International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-4, November 2019

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

Yee-Wai Sim, Lam Hong Lee, Yen Pei Tay, Khang Wen Goh, Dino Isa

Abstract:This paper presents the use of intelligent agent technology, cellular-assisted Global Positioning System (GPS) and data mining for positioning purpose. Due to overlapping coverage areas of cell towers, conventional cell-based positioning techniques have been reported to be inaccurate. Current cell-assisted GPS positioning setup with high accuracy is costly as it requires huge investments on hardware deployments. A new solution of using intelligent agent technology was proposed by the authors for an economical and satisfactory cell-assisted GPS positioning system. Location information in the form of cell identity (ID) and GPS coordinates pairs can be acquired via devices such as smart phones and GPS trackers. The cell ID-GPS coordinates pairs are then grouped by each individual cell ID. An intelligent agent equipped with data mining capabilities is then deployed to computer the optimal GPS coordinates of the cell ID to provide more precise location information. The proposed solution was evaluated via a prototype system. The system was built to collect raw data of cell-ID and GPS coordinate pairs from trackers and mobile phone applications. Using the reference GPS coordinate that was calculated by taking the mean of longitude and mean of latitude for all the GPS coordinates clustered in the same group, the geographical distance between each GPS coordinate and the reference GPS coordinate in the same group was computed to evaluate the performance of the proposed solution.. Experimental results showed that the proposed solution based on intelligent agent equipped with data mining capability helped in improving the prediction of location with sub-kilometer accuracy, in contrast to the conventional cell-assisted GPS positioning system which have low accuracy with distance rate various in kilometers.

Keywords : Intelligent Agent, Cellular-Assisted GPS Positioning, Bayesian, Self-Organizing Map, Data Mining.

Manuscript published on November 30, 2019.

* Correspondence Author

Yee-Wai Sim*, School of Computing and Creative Media, University College of Technology Sarawak, Sibu, Malaysia. Email: yeewai.sim@ucts.edu.my

Lam Hong Lee, School of Computing, Quest International University Perak, Ipoh, Malaysia. Email: lamhong.lee@qiup.edu.my

Yen Pei Tay, School of Computing, Quest International University Perak, Ipoh, Malaysia. Email: yenpei.tay@qiup.edu.my

Khang Wen Goh, School of Computing, Quest International University Perak, Ipoh, Malaysia. Email: khangwen.goh@qiup.edu.my

Dino Isa, Faculty of Engineering, University of Nottingham, Selangor, Malaysia. Email: yenpei.tay@qiup.edu.my

I. INTRODUCTION

Mobile positioning has become a common technology since most modern mobile computing devices are equipped with both Global Positioning System (GPS) and cellular-based positioning features. Tracking of humans and assets has gained importance over the past decade to better monitor movements of targeted people and assets, in both known an unknown environment. Tracking devices that solely rely on GPS technology give out reliable location information under outdoor condition with minimum signal obstruction from buildings, trees, tunnels, etc. Nevertheless, GPS tracking devices will not be able to function when GPS signal is out of coverage, such as at indoor environment. In order to overcome this problem, most modern mobile computing devices, such as smart phones, are equipped with cellular based positioning feature, which enable tracking of the device to work with poor or no GPS signal. However, conventional cellular-based positioning technique has been reported to be less effective with highly degraded accuracy caused by factors such as cellular tower occupancy, cellular towers distributions and the design of cellular chipset in the device [1]. For instance, in real world applications based where cellular based positioning technique is deployed, the tracking devices are not always being connected to the nearest cellular towers, depending on the situations such as the load of the nearest cellular towers, the network switching mechanism implemented on the devices, and the signal strength of the cellular tower nearby. Such situations often lead to inaccurate positioning using conventional cellular-based technology. The problem is further complicated by the overlaps of cellular tower signals, leading to non-reliable location information. There are techniques that can produce high accuracy result, which involve intensive infrastructure and financial investment. However, these techniques are poor in economic efficiency; hence, they failed to gain popularity in the commercial deployment.

In this research work, an intelligent agent equipped with data mining capabilities was developed to improve the accuracy of cellular-based positioning technique. The improvement was made 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 cell towers. In the proposed solution, the raw data of cell-ID and GPS coordinate pairs will be collected from heterogeneous mobile devices, such as smart phones and GPS tracking devices.

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 others. In our study, the feasibility of using Bayesian probabilistic [8] and Self-Organizing Maps (SOM) [9] in processing the raw data was investigated.

Intelligent agent technology was employed in this research work to acquire and process the raw data so that more precise coverage prediction of the cell-IDs can be produced. By having an intelligent agent in carrying out the data mining work in conjunction with the tracking system, data sharing and data mining processes can be conducted through negotiations on a predefined set of criteria without affecting the normal operations which may lead to distortion of the original data in the mobility tracking system. Furthermore, intelligent agent technology was utilized with the main objective of automating the selection and configuration of criteria in processing and optimizing the data. The criteria include, but are not restricted to the following situations: uneven distribution of the cellular towers in due to geographical and obstruction factors; variation in load on the cellular towers at different times and days; and, the abnormal load on certain cellular towers during special occasions. By considering these criteria in the data mining process, the raw data collected in different periods should be handled accordingly. Such solution will produce a more accurate prediction and a more optimal coverage of the cellular based positioning.

The experimental results of this research indicated that the intelligent agent technology coupled with Bayesian probabilistic classification and SOM clustering can successfully improve the performance of the conventional cellular-assisted GPS positioning system. In addition, the proposed solution can produce a more accurate cellular-based positioning information without incurring extensive financial investment on existing hardware and infrastructure, resulting with high economic efficiency.

The next section reviews and presents the related techniques used in this research work. This is followed by a section to report a solution in using intelligent agent for cellular-assisted GPS systems. The results and discussion are presented in Section IV, while the concluding remark is made in Section V.

II. RELATED WORKS AND TECHNIQUES

A. Mobile Positioning Techniques

Various techniques have been introduced over the past decade to improve the accuracy and effectiveness of cellular-based positioning [2]-[5]. However, the reported techniques failed to address the fundamental issue of low accuracy of cellular-based positioning in the absence of GPS and Wireless Local Area Networks. Some research work as reported in [6]-[7], was conducted to overcome the problem by using a Simulated Annealing (SA) method. The method was employed to improve the accuracy of cellular-based positioning by identifying locations and transmission range of base stations in order to accomplish the best promising location accuracy. The work was carried out by using irregular structure of networks compared to the common configurations that focused on the mesh and hexagonal structures, which is challenged with economy efficiency issue.

B. Bayesian Probabilistic Technique

Bayesian approach is interpreted by applying Bayes' theorem as the basis of the theory. Bayes' theorem relates the conditional and marginal probability distributions of random variables. The formula that makes up the theorem is illustrated by the following equation in (1).

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)} \tag{1}$$

Bayes' formula shows that, by observing the value of B, the prior probability $\Pr(A)$ can be converted to a posterior probability $\Pr(A|B)$, where the probability of the state of nature being A given that feature value B has been measured. $\Pr(B|A)$ is called the likelihood of A with respect to B, where a term is chosen to indicate that other things being equal, the category A for which $\Pr(B|A)$ has the highest value is more likely to be the true category. The product of the likelihood $\Pr(B|A)$ and the prior probability $\Pr(A)$ is the most important in determining the posterior probability $\Pr(A|B)$. Meanwhile the evidence factor $\Pr(B)$, which is also known as the normalizing constant, can be viewed as merely a scale factor that guarantees the posterior probabilities are summed to one, as all good probabilities must [8]. Informally, Bayes' formula can be paraphrased as shown in (2).

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Normalizing Constant}} \tag{2}$$

Bayesian probability theory can be derived from the definition of conditional probability. The probability of event A given event B is illustrated as equation in (3).

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} \tag{3}$$

Likewise, the probability of event B given event A is as illustrated in (4).

$$\Pr(B \mid A) = \frac{\Pr(A \cap B)}{\Pr(A)} \tag{4}$$

Equation (5) is found after rearranging and combining the two equations in (3) and (4).

$$\Pr(A \mid B)\Pr(b) = \Pr(A \cap B) = \Pr(B \mid A)\Pr(A) \tag{5}$$

The combined equation in (5) is then being divided by $\Pr(B)$, consider that it is a non-zero value, Bayes' theorem is obtained, as in (6).

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)} \tag{6}$$

Bayes' theorem can hold in problems with more than two variables to handle classification tasks which the cases have multiple features. Equation (7) is an example of Bayes' theorem which holds in a two variable problem.

$$\Pr(A \mid B, C) = \frac{\Pr(A)\Pr(B \mid A)\Pr(C \mid A, B)}{\Pr(B)\Pr(C \mid B)} \qquad (7)$$

Bayesian probabilistic approach makes decision by applying Bayesian decision theory. Bayesian decision theory is fundamental statistical approach to the problem of pattern classification. This approach is based on computing the trade-offs between various classification decisions using probability and the cost that accompany such decision. Bayesian decision theory, which is also known as Bayes decision rule, is a result in probability theory, which relates the conditional and marginal probability distributions of random variables. In some interpretations of probability, Bayesian decision theory tells how to update or revise beliefs in light of new evidences. The decision of classification tasks which regards to Bayesian decision theory can be expressed by the decision rule that:

Decide ***Category1*** if **Pr(*Category1*|*feature*)** > **Pr(*Category2*|*feature*)**, and vice versa.

The rule above states that the decision of classification can naturally be inclined to decide the right category, if the posterior probability of a particular category is larger than the posterior probability of another category in a binary classification. In a multi-category classification, the right category can be justified by selecting category with the highest posterior probability among the others.

The fundamentals of Bayesian decision theory show how it can be viewed as being simply a formalization of common sense procedures. Classifications can be made although problems may arise when the probabilistic structure is not completely known. In the situation that a decision is forced to be made for the prediction of upcoming events without any information, an assumption is made that any mis-classification entails the same cost or consequence. Therefore, the only information which is allowed to be used is the prior probabilities. The logic is to use the decision rule:

Decide ***Category1*** if **Pr(*Category1*)** > **Pr(*Category2*)**, and vice versa.

Bayesian decision theory also expresses that the likelihood of *Category1* with respect to *feature*, a term which has been chosen to indicate that other things being equal, the decision rule below can be obtained:

Decide ***Category1*** if **Pr(*feature*|*Category1*)** > **Pr(*feature*|*Category2*)**, and vice versa.

The evidence, also known as normalizing constant, Pr(feature) is unimportant as far as making a decision is concerned since it basically just a scale factor that guarantees that the posterior probabilities are summed to one and states how frequently we will actually measure a pattern with the feature value. By eliminating this scale vector, the following completely equivalent decision rule can be obtained:

Decide ***Category1*** if **Pr(*feature*|*Category1*).Pr(*Category1*)** > **Pr(*feature*|*Category2*).Pr(*Category2*)**, and vice versa.

C. Self Organizing Map

Self Organizing Map (SOM) is used for organizing and grouping data points with similar properties. Every feature of the data point carries significant information for SOM to cluster them through similarity measurement of these new features. SOM starts its clustering task by deciding on which prominent features are to be used in order to effectively cluster the data into groups. Typically, an iterative process of statistical cluster analysis is implemented to improve feature extraction. Kohonen's principle of topographic map formation, states that the spatial location of an output neuron in the topographic map corresponds to a particular feature of the input pattern [10]. Figure 1 shows the SOM model which provides a map which places a fixed number of input patterns from an input layer into the so called Kohonen layer [11], [12]. The clustering model then learns through self-organization of random neurons whose changeable weights are attached to the layers of neurons [9]. The change of weights depends upon the similarity or neighborhood between the input pattern and the map pattern [13]. The topographic feature maps reduce the dimensions of data to two dimensions simplifying viewing and interpretation.

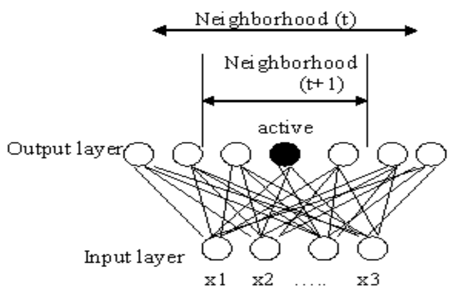

Fig. 1.The SOM model

SOM is constructed iteratively. In each construction step, a sample vector, x from the input data set is chosen randomly and the distance between x and all the weight vectors of the SOM, is calculated by using a Euclidean distance measure. The neuron with the weight vector which is closest to the input vector *x* is called the Best Matching Unit (BMU). The distance between x and weight vectors, is computed by using (8):

$$\|x - m_c\| = \min\{\|x_i - m_i\|\} \qquad (8)$$

where ||.|| is the distance measure, typically Euclidean distance. After the BMU is found, the weight vectors of the SOM are updated so that the BMU is moved closer to the input vector in the input space [10]. The topological neighbors of the BMU are treated similarly.

The update rule for the weight vector of *i* is as illustrated in (9).

$$x_i(t+1) = m_i(t) + \alpha(t) h_{ci}(t)[x(t) - m_i(t)] \quad (9)$$

where $x(t)$ is a vector which is randomly drawn from the input data set, and function $h_{ci}(t)$ is the learning rate and t denotes time [12]. The function $h_{ci}(t)$ is the neighborhood kernel around the winner unit c. The dataset of manufacturing details are fed into the input layer of SOM. Learning parameter is selected between 0.0- 0.9, and the SOM is trained. The training steps will be in the range of 100000 epochs in order to obtain a trained map. These training datasets are coded with reference to their prominent features [12].

SOM produces a map of neurons that represents the dataset. This map size can range from several neurons to a few hundred neurons depending on the characteristics of the dataset. The large number of neurons will potentially lead to a high number of classes when different neurons are fired for every data point. Hence, in order to prevent this, the neurons are first grouped in smaller clusters based on Euclidean distance. Each cluster will consist of several neurons and each cluster is given a number which represents the class. The threshold of the distance that controls the number of cluster is called the map threshold. All data points are then grouped around these clusters based on the Euclidean distance. The threshold value that controls this is called data threshold. Data points with the closest proximity to a particular cluster will be labeled with the number of that cluster.

III. PROPOSED TECHNIQUE

The intelligent agent solution proposed by the authors optimizes the coverage area of cellular towers by implementing Bayesian probabilistic classification and SOM clustering technique as described in the earlier section. Based on the GPS coordinates within the coverage of a particular cellular tower, which are acquired via GPS enabled devices, the intelligent agent optimize the set of GPS coordinates to be annotated to the cellular towers, generating "Virtual Cellular Tower GPS Coordinate" in providing a more precise cellular assisted GPS position information. Fig. 2 illustrates the block diagram of the proposed solution.

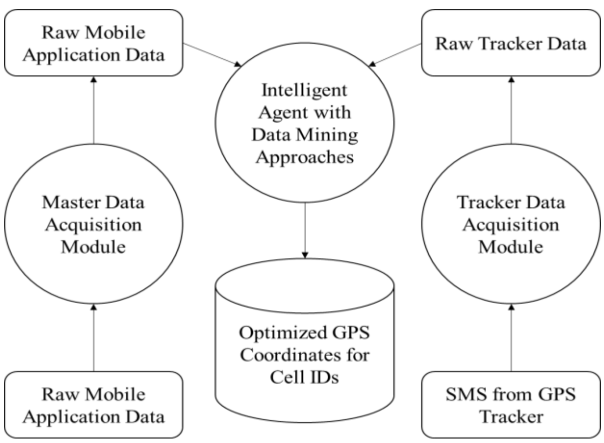

Fig. 2.Block diagram of Intelligent Agent for Enhanced Cellular-Assisted GPS Positioning

The acquired cell-ID and GPS coordinate pairs are grouped to build a training set for the Bayesian optimizer. Each GPS coordinate is then computed using Bayes theorem to get the probabilistic distribution of the GPS coordinate across the available cell-IDs. In order to optimize the coverage of a particular cellular tower, a predefined threshold is 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 is determined by using cross validation process. One of the factors which contributes to the inaccuracy of location returned by cellular-assisted GPS positioning devices is that the GPS enabled 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, we use SOM clustering technique to overcome the problem of having overlapping coverage from neighboring cellular towers. All the GPS coordinates acquired from the GPS enabled devices are fed into the SOM as vectors. The GPS coordinates of the cellular towers are treated as the input vectors of the mapped then the BMUs are computed to distinguish each of the clusters. The number of clusters produced by the SOM is set to be identical to the number of cellular towers captured by the GPS positioning devices. By using the SOM clustering technique, the GPS coordinates which are within close proximity will be grouped under the same cluster and the problem of overlapping coverage from neighboring cell-IDs can be eliminated.

Upon processed by Bayesian probabilistic and SOM, the data are now categorized into groups and labelled with the cell-ID which forms the majority of the GPS coordinates in it. The labels are useful in predicting the location of a tracker where the tracker returns only the cell-ID under the situation where GPS signal is not available.

After the raw data is processed by the intelligent agent, the optimized data are stored in a database for location prediction purpose based on the positioning requests made by the users. A better cellular-assisted GPS positioning accuracy can be obtained by minimizing the impact of 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. All in all, the optimized data produced by the intelligent agent contribute to better performance of cellular-

Table- I: 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	Mobile Country Code	Mobile Network Code	Location Area Code	Predicted Longitude	Predicted Latitude	Area	Max. Distance to Reference Point (meter)	Min. Distance to Reference Point (meter)	Avg. Distance to Reference Point (meter)
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

assisted GPS positioning for tracking systems, when the GPS enabled devices are out of other positioning network coverage (e.g. GPS and WLAN) as compared to the conventional techniques.

IV. RESULT AND DISCUSSION

In this research work, the proposed technique was evaluated by collecting raw data of cell-ID and GPS coordinate pairs from various trackers and mobile phone applications. A prototype system was developed for the data collection purpose from various locations within the area of Klang Valley in Malaysia.

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 obtained by calculating the mean of longitude and the mean of latitude for all the GPS coordinates clustered in the same group. Such solution was employed to mitigate the problem of skewed data due to uneven distribution of trackers. The mean calculation was utilized to cover the area of low tracker population density.

After the reference GPS coordinate of the groups/cell-ID has been identified, the geographical distance between each GPS coordinate and the reference GPS coordinate in the same group was then computed using Haversine distance function as shown in (10), assuming GPS coordinate for Point 1 to be (φ_1,λ_1) and GPS coordinate of Point 2 to be (φ_2,λ_2), and $\Delta\varphi=\varphi2 - \varphi1$, and $\Delta\lambda=\lambda2 - \lambda1$.

$$h = \sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\Delta\lambda}{2}\right) \qquad (10)$$
$$d = 2r\arcsin\left(\sqrt{h}\right)$$

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-I presents the reference GPS coordinate for each of the groups/cell-ID, and the maximum distance, minimum distance and average distance among all the GPS coordinates in the same group to the respective reference GPS coordinate. The table illustrates that the worst case of maximum distance between reference GPS coordinate and the GPS coordinates which have been categorized by Bayesian and SOM into the same group is 1843 meters. This translate to the fact that the biggest error rate of positioning by the authors' proposed technique is less than 2000 meters in the absence of GPS signal or unavailability of WLAN network. The worst case of minimum distance and the average distance between the reference GPS coordinate and the GPS coordinates in the same group were recorded as 260.27 meters and 416.44 meters respectively. In cases for the maximum distance, the minimum distance, and the average distance between the reference GPS coordinate and the GPS coordinates in the same group, the average values were recorded as 246.43 meters, 41.73 meters and 72.48 meters respectively.

The result shows that the proposed technique was able to provide positioning service with a prediction accuracy of within 300 meters, and in most cases, within the accuracy of 100 meters where GPS signal was absence. In other words, the proposed intelligent agent has improved the prediction accuracy of conventional cellular-assisted GPS positioning technique that was reported to have poor accuracy in the range of 200 meters to 1000 meters [1], [14], and even up to 30 kilometers in [15].

V. CONCLUSION

Conventional cellular-assisted GPS positioning techniques have been reported with low accuracy. The proposed intelligent agent equipped with data mining capability improves the location prediction of cellular positioning technique in the absence of GPS and WLAN. Based on the experimental result of this research work, 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.

As for the future works, the authors will consider to include more data such as location history, human and asset mobility pattern, and population density of trackers for use as training criteria to the intelligent agent. By using such an approach, individual profile of each trackers can be built with higher accuracy when the prediction model is trained based on the profiles.

REFERENCES

1. E. Trevisani and A. Vitaletti, "Cell-ID location technique, limits and benefits: An experiment study" in *Proceedings of the 6th IEEE Workshop on Mobile Computing Systems and Applications*, 2004, pp. 51-60.
2. S. Zhang et al, "Distributed Direct Localization Suitable for Dense Networks" *IEEE Transactions on Aerospace and Electronic Systems*, 2019, pp. 1-1.
3. J. Thakkar, S. Vyas, V. Panchal and R. Lokare, "Future Location Estimation of a Mobile User" in *Proceedings of the 2nd International Conference on Advances in Science & Technology*, 2019, available at http://dx.doi.org/10.2139/ssrn.3367689.
4. W. Sakpere, M. A. Oshin, W. Miltwa, "A State-of-the-Art Survey of Indoor Positioning and Navigation Systems and Technologies" *South African Computer Journal*, vol. 29(3), 2017, pp. 145-197.
5. M. D. Firoozjaei, J. Yu, H. Choi, H. Kim, "Privacy-preserving nearest neighbor queries using geographical features of cellular networks" *Computer Communications*, vol. 98(15), 2017, pp. 11-19.
6. M. Ibrahim and M. Youssef, "CallSense: A probabilistic RSSI-based GSM positioning system" in *Proceedings of the IEEE Global Telecommunication Conference*, 2010, pp. 1-5.
7. F. Ming, Y. Yang, Z. Zeng and B. Zhao. "Decomposition of geodetic time series: A combined simulated annealing algorithm and Kalman filter approach" *Advances in Space Research*, vol. 64(5), 2019, pp. 1130-1147.
8. D. Isa, V. P. Kallimani and L. H. Lee, "Using Self Organizing Map for clustering of text documents" *Expert Systems with Applications*, vol. 36(5), 2009, pp. 9584-9591.
9. D. Isa, L. H. Lee, V. P. Kallimani and R. Rajkumar, "Text document preprocessing with the Bayes formula for classification using the Support Vector Machine" *IEEE Transactions on Knowledge and Data Engineering*, vol. 20(9), 2003, pp. 1264-1272.
10. G. Deboeck and T. Kohonen, *Visual explorations in finance with self organizing maps*. New York: Springer-Verlag, 1998.
11. H. P. Kriegel, S. Brechesen, P. Kroger, M. Pfeifle and M. Schbert, "Using sets of features vectors for similarity search on voxelized CAD objects" in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 2003.
12. S. H. Wang, "Cluster analysis using a validated self organizing method: Case of problem identification" *International Journal of Intelligent Systems in Accounting, Finance and Management*, vol. 10(12), 2001, pp. 127-138.
13. R. S. Michalski, I. Bratko and M. Kubat, *Machine learning and data mining methods and applications*. Chichester: Wiley-Verlag, 1998.
14. S. Watzdorf and F. Michahelles, "Accuracy of positioning data on smartphones" in *Proceedings of the 3rd International Workshop on Location and the Web*, 2010, pp. 2.
15. B. Premkumar and A. M. Stanislas, "A survey of positioning algorithms on mobile devices in location based services" *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3(6), 2013, pp. 1779-1784.

AUTHORS PROFILE

Yee-Wai Sim is currently serving as an Associate Professor at School of Computing and Creative Media, University College of Technology Sarawak (UCTS), Malaysia. In addition, he is also the Director of Drone Research and Application Center at UCTS. He graduated with a BEng in Electromechanical Engineering and a PhD degree, both awarded by University of Southampton in 2000 and 2004. His research interests are in the areas of agent modelling and simulation, machine learning and knowledge management. Furthermore, he is expanding his research area into drone technology, studying the use of high altitude platform for broadcasting and communication purposes in rural areas.

Lam Hong Lee received a Bachelor of Computer Science from Universiti Putra Malaysia in 2004, and a Ph.D. degree in Computer Science from the University of Nottingham in 2009. He joined Universiti Tunku Abdul Rahman, Malaysia from 2009 to 2011 as an Assistant Professor. To date, he serves as an Associate Professor at Quest International University Perak, Malaysia since 2013. His current research interest lies in improving text categorization using AI techniques, specifically Support Vector Machines. Besides this, he is also investigating on the implementation of data mining, pattern recognition and machine learning techniques in various kinds of intelligent systems.

Yen Pei Tay is an academician working at Quest International University. He holds a Bachelor degree in Computer Science from Monash University and earned his MBA from University of Manchester. He currently directs the Center for Innovation & Technology at Quest International University Perak and co-founded a company named Simplify, an award-winning startup that is ranked as one of the World's 50 Most Innovative Companies by Fast Co. His main research focus is in the area of building a sustainable networked society.

Khang Wen Goh obtained his PhD from Universiti Teknikal Malaysia Melaka in 2019. He is currently working as Head of Actuarial Science Programme at Quest International University Perak, Malaysia with 10 years of teaching experience. He started his research in the field of numerical optimization toward global optimization since 2006 and obtained his MSc degree from Universiti Malaysia Terrenganu in 2009. He worked through several approaches from deterministic to metaheuristic in global optimization. His current research focus on the application of global optimization approaches in the finance industry.

Dino Isa is currently the Director for CONNECT Initiative. He obtained a PhD from the University of Nottingham UK in 1991. He joined The University of Nottingham Malaysia campus (UNMC) in Department of Electrical & Electronic Engineering in 2001 and was made Professor of Intelligent Systems in 2008. He has been principal investigator or collaborator on 18 grants in the past 18 years and was part of the team originating the BamPRO project which won a Smartfund grant from MESTECC (formerly known as MOSTI) valued at RM 2.33 million. Currently, Prof Isa has more than 100 papers in his field of research.

Retrieval Number: D6753118419 /2019©BEIESP
DOI:10.35940/ijrte.D6753.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication