

# Prediction Model of Ship Traffic Flow Based on MFO-LSSVM

Shan Zhu, Liqian Sun

China Communications Information Technology Group Co. Ltd., Beijing  
Email: zhushan1989@126.com

Received: May 12<sup>th</sup>, 2020; accepted: May 26<sup>th</sup>, 2020; published: Jun. 2<sup>nd</sup>, 2020

---

## Abstract

Current dominant ship traffic flow prediction models don't consider the characteristics of the data and achieve high accuracy in the prediction process, to resolve these problems, a prediction model based on moth flame optimization algorithm and least squares support vector machine is proposed from the perspective of influencing factors of ship traffic flow. The essence of the model is to optimize the internal parameters of LSSVM model by moth flame algorithm and the model is trained based on collected data. To verify the validity of the proposed model, experiments are conducted based on the relevant data of ship traffic flow in Guangdong Province of China, and compared with FOA-LSSVM, PSO-LSSVM and GA-LSSVM models, the experimental results show that the MFO-LSSVM model has higher prediction accuracy and efficiency, the effectiveness of the proposed model is verified and can be used for the prediction of ship traffic flow.

## Keywords

Intelligent Transportation, Traffic Flow Prediction, Moth-Flame Optimization, Ship Traffic Flow, Least Squares Support Vector Machine

---

## 基于MFO-LSSVM的船舶交通流量预测模型

朱 珊, 孙立谦

中国交通信息科技集团有限公司, 北京  
Email: zhushan1989@126.com

收稿日期: 2020年5月12日; 录用日期: 2020年5月26日; 发布日期: 2020年6月2日

---

## 摘 要

针对当前船舶交通流模型没有充分考虑流量数据本身特性、预测方法精度不高的问题, 提出了一种基于

飞蛾火焰优化算法(Moth-flame Optimization, MFO)和最小二乘支持向量机(least squares support vector machine, LSSVM)的预测模型, 该模型主要利用飞蛾火焰算法对LSSVM模型内部参数进行优化, 基于采集的数据进行模型训练和预测。为验证模型有效性, 利用我国广东省船舶交通流量等相关数据进行实验, 并与FOA-LSSVM、PSO-LSSVM和GA-LSSVM等模型进行对比分析, 结果表明MFO-LSSVM模型具有较高的预测精度和预测效率, 验证了方法的有效性, 可以用于船舶交通流量的预测。

## 关键词

智能交通, 交通流量预测, 飞蛾火焰优化算法, 船舶交通流量, 最小二乘支持向量机

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

船舶交通流量作为船舶交通状况的主要反映尺度, 其大小可以反映交通是否拥挤, 只有实时掌握船舶交通流量信息, 才能更准确地进行预测并为船舶交通规划管理提供依据。目前国内外对于船舶交通流量预测方法研究的文献有很多, 主要研究思路分为基于影响因素的预测和基于历史数据的预测, 在这两方面分别采用不同的预测方法进行研究, 并改进相关预测方法来提高船舶交通流量预测准确性。

目前国内外对于基于影响因素的船舶交通流量预测的文献有很多, 总的来说, 船舶交通流量的大小与多方面的因素有关, 如社会经济发展水平、研究的交通流所处地域、当地的物产、货运市场等, 其影响因素又有内因与外因之分。内因即船舶个体特征, 包括船舶种类、船舶尺度、船舶吨级、船舶操纵性能等; 外因包括港口腹地经济发展水平、港口泊位因素、航道条件、锚地条件、航行规定及管理要求、水域分布、自然条件、导助航设施等。冯宏祥和肖英杰[1]于2011年针对船舶交通流的随机性大、影响因素多、船舶交通流量预测的准确度不高等现象, 基于支持向量机理论, 结合长江苏通大桥断面的观测数据, 建立预测模型进行短期的船舶流量预测。白响恩和张浩[2]于2012年提出一种结合粗糙集和支持向量回归智能算法的交通流预测模型, 通过ROSETTA软件进行属性约简预处理, 筛选出影响交通流变化的关键影响因素, 运用Libsvm软件构建基于遗传算法参数寻优的支持向量回归预测模型。张树奎和肖英杰[3]综合考虑季节、气候等因素, 通过分析历史流量数据, 在线性增长模型的基础上构建了考虑周期性波动因素的船舶交通流量预测改进模型, 并运用贝叶斯估计和预测方法求解模型。L Qi等[4]2016年提出了一种称为空间-逻辑映射模型的船舶交通流元胞自动机(CA)模型来提高海上运输效率和安全性, 并以新加坡海峡的船舶交通流为例, 进行了一些模拟和比较。

受到实际影响因素数据无法准确获得的制约, 很多学者尝试从历史数据寻找规律, 对未来交通流量进行准确预测, 开发了很多复杂技术和算法, 这其中主要可以归类为非数学或数学方法[5]。非数学方法包括支持向量机、神经网络等, 数学方法则包括线性和非线性回归、平滑技术和自回归线性过程等。而由于交通状态和流量数据的不确定性、复杂性等特点, 传统的数学方法通常面临着模型难以构建的难题, 而非数学方法由于不需要复杂的模型构建因此也被广泛应用于交通流预测, 这其中主要以神经网络和支持向量机为代表, 相比于神经网络, 支持向量机基于结构风险最小化原则有效解决了小样本、局部最优和维数灾难等问题, 泛化能力得到大大提高[6], 而最小二乘支持向量机作为一种在标准SVM做出改进的机器学习技术, 合理解决了线性KKT系统问题, 并简化计算复杂性, 因而可以从非线性系统建模的角

度进行高精度拟合和预测。随着智能优化技术的进一步发展,越来越多的智能优化算法也被用于 LSSVM 模型的参数确定,如遗传算法,模拟退火算法,粒子群算法等。

船舶运输需求是一种派生性需求,因此影响船舶交通流量的外部环境因素和经济因素众多,单单只通过流量的时间序列数据往往不能进行准确预测,同时由于短期交通流表现出的相关特征,也导致在同一地点连续采集的数据通常会呈现强烈的季节性[7] [8],因此本文通过收集船舶交通流量的季度数据,从流量影响因素的角度对船舶交通流量进行分析,将 LSSVM 与近年来新提出的 MFO 算法相结合,提出了一种新的短期船舶交通流量预测模型,并选择其他算法优化的 LSSVM 模型来进行结果精确度和方法有效性的比较,进而实现流量精确预测。

## 2. 模型构建

船舶交通流量通常呈现出剧烈的波动性和随机性,并在数据的变化趋势中也反映出外部环境(包括港口条件、区域经济等)、突发或随机事件等因素的变化情况,因此本文在构建模型时考虑所预测水域的实际情况,从影响因素的角度构建 MFO-LSSVM 模型进行船舶交通流量预测。

### 2.1. 算法原理

本文应用飞蛾火焰优化算法对 LSSVM 中核函数参数和惩罚参数同时进行优化。飞蛾火焰优化算法是在 2015 年提出的一种新型仿生群体智能算法,算法思路是自然界飞蛾的横向定位导航机制,如图 1 所示,在这种横向导航方式中,飞蛾相对于火焰始终保持着一个固定的角度来飞行,并通过螺旋飞行更新飞蛾和火焰的位置,并最终收敛于火焰位置。相比于传统的粒子群算法和遗传算法,飞蛾火焰优化算法在探索个体最优和开发全局最优方面实现了相对较好的平衡,从而避免算法过早陷入局部最优[9]。最小二乘支持向量机算法于 1999 年由 Suykens [10]提出,本文应用最小二乘支持向量机解决预测过程中过学习、非线性、局部极值点以及高位数等实际问题。

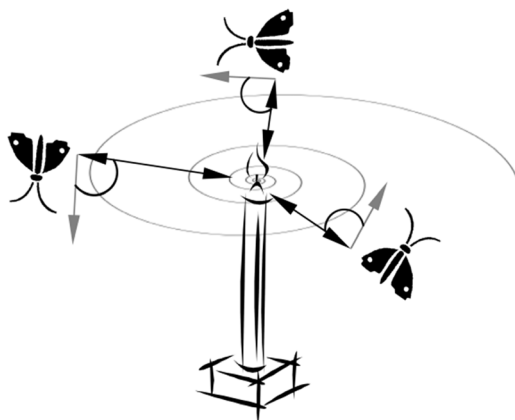


Figure 1. Spiral flying path around close light sources

图 1. 飞蛾靠近火焰的螺旋飞行途经

因此, MFO 算法选择螺旋函数来对飞蛾空间位置进行更新,并在飞行的优化过程中有以下原则: 1) 螺旋的初始点是飞蛾的初始化空间位置,最后一点为火焰的最佳空间位置; 2) 螺旋的波动范围不可以超出搜索空间; 3) 在火焰周围,飞蛾能够随着  $t$  ( $t$  为  $[-1, 1]$  之间的随机数) 的变化从而收敛到任何点,  $t$  越小代表离火焰的距离越近; 4) 当飞蛾逐渐靠近火焰时,火焰两侧的位置更新频率随之增加。MFO 算法数学模型的原理描述如下[11]:

设定待优化问题的候选解为飞蛾所处空间位置, 利用式(1)矩阵表示飞蛾种群规模为  $n$ 、优化问题维度为  $d$  的飞蛾所处的空间位置; 利用式(2)矩阵记录飞蛾个体适应度值。MFO 算法中的另一关键部分则是火焰。其矩阵与飞蛾矩阵类似, 表示为式(3), 并利用式(4)矩阵记录火焰适应度值。

$$M = \begin{bmatrix} m_{1,1} & \cdots & m_{1,d} \\ \vdots & & \vdots \\ m_{n,1} & \cdots & m_{n,d} \end{bmatrix}; \quad (1)$$

$$OM = [OM_1 \quad \cdots \quad OM_n]^T. \quad (2)$$

$$F = \begin{bmatrix} F_{1,1} & \cdots & F_{1,d} \\ \vdots & & \vdots \\ F_{n,1} & \cdots & F_{n,d} \end{bmatrix}; \quad (3)$$

$$OF = [OF_1 \quad \cdots \quad OF_n]^T. \quad (4)$$

式中:  $n$  为飞蛾种群规模;  $d$  为优化问题维度。在这里, 飞蛾和火焰均为候选解, 其区别在于二者位置的更新方式。飞蛾是在搜索空间附近移动的搜索动因, 而火焰是当前迭代所获得的最佳位置, 即通过不断更新火焰位置来获得最佳位置(候选解)。因此, MFO 算法可定义为:

$$\text{MFO} = (I, P, K). \quad (5)$$

式中:  $I$  为随机生成的飞蛾种群规模和相应的适应值函数值;  $P$  为飞蛾在搜索空间移动的函数, 其为  $M$  矩阵和位置更新返回值;  $K$  为终止条件是否满足的判别函数。

用数学模型描述飞蛾—火焰位置更新如下:

$$M_i = S(M_i, F_j). \quad (6)$$

式中:  $M_i$  表示第  $i$  只飞蛾位置;  $F_j$  表示第  $j$  个火焰;  $S$  为螺旋函数。

定义用于模拟飞蛾螺旋飞行路径的 MFO 算法表达式如下:

$$S(M_i, F_j) = D_i e^{bt} \cos(2\pi t) + F_j. \quad (7)$$

式中:  $D_i$  为第  $j$  个火焰到第  $i$  只飞蛾的距离, 表示为  $D_i = |F_j - M_i|$ ;  $b$  为螺旋形状常数;  $t$  为  $[-1, 1]$  之间的随机数。

为保证 MFO 算法获得较快的收敛速度, 提出自适应火焰数量更新机制, 即利用在迭代过程中自适应地减少火焰的数目。其公式为:

$$\text{flame no} = \text{round}\left(N - l \cdot \frac{N-1}{T}\right). \quad (8)$$

式中:  $l$  为当前迭代次数;  $N$  为最大火焰数量;  $T$  为最大迭代次数。

从式(7)可得飞蛾的下一个位置被定义为一个火焰; 参数  $t$  定义了飞蛾的下一个位置应该接近的火焰 ( $t = -1$  表示最接近火焰位置, 而  $t = 1$  表示离火焰位置最远)。因此, 螺旋位置更新算子决定了飞蛾如何在火焰周围进行位置更新; 螺旋函数方程允许不一定在该空间内的飞蛾“围绕”火焰飞翔, 有效保证了 MFO 算法探索和开发空间的能力。

## 2.2. 模型构建

MFO-LSSVM 模型预测流程如图 2 所示。

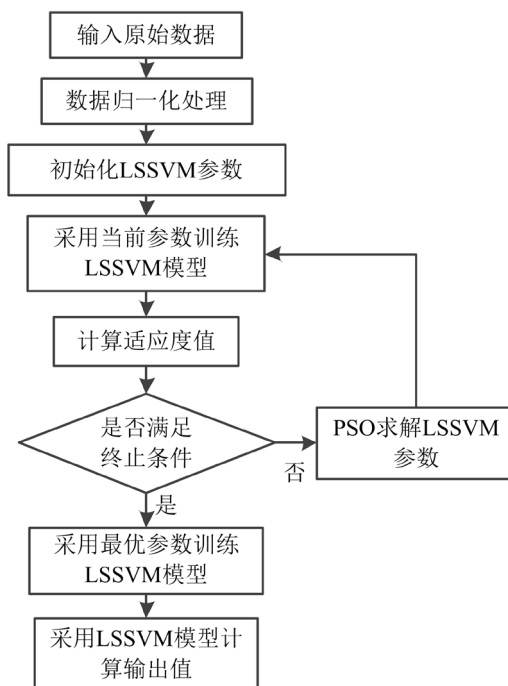


Figure 2. Algorithm process of MFO-LSSVM

图 2. MFO-LSSVM 模型算法流程图

步骤一：初始化参数。设定 MFO 的最大迭代次数  $l$ 、飞蛾和火焰种群规模  $sizepop$  等参数，随机产生初始解。

步骤二：初始计算。将初始飞蛾群代入 LSSVM 模型计算并记录适应度函数，寻找当前最优个体并将其保存进入火焰适应度矩阵，判断是否满足终止条件，满足则转至步骤四。适应度函数为预测值与实际值的均方根误差：

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (l_i - \hat{l}_i)^2}{n}} \quad (9)$$

式中， $l_i$  和  $\hat{l}_i$  分别实际季度船舶交通流量和训练船舶交通流量输出值， $n$  为训练样本的数量。

步骤三：迭代过程。根据式(8)更新火焰数量；计算飞蛾与火焰之间的距离，利用式(7)对飞蛾和火焰的位置进行更新并重新计算适应度函数值，利用式(2)、式(4)分别保存飞蛾及火焰空间位置，并寻找当前最优飞蛾个体，并与前次保留位置进行比较和取舍，判断是否满足终止条件，满足转至步骤四，否则令  $l = l + 1$ ，并执行步骤三。

步骤四：算法终止。得到最佳适应度  $S_i$ ，输出最优个体值和全局极值，并将全局极值代入 LSSVM 模型用于船舶交通流量预测，算法结束。

### 3. 仿真与结果分析

#### 3.1. 数据描述

选取广东省 2010 年第 1 季度到 2017 年第 1 季度共 29 个季度的船舶交通流量，原始数据分布如图 3 所示，这些船舶交通流量的季度数据均值为 743,980 艘，最大为 872,627 艘，最小为 576,308 艘，由图可以看出数据波动幅度较大，同时这些季度数据在一年里的变化情况存在相似性，即存在季节性波动特征。

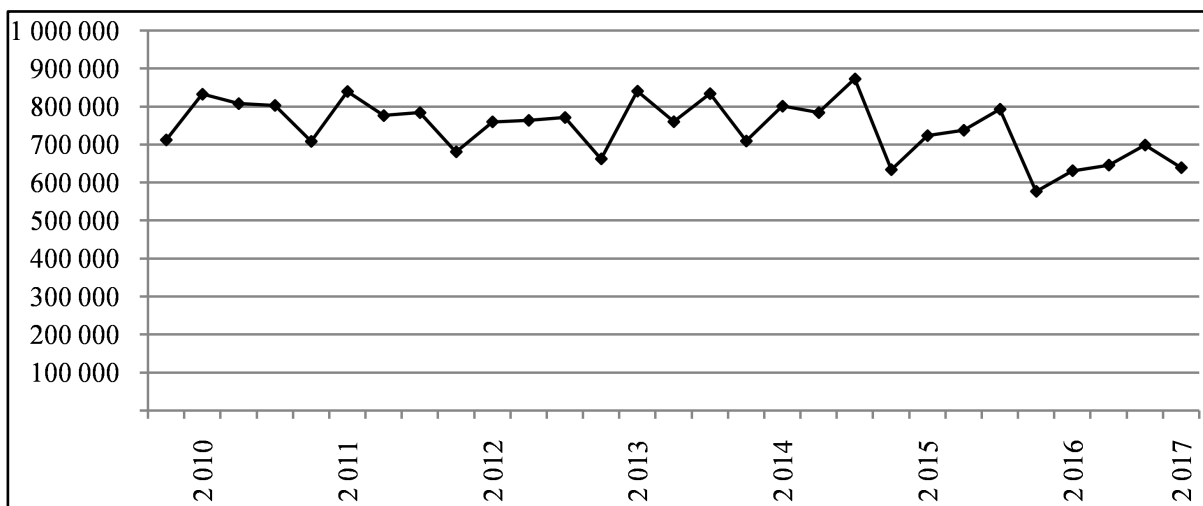


Figure 3. Distribution chart of ship traffic flow

图 3. 广东省船舶交通流量分布图

通过实际调研以及科技资料查阅,发现区域经济、船舶、港口等均能够影响船舶交通流量。通过分析各个因素对于船舶交通流量的影响程度,本文从宏观和微观角度选取广东省 GDP、广东省进出口贸易额、广东省货物吞吐量、广东省旅客吞吐量、广东省船舶总载重量作为主要影响因素来对广东省船舶交通流量进行预测。同时选取 FOA-LSSVM 模型、PSO-LSSVM 模型和 GA-LSSVM 模型来进行结果和方法有效性分析。

### 3.2. 模型训练

在对 LSSVM 进行训练前对数据进行归一化处理,避免数据存在量纲问题而导致训练错误,归一化公式为  $z = (x - x_{\min}) / (x_{\max} - x_{\min})$ , 其中  $x_{\max}$  和  $x_{\min}$  是每一项输入数据和输出数据的最大值和最小值。根据前面的影响因素分析结果,以一组影响因素数据和船舶交通流量数值作为模型的输入和输出,则所有的输入数据可处理为  $5 \times 29$  的矢量,所有的输出数据为  $1 \times 29$  的矢量。基于 LSSVM 模型对样本数据进行训练时,将数据分为两组:训练组和验证组,两组数据占全部样本比例约为 90% 和 10%。按此比例,在 29 组数据中选择 26 组数据作为训练数据,3 组作为检验数据。

采用训练组数据对 MFO-LSSVM 进行训练。核函数和 LSSVM 内部参数  $C$  和  $\sigma$  作为影响 LSSVM 预测精度的主要变量,在训练过程中,因为目前没有权威的方法来确定最佳的核函数,故本文按照一般规律使用 RBF 径向基核函数,另外使用 MFO 算法对模型内部参数  $C, \sigma$  进行优化,对于每组内部参数得到的支持向量机,都采用前文所述的均方根误差对预测后的流量结果进行适应度计算,从而来检验 LSSVM 的训练效果。本文设定 MFO 算法的迭代次数  $l = 100$ , 因为优化对象为模型内部参数  $C$  和  $\sigma$ , 故问题维度为 2, 飞蛾和火焰的群体规模设置为 100, 空间的上下界设置为  $[lb, ub] = [-1000, 1000]$ 。

### 3.3. 仿真实验

经过 MFO 对 LSSVM 内部参数的优化调整,模型适应度变化如图 4 所示,然后以检验数据的输入数据进行模型输入,得到模型预测结果并与检验数据的输出数据进行结果比较,比较预测精度,并采用此预测模型对将这 26 组训练数据进行重新拟合,这 29 组数据的预测值和实际值的拟合效果如图 5 所示。

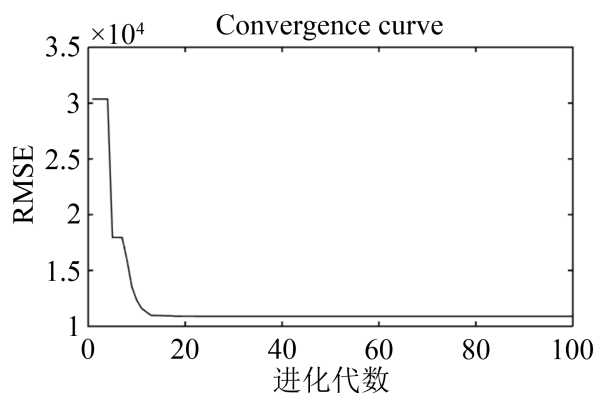


Figure 4. Iteration figure of MFO-LSSVM

图 4. MFO-LSSVM 迭代图

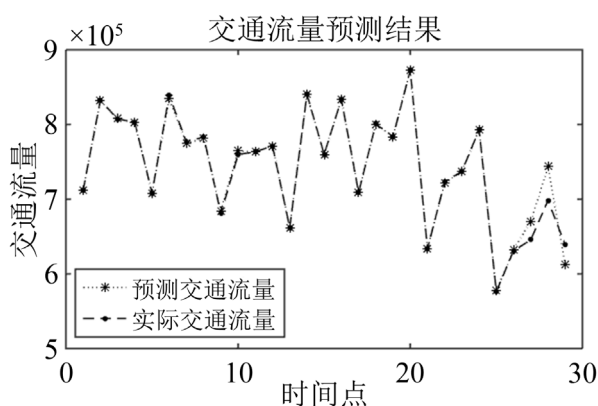


Figure 5. Forecast result of MFO-LSSVM

图 5. MFO-LSSVM 预测结果图

可以看出 MFO-LSSVM 对检验组的广东省船舶交通流量具有良好的预测效果, 同时对 26 组训练数据能够基本实现 0 误差, 通过预测结果可得, 模型在迭代过程中很快就实现基本收敛, 波峰和波谷的吻合度很高, 说明此模型在解决复杂非线性问题上具有优势。

### 3.4. 效果分析

为了评估 MFO-LSSVM 模型对于广东省船舶交通流量的预测效果, 本文采用 FOA-LSSVM 模型、PSO-LSSVM 模型和 GA-LSSVM 模型来进行对比预测, 并采用一些相同的指标对模型预测效果进行比较。

LSSVM 模型的内部参数都由相应的算法进行模型优化。迭代前对相应优化算法的参数进行设置, FOA 如下设置:  $\maxgen = 100$ ,  $sizepop = 100$ ,  $(X\_axis, Y\_axis) \in (-1000, 1000)$ ; PSO 算法如下设置:  $\maxgen = 100$ ,  $swarm\ size = 100$ ,  $particle\_size = 2$ ,  $(popc, popg) \in (-1000, 1000)$ , 学习参数  $c_1 = c_2 = 2$ ; GA 算法如下设置:  $\maxgen = 100$ ,  $sizepop = 100$ ,  $pcross = 0.5$ ,  $pmutation = 0.5$ ,  $numsum = 2$ 。仍然按照 2.2 节的方式进行模型训练和预测, 三个模型的适应度函数收敛图和预测效果图如图 6~11 所示, 模型预测结果如表 1 所示。将图 6~8 与图 4 对比可以得出, MFO-LSSVM 相比于其他三种方法, 在收敛速度上更快, 并且由图 9~11 可以看出三种模型基本实现了对流量数据变化趋势的拟合, 而为了更好地体现出对比效果, 本文采用训练数据的 RMSE、29 组数据拟合效果的总误差、29 组数据拟合效果的 MAPE 和开始收敛代数等四组指标对方法有效性进行比较分析, 结果如表 2 所示。

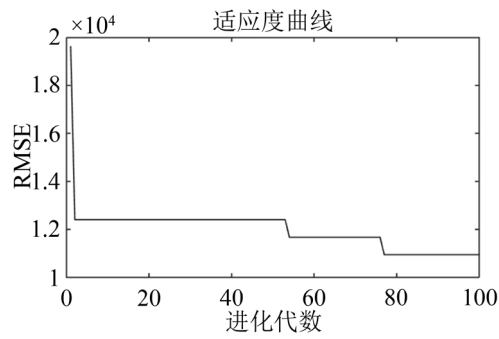


Figure 6. Iteration figure of FOA-LSSVM  
图 6. FOA-LSSVM 迭代图

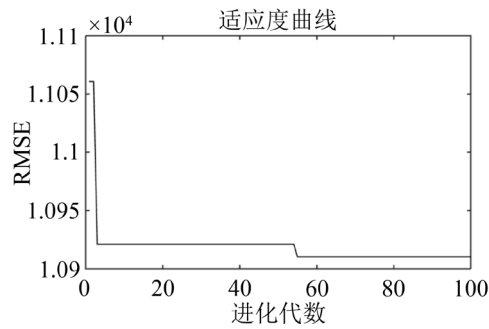


Figure 7. Iteration figure of PSO-LSSVM  
图 7. PSO-LSSVM 迭代图

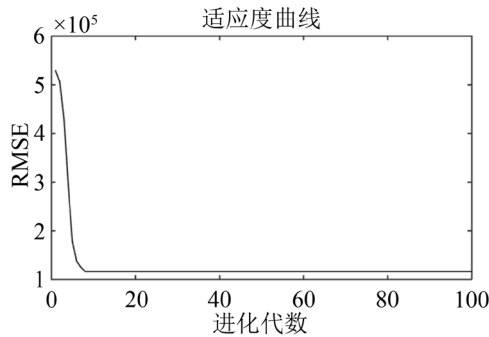


Figure 8. Iteration figure of GA-LSSVM  
图 8. GA-LSSVM 迭代图

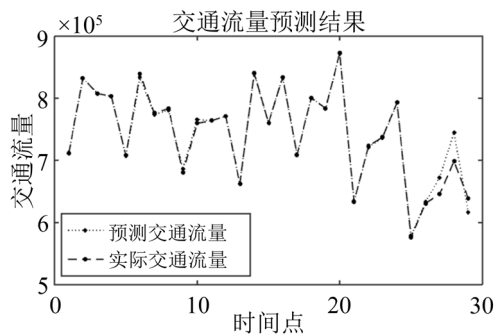


Figure 9. Forecast result of FOA-LSSVM  
图 9. FOA-LSSVM 预测结果图



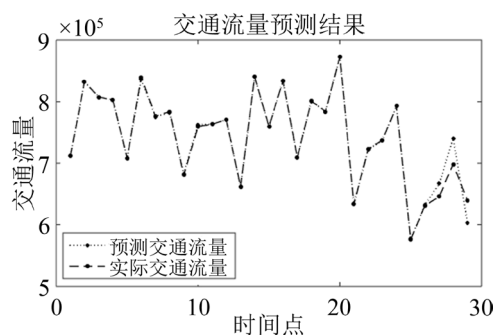


Figure 10. Forecast result of PSO-LSSVM

图 10. PSO-LSSVM 预测结果图

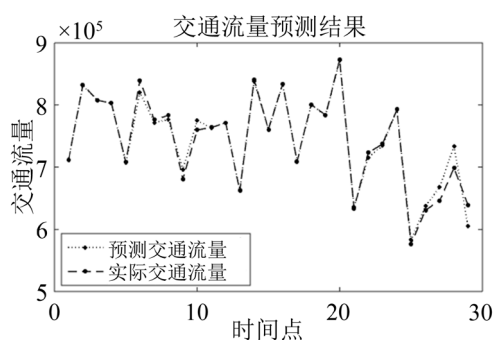


Figure 11. Forecast result of GA-LSSVM

图 11. GA-LSSVM 预测结果图

Table 1. Forecast result

表 1. 模型预测结果

季度	实际值	MFO-LSSVM	FOA-LSSVM	PSO-LSSVM	GA-LSSVM
2016 年 3 季度	645,568	666,650	672,530	670,421	702,839
2016 年 4 季度	698,378	740,267	744,461	742,296	661,123
2017 年 1 季度	638,797	602,799	616,566	611,967	568,828

Table 2. Results of related index

表 2. 模型相关指标结果

模型	MFO-LSSVM	FOA-LSSVM	PSO-LSSVM	GA-LSSVM
MAPE	0.4634	0.7006	0.5771	0.8809
RMSE(10 <sup>4</sup> )	1.0900	1.0940	1.0958	1.1632
总误差	0.1706	0.1876	0.1802	0.8402
开始收敛代数	20	77	71	11

由适应度函数和收敛代数比较可得, MFO-LSSVM 在收敛速度和收敛结果上效果明显, 同时 MFO-LSSVM 模型的 MAPE、RMSE 和总误差指标都明显小于其他三个模型, 这也显示出 MFO 全局启发式算法在 LSSVM 模型优化上的优越性, 同时模型的预测结果也相对更为精确, 因此, 本文提出的 MFO-LSSVM 相比于本文提出的其他三种模型能够更为精确地进行船舶交通流量预测。

## 4. 结论

本文针对船舶交通流量数据存在的波动性和随机性等特点, 从流量影响因素的角度提出了一种飞蛾火焰优化算法-最小二乘支持向量机预测模型, 通过广东省船舶交通流量数据进行实证分析, 结果表明本文建立的 MFO-LSSVM 模型并且相对于 FOA-LSSVM、PSO-LSSVM、GA-LSSVM 模型具有更好的预测性能和预测效果。MFO-LSSVM 模型综合了基于影响因素的预测方法和基于历史数据的预测方法的优势, 不仅可以考虑船舶交通流量历史数据规律特性, 还能够考虑区域经济、港口条件等因素对船舶交通流量的影响, 提高了船舶交通流量预测效率和预测精度。

本预测方法也有进一步改进的空间, 下一步将在飞蛾和火焰的匹配机制上进行创新, 避免算法过早陷入局部最优。

## 基金项目

国家重点研发计划资助(2018YFB1600400)。

## 参考文献

- [1] 冯宏祥, 肖英杰, 孔凡邨. 基于支持向量机的船舶交通流量预测模型[J]. 中国航海, 2011, 34(4): 62-66.
- [2] 张浩, 张晓东, 肖英杰, 等. 组合粗糙集和支持向量回归的船舶交通流预测[J]. 计算机工程与应用, 2012: 251-254.
- [3] 张树奎, 肖英杰. 考虑周期性波动因素的船舶交通流量预测模型[J]. 大连海事大学学报, 2016, 42(4): 41-46.
- [4] Qi, L., Zheng, Z. and Gang, L. (2016) A Cellular Automaton Model for Ship Traffic Flow in Waterways. *Physica A Statistical Mechanics & Its Applications*, **471**, 705-717. <https://doi.org/10.1016/j.physa.2016.12.028>
- [5] Ghosh, B., Basu, B. and O'Mahony, M. (2010) Random Process Model for Urban Traffic Flow Using a Wavelet-Bayesian Hierarchical Technique. *Computer-Aided Civil & Infrastructure Engineering*, **25**, 613-624. <https://doi.org/10.1111/j.1467-8667.2010.00681.x>
- [6] Sujay, R.N. and Deka, P.C. (2014) Support Vector Machine Applications in the Field of Hydrology: A Review. *Applied Soft Computing Journal*, **19**, 372-386. <https://doi.org/10.1016/j.asoc.2014.02.002>
- [7] Xiao, X., Yang, J., Mao, S., et al. (2017) An Improved Seasonal Rolling Grey Forecasting Model Using a Cycle Truncation Accumulated Generating Operation for Traffic Flow. *Applied Mathematical Modelling*, **51**, 386-404. <https://doi.org/10.1016/j.apm.2017.07.010>
- [8] Kumar, S.V. and Vanajakshi, L. (2015) Short-Term Traffic Flow Prediction Using Seasonal ARIMA Model with Limited Input Data. *European Transport Research Review*, **7**, 21. <https://doi.org/10.1007/s12544-015-0170-8>
- [9] Mirjalili, S. (2015) Moth-Flame Optimization Algorithm: A Novel Nature-Inspired Heuristic Paradigm. *Knowledge-Based Systems*, **89**, 228-249. <https://doi.org/10.1016/j.knsys.2015.07.006>
- [10] Suykens, J.A.K. and Vandewalle, J. (1999) Least Squares Support Vector Machine Classifiers. Kluwer Academic Publishers. [https://doi.org/10.1002/\(SICI\)1097-007X\(199911/12\)27:6<605::AID-CTA86>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-007X(199911/12)27:6<605::AID-CTA86>3.0.CO;2-Z)
- [11] 崔东文. 飞蛾火焰优化算法在承压含水层参数反演中的应用[J]. 长江科学院院报, 2016, 33(7): 28-33.