



Asian Journal of Scientific Research

ISSN 1992-1454

science
alert
<http://www.scialert.net>

ANSI*net*
an open access publisher
<http://ansinet.com>

On Data Cleaning with Intelligent Agents to Improve the Accuracy of Wi-Fi Positioning System using GIS

T. Joshva Devadas, C. Seelammal and S. Sadasivam

Department of Information Technology, Sethu Institute of Technology, Pulloor, Kariapatti, Virudhunagar District, 626115, India

Corresponding Author: T. Joshva Devadas, Department of Information Technology, Sethu Institute of Technology, Pulloor, Kariapatti, Virudhunagar District, 626115, India

ABSTRACT

Wi-Fi positioning system uses GIS to achieve higher accuracy by means of comparing the error distance. The objective of this study was to minimize the distance error generated during the process of positioning. Wi-Fi positioning system needs to have a proper placement of those access points which are free from error distance. Data cleaning is introduced in this study to clean the errors that are generated from the experimental results. Agent is an Intelligent Processing system that is situated in the environment and performs autonomous action in order to meet the design objectives. Agents are introduced in this system uses its intelligence in the data cleaning task by using its dynamic and flexible characteristics. This study aimed at describing the functionalities of such a data cleaning procedure along with the agent and improves the performance of Wi-Fi positioning system.

Key words: Wireless positioning, Wi-Fi, GIS, constraint optimization, data cleaning, agents

INTRODUCTION

Wireless Fidelity (Wi-Fi) is to make a connection to internet without the network cable. Wi-Fi or 802.11 b is one of the 802.11 first families to establish wireless standard in the market. Recently a new standard of IEEE 802.11 g offers higher data speed (11 to 54 Mbps) while sharing common operating frequency of 2.4 GHz. Wi-Fi based positioning system has emerged an idea that can solve the positioning in indoor environments by taking the advantage of the rapid growth of wireless access points in urban areas (Parthornratt and Techakittiroj, 2006).

Wi-Fi Technology is evolving technically and practically in past few years leading WLAN to be a common sight at universities, airports, hotels, hospitals, coffee shops, offices and organizations. The localization technique is used for positioning with wireless access points is based on measuring the intensity of the received signal (Parthornratt and Techakittiroj, 2006). The service point or access point is often referred to as “wireless hotspot” or “hotspot” in short. Incentives for developing and standardizing WLAN are definitely mobility and flexibility. The accuracy depends on the number of positions that have been entered into the database. The possible signal fluctuations that may occur can increase errors and inaccuracies in the path of the user. To minimize fluctuations in the received signal, there are certain techniques that can be applied to filter the noise (Galhardas *et al.*, 2005).

Wi-Fi positioning is, however, rapidly gaining acceptance as a complement and supplement to Global Navigation Satellite System (GNSS) positioning for indoor environments. Wi-Fi hotspots are prevalent in the very areas where GNSS starts to struggle and many 'smart' devices are already equipped with Wi-Fi technology that can support positioning applications. Numerous service providers are offering the essential databases that enable the overall positioning capability (Hightower and Borriello, 2001).

Data preprocessing stage is an important part in data mining process (Han and Kamber, 2011). It handles various types of dirty data in large data set. Dirty data consists of noisy, incomplete, inconsistent and lost of data value. Dirty data leads to development of inaccurate knowledge model (Rahm and Do, 2000). Data cleaning is the first phase of data preprocessing and is used to clean data from noise (Galhardas *et al.*, 2005). The data cleaning process needs to analyze the possible occurrence of the noisy data. The noise occurred due to the fault generated during the transmission are resolved by developing a data cleaning system which can eliminate the errors by identifying and removing the outliers identified in the system. In this paper we consider only those error positions which are to be eliminated during the Wi-Fi positioning process (Parthornratt and Techakittiroj, 2006).

Agents, i.e., special types of software applications, have become increasingly popular in computing world in recent years and they perform some set of tasks on behalf of users with some degree of autonomy and react to the environment (Seydim, 1999). In order to work for somebody, an agent has to include some amount of intelligence. Agents have the ability to choose among various courses of action, plan, communicate, adapt to changes in the environment and learn from experience (Kim *et al.*, 2003). Among the advantages of using intelligent agents, one may mention higher work efficiency, meaning that the user saves time, as agents work autonomously and more effectively, as they can search and filter huge amount of information which would be impossible for humans. This opens new approaches for researchers in combining data mining with intelligent agents (Sardinha *et al.*, 2003). Also, the intelligent agent doesn't just react to the changes to its environment but itself takes the initiative under specific circumstance. This characteristics request the agent should have a well defined goal. Agents incorporated in the data cleaning task have a specific goal of removing the errors by identifying the outliers (Cossentino *et al.*, 2006).

Now days the WLAN is the emerging research area that varies from real-world application to pure theoretical aspect. Major issue in wireless security focuses on wireless positioning application. WEP and WPA are encryption algorithms that are to be realized by the real-world globally. Hence, the development for wireless security turns to be the positioning technique (Hightower and Borriello, 2001).

Many researchers use the positioning technique to improve the accuracy of their need with various platform and architecture. The objective of this paper is to improve the accuracy by the means of positioning the access points in two areas with the aid of GIS. The experiment considers the indoor environment with two access points.

There are two types of computation to perform this positioning process. One is location fingerprinting model and the other one is the propagation based computation model (Jan and Lee, 2003). Location fingerprinting requires site survey and to collect signal strength samples at every single equally divided area of that building to build up a signal strength database. Propagation based computation is used to get the strength of signal but it don't have the facility to store the database. These are the some of the computation needed to do the mapping process. The test site shown in Fig. 1 for the system is where we need to position the access points.

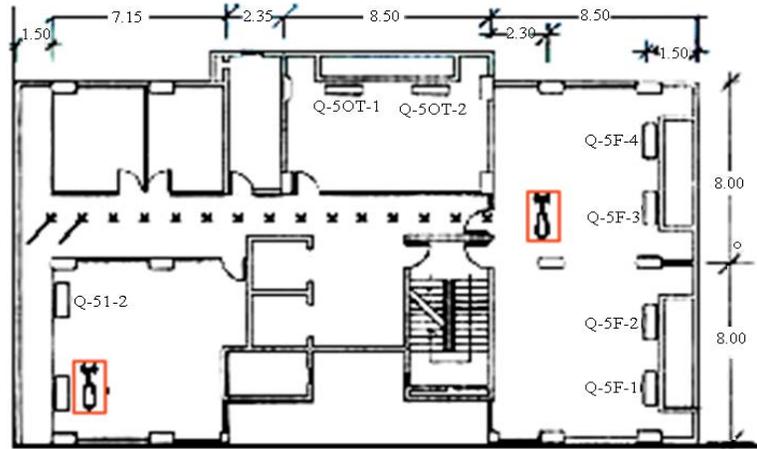


Fig. 1: Floor plan with markers and access points locations

In the test site we need to place the access points in two different areas and the corridor is selected for test process and get the signal from two access points namely NAL-Q51 and DSP-Q52. Triangulation is a basic geometric concept for finding crossing point on the basis of distance or angle (Hightower and Borriello, 2001). Therefore, it is radiating circularly in horizontal direction and so the access point is in the pattern of circle.

The existing system needs to be improved in order to decrease the error distance between the two access points and therefore it needs to perform the operation to obtain the calculated signal strength and error distance (Bahı and Padmanabhan, 2000). The next step in this process is to place the access point in an error free area.

The objective of this paper is to describe the functionalities, the importance of intelligent agents in removing the error distance in positioning using a data cleaning procedure and shows the improvement in Wi-Fi positioning system.

SIGNAL STRENGTH PREPROCESSING

Signal strength collection: The existing system makes use of the NetStumbler software to detect the signal strength. Signals are received from the two access points through Personal Digital Assistant (PDA) or laptop are detected using the NetStumbler software. NetStumbler imports the signal strengths (Fig. 2) and are mapped into excel sheet to calculate the signal average (Netstumbler.com). MATLAB is used to find the erroneous data and later the positioning is done by using the floor position and calculated positions. Predetermined location of the access points named DSP-Q52 and NAL-Q51 is included in Fig. 1 with markers along the corridor. Those fifteen markers are points of interest in our experiment to be used for learning and testing purposes. They are 1.25 meters apart from one another.

Multipath fading minimization: From all the four directions, the signal strength is collected to minimize the multipath fading. The NetStumbler software is used to detect the signal strength. Signal strength samples directly reflect the multipath effect in terms of direction sensitivity and angle of reception. In principle, we should achieve higher resolution of sampling and higher accuracy.

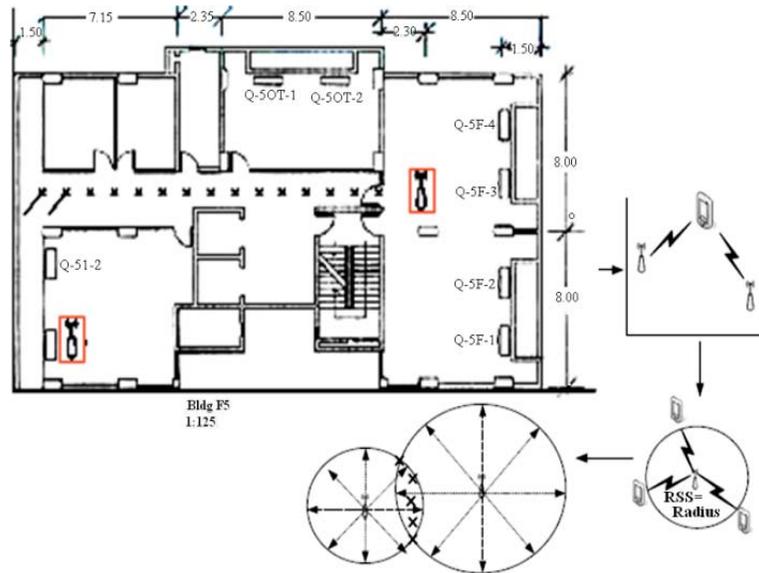


Fig. 3: Position determination overview

Free-space model: This model is used for the worst case consideration. This model can be implemented with GIS but it unveils a trend for further improvement. This model is not suitable to implement because this may disturb the other signals. Also, this model is not appropriate for the indoor environment to do the positioning system. Free-space model or Friis transmission model is described as:

$$P_R = P_T G_T G_R \left\{ \frac{\lambda}{4\pi d} \right\}^2$$

where, P_R and P_T are the receiving and transmitting power in unit of watts. G_T and G_R are the receiving and transmitting antenna gain. λ is a wavelength and d is a distance between transmitter and receiver.

Wall attenuation factor model: This model represents a common and reliable in building characteristics. Attenuation due to an intervening wall or partition is described as (Seidel and Rappaport, 1992):

$$P_R[\text{dBm}] = P_o[\text{dBm}] - 10 \log\{d/d_o\}^n - \text{WAF}$$

where, n is a path loss exponent and WAF is a wall attenuation factor.

Building environment may have the different electrical and physical characteristics which have obstruction that varies from the place to place. These values can be obtained by experiments. Our experiment considers this model because it is considered to be more suitable for indoor environment.

Learning process: Position determination is based on free-space model and it does not require learning process as universal relationship is implied every environment considered. Applying free space model yields higher error distance. WAF model is used to achieve more accurate result and

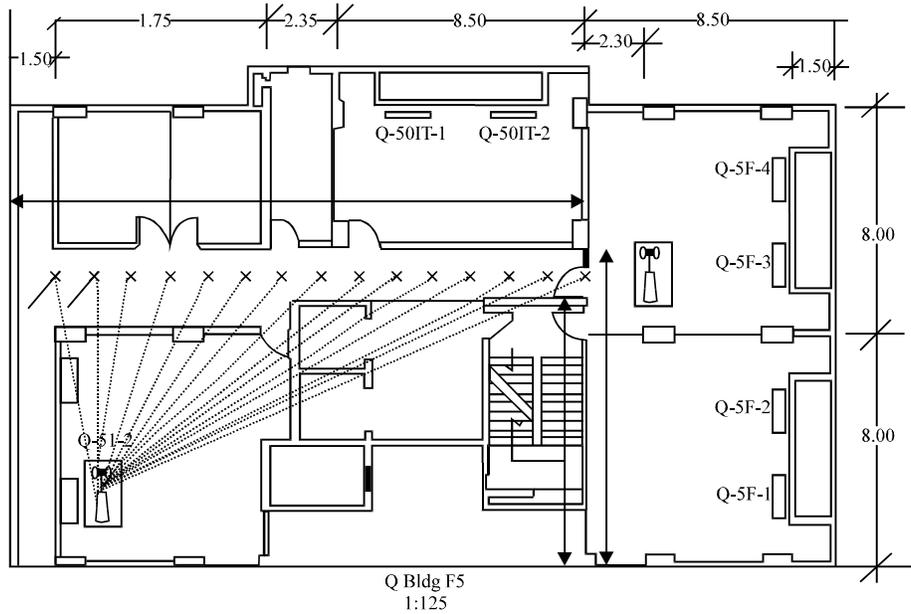


Fig. 4: Line of sight for NAL access point

to represent more realistic indoor environment. The values of WAF and n have been computed from experiment before positioning service. This is considered to be a learning process of environment by which clients are residing in. Once the measurement of signal strength at marking points is done, linear regression is applied to those data sets resulting in WAF and n parameters.

Line joining between the two access points to every single point of interest on the floor plan can at least differentiate amount of signal attenuation due to variety of obstructing objects as shown in Fig. 4. Longer the range from a transmitter, the lower signal strength from a receiver can perceive. Shorter range from transmitter can yields lower signal strength only in the case of presence of a large barrier between transmitter and receiver. Regional division is essential for the case of NAL-Q51 access point as received signal strength does not undergo common obstruction as the case of DSP-Q52 access point. Essentially, Geographical information is now used to separate region under consideration.

A line with rectangle marker in Fig. 5 and 6 represents experimental result, while triangle marker represents calculated result due to linear regression.

System simulation: Signal strength collected is translated into transmitter-receiver distance with the aid of two radio propagation models. Transmitter-receiver distance is the separation distance of access point from user and the distance are visualized as a radius of antenna radiation circle with center at that access point of signal reception. Compute the signal strength for both the access points NAL-Q51 and DSP-Q52 and translate it into radius of antenna radiation circle with the center at the access point itself as depicted in Fig. 3. The next step is to solve these equations and find the two cross points of the circles as illustrated in Fig. 7. Further, use of geographic information system in positioning eliminates the computation of cross points.

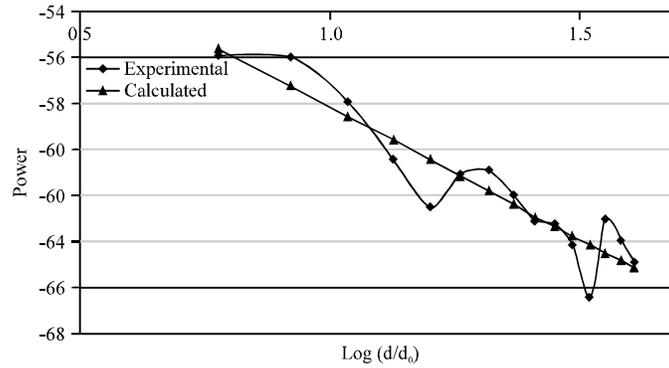


Fig. 5: Signal strength at DSP-Q52 access point

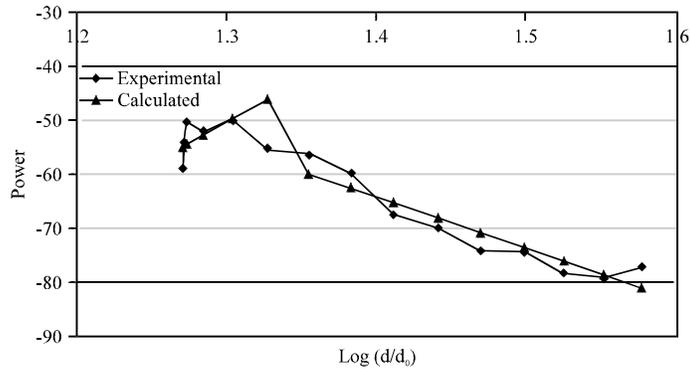


Fig. 6: Signal strength at NAL-Q51 access point

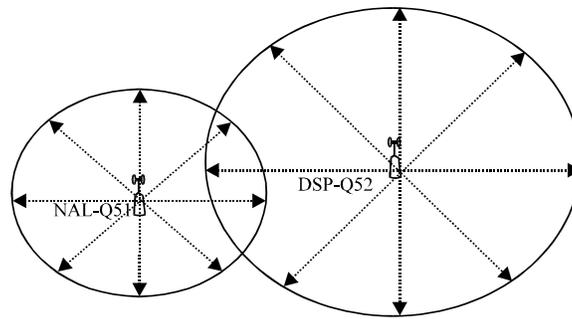


Fig. 7: Crossing of two radiation circles

In general, the mathematical representation of the circle is represented by:

$$(x-x_i)^2+(y-y_j)^2-r_k^2 = 0$$

where, x_i , y_j and r_k are the coordinates and radius of the respective access points NAL-Q51 and DSP-Q52. $|f(x,y)|$ represents the error matrix, where $f(x,y)$ forms a matrix using the mathematical equations. Hallway boundaries for both horizontal ($0 = x = 19.125$) and vertical ($9.25 = y = 11.125$) dimensions are used as constraint for minimization of the error distance. i.e., $\min |f(x,y)|$.

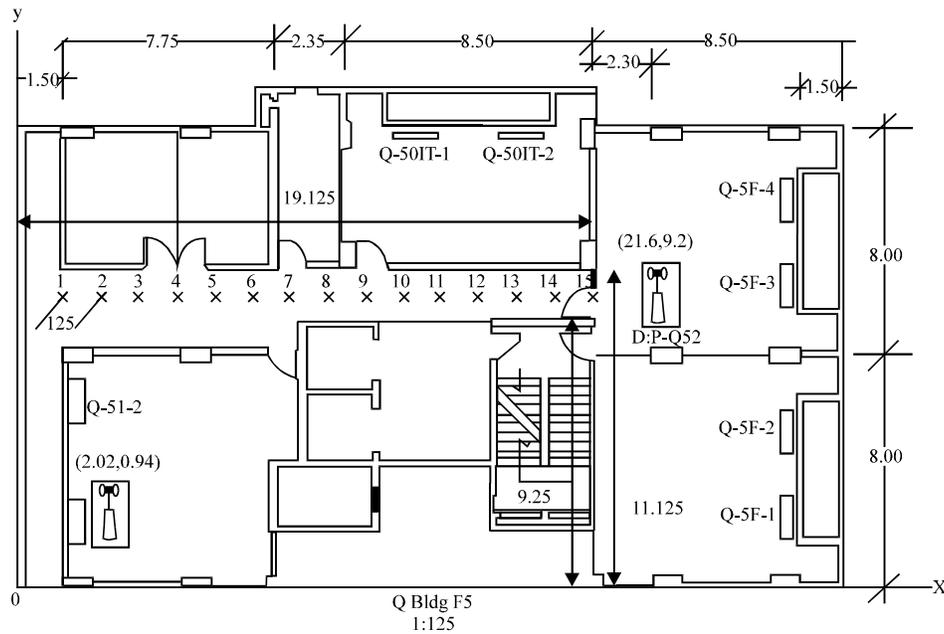


Fig. 8: Floor plan with corridor dimension

A floor plan in Fig. 8 displays our test site along with specific x and y dimensions of corridor section. Now the X and Y ranges from 0 and 19.125 and from 9.25 to 11.125 meters, respectively. Use of agent in positioning system improves the accuracy and reduces the working time of the process. Also, it scopes down and eliminates unlikely intersection area from positioning algorithm in the optimization phase.

DATA CLEANING PROCESS

Data cleaning is a process of identifying and removing the erroneous data. Since, the data is received with poor quality, it is necessary to clean the data before using it for processing. Poor quality of data is obtained due to some noise. The noisy data may have erroneous data, missing values, duplicate/redundant data and useless information. The error occur in the collected data are more common. One way of addressing this problem is to painstakingly check through the data with the help of data cleaning methods.

Data cleaning methods: There exists a multitude of different methods used within the data cleansing process. This section gives a short overview for the most popular of them. Parsing, data transformation, integrity constraint and duplicate elimination are the popular data cleaning methods (Winkler, 2003). Parsing in data cleansing is performed for the detection of syntax errors. A parser for a grammar G is a program that decides for a given string whether it is an element of the language defined by the grammar G . Data transformation intends to map the data from their given format into the format expected by the application. The transformations affect the schema of the tuples as well as the domains of their values. Integrity constraint enforcement ensures the satisfaction of integrity constraints after the transactions modifying a data collection by inserting, deleting or updating tuples have been performed. Duplicate detection method requires an algorithm to determine whether two or more tuples are duplicate representations of the same entity. For efficient duplicate detection every tuple has to be compared to every other tuple using this duplicate detection method.

A noise is nothing but a random error which can be removed either by binning method or clustering method or with the help of human inspection (Han and Kamber, 2011). We consider the clustering method is more suitable for cleaning the data of a Wi-Fi positioning system, because the formation of clusters clearly identifies the outliers. Identification of the outlier will eliminate irrelevant information by forming various clusters.

Need and strength of cleaning: In Wi-Fi positioning system access points PDA or laptop detect the signal strength and export into excel sheet is consider as primary data for data cleaning. At this stage the data collected may have erroneous information. Erroneous data is assumed to be the access points coordinates which causes intersection with the signal strength. Since, the data collected may have many such erroneous data items and the errors are removed by a data cleaning process or method.

Agents: Agents are intelligent processing elements which can work on behalf of others (Brenner and Zarnekow, 2002). An intelligent agent may be categorized as a human agent, or a hardware agent or a software agent. A software agent may in turn be categorized into information agent, cooperation agent and transaction agent. The agents are categorized into two major categories as internal and external properties. The external properties include all those characteristics which are associated with other agents for communication.

Work process of intelligent agent: The functionality of the agent can be described in Fig. 9 as a black box. Input is received through perceptions by the agent. After receiving input, agent use the information received for its intelligent processing and send the output as action performed over the given input (Weiss, 1999). Intelligent processing of the agent is done as a series of actions performed over the input. Interaction component found in this process is responsible of sending and receiving data from or to the external components. Information fusion is responsible for collecting the inputs from the interaction component. The information processing component is responsible of handling the input by carefully analyzing the data to perform action on it. The action component is capable of performing action on the data received and sends the output to the interaction component.

Need and the role of agents in data cleaning: Agents are introduced in the data cleaning process are capable of reacting with the environment with their intelligent behavior and is used to improve performance of Wi-Fi positioning system. The Intelligent behavior is achieved by systematically analyzing the data and store the useful information alone in the knowledge-base. Intelligent agents incorporated in the preprocessing phase of the data cleaning task uses its flexible and dynamic characteristics to clean the data. Agent uses its knowledge-base in the cleaning task

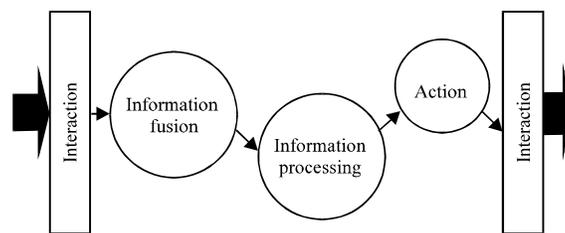


Fig. 9: Workflow of the agent

reacts with the system by verifying the existence of the required information. Use of agents in Wi-Fi positioning system will improve the performance by not repeating the same procedure for the access positions which has intersection in signal strength. Intersection of such signal strengths causes an error and the errors positions are stored in the knowledge-base of an agent. When the agents receive the data, it checks its knowledge-base to verify the existence of erroneous data. If an erroneous position was found then the agent filters this information otherwise it sends the details to perform action. Outliers are identified by using the domain of the erroneous data. Also, Agent based cleaning process distinguishes the access points which do not have the crossing point in their signal strength.

Functionality of intelligent agents in Wi-Fi positioning system: Agents collect data from excel data sheet through the interaction component and send it to the information fusion component which is responsible of receiving data. Data collected through excel data sheet may have some erroneous signal coordinates where, we cannot locate either the DSP or NAL. These signal coordinates collected through information fusion component are transferred to information processing component. Information processing component is responsible of computing the error distance between access points. If there is an error between the access points then those access points are transferred to the action component of the agent which in turn stores the details of the access point coordinates in its knowledge-base as action. The error free access point coordinates alone be transferred as output of the action component through the interaction component. At this stage the data processing handles only those access points which are error free causes improvement in average signal strength computation. Ultimately, the computation time required to accomplish the MATLAB simulation will also be greatly be reduced. Hence, the use of agents in the Wi-Fi positioning system improves the performance in one or more aspects.

Data cleaning agent procedure: PDA receives signals from the access points. The NetStumbler software detects the signal strength from the PDA and explores the signals into the excel data sheet. This data sheet is given in Fig. 10 as an input to the Data Preprocessing component. The preprocessing component is to clean the given data sheet with the help of intelligent agents. Agents are incorporated in the intelligence component to improve the performance of the system. Agents analyze the data access points carefully and compute the error distance to each of the access points. If agents encounters an erroneous data then agents learns to update its knowledge-base. Clusters are formed using the erroneous data that are stored in the knowledge-base. Before sending the details of the access points to the action component, agent stores all such erroneous coordinates of the access points in its knowledge-base.

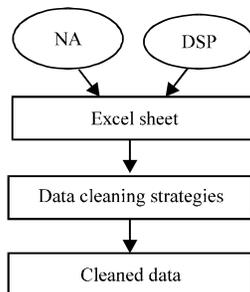


Fig. 10: Data cleaning process

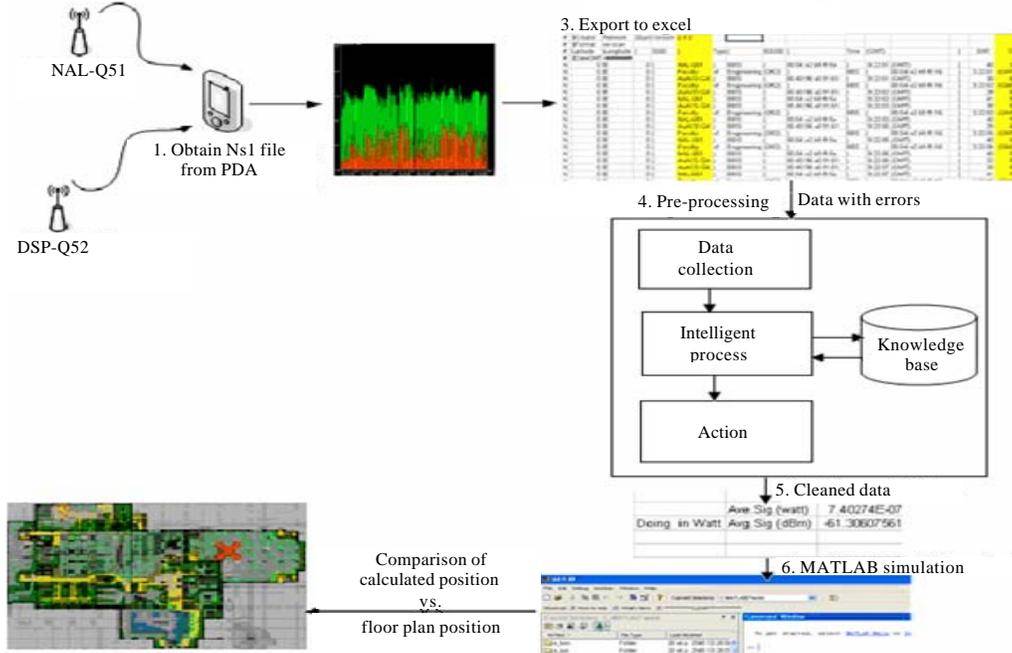


Fig. 11: Agents (intelligent process) in positioning process

Data cleaning is a treatment for decreasing possible negative influence. An outlier is an observation which deviates so much from other observations as to arouse suspicions that was generated by a different mechanism (Bay and Schwabacher, 2003). The goal of outlier detection is to uncover the “different mechanism”. If samples of the different mechanism exist in knowledge-base, then build a classifier to learn from samples. For fine classification, a function that designates the distribution of samples is necessary.

Agents are intelligent systems that verify the input data (access points) from its knowledge-base to identify the outlier. If the access point coordinates are error free then the agent uses the access point coordinates to position the laptop. Otherwise, the agent filters the details of erroneous access points by not sending them to the next layer as illustrated in Fig. 11. Use of agents in this process eliminates the computation time required to find the erroneous data using MATLAB. Agent based data cleaning eliminates the computation time associated with the detection of erroneous data which causes performance improvement in the system.

RESULTS AND PERFORMANCE EVALUATION

In the existing system, access points are positioned with the help of GIS and the errors positions are to be ignored are calculated with the help of MATLAB. All the operations taken individually and the signals are transformed into the excel sheet with the help of the software NetStumbler. Error distance is calculated for each of the floor positions using MATLAB and the error signal strengths are stored in the form of matrix. The next step is to position the access point according to the erroneous distance. This process to done manually and every signal strength are stored separately requires more time and makes the work process as a long one.

Table 1: Cross points and error distance for the case of with and without constraint optimization

Point	DSP signal	NAL signal	Cross 1		Cross 2		Map		Error distance	
			x	y	x	y	x	y	With constraint	No constraint
1	-64.855	-59.0746	2.523629	11.125000	9.666466	-5.80681	1.25	10.375	1.33321149	1.3332114
2	-63.945	-54.6688	5.617157	9.598141	10.73165	-2.52556	2.50	10.375	3.21250374	3.2125038
3	-63.003	-50.3564	8.229423	9.250000	12.30525	-0.41158	3.75	10.375	4.61853348	4.9426032
4	-66.441	-52.3056	0.00000	11.125000	7.905368	-7.61436	5.00	10.375	5.05593710	8.9515312
5	-64.121	-50.2152	5.085383	10.466670	10.98264	-3.51254	6.25	10.375	1.16821932	1.1682201
6	-63.245	-55.3689	7.770637	10.518470	12.89363	-1.62537	7.50	10.375	1.11038730	1.1103867
7	-63.065	-56.4428	8.253113	10.380390	13.13142	-1.18344	8.75	10.375	1.24444172	1.2444414
8	-61.950	-59.7820	10.68006	9.250000	14.01541	1.343695	10.00	10.375	1.20257494	2.8661568
9	-60.932	-67.5720	12.09993	9.250000	15.00625	2.360678	11.25	10.375	1.90083818	2.0942778
10	-61.124	-70.2304	12.53299	9.505269	15.49130	2.492719	12.50	10.375	0.37433131	0.3743317
11	-62.509	-74.3638	11.50825	11.125000	15.93632	0.628428	13.75	10.375	1.73075392	4.9330455
12	-60.459	-74.7163	13.84669	10.900560	17.40744	2.459965	15.00	10.375	1.00701298	1.5815186
13	-57.896	-78.4338	17.01084	10.399650	19.25673	5.075844	16.25	10.375	0.73428422	0.7342800
14	-55.969	-79.2599	15.85924	9.250000	17.62966	5.053288	17.50	10.375	1.21439900	2.6762931
15	-55.894	-77.5262	19.12500	11.125000	21.25162	6.083932	18.75	10.375	1.92088133	2.8488482
			Error distance							

			With constraint				No constraint			
Min			0.37433131				0.3743317			
Max			5.0559371				8.9515312			

Table 1 shows the significance of using constraints to minimize the error distance. Error distance is decreased from 8.95 to 5.05 m by the use of constraints and geographical information systems in improving indoor positioning system accuracy.

The main objective to introduce the agents in the positioning system improves the performance by cleaning the erroneous data and to store it in the knowledge-base for further processing. The accuracy and the time required to position the lap top is reduced. The signals are broadcasted from the access point and the signals make use of the NetStumbler to export the signal into excel sheet. Agent reads the excel sheet and store only the error details in its knowledge-base. Agent makes use of its knowledge-base to verify the error details and quit the work process for erroneous data. Reduction of processing time by the use of agents not only improves the performance but also not repeating the operations which are already done during the cleaning task will also be considered as one of the performance improvement factor. Thus agents present the data cleaning system enable the system with intelligent behavior and improves positioning strategies.

In the existing system the work process is done manually and the errors are detected individually until the positioning is done. Though the manual work progress successfully, the work process is repeated for all the access point signal strength. To overcome this repetition problem and not go beyond the cross point's intelligent agents are introduced in the cleaning system to store the error distances in its knowledge-base and uses it to clean the error signals.

For the given floor plan, the agent based system stores the error distance coordinates in its knowledge-base. Agent uses the knowledge-base information to avoid repeating the computation

process. Hence, time needed to compute the error distance for the subsequent process is eliminated by the use of agents in positioning system. Also, the accuracy of positioning is improved with the aid of geographical information systems. The combined feature of agents and geographical information system further improves the performance of positioning systems.

CONCLUSION

In this study, intelligent agents are introduced to clean the erroneous access point coordinates for the Wi-Fi positioning system that uses the knowledge of geographical information system. The performance variation of the Wi-Fi positioning system is described in terms of agent based and normal positioning. Agents are introduced between the excel data sheet and the MATLAB simulation data processing component. Future enhancement of this study may introduce multi-agents to perform the cleaning task with distinct functionality.

REFERENCES

- Bahi, P. and V.R.A.D.A.R. Padmanabhan, 2000. An in-building RF based user location and tracking system. Proc. IEEE INFOCOM, 2: 775-784.
- Bay, S.D. and M. Schwabacher, 2003. Mining distance-based outliers in near linear time with randomization and a simple pruning rule. Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 24-27, 2003, Washington, DC., USA., pp: 29-38.
- Brener, W. and R. Zarnekow, 2002. Software Agents Foundation and Application. Morgan Publishers, Germany.
- Cossentino, M., A. Chella and U.L. Faso, 2006. Designing agent based systems with UML. Proceedings of the 14th International Conference on Software Engineering, December 19-21, 2006, Cairo, Egypt.
- Galhardas, H., D. Florescu, D. Shasha, E. Simon and C.A. Saita, 2005. Declarative data cleaning: Language, model and algorithms. Proceedings of the 27th Conference on Very Large Data Bases, September 23-27, 2007, University of Vienna, Austria.
- Han, J. and M. Kamber, 2011. Data Mining: Concepts and Techniques. 3rd Edn., Morgan Kaufmann Publishers Inc., San Francisco, CA., Pages: 696.
- Hightower, J. and G. Borriello, 2001. Location systems for ubiquitous computing. IEEE Comput., 34: 57-66.
- Jan, R.H. and Y.R. Lee, 2003. An indoor geolocation system for wireless lans. Proceedings of the International Conference Parallel Process Workshop, October 6-9, 2003, Kaohsiung, Taiwan.
- Kim, D.S., C.S. Kim and K.W. Rim, 2003. Modeling and design of intelligent agent system. Int. J. Control Automation Syst., 1: 257-261.
- Parthornratt, T. and K. Techakittiroj, 2006. Improving accuracy of Wi-Fi positioning system by using Geographical Information System (GIS). AUJT, 10: 38-44.
- Rahm, E. and H.H. Do, 2000. Data cleaning: Problems and current approaches. IEEE Data Eng. Bull., 23: 3-13.
- Sardinha, J.A.R.P., P.C. Ribeiro, C.J.P. Lucena and R.L. Milidui, 2003. An object-oriented framework for building software agents. J. Object Technol., 2: 85-97.

- Seidel, S.Y. and T.S. Rappaport, 1992. A 914 MHz path loss prediction models for indoor wireless communications in multifloored buildings. IEEE. Trans. Antennas Propag., 40: 207-217.
- Seydim, A.Y., 1999. Intelligent agents: A data mining perspective. CiteSeer. IST, Literature Digital Library, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.23.7628>
- Weiss, G., 1999. Multiagent System a Modern Approach to Distributed Artificial Intelligence. The MIT Press, USA.
- Winkler, W.E., 2003. Data cleaning methods. <http://citeseerx.ist.psu.edu/viewdoc/summary?do=10.1.1.1.2066>