Proceedings of the 2011 International Conference on Machine Learning and Cybernetics, Guilin, 10-13 July, 2011

Station Segmentation of Hangzhou Public Free-Bicycle System based on Improved Randomized Algorithm

Hao Wu,Xujian Fang,Haitao Xu

School of Software Engineering, Hangzhou Dianzi University, Hangzhou 310012, China
E-MIAL: xuhaitao@hdu.edu.cn

Abstract

In order to cope with "the last kilometer to end a transit trip", hence to ease traffic and parking difficulties, and to boost a sustainable green transportation system, this paper focuses on a bike sharing program implemented in Hangzhou, Zhejiang. There are many and many technology problems in the decision of intelligent dispatch. In this paper, we consider a special case of the system, it is to segment the stations into two sections. An improved randomized algorithm is proposed to do it. After analysis on the performance of algorithm, the result is 0.6378 and better that the trivial case of 0.5.

Keywords:

data mining; clustering; Segmentation; random algorithm; optimization

1 Introduction

1.1 Brief introduction of Hangzhou Public Free-Bicycle System

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 andto solve road congestion more difficult. Private bicycles 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 a principle "public transit priority" 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.

A complimentary public free-bicycle sysytem, 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. The end of 2010, Hangzhou city public bike has covered 3000 service points, a total of 5 million bicycles.

Fig. 1 Station Photo

In the West Lake region, there are 125 locations, such as Lingyin Temple Bus Interchange, Hangzhou Zoo Station, Nine Creeks Bus Interchange, South Entrance of Su Causeway, Wanghulou Tea House (next to Broken Bridge), car park of Shengtang main entrance, Gate 1 of Liulang Wenying Park (also known as Orioles Singing in the Willows),

978-1-4577-0308-9/11/$26.00 ©2011 IEEE

South Gate of Long Bridge Park, Pinghu Qiuyue Park on Bai Causeway (also known as Autumn Moon over the Calm Lake) and the entrance to Yue Miao (also known as Yuefei's Temple).

20-50 bicycles are available for rental initially and numbers will be adjusted according to the actual operational status at each location.

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.

Bus IC cards (either regular or concession cards) and citizen cards should be presented for bicycle rental. One card is eligible for one bicycle only and you should return the bicycle with the same card.

People should top up the card with 200 yuan as deposit to enjoy the service. The deposit can be refunded in the future if one want to cancel the service.

Tourists or citizens without bus IC cards can purchase a Z-card at 300 yuan, with 200 yuan as deposit and 100 yuan for consumption. You can be refunded the balance remaining in your Z-card within 10 days at any bicycle rental locations or bus IC card top up locations; however, refund over 10 days is only available in the ticketing office at Longxiangqiao Bus Interchange.

Z-card is available in any point of bus IC card sales and the bicycle rental locations. It can be used to pay bus fare at 9% discount, taxi fare (only for those installed with IC card POS) or water bus fare as well.

The first 30 minutes of bicycle rental is for free. 1 yuan is charged for rental from 30 minutes to 90 minutes, 2 yuan from 90 minutes to 150 minutes, 3 yuan per hour for over 150 minutes. It will be deducted from the IC card upon returning the bicycle.

The free rental period is prolonged from 30 minutes to 60 minutes for interchange from bus to bicycle. A discount will also be offered on bus fare for interchange from bicycle to bus once the system is ready.

The primal goal of this work is to find out one algorithm to segment the stations into two sections. In section 2, improved randomized algorithm is introduced. In section 3, the application to segment the stations of the Hangzhou public free-bicycle system is investigated. At the end, we give a conclusion.

2 Improved Randomized Algorithm

In this section, we transfer stations segmentation problem into a max-cut problem by constructing a bipartite graph. The max-cut problem is then relaxed as a semidefinite programming problem. Finally a randomized algorithm is proposed for the solution of resulting semidefinite programming problem.

The bipartite graph $G = (C, E)$ is constructed by the set of station nodes C, and the set of edges E. If station c_i is neighbored with c_j, then there is an edge $e_{ij} \in E$. It is assumed that the weights on all edges in E are the same, that is, $w_{ij} = 1$ for all $e_{ij} \in E$. Then the stations segmentation problem is equivalent to the following problem.

Given a graph $G = (C, E)$ with $|C| = n(n$ is even), C_1 and C_2 of C such that

$$\begin{aligned} \max \quad & \beta(C_1) + \beta(C_2), \\ \text{s.t.} \quad & C_1 \bigcup C_2 = C,\ C_1 \bigcap C_2 = \emptyset, \\ & |C_1| = |C_2| = n/2, \end{aligned}$$

where $\beta(C_k)$, $k = 1, 2$ denotes the sum of edges ,

$$\beta(C_k) = \sum_{c_i \in C_k,\ c_j \in C_k} t_{ij}.$$

Let

$$w_{ij} = \begin{cases} 1, & \text{if edge } e_{ij} \text{ exists} \\ 0, & \text{otherwise} \end{cases}$$

Then t_{ij} can be expressed as

$$t_{ij} = \frac{1}{2} w_{ij}(1 - y_i y_j),$$

$$y_i, y_j \in \{-1, 1\} \text{ for } c_i, c_j \in C.$$

The values of the variables y_i and z_j are defined by

$$y_i = \begin{cases} -1, & \text{if } c_i \in C_1, \\ 1, & \text{if } c_i \in C_2, \end{cases}$$

Then problem (**??**) can be expressed as

$$\begin{aligned} \max \quad & \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij}(1 - y_i y_j), \\ \text{s.t.} \quad & y_i, y_j \in \{-1, 1\}, \\ & \sum_{i=1}^{n} y_i = 0. \end{aligned} \tag{1}$$

where the last constraint comes from the first and the third constraints in problem (**??**).

Let β^* denote the optimal value of problem (**??**). Since, it is stations segmentation problem,(find two subsets C_1 and C_2 of C having the same cardinality $n/2$), it is obvious that we have

$$\beta^* \geq \frac{|E|}{2}$$

where $|E|$ is the number of edges in graph G.

According to the method of Goemans and Willamson [2], problem (2) can be relaxed to be a semidefinite programming in space R^{n+m}. The relaxation is based on the fact that all possible solutions of problem (2) are feasible for the relaxation, and the optimal value of the relaxated problem gives an upper bound on the optimal value of problem (**??**). The relaxation can be realized by simply regarding y_i and z_j as unit vectors in space R^n with Euclidean norm, that is, y_i are replaced by vectors $v_i \in R^n$ with $v_i \in \mathcal{B} = \{ x \in R^n \mid \|x\| = 1 \}$. The objective function and the second constraint in problem (2) are defined as

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}(1 - v_i^T v_j) \quad \text{and} \quad \sum_{j=1}^{m} v_j = 0.$$

in the relaxation. So the relaxed problem is

$$\begin{aligned} \max \quad & \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}(1 - v_i^T v_j), \\ \text{s.t.} \quad & v_i, v_j \in \mathcal{B}, \\ & \sum_{j=1}^{m} v_j = 0. \end{aligned} \tag{2}$$

This is a semidefinite programming problem and can be solved.

We now present the improved randomized algorithm for the stations segmentation problem.

Algorithm IRA

- (1) Solve problem (**??**) to obtain a set of vectors $v_1, v_2, \cdots, v_n \in \mathcal{B} \subset R^n$.
- (2) Give an $\epsilon > 0$, and for $t = 1$ to $T(\epsilon) = \epsilon^{-1}\ln(\epsilon^{-1})$ do:
  - (2.1) Set $C_1^t = \emptyset$, $C_2^t = \emptyset$, $\tilde{C}_1^t = \emptyset$, $\tilde{C}_2^t = \emptyset$, and randomly choose two vectors z_1 and z_2 from the unit sphere $\mathcal{B}$;
  - (2.2) Let $\tilde{S}_t = \{ j \mid v_j^T z_1 \le v_j^T z_2 \}$;
  - (2.3) Suppose $|\tilde{S}_t| = \ell \ge n/2$ (without loss of generality).
    For each $j \in \tilde{S}_t$, define $\Delta(j) = \beta(C, \{c_j\})$, and rearrange the elements in $\tilde{S}_t$ in the decreasing order of the numbers $\Delta(j)$, that is,
    $$\tilde{S}_t = \{j_1, j_2, \cdots, j_\ell\},$$
    with
    $$\Delta(j_1) \ge \Delta(j_2) \ge \cdots \ge \Delta(j_\ell).$$
    Let
    $$S_t = \{j_1, j_2, \cdots, j_{n/2}\}.$$
  - (2.4) For $k = 1$ to n do
    If $\beta(\{c_k\}, C_1^t) \ge \beta(\{c_k\}, C_2^t)$ then $C_1^t = C_1^t \bigcup \{c_k\}$ else $C_2^t = C_2^t \bigcup \{c_k\}$;
    If $\beta(\{c_k\}, \tilde{C}_1^t) \ge \beta(\{c_k\}, \tilde{C}_2^t)$ then $\tilde{C}_1^t = \tilde{C}_1^t \bigcup \{c_k\}$ else $\tilde{C}_2^t = \tilde{C}_2^t \bigcup \{c_k\}$.
  - (2.5) If $t < T$, then $t := t + 1$ and go to (2.1).
- (3) Output C_1^*, C_2^* with $\beta(C_1^*) + \beta(C_2^*) = \max_{1 \le t \le T}[\beta(C_1^t) + \beta(C_2^t)]$.

Note: (1) $\epsilon > 0$ is the control factor of the procedure, and is a given small positive constant. (2) The segmentation $\tilde{C}_1^t$ and $\tilde{C}_2^t$ are only used in the analysis of the algorithm's performance guarantee.

3 Analysis of the algorithm

In this section, we analyze the performance guarantee of the algorithm IRA. Let P_t denote the segmentation C_1^t and C_2^t, and $\tilde{P}_t$ denote the segmentation $\tilde{C}_1^t$ and $\tilde{C}_2^t$. Let W^* denote the maximum value of problem (3). It is obviously that

$$W^* \ge \beta(P_t), \quad W^* \ge \beta(\tilde{P}_t).$$

In order to get the estimation of performance guarantee, we introduce the following random variables,

$$J_t = \beta(\tilde{P}_t) = \beta(\tilde{C}_1^t) + \beta(\tilde{C}_2^t), \ \ 1 \le t \le T.$$

Lemma 1 Let $E(J_t)$ denote the expected value of the random variable J_t. Then

$$E(J_t) \ge \alpha W^*,$$

where α is a positive constant and $\alpha \ge 0.87856$.

The proof of this lemma can be found in [2] (Goemans and Williamson) and [3] (Frieze and Jerrum), and hence is omitted.

Theorem 2 Let W^* be the optimal value of problem (3), and P^* be the optimal value outputed from algorithm IRA, then we have

$$P^* \ \ge \ [2\sqrt{2(1-\epsilon)\alpha} - 2]W^*.$$

Proof From Lemma 1, we have

$$E(Z_t) = \frac{E(J_t)}{W^*} \ge 2\alpha.$$

Since $J_t \le W^*$, we have

$$Z_t = \frac{J_t}{W^*} \le 2.$$

Let

$$P^* = \max_{1 \le t \le T}[\beta(C_1^t) + \beta(C_2^t)].$$

Assume that $|\tilde{C}_1^t| = \delta_t n$, $\frac{1}{2} \le \delta_t < 1$. Then from steps (2.3) and (2.4) of Algorithm IRA, we have

$$\beta(C_1^t) + \beta(C_2^t) \ge \frac{m}{2\delta_t m}[\beta(\tilde{C}_1^t) + \beta(\tilde{C}_2^t)] = \frac{1}{2\delta_t} J_t,$$

which gives

$$P^* \geq \frac{1}{2\delta_t} J_t, \text{ for all } 1 \leq t \leq T.$$

Let

$$Z_\tau = \max_{1 \leq t \leq T} \{Z_t\}.$$

Then it follows from the independency of $Z_1, Z_2, \cdots, Z_T$ that

$$\begin{aligned} & Pr[Z_\tau \leq 2(1-\epsilon)\alpha] \\ = \; & Pr[Z_1 \leq 2(1-\epsilon)\alpha] \cdot Pr[Z_2 \leq 2(1-\epsilon)\alpha] \\ & \cdots Pr[Z_T \leq 2(1-\epsilon)\alpha], \end{aligned}$$

where $Pr(Z_t \leq 2(1-\epsilon)\alpha]$ denotes the probability of $Z_t \leq 2(1-\epsilon)\alpha$. For any given $\epsilon > 0$,

$$Pr[Z_t \leq 2(1-\epsilon)\alpha] \leq \frac{1-\alpha}{1-(1-\epsilon)\alpha} < 1.$$

Then, we have

$$Pr[Z_\tau \leq 2(1-\epsilon)\alpha] \leq (\frac{1-\alpha}{1-(1-\epsilon)\alpha})^T < \epsilon,$$

for given small value of ϵ in Algorithm IRA.
Hence,we can assume that

$$Z_\tau = \frac{J_\tau}{W^*} \geq 2(1-\epsilon)\alpha$$

holds with large probability.

We obtain

$$\lambda \geq 2(1-\epsilon)\alpha - 4\delta_\tau(1-\delta_\tau), \quad \frac{1}{2} \leq \delta_\tau < 1.$$

Thus,

$$\begin{aligned} P^* \; & \geq \; \frac{1}{2\delta_\tau} J_\tau = \frac{\lambda}{2\delta_\tau} W^* \\ & \geq \; \frac{2(1-\epsilon)\alpha - 4\delta_\tau(1-\delta_\tau)}{2\delta_\tau} W^*. \end{aligned}$$

Define $\phi(\delta_\tau) = [2(1-\epsilon)\alpha - 4\delta_\tau(1-\delta_\tau)]/(2\delta_\tau)$, and minimizing the function $\phi(\delta_\tau)$ for $1/2 \leq \delta_\tau < 1$, we obtain

$$P^* \;\; \geq \;\; \phi(\delta_\tau) W^* \geq [2\sqrt{2(1-\epsilon)\alpha} - 2] W^*.$$

This completes the proof.

Since the optimal value W^* of problem (3) gives an upper bound on the optimal value F^* of problem (2) that is the optimal value of problem (1), we get that

$$P^* \geq [2\sqrt{2(1-\epsilon)\alpha} - 2] F^*.$$

If the value of ϵ in the algorithm is less than 0.01, then we have

$$P^* \geq 0.6378 F^*,$$

that is, the preformance guarantee of the algorithm is greater than 0.6378, and improves the trivial 0.5-approximation algorithm.

4 Application to segment the stations of the Hangzhou public free-bicycle system

We investigate the total number of one week. Actually, from 7th March 2010 to 14th March 2010. We select 9 stations near West Lake. The station codes and names are listed in the following table.

Station Code	Station Name
8009	Hangzhou Zoo
8010	Hangzhou Garden
8005	West Gate of Hangzhou Garden
8011	West Gate of Hua Gang
8007	Central Station of Linying Temple
8014	Liu Lang Wen Yin
8006	Man Jue Long
8004	Mulan Mountain Tea
8012	South Side of Suti

The station with station data are presented in the following table. The codes on the top of table (the first line) are borrow stations. The codes on the left side of the table (the first column) are return stations.

	8009	8010	8005	8011	8007	8014	8006	8004	8012
8009	438	41	5	51	9	37	81	7	31
8010	29	651	26	196	37	58	6	49	74
8005	3	21	175	28	38	0	5	49	14
8011	72	256	27	676	31	129	29	29	163
8007	18	19	14	16	375	15	0	101	9
8014	27	31	9	63	1	1505	9	1	200
8006	69	2	5	14	0	9	78	1	14
8004	10	17	30	37	72	5	12	295	12
8012	162	188	12	185	19	255	19	10	1311

We use the improved randomized algorithm to segment the station as two part. The algorithm is done ten times. The result is listed in the following table.

The number of running	Performance(P^*)
1	0.7732
2	0.8513
3	0.7071
4	0.6854
5	0.6461
6	0.7012
7	0.8231
8	0.6917
9	0.6613
10	0.7441

The average performance of improved randomized algorithm is

$$\bar{P}^* = \frac{1}{10} \sum_{i=1}^{10} P_i^* = 0.7284.$$

5 Conclusion Remark

From the experiment results, we can see the proposed algorithm can be used to segment the Hangzhou Public Free-Bicycle System. The average performance is beyond 0.5. It is useful to do it.

Acknowledgment

This research has been supported by the National Natural Science Foundation under Grants(Nos. 61001200) and Natural Science Foundation and Education Department of Zhejiang under Grants(Nos. Y6100010).

References

[1] C. J. C. Burges, *A Tutorial on Support Vector Machines for Pattern Recognition*, Data Mining and Knowledge Discovery, **2** (1998), 121-167.

[2] C. Cortes, V. Vapnik, *Support-vector networks*, Machine Learning, **(3)**20, SEP 1995, 273-297.

[3] V.Vapnik, *Three remarks on support vector function estimation*, IEEE transactions on Neural Networks, **10** (1999), 988-1000.

[4] G. Flake and S. Lawrence, *Efficient SVM Regression Training with SMO*. Machine Learning, **1-3** Volume 46,January 2002,271-290.

[5] Y. Lee and O. L. Mangarasian, *SSVM: A smooth support vector machine for classification*, Computational Optimization and Applications,**22(1)**(2001),5-21.

[6] Y. Yuan, J. Yan and C. Xu, *Polynomial Smooth Support Vector Machine(PSSVM)*, Chinese Journal Of computers, **28 (1)**(2005),9-17.

[7] Y. Yuan and T. Huang, *A Polynomial Smooth Support Vector Machine for Classification*, Lecture Note in Artificial Intelligence, **3584**(2005),157-164.

[8] Y. Yuan, W. G. Fan and D.M. Pu, *Spline Function Smooth Support Vector Machine for Classification*, Journal of Industrial Management and Optimization, **3(3)**(2007),529-542.

[9] Y. Yuan and R. Byrd, *Non-quasi-Newton updates for unconstrained optimization*, J. Comput. Math., **13**(1995),95-107.

[10] Y. Yuan, *A modified BFGS algorithm for unconstrained optimization*, IMA J. Numer. Anal., **11**(1991),325-332.

[11] Z. B. Xu, F. L. Cao, *Simultaneous L^p-approximation order for neural networks*, Neural Networks, 18(2005), 914-923.