International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78. 96 | Impact Factor (2015): 6. 391

Location Big Data Sparse Representation and Wavelet Behavior Prediction for Marketing

Yunting Song

Wuhan University

Abstract: ***Location Based Service (LBS) is a new direction and new growth point in the field of geographic information, navigation and network communication. In order to solve the problems of mixed and complex big data mining. This paper proposed a location big data sparse representation and wavelet behavior prediction for marketing. Based on intensity and time difference, we got indoor positioning data intensity. After data clustering and dimensionality reduction, we built a mining model to analyze and extract the district area gathering and flow characteristics. At last, consumption behaviors of users are forecasted correctly.***

Keywords: LBS, Big data, clustering, indoor position

1. Introduction

Location Based Service (LBS) is a new direction and new growth point in the field of geographic information, navigation and network communication. Through the research of content and structure of location information, it can meet the needs of personalized and intelligent location services in the Internet environment.

The advent of the Internet era causes an explosion of information. It also deals with the problem of useful information. Depth content information is most fundamental location big data model. Euclid surveyed a number of location analysis companies that could collect more than six billion of customer data every day, which measuring thousands of location information.

At present, LBS has been used to provide users with appropriate services through location data. Li Deyi proposed LBS based on cloud computing, combined with traditional geographic information system and navigation system, cloud computing can apply to perpendicular. It gives social position and mobile features with continuous time, space diversity, precision seamless, reliable results and online services data service capability. Tan Mengqian put forward LBS big data mining methods based on spatio-temporal analysis. High dimensional position data use data cleaning frequency pruning algorithm for time and space constraint based on time constraints set in the time dimension. According to the division principle and Euclidean distance clustering similarity, it uses K-Means clustering to achieve the owners of location relationship mining. Qiao Shaojie made big data under the environment of mobile object adaptive trajectory prediction model. Prediction model of SATP adaptive trajectory hidden Markov model HMM based on mobile environment, big data clustering method is with density position and high density zoning segmentation. According to the automatic selection of parameters input trajectory, it used object mass trajectory, which reduced the number of HMM.

The LBS big data service technology is generally different from outdoor location data for low precision. With the development of indoor scenarios, indoor location service needs more in reality, such as shopping, meeting, waiting. There are several problems in the following,

a) According to the flow of users in indoor District, as well as during a brand counter retention time, we combine them with other data (such as purchase, payment etc) to effectively predict users for popularity of a brand in the business district.
b) How to get the distribution of the crowd, the flow of consumer behavior around counters.
c) How to use the precise positioning data to obtain the relationship between different shelves and population in the same store.

The common location of big data has obvious diversity and complexity. Based on the high-precision positioning, the users clustering LBS big data has better aggregation, accuracy and analysis. For large flow of people indoor shopping district, the internal counter distribution, aisle density, WIFI facilities, bathrooms and other settings have a direct or indirect impact on the user's consumption behavior. It can effectively improve the utilization rate of resources, promote consumption and so on, which can be used to predict the users behavior habits.

In other to solve the problems of data diversity and complexity, this paper proposed a location big data sparse representation and wavelet behavior prediction for marketing. Indoor positioning data based on intensity and time difference is firstly analyzed to calculate the location of target. Then, after the original date collected, K-means indoor location data furtherly clusters and achieves dimensionality reduction, At last, consumption behaviors of users are forecasted correctly.

2. Indoor Positioning Databased On Intensity and Time Difference

Indoor LBS is the key to the acquisition of the lacation data. Different positioning methods have different data accuracy, which leads to the result of the LBS big data. It can be applied to the analysis of different scenes. Wi-Fi wireless network and RFID fusion positioning is an important source of data acquisition for indoor location services. In this paper, we use combined signal strength and the arrival time difference of composite positioning scheme.

Before localization of Wi-Fi indoor terminal, firstly, we measure the actual situation of interior to plan terminal distribution room. Those have to be stored in the information processing database. Second, we set the AP access point according to the indoor distribution. At the same time, a fixed reference tag is set. As the measurement reference point, it contributes to the position calibration. The number of setting point settings depends on the specific circumstances of the interior. The location algorithm based on signal strength is to locate the same AP access point with a known reference tag and the label of the Wi-Fi terminal to be located. After comparing the position of the reference label with known value, the correction value of the positioning data of the reader is obtained, and then the correction value is sent to the label to be positioned in time. Fix the location information measured by the label S, which uploads to the reader. This process remains only a reader, the reader in a dormant state. When a reader completes the position information acquisition, the other reader completes the same process. Positioning mode under the need for at least are 3 times position information acquisition. If the signal intensity formula is related to the distance between the reader and the tag, the simple signal propagation model can be expressed as,

$$P(r) = P'(r) - 10alog(r/r_0) \quad (1)$$

Where, $P(r)$ is received signal power, r is the distance between the tag and the reader, r_0 is reference distance to reader, $P'(r)$ is signal power of the reference point, parameter a path loss increases with the increase of the distance r.

The time difference of arrival (TDOA) algorithm is used to measure the time difference of the received signal of the same label, and the distance difference between the tags and the reader is calculated. The difference D between the label to any two readers is a fixed value. The label must be located in the two reader as the focus of the hyperbolic, when there are both a N reader in the distance ($N \geq 3$), by the confluence area between a plurality of hyperbola is to estimate the position of the label.

TDOA is to measure the arrival time difference of the same label location signal received by each reader. The reader participating in the positioning does not require strict synchronization in time.

It is assumed that the measurements of the i reader and the j reader arrival time were TA_i and TA_j. The time difference between the arrival of the signal to the i and the j reader is $TA_{ij} = TA_i - TA_j$. Their distance difference are $R_{ij} = C * TA_{ij}$. There is a relationship between the label coordinates and the reader coordinates,

$$\sqrt{(X_i - x)^2 + (Y_i - Y)^2} - \sqrt{(X_j - x)^2 + (Y_j - Y)^2}^2 = R_{ij}^2 \quad (2)$$

3. K-means Indoor Location Data Clustering and Dimensionality Reduction

A. Multi order features of based on LBD

The primary task of big data mining analysis position from the individual user location and behavior intensity and time difference data is extracted based on value at different time, establish some characteristic models of single area s_i and moving objects p_i. According to the characteristics of indoor location information, we can get the following three kinds of feature models:

a) Zero order characteristic: A moment of user location data, using p to indicate that $p(id, x, y, z)$. Map information within the region with m to express $m(id, x, y, z, k)$, where x, y, z said the regional center coordinates, k said the size of the area. For a specific place (such as airports, conference site, large shopping malls, etc.), will be in all areas of the p point based on the time to gather the space to get a moment of the location data set $P\{p_1, p_2, p_3, \ldots, p_k\}$.
b) One order characteristic: From the specific location data in the specific region, the map data can be simply calculated features, such as the mean, variance, etc. .
c) Second order characteristic: It is necessary to obtain some characteristic information which can be obtained by some high order statistical processing. The aggregation of those information can wipe off the complexity and diversity to some extend.

B. Clustering and dimensionality reduction based on K-means

Because indoor district area is far less than the outdoor area, location characteristics of big data in the indoor user groups have jumbled. However, the situation of indoor location data that the area of it is small, and the similarity of the location data is high. After the indoor location data clustering, the cluster center may be characterized by dense distribution. There is a need to adopt a clustering center dimension reduction model processing. The clustering algorithm can be used to analyze the location of indoor users, which is convenient for visual analysis of map matching. Support clustering is to find out the common features of the data objects through the classification model, and through the semantic analysis of the data item mapping into the corresponding categories.

The consistency preserving K-means algorithm is used in the clustering dimensionality reduction, which is shown in figure 1.

First step: Initialize, randomly select K points as the center of the initial class$(c_1, c_2, \ldots, c_k)$.

Second step: Assign a nearest neighbor set S, assign the S to its nearest class C_p, then,

$$Sp = argmin_{v=1,\ldots,k} \sum_{x_i} (x_i - m_v)^2 \quad (3)$$

Third Step: Update Center to update the clustering center (i. e., centroid) m_v is the center of $C_v, n_k = |C_k|$, is,

$$m_v = \sum_{x_i \epsilon C} x_i / n_v \quad (4)$$

Fourth Step: convergence judgment, through the $J_k = \sum_{v=1,\ldots,k} \sum_{x_i \epsilon C} (x_i - m_v)^2$ convergence judgment, if the centroid is no longer moving, then terminate the algorithm, otherwise return to Second step.

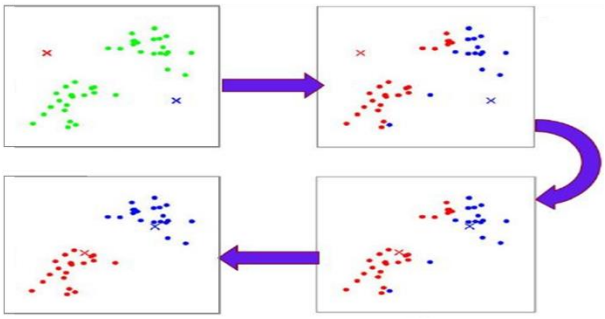

Fig 1: Consistency preserving K-means algorithm

4. Precision marketing based on wavelet behavior prediction

Map matching is a unique part of the preprocessing on big data, which is of great significance to the user experience of data visualization. In this paper, algorithm based on wavelet neural network includes three levels: input layer, hidden layer and output layer. The features extracted from LBS big data in the location are associated with a specific user information, such as personal identity information, health status, behavioral data, etc. It is very easy to set up probabilistic graph model on time scales. Therefore, the collaborative mining on time scale can choose wavelet model to deal with.

Where x represents the hidden state, y represents the output that can be observed, a represents the transition probability of the state, that is, the probability of the feature set on the time scale, and b represents the output probability. According to the experience, the majority of users in the indoor location interaction, high precision position information generated after a series of indoor location feature extraction can be obtained after the hidden state based on x, and the corresponding y can be viewed as a consumer behavior of users.

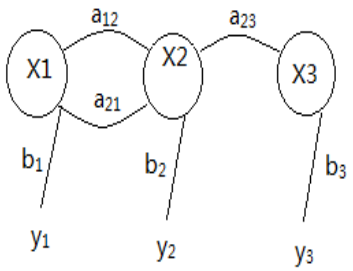

Fig 2: Decision making process of wavelet model

As Figure 3 is the effect of the position of a time, the actual indoor places after data preprocessing, the target time in the user data area is 230, it has reference value for the position analysis of population aggregation at different time. The small points represent the user's real-time location data clustering points. According to the map matching, we can get the specific indoor map and user location model. Further, we can get a higher level of clustering center by the way of cluster mining and importance screening, as shown in Figure 3, the line of three.

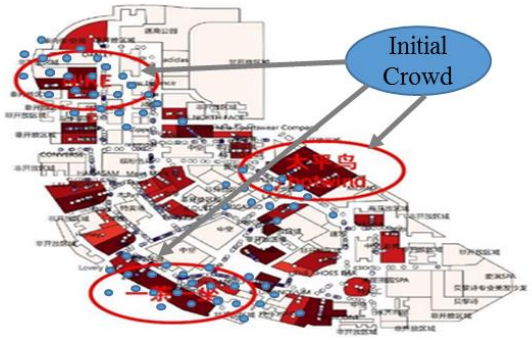

Fig 3: A large shopping mall LBS map for T3 layer

5. Conclusion

This paper proposed a location big data sparse representation and wavelet behavior prediction for marketing. Indoor positioning data based on intensity and time difference is firstly analyzed to calculate the location of target. After data clustering and dimensionality reduction, we built a mining model to analyze and extract. At last, consumption behaviors of users are forecasted correctly to solve the problems of data diversity and complexity

References

[1] Akram Syed Ali, Zachary Zanzinger, Deion DeBose, Brent Stephens. Open Source Building Science Sensors (OSBSS): A low-cost Arduino-based platform for long-term indoor environmental data collection [J]. Building and Environment, 2016
[2] Antonio Ferreira, Z. Daraktchieva, D. Beamish, C. Kirkwood, T. R. Lister, M. Cave, J. Wragg, K. Lee. Indoor radon measurements in south west England explained by topsoil and stream sediment geochemistry, airborne gamma-ray spectroscopy and geology [J]. Journal of Environmental Radioactivity, 2016
[3] Iker Perez, James Pinchin, Michael Brown, Jesse Blum, Sarah Sharples. Unsupervised Labelling of Sequential Data for Location Identification in Indoor Environments [J]. Expert Systems With Applications, 2016
[4] Seung Ho Ryu, Hyeun Jun Moon. Development of an occupancy prediction model using indoor environmental data based on machine learning techniques [J]. Building and Environment, 2016
[5] Y. Wang, L. Shao. Understanding occupancy pattern and improving building energy efficiency through Wi-Fi based indoor positioning [J]. Building and Environment, 2016
[6] Xin Wang, Hai Dong Chi, Tao Zheng, Wen Li Cheng. One Data Transmission System for the Indoor Wireless Communications [J]. Advanced Materials Research, 2013, 2542(760)