Applied Mathematics & Information Sciences
An International Journal

Public Bicycle Traffic Flow Prediction based on a Hybrid Model

Haitao Xu[1,2], *Jing Ying*[2], *Hao Wu*[1] *and Fei Lin*[1]

[1]School of Software Engineering, Hangzhou Dianzi University, 310018 Hangzhou, P.R.China
[2]College of Computer Science and Technology, Zhejiang University, 310012 Hangzhou, P.R.China

Received: 7 Aug. 2012; Revised 28 Nov. 2012; Accepted 23 Dec. 2012
Published online: 1 Mar. 2013

Abstract: In China, Hangzhou is the first city to set up the Public Bicycle System. Now, the system has been the largest bike sharing program in the world. The software of Hangzhou Public Bicycle System was developed by our team. Accurate and precise prediction of public bicycle traffic flow is important in traffic planning, design, operations, etc. According to the highly complexity, nonlinearity and uncertainty of traffic flow, a single prediction model is difficult to ensure the prediction accuracy and efficiency. To overcome the lack of the single prediction method, this paper uses a hybrid model that combining clustering with support vector machine, by exploiting complementary advantages of both approaches. Firstly, this method uses improved k-means algorithm to cluster the original sample set. Secondly, the subset whose character is the most similar to the sample set to be forecasted is chosen. Finally, a polynomial smooth support vector machine uses the subset to forecast the public bicycle traffic flow. The experimental results show that the hybrid model performs higher forecasting accuracy and better generalization ability.

Keywords: Public bicycle system, clustering, k-means algorithm, data mining, Support Vector Machine(SVM)

1. Introduction

With many cities, traffic congestion is a major problem of public transport in Hangzhou. More and more private cars will lead to a big traffic problem and to solve road congestion more difficult. Private bicycles are difficult to be managed and will lead to secure traffic safety. ”Too many private cars, bicycles too chaotic” is traffic problems.

A free public bicycle system, as a part of the public transport, the original intention to promote public bike is to solve the ”last mile” problem. It is ”Too crowd bus ride, too expensive taxi, too far to walk”, through the ”Bicycle-Bus-Bicycle” convenient destination, while promoting the city’s energy reduction of carbon emissions. In China, Hang zhou is the first city to set up the Public Bicycle System. Now, Hangzhou city public bike has covered near 3000 service points, a total of about 60000 bicycles. The Hangzhou Public Bicycle System has surpassed Velib as the largest bike sharing program in the world. Anyone of over the age of 16 and under 70 are eligible for bicycle rental. However, no cash is accepted at the rental service locations.

Because Hangzhou public bicycle is unattended, sometimes it is a common problem to find no bicycle to rent or no place to return at some stations. Hangzhou Government is determined to solve this problem. If we can forecast the public bicycle traffic flow, we can take measures in advance. Also accurate and precise prediction of public bicycle traffic flow is important in traffic planning, design, operations, etc. In this paper, we will introduce a hybrid model to predict the public bicycle traffic flow.

Hangzhou public bicycle system is the largest bike sharing program in the world. Within the limit of our knowledge, the first mention of the intelligent public bicycle traffic flow prediction system appeared in this paper. The system used a new hybrid model. The outline of this paper is as follows. Section 2 reviews the related work of motor vehicle traffic flow prediction. Section 3 outlines the architecture of Hangzhou Intelligent Public Bicycle Traffic Flow Prediction System based on a hybrid model. Section 4, Section 5 and Section 6 respectively describe the three components of the hybrid model. Section 7 discusses the experiments using the hybrid model. Conclusions are given in Section 8.

* Corresponding author: e-mail: xuhaitaohdu@163.com

2. Related Work

Within the limit of our knowledge, there is very little research on the public bicycle traffic flow so far, because public bicycles are an emerging transportation tool. But researchers have built many prediction models of motor vehicle traffic flow. By now there are approximately 30 prediction methods[1]: the dynamic traffic flow distribution methods, the historical-mean methods, the regression analysis methods, the time series methods, the Kalman filtering methods, the neural network methods, the fuzzy neural network method, the fuzzy-neural method, the nonparametric regression methods, the support vector machine(SVM) methods, etc.

Although these methods have alleviated difficulties in traffic modeling and prediction to some extent, from a careful review we can still find some problems. the historical mean methods [2], [3] and regression analysis methods [4]have common drawbacks, which suppose that traffic flow and travel time are both strictly periodic and ignore the uncertainty and nonlinearity of traffic flow. The Kalman filtering methods [2]-[6] are unsuitable for predicting the traffic flow which sample interval is less than 5 min. The nonparametric regression methods [1] need a huge historical database which occupancies many memory and takes much time to predict the traffic flow. The neural network methods suffer from problems like the existence of local minima and the limited generalization ability. In order to improve prediction performance, the hybrid model combined with wavelet analysis and neural network is proposed, but accompanied with low efficiency due to the inherent theory flaw from neural networks. SVM is a relatively new machine learning technique which is used for classification and regression purposes. The fact that SVM has better generalization ability from limited samples than the traditional techniques triggered exploring this technique for short term prediction of traffic parameters. Studies have reported the use of SVM for traffic flow forecasting. When the size of the data set is large, traditional SVM tend to perform worse when trained with the entire data than with a set of fine-quality samples [7].

Based on these insights, we present a novel hybrid prediction model that combines k-means clustering with SVM.

3. Overview

As shown in Figure 1, the architecture of Hangzhou Intelligent Public Bicycle Traffic Flow Prediction System based on a hybrid model consists of three major components: Normalization Processing, K-means Clustering, and SVM Predictor.

- Normalization Processing: For the convenience of data set processing and the acceleration of program convergence, normalization processing is needed.
- K-means Clustering: Before training the support vector machine, we do a preprocessing on the training data set using k-means clustering algorithm. The preprocessing can reduce the time of constructing support vector. Clustering can divide data set into multiple clusters and strengthen data regularity to improve the accuracy of the system. In this component, we employ our improved k-means clustering algorithm to solve the problems that the traditional k-means algorithm has sensitivity to the initial cluster centers.
- SVM Predictor: SVM will be change the actual nonlinear problem from a nonlinear to a high dimensional feature space, ingeniously solved the problem of dimensionality. In this component, we employ our sixth order polynomial smoothing support vector machine [8].

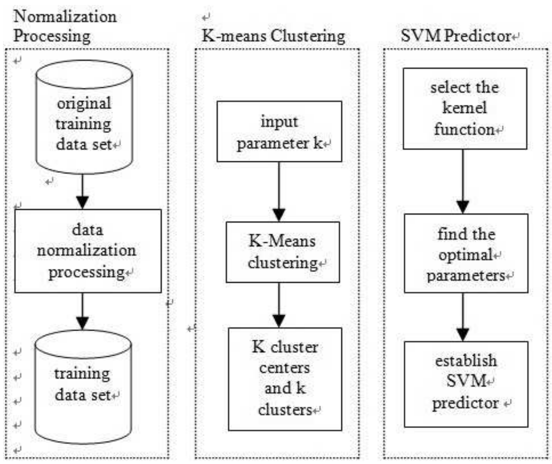

Figure 1 System overview

4. Normalization Processing

We observe from the historical public bicycle rent-return records) that the public bicycle traffic flow has certain regularity. One day's traffic flow data is well correlated with the historical data. It links to 3 days before prediction day and the same time 2 weeks before prediction day. Let P(d,h) is the prediction value, T(d,h) is the actual traffic flow value. T(d-1,h) is the actual value of the day before prediction day at the same time. T(d-2,h) is the actual value of two days before prediction day at the same time. T(d-3,h) is the actual value of three days before prediction day at the same time. T(d-7,h) is the actual value of one week before prediction day at the same time. T(d-14,h) is the actual value of two weeks before prediction day at the same time.

Based on the variation regularity of some special holiday, we add dimension of week and dimension of holidays. The weather is influential in public bicycle traffic flow, so the dimension of weather is added. S(d) represents the special holiday. W(d) represents the weather of prediction day.

The dimension of week is coded as (W3(d),W2(d),W1(d)). For example, (W3(d),W2(d),W1(d)) =(0,0,1) means that the prediction day is Monday.

The problem that prediction system has to solve is to find a optimal function that satisfies P(d,h)=f(W(d),T(d-1,h), T(d-2,h),T(d-3,h),T(d-7,h),T(d-14,h),W3(d),W2(d),W1(d),S(d)). To find the function f, we use k-means clustering algorithm to preprocess and SVM to train the data set.

5. K-means Clustering

5.1. Brief review of k-means clustering

Clustering analysis of data set aims at discovering smaller, more homogeneous groups from a large heterogeneous collection of data points and it is an important unsupervised classification technique used in identifying some inherent structure present in a set of objects.

In some circumstances, the number of clusters, the parameter k, is known as a priori, and clustering may be formulated as distributing m patterns in n-dimensional space among k sets such that the patterns in one set are more similar to each other than to patterns in different sets. This involves minimization of some extrinsic optimization criterion. Agglomerative algorithms, k-means algorithm, fuzzy algorithms, BIRCH and CLARANS are a few of the existing clustering methods.

Among of them, the k-means algorithm is the most basically and widely used one for clustering. Random procedures are used to generate starting clustering centers at the beginning of the k-means algorithm. However, it is known and also can be found from the experiments presented in this paper that the efficiency of the k-means algorithm largely depends on the choice of the clustering centers (Boris Mirkin[10][11][12] has presented this opinion and proposed some intuitions for selection of clustering centers, such as MaxMin for producing deviate centroids, deviate centroids with anomalous pattern, intelligent k-means and so on). In 2004, Shehroz S. Khan and Amir Ahmad [13] also presented that performance of iterative clustering algorithms which would converge to numerous local minima depended highly on Max-Min clustering centers and proposed a clustering center Max-Min algorithm(named CCIA). Their results showed the proposed algorithm could achieve better performance. Also, in 1998, Paul S. Bradley and Usama M. Fayyad [14] had proved that the better Max-Min starting points indeed could lead to improved solutions for clustering problems. In order to improve performance of the k-means method for data clustering, a better centers selection algorithm is proposed in this paper. The idea comes from partition technology according to data distribution. Before k-means algorithm is made, some features of data set for clustering are analyzed, then, the beginning clusters for k-means algorithm are obtained.

5.2. Improved k-means clustering algorithm

Here, we proposed an improved k-means clustering algorithm to do it.

Clustering in n dimensional Euclidean space Rn is the process of partitioning a given set of m points into a number of groups (or, clusters) based on some similarities (or dissimilarities). The similarity establishes a rule for assigning patterns (points) to the domain of a particular cluster center. Let the set of m points be S = x1,x2,o o o,xm with xi being an n-dimensional vector, and k clusters be represented by C1,C2, o o o,Ck . The basic model of describing the clustering problem is given by (can be seen in [7])

$$\begin{cases} \bigcup_{i=1}^{k} C_i = S; \\ C_i \bigcap C_j = \phi,\ i \neq j,\ i,j = 1,2,\cdots,k; \\ C_i \neq \phi, \qquad i = 1,2,\cdots,k. \end{cases} \tag{1}$$

The procedure of finding the k optimal clusters $C_1, C_2, \cdots, C_k$ is equivalent to find k clustering centers, denoted as $\{z_1, z_2, \cdots, z_k\}$. For the swatch set of m points $S = \{x_1, x_2, \cdots, x_m\}$, cluster C_i is determined as follows

$$\begin{aligned} C_i = \{x_j \mid \|x_j - z_i\| \leq \|x_j - z_p\|, \\ p \neq i,\ p = 1,2,\cdots,k,\ x_j \in S\} \end{aligned} \tag{2}$$

where $\|\cdot\|$ is some norm in R^n, that is, C_i is the set of the points that are the closest to the cluster center z_i.

Therefore, the clustering problem is to find k clustering centers $\{z_1, z_2, \cdots, z_k\}$ such that the sum of the distances of each point in the set S to one point in $\{z_1, z_2, \cdots, z_k\}$ is minimized, that is, $\{z_1, z_2, \cdots, z_k\}$ is the solution of the following optimization problem

$$\min_{z_1, z_2, \cdots, z_k} \sum_{j=1}^{n} \min_{1 \leq p \leq k} \|x_j - z_p\|. \tag{3}$$

The objective function in (3) is in general neither convex nor concave, and hence it could be difficult to find the solution by solving the problem. However, based on Lemma 3.1 in [8], problem (3) can be reformulated into the following constrained optimization problem

$$\min_{Z,t} \sum_{j=1}^{n} \sum_{p=1}^{k} t_{jp} \|x_j - z_p\|, \tag{4}$$

$$\text{s.t.} \sum_{p=1}^{k} t_{jp} = 1, t_{jp} \geq 0,, j = 1,2,\cdots,n, p = 1,2,\cdots,k,$$

where $t_{j\bar{p}} = 1$ if $z_{\bar{p}}$ is the closet center to x_j, and $t_{jp} = 0$ for $p = 1,2,\cdots,k,\ p \neq \bar{p}$. If multiple centers have the same minimum distance to x_j, then t_{jp} can be nonzero between x_j and these clustering centers, and form a convex combination of this minimum distance.

Usually, in problem (4), if we employ ℓ_2-norm, the following optimal problem is obtained

$$\min_{Z,t} \quad f(z,t) = \sum_{j=1}^{n}\sum_{p=1}^{k} t_{jp}(\frac{1}{2}\|x_j - z_p\|_2^2), \tag{5}$$

$$\text{s.t.} \quad \sum_{p=1}^{k} t_{jp} = 1, t_{jp} \geq 0,, j = 1,2,\cdots,n, p = 1,2,\cdots,k,$$

and $k-$Means algorithm is one of the widely used clustering techniques for (5). The k-means algorithm is an iterative descent method and can be described as follows:

The $k-$Means Algorithm.
Step1: Generate k initial clustering centers $z_1, z_2, \cdots, z_k$.
Step2: Cluster Assignment: Assign point $x_j, j = 1,2,\cdots,n$, to clusters C_i
with centers $z_i, i = 1,2,\cdots,k$;
Step3: Update clustering centers $z_i^* = \frac{1}{|C_i|}\sum_{x_j \in C_i} x_j$;
Step4: If $z_i^* = z_i, \forall i = 1,2,\cdots,k$ terminate, else $z_i = z_i^*$ and go to step 2;

The $k-$means algorithm generally works well. However, random procedures are used to generate initial clustering centers at the beginning of the traditional k-means algorithm and the algorithm has sensitivity to the initial cluster centers. To solve this problem, the following algorithm is proposed to optimize the initial centers based on the minimum spanning tree. The algorithm partition data points into K initial cluster, and calculate the initial cluster centers.

New Initial Centers Algorithm:.
Input: Data set S which containing n samples, the number of clusters K.
Output: K initial cluster centers.
Step1: Calculate the distance of two data points x_i and $x_j (1 \leq i \leq n, 1 \leq j \leq n)$;
Get the distance matrix D according to dist[x_i, x_j];
Calculate the sum of distance between x_i and other data point;
Calculate the average distance of data set S ;
Step2: S_1=null; For i = 1 to n do if sum[x_i] ¡ avg[S] then S_1=S_1+x_i ;
Step3: Get the distance matrix D_1 according to S_1;
Step4: T=null; flag[1]=True; For i = 2 to n do flag[i]=false;
Step5: Repeat search out the edge e with the minimum dist[x_i, x_j] which connecting point x_i (flag[i]=false) and point x_j (flag[j]=true); T=T+e; flag[i]=true;
Until length(T)==n;
Step6: Split T into K subtrees according to distance descending;
Step7: Calculate average value of data points in each subtrees and get K cluster centers.

6. SVM Predictor

6.1. SVM

SVM is a new statistical learning technique that can be seen as a new method for training classifiers based on polynomial functions, radial basis functions, neural networks, splines or other functions. Mathematically, SVM is a pattern classification problem based on a given classification of m points in the $n-$dimensional space R_n, represented by an $m \times n$ matrix A, given the membership of each data point $A_i, i = 1,2,\cdots,m$ in the classes 1 or -1 as specified by a given $m \times m$ diagonal matrix D with 1 or -1 diagonals.

This problem is given by the following model

$$\begin{aligned} \min_{(\omega,\gamma)\in R^{n+1}} \ & \tfrac{1}{2}\omega^T\omega, \\ s.t. \ & D(A\omega - e\gamma) \geq e. \end{aligned} \tag{6}$$

Model (6) can be seen as the original model of SVM. ω is a vector of separator coefficients (direction vector of classification hyperplane), γ is an offset (the control parameter of the distance of hyperplane plane to the origin) and $e \in R^m$ stands for a vector of ones.

The linear separating hyperplane

$$P = \{x | x \in R^n, x^T\omega = \gamma\}, \tag{7}$$

with normal $\omega \in R^n$ and distance $\frac{|\gamma|}{\|\omega\|_2}$ to the origin.

Mathematically, the model (6) is a quadratical programming with linear inequalities constraints. The most important thing is how to get the optimal solution of (1). One solution method comes from duality theory. Till now, most proposed solution methods are based on the dual method by Lagrange multiplier[15][16].

Another method comes from the approximation solution theory. In 2001[17], Lee formulated (6) into a nonsmooth unconstrained optimization problem. They employed smoothing method to smoothen their proposed model. Their selected function is the integral of the sigmoid function of neural networks. In 2005, Y.Yuan et proposed two classes of polynomial functions[18][19]. In 2007, the spline function was introduced to smoothen the plus function[20]. In this paper,we use an arc smoothing function to smoothen it.

The original model of support vector machine (6) is a special quadratic programming with linear inequalities constraints. In order to solve it, the constraints should be moved away from the optimal model. A slack variable $y \in R^m$ is introduced to do it. With it, the primal model (6) of SVM can be reformulated as following with norm 2

$$\begin{aligned} \min_{(\omega,\gamma,y)\in R^{n+1+m}} \ & \frac{\nu}{2}\|y\|_2^2 + \frac{1}{2}\|\omega\|_2^2, \\ \text{s.t.} \ & D(A\omega - e\gamma) + y \geq e, \\ & y \geq 0. \end{aligned} \tag{8}$$

As a feasible solution of problem (8), y is given by

$$y = (e - (D(A\omega - e\gamma)))_+, \tag{9}$$

where the elements of the vector $(a)_+$ is defined by

$$(a_i)_+ = \begin{cases} a_i, \text{ if } a_i > 0, \\ 0, \text{ if } a_i \leq 0. \end{cases} \quad (10)$$

Substituting y into the objective function of (8) converts problem (8) into an equivalent unconstrained optimization problem

$$\min_{(\omega,\gamma)\in R^{n+1}} \frac{v}{2}\|(e-(D(A\omega-e\gamma)))_+\|_2^2 + \frac{1}{2}\|\omega\|_2^2. \quad (11)$$

This is a strongly convex minimization problem without any constraints and exists a unique solution. However, the objective function in (11) is not differentiable at zero which precludes the use of existing optimization methods using derivatives. In the next section we will introduce a smoothing function with parameter k to smoothen the objective function.

6.2. Sixth Order Polynomial Smoothing Function

In this section, we present smoothing functions to smoothen the first term of (11). Obviously, the plus function is not differentiable at zero. The non-differentiability of $(x)_+$ can be moved away by the smoothing function.

In this paper, we proposed a new smoothing function as following

$$cp(x,k,M) = \begin{cases} x, \text{if } x > \frac{1}{k}; \\ P_6(x,k,M), \text{if } -\frac{1}{k} \leq x \leq \frac{1}{k}; \\ 0, \text{if } x < -\frac{1}{k}. \end{cases} \quad (12)$$

where

$$P_6(x,k,M) = -\frac{k^5}{9M}(5k-4M)x^6 + \frac{k^3}{36M}(68k-49M)x^4 \\ -\frac{k}{36M}(88k-47M)x^2 + \frac{1}{2}x + \frac{1}{M}.$$

Theorem 6.1. If the smoothing function have the formulation as (12), then

i) For any given $x \in R$ and $k \in Z^+, M \in R^+$, $cp(x,k,M)$ is continuous;

ii) $cp(x,k,M)$ is differentiable for any $x \in R$;

iii)For any given $x \in R$ and $k \in Z^+, M \in R^+$, we have

$$cp(x,k,M)^2 - x_+^2 \leq Q_{max}(k,M). \quad (13)$$

Proof. Since

$$\lim_{x\to \frac{1}{k}^-} cp(x,k,M) = \frac{1}{k}, \lim_{x\to(-\frac{1}{k})^+} cp(x,k,M) = 0,$$

and

$$(\frac{d(cp(x,k,M))}{dx})|_{x=-\frac{1}{k}} = 0, (\frac{d(cp(x,k,M))}{dx})|_{x=\frac{1}{k}} = 1.$$

The smoothing function $cp(x,k,M)$ is first-order continuously differentiable for any $x \in R$. i) and ii) hold.

From the definition (12), the result $cp(x,k,M) \geq x_+$ is obviously.

For $x \in [-\frac{1}{k}, \frac{1}{k}]$, we use the method of finding maximize function,

$$\max_{x\in[-\frac{1}{k},\frac{1}{k}]} (cp(x,k,M)^2 - x^2).$$

If the $cp(x,k,M)$ is replaced by (12), then we have

$$\max_{x\in[-\frac{1}{k},\frac{1}{k}]} Q(x,k,M). \quad (14)$$

where

$$Q(x,k,M) = -C_1(k,M)(kx)^6 + C_2(k,M)(kx)^4 \\ -C_3(k,M)(kx)^2 + \frac{1}{2}x + \frac{1}{M},$$

and

$$C_1(k,M) = (\frac{5}{9M} - \frac{4}{9k}), \\ C_2(k,M) = (\frac{68}{36M} - \frac{49}{36k}), \\ C_3(k,M) = (\frac{88}{36M} - \frac{47}{36k} + \frac{1}{k^2}).$$

Actually, the problem (14) is a basic optimal one. The objective function of it is continuous on the interval $[-\frac{1}{k}, \frac{1}{k}]$. According to the min-max value existence theorem. The maximal must can be checked out. Here, we denote the maximal of objective function on the interval $[-\frac{1}{k}, \frac{1}{k}]$ as $Q_{max}(k,M)$.

Figure 2 shows that the maximal distance between plus and one control parameter smoothing function can be controlled by the parameter M.The top blue solid line is the smoothing function with $M = k$. The other blue lines are the smoothing functions with $M = k*N$ in turn from top to bottom. Doted lines are the quadratic, forth polynomial and plus functions.

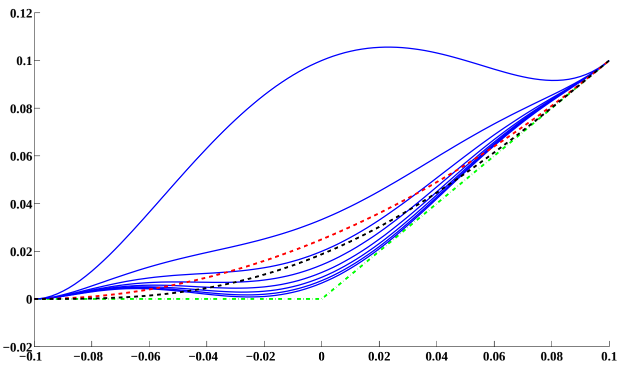

Figure 2 Maximal distance between plus and smoothing function

If we select an enough large number M, we can make sure that

$$Q_{max}(k,M) \leq \frac{1}{20k^2}.$$

iii) holds. The proof of Theorem 6.1 is ended.

6.3. Solution of the Proposed Model

If we employ the sixth order smoothing function in (12), the following unconstrained optimal model of SVM can be obtained

$$\min_{(\omega,\gamma)\in R^{n+1}} \frac{\nu}{2}\|cp((e-(D(A\omega-e\gamma))),k)\|_2^2+\frac{1}{2}\|\omega\|_2^2. \quad (15)$$

Theorem 6.2. There exists a unique solution x^* of (11) and a unique solution $x(k)^*$ of (15).

Proof. Let

$$f_1(x)=\frac{\nu}{2}\|(e-(D(A\omega-e\gamma)))_+\|_2^2+\frac{1}{2}\|\omega\|_2^2,$$

$$f_2(x,k,M)=\frac{\nu}{2}\|cp((e-(D(A\omega-e\gamma))),k,M)\|_2^2+\frac{1}{2}\|\omega\|_2^2.$$

It is obvious that $f_1(x)$ and $f_2(x,k,M)$ are strongly convex functions. The optimal solutions of (11) and (15) are existed.

Since $cp(x,k,M)\geq(x)_+$, the level sets

$$L_S(f_2(x,k,M))=\{x|x\in R^n, f_2(x,k,M)\leq S\},$$

$$L_S(f_1(x))=\{x|x\in R^n, f(x)\leq S\}$$

satisfy

$$L_S(f_2(x,k,M))\subset L_S(f_1(x))\subset\{x|\|x\|_2^2\leq 2S\}, \quad (16)$$

for $S\geq 0$. Hence $L_S(f_2(x,k,M))$ and $L_S(f_1(x))$ are compact subsets in R^n. The uniqueness of these solutions comes from the strong convexity of functions $f_1(x)$ and $f_2(x,k,M)$ for all $k\in Z^+$.

Theorem 6.3. For any $k\in Z^+$, we have the following inequality

$$\|x(k)^*-x^*\|_2^2\leq\frac{m}{2Q_{max}(k,M)}\leq\frac{1}{40k^2}m. \quad (17)$$

and $\lim_{k\to\infty}x(k)^*=x^*$.

Proof. It follows from the strong convexity of $f_1(x)$ and $f_2(x,k,M)$ that we have

$$f_1(x(k)^*)-f_1(x^*)\geq\nabla f_1(x^*)(x(k)^*-x^*)+\frac{1}{2}\|(x(k)^*-x^*)\|_2^2, \quad (18)$$

and

$$f_2(x^*,k,M)-f_2(x(k)^*,k,M)\geq \nabla f_2(x(k)^*,k,M)(x^*-x(k)^*)+\frac{1}{2}\|(x^*-x(k)^*)\|_2^2.$$

Since the $cp(x,k,M)\geq x_+$ we have

$$f_2(x,k,M)\geq f_1(x)\geq 0$$

for all $k\in Z^+$. Then, we have

$$\begin{aligned}\|x(k)^*-x^*\|_2^2 &\leq (f_2(x^*,k,M)-f_1(x^*))- \\ &\quad (f_2(x(k)^*,k,M)-f_1(x(k)^*)) \\ &\leq (f_2(x^*,k,M)-f_1(x^*))\end{aligned} \quad (19)$$

It follows from iii) of Theorem 6.1 that

$$\|x(k)^*-x^*\|_2^2\leq\frac{m}{2Q_{max}(k,M)}\leq\frac{1}{40k^2}m.$$

Based on the above inequality, it is easy to verify that $\lim_{k\to\infty}x(k)^*=x^*$. □

Theorem 6.3 shows that the optimal solution of SVM model (6) can be obtained by successively solving problem (15). So we use this SVM to predict public bicycle traffic flow.

7. Practical Application

The software of Hangzhou Public Bicycle System was developed by our team. We get all of the rent-return records in half a year. Actually, from 1th July 2011 to 31th December 2011. We selected 20% randomly from the actual historical records as the modeling data. The prediction result is compared with the actual result and other methods.

At first, we need to determine the weights for each dimension before carrying out clustering. In our application, weight coefficients of all dimensions are shown in table 7.1. It can be seen from the table, the biggest weight is the weather, the day before prediction day and one week before prediction day at the same time.

Table 1 weight coefficients of all dimensions

Dimension	Weight	Dimension	Weight
W(d)	1.000	T(d-14,h)	0.802
T(d-1,h)	1.000	W3(d)	0.330
T(d-2,h)	0.913	W2(d)	0.272
T(d-3,h)	0.783	W1(d)	0.272
T(d-7,h)	0.990	S(d)	0.504

Secondly, we do a preprocessing on the training data set using $k-$means clustering algorithm before training the support vector machine. Through experiments, we found that our application get the best results when K = 24. In this case, we get 24 clustering centers.

Finally, we use the sixth order polynomial smoothing support vector machine to predict the public bicycle traffic flow data.

7.1. Comparison with Actual Value

The prediction results of the intelligent public bicycle traffic flow prediction system based on above methods show as follows. We compared the prediction value with the actual value. The comparison results are shown in Figure 3, Figure 4 and Figure 5.

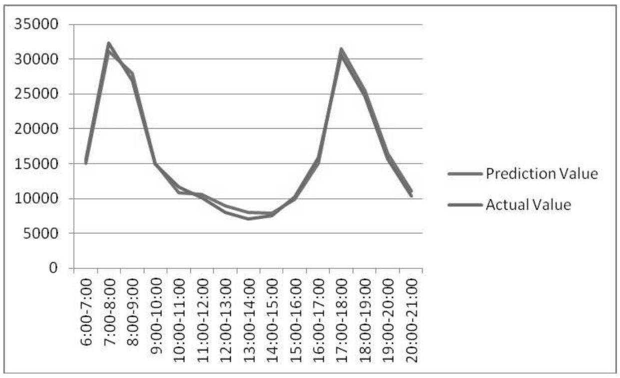

Figure 3 April 2, 2012(Monday) predictions

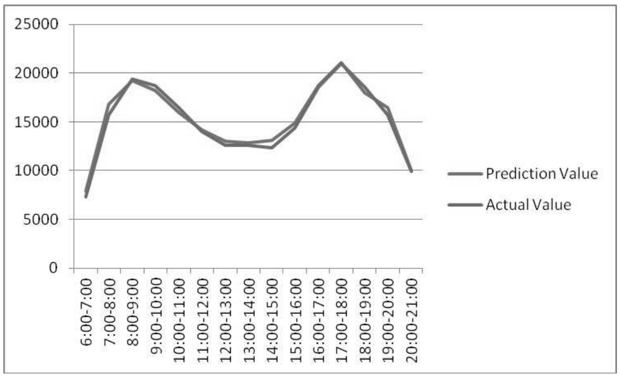

Figure 4 April 7, 2012(Saturday) predictions

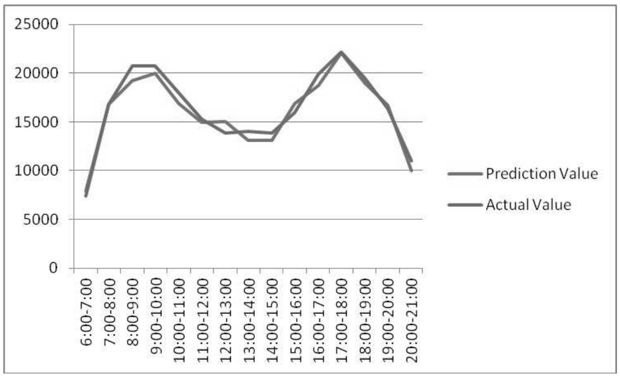

Figure 5 May 1, 2012(May Day) predictions.

7.2. Comparison with other methods

Using the same data and the environment, we have established the BP neural network model and the pure SVM model. The BP neural network has three-layer structure. Hidden layer has 12 nodes and use the S-type function. The input layer has 10 input. The pure SVM model only use the SVM prescribed in section VI. The comparison results of these three methods are shown in Table 2.

Table 2 Error Rate Comparison

Date	BP Error Rate	SVM Error Rate	Our Hybrid Model Error Rate
April 3, 2012(Tuesday)	8.23%	5.17%	3.57%
T(d-1,h)	1.000	W3(d)	0.330
T(d-2,h)	0.913	W2(d)	0.272
T(d-3,h)	0.783	W1(d)	0.272
T(d-7,h)	0.990	S(d)	0.504

8. Conclusion

This paper presents a hybrid model for public bicycle traffic flow prediction. In this model, there are three major components: Normalization Processing, K-means Clustering, and SVM Predictor. For the convenience of data set processing and the acceleration of program convergence, normalization processing is needed. Then the hybrid model uses our improved K-means algorithm to cluster the original sample set. Finally, we proposed a sixth order polynomial smooth support vector machine. It is used to forecast the public bicycle traffic flow. Experimental results are presented to show the effectiveness of the proposed model. Now, an Intelligent Public Bicycle Traffic Flow Prediction System based on the hybrid model has been applied in Hangzhou.

Acknowledgement

This work was financially supported by Research Project of Zhejiang Provincial Education Department (Y201120473), Science Foundation of Hangzhou Dianzi University (KYF215610034) and Zhejiang Provincial Natural Science Foundation of China (LY12F02017).
The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

References

[1] Dehuai Zeng, Jianmin Xu, Jianwei Gu , Liyan Liu , and Gang Xu, Short Term Traffic Flow Prediction Based on Online Learning SVR, Workshop on Power Electronics and Intelligent Transportation System, Aug , 616-620 (2008).
[2] Liang Zhao, Fei-Yue Wang, Short-term traffic flow prediction based on ratio-median lengths of intervals two-factors high-order fuzzy time series, Proc. of Conf. on Vehicular Electronics and Safety, Dec , 1-7 (2007).

[3] Smith. B. L. and Demetsky. M. J., Traffic flow forecasting: comparison of modeling approaches, Journal of Transportation Engineering, vol. 123, **4**, 261-266 (1997).
[4] Kaysi. I., Ben-Akiva. M. and Koutsopoulos. H., An integrated approach to vehicle routing and congestion prediction for real-time driver guidance, Transportation Research Record **1408**, TRB, Washington, D.C., 66-74 (1993).
[5] Hobeika A.G,Chang kyun Kim, Traffic-flow-prediction system based on upstream traffic [A], Vehical Navigation and Information System Conference [C],1994.
[6] Ben-Akiva. M., Cascetta. E. and Whittaker. J., Recent progress in short-range traffic prediction, Compendium of Technical Papers, Institute of Transportation Engineers (ITE), 262-265 (1993).
[7] Hwanjo Yu, Jiong Yang, Jiawei Han, Xiaolei Li, Making SVMs Scalable to Large Data Sets using Hierarchical Cluster Indexing, Submission to Data Mining and Knowledge Dsicovery: an International Journal, May ,100-128 (2005).
[8] Haitao Xu, Hao Wu,Wanjun Zhang, Ning Zheng, Busy Stations Recognition of Hangzhou Public Free-Bicycle System based on Sixth Order Polynomial Smoothing Support Vector Machine, Proceedings of the 10th International Conference on Machine Learning and Cybernetics, July , 699-704 (2011).
[9] Haitao Xu, Hao Wu, XuJian Fang, and Wanjun Zhang, Finding Key Stations of Hangzhou Public Bicycle System by a Improved K-Means Algorithm, Applied Mechanics and Materials, Vols.209-211, 925-929 (2012).
[10] Boris Mirkin, Clustering Algorithms: a review, Mathematical Classification and Clustering, Chapter 3,Kluwer Academic Publishers, 109-169 (1996).
[11] Boris Mirkin, K-Means Clustering , Clustering for Data Mining, Taylor & Francis Group, 75-110 (2005).
[12] Boris Mirkin, Concept Learning and Feature Selection Based on Square-Error Clustering, Machine Learning, series **35**(1), 25-39 (1999).
[13] Shehroz S. Khan and Amir Ahmad, Cluster center Max-Minization algorithm for K-means clustering, Pattern Recognition Letters, series **25**(11), 1293-1302 (2004).
[14] Bjarni Bodvarsson, M. Morkebjerg, L.K. Hansen,G.M. Knudsen and C. Svarer, Extraction of time activity curves from positron emission tomography:K-means clustering or non-negative matrix factorization,NeuroImage, series **31**(2), 185-186 (2006).
[15] C. J. C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery, 121-167 (1998).
[16] V.Vapnik, Three remarks on support vector function estimation, IEEE transactions on Neural Networks, 988-1000(1999).
[17] Y. Lee, O. L. Mangarasian, SSVM: A smooth support vector machine for classification, Computational Optimization and Applications, series 22(1), 5-21 (2001).
[18] Y. Yuan, J. Yan and C. Xu, Polynomial Smooth Support Vector Machine(PSSVM), Chinese Journal Of computers, series **28** (1), 9-17 (2005).
[19] Y. Yuan, T. Huang, A Polynomial Smooth Support Vector Machine for Classification, Lecture Note in Artificial Intelligence, 3584, 157-164 (2005).
[20] Y. Yuan, W. G. Fan, and D.M. Pu, Spline Function Smooth Support Vector Machine for Classification, Journal of Industrial Management and Optimization, series **3**(3), 529-542 (2007).

Haitao Xu received the MS degree in computer software from Hangzhou Dianzi University in 2005, and now is a PhD candidate of Zhejiang University.His research interests are in the areas of machine learning and data mining.