

基于优化组合核极限学习机的网络流量预测

刘悦,王芳

(开封大学信息工程学院,河南开封 475004)

摘要:为了提高网络流量预测的精度,针对网络流量数据具有非线性、非平稳的特点,提出一种基于经验模态分解(EMD)和混沌粒子群算法优化组合核极限学习机的网络流量预测模型。首先将网络流量时间序列进行EMD分解,提取网络流量数据的各个分量,然后分别对各个分量采用核极限学习机进行预测,最后重构出预测结果。针对传统核极限学习机拟合能力的不足,提出一种基于高斯核和多项式核组合的组合核极限学习机,并且采用改进的混沌粒子群算法优化组合核的核参数组合权值以及惩罚因子,并将其应用到网络流量预测中。实验结果表明,该方法可以有效提高网络流量预测的精度,有助于指导网络资源的合理分配和规划。

关键词:网络流量预测;核极限学习机;组合核函数;混沌粒子群;经验模态分解

中图分类号:TP391.9

文献标识码:A

文章编号:1673-629X(2016)06-0073-05

doi:10.3969/j.issn.1673-629X.2016.06.016

Network Flow Prediction Based on Optimization Combined Kernel Extreme Learning Machine

LIU Yue, WANG Fang

(College of Information Engineering, Kaifeng University, Kaifeng 475004, China)

Abstract: In order to improve precision of network flow prediction, a prediction model is proposed in this paper based on Empirical Mode Decomposition (EMD) and chaos particle swarm optimization combined kernel extreme learning machine aiming at the features of non-linear and non-stationary for network flow data. Unit flow is obtained through EMD on the network flow in time sequence, then each unit data is predicted with kernel extreme learning machine. Finally, the prediction result is reconstructed. In view of the inadequate fitting capacity of traditional kernel extreme learning machine, a machine combining Gaussian kernel and multinomial kernel is proposed and the improved kernel parameter combination and penalty factor of chaos particle swarm optimization with combined kernel are applied in the prediction of network flow. The experiment shows that this method can improve the accuracy of network prediction effectively, and help guide the rational allocation and planning of network resources.

Key words: network flow prediction; kernel extreme learning machine; combined kernel function; chaos particle swarm; empirical mode decomposition

0 引言

随着互联网技术的飞速发展,流量和带宽不断增加,网络规模和网络的复杂程度日益增大,导致网络管理工作的繁重程度急剧上升,网络故障频现。为了保证互联网数据的可靠传输和资源的合理分配,高质量的网络流量预测对网络的规划、管理和设计具有重要的理论意义和实际价值^[1-2]。

核极限学习机算法是新加坡学者黄广斌提出的一种新的前馈神经网络学习算法^[3]。它计算速度快,泛化性好,得到了广泛应用^[4-6]。文中针对网络流量数

据的非平稳和非线性的特点,提出了结合经验模态分解(EMD)和核极限学习机的网络流量预测算法。首先采用EMD对原始时间序列进行分解,然后将分解出来的各个分量分别采用核极限学习机进行预测,最后重构出预测结果。针对核极限学习机逼近能力的不足,提出一种基于径向基核函数和多项式核函数组合的多核极限学习机方法,并且采用改进的混沌粒子群算法去优化多核极限学习机的参数。该方法为网络资源的合理配置和可靠传输提供决策的依据。

收稿日期:2015-10-01

修回日期:2016-01-05

网络出版时间:2016-05-05

基金项目:河南省教育科学技术研究重点项目(15C520016);开封市科技攻关计划项目(130145)

作者简介:刘悦(1977-),男,硕士,讲师,研究方向为机器学习、网络安全。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160505.0831.112.html>

1 经验模态分解

EMD 是一种自适应信号分解法,其吸取小波变换多分辨的优点的同时,克服了小波基选择和分解尺度难以确定的缺点,因此非常适合非线性非平稳信号的分解^[7-8]。网络流量时间序列是一个典型的非平稳时间序列^[9],文中采用先 EMD 分解,再预测、重构的方法。

2 CPSO 组合核极限学习机

2.1 核极限学习机

Huang 等研究前馈神经网络无需像 BP 那样迭代求解,而是可以通过随机设置前馈神经网络的输入权值,通过求解输出权值的最小二乘解来完成训练^[10]。极限学习机因其极快的训练速度和良好的泛化性,已经受到很多学者的关注。在求解极限学习机的过程中,既要考虑使训练误差达到最小,还要考虑到结构风险最小化。因此需要在最小化的输出权值和最小化的误差之间做出折中,可以构造如下的计算公式:

$$\min: \|\boldsymbol{\beta}\|^2, \sum_{i=1}^N \|\boldsymbol{\beta} \cdot h(x_i) - t_i\| \quad (1)$$

也可以表达如下:

$$\min: L = \frac{1}{2} \|\boldsymbol{\beta}\|^2 + \frac{1}{2} C \sum_{i=1}^N \|\xi_i\|^2 \quad (2)$$

$$\text{s. t. } : h(x_i)\boldsymbol{\beta} = \mathbf{t}_i^T - \boldsymbol{\xi}_i^T, i = 1, 2, \dots, N$$

其中, $\boldsymbol{\xi}_i = [\xi_{i,1}, \xi_{i,2}, \dots, \xi_{i,m}]$ 是训练样本 x_i 对应的网络输出和真实值之间的误差。

根据 KKT 条件,可以定义 Lagrange 函数求解上面的问题,也就是说上面的问题可以等效为求解下面的公式:

$$\begin{aligned} \min: L_{\text{DELM}} = & \frac{1}{2} \|\boldsymbol{\beta}\|^2 + \frac{1}{2} C \sum_{i=1}^N \|\xi_i\|^2 - \\ & \sum_{i=1}^N \sum_{j=1}^m \alpha_{i,j} (h(x_i)\beta_j - t_{i,j} + \xi_{i,j}) \end{aligned} \quad (3)$$

其中: α_i 为 Lagrange 乘数,且是非负数; β_j 是连接隐含层和第 j 个输出节点的权值, $\boldsymbol{\beta} = [\beta_1, \dots, \beta_m]$ 。

相应的优化限制条件为:

$$\begin{aligned} \frac{\partial L_{\text{DELM}}}{\partial \beta_j} = 0 \rightarrow \boldsymbol{\beta}_j &= \sum_{i=1}^N \alpha_i h(x_i)^T \rightarrow \boldsymbol{\beta} = \mathbf{H}^T \boldsymbol{\alpha} \\ \frac{\partial L_{\text{DELM}}}{\partial \xi_i} = 0 \rightarrow \boldsymbol{\alpha}_i &= C \boldsymbol{\xi}_i, i = 1, 2, \dots, N \\ \frac{\partial L_{\text{DELM}}}{\partial \alpha_i} = 0 \rightarrow h(x_i)\boldsymbol{\beta} - \mathbf{t}_i^T + \boldsymbol{\xi}_i^T &= 0, i = 1, 2, \dots, N \end{aligned} \quad (4)$$

式中, \mathbf{H} 是隐含层输出矩阵的权值,随机给定以后它就是一个固定的矩阵,它的数值仅仅与输入样本

的个数和设置的隐层节点数有关。

把上述公式进行组合,可以得到:

$$\begin{aligned} h(x_1)\mathbf{H}^T C \boldsymbol{\xi}_1 - \mathbf{t}_1^T + \boldsymbol{\xi}_1^T &= 0 \\ &\vdots \end{aligned} \quad (5)$$

$$h(x_N)\mathbf{H}^T C \boldsymbol{\xi}_N - \mathbf{t}_N^T + \boldsymbol{\xi}_N^T = 0$$

其中:

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & \cdots & t_{1m} \\ \vdots & \vdots & \vdots \\ t_{N1} & \cdots & t_{Nm} \end{bmatrix} \quad (6)$$

$$\mathbf{H} = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix}$$

于是上面公式可以合并写成:

$$\left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T\right)\boldsymbol{\alpha} = \mathbf{T} \quad (7)$$

最终可以推导得到:

$$\boldsymbol{\beta} = \mathbf{H}^T (\mathbf{H}\mathbf{H}^T + \frac{\mathbf{I}}{C})^{-1} \mathbf{T} \quad (8)$$

于是极限学习机的逼近函数可以写成:

$$f(x_i) = h(x_i)\mathbf{H}^T (\mathbf{H}\mathbf{H}^T + \frac{\mathbf{I}}{C})^{-1} \mathbf{T} \quad (9)$$

可以看出,无论是 $h(x_i)\mathbf{H}^T$ 还是 $\mathbf{H}\mathbf{H}^T$ 都是矩阵内积的形式,可以构造一个满足 mercer 定理的核函数来代替这个内积,于是有如下公式:

$$\begin{aligned} \mathbf{H}\mathbf{H}^T = \boldsymbol{\Omega}_{\text{ELM}} &= \begin{bmatrix} K(\mathbf{x}_1, \mathbf{x}_1) & \cdots & K(\mathbf{x}_1, \mathbf{x}_j) \\ \vdots & \ddots & \vdots \\ K(\mathbf{x}_N, \mathbf{x}_1) & \cdots & K(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix} \\ h(\mathbf{x})\mathbf{H}^T &= \begin{bmatrix} K(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}, \mathbf{x}_N) \end{bmatrix} \end{aligned} \quad (10)$$

于是核极限学习机的逼近函数可以写成:

$$f(\mathbf{x}) = \begin{bmatrix} K(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}, \mathbf{x}_N) \end{bmatrix}^T \left(\frac{\mathbf{I}}{C} + \boldsymbol{\Omega}_{\text{ELM}}\right)^{-1} \mathbf{T} \quad (11)$$

2.2 组合核极限学习机

只要满足 mercer 定理的核函数都可以作为极限学习机的核函数,如高斯径向基核函数(RBF)、多项式核函数(polynomial)等,但是往往每个核函数在不同的应用领域都各有它的优势,而且单独的一个核函数往往并不能达到最佳的逼近效果^[10-14],所以文中提出了一种基于组合核的极限学习机算法,可以构造权重对不同的核函数进行加权组合。

文中采用径向基核和多项式核加权组合的方式来构造多核核函数。RBF 核是一种局部核函数,多项式

核是一种全局核函数,局部核函数的学习力比较强,而全局核函数泛化力强,学习能力弱,因此两者组合能发挥出比较好的效果。

RBF 核公式为:

$$k(x, x_i) = \exp(-\|x - x_i\|^2/a)$$

多项式核公式为:

$$k(x, x_i) = (xx_i + 1)^q$$

混合核函数的公式为:

$$k(x, x_i) = p (xx_i + 1)^q + (1 - p) \exp(-\|x - x_i\|^2/a) \quad (12)$$

文中把这种采用了多核组合的核极限学习机称之为多核极限学习机 (Multi Kernel Extreme Learning Machine, MKELM)。

2.3 混沌粒子群算法 (CPSO)

针对传统 PSO 算法存在早熟的问题,将混沌理论引入粒子群算法对其进行改进,其算法具体流程如下:

(1)混沌初始化。假设优化变量为 D 维。随机生成 D 维向量 $z_1 = [z_{11}, z_{12}, \dots, z_{1D}]$, 每个分量均处于 $[0, 1]$, 依据 Logistic 方程获得 M 个分量, z_1, z_2, \dots, z_M :

$$z_{n+1} = \mu z_n (1 - z_n), n = 0, 1, \dots; 0 < z_n < 1; \mu \in [0, 4] \quad (13)$$

运用式 (13) 实现混沌区间和变量范围的映射:

$$x_{ij} = a_j + (b_j - a_j) z_{ij}$$

其中, b_j, a_j 分别表示所需优化变量的上界和下界。

(2)依据适应度函数评估每个粒子的适应度函数值,从 M 个初始粒子群中选择 N 个作为初始解,粒子的速度随机生成。

(3)设定初始个体极值和全局极值。设定各粒子的当前位置为个体极值 P_i , 依据适应度函数评估各个个体极值 P_i 的适应度函数值,选择最优值的粒子所在位置定义为全局极值 P_g 。

(4)根据式 (14) 和式 (15) 实现粒子飞行速度和位置的更新。

$$v_{id}(t+1) = v_{id}(t) + c_1 \text{rand}_1 * (p_{id} - x_{id}(t)) + c_2 \text{rand}_2 * (p_{gd} - x_{id}(t)) \quad (14)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \quad (15)$$

(5)混沌优化最优位置 P_g 。先将最优位置映射成为 Logistic 方程的取值范围 $[0, 1]$, 再根据 Logistic 方程生成 m 个混沌变量序列,最后将生成的混沌变量序列映射到优化变量,获得 m 个粒子,评估每个粒子的适应度函数值,获得最优解 p' 。

(6)用最优解 p' 更新当前粒子搜索群体中任意粒子的位置。

(7)转到步骤 (4), 若粒子群终止条件满足,则输出最优结果;反之,继续迭代。

2.4 混沌粒子群算法优化组合核极限学习机

需要优化的参数为组合核的权值 p , 高斯核的参数 a , 多项式核的参数 q , 以及惩罚因子 C 。

构造的适应度函数如下所示:

$$\min \text{Fitness}(t) = \frac{1}{2n} \sum_{i=1}^n [e(t)]^2 \quad (16)$$

采用 2.3 所述算法进行优化,最终输出 4 个参数的最优化结果。

3 仿真实验

文中算法步骤如下:先采用 EMD 算法将时间序列分解成几个分量,然后分别采用 CPSO-MKELM 进行预测,最后将各个预测结果进行相加组合重构出最终的预测结果。

3.1 数据来源

文中数据来源于流量文库,收集自 2014 年 11 月 7 日—2014 年 11 月 21 日一共 15 天的真实流量数据为研究对象,每天每间隔 1 小时采集一次网络访问流量,一共采集 $15 * 24 = 360$ 组数据。网络流量的原始网络流量数据如图 1 所示。

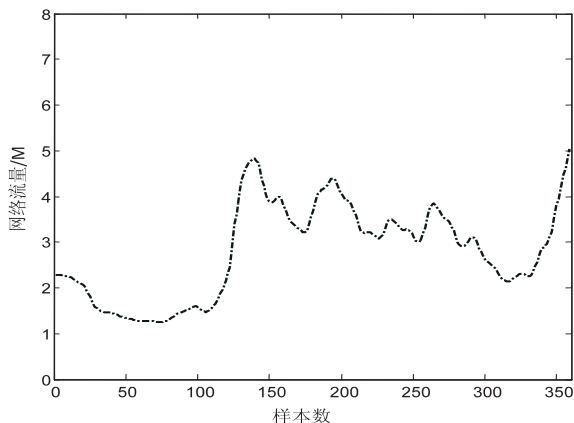


图1 原始网络流量

3.2 数据处理

对原始网络流量数据序列进行 EMD 分解,依次分解出不同的 IMF 分量,网络流量数据被分解成 4 个波动较小的分量和 1 个剩余分量。

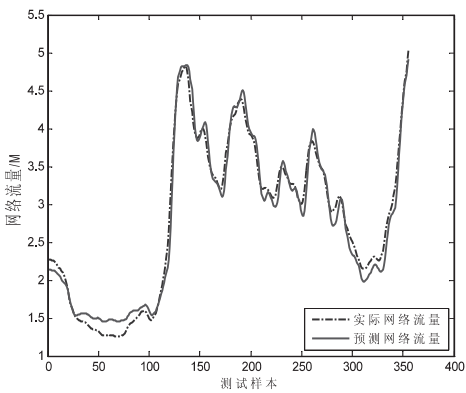
针对每个分量采用时间序列的方法进行预测,以其中一个分量 IMF_1 为例,采用 $\text{IMF}_1(t-n), \text{IMF}_1(t-n+1), \dots, \text{IMF}_1(t)$ 来预测 $\text{IMF}_1(t+1)$ 。其中, n 为时间序列预测的步长。

运用 CPSO 优化 MKELM 的模型分别对各个分量进行预测,最后通过各个分量的预测结果重构出网络流量的预测结果。结果采用均方误差来评价算法的性能。

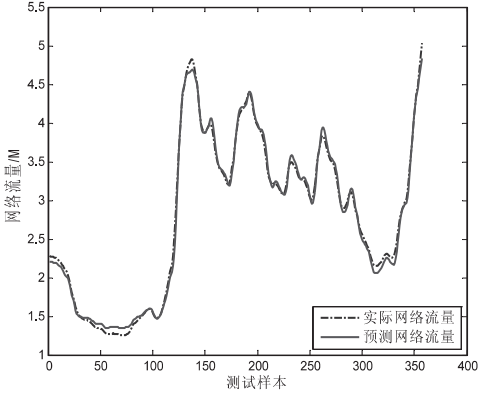
3.3 实验结果分析

将 360 组网络流量数据分成训练样本和测试样本,将 336 组数据作为训练数据,后面 24 组作为测试数据,用于验证预测结果的好坏。设定 CPSO 算法的

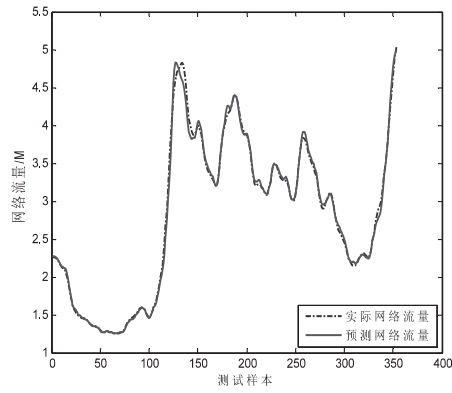
最大迭代次数为 100,种群大小为 20,其预测结果如图 2 所示,分别表示单步预测、3 步预测、5 步预测和 7 步预测。



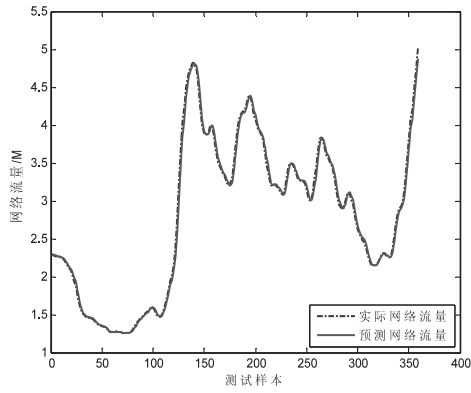
(a) 单步预测结果



(b) 3 步预测结果



(c) 5 步预测结果



(d) 7 步预测结果

图 2 基于 EMD-CPSO-MKELM 算法的 7 步网络流量预测

由 EMD-CPSO-MKELM 算法的单步预测、3 步预测、5 步预测和 7 步预测结果图可知,随着预测步长的增加,EMD-CPSO-MKELM 算法的预测精度不断提高,效果较好。图 3 是 CPSO 算法优化 MKELM 的适应度曲线。

表 1 四种算法预测的 MSE 误差对比

运行次数	SVM	CPSO-KELM	CPSO-MKELM	EMD-CPSO-MKELM
1	0.008 8	0.006 0	0.005 6	0.004 0
2	0.007 2	0.005 5	0.005 1	0.004 6
3	0.006 8	0.005 0	0.004 9	0.004 0
4	0.008 4	0.006 4	0.005 1	0.004 2
5	0.006 5	0.005 6	0.004 7	0.003 8
6	0.007 4	0.006 1	0.005 8	0.004 3
7	0.007 6	0.005 6	0.004 6	0.003 5
8	0.006 4	0.004 6	0.004 3	0.004 2
9	0.007 2	0.007 6	0.005 4	0.004 0
10	0.006 7	0.005 1	0.004 8	0.004 7
均值	0.007 2	0.005 8	0.005 2	0.004 1

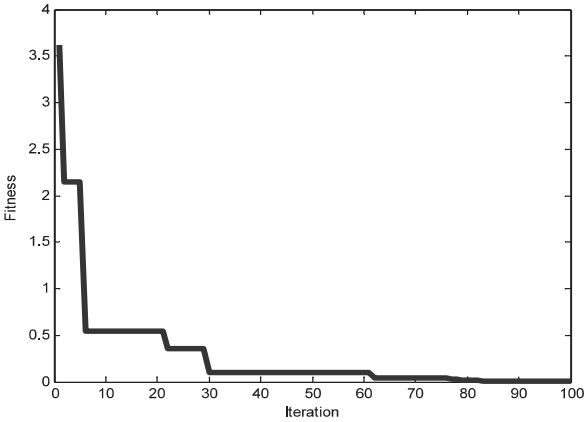


图 3 CPSO 优化 MKELM 的适应度曲线图

为了对比 EMD-CPSO-MKELM 算法的优越性,将 EMD-CPSO-MKELM 算法、CPSO-MKELM 算法、CPSO-KELM 算法和普通的 SVM 算法四者的预测结果进行对比,运行 10 次,其对比结果如表 1 所示。

从表 1 中可以看出,文中的多核极限学习机算法要好于普通的核极限学习机算法,同时由 MSE 误差对比结果可知,文中的 EMD-CPSO-MKELM 算法的预测效果最好。

4 结束语

针对核极限学习机逼近能力的不足,文中提出了一种高斯核函数和多项式核函数加权组合的核极限学习机方法。针对该算法参数选择对识别性能的影响,文中运用混沌粒子群算法对组合核极限学习机的核参数、权值以及惩罚因子进行优化,同时结合 EMD 提取网络流量的细节特征和趋势特征,构建出基于 EMD-CPSO-MKELM 的网络流量预测模型,分别进行单步、3 步、5 步和 7 步预测。通过对比不同网络流量预测模型的预测均方误差发现,文中提出的算法预测精度要优于其他模型,从而为网络资源的合理配置和可靠传输提供决策的依据。

参考文献:

[1] 雷 霆,余镇危. 一种网络流量预测的小波神经网络模型[J]. 计算机应用,2012,26(3):526-528.

[2] 杨 光,张国梅,刘星宇. 基于小波核 LS-SVM 的网络流量预测[J]. 微机发展(现更名:计算机技术与发展),2005,15(12):125-128.

[3] Huang G B,Zhu Q Y,Siew C K. Extreme learning machine:a new learning scheme of feedforward neural networks[J]. Neurocomputing,2004,2(2):985-990.

[4] Huang G B,Ding X,Zhou H. Optimization method based extreme learning machine for classification[J]. Neurocomput-

ing,2010,74(1-3):155-163.

[5] 武峰雨,乐秀璠,南东亮. 相空间重构的极端学习机短期风速预测模型[J]. 电力系统及其自动化学报,2013,25(1):136-141.

[6] 丁金林,王 峰,孙 洪,等. 改进 RELM 在多变量解耦控制中的应用[J]. 微电子学与计算机,2012,29(11):70-73.

[7] 肖志勇,杨小玲,刘爱伦. 基于改进 EMD 和 LSSVM 的机械故障诊断[J]. 自动化仪表,2008,29(6):24-26.

[8] 冯 平,丁志宏,韩瑞光,等. 基于 EMD 的降雨径流神经网络预测模型[J]. 系统工程理论与实践,2009,29(1):152-158.

[9] 徐晓刚,徐冠雷,王孝通,等. 经验模式分解(EMD)及其应用[J]. 电子学报,2009,37(3):581-585.

[10] 郑志成,徐卫亚,徐 飞,等. 基于混合核函数 PSO-LSSVM 的边坡变形预测[J]. 岩土力学,2012,33(5):1421-1426.

[11] Han Jianwei,Kamber M. 数据挖掘概念与技术[M]. 北京:机械工业出版社,2007.

[12] Freund Y,Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. Journal of Computer and System Sciences,1997,55(1):119-139.

[13] Moore A W,Zuev D. Internet traffic classification using Bayesian analysis techniques[C]//Proc of SIGMETRICS. [s. l.]:[s. n.],2005:50-60.

[14] 杨家海,吴建平,安常青. 互联网络测量理论与应用[M]. 北京:人民邮电出版社,2009.

(上接第 72 页)

CVPR. Piscataway,NJ:IEEE,2011:9-16.

[3] Kosiba D A,Devaux P M,Balasubramanian S,et al. An automatic jigsaw puzzle solver[C]//Proceedings of the 12th International conference on image processing. Piscataway,NJ:IEEE,1994:616-618.

[4] Cho T S,Avidan S,Freeman W T. A probabilistic image jigsaw puzzle solver[C]//Proceedings of CVPR. Piscataway,NJ:IEEE,2010:183-190.

[5] Yang X,Adluru N,Latecki L J. Particle filter with state permutations for solving image jigsaw puzzles[C]//Proceedings of CVPR. Piscataway,NJ:IEEE,2011:2873-2880.

[6] Alajlan N. Solving square jigsaw puzzles using dynamic programming and the hungarian procedure[J]. American Journal of Applied Sciences,2009,5(11):1941-1947.

[7] Sholomon D,David O,Netanyahu N S. A genetic algorithm-based solver for very large jigsaw puzzles[C]//Proceedings of CVPR. Piscataway,NJ:IEEE,2013:1767-1774.

[8] Toyama F,Fujiki Y,Shoji K,et al. Assembly of puzzles using a

genetic algorithm[C]//Proceedings of CVPR. Piscataway,NJ:IEEE,2002.

[9] Held M,Karp R M. The traveling-salesman problem and minimum spanning trees:Part II[J]. Mathematical Programming,1971,1(1):6-25.

[10] 朱 青. 计算机算法与程序设计[M]. 北京:清华大学出版社,2009:122-126.

[11] Kruskal J B. On the shortest spanning subtree of a graph and the travelling salesman problem[J]. Proceedings of the American Mathematical Society,1956,7:48-50.

[12] Cormen T H,Leiserson C E,Rivest R L,et al. Introduction to algorithms[M]. 2nd ed. [s. l.]:McGraw-Hill,2001:1297-1305.

[13] 朱清新,杨 凡,钟黔川. 计算机算法设计与分析导论[M]. 北京:人民邮电出版社,2008.

[14] Matsui T. The minimum spanning tree problem on a planar graph[J]. Discrete Applied Mathematics,1995,58(1):91-94.