

Station Segmentation with an Improved K-Means Algorithm for Hangzhou Public Bicycle System

Haitao Xu
Hangzhou Dianzi University, Hangzhou, China
Email: xuhaitaohdu@163.com

Jing Ying
Zhejiang University, Hangzhou, China
Email: ying_2000@yahoo.com

Fei Lin
Hangzhou Dianzi University, Hangzhou, China
Email: linfei@hdu.edu.cn

Yubo Yuan
East China University of Science and Technology, Shanghai, China
Email: yuanyb@163.com

***Abstract*—In China, Hangzhou is the first city to establish 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. There are many and many technology problems in the decision of intelligent dispatch. Among of these problems, determining how to segment the stations into several sections to give different care is very important. In this paper, an improved k-means algorithm based on optimized simulated annealing is used to segment the stations of Hangzhou Public Bicycle System. The optimized simulated annealing(SA) algorithm is used to assign k-means initial cluster centers. Practice examples and comparison with the traditional k-means algorithm are made. The results show that the proposed algorithm is efficient and robust. The research result has been applied in Hangzhou.**

***Index Terms*—Public Bicycle System, K-means algorithm, data mining, Segmentation**

I. Introduction

In the 1970s, China was named the Kingdom of Bicycles because of the nation's heavy reliance on cycling for mobility. China's citizens relied on bicycles because of their relatively low income, the country's compact urban development, and the short trip distances. Over the past 20 years, however, bicycle use has steadily declined because of economic growth, rapid motorization, longer trip distances, and a gradually deteriorating cycling environment. For instance, average bicycle ownership in Chinese cities declined from 197 bikes per 100 households in 1993 to 113 bikes per 100 households in 2007. Even some traditional cycling cities, in which the topography and weather are suitable for biking, also experienced decline. In Hangzhou, with a flat topography and an annual average temperature of 17.5°C, bicycle modal share has decreased from 60.78% in 1997 to 33.5% in 2007.

In light of growing traffic congestion and environmental concerns, the Chinese Ministry of Housing and Urban-Rural Development recently opposed bicycle use restrictions and supported tackling cycling barriers. 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 bicycle is difficult to be managed and will lead to secure traffic safety. "Too many private cars, bicycles too chaotic" is traffic problems.

Over the years, Hangzhou has set up "public transit priority" principle to ease the pressure on public transport. But, it is very difficult to get a good performance because of so many traffic jams. As road congestion problems are conspicuous, the average operating speed of public transport decline year after year, punctuality rate of less than 30 percent. "fast bus is not up and not on time" is the main reason for increased attractiveness of public transport.

Bike sharing (or short-term public use of a shared bicycle fleet) is one governmental initiative that supports this goal. On May 1, 2008, the Hangzhou city government launched the first information technology–based public bicycle system in mainland China.

The 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, Hangzhou 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. Hangzhou public bicycle system has

doi:10.4304/jsw.8.9.2289-2296

surpassed Vélib's bicycle system as the largest bike sharing program in the world.

For renting a bicycle, one need a Transportation Smart Card T or Z. Z card is specially designed for visitors. It can be applied at the Smart Card Center at 20 Long Xiang Lu, Shang Cheng District. You need to show your ID and store at least CNY300 in your new card.

After getting the card, you can rent a bicycle at any public bicycle spot, and return it at any service spot. When renting bike, you just need to put your card on the electric locker. When the light turns to green and the buzzer beeps you need to take the bike out of the locker within 30 seconds. A CNY200 deposit is required to rent the bike. This deposit minus the rental fee is returned when the bike is secured a locker at any service spot.

When returning the bike, you need to firstly make the bike locked by the electric locker, and put your card on the locker when the green light is on. The return of the bike is a success when the green light stops shining and the buzzer beeps.

The first 60 minutes of bicycle rental is for free. 1 RMB is charged for rental from 60 minutes to 120 minutes, 2 RMB from 120 minutes to 180 minutes, 3 RMB per hour for over 180 minutes. It will be deducted from the IC card upon returning the bicycle.

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. There are near 3000 bicycle stations, it is difficult to solve the problem of all stations. At first, we should choose some key stations to dispatch in the case of limited human resources.

Figure 1.Staion Photo

The primal goal of this work is to find out one algorithm to segment the stations into several sections. This paper includes four parts: The second part details the k-means algorithm. The third part presents the improved k-means clustering algorithm, the last part of this paper describes the experimental results and conclusions through experimenting with rent-return data sets of Hangzhou public bicycle system.

II. K-MEANS CLUSTERING

K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The clustering problem has been addressed in many articles and by researchers in many disciplines. This reflects its broad appeal and usefulness as one of the steps in exploratory data analysis. Clustering is used in many fields including data mining, statistical data analysis, image segmentation, pattern recognition, bio-informatics, financial series analysis, disease assistant diagnosis (such as cancer, heart disease), vector quantization and various business applications and so on.

The k-means algorithm applies to objects that are represented by points in a d-dimensional vector space. Thus, it clusters a set of d-dimensional vectors, $D = \{x_i | i = 1,\ldots,N\}$, where $x_i \in R^d$ denotes the ith object or "data point." K-means is a clustering algorithm that partitions D into k clusters of points. That is, the k-means algorithm clusters all of the data points in D such that each point x_i falls in one and only one of the k partitions. One can keep track of which point is in which cluster by assigning each point a cluster ID. Points with the same cluster ID are in the same cluster, while points with different cluster IDs are in different clusters. One can denote this with a cluster membership vector m of length N, where m_i is the cluster ID of x_i[1].

The value of k is an input to the base algorithm. Typically, the value for k is based on criteria such as prior knowledge of how many clusters actually appear in D, how many clusters are desired for the current application, or the types of clusters found by exploring/experimenting with different values of k[1].

In k-means, every of the k clusters are represented by a single point in R^d. We can denote this set of cluster representatives as the set $C = \{c_j | j = 1,\ldots,k\}$. These k cluster representatives are also called the cluster means or cluster centroids.

In clustering algorithms, points are grouped by some notion of "closeness" or "similarity." In k-means, the default measure of closeness is the Euclidean distance. The Euclidean distance $Ed(x_i, y_i)$ can be obtained as follow:

$$Ed(x_i, y_i) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

In particular, one can readily show that k-means attempts to minimize the following nonnegative cost function:

$$Cost = \sum_{i=1}^{N} (\mathrm{argmin}_j ||x_i - c_j||_2^2)$$

In other words, k-means attempts to minimize the total squared Euclidean distance between each point x_i and its closest cluster representative c_j. The above equation is often referred to as the k-means objective function.

The process of k-means algorithm can be described as follows[5]:

Input:

Number of desired clusters, k, and a database $D=\{d_1, d_2,\ldots d_n\}$ containing n data objects.

Output:
A set of k clusters.
Steps:
1) Randomly select k data objects from dataset D as initial cluster centers.
2) Repeat;
3) Calculate the distance between each data object d_i (1<=i<=n) and all k cluster centers C_j(1<=j<=k) and assign data object di to the nearest cluster.
4) For each cluster j (1<=j<=k), recalculate the cluster center.
5) Until no changing in the center of clusters.

The k-means clustering algorithm always converges to local minimum[5]. The particular local minimum found depends on the starting cluster centers. The k-means algorithm updates cluster centers till local minimum is found. Before the k-means algorithm converges, distance and center calculations are done while loops are executed a number of times.

III. IMPROVED K-MEANS CLUSTERING ALGORITHM

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 [5]. To solve this problem, we propose an improved k-means clustering algorithm based on optimized simulated annealing. We use the optimized simulated annealing (SA) algorithm to assign k-means initial cluster centers for Hangzhou public bicycle system dataset.

A. Simulated Annealing

Simulated annealing (SA) is a method that has been widely applied to deal with different combinatorial optimization problems. The main idea of simulated annealing comes from the analogy with thermodynamics of metal cooling and annealing. The Boltzmann probability distribution, $P(E)\sim\exp(\frac{-\Delta E}{cT})$, where c is a constant of nature that relate temperature to energy, shows that a system in thermal equilibrium at temperature T has its energy probabilistically distributed among all different energy states, Therefore, if there is a chance to being in a lower energy state, we can replace the local optimum by a worse solution. Metropolis et al. first introduced these ideas into numerical calculations. In this algorithm, for each possible solution, the measure calculated as an objective function is called "energy". The solution with the best measure is the one we are seeking [6].

This algorithm moves from the current solution to a new one with probability $P = \exp[-\frac{E_{new}-E_{current}}{T}]$, called the Metropolis acceptance rule, in which T is a sequence of temperature values, called annealing cooling schedule, and $E_{current}$ and E_{new} are the energies of the corresponding solutions. If $E_{new} < E_{current}$, then the algorithm selects E_{new} as the new current solution, otherwise, E_{new} is selected only with probability P. In other words, the system always moves downhill towards a better solution, while sometimes it takes an uphill step with a chance towards a better one. Based on the acceptance rule, when T is very high at the beginning, most of the moves, even toward a worse one, will be accepted, but while T approaches 0, most of the uphill moves will be rejected. There are two decisions that we have to make before the algorithm starts working: set an optimality criterion and an annealing schedule [6].

For optimality criterion, there are, usually, several optimization criteria, such as the probability to be in the ground state, the final energy, and the best-so-far energy, introduced in [7]. The best-so-far energy represents the lowest energy found in the solution path.

The annealing schedule is one of the important aspects for the efficiency of the algorithm. There are two kinds of annealing schedules, fixed and adaptive decrement rules. The fixed decrement rule is independent of the algorithm itself. The temperature is decreased in proportion to a constant over the course of the algorithm. There are several choices, such as the geometric cooling schedule, $T(t) = \alpha^t T_0$, the logarithmic cooling schedule, $T(t) = \frac{T_0}{\log(t)}$, and the exponential schedule, $T(t) = T_0^{(c-1)t}$, where t represents the annealing step. The adaptive decrement rule dynamically varies the proportional scale of the temperature decrements over the course of the algorithm. The fixed decrement rule is widely used because of its simplicity and the requirement of lower time complexity. The adaptive decrement rule can obtain a better performance, however, with more time complexity.

The steps of Simulated Annealing (SA) algorithm are as follow:

1) Initialization(Current_solution, Temperature);
2) Calculation of the Current_Cost;
3) Loop
   New_State;
   Calculation of the now_cost;
   If $(\Delta(\text{Current\_Cost} - \text{New\_Cost})<=0)$
     Current_State=New_State;
   Else
     If (Exp(($\text{Current}_{\text{Cost}}$ − New_Cost)/Temparature)>Random(0,1))
   //Accept;
   Current_State=New_State;
  else
    //Reject;
  Decrease the temperature;
  Exit when STOP_CRITERION
  End Loop

SA uses for optimization problems because all of optimization problems have a final state which has minimum energy. The goal of each optimization problem is reaching to the final state. The system will reach to a stable state in the final state. Thus, SA can be used for optimization problems, because SA reaches a system to a stable state.

B. K-means Clustering based on Simulated Annealing

The traditional k-means algorithm has sensitivity to the initial start center. To solve this problem, the following

algorithm is proposed to optimize the initial centers based on simulated annealing. The algorithm partition data points into K initial cluster, and calculate the initial cluster centers.

New Initial Centers Algorithm as follow:

Input:

Data set S which containing n samples and K which is the number of clusters.

Output:

K initial cluster centers C.

Steps:

1) $T=T_0$;
2) P_M=Random_partition(S);
3) $E_{object}=E(P_M)$; //E_{object} is the objective function of current partition P_M
4) C=Center(P_M); //compute the cluster centers of current partition P_M
5) $E_{best}= E_{object}$; // E_{best} is the value of best objective function
6) While (T>1)

```
{
  For (i=0;i<K;i++)
  {
    num=0;
    while(num<X)//X is iteration time
    {
        K number of objects which are farthest
        from the cluster centers are selected from
        the partition PM. The objects will be
        randomly assigned to other clusters, a
        new partition P'M is formed.
        C' = Center(P'M);
        E'object = E(P'M);
        ΔE = Eobject − E'object;
        if (ΔE)>0
        {
           PM=P'M;
           Eobject = E'object;
           C=C';
           if (Ebest>E'object) num=0;
           else num++;
        }
        else
        {
           if (random[0,1]<=min(1,exp[-ΔE/T]))
           {
              PM=P'M;
              C=C';
           }
        }
    }
  }
  T=α*T;
}
```

7) return C;

Different with the classical simulated annealing algorithm, the improved simulated annealing algorithm consists of three loops. The outermost loop is simulated annealing cooling process. The middle loop is the transformation rule of generating new solutions. The innermost loop is used to maintain sample's stability.

This algorithm is a problem of set partition, the rule of generating new solutions is random insertion. K number of objects which are farthest from the cluster centers are selected from the partition, and the objects will be randomly assigned to other clusters. By random insertion to accept worse solution, the algorithm is able to jump out of the local optimal solution and obtain the global optimal solution finally.

Using the initial cluster centers obtained by the improved algorithm, k-means algorithm avoids the blind search in the initial stage. The proposed method is more advantageous than K-means, which is not sensitive to the initial clustering center. Meanwhile, it reduces the number of iterations of the K-means algorithm and improves the efficiency of the algorithm.

C. Procedure and Setting Parameter of Improved Algorithm

In the k-means clustering based on simulated annealing, objective function E and parameters must be set. In this paper, the setting as follow:

1) Objective function:

In the new initial centers algorithm of Hangzhou public bicycle system, the objective function E is defined as follows:

$$E = \frac{\sum_{i=1}^{K-1}\sum_{j=i+1}^{K} d(x_i, x_j)}{\sum_{i=1}^{K} d(x_i)}$$

In the objective function, $d(x_i)$ is inner class distance of x_i, $\sum_{i=1}^{K} d(x_i)$ is the sum of inner class distance, $d(x_i, x_j)$ is the distance between x_i and x_j, $\sum_{i=1}^{K-1}\sum_{j=i+1}^{K} d(x_i, x_j)$ is sum of class-class distance.

$d(x_i, x_j)$ is the distance between x_i and x_j, d_i is defined as:

$$d_i = \sum_{c_i \in X_i} \mathrm{Dist}(c_i, \overline{c_\iota})$$

In the above formula, $\overline{c_\iota}$ is the center of X_i. $\overline{c_\iota}$ is defined as

$$\overline{c_\iota} = \frac{1}{|X_i|}\sum_{\mu \in X_i} \mu$$

$|X_i|$ is the number of elements in X_i.

$d(x_i, x_j)$ is defined as:

$$d(x_i, x_j) = \frac{1}{|X_i|}\sum_{\mu \in X_i} Dist\left(c_i, \overline{c_j}\right) (i \neq j)$$

$\overline{c_j}$ is the cluster center of X_j. The function of Dist is defined as:

$$\mathrm{Dist}(c_i, c_j) = \gamma D_1(c_i, c_j) + (1-\gamma) D_2(c_i, c_j), 1 \leq i, j \leq K, 0 < \gamma < 1$$

In the above formula, $D_1(c_i, c_j)$ is the distance of elements computed by the continuous properties. $D_2(c_i, c_j)$ is the distance of elements computed by discrete properties. γ is the weighting coefficient, it is defined as:

$$\gamma = \frac{\sqrt{num(A_1)}}{\sqrt{num(A_1)} + \sqrt{num(A_2)}}$$

$num(A_1)$ is the number of continuous properties. $num(A_2)$ is the number of discrete properties.

The formula of $D_1(c_i, c_j)$ is proposed as:

$$D_1(c_i, c_j) = \sqrt{\sum_{k=1}^{\alpha} (c_{i,k} - c_{j,k})^2}$$

$$\alpha = num(A_1)$$

The formula of $D_2(c_i, c_j)$ is proposed as:

$$D_2(c_i, c_j) = \sqrt{\sum_{k=\alpha}^{\beta} (c_{i,k} - c_{j,k})^2}$$

$$\alpha = num(A_1)$$

$$\beta = num(A_2)$$

Because continuous properties and discrete properties have different impacts on the computing of distance , we adopt the approach of summing weighting coefficient.

2) Initial temperature:

The initial temperature setting is one of the important factors that affect the simulated annealing algorithm for global search performance.

The probability of accepting a non-optimal solution is proportional to the temperature. Often, the higher temperature is used, the greater the chance of accepting the non-optimal solution, and the higher possibility of getting the global optimal solution. But this will take a lot of time to run. Conversely, it can save execution time, but global search performance may be affected.

In the practical application, initial temperature needs a number of adjustments based on the experimental results.

3) Annealing manner:

Annealing manner controls the decreasing speed of the temperature. Increasing the temperature reduction rate, the algorithm reduces the number of searches and the algorithm solution quality will be impacted. On the contrary, the algorithm solution quality may be improved, but it will take much more time.

Through testing in Hangzhou public bicycle system, the new temperature is generally between 90% -95% of the old temperature.

4) Disturbance manner:

Our algorithm uses method of random disturbance. When a sample's clustering partition of current solution is changed randomly, so that the new clustering partition is obtained, this makes our algorithm jump out the local minimum.

5) Iteration time:

Iteration time determines whether or not the innermost loop of the simulated annealing algorithm terminated. If Iteration time is too small, then the algorithm will search inadequately. If X is too large, the execution time of the algorithm is too long.

IV. Practical Examples

To practice the effectiveness of the improved algorithm, we performed extensive practical experiments. Table1 shows the number of borrowed and returned bicycle. In table 1, n is the number of stations and m is the rent-return records number.

TABLE 1.

DATABASE SCALE

Practical Example	n	m
E1	100	4981
E2	500	55874
E3	1000	134789
E4	1500	213488
E5	2000	295696

We can know from above introduction, simulated annealing algorithm is used in the first stage of the improved algorithm. One quality of simulated annealing algorithm is that its execution result is closely related to the selection of algorithm parameter. Better execution results can be obtained by selecting the appropriate algorithm parameters depending on the application environment.

The above introduction shows that the algorithm has the following parameters: initial temperature T0, the decreasing speed of temperature ΔT and iteration time X. Table 2 shows the parameters value in our practical examples.

TABLE 2.

PARAMETER SETTING OF IMPROVED ALGORITHM

Parameter	value
initial temperature T0	$T_0 = 10\Delta E_0$ /K
the decreasing speed of temperature α	α
iteration time X	X=7

To test the validity of the proposed algorithm, we compare the convergence time of k-means algorithm and the improved algorithm. Figure 2 shows the comparison results. X axis indicates the number of clusters. Y axis indicates the convergence time of algorithm. We can see that as the number of clusters increased, the convergence time of algorithm grew. Convergence time of the proposed algorithm is significantly less than that of the K-means algorithm.

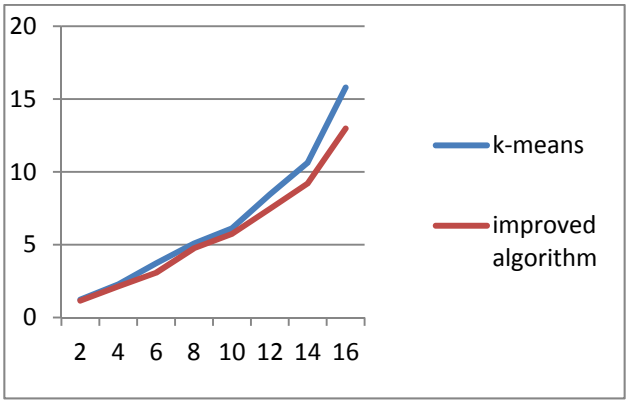

Figure 2.Convergence Time Comparison

Table3 shows the comparison result between the traditional k-means algorithm and our improved algorithm with CPU times to segment the stations as the number of stations increased. The practice is done on a pc with CPU 2.4G, 1.0 G DDR.

We can see from Table 3 that the convergence time of the algorithm increased with the growth of the number of stations. Convergence time of the proposed algorithm is significantly less than that of the K-means algorithm. The main reasons may include these aspects. Firstly, the proposed algorithm can jump out of local optimal solution based on the simulated annealing probabilistic search algorithm in the first stage. The initial clustering center can quickly be found. The k-means algorithm's circuitous search on the local optimal solution is avoided. Secondly, in the second phase of improved algorithm, we use the initial cluster centers which were found in simulated annealing algorithm instead of the random initial centers. This can make the clustering be obtained within a short period of time.

TABLE 3.
EXPERIMENTAL RESULTS

Practical Example	k-means	improved algorithm
E1	2.23	2.06
E2	4.44	3.26
E3	10.92	7.03
E4	30.83	26.92
E5	70.43	63.67

In the West Lake region, there are 529 public bicycle stations. We get all of the rent-return records in one month. Actually, from 1th December 2012 to 31th December 2012. According to the management office, the rank should be arranged as "ABCDE". The centers number is 5.

We employ the improved k-means clustering to cluster the 529 stations. Finally, Table 4 shows the centers.

TABLE 4.
CLUSTER CENTERS

Rank	Borrow	Return
A	9237.2	9347.3
B	7001.6	6903.9
C	5598.1	5713.4
D	2487.9	2497.3
E	523.2	548.7

There are near 3000 bicycle stations in Hangzhou, it is difficult to solve the dispatch problem of all stations. At first, we should choose some key stations from stations that ranked A to dispatch in the case of limited human resources.

Table 5 shows the key stations that we choose in Hangzhou public bicycle system.

TABLE 5
KEY STATIONS

No.	Station Num
1	5160
2	5161
3	5162
4	5163
5	5168
6	5169
7	5170
8	5211
9	5225
10	5044
11	5045
12	5062
13	5064
14	5071
15	5072
16	5129
17	5130
18	5141
19	5148
20	5369
21	5389
22	5421
23	5424
24	5361
25	5464
26	5476
27	5382
28	5378
29	8051
30	5481

Now, the staff of Hangzhou public bicycle company gave these 30 key stations special care. The problem to find no bicycle to rent or no place to return has be effectively solved.

V. CONCLUSION

Hangzhou public bicycle is unattended, sometimes someone find no bicycle to rent or no place to return at some stations. There are near 3000 bicycle stations, it is difficult to solve the problem of all stations.

At first, we should choose some key stations to dispatch in the case of limited human resources. To find the key stations, we proposed an improved k-means algorithm. The improved k-means algorithm involves two major processes. In the first process, simulated annealing

algorithm is used to find the initial cluster centers. In the second process, we use the initial cluster centers instead of the random initial centers.

In this way, k-means algorithm avoids the blind search in the initial stage. The proposed method is more advantageous than K-means, which is not sensitive to the initial clustering center. Meanwhile, it reduces the number of iterations of the K-means algorithm and improves the efficiency of the algorithm. Now, the research result has been applied in Hangzhou.

ACKNOWLEDGMENT

This work was financially supported by Research Project of Zhejiang Provincial Education Department (Y201120473), the National Natural Science Foundation of China (Nos. 61001200) and Zhejiang Provincial Natural Science Foundation of China (LY12F02017).

REFERENCES

[1] Xindong Wu, Vipin Kumar, *The Top Ten Algorithms in Data Mining*, CRC Press, Boca Raton, 2009, pp.21–25.
[2] Sun Jigui, Liu Jie, Zhao Lianyu, “Clustering algorithms Research”,Journal of Software ,Vol 19,No 1, pp.48-61,January 2008.
[3] Fahim A M,Salem A M,Torkey F A, “An efficient enhanced k-means clustering algorithm” Journal of Zhejiang University Science A, Vol.10,pp:1626-1633,July 2006.
[4] K.A.Abdul Nazeer, M.P.Sebastian, “Improving the Accuracy and Efficiency of the k-means Clustering Algorithm”,Proceeding of the World Congress on Engineering, vol 1,london, July 2009.
[5] Shi Na; Liu Xumin; Guan Yong, "Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm", Third International Symposium on Intelligent Information Technology and Security Informatics,pp 63-67,2010.
[6] 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, pp.925-929,2012.
[7] Mohammad Mehdi Keikha. "Improved Simulated Annealing using Momentum Terms". In Proceeding(s) of 2011 Second International Conference on Intelligent Systems, pp.44-48, 2011.
[8] K. Hoffmann and P. Salamon. “The optimal simulated annealing schedule for a simple model”. J. Phys. A: Math. Gen,23:3511–3523, 1989.
[9] QI Ji-Yang. "Application of Improved SimulatedAnnealing Algorithm in Facility Layout Design". In Proceeding(s) of the 29th Chinese Control Conference, pp.5224-5227, 2010.
[10] Wei Yang, Luis Rueda, Alioune Ngom. "A Simulated Annealing Approach to Find the Optimal Parameters for Fuzzy Clustering Microarray Data". In Proceeding(s) of the XXV International Conference of the Chilean Computer Science Society (SCCC’05), 2005.
[11] Haitao Xu, Jing Ying, Hao Wu and Fei Lin. "Public Bicycle Traffic Flow Prediction based on a Hybrid Model". Appl. Math. Inf. Sci.7,No.2, pp.667-674, 2013.
[12] 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". In Proceedings of the 10th International Conference on Machine Learning and Cybernetics, pp. 699-704, 2011.
[13] 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), pp.529-542, 2007.
[14] Sun Jigui, Liu Jie, Zhao Lianyu, “Clustering algorithms Research”,Journal of Software ,Vol 19,No 1, pp.48-61,January 2008.
[15] Xuesong Yin, Songcan Chen, Enliang Hu. "Regularized soft K-means for discriminant analysis ". Neurocomputing, Volume 103, pp. 29–42, 2013.
[16] Craig J. Bennetts, Tammy M. Owings, Ahmet Erdemir, Georgeanne Botek, Peter R. Cavanagh. "Clustering and classification of regional peak plantar pressures of diabetic feet". Journal of Biomechanics, Volume 46, pp. 19–25, 2013.
[17] D. Tefankovi et al. “Adaptive simulated annealing: A near-optimal connection between sampling and counting,” J. ACM, vol. 56, no. 3, May, 2009, pp. 1-36.
[18] L. Lamberti, “An efficient simulated annealing algorithm for design optimization of truss structures,” Computers & Structures, vol. 86, October 2008, pp. 1936-1953.
[19] Raman, Dhamodharan, Nagalingam, Sev V.; Gurd, Bruce W. “A genetic algorithm and queuing theory based methodology for facilities layout problem”. International Journal of Production Research, 2009,47(20),pp.5611-5635.
[20] Asha Gowda Karegowda, Vidya T., Shama, M.A. Jayaram, and A.S. Manjunath. "Improving Performance of K-Means Clustering by Initializing Cluster Centers Using Genetic Algorithm and Entropy Based Fuzzy Clustering for Categorization of Diabetic Patients", In Proceedings of ICAdC, AISC 174, pp. 899–904, 2013.

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.

He joined in the research project named “The Data Fusion of Public Transport Resources and Intelligent System based on things” in 2009. In 2012, the project won the second prize of Zhejiang Science and Technology Award. In addition, the project named “Research and Development of Public Bicycle Transportation System” won the first prize of Zhejiang Higher scientific research achievement awards in 2010. In the award, he ranked second.

Dr. Jing Ying, born in October 1971, Professor. He is deputy director of College of Computer Science and Technology, Zhejiang University. In 2003, he won the second prize of Zhejiang Province Award for excellence in the Eleventh teenagers (youth group). As a member of the chinese institute of computer software engineering committee, he mainly engaged in software development, software architecture, software engineering specifications. He has accomplished more than 20 scientific research projects at national and ministerial level. And he has published over 30 theses in important academic periodicals home and abroad.

Fei Lin completed her M.S. in Circuit and System from Hangzhou Dianzi University, Hangzhou, Zhejiang, China in 2002. She is currently working as Associate Professor and Assistant Dean in the Institute of Software Engineer, Hangzhou Dianzi University, Hangzhou, Zhejiang, China.

She has more than 15 research publications in national and international journals and conferences. Her research interests are in the area of large-scale distributed systems,specifically storing and computing , cloud computing and SaaS.

Yubo Yuan received the B.Sc. and M.Sc. degrees in information and computational science from Lanzhou University, PR China, in 1997 and 2000, and Ph.D. degree in information and computational science from Xi'an Jiaotong University, PR China, in 2003. From November 2003 to June 2006, he worked as Lecture and Associate Professor in University of Electronic Science and Technology of China. From July 2006 to August 2008, he worked as Research Assistant Professor in Mathematics Department of Virginia Tech, USA. From February 2009, he has been working as director in the Institute of Metrology and Computational Science, China Jiliang University, PR China. His current research include optimization, data mining, machine learning, computational intelligence and extreme learning machine.