Indoor Positioning Using Periodical Analysis

Yansha Guo[1,2]

[1]*School of Information Technology Engineering, Tianjin University of Technology and Education, Tianjin 300222, China*

[2]*Institute of Computing Technology, Chinese Academy of Sciences, Beijing 10090, China*

Yanni Guo[*]

College of Science, Civil Aviation University of China, Tianjin 300300, China

ABSTRACT: LBS (Location Based Service) has been more and more integrated into the various industries, simultaneously, the development of the industry further promote the intensive research and widespread application of LBS. In this paper, we present an excellent resolution for indoor location estimation based on local geomagnetic features. Multi-stage periodical analysis can sufficiently mine the different features for high precision. Initial stage: Kalman filter was used to smoothing the fluctuation of magnetic measurements; Second stage: the filtered data were normalized to construct mean generation matrix of periodical extension, which was the basis for multiple stepwise regression modeling; Third stage: location estimation was finally speculated by KNN (K-nearest neighbor) method in view of the local magnetic characteristics. The magnetic measurements (X,Y,Z) over 82 days, but only the data of 46 days were applied in this paper, in entire implementing processes were from an office environment in Institute of Computing Technology Chinese Academy of Sciences. Our results showed that multi-stage analysis can efficiently estimate the indoor location with an overall accuracy of 83%, completely incorporated the superiority of smart-phone and natural resources, and can be conveniently extended to the relative LBS fields.

KEYWORD: Indoor Positioning; Magnetic Field; Periodical Analysis; Multiple Stepwise Regression

1 INTRODUCTION

With more and more demands of indoor positioning, the research of which is also increasingly become a hot spot. In recent years, scholars home and abroad tried to use infrared, ultrasonic, Bluetooth and WIFI technology for estimating a concrete location based on signal propagation and fingerprint algorithm (Mori S., et al., 2008; Garcia E., et al., 2009; Chen L., et al., 2011; Ouyang R.W., 2012), and many achievements had been made. However, these techniques have one thing in common that a wireless network or additional equipments are required to configure for transmitting and receiving signals, these works maybe increase the device costs and workload before positioning to some extent. Magnetic field as a ubiquitous natural resource, which can be conveniently collected by a magnetometer in smart-phone, may be influenced by some man-made factors (e.g. steel and concrete structure, electric appliances or devices) in addition to its periodical changes. Even so, the magnetic field can be kept relatively stable in a long time; but there also exists a problem that different positions (they may be adjacent or be far apart) may have similar magnetic measurements in one or several dimensions so that the confused location identifications are caused and result in accuracy reduction. In view of it, this paper explores a study on localization estimation through multi-stage periodical analysis: data preprocessing was firstly carried out by using Kalman filtering, and the results of which were standardized for reducing difference range among the dimensions, then mean matrixes were generated by using mean generating function and periodical extension method, prediction model was subsequently constructed by multiple stepwise regression, and the best location can be finally gained by distinguishing the special magnetic characteristics from potential locations.

2 METHODOLOGIES

2.1 *Mean generating function*

Time series: $y(k) = \{y(1), y(2), \cdots, y(n)\}$, and the mean of which can be defined as:

$$\bar{y} = \frac{1}{n}\sum_{k=1}^{n} y(k) \qquad (1)$$

Mean generating function of time series is deferent from the above mean, and can be expressed as the followings (Yang X., et al., 1998):

$$\bar{y}_l(i) = \frac{1}{n_l}\sum_{j=0}^{n_l-1} y(i + j \times l) \quad (i = 1,2,\cdots,l; 1 \le l \le m) \qquad (2)$$

Where n_l is the maximum integer of $INT(n/l)$ and m is the maximum integer of $INT(n/2)$.

$$l = 1: \quad \bar{y}_1(1) = \bar{y} = \frac{1}{n}\sum_{k=1}^{n} y(k) \qquad (3)$$

$$l = 2:$$

$$\begin{aligned}
\bar{y}_2(1) &= \frac{1}{n_2}\sum_{j=0}^{n_2-1} y(1+2j) \\
&= \frac{1}{n_2}[y(1) + y(3) + \cdots + y(1 + 2(n_2-1))]
\end{aligned} \qquad (4)$$

$$\begin{aligned}
\bar{y}_2(2) &= \frac{1}{n_2}\sum_{j=0}^{n_2-1} y(2+2j) \\
&= \frac{1}{n_2}[y(2) + y(4) + \cdots + y(2 + 2(n_2-1))]
\end{aligned} \qquad (5)$$

$\bar{y}_3(i), \bar{y}_4(i), \cdots, \bar{y}_m(i)$ can be achieved in turn according to these ways. $\bar{y}_1(1)$ denotes the mean generating function of 1 interval, $\bar{y}_2(i)$ denotes the mean generating function of 2 intervals. Thus it can be seen that the interval number also indicates the mean count, i.e. 1 interval includes 1 mean and 2 intervals owns 2 means.

2.2 *Mean generating function of periodical extension*

This part will implement the periodical extension based on those deferent means obtained in 2.1 for multiple regression analysis. Eqs.(6) shows the concrete process (Zhao H.Q., 2002).

$$M = \begin{bmatrix}
\bar{y}_1(1) & \bar{y}_2(1) & \bar{y}_3(1) & \cdots & \bar{y}_m(1) \\
\bar{y}_1(1) & \bar{y}_2(2) & \bar{y}_3(2) & \cdots & \bar{y}_m(2) \\
\bar{y}_1(1) & \bar{y}_2(1) & \bar{y}_3(3) & \cdots & \bar{y}_m(3) \\
\vdots & \vdots & \vdots & \cdots & \vdots \\
\bar{y}_1(1) & \bar{y}_2(i_2) & \bar{y}_3(i_3) & \cdots & \bar{y}_m(i_m)
\end{bmatrix} \qquad (6)$$

Where $\bar{y}_2(i_2)$ denotes $\bar{y}_2(1)$ or $\bar{y}_2(2)$, $\bar{y}_3(i_3)$ denotes $\bar{y}_3(1)$ or $\bar{y}_3(2)$ or $\bar{y}_3(3)$,......, $\bar{y}_m(i_m)$ denotes one of $\bar{y}_m(1), \bar{y}_m(2),......, \bar{y}_m(m)$.

2.3 *Multiple regression modeling*

Principle: suppose p independent variables $x_1, x_2, \cdots, x_p$ and one dependent variable y, lineal relationship is existed between them and can be expressed as Eqs. (7).

$$y = c_0 + c_1 x_1 + c_2 x_2 + \cdots + c_p x_p + \varepsilon \qquad (7)$$

Where ε is random error and $\varepsilon \sim N(0,\sigma^2)$, ε, c_0,......, c_p are the estimated parameters. n measurements and the corresponding equations are needed for calculating these parameters. In Eqs. (8), y_i is the i observed value (dependent variable), $x_{1i}, x_{2i}, \cdots, x_{pi}$ (independent variables) are the i observed values, ε_i is the i random error.

$$y_i = c_0 + c_1 x_{1i} + \cdots + c_p x_{pi} + \varepsilon_i, \ i = 1,\cdots,n \qquad (8)$$

3 EXPERIMENTS AND RESULTS

3.1 *Study area and data conditions*

Eight neighboring stations of an office building in Institute of Computing Technology Chinese Academy of Sciences were selected as the experiment sites. The environment is a representative working space with computers, desks, chairs, printers, and some other things. All these items may cause the fluctuated magnetic field in any station at some time. The minimum distance between any two neighbors is about 1m and the maximum distance between them is about 2.4m. The magnetic measurements were collected only with the help of a magnetometer in smart-phone (HTC).

Every sampling started on an hour and obtained 25 records lasting about five seconds in each station. Magnetic measurements in this paper, including three indicators (*X, Y, Z*), were achieved from 9 a.m. to 21 p.m. in weekdays and from 11 a.m. to 16 p.m. at weekend over 82 days. Nevertheless, only the data of 46 days can be used, where the data of previous 37 days (over 73 days) were employed to construct the predict models and the others were as testing data for validation.

3.2 *Analysis process*

Step1: Data preprocess. In order to gain the correspondingly stable data, a series, a total of 325

records with three indicators (X,Y,Z), of 1, 2,…, 25 records of original magnetic measurements in every station at a time point were organized to be preprocessed by Kalman filter.

Step2: Standardization (z-score). Reduce the otherness of three dimensions: X, Y, Z.

Step3: Mean generation of periodical extension. (1) Generating the original periodical mean matrix (73*36, 73 was the number of days and 36 = INT(73/2) was the maximum period or internal days) according to 2.1. (2) Determine the input matrix that will participate in multiple regression modeling: the collected data of real 37 days were selected from 73 generated means by comparing the date between them.

Step4: Multiple regression modeling. The periodical matrix of each dimension in every station at each time of day was as independent variables, i.e. each column was an independent variable; the standardized magnetic measurement was the corresponding dependent variable; and then the predicted model can be constructed by stepwise regression based on 2.3.

Step5: Prediction and comparisons. A location aiming at the current measurements was predicted based on the obtained model, but this result wasn't nicety (experiment had validated it); KNN was introduced to select four locations that were nearest with the predicted value, then one optimal location was finally achieved on the grounds of given magnetic characteristics.

Four magnetic characteristics were chose for distinguishing different stations. In this process, determinations of upper and lower limit of measurements were the key of feature extractions. Because the variation range of magnetic measurements in test environment of this paper was from -300 to 200, 20 units were divided into a group and there were 25 groups. Part of Matlab codes are as followings:

```
function [low,upper]=predictRange(value)
    if value>=0 && value<20
        low=0; upper=20;
    ……
  elseif value<-280 && value>=-200
        low=-300; upper=-280;
    end
  end
```

(1) The first group of feature extractions were on the basis of X dimension

The feature came down to two factors, one was frequency (got the upper and lower limit for current measurements; then computed the count, nx_i, of existed data including in this range; frequency, px_i, can be attained by $px_i = nx_i/Nx_i$, Nx_i is total

existed data) ; the other factor was difference extent, $errorx_i = |x - meanx_i|$, x was the real measurement and $meanx_i$ was the mean of location i. So:

$$featurex_i = pSx_i/errorSx_i \qquad (9)$$

$$pSx_i = (px_i - \min px_i)/(\max px_i - \min px_i)+1 \quad (10)$$

$$errorSx_i = (errorx_i - \min errorx_i) \\ /(\max errorx_i - \min errorx_i)+1 \qquad (11)$$

(2) The second and third groups of feature extractions were based on Y and Z dimensions.

(3) The fourth group of feature extractions were founded on X, Y and Z dimensions.

$$featurexyz_i = pSxyz_i/errorSxyz_i \qquad (12)$$

$pSxyz_i$ was a value that satisfied the upper and lower limit of X, Y and Z dimensions, $errorSxyz_i$ was the mean of $errorSx_i$, $errorSy_i$ and $errorSz_i$.

Rearrange four locations according to those features by Eqs. (13), where d_i got by KNN was the distance of location i and the maximum, r_i, was the optimal location for current measurements.

$$r_i = (pSx_i/errorSx_i + pSy_i/errorSy_i \\ + pSz_i/errorSz_i + pSxyz_i/errorSxyz_i)/d_i \qquad (13)$$

Table 1 showed precision comparisons between the predicted and optimal sites among four locations. In table 1, S denoted "Station", P_L expressed the precision of predicted value, O_L indicated the precision of optimal site that selected from the nearest four locations; "_3, _4 and _5" denoted the set value of introduced and deleted variables, and the both of which were same in the experiments.

Table 1 Precision comparisons between the predicted and optimal sites among four locations

S 1-4	P_L	O_L	S 5-8	P_L	O_L
S1_3	0.409	1	S5_3	0.635	0.962
S1_4	0.442	0.981	S5_4	0.596	0.885
S 1_5	0.462	1	S5_5	0.615	0.885
S 2_3	1	1	S6_3	0.635	0.981
S 2_4	1	1	S6_4	0.75	0.981
S 2_5	1	1	S6_5	0.712	0.981
S 3_3	0.942	0.962	S7_3	0.481	0.981
S 3_4	1	1	S7_4	0.462	0.942
S 3_5	1	1	S7_5	0.5	0.942
S 4_3	0.885	0.981	S8_3	0.712	0.942
S 4_4	0.942	1	S8_4	0.75	0.885
S 4_5	0.981	1	S8_5	0.808	0.885

3.3 *Results*

According to the steps in 3.2, some experiments were executed. One used the original filtered data (no standardization, Ori_Data), the other (SOri_Data) strictly abided the concrete processes. Table 2 displayed the comparisons of both results; "_3, _4 and _5" were the same as Table 1.

Table 2 Results based on data of each hour and mean of hours

Data type	Ori_Data	SOri_Data
Hour_3	0.534	**0.834**
Hour_4	0.534	0.772
Hour_5	0.517	0.769
Mean_3	**0.798**	**0.798**
Mean_4	0.791	0.798
Mean_5	0.796	0.798

In view of Table 2, the result (0.834) of proposed method in this paper owned obviously superiority in indoor positioning. The results based on the mean of hours were relatively stable in despite of the original filtered or standardized data, but the precision was clearly high when the standardized data of each hour were employed. To sum up, standardization can improve the positioning precision by reducing the dimension difference to a certain extent, and the exactness was higher when the introduced and deleted variables were set to 3.

4 CONCLUSIONS

In this paper, a new idea of indoor positioning was proposed: (1) Kalman filter was firstly imported to stabilize magnetic measurements; (2)Standardization would reduce the dimension difference; (3) Mean generation matrixes of periodical extension prepared the independent and dependent variables; (4) Multiple stepwise regression used the results of (3) to construct the prediction models for test data; (5) KNN was introduced to choose an optimal site. In order to validate the whole process, some experiments were implemented and the results depicted the dominance of this proposed method.

Future researches: (1) Representative stations will be determined to collect data by analyzing the environment features; (2) Effectively differentiate the locations with similar magnetic measurements for improving indoor positioning precision.

REFERENCES

[1] Garcia E., Garcia J.J., Hernandez A., et al., "Ultrasonic Positioning System by using UWB techniques," Proceeding ETFA'09 Proceedings of the 14th IEEE international conference on Emerging technologies & factory automation, 2009, 1734-1737.

[2] Chen L., Kuusniemi H., Chen Y.W., et al., "Information Filter with Speed Detection for Indoor Bluetooth positioning," 2011 International Conference on Localization and GNSS (ICL-GNSS 2011) , 2011, 47-52.

[3] Zhao H.Q.. Prediction model of periodical analysis based on the dynamic time series. Dissertation for the Degree of Master, 2002.

[4] Ouyang R.W., Wong A.K.S., Lea C.T., et al., "Indoor location estimation with reduced calibration exploiting unlabeled data via hybrid generative/discriminative learning," IEEE transactions on mobile computing, 2012, 11(11): 1613-1625.

[5] Mori S., Hida K., Sawada K., "Advanced Car Positioning Method Using Infrared Beacon," The 8th International Conference on ITS Telecommunications, 2008,45-50.

[6] Yang X., Zhang R.J.. The problem of periodic extension of mean generating function in regression analysis and its resolution scheme. ACTA METEOROLOGICA SINICA, 1998, 56(4): 493-499.