Indoor-Outdoor Detection Using a Smart Phone Sensor

Weiping Wang, Qiang Chang *, Qun Li, Zesen Shi and Wei Chen

College of Information Systems and Management, National University of Defense Technology,
Changsha 410073, China; wang.wp2010@gmail.com (W.W.); liqun@nudt.edu.cn (Q.L.);
nudtshizesen@126.com (Z.S.); weichen@nudt.edu.cn (W.C.)
* Correspondence: changqiang@nudt.edu.cn; Tel.: +86-187-1108-3508

Abstract: In the era of mobile internet, Location Based Services (LBS) have developed dramatically. Seamless Indoor and Outdoor Navigation and Localization (SNAL) has attracted a lot of attention. No single positioning technology was capable of meeting the various positioning requirements in different environments. Selecting different positioning techniques for different environments is an alternative method. Detecting the users' current environment is crucial for this technique. In this paper, we proposed to detect the indoor/outdoor environment automatically without high energy consumption. The basic idea was simple: we applied a machine learning algorithm to classify the neighboring Global System for Mobile (GSM) communication cellular base station's signal strength in different environments, and identified the users' current context by signal pattern recognition. We tested the algorithm in four different environments. The results showed that the proposed algorithm was capable of identifying open outdoors, semi-outdoors, light indoors and deep indoors environments with 100% accuracy using the signal strength of four nearby GSM stations. The required hardware and signal are widely available in our daily lives, implying its high compatibility and availability.

Keywords: seamless positioning; indoor/outdoor detection; machine learning; GSM
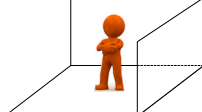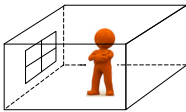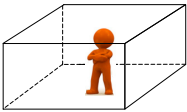
---

1. Introduction

In the era of mobile internet, we can connect to the internet any place and any time with portable devices. As a result, Location Based Services (LBS) have developed dramatically. The integration of traditional services and location information has brought about a great deal of innovative applications. All of these LBS applications have the same requirement: the user's current position.

The users may by in any place, such as in the open outdoors, crowded avenues, deep indoors and so on. The next generation of positioning systems must work well in a range of environments to meet the needs of a variety of LBS applications. Researchers have developed many positioning technologies for different environments. No single positioning technology is robust enough to perform well in all of these environments. However, there is an accurate enough positioning technique for any specific environment. For example, Global Navigation Satellite Systems (GNSS) performs well in open sky environments, while Shadow Matching [1] is enough for urban canyon environments, and Wi-Fi fingerprint positioning is suitable for indoor environments. This implies that we can integrate all these positioning techniques for Seamless indoor and outdoor Navigation and Localization (SNAL) applications [2]. However, the battery power is a limitation for any portable smart device. Turning on all of the required sensors is energy consuming. Letting the user manually choose different techniques according to different environments is not user friendly. Detecting the users' current environments

automatically and efficiently with low battery consumption is crucial for SNAL. In this paper, we focus on the automatic indoor/outdoor detection problem in SNAL.

There are a lot of previous works focused on indoor/outdoor detection to provide essential information for upper layer applications [3]. They mainly work in two types of environment: indoor and outdoor environments [4], but for SNAL, these two environments are not enough. In this paper, we focused on four types of environments: open outdoors, semi-outdoors, light indoors and deep indoors. These environments are shown in Table 1.

Table 1. Four indoor/outdoor environments.

Environment	Open Outdoors	Semi-Outdoors	Light Indoors	Deep Indoors
Definition	Outside a building	Near a building	In a room with windows	In a room without windows
Example				

In the open outdoors environment, at least four satellites are available for positioning because of the open sky condition. Semi-outdoors represents a GNSS-hostile outdoor environment such as an urban canyon or a wooded area, where there are not enough satellites for positioning. Light indoors is similar to a semi-outdoors environment. Deep indoors environment refers to a place with no satellite in view.

In each of these four environments, there was a well-developed positioning algorithm for localization. In the open outdoors environment, no matter whether on water or on a highway, we can localize ourselves using GNSS [5]. In the semi-outdoors environment, urban canyons and wooded areas are both capable of using GNSS shadow matching for positioning [1]. In the light indoors environment, cooperative positioning was suitable [6], while in deep indoors environments, a fingerprint positioning algorithm is the most commonly used technique [7,8].

The difference between open outdoors and semi-outdoors is the number of satellites in view. In open outdoor environments, at least four satellites are available, which means the user can be localized by GNSS, while in semi-outdoors environments, the number of visible satellites is not enough for GNSS localization. The difference between light indoors and deep indoors is the visibility of navigation satellites. In light indoors environments, users may receive navigation signals from several satellites. As a result, the users in light indoor environments can be localized using peer to peer cooperative positioning, but in deep indoor environments, no navigation satellites are available. In that case, cooperative positioning fails.

Many works address the issue of detecting indoor/outdoor environments. Walter et al. identified a set of novel environmental features that could be used for environment detection [9]. These features included gravity, ambient light, magnetic fields, scents, road signs, temperature, terrain height, road texture and so on. Some of the related works focused on environment sensing using mobile device's sensors based on these features. The sensors used included Global Positioning System (GPS), accelerometers, gyroscopes, barometers, geomagnetic sensors, Wi-Fi cards and so on. In Groves's work, environmental context detection using GNSS and Wi-Fi were examined [10]. The results showed that GNSS C/No measurement can be used to distinguish indoor from outdoor environments and to distinguish different types of outdoor environment, such as urban and open. Wi-Fi measurements have been shown to be unreliable for distinguishing betweenindoor and outdoor environments, but good for distinguishing different outdoor types, such as residential and business districts. Muralidharan proposed to use a barometer and GPS to identify different floors in indoor environments [11]. Ravindranath et al. showed that GPS lock status can be used to indirectly infer the ambient environment [12]. The GPS, however, consumed too much energy to be useful for many applications. A smart phone may run out of energy in about 6 h if the GPS is running continuously [13],

and furthermore, GPS was only available in open sky environments [14]. As a result, researchers proposed to use the low energy consumption sensors available in smart phones for detecting the environment, such as accelerometers, gyroscopes, barometers and magnetic sensors. Mostafa et al. made use of accelerometers, gyroscopes, barometers, and magnetic sensors to detect the height changing modes of the user in indoor environments [15–17]. However, accelerometers are orientation and position-dependent, a require a high sampling rate to achieve good accuracy. Researchers sought to use alternative sensors. For example, IODetector detected indoors, outdoors and semi-outdoors environments using cell signals, light and magnetic intensity [2]. Vanini proposed to use barometers to detect the change of floors [18]. TempIO classified the indoor/outdoor environment by comparing the temperature [19]. Wu et al. showed that barometers were capable of monitoring door events [20]. Barometer and temperature sensors are not widely available in current mobile phones. Radu et al. presented a general method employing semi-supervised machine learning and used light intensity, cellular signal strength, magnetic variance and sound intensity [3]. They provided a detection accuracy exceeding 90%, but their algorithm relied on several sensors, which is not energy efficient. Detecting indoor/outdoor environments using widely available sensors with low energy consumption remains a challenge for researchers.

Our research was motivated by these pioneering works, but we went further. We proposed to use the GSM signal strength to detect four types of indoor/outdoor environments. Firstly, GSM is available on all GSM-based smart phones. Secondly, it consumes minimal energy in addition to standard cell-phone operation [21]. The basic idea was simple: the propagation of radio signals is affected by the environment. Different environments result in different signal strength characteristics. By identifying the signal strength's characteristics, we can determine the user's environment. We have investigated a wide range of machine learning algorithms for classification, including Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), K Nearest Neighbor (KNN), Logistic Regression (LR), Naive Bayesian (NB) [22], and Neural Network (NN). The classifiers were applied for recognition.

Experimental results showed that the proposed algorithm was capable of detecting open outdoors, semi-outdoors, light indoors and deep indoors environments with 100% accuracy using four nearby GSM stations' signal strength. The required hardware and signals are widely available in our daily lives, implying its high compatibility and availability.

2. Methodology

The proposed algorithm contains three processes: Data Input, Training, and Testing. An overview of the indoor/outdoor detection process is illustrated in Figure 1.
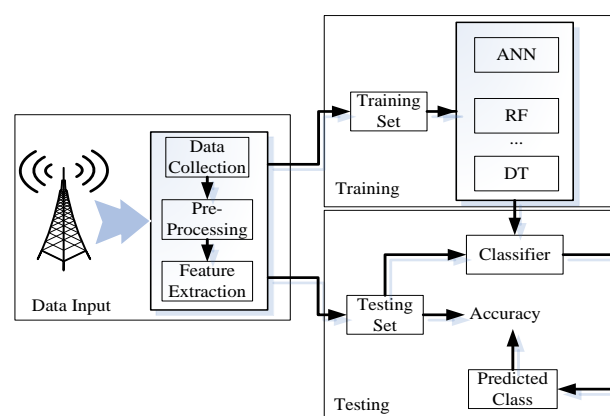

Figure 1. Framework of indoor/outdoor detection. This framework contains three processes: Data Input, Training, and Testing. In the Data Input process, neighboring cellular base stations' signal strength is recorded, and features are extracted. The data are classified in the Training phase. The classifier is applied to detect the user's current environment in the Testing phase.

2.1. Data Input

In the Data Input process, three sub-processes are included: Data Collection, Pre-Processing, and Feature Extraction.

2.1.1. Data Collection

In this process, GSM signal strength is recorded in different environments. According to the GSM standard, each smart phone records a vector of the detected power levels of the pilot signal from at most seven cellular base stations, one of which the smart phone is associated with [21]. For the Android Operating System, the received signal strength index (RSSI) for GSM is in asu, where 0 asu means -113 dBm and 31 asu means -51 dBm. So we have:

$$rss = -113 + 2 * rssi \tag{1}$$

The recorded signal strength is arranged from the strongest to the weakest. Assuming that the sampling interval is τ (s), the sampling frequency is $1/\tau$ Hz. The data collected during T(s) is denoted as S, where:

$$S = \{s_i | i = 0, 1, \cdots, \lfloor T/\tau \rfloor\}, \; s_i = \{rss_{i,1}, rss_{i,2}, \cdots, rss_{i,n_i}\}, rss_{i,m} \geq rss_{i,n} \forall \text{m} > \text{n} \tag{2}$$

where n_i is the number of visible stations at time i.

2.1.2. Data Pre-Processing

The environment is changing continuously, therefore, we need to look at how a signal changes over a specific period of time to identify the environment. A window is the most basic step and is used by almost all researchers. At each time epoch, the signal involving the current measurement and the previous $N - 1$ epochs, where N is a chosen integer that represents the length of the window. Assuming the window length is ΔT(s), so we have $N = \lfloor \Delta T/\tau \rfloor$, the measurement is arranged as:

$$W = \{w_i | i = 0, 1, \cdots, \lfloor T/ \triangle T \rfloor\}, \text{where } w_t = \{s_{t-i} | i = 0, 1, \cdots, \lfloor \Delta T/\tau \rfloor - 1\} \tag{3}$$

Different window lengths contain different information used for detection. The wider the window is, the more descriptive the features extracted from it can be. While the shorter the window, the less information it contains for detection.

2.1.3. Feature Extraction

In each window, we can extract several features to describe the character of the environment. These features include Mean, Standard Deviation, Maximum, Minimum and Range.

(1) Mean

Mean is the most basic character of a signal. It is calculated by summing the values and dividing the numbers:

$$mean(w_t) = \sum w_t / |w_t| \tag{4}$$

where $|*|$ is the number getting operation. Mean is a measure of the middle value of a signal.

(2) Standard Deviation

Standard Deviation is an indicator of how much a signal is dispersed around its mean. It is calculated as follows:

$$Std(w_t) = \sqrt{\sum (s_i - \text{mean}(w_t))^2 / |w_t|} \tag{5}$$

(3) Maximum and Minimum

Maximum and Minimum values are the extreme values in the window:

$$Max(w_t) = s_i \in w_t, \forall s_j \in w_t, s_j < s_i, Min(w_t) = s_i \in w_t, \forall s_j \in w_t, s_j > s_i \tag{6}$$

(4) Range

Range is the difference between the Maximum and Minimum values:

$$Range(w_t) = Max(w_t) - Min(w_t) \tag{7}$$

2.2. Training

In the Training Phase, we investigate a wide range of machine learning algorithms to classify the training data, including Decision Tree, Random Forest, Support Vector Machine, K Nearest Neighbor, Logistic Regression, Naive Bayesian, and Neural Network. The classifiers are applied for recognition. The training data can be raw data from the sensors, or the features in different window lengths. The best classifier will be selected for indoors/outdoors detection.

2.3. Testing

In the testing phase, new samples are categorized using the classifiers created by the machine learning algorithms. Different algorithms perform differently. Researchers have proposed a lot of performance measures for multi-class classification.

Confusion matrix was proposed to evaluate the performance of a classification system for both 2-class and multi-class samples. Table 2 is a template of confusion matrix for a 3-class classifier.

Table 2. Template of a confusion matrix for a 3-class classifier.

Class	Class 1	Class 2	Class 3
Class 1	n_{11}	n_{12}	n_{13}
Class 2	n_{21}	n_{22}	n_{23}
Class 3	n_{31}	n_{32}	n_{33}

In Table 2, the rows represent the true classes of the tested samples, and the columns represent the predicted classes. n_{ij} is the number of test samples of a class i recognized as class j. Sometimes n_{ij} would also be the percentage of the classification result. The nearer the confusion matrix is to a diagonal matrix, the better the classification algorithm is. Based on the confusion matrix, there are several widely used performance measures. Here we just introduce in Table 3 the ones we will use in the experiments.

Table 3. Template of a confusion matrix for a 3-class classifier.

	Definition	Formula
True Positive (TP)	The number of samples of a class which have been correctly classified	$TP_i = n_{ii}$
True Negative (TN)	The number of samples of other classes which has been correctly classified	$TN_i = \sum_{j \neq i} \sum_{k \neq i} n_{jk}$
False Positive (FP)	The number of samples not belongs to a class which has been incorrectly classified as belonging to it	$FP_i = \sum_{k \neq i} n_{ki}$
False Negative (FN)	The number of samples belonging to a class which have been incorrectly classified as belong to other class	$FN_i = \sum_{k \neq i} n_{ik}$
Accuracy	The proportion of all samples which have been correctly classified	$Acc = \frac{\sum_i n_{ii}}{TP_i + TN_i + FP_i + FN_i}$

Table 3. *Cont.*

	Definition	Formula
Sensitivity	The proportion of samples which have been correctly classified	$Sens_i = \frac{TP_i}{TP_i + FP_i}$
Precision	The proportion of sample predicted to belong to a class which is correct	$Prec_i = \frac{TP_i}{TP_i + FP_i}$
Specificity	The proportion of negative samples which have been correctly classified to be negative	$Spec_i = \frac{TN_i}{FP_i + TN_i}$
F-Measure	the weighted average of the precision and sensitivity	$F1 = \frac{2}{1/Sens_i + 1/Prec_i}$

3. Experiments

To test the context sensing algorithm, we implemented an Android smart phone application capable of capturing nearby cellular base stations' signal strength. The application is called DrawRSS. This application was tested on a Meizu MX3 smart phone (MeiZu, ZhuHai, China), which is running Android OS Version 4.4 (KitKat) (Google, Mountain View, CA, USA). A screen-shot of this application is shown in Figure 2.
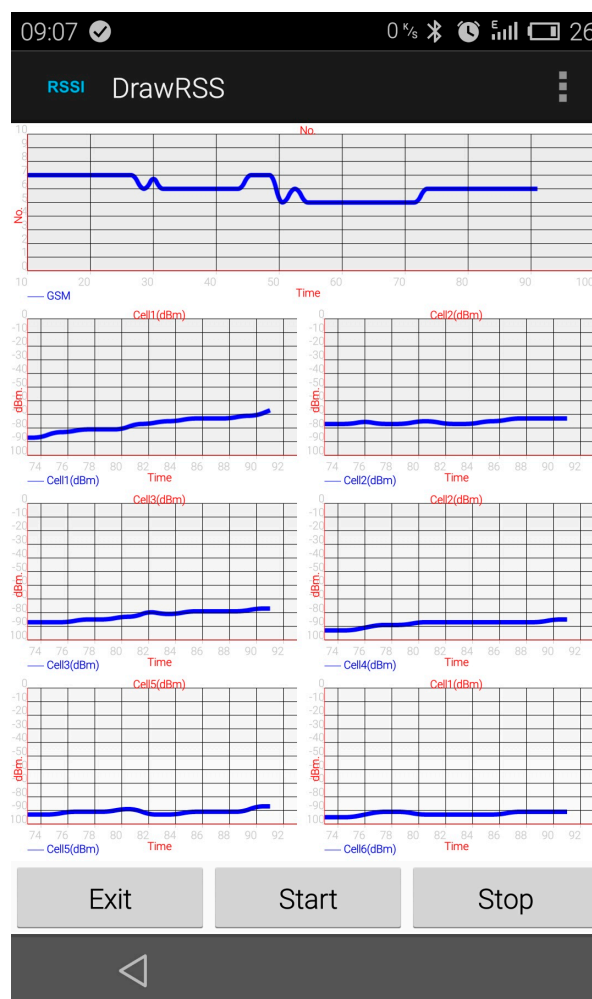

Figure 2. Screenshot of the DrawRSS application. In this application, all the neighboring cellular base stations' signal strengths are drawn on the canvas and recorded in the phone's memory.

The signal strength is recorded every 0.5 s, and then drawn oin the user interface, so the sampling rate is 2 Hz. The application records the number of cellular base stations and their signal strengths. The data are called training data. The testing environments in this experiment are shown in Figure 3.

(**a**) open outdoors

(**b**) semi-outdoors

(**c**) light indoors

(**d**) deep indoors

Figure 3. The four testing environments. These four environments are located on the same campus. The largest distance between them is less than 400 m.

During the experiment, the volunteer stayed in the four environments mentioned above for about 10 min. He could use his phone in the same manner he naturally would. He kept moving around in the environment. For example, in the light indoors environment, he walked in the room randomly. The training data are collected during these 10 min. In the deep indoors environment (Figure 3d), the volunteer moved from one side of the corridor to the other side several times.

Figure 4 is the number of cellular base stations in different environments recorded by the application. This figure only shows the result beyond the first 200 s.
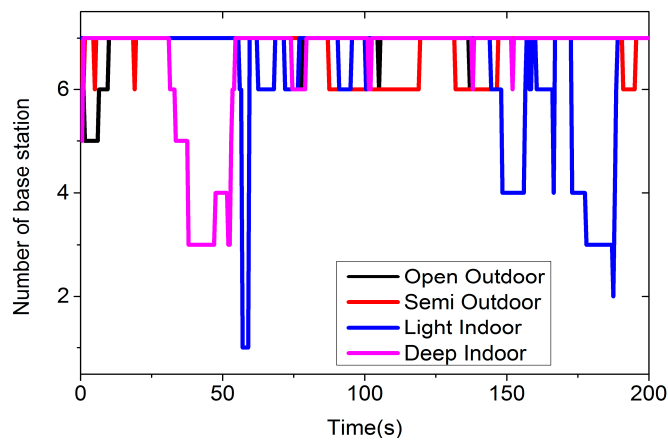

Figure 4. The number of cellular base stations in different environments.

In Figure 4, we can see that the number of cellular base stations is quite different from the open outdoors to deep indoors environment. In the open outdoors environment, we have seven stations

most of the time, while in the semi-outdoors environment, we often have six stations. In the indoors environment, the number of stations varies from 1 to 7. Figure 5 is the probability curves for different environments. Every node (*x*, *y*) in Figure 5 means there is *y*% probability of having at least *x* cellular base stations.
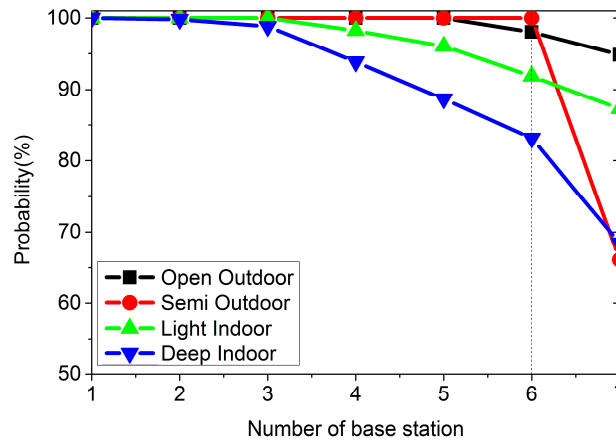

Figure 5. The probability curve of the number of GSM cellular base stations.

From Figure 5, we can see that we have at least one station in the specified four environments. If we want six cellular base stations, the probability is 98.02%, 100%, 91.87%, and 83.1% in the open outdoors, semi-outdoors, light indoors and deep indoors environments, separately. On average, the probability of having at least 1 to 7 base stations in any environment is 100%, 99.95%, 99.70%, 98.02%, 96.18%, 93.25%, and 79.17%, respectively.

We filtered out the samples with less than six cellular base stations and compare the signal strength received in the different environments in Figure 6. The widely used Log-Distance Path Loss (LDPL) signal propagation model [23] tells us that the signal strength is affected mainly by the distance. As a result, we will have the same signal strength measures in different environments. In Figure 6, we can find many examples, but if we take more stations' signal strength into consideration, and we look at how the signal strength changes with time, we will find that different environments show different patterns. From Figure 6, we can see that the signal strength received in these four environments is quite different. These results imply that it is possible to identify different environments using the cellular base stations' signal strength.

Before performing a classification using the raw data, we pre-process these data as mentioned above. The raw data are grouped according to different windows varies from 1 to 20 s. We calculate the features in each window, including Mean, Standard Deviation, Maximum, Minimum, and Range.
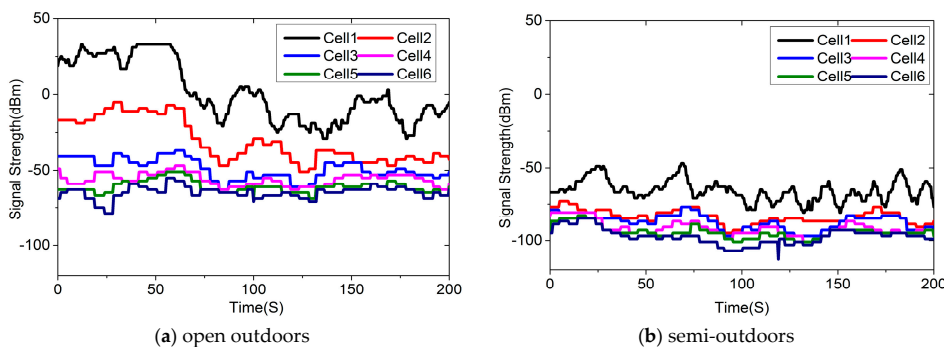

(**a**) open outdoors   (**b**) semi-outdoors

Figure 6. *Cont.*

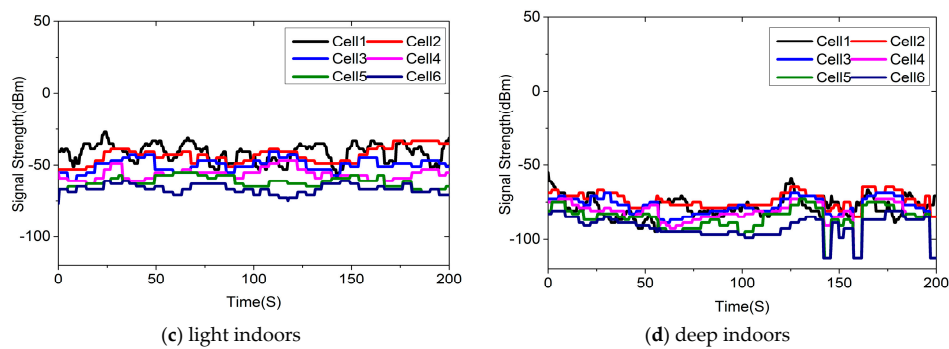(**c**) light indoors　　　　　　　　　　　　　　　　(**d**) deep indoors

Figure 6. Six cellular base stations' signal strength received in different environments.

4. Data Analysis

The input data, including measured data and the pre-processed data are both ready for classification using the machine learning algorithms. In this section, we investigate the use of a wide range of machine learning algorithms to classify the training data, including Decision Tree, Random Forest, Support Vector Machine, K Nearest Neighbor, Logistic Regression, Naive Bayesian and Neural Network. The classifiers are applied for recognition. The classification performance will be compared between these Machine Learning algorithms.

In this paper, learning and classification is conducted by the Orange data mining toolkit [24]. Orange is an easy-to-use machine learning toolkit, which allows us to perform repeat model training using a wide range of machine learning algorithms and employing standard performance testing techniques, such as cross-validation without programming. Figure 7 shows the work flow of this experiment.
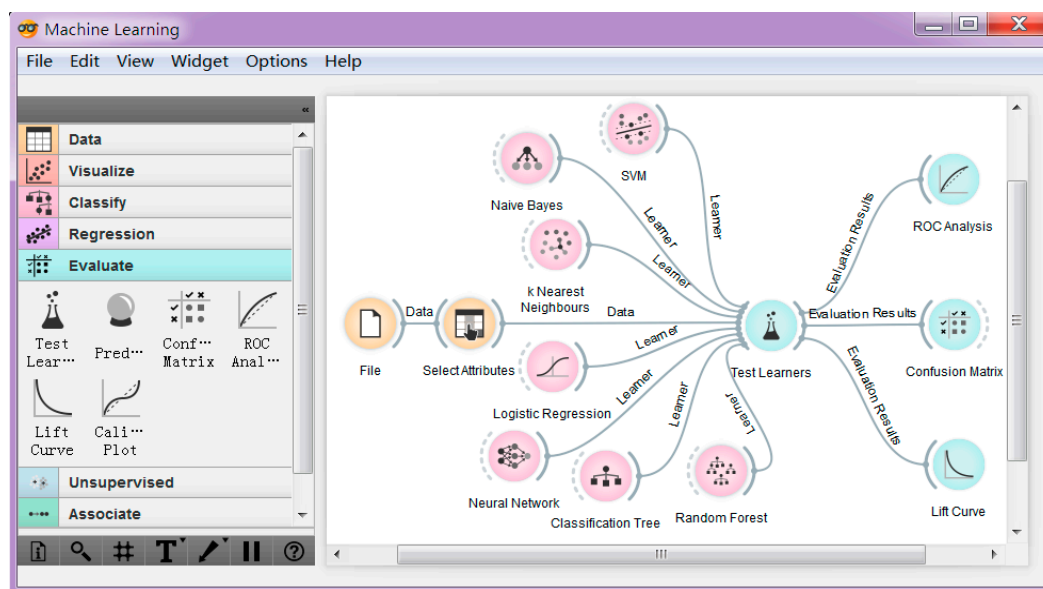

Figure 7. Workflow of the data classification using Orange toolkit.

In Figure 7, seven classification algorithms are created by dragging the corresponding widgets to the canvas. The file widget reads data from disk. Firstly, we apply the raw data for classification. The data are collected every 0.5 s in the four environments for 10 min, respectively. Figure 8 shows the classification accuracy. In Figure 8, we can see that for most of the classification algorithms, the more stations used, the better the accuracy is. KNN performs the best among the seven algorithms, followed by Decision Tree and Random Forest algorithm. Neural Network performs the worst. We just need four cells' signal strength to get the best accuracy using KNN.
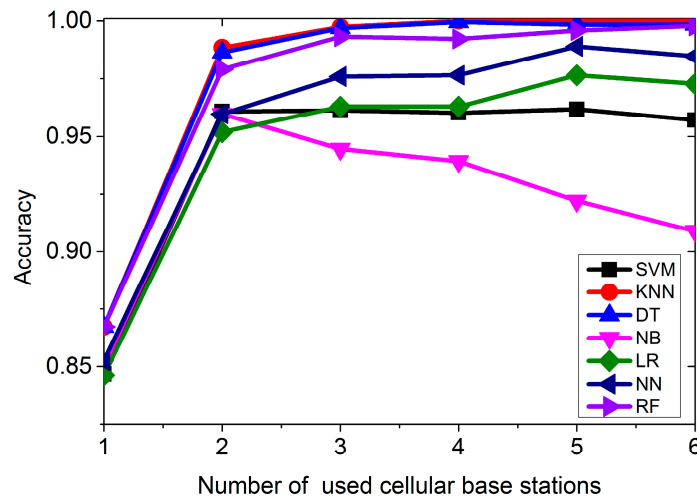
Figure 8. Classification accuracy using raw data.

However, from Figure 5 we know that we cannot always receive signal from at least four stations. On average, the probability of receiving at least four cellular stations is 98.02% for all the cases. However, we will surely have at least one station in any case. In the following part of this section, we are going to using just one cellular base station's signal strength for classification.

As mentioned above, we can calculate the features for different window lengths and apply the features for classification. The window length is varied from 1 to 20 s and we calculate the features in each window length, including Mean, Standard Deviation, Maximum, Minimum, and Range. In each window length, the number of instance is different. For example, there are 600 instances if the window length is 1 s. Figure 9 shows the accuracy of classifying the features.
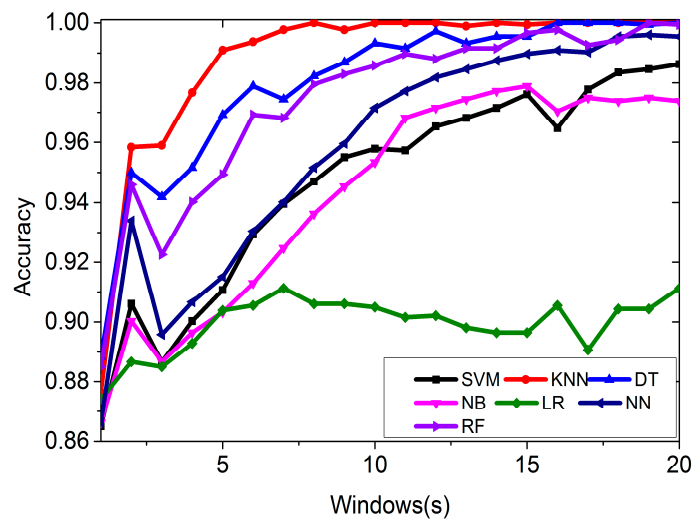

Figure 9. Classification accuracy using features varying with window length.

From Figure 9, we can see that in most of the cases, the longer the window length is, the better the accuracy gets. Once again, KNN performs the best among all the algorithms, followed by the Decision Tree and Random Forest algorithms. Logistic Regression performs the worst. When the window length is 8 s, we can correctly classify all four environments using the KNN algorithm.

For testing this algorithm under different conditions, we experimented during one week with five different walking traces during the period from 9:00 to 17:00 under different weather conditions. These traces are different from the environments where we collected the data to generate the classifier.

These walking traces contain nine open outdoors segments, 11 semi-outdoors segments, 10 light indoors segments, and 11 deep indoors segments. Figure 10 shows one of the walking traces that we experimented with.
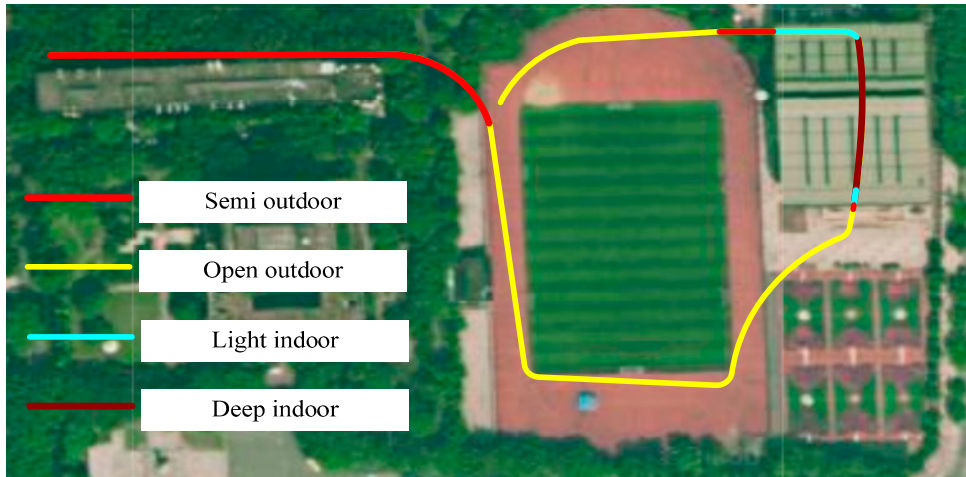

Figure 10. Testing trace located in the university campus. This trace contains two open outdoors segments, three semi-outdoors segments, two light indoors segments, and one deep indoors segment. The volunteer walks along the path in 10 min.

In the experiments, the volunteer walks along this path while using his phone in the same manner he naturally would. The true environment type is manually labeled. Each day, these traces are tested three times in the morning, noon and afternoon, respectively.

We first use the four cellular base stations' signal strength for testing. The instances with less than four stations are filtered out. Figure 11 shows the average accuracy for different classifiers.
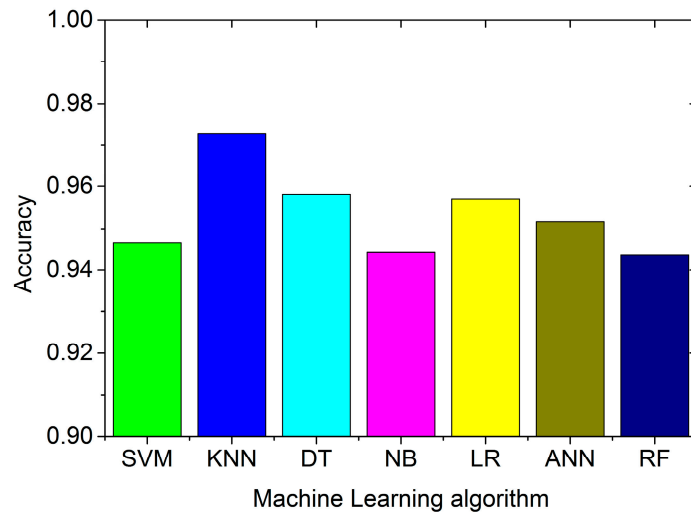

Figure 11. Classification accuracy using four stations' signal strength with different classifiers created by different machine learning algorithms.

From Figure 11, we find that all the algorithms have accuracies better than 94%. KNN performs the best, with an average accuracy is 97.27%. Random Forest performs the worst, as its accuracy is

94.36%. The confusion matrix for KNN algorithm is shown in Table 4. The value a_{ij} in this confusion matrix is not the number of the sample, but the percentage, given by the following expression:

$$a_{ij} = n_{ij}/n_i \times 100\% \tag{8}$$

Table 4. Confusion matrix using four stations' signal strength with KNN classifier.

Environment	Deep Indoors	Semi-Outdoors	Light Indoors	Open Outdoors
Deep Indoors	97.1%	0	2.9%	0
Semi Outdoors	0	98.6%	0	1.4%
Light Indoors	6.5%	0	93.5%	0
Open Outdoors	0	0	0	100%

Table 4 shows that we can detect open outdoors environments correctly. There is a 2.9% possibility of identifying deep indoors as light indoors environments, and a 1.4% possibility of identifying semi-outdoors as open outdoors environments. A light indoors environment might be classified as deep indoors with a possibility of 6.5%. However, we can't always have at least four cellular stations. Figure 9 show that when the window length is 8 s, we can correctly classify all four environments using the KNN algorithm. In Table 5, we apply the different classifiers to detect the environment using 8 s window length.

From Table 5, we can see that Random Forest algorithm performs the best in all the measures. In the following comparison, we will apply the Random Forest algorithm for classification. Finally, we compare the proposed indoor/outdoor detection algorithm with the IODetector [2], Co-Training [3] and GPS based detection in terms of accuracy and energy consumption.

Table 5. Performance measures using 8 s window length with different algorithms.

Algorithm	Accuracy	Sensitivity	Specificity	F-Measure	Precision
SVM	0.8095	0.8095	0.9365	0.8077	0.8087
KNN	0.8652	0.8652	0.9551	0.8638	0.8634
DT	0.8861	0.8861	0.9621	0.8856	0.8901
NB	0.8734	0.8734	0.9578	0.8719	0.8717
LR	0.7994	0.7994	0.9331	0.7932	0.8051
NN	0.8677	0.8677	0.9559	0.8658	0.8652
RF	0.8943	0.8943	0.9647	0.8938	0.8933

The app DrawRSS is updated to collect other required signal strengths, including light intensity, cellular signal strength, magnetic variance, sound intensity, and visible GPS satellites. A screen shot of the application is given in Figure 12.

We collect the required signal strength from the four environments for 10 min. The sampling rate is 2 Hz. The IODetector use the light intensity, cellular signal strength and magnetic variance to detect the indoor/outdoor environment. Co-Training uses the light intensity, cellular signal strength and sound intensity for detection. Our proposed algorithm uses the Random Forest algorithm to classify 8 s window length features from one cellular station's signal strength. These four algorithms can detect different indoor/outdoor environments: co-Training and a GPS-based algorithm are proposed to detect indoor and outdoor environments, while IODetector is capable of detecting outdoors, semi-outdoors, and indoors; our proposed algorithm can detect four kind of different indoor/outdoor environment. In this experiment, all four of these algorithms only detect indoors and outdoors environments. Table 6 shows the results, confirming that the proposed algorithm performs the best among the four algorithms.
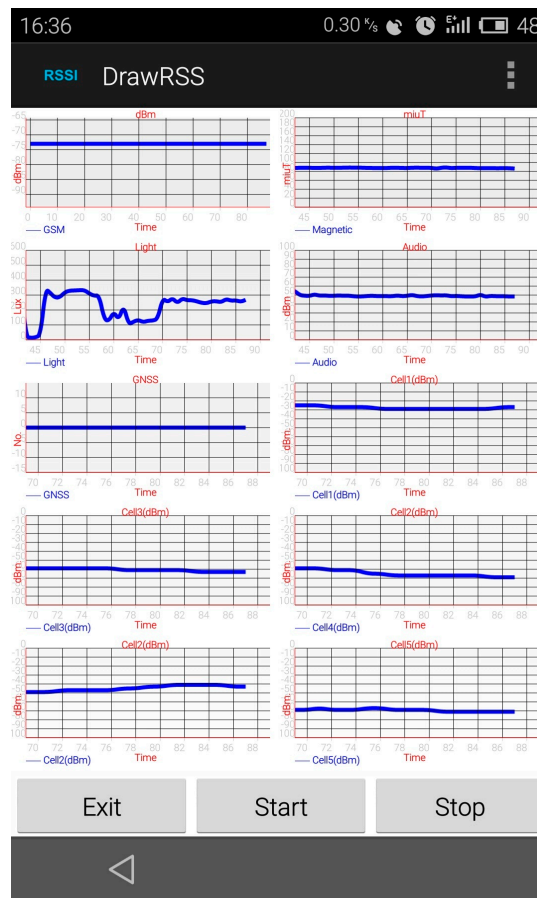
Figure 12. Screen shot of the updated application. The updated application records the light intensity, cellular signal strength, magnetic variance, sound intensity, and visible GPS satellites in the memory.

Table 6. Accuracy comparison among the proposed algorithm, IODetector, Co-Training and GPS based detection.

	Random Forest	IODetector	Co-Training	GPS
Accuracy	95.3%	61.2%	93.14%	71.6%

Different indoor/outdoor algorithms require different sensors. As a result, they consume different amounts of energy. According to [4], GPS consumes the most energy, followed by microphone, light sensor, and magnetic sensor. GSM consumes the least energy. Compared with the other indoor/outdoor detection algorithms, our proposed algorithm only needs the GSM sensor, which consumes minimal energy in addition to standard cell-phone operation, so the proposed algorithm is the most energy efficient among the four detection algorithms.

5. Discussion and Conclusions

Detecting the users' current environment automatically is crucial for SNAL. In this paper we propose a GSM-based indoor/outdoor detection algorithm. Some studies [25,26] show that both the distance from the base station and the difference in a height of mobile phones influence the received signal strength. Our idea is to not just rely on the signal strength. We apply a machine learning algorithm to classify the neighboring GSM station's signal in different environments, and identify the users' current context by signal recognition. We test the algorithm in four different environments. The results show that the proposed algorithm is capable of identifying open outdoors, semi-outdoors,

light indoors and deep indoors environments with 100% accuracy using four nearby GSM stations' signal strength. Large scale experiments have proved the efficiency of the algorithm. The required hardware and signals are widely available in our daily lives, implying its high compatibility and availability. Future work will concentrate on real-time indoors/outdoors detection by introducing new low energy consumption signals that exist pervasively. The proposed approach was tested on only one phone. However, GSM RSSI may be different on different phones even for the same location. The device diversity problem is the one we must focus on in the future.

Author Contributions: Li Qun conceived and designed the experiments; Wang Weiping performed the experiments; Shi Zesen and Chen Wei analyzed the data. Chang Qiang wrote the paper.

References

1. Groves, P.D. Shadow matching: A new GNSS positioning technique for urban canyons. *J. Navig.* **2011**, *64*, 417–430. [CrossRef]

2. Cheng, J.; Yang, L.; Li, Y.; Zhang, W. Seamless outdoor/indoor navigation with WIFI/GPS aided low cost Inertial Navigation System. *Phys. Commun.* **2014**, *13*, 31–43. [CrossRef]

3. Li, M.; Zhou, P.; Zheng, Y. IODetector: A Generic Service for Indoor/Outdoor Detection. *ACM Trans. Sens. Netw.* **2014**, *11*. [CrossRef]

4. Dadu, V.; Katasikouli, P.; Sarkar, R.; Marina, M.K. A Semi-Supervised Learning Approach for Robust Indoor-Outdoor Detection with Smartphones. In Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems, Memphis, TN, USA, 3–6 November 2014.

5. Kaplan, E.D. *Understanding GPS: Principles and Applications*; Artech House: London, UK, 2005.

6. Caceres, M.A.; Penna, F.; Wymeersch, H.; Garello, R. Hybrid cooperative positioning based on distributed belief propagation. *IEEE J. Sel. Areas Commun.* **2011**, *29*, 1948–1958. [CrossRef]

7. Chen, W.; Wang, W.; Li, Q.; Chang, Q.; Hou, H. A Crowd-Sourcing Indoor Localization Algorithm via Optical Camera on a Smartphone Assisted by Wi-Fi Fingerprint RSSI. *Sensors* **2016**, *16*, 410. [CrossRef] [PubMed]

8. Bahl, P.; Padmanabhan, V.N. RADAR: An in-building RF-based user location and tracking system. In Proceedings of Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, Tel Aviv, Israel, 26–30 March 2000.

9. Walter, D.J.; Groves, P.D.; Mason, R.J.; Harrison, J.; Woodward, J.; Wright, P. Novel Environmental Features for Robust Multisensor Navigation. In Proceedings of the 26th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2013), Nashville, TN, USA, 16–20 September 2013.

10. Groves, P.D.; Martin, H.F.S.; Voutsis, K.; Walter, D.J.; Wang, L. Context detection, categorization and connectivity for advanced adaptive integrated navigation. In Proceedings of the 26th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2013), Nashville, TN, USA, 16–20 September 2013.

11. Muralidharan, K.; Khan, A.J.; Misra, A.; Balan, R.K.; Agarwal, S. Barometric phone sensors: More hype than hope! In Proceedings of the 15th Workshop on Mobile Computing Systems and Applications, Santa Barbara, CA, USA, 26–27 February 2014.

12. Ravindranath, L.; Newport, C.; Balakrishnan, H.; Madden, S. Improving wireless network performance using sensor hints. In Proceedings of the 8th USENIX Conference on Networked Systems Design and Implementation, Boston, MA, USA, 30 March–1 April 2011.

13. Liu, J.; Priyantha, B.; Hart, T.; Ramos, H.S.; Loureiro, A.A.F.; Wang, Q. Energy efficient GPS sensing with cloud offloading. In Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems, Toronto, ON, Canada, 6–9 November 2012; pp. 85–98.

14. Lin, K.; Kansal, A.; Lymberopoulos, D.; Zhao, F. Energy-accuracy trade-off for continuous mobile device location. In Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, San Francisco, CA, USA, 15–18 June 2010; pp. 285–298.

15. Elhoushi, M.; Georgy, J.; Wahdan, A.; Korenberg, M.; Noureldin, A. Using portable device sensors to recognize height changing modes of motion. In Proceedings of the 2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Montevideo, Uruguay, 12–15 May 2014; pp. 477–481.

16. Elhoushi, M. Advanced Motion Mode Recognition for Portable Navigation. Ph.D. Thesis, Queen's University, Kingston, ON, Canada, 2015.

17. Elhoushi, M.; Georgy, J.; Korenberg, M.; Noureldin, A. Robust motion mode recognition for portable navigation independent on device usage. In Proceedings of the 2014 IEEE/ION Position, Location and Navigation Symposium (PLANS 2004), Monterey, CA, USA, 5–8 May 2014; pp. 158–163.

18. Vanini, S.; Giordano, S. Adaptive context-agnostic floor transition detection on smart mobile devices. In Proceedings of the 2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), San Diego, CA, USA, 18–22 March 2013; pp. 2–7.

19. Krumm, J.; Hariharan, R. TempIO: Inside/outside classification with temperature. In Proceedings of the 2nd Workshop on Man-Machine Symbiotic Systems, Kyoto, Japan, 23–24 November 2004.

20. Wu, M.; Pathak, P.H.; Mohapatra, P. Monitoring building door events using barometer sensor in smartphones. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan, 7–11 September 2015; pp. 319–323.

21. Ergen, S.C.; Tetikol, H.S.; Kontik, M.; Sevlian, R.; Rajagopal, R.; Varaiya, P. RSSI-fingerprinting-based mobile phone localization with route constraints. *IEEE Trans. Veh. Technol.* **2014**, *63*, 423–428. [CrossRef]

22. Chen, R.; Chu, T.; Liu, K.; Liu, J.; Chen, Y. Inferring Human Activity in Mobile Devices by Computing Multiple Contexts. *Sensors* **2015**, *15*, 21219–21238. [CrossRef] [PubMed]

23. Rappaport, T.S. *Wireless Communications: Principles and Practice*; Prentice Hall PTR: Upper Saddle River, NJ, USA, 1996; Volume 2.

24. Orange. Available online: http://www.ailab.si/orange/ (accessed on 3 March 2016).

25. Stanivuk, V. Measurements of the GSM signal strength by mobile phone. In Proceedings of the 2012 20th Telecommunications Forum (TELFOR), Belgrade, Serbia, 20–22 November 2012; pp. 1784–1787.

26. Kyritsi, P.; Cox, D.C. Propagation characteristics of horizontally and vertically polarized electric fields in an indoor environment: Simple model results. In Proceedings of the IEEE 54th Vehicular Technology Conference VTC Fall 2001, Atlantic City, NJ, USA, 7–11 October 2001; Volume 3, pp. 1422–1426.