k 平均法を用いたBLEデバイスの配置手法

大西 建輔[1,a)]

概要：Bluetooth low-energy(BLE) を用いたデバイスを用いた事例は, 急速に増加している. BLE デバイスの価格は減少しているが, その設置コストは高いままである. これは, BLE デバイスを設置する際, それぞれのデバイスの設置場所を決めることが, 複雑で時間を必要とする作業であるからである.
本研究では, k 平均法を用いて, 設置領域に対し, BLE デバイスの設置場所を決定する手法を提案する. また, 現実の設置領域での配置場所の計算機実験をおこなった. このとき, 提案手法の中で用いる生成点の数と, 設置する BLE デバイスの数を変更させて実験をおこなった. さらに, その設置領域に実際に BLE デバイスを設置した際の計測実験についても説明する.

キーワード：Bluetooth Low Energy, k 平均法, k-means++ method

A computation algorithm for the configuration of BLE devices using k-means method

Kensuke Onishi[1,a)]

Abstract: The number of devices which use Bluetooth low-energy (BLE) technology is increasing rapidly. Although the cost of the BLE components is decreasing, the installation costs remain high. When many BLE devices are installed, determination of the locations for installation, referred to as the configuration, becomes complicated and labor intensive.
In this research, we investigate computational methods for configuring BLE devices in a given region. We propose a computation method using k-means method, and present numerical results. We then evaluate the proposed method in terms of the number of generation points and the number of BLE devices to be installed. We also introduce a measure for the configuration.

Keywords: Bluetooth Low Energy, k-means method, k-means++ method

1. Introduction

Bluetooth low energy (BLE), is a low-energy communication mechanism [1]. Various devices are equipped to use BLE signals, and are referred to here as *BLE devices*.

BLE devices have a wide range of applications, for example, in the guide service system in Ueno Zoological Gardens and Hama-rikyu Gardens [2], the guide system for Himeji [3], the interactive security system for Narita Airport [4], and the parking system at Nagoya [5].

Research has been conducted on using BLE devices for indoor position detection, such as Ishizuka et al. [6], who proposed a pedestrian dead reckoning method. Onishi [7] proposed a method using the ordered order-k Voronoi diagrams [8], pp.144-151.

The cost of manufacturing BLE devices is decreasing as the production volume increases, however the labor cost for installation remains high, or is even increasing. The reason for this is the work flow for installation. Currently, the devices are installed by trial and error i.e., the installation of BLE devices (beacons) and measurement of

[1] 東海大学理学部情報数理学科
Tokai University, School of Science, Department of Mathematical Sciences
[a)] onishi@tokai-u.jp

signals is repeated until a good configuration is obtained.

Here, we address the following problem.

[Problem] Given a connected region $\mathcal{R}$ in the Euclidean plane and a positive integer k, we compute a set $SP = \{sp_1, \ldots, sp_k\}$ of k points in $\mathcal{R}$, where we minimize the sum of the squares of the distances between any two points in $\mathcal{R}$ and the nearest sp_i from the points.

To formulate the above problem as an optimization problem, we introduce a set $P \subset \mathcal{R}$ of points and an objective function as follows.

$$\varphi(P) = \sum_{i=1}^{k} \sum_{p \in P_i} \mathrm{dist}(p, sp_i)^2,$$

where P_i is the set of points in P nearest to p_i and $\mathrm{dist}(\cdot, \cdot)$ is the Euclidean distance. We also define the *normalized objective function* φ.

$$\varphi = \frac{1}{|P|}\varphi(P).$$

Here φ is the mean of the squares of the Euclidean distances and is independent of the size of P. So, we can compare φ for different sets.

The problem above is similar to the facility location problem, which considers the minimization of the sum of the square of the distances between points and the nearest facility [9]. In [9], a region, a density function over the region and a positive integer k are given, and the (local) optimal value of the objective function is then computed by the steepest descent method. Since the facility location problem is $\mathcal{NP}$-hard, many algorithms for approximating the solution have been proposed, and are reviewed in [10].

Our contributions are (1) we propose a method that requires only the connected region and a positive integer $k > 0$; (2) we show that the number k is dependent on the number of BLE devices used. This paper is organized as follows. Two related works on the k-means and k-means++ method are introduced in Section 2. The proposed method is described in Section 3. Numerical experiments are shown in Section 4. We discuss the experiments in Section 5.

2. Related Work

2.1 k-means method

MacQueen proposed the k-means method for dividing the set P into k clusters, which is called non-hierarchical clustering [11]. The k-means method is outlined in Algorithm 1.

Algorithm 1 k-means method

Require: $P = \{p_1, p_2, \ldots, p_n\}$;
Require: k;
Select k points from P, called $SP = \{sp_1, \ldots, sp_k\}$;
Compute $\varphi(P)$;
repeat
  $\varphi' \leftarrow \varphi(P)$;
  Divide P into $P_i (i = 1, \ldots k)$ s.t. $P_i = \{p_j \in P : \mathrm{dist}(p_j, sp_i) < \mathrm{dist}(p_j, sp_l) \quad l \neq i\}$, where $\mathrm{dist}(p, q)$ is a distance function between p, q.
  for $i = 1$ to k **do**
    $sp_i \leftarrow \dfrac{1}{|P_i|} \sum_{p \in P_i} p$;
  end for
  Compute the objective function $\varphi(P) \leftarrow \sum_{i=1}^{k} \sum_{p \in P_i} \mathrm{dist}(p, sp_i)^2$;
until $(\varphi' \leq \varphi(P))$
return $P_i (i = 1, \ldots k)$;

Algorithm 1 is used to iteratively calculate and minimize the objective function φ. Calculating the minimum objective function explicitly using the k-means method requires $O(n^{kd})$ computation time [12], where n is the number of points and d is the dimension of the points. Various polynomial-time approximation scheme algorithms have been proposed [13], [14]. It is known that the k-means method outputs the local optimal solution from the initial points. A method for iterating the k-means method for different initial points is called the *random method* in this paper.

2.2 Selection of initial points

There are a number of methods for selecting the initial points. Ostrovsky et al. proposed a selection method [15]; if the set P of points satisfies the conditions for the optimal k-means solution, the random selection from P is $O(1)$-competitive [*1].

Arthur and Sergei proposed the k-means++ method [16]. The k-means++ method is outlined in Algorithm 2. For any set of points, the algorithm always computes a value of φ less than $\Theta(\log k)$ times the optimal value.

In Algorithm 2, the selection of a sp_i in Step 3 is dependent on the distance from the nearest sp_i of the previous step.

3. Proposed Method

In this section, we present our proposed method. Let $\mathcal{R}$ and k be the connected region and a positive integer,

[*1] An approximate algorithm is $O(f)$-competitive when the value of φ of the algorithm always satisfies $E[\varphi] \leq O(f)\cdot$ (optimal value).

Algorithm 2 k-means++ method

Require: $P = \{p_1, p_2, \ldots, p_n\}$;
Require: k;
1: Select one point sp_1 from P randomly;
2: **repeat**
3: Choose the next point sp_i from P with probability $\dfrac{\mathrm{dist}(p, sp_l)}{\sum_p \mathrm{dist}(p, sp_l)}$, where $\mathrm{dist}(p, q)$ is the distance function and sp_l is the nearest point to p among the selected points;
4: **until** (k points are selected)
5: apply the k-means method, using $SP = \{sp_1, \ldots, sp_k\}$ as the initial points;

respectively. We propose Algorithm 3 for computing a set SP of k points from $\mathcal{R}$ and k.

Algorithm 3 Proposed method for computing the configuration of BLE devices

Require: Connected region $\mathcal{R}$;
Require: k;
Fix a positive integer n;
for randomly generate sets of n points $P = \{p_i \mid p_i \in \mathcal{R}, i = 1, \ldots, n\}$ **do**
  Apply the k-means method to P;
end for
Choose the cluster with the smallest value of the objective function $\varphi(P)$;
return all centroids of the cluster;

In Algorithm 3, the number n is fixed and sets of n points are simulated. We repeat the simulation of points until the estimate for $\varphi(P)$ is within a predefined tolerance level. Then, we select the cluster with the smallest value of $\phi(P)$.

The k-means method also introduces randomness into the selection of points in Step 1, Algorithm 1. The k-means++ method has randomness in Steps 1 and 3 of Algorithm 2. We use some sets SP of selection points from P to obtain a better configuration.

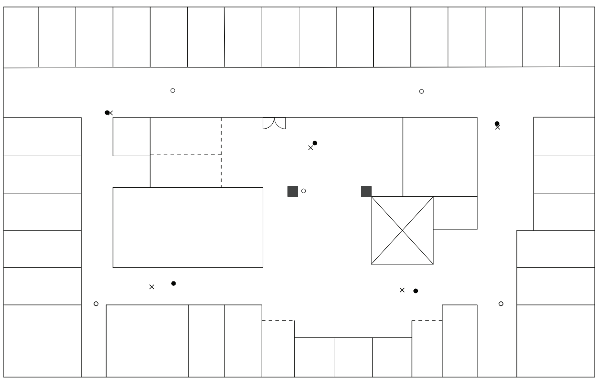

図 1 Configurations of BLE devices (○: old, ●: best for 1000 points, ×: best for 7000 points)

Figure 1 shows different configurations of the BLE devices. The configuration ○ was determined manually by the author, and the others were computed by the proposed algorithm. In Fig. 1, ● and × represent the best configurations among the 100 configurations for 5 BLE devices with 1000 and 7000 points, respectively.

4. Experimental Results

4.1 Overview of experiments

In this section, we describe the performance of our method for a real-world application.

Our experiments were divided into two cases. One considered different numbers of generation points, n. The second considered different numbers of BLE devices, k. The results of these experiments are shown in Sections 4.2 and 4.3, respectively. We present the results for the application of our method to a real-world example in Section 4.4.

表 1 Computer specifications.

Model Name	Think Pad X230
Manufacturer	Lenovo
OS	Windows 7
CPU	Intel(R) Core(TM) i7-3520M (2.90 GHz)
Main Memory	16 GB
Language	Perl v5.14.2

Table 1 gives the specifications of the computer on which the experiments were conducted. The experiment was conducted as follows. First, for fixed n and k, generate a set P of n points in $\mathcal{R}$ randomly. Then, select 10 different sets SP of k points from P using the random and k-means++ methods. For each pair P and SP, apply the k-means method. Repeat for 10 different P. A total of 100 experiments were executed for pairs of n and k. The mean, minimum and maximum of the objective function for these 100 experiments are presented in from Table 2 to Table 7.

4.2 Number of points

In this section, we describe the results of our numerical experiments on varying the number n of generation points. We fixed the number of BLE devices at $k = 5$, and increased n from 1000 to 10000 in steps of 1000.

Table 2 shows the mean computation time (msec) per experiment. The second to fourth rows are the calculation times for the random method. The second row is the selection time for the initial points SP. The third row is the computation time for the k-means method. The fourth

表 2 Mean of computation time (msec), $k = 5$

method	n	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
random	selection	42	51	53	56	62	66	85	81	90	84
	k-means	330	631	1017	1640	1970	2461	2617	3029	3343	3827
	total	372	682	1070	1696	2032	2527	2702	3110	3433	3911
k-means++	selection	78	123	155	193	232	279	327	384	427	479
	k-means	368	586	1170	1422	1762	2239	2894	2991	3182	3853
	total	446	709	1325	1615	1993	2518	3221	3375	3609	4332

row is the sum of the selection time and the computation times in the above rows. The fifth to seventh rows are the computation times for the k-means++ method. The interpretation of these rows is the same as for the random method.

Table 3 is the average number of repetitions required to reach convergence for the k-means method. The second and third rows are the averages for the random method and for the k-means method, respectively.

Table 4 presents the simulated values of the normalized objective function φ. The second to fourth lines are the values for the random method. The second line is the average values of φ. The third and fourth lines are the minimum and the maximum values of φ, respectively. The fifth to seventh lines are for k-means++ method. Each line is the same as that for random method.

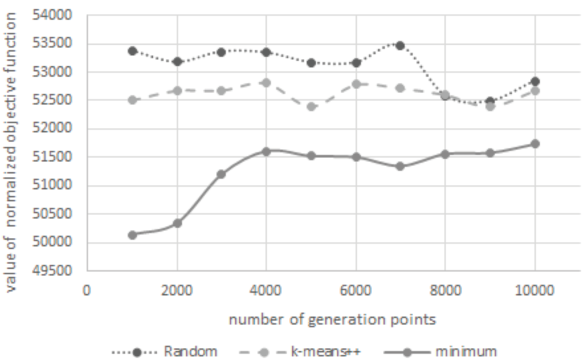

図 2 Values of the normalized objective function

Figure 2 is a graph of the second, fifth and sixth rows of Table 4. The x-axis of Fig. 2 is the number of points n and the y-axis is the value of φ.

4.3 Number of beacons

In this section, we show the results for different numbers of BLE devices, k, in the numerical experiments. We fixed $n = 5000$ and set k at $5, 6, 7, 8, 9, 10, 15$ and 20.

Table 5 shows the mean computation time (msec) for the experiments. The interpretation of the rows is the same as for Table 2.

Table 6 is the average number of repetitions required for convergence with the k-means method. The interpretation of the rows is the same as for Table 3.

Table 7 shows the simulated values of the normalized objective function φ. The interpretation of the rows is the same as for Table4.

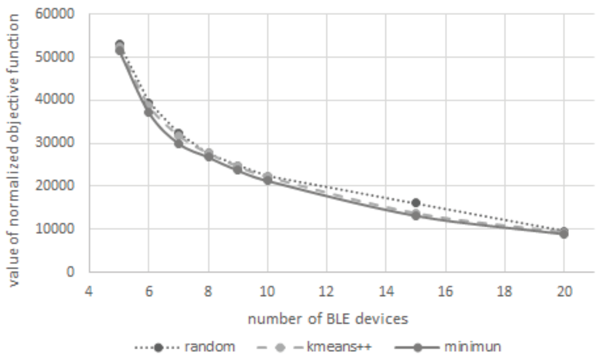

図 3 Values of the normalized objective function

Figure 3 is a graph of the second, fifth and sixth rows of Table 7. The x-axis of Fig. 3 is the number k of BLE devices and the y-axis is the value of φ.

4.4 Validation in a real-world application

We installed 5 BLE devices on the 7th floor of Building 18 in Tokai University. We tried two configurations of BLE devices to evaluate the efficiency of our method. One configuration was ○ as shown in Fig. 1, and the other was ●, which had the minimum value for the objective function among all experiments.

表 8 Three means of accuracy rates

method	first	second	third	#measurement place
○ [7]	0.855	0.489	0.286	136
● (proposed)	0.916	0.533	0.291	126

Table 8 shows the mean accuracy at each location. At each location, we measure the RSSI from the BLE devices and sort the devices based on the RSSI. When the sort order is the same as the order of distance between the place and the beacons, the measurement is correct. The detail of correctness is in [7].

表 3 Mean of repetitions, $k = 5$

n	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
random	17.57	18.22	21.58	23.18	23.94	24.27	23.54	23.54	23.55	23.47
k-means++	16.03	17.32	20.29	21.36	21.56	23.12	21.54	23.03	22.67	23.47

表 4 Values of the objective function, $k = 5$

method	n	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
random	average	53378	53177	53362	53345	53171	53170	53469	52584	52490	52847
	minimum	50144	50348	51206	51613	51527	51507	51352	52584	52490	52847
	maximum	66392	65187	70630	65149	66817	56740	64727	55616	55800	56051
k-means++	average	52509	52680	52685	52813	52398	52784	52721	52605	52397	52677
	minimum	50144	50347	51206	51613	51527	51507	51352	51563	51582	51736
	maximum	63296	65395	64618	65149	56061	56740	64344	55616	55800	56051

5. Discussion

5.1 Number of points

5.1.1 Computation time.

We discuss the computation time for different numbers of points, n. We compare the second (random method) and fifth rows (k-means method) in Table 2, which is the average time for selecting initial points SP from the set P of generation points. While the times for the random method are almost unchanged, those for the k-means++ method are proportional to n, because there are repetitions of the computed distances among P in Step 3, Algorithm 2. Moreover, the computation times of the k-means method for the k-means++ method (sixth row) are almost the same as those for the random method (third row), which also follows from the average number of repetitions (Table 3). Since the computation time for the k-means method is much larger than the selection time, the total computation time for the k-means++ method is almost the same as that for the random method.

5.1.2 Value of the normalized objective function.

The mean values of the normalized objective function for the k-means++ method are smaller than those for the random method. The minimum values for these methods are almost the same between the third and sixth rows of Table 4.

The curve of the minimum values in Fig 2 increases when $n \leq 4000$ and has little change when $n > 4000$. The curve has two local minimum values; $50144(n = 1000)$ and $51352(n = 7000)$. The configuration for the case of $n = 7000$ is illustrated for $\times$ in Fig. 1. The upper three BLE devices had the same position and the lower two were slightly shifted to the left. The value of the objective function for $n = 1000$ reached the minimum for all experiments, illustrated for $\bullet$ in Fig. 1. We approximate the given region $\mathcal{R}$ with points in the proposed method. Therefore, a sufficient number of points are needed to obtain the optimal solution. Methods for determining the value for n are left as an area for future work.

5.2 Number of beacons

5.2.1 Computation time.

The average selection time for the k-means++ method (fourth row, Table 5) increases when the number k of beacons increases. The total times for the k-means++ method are smaller than those for the random method. The reason is that the times of the k-means method for the k-means++ method are smaller than the times for the random method. The difference in these computation times follows from the mean number of repetitions required in the k-means method. Each repetition of the k-means method requires $O(kn)$ distance calculations. The larger the number of repetitions required, the larger the total computation time.

5.2.2 Value of the normalized objective function.

The average values of the normalized objective function decreased when k increased in our experiments (Figure 3). Thus, we obtain the following fact.

Fact: The normalized objective function φ monotonically decreases with increasing number of beacons k.

From this fact, we can vary k to determine the power of the signal, for a given BLE device location. Since the power is inversely proportional to the distance squared, we can calculate a threshold distance from the BLE device within which the device will have sufficient power. The value of φ is the square of the mean distance from the BLE device. Therefore, we increase k until φ is smaller than the square of the threshold.

The mean values for the k-means++ method are smaller

表 5 Mean of computation time (msec), $n = 5000$

method	k	5	6	7	8	9	10	15	20
random	selection	62	49	49	49	48	48	49	50
	k-means	1970	1758	1799	2247	2791	3126	6266	9979
	total	2032	1807	1848	2296	2839	3174	6315	10029
k-means++	selection	232	227	280	327	392	452	892	1413
	k-means	1762	1579	1536	1990	2199	2850	5460	7355
	total	1994	1806	1816	2317	2591	3302	6352	8768

表 6 Number of repetitions, $n = 5000$

k	5	6	7	8	9	10	15	20
random	23.94	26.63	24.02	29.21	33.32	33.99	51.3	64.99
k-means++	21.56	23.57	20.85	29.21	28.07	31.63	43.54	47.15

than those for the random method (Figure 3). The minimum values for the k-means++ method are almost the same as those for the random method when $k = 5, 6, \ldots, 10$. The minimum values for the k-means++ method are smaller than those for the random method when $k = 15, 20$ (third and sixth rows of Table 7).

The maximum values for the k-means++ method (seventh row, Table 7) are much smaller than those for the random method (fourth row) when $k = 15, 20$. This means that the k-means++ method is $\Theta(\log k)$-competitive.

5.3 Measurement accuracy rate

From Table 8, the accuracy rate for the nearest and the second nearest points in the measurement locations for the proposed method is larger than that for the previous method [7]. The increase in the accuracy rate is 6.1% for the nearest and 4.4% for the second nearest points. While the nearest BLE device is useful for position detection, the second nearest BLE device is unreliable. The results suggest that the accuracy of position detection can be increased using ordered order-k Voronoi diagrams and/or the installation of more BLE devices for the real place.

6. Conclusion

We proposed an algorithm for computing $k\,(> 0)$ points in a region $\mathcal{R}$ s.t. the objective function φ has a minimum value for the computed k points (Algorithm 3). This algorithm is based on the k-means method. So, we generate a set P of n points in $\mathcal{R}$, then we apply the k-means method to P. Since the k-means method is approximate, the proposed method also represents an approximation. We numerically evaluated the performance of the proposed algorithm.

We conducted the experiments using two selection methods for the initial points: the random method and the k-means++ method. We varied two parameters in the experiments: the number of generation points n, or the number of BLE devices, k.

Two approximations were considered for the proposed algorithm: an approximation of $\mathcal{R}$, and that in the k-means method. The former approximation is outlined in Step 2, Algorithm 3. First we generated some sets of n points. The larger the number n is, the better the approximation of $\mathcal{R}$ is. In our experiments, we found a local optimal value when $n = 7,000$. Thus, we need to try some n to compute φ for suitable $\mathcal{R}$. The approximation in k-means method considers the selection of initial points. So, we try to use k-means++ method, or 10 SPs in our experiments.

The computation time for the k-means method is short. The execution time for one approximation using the k-means method is about 10.0 seconds for $n = 5,000$, $k = 20$ (random method) and 4.33 seconds for $n = 10,000$, $k = 5$ (k-means++ method). We can thus use much of the available computation time for simulating P, or selection of SP in the k-means method.

When k increases, the computation time is large and the value of φ is small. It was necessary to introduce a method for determining a k which provides an appropriate balance between computation time and accuracy. Using the binary search method, we determined a number k, then computed a configuration of k BLE devices and φ. If φ is smaller than the distance squared depending on the BLE devices, then k is twice. We repeat this step until φ is larger than the square of the threshold distance T. Let K be the number of BLE devices when $\varphi > T$. The optimal number of BLE devices is in between $K/2$ and K. So, we compute the φ for $3K/4 (= (K/2+K)/2)$ BLE devices. If the φ is smaller than T, then the φ for $(3K/4+K/2)/2$

表 7 Values of the objective function, $n = 5000$

method	k	5	6	7	8	9	10	15	20
random	average	53171	39384	32337	27655	24715	22349	16122	9571
	minimum	51527	37127	29931	26739	23656	21258	13195	9004
	maximum	66817	51389	44026	52557	30082	24442	55838	23834
k-means++	average	52398	38860	31842	27655	24567	22138	13647	9204
	minimum	51527	37127	29931	26739	23656	21258	13146	8907
	maximum	56061	52897	44026	52557	26697	23846	15580	10036

is calculated. Otherwise φ for $(K/2 + K)/2$, recursively.

Acknowledgments I thank Professor Yoshimi ISHIHARA, Dean of the School of Science, Tokai University, for allowing the installation of beacons in the Building 18 and enabling this work to be conducted. I also thank Professor Masanori ITAI for his advice regarding the installation of devices. The beacons were supplied by OBFT(http://openbeacon.android-group.jp/). Special thanks to Tomio Ishikawa, a student in my laboratory, for measuring the RSSI values for the beacons.

参考文献

[1] "Bluetooth specifications version 4.0", https://www.bluetooth.org/, June, 2010.

[2] "Tokyo Parks navi", Tokyo Metropolitan Government and Fujitsu, http://www.fujitsu.com/global/about/resources/news/press-releases/2016/0322-02.html .

[3] "Himeji Navi", Available at http://www.himeji navi.net/ .

[4] "Demonstration Experiment of Interactive Security System", http://news.panasonic.com/jp/press/data/2016/02/jn160215-1/jn160215-1.html, *in Japanese*.

[5] "Smart Parking", http://smart-parking.jp/.

[6] Hiroki Ishizuka, Daisuke Kamisaka, Mori Kurokawa, Takafumi Watanabe, Shigeki Muramatsu and Chihiro Ono, "A Fundamental Study on a Indoor localization method using BLE signals and PDR for a smart phone", IEICE Technical Report vol. 114, no. 31, MoNA 2014-10, pp.133-138, *in Japanese*.

[7] Kensuke Onishi, "Indoor Position Detection Using BLE Signals based on Voronoi Diagram", in Proceedings of the 14th International Conference, SoMeT 2015, Communications in Computer and Information System Vol. 532, pp.18-29, Naples, Italy, September, 2015.

[8] Atsuyuki Okabe, Barry Boots, Kokichi Sugihara and Sung Nok Chiu, "SPATIAL TESSELLATIONS Concepts and Applications of Voronoi Diagrams" 2nd Edition, John Wiley & Sons, 2000.

[9] Masao Iri, Kazuo Murota and Takao Ohya, "A fast Voronoi-diagram algorithm with applications to geographical optimization problems", in: P. Thoft-Christensen (ed.), "System Modelling and Optimization", (Proceedings of 11th IFIP Conference, Copenhagen), Lecture Notes in Control and Information Science 59, Springer-Verlag, Berlin, pp.273-288, 1984.

[10] Madhukar R. Korupolu, C. Gerg Plaxton and Rajmohan Rjaraman, "Analysis of local search heuristic for facility location problems", Journal of Algorithms, Vol.37, Issue 1, October, pp.146-188, 2000.

[11] J. B. MacQueen, "Some Methods for classification analysis of multivariate Observations", In Proceedings of the 5th Berkeley Symposium on math. stat. and prob, vol.1, pp.281-297, AD669871, 1967.

[12] Mary Inaba, Naoki Katoh and Hiroshi Imai, "Applications of weighted Voronoi diagrams and randomization to variance-based k-clustering: (extended abstract)". in Proceedings of the tenth annual Symposium on Computational geometry, SoCG 1994, pp.332-339, New York, NY, USA, 1994, ACM Press.

[13] J. Matousek, "On approximate geometric k-clustering", Discrete & Computational Geometry, Vol.24, pp.61-84, 2000.

[14] A. Kumar, Y. Sabharwal, and S. Sen, "A simple linear time $(1 + \varepsilon)$-approximation algorithm for k-means clustering in any dimensions", in Proceedings of 45th Symposium on Foundations of Computer Science, FOCS 2004, pp.454-462, 2004.

[15] R. Ostrovsky, Y. Rabani, L. Schulman, and C. Swamy. "The effectiveness of Lloyd-type methods for the k-Means problem", in Proceedings of 47th Symposium on Foundations of Computer Science, FOCS 2006, pp.165-176, 2006.

[16] David Arthur and Sergei Vassilvitskii, "k-means++: the advantages of careful seeding" in Proceedings of the eighteenth annual ACM-SIAM Symposium on Discrete algorithms, SODA 2007, pp.1027-1035, 2007.

[17] Open Beacon Field Trial, "The First Open Beacon Field Trial", Available at http://openbeacon.android-group.jp/, *in Japanese*.