A Weighted-Range Classification Model for Localizing Cell using Crowdsource Data

Aaron Franklin Soon, Siti Nurulain Mohd Rum, Hamidah Ibrahim, Rohaya Latip , Razali Yaakob , Lilly Suriani Affendey

Abstract: *The vast amount of mobile smartphone users provides an infinite source of data for crowdsourcing. Crowdsourcing provides an economical method of gathering data to cover a large geographical area compared to traditional methods. However, the inaccurate predictions for base station localization derived from mobile crowdsourcing impacts its effectiveness for use in radio planning. Therefore, the purpose of this study is to design a model that can yield a more accurate localization through the introduction of a rule-based weighted classification. The methodology deployed is a permutation series based on fingerprint of the cell site with weightage derived from rule-based classification. DeLaunay triangulation and Voronoi diagrams are used to determine the positions of the existing base stations and the prediction of new site location respectively. The expected results are better accuracy of the classification model in the localization prediction of the base station leading to a more accurate prediction of new site location.*

Index Terms: Crowdsourcing, Triangulation, Fingerprint, Weighted range

I. INTRODUCTION

To date, there are approximately five billion unique mobile subscribers on the planet and this number is expected to rise to approximately six billion by 2025. According to GSMA Mobile Economy 2018, 57% of these are smartphone users and the figures are poised to grow to 77% by 2025. These figures are further enhanced by the actual number of SIM (Subscriber Identity Module) connections, which stands at a staggering 7.8 billion as of 2017 and expected to reach 9 billion by 2025. The mobile industry contributes about USD3.6 trillion or 4.5% to the world GDP. These mobile subscribers and the services involved are supported by enablers which are the cellular base stations propagating its coverage using several technologies such as 2G, 3G, and 4G. There are thousands of these base stations deployed worldwide however these are well kept secret by the MNOs (Mobile Network Operator) as these are used for mobile radio

Aaron Franklin Soon, **Aaron Franklin Soon** is a Ph.D. Student, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia.

Siti Nurulain Mohd Rum, Senior Lecturer , Faculty of Computer Science and Information Technology, Universiti Putra Malaysia.

Hamidah Ibrahim, Professor, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia

Rohaya Latip, Associate Professor, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia.

Razali Yaakob, Associate Professor, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia.

Lilly Suriani Affendey, Associate Professor, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia.

planning to create competitive advantage [1, 2]. Official public databases containing positions of these cellular base stations are generally non-existent and their accuracy has been the subject of several studies [2-5]. Therefore, this study shall focus on improving the inaccuracy of geo-positioning localization of signal source derived from mobile crowdsourcing data. Many studies on localization have concentrated on outdoor and indoor localization of mobile devices [6-10]. Many of these studies involve crowdsourcing as it is a promising method to gather samples to create an outdoor radio map [11]. Crowdsourcing is a term coined to indicate work performed by a large group of people through voluntary basis compared to the traditional method of self-sufficient engagement of work [12]. The approach can generate a large amount of data sufficient to generate a sizeable radio map. However, effort is taken to manage its outliers due to radio propagation properties which are susceptible to terrain, Rayleigh fading, multipath fading, reflections and obstacles which may affect the samples [13]. Other issues such as accuracy of user equipment and errors generated by these volunteers are to be noted. Accordingly, studies done with crowdsourcing does show inaccuracies related to prediction of the intended position whether mobile device or cell site location for both indoor and outdoor environment [2-5, 8]. Nevertheless, crowdsourcing is a good alternative to traditional method such as wardriving and warwalking, which requires lots of time, cost and effort. In contrast, though traditional methods describe a systematic approach, however, it is a cross sectional method and does not offer the data coverage that crowdsourcing can generate.

Some of the available public databases used by researchers are OpenCellID, OpenBMap OpenMobileNetwork and OpenSignal [2].These large databases are loaded with datasets gathered by crowdsourcing approach. Studies by [2, 14] indicate that inaccuracies exist within the dataset when estimation algorithms are applied onto it. Some of the known approaches studied are centroid-based approach, weighted centroid based approach, minimum enclosing circle, grid based approach, coverage area estimation (model based), and highest RSS (Radio Signal Strength). A summary of seven collective approaches is shown in Table 1.

These approaches have their strengths and weaknesses. Findings from the studies indicate that there is no algorithm that can perform consistently due to variances in the dataset.

1351

However, studies do point to weighted centroid as being stable overall [14]. These variances affect the accuracy of the estimated base station localization. Therefore, the purpose of this study is to improve the accuracy of the current method of predicting localization of cell sites using crowdsource data. To achieve this, several issues need to be investigated:

a. Further exploration and analysis on quantified crowdsourced geolocation data.

b. Variances in the crowdsource data.

c. Algorithm used for accuracy.

d. The effectiveness of using crowdsource data for policymakers to make informed decisions.

e.

Table 1. Summary of seven collectively used localization approaches

#	Localization Approach	Estimated Basestation Position
1	Centroid [15]	Middle of geometric measurement
2	Minimum Enclosing Circle [16]	Middle of minimum enclosing circle of total samples
3	Weighted Centroid [17]	Middle of RSS-weighted geometric samples
4	Grid based [18]	Middle of grid within likelihood of strongest RSS
5	Ecolocation [19]	Position with the highest statistical value for RSS-distance relationship heatmap
6	Highest RSS [20]	Position of sample with the highest RSS value
7	Model based [21-23]	Position of the strongest predicted RSS using a model tuned propagation model

II. LITERATURE REVIEW

In this section, we explored relevant sources in relation to our research problem.

.Wardriving and warwalking

Wardriving is a process where a vehicle drives within a selected area while recording measurements from cell towers within its coverage proximity [24]. Warwalking is generally referred to the same task however; it is carried out on foot typically in malls or high rise buildings. Traditionally, this task is carried out either by the MNO, its vendor or both together. These tasks are routinely carried out by regulators worldwide to benchmark each service provider's service quality. In contrast to wardriving by crowdsourcing, the approach offers large scale measurements, however; the measurements are often filled with variances. Some of the common reasons are unstandardized equipment and methods of collection resulting in uneven datasets.

The task is carried out using one or several test phones, recording software equipped with a GPS (Global Positioning System) for positioning. These tests are pre-determined to measure certain parameters defined and agreed by both the vendor and MNO [25]. The task requires a considerable amount of resources such as vehicles, radio engineers, drivers, time, software and expensive test equipment [26, 27] . The cost of conducting these works can be astronomical when a large area is to be covered and has risks such as equipment failure and site outage leading to repetitive work to be done. Crowdsourcing was an approach used by researchers to gather RSS measurement in large areas with a much

reduced cost. Some of the available projects are seen from databases such as OpenCellID, OpenBMap, and OpenSignal. These databases offer a vast amount of dataset for research use as carried out by [2, 14] and many more. The studies done involved datasets from various countries and offer insight into how its radio mapping look like while contributing erroneous localization information for future researchers to work on.

Databases

There are several public databases used by many researchers. Among the common ones are OpenCellID, OpenBMap, and OpenSignal [2]. These databases contain datasets gathered from crowdsourcing for both mobile and WiFi. Works by [2, 14] utilised these databases. Neidhardt [2] compiled a list of available databases however, only the open cellular databases are used for the study as shown in Fig 1.

COMMERCIAL CELLULAR DATABASE						
Database	Operator	Since	Cells	Countries	Operators	Data Quality
MyLocation	Google	2007	many	Global	Global	High
NA	Apple	2010	NA	NA	NA	NA
NA	Microsoft & Nokia	2011	NA	NA	NA	NA
WPS	Skyhook Wireless	NA	NA	NA	NA	NA
location-api.com	Combain Mobile AB	NA	>32mil	>200	NA	96% success rate
NA	Navizon	NA	NA	NA	NA	NA
opensignal.com	OpenSignal	2010	830k	Global	825	NA
NA	Cellspotting	NA	NA	NA	NA	Symbian only
NA	Ericsson	NA	>18.3mil	Global	NA	offline
NA	cellid.telin.nl	NA	NA	NA	NA	offline

Fig 1. Overview of available databases [2]

Li [14] produced their findings using open cellular databases as well. A study conducted to compare the investments of MNO in the provisioning of mobile coverage in the city against its rural counterpart against income-level indicated that the latter are often excluded or received less attention. The result of the studies were made available through the availability of crowdsource data from OpenSignal [1]. Another option to gather large scale data is through the use of CDR (Call Detail Record) however; this method requires arrangement with the MNO. A study was done to identify and locate places where people frequent such as restaurants, malls and nightspots [28]. Other approaches include extracting data from base stations and internal databases in XML format. A study was done using XML data to visualize good cell and bad cell using Google map to relate mobile user behavior and cell performance [29]. However,

Retrieval Number: B10660882S819/2019©BEIESP
DOI:10.35940/ijrte.B1066.0882S819

1352

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

both XML and CDR methods require permission from the MNO and adherence to privacy laws. In summary, public databases are a good source of data as it contains a large set of data from crowdsourcing which offer researchers volume across different terrain, countries, and technology.

Crowdsourcing

Crowdsourcing is an outsourced task towards the crowd instead of regular workers [12]. Achtzehn [26], in their research work, used mobile crowdsourcing data to monitor the radio environment met with limitations in positioning accuracy. However, the research proved the worth of crowdsourcing in gathering data from across several operators instead of traditional methods where the focus is usually on a single operator. Li [14] utilized public crowdsourcing database to perform localization prediction also met with poor results and very high variances. However, it is noteworthy to say that the analysis was across several countries and could not have been accomplished without the help of crowdsourcing. This is similarly echoed by earlier research done by [2]. Among the other usage of crowdsourcing for the mobile sector is for benchmarking several mobile networks together. The study done to pursue this involved dataset with over two million instances. The study showed a tremendous potential of the importance of using this paradigm for the purpose of benchmarking [30]. Regardless of the known variances from crowdsourcing data, researches still embark on utilizing the model to study localization even for indoor environment. Many end up in concluding the necessity of outlier removal and validation of their proposed schemes [13]. As a conclusion, crowdsourcing data can be used for the purpose of the study however; emphasis should be placed on its outliers which causes high variances.

Outliers

Outlier detection is the process to detect deviation and to clean them. It is an essential stage in any data analysis methodology [31]. The process involves identifying data that deviate from the patterns shown by other data in the dataset. It can be measured by standard deviation, min, max and means values statistically [2-5, 8, 14]. These basic measurements are important for reliability and validity purposes. A study done to localize the positions of small cells using network measurement reports face erroneous location information due to outliers. The study suggested using RF (Radio Frequency) fingerprints gathered from these measurements to be used to eliminate UE (User Equipment) measurement reports with erroneous information on locations. These UE measurements with erroneous locations showed variances as large as 200km for rural areas and as small as 50m. The finding of the study proved that discarding these erroneous UE measurements increased the effectiveness of the small cell localization algorithm where large location errors (larger than 100 meters) can be eliminated compared to non RF fingerprint methods [32].The study claims to provide an accuracy of 20-50% improvement of the serving cell size where a typical cell radius of 250m yields location accuracy of approximately 50m to 125m. The result is consistent with other studies where outliers proved to be the concentration of other studies

regardless of the algorithm used [2, 10, 14, 18, 24, 32, 33]. Li [14] approach of handling outliers was to characterize the crowdsourced datasets into spatial, localization algorithm and combined RSS with spatial. These are illustrated in Fig 2.

Similar issues on outlier variances were faced by studies to predict outdoor base station positions [2, 4, 26] and WiFi positions [3, 13]. The common methods used to discard outliers are dependent on the algorithm used. Table 2 shows the drawbacks of several ways when managing outliers.

Dimension	Feature Class	Details
Spatial	measurement count	
	diameter	
	clustering: index of dispersion [18]	
	clustering: nearest-neighbor index	
	minimum enclosing circle radius	
	dispersion	max, min, median, mean, StdDev, coefficient of variance
	angular coverage	
	standard deviational ellipse [51]	StdDev(major), StdDev(minor), StdDev(major)/StdDev(minor)
RSS	RSS	max, min, median, mean, StdDev, coefficient of variance
	RSS (power level, dB)	
	% of RSS (power level)> γ	γ=-55, -60, -65, -70 and -80dB
	# of RSS (power level)> γ	
Algorithm	distance between algorithm M and N's location estimates	
RSS-Spatial	RSS-weighted dispersion	max, min, median, mean, StdDev, coefficient of variance
	RSS-weighted standard deviational ellipse	StdDev(major), StdDev(minor), StdDev(major)/StdDev(minor)
	correlation between measurement distance to center and RSS	
	spatial autocorrelation [20]	
	estimated path loss exponent	

Fig 2. Features handling outliers [14]

A study on radio coverage threshold observed that coverage can be categorised into several categories using live random drive test technique as suggested by [37]. The key performance indicator to measure the level of radio signal with 2G Global System for Mobile communication (GSM) is shown in Fig 3 as RxLev.

	Signal Strength			Signal Quality		
	2G	3G (FDD)	4G	2G	3G	4G
Coverage	RxLev (dBm)	RSCP (dBm)	RSRP (dBm)		Ec/No (dB)	RSRQ (dB)
Excellent	-48 ≤ RxLevel	-25 ≤ RSCP	-44 ≤ RSRP	NA	-6 ≤ Ec/No ≤-0	RSRQ >-3
Good	-70 ≤ RxLevel ≤ -48	-95 ≤ RSCP ≤-25	-110 ≤ RSRP ≤ -44	NA	-18 ≤ Ec/No ≤ -6	-12 ≤ RSRQ ≤ -3
Bad	-110 ≤ RxLevel ≤ -70	-115 ≤ RSCP ≤ -95	-140 ≤ RSRP ≤ -110	NA	-20 ≤ Ec/No ≤ -18	-19.5 ≤ RSRQ ≤ -12
Non-existent	RxLevel <-110	RSCP <-115	RSRP <-140	NA	Ec/No <-20	RSRQ <-19.5

Fig 3. Signal strength categories of several mobile generation technologies [34]

RxLev yields a negative value in dBm corresponding to the strength of the downlink radio signal. A mobile station (MS) must receive a signal greater than the parameter RXLEV_ACCESS_MIN [37, 38] to be able to access the mobile service. Third generation or 3G Universal Mobile Telecommunications System (UMTS) networks measures its radio signal using CPICH RSCP (Common Pilot Channel Received Signal Code Power) and Ec/No (energy per chip over noise ratio). Similarly the CPICH RSCP is a negative value measured in dBm [37, 38]. Fourth generation or 4G measures its radio signals using RSRP (Reference Signal Received Power). It is defined as the linear average over power contributions of the resource elements that carry cell-specific reference signals within the considered measurement frequency bandwidth [37, 38].

1353

A Weighted-Range Classification Model for Localizing Cell using Crowdsource Data

The study on autonomous coverage prediction comparing traditional drive test versus crowdsource indicated that better results can be achieved if a better prediction technique based on measurement distribution, environmental information and measurement density can be included in its algorithm [4, 39]. This is echoed by the study done to differentiate the impact of indoor and outdoor signal measurements from crowdsourced data. The attempt suggested that mixing indoor and outdoor measurements leads to unreliable results [40]. Other similar researches also indicated performance variances in prediction accuracy for localization due to high amount of variance in outliers in the crowdsourcing instances [2, 14]. Overall finding suggests that total removal of outliers may not be a solution however, a new way to utilize these data by classification could be an answer.
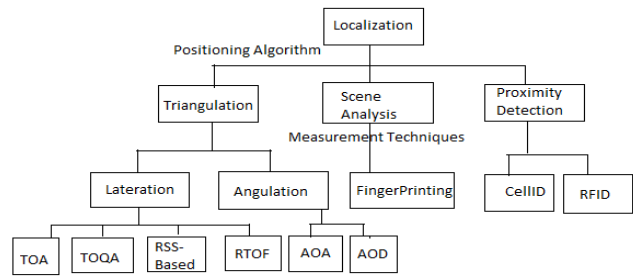

Fig 4: Classification of localization methods [10]

Triangulation, Lateration and Angulation

Lateration is an approach to determine the position of an object using distances between nodes while angulation uses the angles between the nodes. Lateration is divided into two algorithms which are trilateration and multilateration [41].

Table 2. Drawbacks of common approaches

Algorithm	Approach	Disadvantages
Centroid[2, 15]	All measurements inside polygon are equal weight.	Include outliers inside measurements. Prone to high inaccuracy due to method.
Weighted Centroid[2, 14, 17]	Classify each measurement to a weight of 0 and 1 using formula below: $w_k = \dfrac{rssi_k - rssi_{min}}{rssi_{max} - rssi_{min}}$ where: wk = weightage rssik = signal instance (k) rssimin = minimum signal in group rssimax = maximum signal in group	Dependant on size of sample. Each group may yield different minimum and maximum values. Weak RSS have high tendency of variances and error due to radio propagation properties [34].
Grid based[2, 35]	Modified fingerprinting method based on grid mapping mean and standard deviation of RSS inside the assigned grid. The probability of observing cell j with signal intensity of s in grid i is given as $p(i\|j, s) = \dfrac{1}{\sqrt{2\pi\sigma_{i,j}^2}} \exp\left(-\dfrac{(s - \mu_{i,j})^2}{2\pi\sigma_{i,j}^2}\right)$ where: μij = mean Ơij = standard deviation j = observed cell s = signal intensity i = grid	Middle of RSS-weighted geometric samples
Highest RSS[10, 20, 24, 36]	Strongest RSS estimates a cell's location as the location of the measurement with the strongest observed RSS from that cell. This approach works well when a cell is located close to the road where wardriving measurements were collected, but often fails otherwise.	Middle of grid within likelihood of strongest RSS

Radio Positioning and Localization Methods

Radio positioning is generally recognized as a method of determining the geographic position of a radio device using the properties of radio signals. Positioning systems determine coordinates of an object whereas localization systems place these objects on the map [6]. The methods used are classified into the categories as illustrated in Fig 4.

Multilateration (MLAT) is a surveillance method which measures the difference in distance to two stations at identified positions by broadcast signals at identified times. The technique yields an infinite number of locations that satisfies the measurement. In comparison, trilateration is the method to determine absolute or relative position of points by measuring distances through overlapping circles, spheres, and

triangles. Triangulation differentiates itself by not involving the measurement of angles. A known method is Delaunay triangulation and Voronoi diagram [42]. A study on WSN (Wireless Sensor Networks) used Delaunay triangulation and Voronoi diagram to locate and discover blind zone which represents a coverage hole within a group of clusters [43]. Similarly, other researchers used Delaunay triangulation and Voronoi diagram (see Fig 5) to generalize point features arranged in clusters such as thematic dot maps on cartographic maps [44]. Another study also indicated that Voronoi-based cell plans are better for location positioning however adding Bayesian method over it may not improve its accuracy for coverage prediction purposes [45].
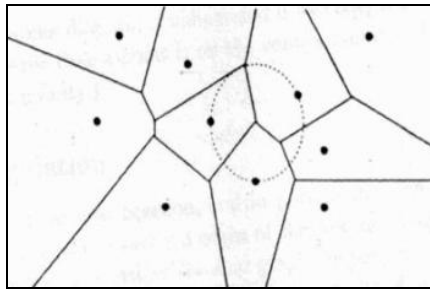

Fig 5: Voronoi diagram [23]

Angle measurements are positional solutions obtained by calculating geometric intersection points using circles derived from measurements such as time of arrival (ToA), time difference of arrival (TDoA) or even received signal strength (RSS). A study done using circle and DeLaunay triangle to locate coverage hole and boundary nodes for wireless sensor nodes (WSN) suggested that the algorithm can provide good performance [46]. Akbari [39] argued that the method of coverage estimation using geometry coordinates obtained from positioning techniques such as OTDOA (observed time difference of arrival) and A-GPS are prone to errors, therefore, the positions are tagged wrongly. Indoor environment, positioning techniques use WLAN (Wireless LAN) and RFID (Radio Frequency Identification). The study concluded that as long as these errors are not corrected the accuracy of prediction is affected hence requiring margins to be introduced. A study on location based services using TOA to calculate the location of a target mobile using nonlinear least squares indicated that prediction accuracy decreases when the number of base station is larger than five [3]. Similar studies using TDoA and AOA to locate a point source in a three dimensional space indicated that a hybrid method increases prediction accuracy. However, the study only uses two stations for prediction and requires known positions for both sources to improve accuracy [47].

Scene Analysis and Fingerprinting

Leca [3] studied characteristics of outdoor radio fingerprinting positioning using WiFi crowdsourcing data. The research found that fading influence caused by the environment results in high variances over 6dB for RSS values below -80dBm while better RSS values have standard deviation below 5dB. Similarly, studies by [32] to determine small cell site locations using geo-located UE (user equipment) measurement reports suggested that RF fingerprinting method improves accuracy compared to other techniques. However, the method only considered reliable UE geo-location measurements while all outliers are removed. Neidhart [2] used a grid based approach combined with fingerprinting to estimate a cell site position. According to [8] fingerprinting can be divided into four types i.e. visual, motion, signal and hybrid while localization objectives are categorized into three i.e. accuracy, energy efficiency, and latency. Each type and objectives are plagued with problems related to high variances and prediction accuracies in line with findings of other studies such as [2-4, 8, 14].

Proximity detection

A study done to characterize mobile phone localization using public dataset by OpenCellID resulted in high amount of errors in the form of erroneous cell IDs, antenna dragging, severe outliers, and unrealistic cell sizes [33]. Similarly, a study done to localize Wi-Fi positions in indoor environment met with challenges due to sparse coverage [48] . However, the study also found that using AP (Access Point) ranking and BSSID (Base Station ID) as fingerprint encourages good results. Li [14] shared similar findings where outliers were a problem even when a combination of techniques was used including machine learning. The observation showed that the combination technique has higher accuracy compared to the other methods however; outliers were a major challenge. The review indicates that a combination of methods works best to achieve better prediction leading to higher accuracy. The management of outliers play an important role in the goal to achieve better accuracy.

III. METHODS

The purpose of the study is to determine the best method to improve the prediction accuracy of base station location using crowdsource data. This study shall utilize data from selected public open source database. A control data setup shall be in place to act as a reference for signal strength versus distance from a known base station location. The predicted base station position shall be evaluated and validated with ground based data to test the accuracy of the prediction. The general overview of the experiment is illustrated in Fig6. The proposed tools required for input and data collection are:

a. Internet access to download public database data in text or CSV format.

b. Industry standard drive test kit to do the drive test.

c. Apache, MySQL webserver.

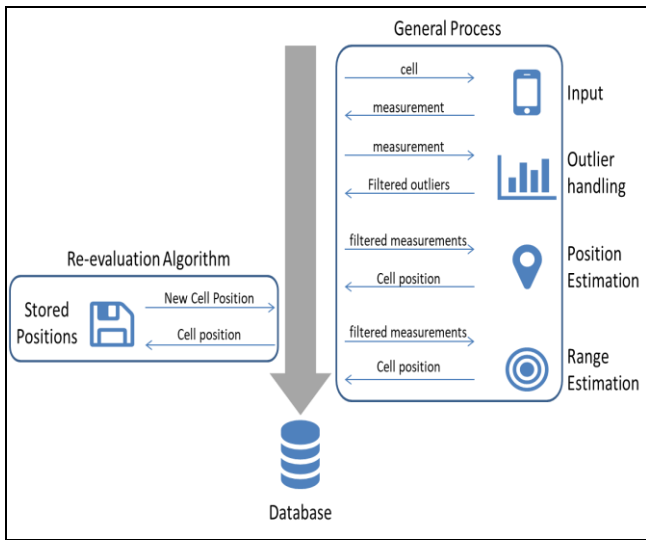d. Anaconda software with Python version 3 (i.e. Spyder)

Fig 6: Overview of total prediction process

Preparation of data involves the detection and handling of outliers. The proposed method is rule based weightage classification as shown in Table 3.

Table 3: Proposed Rule Based Weightage Classification Table

Technology	Signal Strength	Range	Weight	Classification
2G	RxLev	-46dBm - 70dBm	1	Excellent
		-71dBm - 85dBm	2	Good
		-85dBm - 110dBm	3	Average
		-111dBm - 120dBm	4	Bad
3G	RSSI	-46dBm - 70dBm	1	Excellent
		-71dBm - 85dBm	2	Good
		-85dBm - 110dBm	3	Average
		-111dBm - 120dBm	4	Bad
4G	RSRP	-46dBm - 70dBm	1	Excellent
		-71dBm - 85dBm	2	Good
		-85dBm - 110dBm	3	Average
		-111dBm - 120dBm	4	Bad

Using highest RSS technique, the signals are classified into categories of excellent, good, average, and bad. From literature review, it can be concluded that these ranges are valid and applicable to all technology [37]. Though there are improvements in receiver sensitivity for newer technology however, the relation of these signal strengths towards outliers is a known fact. Each measurement shall be marked with their weightage indicators together with its corresponding fingerprint of CellID. A CellID is a unique identifier of a base station and has 3 sectors as a standard. These are similar across all technologies. An example of the combination of these measurements and weight is shown in Table 4.

Table 4: Composition of Weight per instance

Signal Strength	CellID* *ID=Identification	SiteID*	CGI (Common Group Identifier)	Weight	Longitude	Latitude	Cluster
-55	12341	1234	Mnc+mcc+Cellid+lac	1	X1	Y1	1
-78	12342	1234	Mnc+mcc+Cellid+lac	2	X2	Y2	2
-90	12343	1234	Mnc+mcc+Cellid+lac	3	X3	Y3	3
-55	12341	1234	Mnc+mcc+Cellid+lac	1	X4	Y4	1
-80	12342	1234	Mnc+mcc+Cellid+lac	2	X5	Y5	2
-95	12343	1234	Mnc+mcc+Cellid+lac	3	X6	Y6	3
....	....	....	....	....	....	....	....

The final stage introduces clustering of the weighted data from Table 4. The clustered plot of these data is then passed through the triangulation stage where the center of the predicted position denotes the site location. Accuracy of the prediction is based on the permutation extracted from the weightage and clustering method. A higher weightage matrix for each cluster is expected to yield more accurate results.

IV. RESULTS AND DISCUSSION

In this research work, we have explored relevant research work done by other researchers to construct our proposed model. In summary, the overall processes in our proposed model consists of six steps which involves; (1). Extract known position, (2) Map with fingerprint, (3) Classify the data, (4) Rank the data, (5) Triangulate (6) Report and, update the database; lastly the process will be repeated from process (1) to (6). With the introduction of rule based classifiers, this research work is expected to yield better results as the instances are further narrowed down closer to the imagined base station position. As the classification yields a possible of 81 permutations per technology, it also permits a possible combination of technology dependent results to be compared with. Here, the likelihood of all technologies is located at the same site is relatively high therefore the absence of a good weightage set can be replaced or overcome by other permutations from another technology. It is also suggested for future work to emphasize on other outlier techniques with the combination of satellite imagery to predict the actual position of the base station.

REFERENCES

1. P. Koutroumpis and A. Leiponen, "Crowdsourcing mobile coverage," *Telecommunications Policy,* vol. 40, pp. 532-544, 2016.
2. E. Neidhardt, A. Uzun, U. Bareth, and A. Küpper, "Estimating locations and coverage areas of mobile network cells based on crowdsourced data," in *Wireless and Mobile Networking Conference (WMNC), 2013 6th Joint IFIP*, 2013, pp. 1-8.
3. C. L. Leca, P. Ciotirnae, C. I. Rincu, and I. Nicolaescu, "Characteristics of crowdsourcing for outdoor radio fingerprinting positioning," in *Electronics, Computers and Artificial Intelligence (ECAI), 2017 9th International Conference on*, 2017, pp. 1-4.
4. M. Molinari, M.-R. Fida, M. K. Marina, and A. Pescape, "Spatial interpolation based cellular coverage prediction with crowdsourced measurements," in *Proceedings of the 2015 ACM SIGCOMM Workshop on Crowdsourcing and Crowdsharing of Big (Internet) Data*, 2015, pp. 33-38.
5. R. Wang, C.-Y. Chow, Y. Lyu, V. C. Lee, S. Nutanong, Y. Li, *et al.*, "Exploring cell tower data dumps for supervised learning-based point-of-interest prediction (industrial paper)," *GeoInformatica,* vol. 20, pp. 327-349, 2016.
6. J. A. del Peral-Rosado, R. Raulefs, J. A. López-Salcedo, and G. Seco-Granados, "Survey of cellular mobile radio localization methods: from 1G to 5G," *IEEE Communications Surveys & Tutorials,* vol. 20, pp. 1124-1148, 2017.
7. S. M. Mohamed, H. S. Hamza, and I. A. Saroit, "Coverage in mobile wireless sensor networks (M-WSN): A survey," *Computer Communications,* vol. 110, pp. 133-150, 2017.
8. Q. D. Vo and P. De, "A Survey of Fingerprint-Based Outdoor Localization," *IEEE Communications Surveys & Tutorials,* vol. 18, pp. 491-506, 2016.
9. Y. Wang, S. Wu, Z. Chen, X. Gao, and G. Chen, "Coverage problem with uncertain properties in wireless sensor networks: A survey," *Computer Networks,* vol. 123, pp. 200-232, 2017.
10. A. Yassin, Y. Nasser, M. Awad, A. Al-Dubai, R. Liu, C. Yuen, *et al.*, "Recent advances in indoor localization: A survey on theoretical approaches and applications," *IEEE Communications Surveys & Tutorials,* vol. 19, pp. 1327-1346, 2016.
11. J. Howe, "The rise of crowdsourcing," *Wired magazine,* vol. 14, pp. 1-4, 2006.
12. J. Howe, "Crowdsourcing: A definition," 2006.
13. Y. Ye and B. Wang, "On Building an Indoor Radio Map from Crowdsourced Samples with Annotation Errors," in *Local Computer Networks (LCN), 2017 IEEE 42nd Conference on*, 2017, pp. 401-408.
14. Z. Li, A. Nika, X. Zhang, Y. Zhu, Y. Yao, B. Y. Zhao, *et al.*, "Identifying value in crowdsourced wireless signal measurements," in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 607-616.
15. Y.-C. Cheng, Y. Chawathe, A. LaMarca, and J. Krumm, "Accuracy characterization for metropolitan-scale Wi-Fi localization," in *Proceedings of the 3rd international conference on Mobile systems, applications, and services*, 2005, pp. 233-245.
16. N. Megiddo, "Linear-time algorithms for linear programming in R^3 and related problems," *SIAM journal on computing,* vol. 12, pp. 759-776, 1983.
17. M. Kim, J. J. Fielding, and D. Kotz, "Risks of using AP locations discovered through war driving," in *International Conference on Pervasive Computing*, 2006, pp. 67-82.
18. H. Nurminen, M. Dashti, and R. Piché, "A survey on wireless transmitter localization using signal strength measurements," *Wireless Communications and Mobile Computing,* vol. 2017, 2017.
19. K. Yedavalli, B. Krishnamachari, S. Ravula, and B. Srinivasan, "Ecolocation: a sequence based technique for RF localization in wireless sensor networks," in *Proceedings of the 4th international symposium on Information processing in sensor networks*, 2005, p. 38.
20. M. B. Jamˆaa, A. Koubˆaa, and Y. Kayani, "Easyloc: Rss-based localization made easy," *Procedia Computer Science,* vol. 10, pp. 1127-1133, 2012.
21. P. Bernardin, M. Yee, and T. Ellis, "Estimating the range to the cell edge from signal strength measurements," in *Vehicular Technology Conference, 1997, IEEE 47th*, 1997, pp. 266-270.
22. H. Lim, L.-C. Kung, J. C. Hou, and H. Luo, "Zero-configuration, robust indoor localization: Theory and experimentation," 2005.
23. N. Panigrahi, *Computing in geographic information systems*: CRC Press, 2014.
24. J. Yang, A. Varshavsky, H. Liu, Y. Chen, and M. Gruteser, "Accuracy characterization of cell tower localization," in *Proceedings of the 12th ACM international conference on Ubiquitous computing*, 2010, pp. 223-226.
25. A. Lutu, Y. R. Siwakoti, Ö. Alay, D. Baltrūnas, and A. Elmokashfi, "The good, the bad and the implications of profiling mobile broadband coverage," *Computer Networks,* vol. 107, pp. 76-93, 2016.
26. A. Achtzehn, J. Riihijärvi, I. A. Barriía Castillo, M. Petrova, and P. Mähönen, "Crowdrem: Harnessing the power of the mobile crowd for flexible wireless network monitoring," in *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*, 2015, pp. 63-68.
27. J. D. Mankowitz and A. J. Paverd, "Mobile device-based cellular network coverage analysis using crowd sourcing," in *EUROCON-International Conference on Computer as a Tool (EUROCON), 2011 IEEE*, 2011, pp. 1-6.
28. M. Mamei, M. Colonna, and M. Galassi, "Automatic identification of relevant places from cellular network data," *Pervasive and Mobile Computing,* vol. 31, pp. 147-158, 2016.
29. R. A. Talukder, K. Andersson, and M. Ahsan, "Visualization and cell data analysis tool based on XML log files," in *Local Computer Networks Conference Workshops (LCN Workshops), 2015 IEEE 40th*, 2015, pp. 878-883.
30. C. Midoglu and P. Svoboda, "Opportunities and challenges of using crowdsourced measurements for mobile network benchmarking a case study on RTR open data," in *SAI Computing Conference (SAI), 2016*, 2016, pp. 996-1005.
31. A. Fawzy, H. M. Mokhtar, and O. Hegazy, "Outliers detection and classification in wireless sensor networks," *Egyptian Informatics Journal,* vol. 14, pp. 157-164, 2013.
32. R. Joyce and L. Zhang, "Locating small cells using geo-located UE measurement reports & RF fingerprinting," in *Communications (ICC), 2015 IEEE International Conference on*, 2015, pp. 3275-3280.
33. M. Ulm, P. Widhalm, and N. Brändle, "Characterization of mobile phone localization errors with opencellid data," in *Advanced Logistics and Transport (ICALT), 2015 4th International Conference on*, 2015, pp. 100-104.
34. L. Ferreira, M. Kuipers, C. Rodrigues, and L. M. Correia, "Characterisation of signal penetration into buildings for GSM and UMTS," 2006.
35. P. Nurmi, S. Bhattacharya, and J. Kukkonen, "A grid-based algorithm for on-device GSM positioning," in *Proceedings of the 12th ACM international conference on Ubiquitous computing*, 2010, pp. 227-236.
36. L. Cano, A. Capone, G. Carello, M. Cesana, and M. Passacantando, "On Optimal Infrastructure Sharing Strategies in Mobile Radio Networks," *IEEE Transactions on Wireless Communications,* vol. 16, pp. 3003-3016, 2017.
37. A. D. Kora, B. A. E. Ongbwa, J.-P. Cances, and V. Meghdadi, "Accurate radio coverage assessment methods investigation for 3G/4G networks," *Computer Networks,* vol. 107, pp. 246-257, 2016.
38. M. S. Sharawi, "RF planning and optimization for LTE Networks," *Evolved Cellular Network Planning and Optimization for UMTS and LTE,* p. 399, 2010.
39. I. Akbari, O. Onireti, A. Imran, M. A. Imran, and R. Tafazolli, "Impact of inaccurate user and base station positioning on autonomous coverage estimation," 2015.
40. M. K. Marina, V. Radu, and K. Balampekos, "Impact of indoor-outdoor context on crowdsourcing based mobile coverage analysis," in *Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges*, 2015, pp. 45-50.
41. P. Gaal, S. Fischer, and S. W. Edge, "Global navigation satellite system," ed: Google Patents, 2012.
42. D. Satyanarayana and S. Rao, "Local Delaunay triangulation for mobile nodes," in *Emerging Trends in Engineering and Technology, 2008. ICETET'08. First International Conference on*, 2008, pp. 282-287.
43. W. Fang, X. Song, X. Wu, J. Sun, and M. Hu, "Novel efficient deployment schemes for sensor coverage in mobile wireless sensor networks," *Information Fusion,* vol. 41, pp. 25-36, 2018.
44. H. Yan and R. Weibel, "An algorithm for point cluster generalization based on the Voronoi diagram," *Computers & Geosciences,* vol. 34, pp. 939-954, 2008.
45. J. Rõõmusaare, "Passiivsete mobiilipositsiooni sündmuste tõenäosuslik asukoha hinnang," 2016.

46. W. Li and W. Zhang, "Coverage hole and boundary nodes detection in wireless sensor networks," *Journal of Network and Computer Applications,* vol. 48, pp. 35-43, 2015.
47. J. Yin, Q. Wan, S. Yang, and K. Ho, "A simple and accurate TDOA-AOA localization method using two stations," *IEEE Signal Processing Letters,* vol. 23, pp. 144-148, 2016.
48. W. Waqar, Y. Chen, and A. Vardy, "Smartphone positioning in sparse Wi-Fi environments," *Computer Communications,* vol. 73, pp. 108-117, 2016.

AUTHORS PROFILE

Aaron Franklin Soon is a Ph.D. candidate from the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia. He received his B.E. in Electronics/Computer in 1997 from Universiti Pertanian Malaysia and MBA from Coventry University, UK in 2018. He has over 20 years industry experience in Telecommunications covering Radio Planning, Optimization and Implementation working with local and overseas mobile operators and vendors. His research interest includes mobile customer complaints and loyalty, network optimization problems, approximation algorithm, data mining and analysis.

Siti Nurulain Mohd Rum is a senior lecturer from Faculty of Computer Science and Information Technology, Universiti Putra Malaysia. She received her Master Degree in 2012 and PhD in 2016 from University of Malaya. Her Bachelor Degree is from Universiti Teknologi Malaysia. Siti was involved in various researches, leading to publication of a number of academic papers in the areas of Social Network Analysis, Semantic Web and Educational Technology. Her research interest includes Database, Semantic Web, Big Data Analytics and Social Network Analysis.

Prof. Hamidah Ibrahim is currently a professor at the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia. She obtained her PhD in computer science from the University of Wales Cardiff, UK in 1998. Her current research interests include databases (distributed, parallel, mobile, bio-medical, XML) focusing on issues related to integrity constraints checking, cache strategies, integration, access control, transaction processing, and query processing and optimization; data management in grid and knowledge-based systems. (e-mail: hamidah.ibrahim@upm.edu.my).

Assoc. Prof. Dr. Rohaya Latip is an Associate Professor at Faculty of Computer Science and Information Technology, University Putra Malaysia. She holds a Ph. D in Distributed Database and Msc. in Distributed System from University Putra Malaysia. She graduated her Bachelor of Computer Science from University Technology Malaysia, Malaysia.
She is currently the Head of Department of Communication Technology and Network. She served as an Associate Professor at Najran University, Kingdom of Arab Saudi (2012-2013). She is the Head of HPC section in University Putra Malaysia (2011-2012) and consulted the Campus Grid project and also the Wireless for hostel in Campus UPM project. She is also a Co-researcher at Institute for Mathematic Research (INSPEM). She is the Editor in Chief of International Journal of New Computer Architectures and their Applications (IJNCAA), Editor in Chief International Journal of Digital Information and Wireless Communication (IJDIWC) and editorial board of **International Journal of Computer Networks and Communications Security (IJCNCS), editorial board of** International Journal of Digital Contents and Applications (IJDCA) **and editorial board for International Journal of Computer Networks and Applications (IJCNA).**

Her research interests include Big Data, Cloud and Grid Computing, Network management, and Distributed database. For her research work, she won two medals at The World Inventor Award Festival (WIAF) 2014 organized by Korea Invention News. She was awarded Gold medal at IMIT-SIC Innovation Expo in 2018, Riau Indonesia, I-RIA 2018 (Best of the best award), Malaysia Technology Expo (MTE2014) and Malaysian Innovation Expo (MiExpo2013). She also won Silver medal at National Design, Research and Innovation Expo (PRPI) 2010 and Bronze medal at National Design, Research and Innovation Expo (PRPI) 2007 and 2006 respectively. She has published more than 80 papers in international and national journals, proceedings and posters.

Assoc. Prof. Razali Yaakob received the Bachelor Degree in Computer Science in 1996 and Master in Computer Science from in 1999, from Universiti Putra Malaysia, and PhD from University of Nottingham, United Kingdom in 2008. Currently, he is a lecturer at the Faculty of Computer Science and IT, Universiti Putra Malaysia. His research areas include artificial neural network, pattern recognition, and evolutionary computation in game playing. He is a member of the Intelligent Computing Group at the faculty.

Lilly Suriani Affendey is an Associate Professor in the Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia. She received her Bachelor of Computer Science degree in 1991 from the Universiti Pertanian Malaysia and in 1994 received her MSc in Computing degree from the University of Bradford, UK. In 2007, she received her PhD degree from Universiti Putra Malaysia. Her current research interest is in Multimedia Databases, Video Content-based Retrieval, Data Science and Big Data Analytics.