센서 레지스트리 시스템에서 효율적인 센서 필터링을 위한 LSTM 기반 모델

천하오티엔[1,2] · 정현준[1] · 이석훈[1,*] · 온병원[1] · 정동원[1,*]
[1]군산대학교 · [2]하북지질대학교

LSTM-based Model for Effective Sensor Filtering in Sensor Registry System

Haotian Chen[1,2] · Hyunjun Jung[1] · Sukhoon Lee[1*] · Byung-Won On[1] · Dongwon Jeong[1*]
[1]Kunsan National University · [2]Hebei GEO University
E-mail : chenhaotian@hgu.edu.cn / {junghj85, leha82, bwon, djeong}@kunsan.ac.kr

요 약

센서 레지스트리 시스템은 센서와 디바이스 간 상호운용성 문제를 해결하기 위하여 모바일 디바이스의 위치 정보에 기반하여 센서에 대한 의미적 메타데이터를 제공한다. 하지만 모바일 디바이스의 GPS가 잘못 수신되면 센서 레지스트리 시스템은 잘못된 센서 정보를 받게 되며 센서와 연결할 수 없다는 문제를 지닌다. 이 논문에서는 이러한 문제를 해결하기 위해 모바일 장치와 센서 간의 성공적인 요청 확률을 향상시키기 위해 지리적 임베딩 및 LSTM 기반 경로 예측을 기반으로 한 이중 협업 전략을 제안하고 몬테카를로 방법을 이용하여 평가한다. 실험을 통하여, 제안한 방법이 위치 이상 문제를 개선하고 효과적인 멀티캐스팅 메카니즘임을 보였다.

ABSTRACT

A sensor registry system (SRS) provides semantic metadata about a sensor based on location information of a mobile device in order to solve a problem of interoperability between a sensor and a device. However, if the GPS of the mobile device is incorrectly received, the SRS receives incorrect sensor information and has a problem in that it cannot connect with the sensor. This paper proposes a dual collaboration strategy based on geographical embedding and LSTM-based path prediction to improve the probability of successful requests between mobile devices and sensors to address this problem and evaluate with the Monte Carlo approach. Through experiments, it was shown that the proposed method can compensate for location abnormalities and is an effective multicasting mechanism.

키워드

Path Prediction, LSTM, Monte Carlo, Sensor Registry System

Ⅰ. Introduction

In recent years, intelligent services based on ambient sensors for mobile devices have developed rapidly. For such services, the Sensor Registry System (SRS) based on ISO/IEC 11179 provides semantic metadata between sensors and devices to resolve the interoperability problem [1]. In such a mobile computing environment, there are two connections in the data delivery process. When a user makes a service request, the SRS sends semantically consistent metadata of sensors to the user's mobile device through the first connection and the mobile device, through multicasting, collects data from the sensors via the second connection. With these two connections, the mobile device can acquire heterogeneous data by metadata for providing intelligent services.

Existing optimization methods focus on the first connection. There is a denial-of-service area in which the mobile device cannot communicate with the SRS. Suppose the mobile device cannot obtain metadata

* corresponding author

from the SRS, it also cannot acquire data from the ambient sensors, and thus the service request fails. Existing studies [2-4] focus on path prediction algorithms to improve the reliability of acquiring metadata. The trajectory history is used to predict the next position, and the metadata of the filtered sensors are obtained before entering the denial-of-service area. However, these efforts do not consider the fact that the position obtained from GPS is potentially wrong. Meanwhile, the performance of the SRS is not quantitatively evaluated.

In this work, we focus on the multicasting mechanism of the second connection. We propose a dual collaboration strategy based on geographical embedding [5] and LSTM-based path prediction to improve the probability of successful request between mobile devices and sensors. The position from the GPS and that obtained by the path prediction algorithm are simultaneously used to send the request. In this way, if one of the two positions is correct, the request will be successfully served. We also propose a new approach for estimating service provision rates based on Monte Carlo to measure the performance of the SRS.

II. Methodology

The path prediction algorithm is an essential part of location-based intelligent services. We capture participants' positions in their daily lives to compose the dataset, a collection of spatial-temporal points. It is further mapped to segments [2] or cells [4], thus converting the task into a time series multi-classification problem. Data preprocessing techniques, such as noise filtering and sliding windows, are also necessary for path prediction tasks.

In this work, we employ LSTM, which is designed to solve the gradient vanishing problem for processing sequential data. The model consists of a geographical embedding layer [5], an LSTM layer, and a softmax layer.

Service provision rate is the probability of successfully provided services to a mobile device when they are requested. To measure the service provision rate, we have developed a simulator base on the Monte Carlo method, which uses historical trajectories to simulate the performance of the SRS. The communication range of the sensors is simulated by randomly generating fixed radius circles in Geographic Information System (GIS). If the position of a mobile device is within the circle, it can communicate with the sensor. It implies that we can use the percentage of the shaded area inside the cell to indicate the probability that the user gets a provided service in that cell. Following the dual collaboration strategy, the cell from the GPS and the cell obtained from the path prediction are used simultaneously to request the service with the sensor corresponding to the actual position. A successful one of them indicates a successful service. The simulation process using historical trajectories is repeated several times by randomly generating sensing networks. Finally, statistically average the results of the random investigation to obtain the service provision rate.

III. Experiment

The ubiquitous smart devices with GPS make the collection of personal trajectories more efficient and effective. The entire GPS trajectory dataset collected by 59 users in six months contains over 900 million spatial-temporal points. After data preprocessing [6], the filtered data consisted of 240 million spatial-temporal points that make up 35,234 trajectories with a total distance of 1,588 km and 3,300+ hours. It recorded a broad range of the participants' daily activities.

We evaluate the LSTM model with vanilla RNN, CBP [2], and GatedCNN [7]. LSTM outperforms the other methods, yields about 23% improvement than CBP, which is the baseline for path prediction-based SRS in previous work. Compared with the performance of the vanilla RNN, we find that the macro-recall and weighted-call are about 50% and 38% better, respectively. For directly stacking GatedCNN layers, the LSTM model also has approximately 15% improvement on weighted-recall.

CBP has reached the useable level. The values of weighted-recall and macro-recall are significantly different. The distribution of the dataset is imbalanced. Recall of cells that frequently appear in the dataset is higher than the recall of cells that occur less often. The difference will be produced after multiplying by the weight. GatedCNN is also a feasible method. However, the weighted-recall of GatedCNN is higher than CBP, and the macro-recall is lower. It shows that the two ways capture different movement patterns. The CBP method is based on the greedy algorithm and only considers the current cell, and does not dependent on the whole trajectory sequence. In contrast, GatedCNN extracts all the information of the entire trajectory through staked convolutional layers, overfits more than CBP and LSTM on small datasets, improves the overall accuracy but loses the ability to generalize various behavior patterns. LSTM determined a further increment not only in weighted-recall but also in macro-recall, greatly improved performances.

Besides, we examine the performance by the Monte Carlo method. The ideal scenario of a completely accurate mobile device position determines the upper limits of the service provision rate, and the lower limit is the case when GPS positioning is sometimes wrong. The service provision rate reinforced by the dual collaboration strategy is distributed between the ideal scenario and the actual situation.

The four path prediction methods exhibit distinct characteristics. LSTM has the best performance. By applying the dual collaboration strategy with LSTM, the effect of mobile device position error can be compensatory. The evaluations are close to ideal under a variety of coverage conditions. The predictions of the CBP algorithm are chosen from the eight cells surrounding the input cell. The true cell is not always adjacent to the input cell or the most frequently visited one. The assumption of the frequency of nearest cell occurs limits the accuracy of CPB. However, considering the movement speed of participants, the true cell is generally near the input cell. Even if the predicted cell is not the same as the true cell, it is still near the true cell in geographical space. The true sensor may cover both true and predicted cells. In evaluating service provision rate, the possibility improves the performance of CBP, which is second to the ideal scenario and LSTM model, higher than GatedCNN and vanilla RNN.

Ⅳ. Conclusion

In this paper, we investigated the sensor filtering problem of GPS displaying the wrong position. We proposed a dual collaboration strategy in multicasting mechanisms in the mobile computing environment. It integrates positions obtained from GPS and predicted by past trajectory to enhance the service provision rate in the SRS. The LSTM model learns movement features from long-range cell sequences and is an effective grid-based path prediction algorithm. A Monte Carlo-based simulation flow is presented for evaluating the service provision rate. Path prediction algorithms with various ideas and structures focus on different characteristics of historical data. These features will produce results in the SRS with a different performance from accuracy. For our dataset, LSTM and CBP have the best performance.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIP) (No. 2019R1I1A3A01060826).

References

[1] D. Jeong, “Framework for seamless interpretation of semantics in heterogeneous ubiquitous sensor networks,” *International Journal of Software Engineering and Its Applications*, Vol. 6, No. 3, pp. 9-16, Jul. 2012.

[2] S. Lee, D. Jeong, D. Baik, and D. Kim, “Path Prediction Method for Effective Sensor Filtering in Sensor Registry System,” *International Journal of Distributed Sensor Networks*, Vol. 11, No. 7, Jun. 2015.

[3] H. Jung, D. Jeong, S. Lee, B. On, and D. Baik, “A Network Coverage Information-Based Sensor Registry System for IoT Environments,” *sensors*, Vol. 16, No. 8, Jul 2016.

[4] H. Jung, D. Jeong, and S. Lee, “Development of Sensor Registry System-Based Predictive Information Service Using a Grid,” *sensors*, Vol. 18, No. 11, Oct. 2018.

[5] A. Crivellari and E. Beinat, “From Motion Activity to Geo-Embeddings: Generating and Exploring Vector Representations of Locations, Traces and Visitors through Large-Scale Mobility Data,” *ISPRS International Journal of Geo-Information*, Vol. 8, No. 3, Mar. 2019.

[6] S. Kim and S. Lee, “Clustered Noise Filtering Based Trajectory Refinement Method,” *The Journal of Korean Institute of Information Technology*, Vol. 18, No. 3, Mar. 2020.

[7] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, “Language Modeling with Gated Convolutional Networks,” in *Proc of The 34th International Conference on Machine Learning*, Sydney: NSW, pp. 933-941, 2017.