

文章编号:1001-9081(2008)05-1200-04

基于粗糙集和分形理论的交通流优化控制设计模型

王家伟¹, 黄大荣², 雷鸣²

(1. 重庆交通大学 计算机与信息学院, 重庆 400074; 2. 重庆交通大学 信息与计算科学研究所, 重庆 400074)
(wzybq123@163.com)

摘要:为解决交通预测的全局优化控制问题,在运用粗糙集理论对路网节点所测得的历史交通流量进行量化分析的基础上,建立了规则简化的数据清洗模型;基于分形理论的相似性原理建立了交通流量的预测模型,并给出了交通流优化控制系统的运行机制和相应的网络拓扑结构。用某交通观测站的实测数据对模型进行仿真,仿真结果表明,文中所设计的模型和算法是有效的。

关键词:交通流预测;优化控制;粗糙集理论;分形理论;相似性原理

中图分类号: TP391 文献标志码:A

Optimal control of traffic flow based on rough theory and fractal theory

WANG Jia-wei¹, HUANG Da-rong², LEI Ming²

(1. Computer and Information School, Chongqing Jiaotong University, Chongqing 400074, China;
2. Institute of Information and Computational Science, Chongqing Jiaotong University, Chongqing 400074, China)

Abstract: Accurate and timely prediction of the traffic flow in city is a key factor for traffic control and traffic-induced system. To solve the global optimal controlling problem in traffic flow prediction, we constructed a rule-simplified data-cleaning model by analyzing the real traffic flow based on rough theory, presented a traffic-flow forecasting model according to comparability in fractal theory and designed the corresponding traffic flow net topology and the running mechanism of traffic flow optimal control system. The simulation result shows that the technique is effective.

Key words: forecasting traffic flow; optimal control; rough theory; fractal theory; comparability

0 引言

近年来,随着智能交通系统(Intelligent Transportation System, ITS)的蓬勃发展,智能交通控制与诱导系统已成为ITS研究的热门课题。无论是交通控制还是交通诱导系统,实时准确的城市路网系统交通流量预测是这些系统实现的前提及关键,也是实现城市交通控制系统优化配置的必要步骤,其交通流量预测结果的好坏将直接影响到交通控制和诱导的效果。

鉴于此,找到一个实时、准确的交通流预测模型,便成了在智能交通系统中亟需解决的难点。为了寻求合适的预测模型,国内外学者从不同的角度进行了大量的研究和探讨:文献[1-4]基于卡尔曼滤波理论、模糊数学和混沌理论提出构造基准函数的方法,即从过去的最近一周内选择一个与当日最接近的曲线作为基准函数曲线,采用状态方程和观测方程组成的状态空间模型来描述滤波器结构,从而得出相应的交通流局部预测模型;文献[5-8]则基于神经网络模型和相空间重构理论分别研究了交通流量短期预测模型,均取得了较好的效果。但是这些拟合模型所解决的问题都是局部最优而不是全局最优,这对于日益复杂的城市交通路网系统来说具有极大的不安全性,为了克服这个问题,文献[9,10]结合小波理论和自回归滑动平均(Auto-Regressive and Moving Average Model, ARMA)以及分形理论研究了全局优化控制的交通流预测问题,取得了极大的成效。

然而,由于道路交通系统是一个有人参与的、时变的复杂系统,因此它具有高度的不确定性,这种不确定性不但有来自自然界方面的原因(如季节因素等),也有来自人为因素的原因(如交通事故、突发事件、司机的心理状态)。这些不确定干扰因素的存在给交通流量的预测尤其是短时交通流量预测带来了极大的困难。显然,建立包含所有干扰因素的交通流预测模型对于智能化程度越来越高的交通流优化控制的需求来说数据显得过于庞大,而且由于交通流信号具有不规则性和不对称性的特性,传统的交通流数据处理方法已经不能满足交通优化控制的需要,因此其准确性和实时性都有待提高。这样,如何在这些大量的影响因素中找到合理的指标体系,建立更加符合实际路况的交通流预测模型就成了当前需要着手解决的关键。因此,本文在前期研究的基础上,采用粗糙集理论对监测数据进行量化分析,简化了交通流优化控制的影响因素,然后根据路网节点所测得的历史交通流量,利用分形理论的相似性原理,研究了实时动态交通流的模型结构并给出了交通流优化控制方法。

1 交通流预测模型的建立和分析

由于交通流信号是交通控制中心实时监测得到的,其监测过程中受到各种因素的干扰,但并不是所有的因素对实际交通流量控制都有着至关重要的作用,需要从所监测网络节点的交通流信号中提取出有用的信息进行信息编码,才能进行有效的交通流量的优化控制。而上个世纪 60 年代发

收稿日期:2007-11-17;修回日期:2008-01-28。 基金项目:重庆市自然科学基金项目(CSTC,2006BB2422);重庆市教委科学技术研究项目(KJ060414,KJ070403);重庆市自然科学基金项目(KJ070403)。

作者简介:王家伟(1971-),男,四川宣汉人,讲师,硕士,主要研究方向:智能控制与数据挖掘; 黄大荣(1978-),男,湖北恩施人,副教授,博士,主要研究方向:随机控制理论、信息融合理论、故障预测; 雷鸣(1961-),男,重庆人,副教授,主要研究方向:应用数学。

展起来的粗糙集理论由于具有分析和处理不规则和不对称信号的能力,因此在我们的实验中借助粗糙集的相关理论对所测得的交通流信号进行分解处理,过滤掉其中的扰动信号;然后根据这些有用的信息建立分形预测模型来进行实时交通流的优化控制。

1.1 基于粗糙集理论的数据清洗模型

粗糙集理论是一种新的处理模糊和不确定知识的数学工具。目前,粗糙集理论已被成功地运用于机器学习、决策分析、过程控制、模式识别与数据挖掘等领域,其主要的思想就是在保持分类能力不变的前提下,通过知识约简,导出问题的决策或分类规则^[1]。这种思想表明,在交通流优化控制过程中所出现的庞大数据量面前,粗糙集理论在简化ITS变量,导出交通路网系统优化控制规则方面有很大的应用价值。

假设影响交通流监测数据的干扰因素有n个,记为 $X = \{X_1, X_2, \dots, X_n\}$,其实时监测数据为Y,因而对整个交通路网监测系统来说,可以把状态变量集合X看作是条件属性,实时监测数据Y看作是结论属性。由于在不同的时间段可以观测到不同的交通流量数据,假设观测的次数是m次,则在对每一个状态变量确定值域赋值后,将全部的观测结果用一个n×m阶矩阵 \bar{X} 表示,得到下面的数据矩阵:

$$\bar{X} = \begin{pmatrix} X_1 & X_2 & \cdots & X_n & Y \\ X_{11} & X_{21} & \cdots & X_{n1} & Y_1 \\ X_{12} & X_{22} & \cdots & X_{n2} & Y_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ X_{1m} & X_{2m} & \cdots & X_{nm} & Y_m \end{pmatrix} \quad (1)$$

显然,并不是所有的条件属性都对交通路网监测系统的正常运行重要,有些数据属性对交通流优化控制来说并不重要,甚至可以造成混淆视听的后果,这些数据属性应该从数据库中清除掉,从而保证不会因为数据属性的庞大而造成决策的失误。为了分析这个问题,首先引入几个定义^[1]:

定义1 设 $K = (U, R)$ 为一个知识库,定义 $ind(K)$ 为K中所有的等价关系的族,且 $ind(K) = \{ind(P) \mid \Phi \neq P \subseteq R\}$ 。

定义2 P 和 Q 为 U 的等价关系, Q 的 P 正域记为 $Pos_p(Q) = \bigcup_{x \in U/Q} PX$,也即 U 中所有根据分类 U/P 的信息可以准确地划分到关系 Q 的等价类中去的对象的集合。

定义3 给定一知识库 $K = (U, R)$,且 $P, Q \subseteq R$,当 $k = \gamma_P(Q) = |Pos_p(Q)| / |U|$ 时,称知识 Q 是 $k(0 \leq k \leq 1)$ 度依赖于知识 P 的。

定义4 假设 C 和 D 分别为条件属性集和决策属性集,属性子集 $C' \subseteq C$ 关于 D 的重要性定义为 $\sigma_{CD}(C') = \gamma_C(D) - \gamma_{C-C'}(D)$ 。

为了对干扰因素进行清洗,得到合理的能反应实际路况的交通流监测数据,需要从决策表中去掉一些属性。为此,我们采取如下的措施:去掉决策表中的某个属性,考查没有这个属性分类的变化是怎样的。如果去掉该属性后相应的分类变化很大,则说明这个数据属性的强度大,重要性高;反之,说明该属性的强度小,即重要性低。因而把依赖度 $k = \gamma_P(Q)$ 和重要性量度 $\sigma_{CD}(C') = \gamma_C(D) - \gamma_{C-C'}(D)$ 作为路网系统数据简化的量化指标。具体的操作步骤是对每一个路网交通流量影响因素设置一个重要容许值 $\sigma_{0i}(i = 1, 2, \dots, n)$,一旦计算得到的重要性 $\sigma_i(i = 1, 2, \dots, n)$ 低于这个估计值,就应该从数据库中删除这个数据,达到简化交通路网系统复杂度

的目的,得出相应的交通路网优化控制规则,利用这些规则就可以根据实时的监测数据来对路网交通流进行优化配置。

如此,就得到对建立合适的交通流预测模型有效的交通流信号。

1.2 基于分形理论的交通流预测模型算法和程序实现

为了实施交通流的优化控制,首先对所测得的交通流量观测值 $\{f(t), t = 1, 2, \dots, n\}$ 进行粗糙清洗处理,得到的新信号流 $\{(t, Y_t), t = 1, 2, \dots, n\}$ 。粗糙处理后得到的新的交通流量由于删除了其中的相对无关的干扰信号,保留了一些比较重要的干扰信号,因而可以用于建立更加符合实际情形的交通流预测模型。

同时,在分形理论中,一般利用如下的Hurst指数来判定时间序列是否具有分形特性。

$$R/S = CT^H \quad (2)$$

其中 R 为时间序列的极差, S 为标准差, T 为时间间隔, C 为常数, H 为Hurst指数,其定义如下:

$$EY_t = (1/T) \sum_{i=1}^n Y_i \quad (3)$$

$$X(i, T) = \sum_{t=1}^i [Y_t - EY_t] \quad (4)$$

$$R = \max X(i, T) - \min X(i, T) \quad (5)$$

$$S = \left(1/T \sum_{t=1}^n (Y_t - EY_t)^2 \right)^{1/2} \quad (6)$$

根据式(2)~(6)先估计出 R/S ,然后利用最小二乘法估计得到Hurst指数 H :如果满足 $H > 1/2$,则说明数据具有分形特性,因此可以利用分形理论建立交通流优化控制的预测模型,实现优化分配。

由于处理后的数据信号只给出了一些数据点 $\{(t, Y_t), t = 1, 2, \dots, n\}$,而函数关系则是未知的,因此交通流的数据关系的具体形式无法求出。在这种情况下,本文采用线性分形插值函数法来建立交通流预测模型,具体方法是:先将数据进行一系列变换,从中选出一种变换,使变换后的数据能与分形分布模型拟合程度达到一致。具体的建模过程如下:

1) 将原始数据点 $\{(t, Y_t), t = 1, 2, \dots, n\}$ 绘于双对数坐标上,如果它们不能与分形分布模型符合良好,于是可将 Y_t 排成一个基本序列,即有:

$$\{Y_t\} = \{Y_1, Y_2, Y_3, \dots\}; t = 1, 2, \dots, n \quad (7)$$

然后根据基本序列构造。例如构造一阶、二阶、三阶累计和序列,即:

$$\{s1_t\} = \{Y_1, Y_1 + Y_2, Y_1 + Y_2 + Y_3, \dots\}; t = 1, 2, \dots, n \quad (8)$$

$$\{s2_t\} = \{s1_1, s1_1 + s1_2, s1_1 + s1_2 + s1_3, \dots\}; t = 1, 2, \dots, n \quad (9)$$

$$\{s3_t\} = \{s2_1, s2_1 + s2_2, s2_1 + s2_2 + s2_3, \dots\}; t = 1, 2, \dots, n \quad (10)$$

2) 建立各阶累计和的分形模型。

具体方法是将数据点 $(sl_i, t_i)(l = 1, 2, 3; i = 1, 2, \dots, n)$ 绘于双对数坐标上,并根据分形模型(11)和(12)计算分维数 D 和常数 C :

$$sl_i = Ct_i^D \quad (11)$$

$$D = \ln(sl_i / sl_j) / \ln(t_j / t_i) \quad (12)$$

3) 将各阶累计和构成的数据点分别绘于双对数坐标上,然后和某一分形分布模型(本文应用的是由已知数据点构成的最后一段分形模型)进行对比,选择效果最好的变换并确

定其相应的分形参数。

4) 将累计和的预测结果再换算成 Y 值的预测结果,进而对交通流数据进行优化控制。

算法实现的基本流程如图 1 所示:

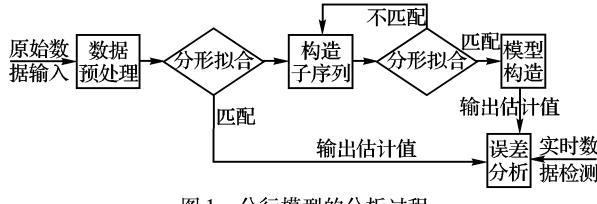


图 1 分形模型的分析过程

这样就确定了模型的系统参数 D 和 C ,因此可以根据式(11)进行实时交通的优化控制。

同时,由于模型预测结果的好坏影响着实时交通优化控制效果的好坏,故而选取 5 个指标来评价模型的效果。

平均绝对误差:

$$MAE = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t| \quad (13)$$

均方误差:

$$MSE = \frac{1}{n} \sqrt{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \quad (14)$$

平均绝对百分比误差:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \quad (15)$$

均方百分比误差:

$$MSP = \frac{1}{n} \sqrt{\sum_{t=1}^n \left(\frac{Y_t - \hat{Y}_t}{Y_t} \right)^2} \quad (16)$$

均等系数:

$$EC = 1 - \frac{\sqrt{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}}{\sqrt{\sum_{t=1}^n Y_t^2} + \sqrt{\sum_{t=1}^n \hat{Y}_t^2}} \quad (17)$$

2 交通流预测的算法设计和实现

在实际应用过程中,需要对交通流的信号给出实时优化配置,故而根据以上分析给出交通流优化控制系统的运行机制如下:

- 1) 系统初始化:装载历史交通流数据 $\{f(t), t = 1, 2, \dots, n\}$,存储在计算机中;
- 2) 对数据进行粗糙处理,输出新信号 $\{(t, Y_t), t = 1, 2, \dots, n\}$;
- 3) 通过式(2)~式(12)对粗糙处理的数据进行分形拟合,如果拟合程度合理,建立分形预测模型,直接进入 5),否则进入 4);
- 4) 构造累加和序列,对重构的序列进行分形拟合,如果合理,建立预测模型,进入 5),否则循环运行;
- 5) 利用历史数据预测未来的交通流量,进行误差分析,如果评价性能指标优秀,直接回到第 2),否则进行优化控制方案设计回到 2);
- 6) 输出结果,得到交通流优化控制的分析报告。

流程图描述如图 2 所示。

从图 2 的网络拓扑结构可以看出,这个系统可以简化为信号粗糙处理模块、分形拟合预测处理模块、评价模块、决策模块以及仿真模块,其中基础是信号粗糙处理模块和分形拟

合预测处理模块,核心是评价模块。显然,交通流优化控制系统中的所有的监测数据都可以返回到控制中心进行实时控制,能有效地进行交通流的优化控制。

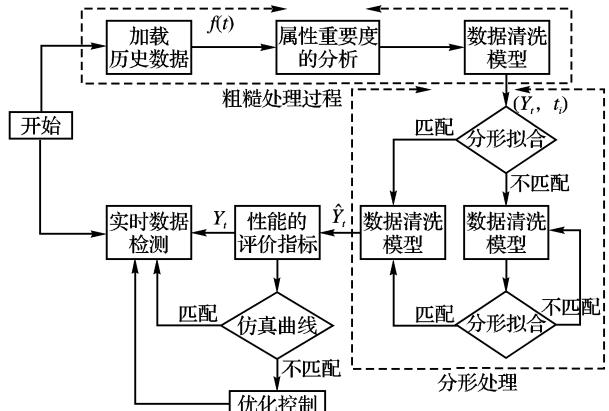


图 2 交通流预测优化控制的网络拓扑结构

3 仿真实验

本文把所设计的基于粗糙集理论和分形理论的交通流预测优化控制系统应用到实际的交通流优化控制中去。在 2005 年 7 月 18 日,在某高速公路检测站对交通流进行了实时监控,采取异常交通预测:异常交通的观测采用封闭一个车道的方式,车道封闭的时间为:8:00 ~ 8:10,9:30 ~ 9:40,11:00 ~ 11:10,12:30 ~ 12:40,14:00 ~ 14:10,15:30 ~ 15:40,17:00 ~ 17:10,18:30 ~ 18:40。

观测时间从早上 7:00 到晚上 8:025,检测时间间隔为 10 s,所测的数据交通流量数据为 780 个。其影响因素包括天气、车辆大小等 40 多种因素,我们用粗糙处理技术把数据信号最终的影响因素确定为 21 种。前 100 个数据进行粗糙处理后其原始数据和拟合数据的分布图分别如图 3 和图 4 所示(流量单位:辆/s)。

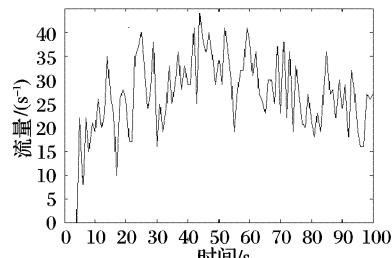


图 3 初始监测数据的分布图形(1 ~ 100)

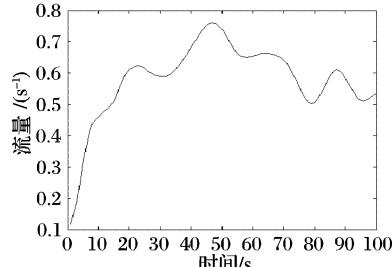


图 4 粗糙处理后的数据分布图(1 ~ 100)

根据粗糙处理后的数据计算得到 $H > 1/2$,这说明交通流数据具有分形特性,故而可建立分形预测模型。选取线性分形插入函数为 $Y_t = 0.1 + 0.8(data - \min)/(max - \min)$

进而对处理数据进行累加和序列运算,在双对数条件下取定阶数为 3,其指数分布图如图 5 所示。

这样利用这前 100 个交通流数据可以确定得到 $C =$

3.1744 , $D = -3.2544$, 从而可以根据式(11)进行预测, 所得到的预测结果和实际监测值的拟合曲线如图6所示(注意: 曲线中“+”表示预测值, “○”表示原始数据)。

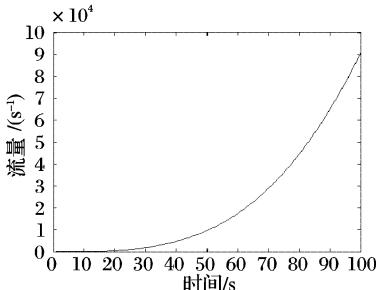


图5 3阶累加和序列的据分布图(1~100)

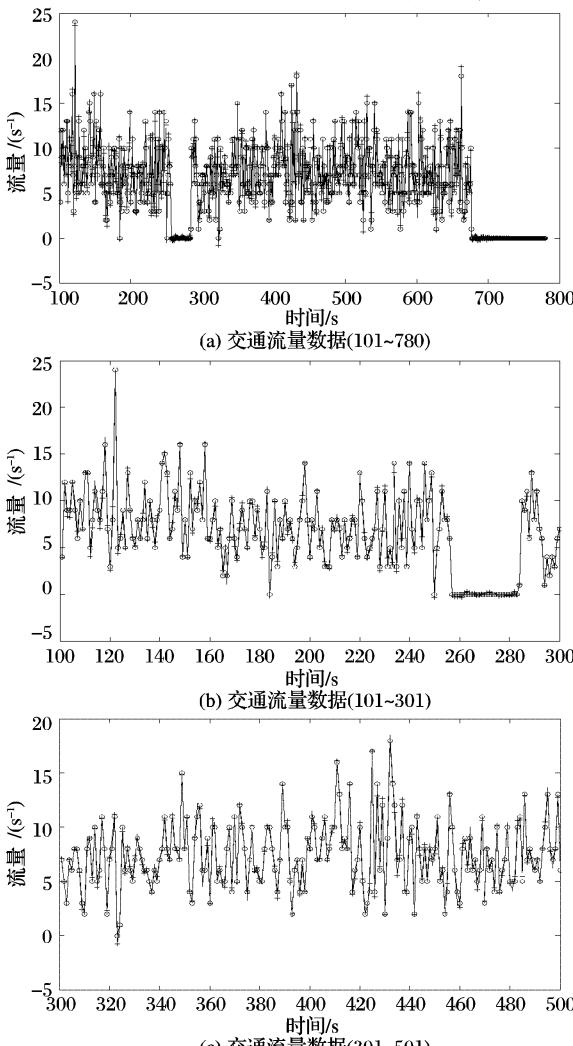


图6 数据预测值和实际监测值拟合曲线

图6(a)显示了所研究模型预测结果和实际监测结果的有效性, 图6(b)和图6(c)分别表示了部分数据点的拟合模型, 对前100个、500个以及检测的总共780个数据来说, 其评价指标分别为: $MAE_{100} = 3.0287$, $MSE_{100} = 42.7248$, $MAPE_{100} = 32.423$, $MSP_{100} = 2.4325$, $EC_{100} = 0.8565$; $MAE_{500} = 4.8036$, $MSE_{500} = 97.2343$, $MAPE_{500} = 34.505$, $MSP_{500} = 2.4354$, $EC_{500} = 0.8658$; $MAE_{780} = 6.4033$, $MSE_{780} = 162.8236$, $MAPE_{780} = 48.41$, $MSP_{780} = 2.2297$, $EC_{780} = 0.9243$ (评价指标含义分别是: MAE 为平均绝对误差, MSE 为均方误差, $MAPE$ 为平均绝对百分比误差, MSP 为均方百分比误差, EC 为均等系数)。实验分析结果表明, 我们

所建立的模型能进行不同状态的交通流数据的优化控制, 实现有效的资源配置。

4 结语

本文根据实验所测得的交通流量数据, 通过对流量数据序列进行粗糙处理, 结合分形理论建立了交通流时间序列分形预测模型, 并用实测数据进行预测, 仿真结果表明该分形预测模型能够较好地拟合交通流时间序列并可获得较高的中短期预测精度, 因而可用于动态交通信号控制。

同时, 粗糙集理论和分形理论相结合的交通流预测模型是根据路网节点过去和现在的交通流量, 运用相似性的原理科学地预测未来时段的交通流量, 要求的数据样本量不大, 并且有明确的收敛准则, 算法本身决定了这是一种全局寻优而不是局部搜索法, 因此准确性和实时性都比较高, 适合实时动态预测交通流。

参考文献:

- [1] JIANG Y. Dynamic prediction of traffic flow and congestion at freeway construction zones[J]. International Journal of Construction Education, Spring 2002, 7(1): 45–57.
- [2] KIRSCHFINK H, CHADENAS C. Traffic situation prediction applying pattern matching and fuzzy classification [EB/OL]. [2007-10-30]. http://www.erudit.de/erudit/events/esit99/12895_p.pdf.
- [3] HU JIAN-MING, SONG JING-YAN, ZHANG YI, et al. Modeling and analysis for self-organization of urban traffic flow[C]// Proceedings of the 2005 IEEE International Conference on Intelligent Transportation Systems. Washington, DC: IEEE Computer Society, 2005: 237–242.
- [4] HU JIAN-MING, ZONG CHUN-GUANG, SONG JING-YAN, et al. An applicable short-term traffic flow forecasting method based on chaotic theory[C]// Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems. Washington, DC: IEEE Computer Society, 2003: 608–613.
- [5] HU JIAN-MING, SONG JING-YAN, YU GUO-QIANG, et al. A novel networked traffic parameter forecasting method based on Markov chain model[C]// Proceedings of the 2003 IEEE International Conference on Systems, Man and Cybernetics. Washington, DC: IEEE Computer Society, 2003: 3595–3600.
- [6] HIPPERT H S, PEREIRA C E, SOUZA R C. Neural networks for short-term load forecasting: A Review and evaluation[J]. IEEE Transactions on Power Systems, 2001, 16(1): 44–55.
- [7] DOULAMIS A D, DOULAMIS N D, KOLLIAS S D. An adaptable neural-network model for recursive nonlinear traffic prediction and modeling of MPEG video sources[J]. IEEE Transactions on Neural Network, 2003, 14(1): 150–166.
- [8] LEE E B, KIM C, Harvey J C, et al. Implementation of automated travel-time information and public reaction on urban highway rehabilitation[J]. Journal of Transportation Engineering, 2006, 132(10): 808–816.
- [9] HUANG DA-RONG, WANG DA-CHENG, SONG JUN, et al. Forecasting model of traffic flow based wavelet transform and fractal theory[C]. Proceeding of International Conference on Sensing, Computing and Automation (ICSCA'06). WATERLOO, CANADA: Watam Press 2006: 494–498.
- [10] HUANG DA-RONG, WANG DA-CHENG, SONG JUN, et al. Forecasting model of traffic flow based ARMA and wavelet transform [C]// Proceeding of the 3rd International Conference on Impulsive Dynamical Systems and Applications (ICIDSA'06). Qingdao: [s. n.], 2006: 860–869.
- [11] PAWK Z. Rough set theory and its applications to data analysis[J]. Cybernetics and systems, 1998, 29(7): 661–668.