设计

在引入了改进 PSO 算法之后,采用实数编码将 (γ, σ) 组成一个粒子,随后结合式(9)至式(12)的推导过程,构建出优化后的预测模型,如式(16)所示:

$$\begin{cases} \min J = \min f(z_i) = \frac{1}{L} \sum_{i=1}^n (y_i - \hat{y}_i) \\ \text{s.t. } \gamma \in [\gamma_{\min}, \gamma_{\max}], \sigma \in [\sigma_{\min}, \sigma_{\max}] \end{cases} \quad (16)$$

适应度函数取预测值与真实值之间的误差总和,即:

$$f(\cdot) = \sum_{i=1}^n |y_i - \hat{y}_i| \quad (17)$$

其中, y_i 为当前样本预测值,而 \hat{y}_i 为对应样本的真实值。算法流程如图1所示。

对算法各主要环节描述如下:

Step 1. 采集 SDN 网络流量历史数据,构建时间序列的样本群,并采用 C-C 算法获取最佳的嵌入维和重构时延,在此基础上完成对原样本群的相空间重构,得到规范化的数据序列;

Step 2. 对 PSO 算法执行初始化操作, 确定样本群规模、迭代轮次数、粒子初始状态 (位置 z_i 、速度 v_i)、个体最佳位置 p_i 、群体最佳位置 p_g 以及 ω 、 c_1 和 c_2 等关键参数的初值, 完成参数寻优的准备工作;

Step 3. 将 (γ, σ) 参数组成粒子个体, 并采用随机取值法构建初代粒子群体;

Step 4. 根据式(17)计算群体中粒子个体的适应度, 调整本轮次中 ω 、 c_1 和 c_2 的取值, 根据式(13)更新个体粒子的 z_i 与 v_i , 产生本轮次的粒子群体, 并计算出新群体的 p_i 和 p_g ;

Step 5. 依据适应度值判断参数寻优目标是否完成, 若达要求则将本轮次寻得的最优粒子取出, 按照实数编码规则解析出对应的 (γ, σ) 作为 LSSVM 的优化参数, 否则返回 Step 4, 进入下一轮次的迭代训练;

Step 6. 利用求得的参数和重构的训练样本对 LSSVM 进行求解, 构建网络流量预测模型, 并使用验证样本群进行对照分析, 输出最终的预测结果.

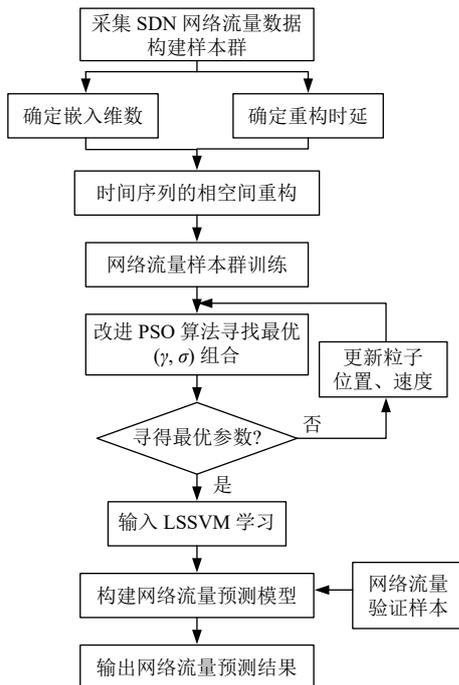


图1 预测算法执行流程

4.2 实验设计与结果分析

在数据样本的选择上, 采用在美国 Abilene 网络上采集到的真实数据展开仿真实验分析. 该数据集汇聚了一个包括 30 条链路在内的 SDN-IP 混合网络中的流量信息, 其中 SDN-OD 流 46 条, IP-OD 流为 108 条, 数

据采样周期为 5 min, 选择其中 1000 条流量数据, 前 800 条数据作为训练样本群, 后 200 条作为验证样本群, 其中部分样本信息如图 2 所示.

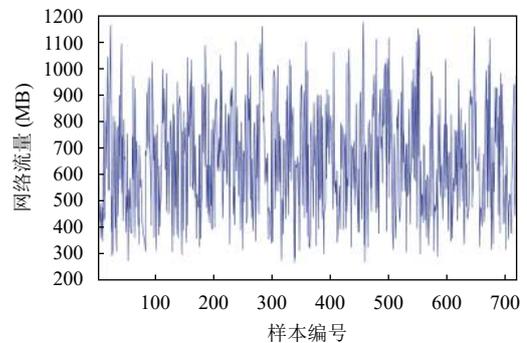


图2 实验流量数据样本

本次仿真实验共分为两组, 分别为 1 步预测仿真和 3 步预测仿真. 首先对样本群进行相空间重构, 采用 C-C 法对流量样本进行处理, 求得最优的重构时延 $\tau=2$, 嵌入维 $m=10$, 完成重构后将前 800 条样本输入 LSSVM 训练, 并采用改进 PSO 算法进行参数寻优. 得到最优参数为: $\gamma=102.83$, $\sigma=1.35$, 最终构建针对 SDN 网络流量的预测模型. 在经过前 800 条样本的训练后, 对后 200 条验证样本进行预测对比, 其中 1 步预测实验结果如图 3 所示.

观察图 3(a) 中预测结果跟踪曲线可以发现, 在 1 步预测仿真实验中, 基于改进 PSO-LSSVM 的预测模型表现出了良好的预测精度, 各预测值与真实值拟合程度高, 预测结果跟踪曲线与真实流量变化曲线基本重合, 证明该模型很好地模拟了 SDN 网络流量所具备的周期性和非线性变化特征, 具有可靠的泛化性能; 从图 3(b) 中可以看出, 预测结果的偏差十分稳定, 且均在狭小范围内波动. 在 200 次的预测结果与真实数据的比对中, 167 次预测结果的相对误差小于 6 MB, 其中更有 94 次相对误差小于 2 MB, 而误差超过 8 MB 仅有 9 次, 且所有预测误差均在 10 MB 以内, 预测结果表现出了良好的精确性和稳定性.

网络流量预测模型需要真实反映网络流量变化规律, 其预测精度越高、提前量越大, 模型应用价值就越高. 1 步预测仅能对时间序列中下一个临近点的流量进行预测, 并不能客观的验证模型的预测性能^[21], 为了进一步考察本模型在网络流量多步预测方面的表现, 设计了 3 步预测仿真实验, 其结果如图 4 所示.

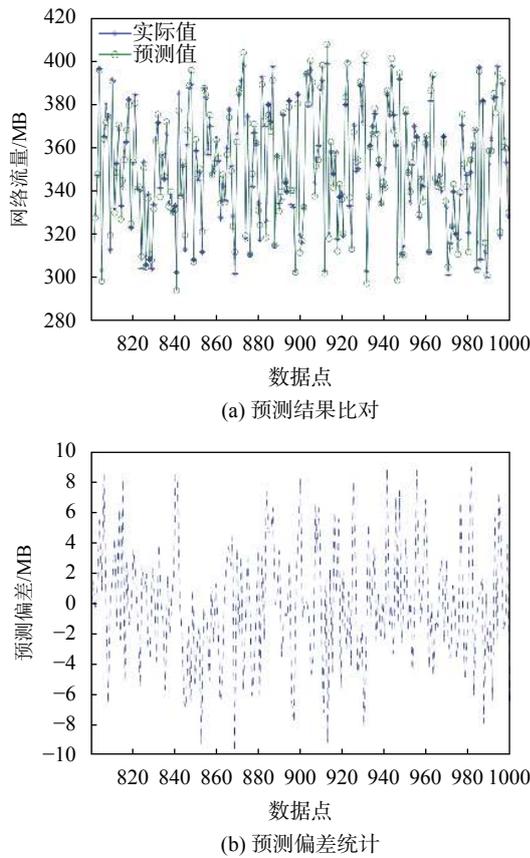


图3 SDN网络流量1步预测实验结果

观察图4(a)可以发现,在增加了预测提前量之后,模型对SDN网络流量变化的拟合性能有了一定程度的下降,且图4(b)中的预测偏差也有所增大,但预测质量整体上仍处于良好水平,其中154次预测相对误差在20MB以内,仅有4次超过40MB,其中1次达到了最高的50.13MB.从整体来看,误差波动范围稳定,基本控制在10%以内,预测曲线与实际流量曲线的波动方向吻合,能够在实际应用过程中对网络流量监控工作提供可靠的参考.

为了对比分析预测模型的运行质量,同时选取了BP神经网络(BPNN)、自回归滑动平均(ARIMA)和基于遗传算法的最小二乘支持向量机(GA-LSSVM)三种预测模型参与实验,并选取相对平均误差(MPAE)和均方根误差(RMSE)两种指标对预测结果进行评价,其计算公式分别为:

$$MPAE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (18)$$

$$RMSE = \sqrt{\frac{\sum_{k=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (19)$$

使用以上3种网络流量预测模型对同组样本进行预测,并将统计结果与本文提出的改进PSO-LSSVM模型的预测结果进行对比分析,其统计结果如表1所示.

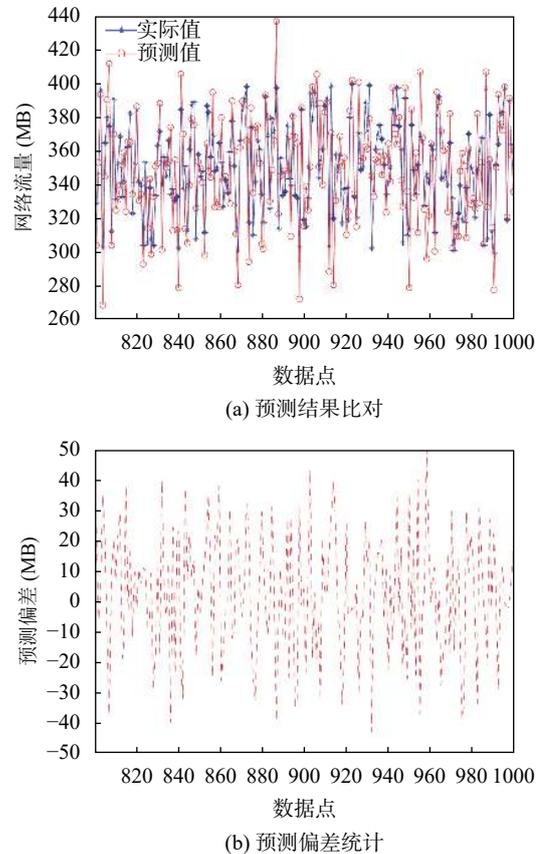


图4 SDN网络流量3步预测实验结果

表1 多模型预测结果统计信息

预测模型	1步预测统计		3步预测统计	
	MPAE(%)	RMSE	MPAE(%)	RMSE
BPNN	4.06	18.45	8.63	41.79
ARIME	5.91	26.34	11.24	53.66
GA-LSSVM	3.42	17.43	7.94	34.61
PSO-LSSVM	0.98	4.14	5.04	21.03

对比4种模型对SDN网络流量的预测统计数据可以看出,相较于BPNN模型和ARIME模型,改进后的PSO-LSSVM模型在1步与3步预测仿真实验中表现出明显优势,其预测结果的拟合精度和误差控制水平平均大幅领先,这也证明了基于最小二乘向量机LSSVM技术构建的预测模型具有更好的泛化性能,尤其适合针对大规模样本群体的分析与处理场合;而与采用了同类型技术的GA-LSSVM模型相比,本模型1步预测的质量依旧明显领先于前者,这说明改进后的PSO算

法对惯性权重和学习因子实施自适应动态调节后,显著地提高了粒子的全局搜索能力,更有效的避免了局部极值问题,从而获得了比遗传算法更加突出的参数寻优效果.在3步预测仿真实验中,本模型表现出的优势虽有所减少,但仍要优于GA-LSSVM模型,这表明随着预测步数(即预测提前量)的增加,SDN网络流量的非线性和时变特征对模型构成的干扰效应会持续放大,从而导致所有模型的预测质量均会下降,彼此之间的表现差异也会相对缩小,这一规律在其余两种模型的预测结果中也得到了充分的体现.

5 结束语

本文针对SDN网络流量预测问题展开研究,首先提出对SDN网络流量时间序列进行混沌处理,实现在多维相空间中的重构,减少了网络流量的复杂性特征对预测模型造成的负面影响;随后采用自适应方法对PSO算法中的惯性权重及学习因子进行动态调整,显著改善了该算法的参数寻优能力,并将其引入到LSSVM中,最终构建出SDN网络流量预测模型.通过仿真实验证明,该模型与其他3种经典预测模型相比,在预测精度和误差控制等方面具有明显的优势,应用前景十分广泛.

参考文献

- 1 Liu J, Li J, Shou G C, *et al.* SDN based load balancing mechanism for elephant flow in data center networks. Proceedings of 2014 International Symposium on Wireless Personal Multimedia Communications. Sydney, NSW, Australia. 2015. 486–490.
- 2 朱格,李明政,刘童,等.利用SDN实现可编程无线局域网多径接入管理.计算机系统应用,2019,28(9):125–132. [doi: 10.15888/j.cnki.csa.007075]
- 3 王逸兮,余铮,查志勇,等.基于改进SVM的电力企业信息系统异常检测方案的优化.计算机与数字工程,2020,48(3):567–570,662. [doi: 10.3969/j.issn.1672-9722.2020.03.013]
- 4 Auld T, Moore AW, Gull SF. Bayesian neural networks for Internet traffic classification. IEEE Transactions on Neural Network, 2007, 18(1): 223–239. [doi: 10.1109/TNN.2006.883010]
- 5 叶德忠,巫忠正,蒋勇.基于马尔可夫时变模型的流量数据挖掘.软件,2017,38(9):8–11. [doi: 10.3969/j.issn.1003-6970.2017.09.002]
- 6 Lu Y, Fu BC, Xi XF, *et al.* An SDN-based flow control mechanism for guaranteeing QoS and maximizing throughput. Wireless Personal Communications, 2017, 97(1): 417–442. [doi: 10.1007/s11277-017-4512-9]
- 7 李晓方,王子磊,奚宏生.混合SDN的自适应流量估计方法.计算机工程,2016,42(3):103–110. [doi: 10.3969/j.issn.1000-3428.2016.03.019]
- 8 于静,王辉.基于组合模型的网络流量预测.计算机工程与应用,2013,49(8):92–95. [doi: 10.3778/j.issn.1002-8331.1109-0531]
- 9 李亦滔.基于支持向量机的改进分类算法.计算机系统应用,2019,28(10):145–151. [doi: 10.15888/j.cnki.csa.007080]
- 10 房汉鸣,税爱社,汪辉,等.支持向量机动态多分类方法.后勤工程学院学报,2017,33(2):90–96. [doi: 10.3969/j.issn.1672-7843.2017.02.015]
- 11 王鸣,孙奕鸣.小波支持向量机的网络流量预测研究.计算机仿真,2012,29(11):198–201. [doi: 10.3969/j.issn.1006-9348.2012.11.046]
- 12 周向军.基于最小二乘支持向量机的无线网络信道检测.计算机系统应用,2018,27(5):151–155. [doi: 10.15888/j.cnki.csa.006339]
- 13 Nguyen TTT, Armitage G. A survey of techniques for Internet traffic classification using machine learning. IEEE Communications Surveys & Tutorials, 2008, 10(4): 56–76.
- 14 阳凯,林海涛,黎海雪.基于SDN的流量控制算法综述.通信技术,2019,52(4):773–781. [doi: 10.3969/j.issn.1002-0802.2019.04.001]
- 15 Este A, Gringoli F, Salgarelli L. Support vector machines for TCP traffic classification. Computer Networks, 2009, 53(14): 2476–2490. [doi: 10.1016/j.comnet.2009.05.003]
- 16 Sequeira L, de la Cruz JL, Ruiz-Mas J, *et al.* Building an SDN enterprise WLAN based on virtual APs. IEEE Communications Letters, 2017, 21(2): 374–377. [doi: 10.1109/LCOMM.2016.2623602]
- 17 黄国权,尤新华.改进粒子群算法优化最小二乘支持向量机的网络流量混沌预测.激光杂志,2015,36(3):96–99.
- 18 陈鸿星.基于ELM-LSSVM的网络流量预测.计算机工程与应用,2015,51(24):73–77. [doi: 10.3778/j.issn.1002-8331.1312-0119]
- 19 李昆仑,张亚欣,刘利利,等.基于改进PCA和支持向量机的掌纹识别.计算机科学,2015,42(S2):146–150.
- 20 Wang XD, Tian Y, Zhao M, *et al.* PNPL: Simplifying programming for protocol-oblivious SDN networks. Computer Networks, 2018, 147: 64–80. [doi: 10.1016/j.comnet.2018.09.018]
- 21 王雪松.改进支持向量机的网络流量预测.计算机系统应用,2017,26(3):230–233. [doi: 10.15888/j.cnki.csa.005668]