http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (18): 4544-4548, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.4544.4548

Rss-based Affinity Propagation Algorithm Accuracy Improvement Considering Physical Location

[1]X.H. Chen, L. Niu, Y.J. Zhou, Z. Bi, G. Ding and D. Liang
[1]Beijing University of Posts and Telecommunications, 100876, Beijing, China

Abstract: Affinity propagation is an algorithm which is proposed in recent years. It is an algorithm of broad application and high accuracy. From the perspective of statistic, the clustering error falls into two major categories. One results from environmental distraction, which belongs to a random error. Another part is system error resulting from the shortcoming of the fingerprint-based location technique. We refine this shortcoming as aliasing. Aliasing means that there are several distinct locations but having the same RSS vector. When there are clusters having the same vector as a certain input RSS vector, fingerprint-based technique is not able to indicate which cluster is the correct one. So, how to avoid aliasing error has become a crucial problem to solve. In this study, we proposed an innovative approach inspired by Affinity Propagation and based upon the log-normal shadowing model. The experimental results demonstrate that this approach not only achieves lower location error and quicker coverage speed, but also avoid great error to some extent.

Key words: Indoor positioning, received signal strength (RSS), affinity propagation, path loss model, physical location

INTRODUCTION

In recent years, due to the proliferation of short-distance wireless communications and Wireless Local Area Network (WLAN) technologies, there is a significant growth in Location Based Services (LBS). LBS have been widely used in many areas such as health care, navigation and so on. Accurate positioning technique is a prerequisite and critical component to LBS. (Shravan *et al.,* 2008; Miluzzo *et al.,* 2008).

In the open outdoor environments, the Global Position System (GPS) and cellular wireless communication system can provide highly-accurate and reliable position information. However, because of the serious signal shadowing fading and multipath interference, it can hardly work in indoor environment surrounded by the concrete. (Varshavsky *et al.*, 2007). Under this circumstance, it has attracted a lot of researches from both academia and industry and much substantial work has already been done in the area of indoor positioning. Since the deployment of WLAN infrastructures are widely adopted and the Received Signal Strength (RSS) can be easily obtained and stabilized in multipath and none-line-of-sight-conditions, WLAN based fingerprinting approach is considered now as the best choice for indoor positioning.

Fingerprint-based location technology usually works in two phases: A off-line phase and a location phase. During the off-line phase, the system tabulates the signal strength received from Access Points(AP), stores fingerprint information and establishes the model. During the location phase, the system collects samples of the client in real-time and applies special method in the searching for matching location map.

Clustering algorithm is always used in the off-line phase to group a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups , which in turns reduces the run time in the location phase. K-means algorithm is one of the basic and most simple partitioning clustering techniques. In comparison with K-means algorithm, Affinity propagation takes a different approach to clustering.

In this study, we focus on working out how to improve the accuracy by combing Received-Signal-Strength (RSS) and physical location. Furthermore, we propose an improved affinity propagation algorithm based upon the consideration of physical location. Experimental results show that much higher positioning accuracy is obtained by the proposed algorithm than traditional positioning algorithm. Meanwhile, it can avoid causing great error.

The outline of the rest of the study is as follows. In Section II, problem statement is represented. Section III elaborates on the fundamental and application of affinity propagation algorithm and proposes improvement based upon physical location in detail. The experimental test

Corresponding Author: X.H. Chen, Beijing University of Posts and Telecommunications, 100876, Beijing, China

results are then briefly addressed in Section IV and conclusion will be given in Section V.

PROBLEM STATEMENT

Indoor location system research has mainly focused on maximum matching algorithm. From the perspective of statistic, the location error falls into two major categories. One results from environmental distraction, which belongs to a random error. There are many factors causing impact on this kind of error, such as real environment (open or complex), the time of day(morning, afternoon or night), the density of people, the AP orientation, the acceptance of portable handsets and distance between reference node. This kind of error caused by these factors is time-varying and uncertain. We hence cannot completely eliminate such error. Another part is system error resulting from the shortcoming of the fingerprint-based location technique. In this study, we refine this shortcoming as aliasing. Aliasing means that there are several distinct locations but having the same RSS vector. As is well-known, fingerprint technique relies on the best match between the sample of client RSS vector and the RSS vectors stored in the radio map. So, when there are clusters having the same vector as a certain input RSS vector, fingerprint-based technique is not able to indicate which cluster is the correct one. In other word, aliasing is the source of clustering and positioning errors.(You *et al*., 2010).

Affinity propagation is an algorithm which is proposed in recent years. In comparison with K-means algorithm, Affinity propagation has higher robustness. But the development of more effective affinity propagation algorithm for clustering, other than a few preliminary solutions, is still an open problem. This study mainly concentrates improving the accuracy and robustness of Affinity propagation algorithm by reducing the system error on the consideration of physical location.

AFFINITY PROPAGATION

Affinity propagation is a new and popular algorithm, which makes clustering based upon the similarity between each data point. Rather than specify the number of cluster in advance. Affinity Propagation discovers the number of cluster automatically and simultaneously regards all sample data points as potential centroids.

Affinity propagation implements measurement of similarity between pairs of sample data points, where the similarity indicates how well the sample data point with the index k s suited to be the center of data point i. The most widely used criterion for weighing similarities is the negative Euclidean distance between RSS vectors: For sample data point x_i and x_j:

$$s\ (i, j) = -\|Ri-Rj\|, \forall i, j \in \{1, 2 \ldots, n\} j \neq i \qquad (1)$$

Besides, each data point relates to a preference value that describes how likely the each sample data point is to be a centroid. The greater preference value is, the more possible corresponding sample data point can be selected as a centroid. In addition, the value of preference definitely causes significant impact on the number of clustering. Furthermore, we can denote as preference value. For the lack of valuable prior knowledge, each data point can be as a potential centroid. Hence the preference of each sample data point is set to median similarity, which can be written as follows:

$$P = \text{median}\ \{s(i, j), \forall I, j \in \{1, 2, \ldots, n\}, j \neq i\} \qquad (2)$$

Affinity propagation can be viewed as searching configurations of the labels $C = \{c_1, c_2, \ldots, c_n\}$ to promote the effect of clustering, where the criterion function can be written as follows:

$$\text{Effect}(c_i) = -\sum_{i=1}^{n} s(i, c_i) \qquad (3)$$

The process of message communication between pairs of data point can be viewed as the core of the implementation of affinity propagation algorithm. There are two main kinds of information exchanged between pairs of data point i.e., 'responsibility' and 'availability'. The message of responsibility sent from data point i to potential center j specifies how well data point j suits to serves as the center of point i. The message of availability sent from potential center j to data point i indicates how proper it would be for data point i to select data point k as its center. So data point j with larger values of and are more likely to be chosen as the center of data point i. (Frey and Dueck, 2006).

Affinity propagation is an iterative method for generating high-quality centers by repeatedly transferring message and updating the 'responsibility' and 'availability' of each data point. Subsequently the remaining data points are assigned to the corresponding cluster. Generally, the process of affinity propagation can be carried in three steps:

- **Step 1:** Initialize the availability to zero and update the message responsibilities by the equation:

$$a\ (i, j) = 0 \qquad (4)$$

$$r(i, j) = s(i, j) - \max\{a(i, j) + s(i, j)\} \quad (5)$$

- **Step 2:** Updating the message availabilities by the equation:

$$\alpha(i, j) = \min\{0, r(i, j) + \sum_{i' \neq i, j} \max\{0, r(i', j)\}\}, i \neq j \quad (6)$$

Notably, the self-availability is updated differently:

$$a(j, j) = \sum_{i' \neq j} \max\{0, r(i', j)\} \quad (7)$$

- **Step 3:** Making center decision by combing responsibilities and availabilities. The process of affinity propagation will be terminated after a fixed number of iterations if there is no sufficient change in local decision or the number of iterations reaches the threshold value

The phenomenon of numerical oscillation may result in infinite iteration in affinity propagation clustering. It must be accounted for as in each time of updating message. The rule of numerical oscillation can be written as follows:

$$R_i = (1\text{-lam})^* R_i + \text{lam}^* R_{i-1} \quad (8)$$

$$A_i = (1\text{-lam})^* A_i + \text{lam}^* A_{i-1} \quad (9)$$

where R_i is the current 'responsibility', R_{i-1} is the 'responsibility' in last time, A_I is the current 'responsibility', A_{i-1} is the 'responsibility' in last time and lam$\in$(0.5, 1) is the damp factor of numerical oscillation. For avoiding numerical oscillation as possible, it is important to determine the value of damp factor.

PROPOSED ALGORITHM

In some sense, clustering is the task of grouping in such a way that objects in the same group are more similar to each other than to those in other group. While in data mining, the resulting groups are the matter of interest, their primarily discriminative power is of interest in clustering. In this section, we describe the improved Affinity Propagation algorithm based upon the consideration of physical location. We start by recalling the concept of the path loss prediction model, which is at the basis of our improved propagation.

Path loss is the major component in the analysis of indoor localization in complicated surrounding condition. Path loss normally includes propagation losses, absorption losses, diffraction losses and losses caused by other phenomena. Theoretically, the path loss can be represented as a one-variable function of variable d according to a log-normal shadowing model, is given as (Wang *et al.*, 2009)

$$L = C + 10n\log 10\,(d) + \chi \quad (10)$$

where, L is the path loss in decibels, n is the path loss exponent, d is the distance between the transmitter and the receiver, usually measured in meters, C is a constant which accounts for system losses and χ is the shadowing fading variation and standard deviation is σ.

From Eq. 10, we can find that there must be a definite link between the fluctuations of RSS and physical distance. Under this circumstance, we not only reflect on the factor of RSS but also take the influence of physical location into consideration when computing the similarity between the each pair of data points. The new equation is expressed as follows:

$$\begin{aligned} &\forall i, j \neq i\{1, 2, \ldots, n\}, \\ &s(i, j) = -\| R_i - R_j \|^2 * (10\log_{10}(\| y_i - y_j \|^2)) \end{aligned} \quad (11)$$

where, is the similarity between data point i and j, the x_i, x_j are the n-dimensional RSS vector respectively, y_i and y_j are the actual physical location, $\|R_i\text{-}R_j\|$ is the difference of RSS and $\|y_i\text{-}y_j\|$ is the distance of physical location. Only when the RSS characteristic of data point is similar and its physical location is close can data point can be clustered into the same category. From this perspective, affinity propagation algorithm automatically classifies a collection of data points in such a way that objects in the same cluster are more similar to each other than to those in other cluster.

EXPERIMENTAL RESULT

Test environment: The testbed runs in a 25×28 m laboratory room with two separate rooms in the left-above corner and 3×28 m passage, which is shown in Fig.1.

Test results: This section reports on a number of experiments we conducted to evaluate the proposed algorithm. Our experiments were conducted using a number of synthetic and real datasets. Our main goal was to compare the effect of proposed algorithm with original one.

As is shown in Fig. 2, a collection of data points are grouped more compact than those without regard to the impact of physical location. In other words, the result of the classification of proposed algorithm is more scientific

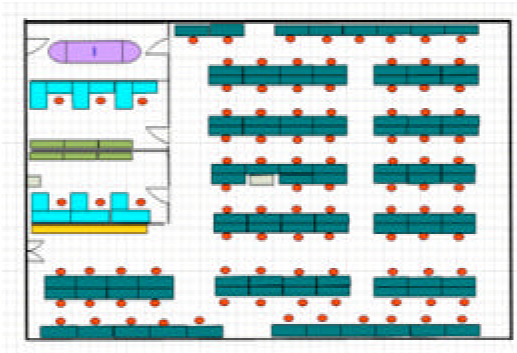

Fig. 1: Topology of the test building

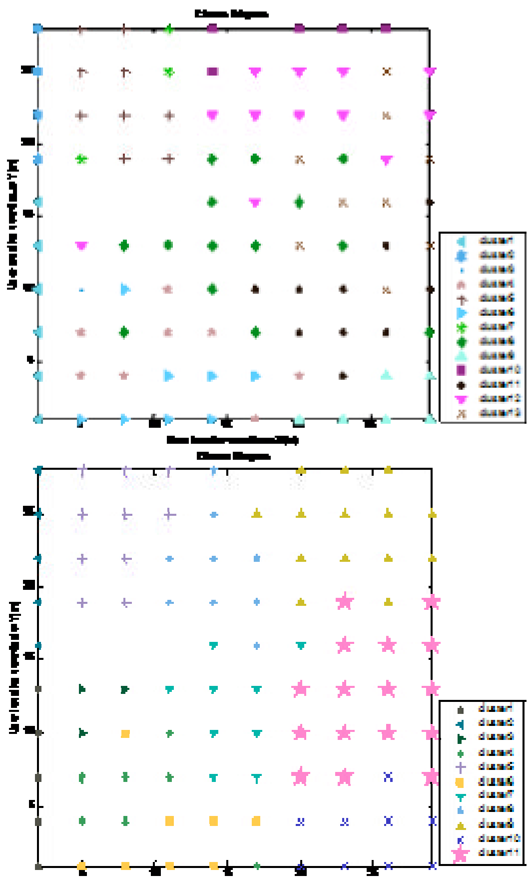

Fig. 2(a-b): (a) Original Affinity propagation and (b) Proposed algorithm, E ach cluster is represented with different color

and reasonable. The negative effect of aliasing which leads to huge error has been avoided to some extent.

Figure 3 shows the CDF (Cumulative Distribution Function) plot of the location error of all the small areas. From the perspective of accuracy of positioning, the

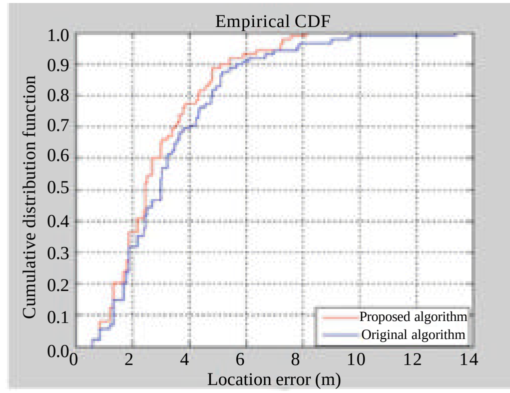

Fig. 3: CDF Diagram

proposed algorithm indeed appears better. Meanwhile, the coverage of proposed algorithm has a significant increase. Besides, it reduces the maximum error.

CONCLUSION

In this study, we have proposed a improved clustering method inspired by Affinity Propagation and based upon the path loss model. The experimental results show that proposed algorithm not only achieves lower location error and quicker coverage speed, but also avoid great error to some extent. Future work includes the theoretical properties of the method and extension of experimental validation to computer vision application.

ACKNOWLEDEGMENTS

This work was supported by the Program for Research Innovation Fund for College Student of Beijing University of Posts and Telecommunications under the Grant 201310013077 and the Fundamental Research Funds for the Central Universities under the Grant2012RC0113.

REFERENCES

Frey, B.J. and D. Dueck, 2006. Mixture modeling by affinity propagation. Proceedings of the Neural Information Processing Neural Information Processing System, December 4-7, 2006, Vancouver, British Columbia.

Miluzzo, E., N.D. Lane, K. Fodor, R. Peterson and H. Lu *et al*., 2008. Sensing meets mobile social networks: the design, implementation and evaluation of the CenceMe application. Proceedings of the International Conference on Embedded Networked Sensor Systems, June 2008, New York, USA., pp: 337-350.

Shravan, G., J. Li, R.R. Choudhury, L. Cox and A. Schmidt, 2008. Micro-blog: Sharing and querying content through mobile phones and social participation. Proceedings of the 6th International Conference on Mobile Systems, Applications and Services, June 2008, Breckenridge, CO., pp: 174-186.

Varshavsky, A., A. LaMarca, J. Hightower and E. de Lara, 2007. The skyloc floor localization system. Proceedings of the 5th Annual IEEE International Conference on Pervasive Computing and Communications, March 19-23, 2007, White Plains, NY., pp: 125-134.

Wang, P., Y.Z. Li and K. Sun, 2009. Path loss modelling and comparison based on the radio propagation measurement at 3.5 Ghz. High Technol. Lett., 15: 272-276.

You, Z., O. Baala and A. Caminada, 2010. Optimization model for an indoor WLAN-based positioning system. Proceedings of the International Conference on Indoor Positioning and Indoor Navigation, September 15-17, 2010, Zurich, Switzerland, pp: 1-7.