eISSN:2600-7920
INTI JOURNAL
Vol.1, 2018 (28)

Intelligent Agent Technology for Cellular-Assisted GPS Positioning using Bayesian and Self-Organizing Map Approaches

Wou Onn Choo[1], Lam Hong Lee[2], Yen Pei Tay[2], Khang Wen Goh[2], Wen Yeen Chue[3], Suliman Mohamed Fati[1]

[1]Faulty of Information Technology and Science, INTI International University, Nilai, Negeri Sembilan, 71800, Malaysia
[2]Faculty of Science and Technology, Quest International University Perak, 30250 Ipoh, Perak, Malaysia.
[3]Faculty of Business and Information Science, UCSI University, 56000 Kuala Lumpur, Malaysia.

Corresponding Author: wouonn.choo@newinti.edu.my, suliman.gaber@newinti.edu.my

Abstract

An intelligent agent equipped with cellular-assisted Global Positioning System (GPS) positioning has been proposed in this paper. The positioning technique has been enhanced by using Bayesian and Self-Organizing Maps approaches. Due to the overlapping of coverage areas of cellular towers, conventional cellular-based positioning techniques have been reported to be inaccurate. Current cellular-assisted GPS positioning techniques are cost ineffective as extensive investments on hardware deployments are required in order to achieve the highly accurate positioning performance. A relatively low cost approach is presented in this paper for more economical and satisfactory cellular-assisted GPS positioning. Raw location information, in the form of cellular identity (ID) and GPS coordinate pairs, are acquired by using equipment such as smart phones and GPS trackers. These raw information were categorized into categories according to the distribution patterns of cellular towers. The cellular ID and GPS coordinate pairs were further grouped within each of the individual cellular IDs. An intelligent software agent equipped with data mining capabilities was then deployed to collect and process the device-coordinates in order to predict the optimal GPS coordinates of the cellular ID. This results to the determination of a virtual cellular tower location for each cellular ID, to provide more precise location positioning. Experimental results show that the prediction of location using GPS coordinate of cellular IDs helps in improving the contemporary cellular-assisted GPS positioning technique to sub-kilometre accuracy.

Keywords

Cellular-Assisted GPS Positioning; Bayesian; Self-Organizing Map

Introduction

Positioning mobile devices had become popular for mobility tracking, since most modern mobile computing devices are equipped with both Global Positioning System (GPS) and cellular-based positioning. Due to high-rising crime rate in many countries, the tracking of humans and assets have gained importance over the past decades by providing better monitoring of movements of targeted people and goods, in both known and unknown environment. Human and asset mobility tracking devices which are specifically designed and developed for this purpose, have resulted in sufficient tracking using conventional GPS technology. However, most of the conventional GPS trackers do not cater for accurate cellular-based positioning, which assists the tracking process in the situation where GPS signal is out of coverage, such as indoor environment. Most modern mobile computing devices such as smart phones come with cellular-based positioning feature, which still enable tracking of the device even when GPS links cannot be established. However, conventional cellular-based positioning technique has been reported to be less effective, with highly degraded accuracy by factors such as cellular tower occupancy, cellular towers distributions and the design of cellular chipset in the device [1]. This problem is due to the fact that there are overlaps in the signal coverage of cellular towers, hence leading to non-reliable location information resulted from conventional cellular-based positioning technique, i.e. when no GPS is involved. In real world applications where cellular-based positioning techniques are deployed, the tracking devices are not always being connected to the nearest cellular towers. This situation is caused by several factors, such as the load of the nearest cellular towers, the network switching mechanism implemented on the devices, and the signal strength of the cellular towers nearby. Such situations often lead to inaccurate positioning using cellular-based technology. Techniques which involving intensive infrastructure and financial investments are accurate, but they are poor in economic efficiency, hence these techniques failed to gain popularity in the commercial deployment.

Various techniques have been introduced over the past decades to improve the accuracy and effectiveness of cellular-based positioning [2-5]. However, many of them failed to address the fundamental issue of low accuracy of cellular-based positioning in the absence of GPS and WLAN networks. Simulated Annealing method was used to improve the accuracy of cellular-based positioning to identify the locations and ranges of transmission of base stations to accomplish the best promising location accuracy [6-7]. The work was carried out by using an irregular structure of networks compared to the previous work which focused on the mesh and hexagonal structures. A probabilistic fingerprinting localization technique was implemented which is more suitable for cellular phone localization [6].

An intelligence software agent which is equipped with data mining capabilities, is proposed to improve the accuracy of cellular-based positioning technique. This can be achieved by analyzing the GPS coordinates within the coverage of the cellular towers, predicting and extracting the significant GPS coordinates for each cellular identity (cell-ID), hence minimizing the overlapping areas within the signal coverage of different cellular towers. The raw data of cell-ID and GPS coordinate pairs are collected from heterogeneous mobile devices such as smart phones and GPS trackers, among the community of the system users. The collected GPS coordinates from each cell-ID are then processed using probabilistic classification and unsupervised clustering techniques, in order to predict the optimal coverage of the cell-ID with minimum overlapping with the others. In our study, the feasibility of Bayesian probabilistic approach [8] and Self-Organizing Maps (SOM) [9] in processing the raw data is investigated. Intelligent agent technology is utilized in acquiring raw data, processing the data and resulting in more precise coverage prediction of the cell-IDs. By having an intelligent agent software in carrying out the data mining process in conjunction to the human and asset tracking system, data sharing and data mining processes can be conducted through negotiations on a predefined set of criteria, without affecting the normal operations and distorting the original data of the human and asset mobility tracking system. Intelligent agent technology is introduced with the main objective of automating the selection and configuration of criteria in processing and optimizing the data. These criteria include, but are not restricted to the difference in distribution of the cellular towers in different areas, the difference in load on the cellular towers at different times and days, and the abnormal load on certain cellular towers during special events. By considering these criteria in the data mining process, raw data which are collected in different periods can be handled differently, hence resulting in more accurate prediction and more optimal coverage of the cellular based positioning.

The experimental results show that the intelligent software agent which is equipped with Bayesian probabilistic classification technique and SOM clustering technique successfully improve the performance of the conventional cellular-assisted GPS positioning technique. This technique contributes to a more accurate cellular-based positioning, without incurring extensive financial investment on the existing hardware and infrastructure, hence results to a cellular-assisted GPS technique with high economic efficiency.

Intelligent Agent for Cellular-Assisted GPS Technique

The proposed intelligent agent optimizes the coverage area of cellular towers by implementing Bayesian probabilistic classification approach and SOM clustering technique. Based on the GPS coordinates within the coverage of a particular cellular tower, which were acquired from the GPS devices, the intelligent agent software optimizes the set of GPS coordinates to be annotated to the cellular towers, resulted to the "Virtual Cellular Tower GPS Coordinate" in providing a more precise cellular-assisted GPS positioning technique. Figure 1 illustrates the block diagram of the proposed intelligent agent.

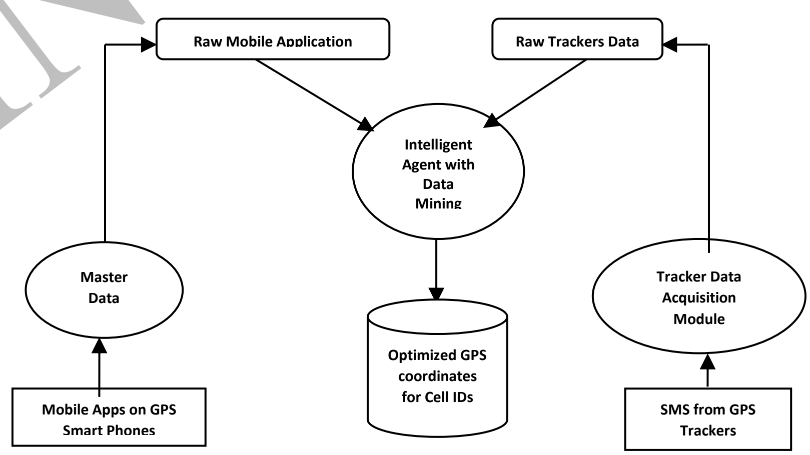

Figure 1: Block diagram of Intelligent Agent for Enhanced Cellular-Assisted GPS Positioning

The acquired cell-ID and GPS coordinate pairs were grouped to build a training set for the Bayesian classifier. Each GPS coordinate was then being computed using Bayes theorem to get the probabilistic distribution of the GPS coordinate across the available cell-IDs. To optimize the coverage area of a particular cellular tower, a predefined threshold was configured to eliminate the GPS coordinates with low probability to the cellular tower and retain those with high probability. The optimum value of the threshold was determined by using cross validation process. One of the reasons which contributes to the inaccuracy of location returned by cellular-assisted GPS positioning devices is that, these devices are not connected to their nearest cellular towers. By eliminating the GPS coordinates with low probability of annotation to a particular cell-ID, the problem of having inaccurate cell-ID information returned by the devices can be minimized.

In this study, SOM clustering technique was used to overcome the problem of having overlapping coverage from neighboring cellular towers. All the GPS coordinates acquired from the devices were fed into the SOM model as vectors. The GPS coordinates of the cellular towers were treated as the input vectors of the map and then the Best Matching Units (BMUs) were computed to distinguish each of the clusters. The number of clusters produced by the SOM model was set to be identical to the number of cellular towers captured by the GPS positioning devices. By using SOM approach, the GPS coordinates which are within close proximity will be grouped under the same cluster and the problem of overlapping coverage from different cell-IDs can be eliminated.

The data were then categorized into groups after being processed by Bayesian approach and SOM, and later be labeled with the cell-ID which forms the majority of the GPS coordinates in it. These labels are useful in predicting the location of a tracker, when the tracker returns only the cell-ID, under the situations where GPS signal is not available.

After the raw data was processed by the intelligent agent, the optimized data were stored in a database for location prediction based on the positioning requests from the users. The better cellular-assisted GPS positioning accuracy can be obtained by minimizing the impact of the factors such as overloaded cellular towers, improper network switching mechanism of the devices, weak signal strength of the cellular towers, and overlapping in coverage areas of the neighboring cellular towers. The optimized data resulted from the intelligent agent contribute to better performance of cellular-assisted GPS positioning for the tracking system, when the trackers are out of other positioning network coverage, such as GPS and WLAN, as compared to the conventional techniques.

Experiments and Evaluations

In the experiments and evaluations of the performance of the proposed intelligent agent, raw data of cell-ID and GPS coordinate pairs have been collected from various trackers and mobile phone applications from various locations within Klang Valley, Malaysia. A prototype system has been developed for this purpose.

In order to evaluate the performance of the proposed intelligent agent, a reference GPS coordinate has been computed for each of the groups/cell-IDs. This reference GPS coordinate is computed by calculating the mean of longitude and the mean of latitude for all the GPS coordinates clustered in the same group. This is due to the reason that the collected data are skewed due to the uneven distribution of the trackers. This situation reflects the real life situation of uneven distribution of human and asset under the coverage of a cellular tower. Hence, mean calculation is

eISSN:2600-7920
INTI JOURNAL
Vol.1, 2018 (28)

selected to cover the area of low tracker population density.

After the reference GPS coordinate of the groups/cell-IDs has been identified, the geographical distance between each GPS coordinate and the reference GPS coordinate in the same group was then computed by using Haversine distance function as described below. Let GPS coordinate for Point 1 be (φ_1,λ_1) and GPS coordinate of Point 2 be (φ_2,λ_2),

$$\Delta\varphi=\varphi_2 - \varphi_1$$
$$\Delta\lambda=\lambda_2 - \lambda_1$$
$$a = \sin^2(\Delta\varphi/2) + \cos \varphi_1 \cdot \cos \varphi_2 \cdot \sin^2(\Delta\lambda/2)$$
$$D = 2R \cdot \arcsin(\sqrt{a})$$

where D is the distance in meters between Point 1 and Point 2, and R is the radius of the earth, which carries the value of 6371009 meters.

Table 1 presents the reference GPS coordinate for each of the groups/cell-IDs, and the maximum distance, minimum distance and average distance among all the GPS coordinates in the same group to the respective reference GPS coordinate.

Table 1: Reference point for each cell ID, and its maximum distance, minimum distance and average distance among all the GPS coordinates in the same group

Cell ID	MCC	MNC	LAC	Pred. Long.	Pred. Lat.	Area	Max. Dist. (m)	Min. Dist. (m)	Avg. Dist.(m)
7107556	502	12	23500	101.6233	3.0257	Bdr Puteri Puchong	313.21	5.54	17.59
7107551	502	12	23500	101.6233	3.0256	Bdr Puteri Puchong	5.39	0.24	2.32
7107554	502	12	23500	101.6233	3.0257	Bdr Puteri Puchong	60.68	0.55	2.79
7107549	502	12	23500	101.6233	3.0257	Bdr Puteri Puchong	294.03	3.35	10.33
7091561	502	12	23500	101.5701	2.9957	Putra Height	201.94	74.92	148.54
7091565	502	12	23500	101.5695	2.9932	Putra Height	337.04	44.54	116.89
7033329	502	12	22500	101.7003	3.0706	Sri Petaling	231.56	6.40	23.26
59307	502	12	2111	101.5692	2.9927	Putra Height	1303.71	19.32	59.93
7033324	502	12	22500	101.7003	3.0707	Sri Petaling	9.66	2.51	4.46
7097571	502	12	23500	101.5691	2.9931	Putra Height	13.00	0.37	0.91
7033308	502	12	22500	101.6996	3.0706	Sri Petaling	473.25	74.42	134.63
7091560	502	12	23500	101.5691	2.9930	Putra Height	377.25	1.61	6.81
1708646	502	16	33261	101.5844	2.9574	Saujana Putra	96.07	5.22	22.03
1704996	502	16	33261	101.6234	3.0259	Taman WawasanPuchong	142.92	24.50	49.00
1714996	502	16	33261	101.6233	3.0259	Taman WawasanPuchong	116.61	18.25	41.11
1718646	502	16	33261	101.5854	2.9583	Saujana Putra	1848.30	126.61	247.81
8203301	502	19	1251	101.6233	3.0257	Bdr Puteri Puchong	5.23	0.81	2.92
8213301	502	19	1251	101.6244	3.0254	Bdr Puteri Puchong	114.74	20.85	52.15
8203263	502	19	1251	101.5974	3.0206	Kampung Bersatu, Puchong	215.54	53.89	86.22
8202283	502	19	1251	101.5822	3.0239	USJ 23	339.51	189.49	283.17
8202223	502	19	1251	101.5762	3.0284	USJ 23	4.29	4.29	4.29
7753810	502	19	1181	101.5399	3.0195	Kemuning Bayu	65.81	21.94	32.90
7745831	502	19	1181	101.5192	2.9944	Bukit Kemuning	2.50	1.87	2.14
7744081	502	19	1181	101.5144	2.9903	Bukit Kemuning	291.83	116.73	166.76
7745351	502	19	1181	101.4949	2.9857	Bandar Putera	15.94	5.31	7.97
15341	502	19	12501	101.4739	2.9864	Bandar Puteri Klang	1041.09	260.27	416.44
7780149	502	19	1181	101.4668	2.9860	Bandar Puteri Klang	266.61	114.62	159.83
7753899	502	19	1181	101.4495	2.9901	Ambang Botani	316.76	77.02	126.86
7744010	502	19	1181	101.4441	2.9971	Bukit Tinggi Klang	33.24	7.20	12.13
7754320	502	19	1181	101.4281	3.0276	Taman Palm Groove Klang	0.36	0.24	0.29
7744319	502	19	1181	101.4319	3.0307	Taman Palm Groove Klang	712.73	133.67	245.20
7744321	502	19	1181	101.4286	3.0311	Taman Palm Groove Klang	295.89	82.95	193.20
7097576	502	12	23500	101.5691	2.9930	Putra Height	12.60	1.28	3.29
6842862	502	12	25200	101.7052	3.1334	Jalan San Peng KL	246.46	37.84	99.23
7744031	502	19	1181	101.4523	2.9852	Bandar Botanic Klang	645.37	213.64	323.48
7754319	502	19	1181	101.4317	3.0301	Taman Palm Groove Klang	284.07	121.26	190.71
7107552	502	12	23500	101.6233	3.0257	Taman Wawasan, Puchong	7.85	0.73	2.26
641	502	12	2023	101.6076	3.0718	Sumway Pyramid	16.43	5.98	11.06
7107557	502	12	23500	101.6233	3.0257	Taman Wawasan, Puchong	293.20	0.99	11.75
8203301	502	19	1252	101.6233	3.0257	Taman Wawasan, Puchong	315.29	1.32	7.44
1724996	502	16	33261	101.6233	3.0257	Taman Wawasan, Puchong	6.62	1.14	3.66
6770462	502	12	10400	101.6568	3.1806	Segambut	65.41	23.51	36.40
7754321	502	19	1181	101.4292	3.0313	Teluk Gadong Klang	312.56	191.58	248.57

7033327	502	12	22500	101.7003	3.0701	Sri Petaling	595.96	52.19	108.37
7827843	502	12	10023	101.5932	3.0784	SS15	92.15	24.83	48.42
17821	502	12	2023	101.5938	3.0593	Subang Light Industrial Park	15.15	5.11	8.92
17827	502	12	2023	101.5929	3.0609	Subang Light Industrial Park	291.95	49.69	84.93
4271	502	12	2111	101.5691	2.9930	Putra Height	6.25	0.99	3.77
4262	502	12	2111	101.5691	2.9930	Putra Height	3.80	0.64	1.27
7107735	502	12	23500	101.6247	2.9813	Bukit Puchong	0.40	0.04	0.08
1562	502	12	2242	101.6247	2.9813	Bukit Puchong	2.03	0.13	0.24
8213301	502	19	1252	101.6244	3.0254	Taman Wawasan, Puchong	64.51	11.33	32.55
7284658	502	19	1111	101.6057	3.1210	Taman SEA, PJ	2.03	1.60	1.78
7033239	502	12	22500	101.6729	3.0727	OUG	17.00	1.19	6.22
6842531	502	12	25200	101.7113	3.1391	Pudu, KL	48.87	21.46	31.69
7232042	502	12	10024	101.6117	3.1581	Mutiara Damansara	36.27	23.56	28.29
7033234	502	12	22500	101.6730	3.0725	OUG	7.82	0.97	5.98
7821437	502	19	1192	101.7386	3.0722	Alam Damai	14.19	9.46	11.36
18387	502	16	22042	101.6632	2.9268	Cyberjaya	63.06	53.70	58.89
42426	502	12	2025	101.6633	2.9272	Cyberjaya	456.17	31.55	86.99
40173	502	12	2025	101.6628	2.9270	Cyberjaya	11.74	5.93	7.13
42428	502	12	2025	101.6573	2.9262	Cyberjaya	630.13	156.74	252.02
40171	502	12	2025	101.6629	2.9270	Cyberjaya	10.02	2.95	6.81
7744079	502	12	10026	101.4410	3.0252	Taman Selatan, Klang	7.72	0.23	3.15
7744082	502	12	10026	101.4411	3.0254	Taman Selatan, Klang	310.70	21.94	44.63
7744084	502	12	10026	101.4410	3.0235	Taman Chi Liung, Klang	1823.76	179.12	331.61

Note: MCC: Mobile Country Code; MNC: Mobile Network Code; LAC: Location Area Code; Area is by referencing to Google Maps. All distances are to Reference Point

Table 1 illustrates that the worst case of maximum distance between the reference GPS coordinate and the GPS coordinates which have been categorized by Bayesian and SOM into the same group, is 1848.3m. This translates to the fact that the biggest error rate of positioning by our proposed technique, in the absence of GPS signal or unavailability of WLAN network, is approximately 2km. The worst case of minimum distance and the average distance between the reference GPS coordinate and the GPS coordinates in the same group, have been recorded as 260.27m and 416.44m respectively. On average case, the maximum distance, the minimum distance and the average distance between the reference GPS coordinate and the GPS coordinates in the same group, have been recorded as 246.43m, 41.73m and 72.48m respectively. The results in Table 1 further shows that our proposed technique is able to provide positioning service with the prediction accuracy of within 300m, and in most cases, within the accuracy of 100m, under the condition where GPS signal cannot be established. The proposed intelligent agent has improved the prediction accuracy of conventional cellular-assisted GPS positioning techniques which have been reported to have the accuracy of approximately 200m to 1000m [1, 14], and even up to 30km [15].

Conclusions

Conventional cellular-assisted GPS positioning techniques have been reported with low accuracy. The proposed intelligent agent equipped with data mining approaches improves the location prediction of cellular positioning technique in the absence of GPS and WLAN. Based on our experimental results, sub-kilometer position prediction can often be achieved, in contrast to the conventional techniques which have been reported to have low accuracy with distance rate varies in kilometers. However, sub-kilometer accuracy is still insignificant in effective positioning. As for the future works, location history, human and asset mobility pattern, and population density of trackers shall be obtained and made as the training criteria to the software agent. By doing this, individual profile of each of the trackers can be built with improved accuracy when the prediction model is trained based on users' profile.

References

[1] Trevisani, E., & Vitaletti, A. (2004, December). Cell-ID location technique, limits and benefits: an experimental study. In Mobile computing systems and applications, 2004. WMCSA 2004. Sixth IEEE workshop on (pp. 51-60). IEEE.
[2] Fox, V., Hightower, J., Liao, L., Schulz, D., & Borriello, G. (2003). Bayesian filtering for location estimation. IEEE pervasive computing, 2(3), 24-33.
[3] McGuire, M., & Plataniotis, K. N. (2003). Dynamic model-based filtering for mobile terminal location estimation. IEEE Transactions on Vehicular Technology, 52(4), 1012-1031.
[4] Najar, M., & Vidal, J. (2003, September). Kalman tracking for mobile location in NLOS situations. In Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003. 14th IEEE Proceedings on (Vol. 3, pp. 2203-2207). IEEE.
[5] Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Transactions on signal processing, 50(2), 174-188.
[6] Ibrahim, M., & Youssef, M. (2010, December). CellSense: A probabilistic RSSI-based GSM positioning system. In Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE (pp. 1-5). IEEE.
[7] Jan, R. H., Chu, H. C., & Lee, Y. F. (2004). Improving the accuracy of cell-based positioning for wireless networks. Computer Networks, 46(6), 817-827.
[8] Isa, D., Kallimani, V. P., & Lee, L. H. (2009). Using the self-organizing map for clustering of text documents. Expert Systems with Applications, 36(5), 9584-9591.
[9] Isa, D., Lee, L. H., Kallimani, V. P., & Rajkumar, R. (2008). Text document preprocessing with the Bayes formula for classification using the support vector machine. IEEE Transactions on Knowledge and Data engineering, 20(9), 1264-1272.
[10] Deboeck, G., & Kohonen, T. (Eds.). (2013). Visual explorations in finance: with self-organizing maps. Springer Science & Business Media.
[11] Kriegel, H. P., Brecheisen, S., Kröger, P., Pfeifle, M., & Schubert, M. (2003, June). Using sets of feature vectors for similarity search on voxelized CAD objects. In Proceedings of the 2003 ACM SIGMOD international conference on Management of data (pp. 587-598). ACM.
[12] Wang, S. (2001). Cluster analysis using a validated self- organizing method: cases of problem identification. Intelligent Systems in Accounting, Finance and Management, 10(2), 127-138.
[13] Michalski, R. S., Bratko, I., & Bratko, A. (1998). Machine learning and data mining; methods and applications. John Wiley & Sons, Inc.
[14] Von Watzdorf, S., & Michahelles, F. (2010, November). Accuracy of positioning data on smartphones. In Proceedings of the 3rd International Workshop on Location and the Web (p. 2). ACM.
[15] Ashok, P., & Marie, K. B. (2013). A Survey of positioning algorithms on mobile devices in location based services. Int J Adv Res Comput Sci Softw Eng, 3(6), 1778-1784.