

Toward an Easy Deployable Outdoor Parking System – Lessons from Long-term Deployment

Yu Huang*, Yi-Chao Chen[†], Chuang-Wen You*, Dian-Xuan Wu^{*,‡}, Yi-Ling Chen*, Kai-Lung Hua[‡],
Jane Yung-Jen Hsu*

*National Taiwan University, Taipei, Taiwan

Email: {b02902084, cwyou}@ntu.edu.tw, yiling@arbor.ee.ntu.edu.tw, yjhsu@csie.ntu.edu.tw

[†]Shanghai Jiao Tong University, Shanghai, China

Email: yichao@cs.utexas.edu

[‡]National Taiwan University of Science & Technology, Taipei, Taiwan

Email: {m10315014, hua}@mail.ntust.edu.tw

Abstract—Data pertaining to the availability of parking slots is crucial to the efficient operation of systems designed to monitor the state of parking spaces. Outdoor parking systems have been developed using wireless sensors, Internet of Things (IoT) technology, and cameras. Unfortunately, interference from electromagnetic fields complicates the tuning of parameters for detection algorithms and limits accuracy to only 90 percent. In this study, we investigated these problems by collecting data from magnetic sensors, light sensors, and LoRa wireless modules used in the detection transient events (car arrivals and departures) over a period of 13 months. This led to the design an adaptive occupancy detection system using a variety of sensors, which can be deployed with only minimal calibration.

I. INTRODUCTION

Data pertaining to the availability of parking slots is crucial to the efficient locating of empty parking spaces [1]. Existing outdoor parking systems are based on wireless sensors, Internet of Things (IoT) technology [12], and cameras [24]. IoT-based systems developed by Streetline Inc. [17] and in the SFpark project [14] have been deployed in North America and European cities. However, interference from electromagnetic fields [12] causes frequent false positives; i.e., degraded accuracy. Discussions with experts in IoT technology revealed that magnetic sensors are prone to false detections when deployed as parking nodes. Thus, systems tend to be retrained only in response to complaints from customers or within a schedule wherein updates are conducted more frequently than necessary.

Most existing parking solutions dealing with parking occupancy rely on magnetic sensors to identify the arrival and/or departure of vehicles. Current state-of-the-art parking solutions, such as FastPrk [5] and Fastprk2 [20] are reported to provide accuracy of 95% [6] ~ 98% [7]; however, these estimates are based on short-term data collection often within a laboratory environment [6], [13]. Furthermore, administrators replaced a FastPrk system deployed in Moscow with another system after the initial trials [15], citing shortcomings in the system. The three main challenges in the design of parking solutions based on magnetic devices are (1) environmental changes, (2) the spatial characteristics of individual parking slots, and (3) time-varying factors, such as variations in the geomagnetic field or dwindling battery life. Maintaining the

accuracy of such systems requires that the model be retrained; however, existing systems lack a mechanism by which to alert operators when retraining is required. Thus, we conducted an empirical study to identify challenges and limitations in the use of magnetic, light-based, and LoRa wireless modules for the detection of *transient events* (including events of car arrivals and departures) in a real-world environment. Long-term multi-slot deployment was conducted to reveal weaknesses in current learning-based parking systems. To the best of our knowledge, previous reports on these systems employed learning-based methods, and none of them discussed system performance over extended durations or in different locations. Our long-term assessment of sensors in multiple parking spaces revealed the need for an adaptive occupancy detection algorithm to complement current state-of-the-art parking solutions.

In this study, we explored the challenges and limitation to understand the potential design considerations of designing easy-deployable and manageable IoT-based parking systems. Through collecting and analyzing the 13-month data collected from magnetic sensing, light sensing, and LoRa wireless modules, we propose a feasible adaptive design to enable easy-deployable parking solutions. The goal of this paper is to highlight the need of adaptive machine learning schemes when designing parking occupancy detection system to encourage major outdoor parking solution providers continuing to enable future easy-deployable solutions.

The contributions of this study are three-fold. First, we elucidated the strengths and weaknesses of three major sensing modalities in real-world scenarios. Second, to overcome imprecision in the collected data, we developed three adaptive machine learning schemes: model selection, sensor selection, and failure detection. These machine learning schemes led to the development of an adaptive occupancy detection system using three types of sensors in order to eliminate the need for major calibration. Third, we conducted a series of evaluations using a variety of metrics (accuracy, precision, recall, and F1-score) to validate the performance of the proposed adaptive machine learning schemes in the detection of transient events.

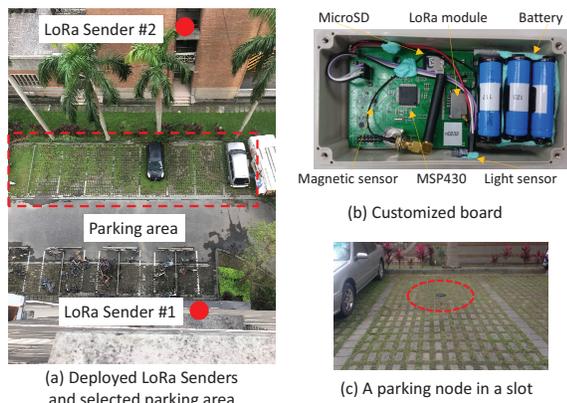


Fig. 1: LoRa sensors and proposed parking nodes. In (a), only Sender #1 (2nd floor of the building: top) and Sender #2 (2nd floor of another building: bottom) are presented. In (c), the node is half-buried in the center of the slot.

II. RELATED WORK

Existing smart parking solutions for outdoor environments [17], [14], [22], [5] use magnetic sensors to identify the arrival or departure of cars. In [30], researchers of the pilot SFpark project listed car models and environmental interference as challenges in the use of magnetic sensors to detect vehicle arrivals and departures. Streetline [17] uses camera sensors in conjunction with or in lieu of in-ground magnetic sensors to reduce the effects of electrical interference [16]. Unfortunately, this approach is limited to line-of-sight scenarios and is susceptible to changes in lighting conditions, thereby necessitating the deployment of a large number of cameras. ADEC technologies [2] recently announced a parking occupancy detector that incorporates additional sensors; however, they have released no details of the proposed system. The inclusion of additional sensors no doubt improves detection accuracy; however, it greatly increases power consumption and capital overhead. Worldsensing [23] recently incorporated server-side service layers in the development of a new generation parking system (i.e., Fastprk-2 [20]). This system is based on the occupancy sensors used in the Fastprk system in conjunction with advanced signal processing algorithms for the detection of occupancy using information obtained via data mining. It is reported to have accuracy of 95% ~ 98%; however, these results are based on short-term evaluations under laboratory conditions [15].

Researchers have used a variety of alternative sensing techniques, such as passive acoustic array sensors, passive infrared sensor, RFID, ultrasonic and video image processing [26], [28], [29]. Unfortunately, some of these devices (e.g., RFID, ultrasonic signals, or visual signals) require expensive overhead installations and on-going maintenance. The fact that most situations preclude the installation of overhead devices has led to the development of inductive loops, piezoelectric cables, and weigh-in-motion sensors [26]. This commonly requires the cutting of pavement for installation and necessitates the use of multiple detectors. Other solutions involve the

Module	Average current	Sampling energy (time)
System	Low-power: 22.06 μ A Active: 5.54mA	-
Magnetic	2.40mA	12.68mJ (1.6ms)
LoRa	Receiving: 35.88mA Sending: 76.00mA	Receiving: 1.78J (15ms) Sending: 3.16J (12.6ms)
Light	3.75mA	745.63mJ (250ms)

TABLE I: Average current and energy drawn by the system, magnetic sensing, LoRa wireless, and light sensing module.

detection of signals reflected from vehicles (e.g., microwave radar or ultrasonic ranging sensors); however, this approach greatly increases power consumption and the such systems are highly susceptible to environmental interference.

ParkNet [28] uses an ultrasonic rangefinder to detect parking spot occupancy. Each location is associated with GPS coordinates, thereby making it possible to pinpoint the location of each car and generate an occupancy map of the entire city. ParkSense [29] infers unparking events by observing changes in the Wi-Fi signatures from the mobile phones of drivers. Unfortunately, these crowd-sourcing approaches require participation by a substantial proportion of drivers and are generally less reliable than other methods.

III. EXPERIMENT: LONG-TERM DATA COLLECTION

Our aim was to identify the factors that affect accuracy in determining the occupancy of parking spaces. Long-term readings were obtained from magnetic sensors and other low-cost devices.

A. Experimental Setup

Customized parking nodes were deployed in four adjacent parking lots on the campus of National Taiwan University (Fig. 1(a)). To ensure easy deployment and maintenance, we opted for the single-point deployment of low-cost, low-power sensors for the detection of transient events. The parking nodes were housed in a rectangular water-proof case containing a circuit board powered by four Lithium-ion batteries.

As shown in Fig. 1(b), each circuit board was equipped with a low-power TI MSP430 microcontroller [21], a Honeywell HMC5883 3-axis compass sensor [9], a TSL2561 light-to-digital converter [19], and an ARCT TM1276 LoRa transceiver module [4]. Table I lists the average current and energy consumption of the bare system and sensing modules. The magnetic sensors and light sensors consume far less energy (12.68 and 745.63 milli-joules) than do the LoRa wireless modules (3.16 joules for sending and 1.78 joules for receiving a 4-byte packet). Each parking node was placed in the center of the parking lot to reduce interference from vehicles in adjacent lots. Note that numerous low-cost sensing alternatives are available; however, we selected a combination of sensors (i.e., a magnetic sensor, a light sensor, and the wireless LoRa modules) to demonstrate the application of a sensor selection algorithm to achieve a trade-off between high accuracy and low power consumption. In the event that other sensors are used in the parking nodes, it should be possible for the

same sensor selection algorithm to switch between any new combination of sensors.

A surveillance camera was mounted on an outside wall on the 7th floor of a nearby building to monitor the parking area. This data was used as ground-truth information pertaining to parking events. All recorded footage was streamed back to a server via a wired link and stored. We envision that a few internet-connected gateways (IoT gateways) could be scattered throughout the city in the future, thereby ensuring reliability in the relaying of data to a central network server. To emulate this situation, we installed a number of IoT gateways around the selected parking area to enable the exchange of messages between nodes and the backend. Three LoRa senders deployed in two buildings surrounding the area broadcast messages every 500ms to be received by LoRa modules on the parking nodes. LoRa Senders #1 and #2 were located on the 2nd floor of the two building 15 meters from the closest parking node to emulate gateways deployed in positions at the same height as street lights. LoRa Sender #3 was located on the 5th floor of a building at a distance of 40 meters from the closest parking node. All readings from the sensing and communications modules were recorded on an on-board micro SD card and then manually retrieved via a serial interface every two days

B. Characterizing Transient Events

Between April 2015 and April 2016, we collected 7175 hours of parking data involving 1276 transient events; i.e., 666 arrivals and 610 departures for an average of 51 arrival events and 46 departure events per month. The MSP430 micro-controller sampled readings from magnetic sensors and light sensors. RSSI values in the header of LoRa packets achieved a best-effort sampling rate of 250 ms per second. This was limited by the intensity of the light received by the sensor [19]. Each transient event was associated with changes in the patterns read from magnetic sensors, light sensors, and LoRa wireless modules.

Magnetic sensor: Magnet-based detection methods can be used identify parking events by responding to changes in the patterns of signals. Typically, changes in the magnetic field can be attributed to the movement of a car and the engine being shut on/off. The large amount of ferrous materials in a car body means that changes in the magnetic field are easily affected by vehicles entering or leaving, but also by nearby cars [3]. In obtaining samples, the sensor decomposes the Earth's magnetic field and outputs individual components over the x , y , and z axes relative to the local coordinate frame of a compass. Combining the individual components into a vector makes it possible to obtain an aggregate vector.

Light sensor: Changes in the intensity of incident light at parking nodes makes it possible to infer the departure or arrival of a car. This study used a wide light spectrum containing visible and infrared light, using digital output signals between 0 ~ 65535 from the broadband channel of a TSL2561 [19]. Placement of a light sensor in the center of a slot would mean that is intermittently covered by cars entering or leaving the

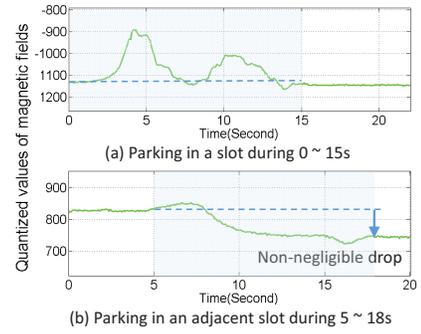


Fig. 2: Readings collected from a magnetic sensor along the z axis in two scenarios.

slot. As indicated by the red line in Fig. 3, the digital output decreased by more than 30000 when a car parked at 47 s.

LoRa module: When a parking node is deployed in the center of a parking slot, the transmission path between surrounding LoRa senders and the parking node is blocked by passing cars, resulting in the attenuation of the received signal strength (RSSI). Thus, the passage of cars can be interpreted as transient events. Fig. 4(a) shows RSSI values received by LoRa Sender #1 at a distance of 15 meters, when the parking slot is occupied, empty, or soon to be occupied by a car (e.g. the 12 dBm difference in signal strength at 50s).

Characterization of changes in the patterns collected from the sensors began with a moving average pre-processing step to filter out corrupted samples. We then extracted feature segments using a sliding window (empirical-determined 15-second duration, with 50% overlap). The presence of a car is determined through the extraction of statistical features (mean, median, mode, and range). Our aim was to express the central tendency and statistical dispersion of the magnetic field, light intensity, and RSSI values as well as all derivatives of the individual components (or aggregated vectors) within a data segment. This included a total of 108 magnetic features, 30 light features, and 90 LoRa features (i.e., 30 LoRa features for each sender). Based on these features, the proposed system adaptively selects a sub-optimal set of features as an indication of a transient event, as described in Section V-C.

IV. INITIAL OBSERVATIONS: CHALLENGES AND LIMITATIONS OF EACH SENSING MODALITY

An easy-deployable parking system should provide 24/7 operational coverage and enable the accurate detection of a range of vehicle models. The long-term data set described in Section III made it possible to identify factors capable of interfering with sensor readings: (1) environmental factors, (2) deployment factors, and (3) target-vehicle factors.

A. Interference Cases Caused by Environmental Factors

Environmental factors, such as weather conditions or nearby vehicles, can affect magnetic sensors, light sensors, and LoRa wireless modules.

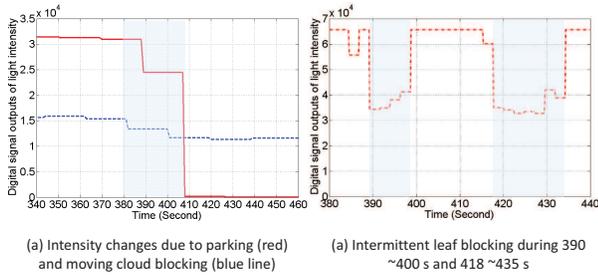


Fig. 3: Readings collected from light sensors when a car was parking in a slot (the red line in (a) and (b)), in which sunlight is gradually blocked by slow-moving clouds (the blue line in (a)), and intermittently blocked by leaves moving in a nearby tree (the red line in (b)).

1) *Magnetic sensor*: The earth’s magnetic field varies with location and over time. It is also influenced by nearby ferrous-metal objects (e.g., bicycles or adjacent vehicles) [3]. The impact on the magnetic field varies according to the amount of ferrous metal in vehicles and the distance from the parking node. Fig. 2(b) illustrates the effects of interference from a car parking in an adjacent slot. Changes in measurements along the z -axis are non-negligible, which means that they could be interpreted as a non-normal transient event (Fig. 2(a)).

2) *Light sensor*: Outdoor lighting conditions are highly variable. The arrival and departure of cars during the day can result in dramatic changes in lighting intensity over a very brief period, as shown by the blue line in Fig. 3(a). As the sun moves across the sky, the light can also be blocked by stationary objects. Readings collected by light sensors when a car was parking in a slot (the red line in Fig. 3(a)), show how sunlight was blocked by slow-moving clouds (the blue line in Fig. 3(a)), or intermittently blocked by leaves moving in a nearby tree (the red line in 3(b)). Observations of previous detection failure demonstrate that existing methods often produce patterns that do not necessarily correspond to an actual parking event.

3) *LoRa module*: Due to multipath effects in the transmission of radio signals, any objects (e.g., nearby humans or cars) within a given environment can have an impact on the received signal strength of LoRa modules. The degree to which RSSI is altered by vehicles passing nearby (red cross in Fig. 4(b)) tends to be less pronounced than that caused by the arrival or departure of cars to/from the target slot (blue circles in Fig. 4(b)). Conversely, the degree to which RSSI is altered by the arrival or departure of target vehicles increases with a higher *baseline RSSI* (defined as the RSSI measured by a parking node that is not covered by vehicles in the target slot).

Fig. 4(b) shows that the change in RSSI caused by the arrival or departure of a car is more pronounced than that caused by the arrival or departure of cars in nearby slots when the baseline RSSI exceeds -90dBm. This is an indication that RSSI data could be used to indicate the occurrence of transient events. The presence of human or other environmental objects can also affect the received signal strength. The degree to which RSSI changes is determined by the size and location

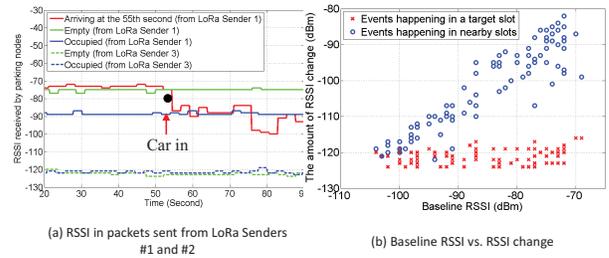


Fig. 4: Values of RSSI in packets collected by LoRa modules in a parking node deployed in a target slot under the following scenarios: (a) RSSI in packets sent by Sender #1 when the slot was empty (green line), occupied (blue line), or initially empty and later occupied (red line), and RSSI in packets sent by Sender #2 when the slot was empty (dotted green line) or occupied (dotted blue line). (b) Baseline RSSI vs. change in RSSI of a packet sent from senders caused by target vehicles or nearby vehicles.

of objects, and this commonly leads to false detections. This means that a sensor selection algorithm is required to avoid the effects of unreliable sensing readings, as outlined in the following section.

B. Interference Caused by Deployment Factors

The re-deployment of the parking nodes every two days led to slight variations in the orientation of the sensors, and these effects were particularly pronounced when the devices were placed in a new parking area.

1) *Magnetic sensor*: Slight differences in the orientation of a parking node produces small changes in the magnetic field projected onto individual components (i.e., x , y , and z axes of a compass local coordinate frame). Local changes in the local environment (e.g., nearby bicycles or parked cars) or global shifts in the earth’s magnetic field can also alter the readings. A review of the raw data from each deployment revealed slight variations in the baseline magnetic field readings (i.e., the magnetic field changed when the slot was empty).

2) *Light sensor*: The light sensors were housed behind a transparent acrylic window to protect the parking nodes from water. Unfortunately, the effects of morning dew or dust on the window altered the transmittance of light. When transmittance dropped considerably, the intensity of light reaching the sensor was insufficient to differentiate transient from non-transient events.

3) *LoRa module*: A dipole antenna (as shown in Fig. 1(b)) was placed on the Lora module of the parking node to transmit/receive signals. Unfortunately, slight variations in the orientation of nodes resulted in different baseline RSSI values. Even when we used an omni-directional patch antenna [18], the location and orientation of the antenna on LoRa senders (i.e., LoRa gateways) tended to alter the transmission of signals between senders and parking nodes, which had a subsequent impact on the RSSI received by the parking nodes. Fig. 4(a) shows RSSI values received by Sender #1 and Sender #3, which were respectively deployed 15 or 40

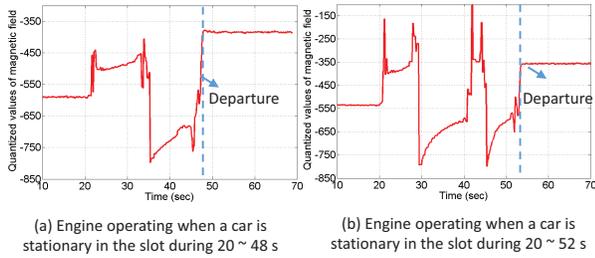


Fig. 5: Readings collected from a magnetic sensor along the z axis: Two engine-operating scenarios with cars stationary in a slot.

meters from the parking slot. The difference in RSSI values received by Sender #1 between occupied and empty states was approximately 12 dBm. However, deploying the gateway at a greater distance reduced the difference in signals, as indicated by the two dotted lines at the bottom of Fig. 4(a). In all such situations, the RSSI values was scattered across a wider range. In this situation, differentiating between parking and non-parking events could only be achieved when the baseline RSSI value (i.e., the average RSSI value received by a node in an empty lot) exceeded -80 dBm.

To elucidate the impact of interference caused by deployment factors, we reviewed the distribution of long-term data. In Section IV-D, we also propose solutions to reduce the impact of these effects.

C. Interference Caused by Target-vehicle Factors

Variations in the components used in the manufacture of vehicles can have a profound impact on readings from magnetic sensors and LoRa wireless modules.

1) *Magnetic sensor*: Changes in the magnetic field caused by the arrival and departure of cars varies with the design of the vehicle. Fig. 5(a) and (b) present data along the z axis collected for two Lexus CT 200H while idling, respectively. The magnetic field generated by the electric motors caused fluctuations in the readings along each of the axes when the vehicles were switched off. We also observed a jump in the readings when the cars left the parking slot. Nonetheless, the readings revealed a number of similarities indicating the need to collect additional data from a variety of vehicles in order to identify abnormal fluctuations caused by electrical components.

2) *LoRa module*: The bodies of cars with a low-slung design are closer to the parking node, which means that they are more effective in blocking the signals. Furthermore, vehicles with a greater proportion of ferrous materials block signals far more effectively than do vehicles with a greater content of other materials, such as aluminum. Cars inducing only a small drop in the RSSI value can confuse the system, thereby necessitating a sensor selection algorithm to avoid unreliable sensing readings.

D. Data Accuracy over the Long-term

Time

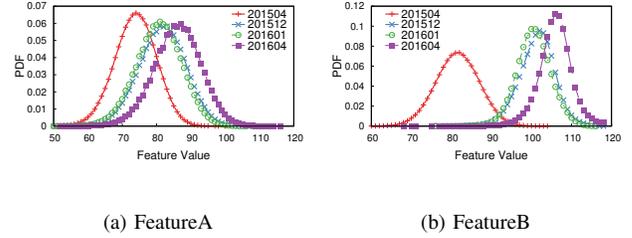


Fig. 6: Distribution of two features collected in different months.

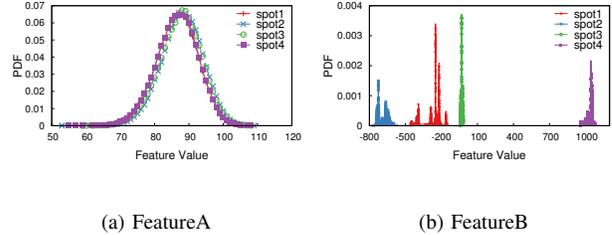


Fig. 7: Distribution of two features collected from different parking slots.

Fig. 6 illustrates the distribution of two features collected in different months. We can see similarities between the values of Feature A at 2015.12 and 2016.01 and considerable fluctuations in the other two months. The fluctuations in Feature B are less pronounced.

Space

Fig. 7 presents the distribution of two features in the same month but different parking slots, in which similarities were observed in some feature distributions and differences were observed in others. Any difference in data distribution can confuse classifiers and thereby undermine prediction accuracy. In this situation, increasing the size of training data set would not necessarily improve classification accuracy.

In the data collected over a period of 13 months, we observed that a few of the features presented consistent distributions across space and time. Researchers [31] have shown that any difference in data distribution can confuse classifiers and thereby undermine prediction accuracy. In this situation, increasing the size of training data set would not necessarily improve classification accuracy. This conclusion is further validated in Section VI.

Instead of simply measuring changes in magnitude, the system must learn to discern changes in the patterns associated with actual transient events. The results obtained in the pilot phase of data collection were used to guide the subsequent design of the system, as outlined in the following.

V. METHODOLOGY

We propose three schemes to handle uncertainty in the collection of data: model selection, sensor selection, and failure detection. *Model selection* refers to the identification

of the most appropriate model (i.e., with feature distribution similar to the testing data). *Failure detection* is used to monitor changes in the distribution of incoming data in order to respond in a timely manner. In the event that a failure is detected, the system will select another suitable trained model to keep providing prediction with high accuracy. Otherwise, the selected model outputs classification results as well as a corresponding confidence interval. During *sensor selection*, the confidence value is taken into consideration to facilitate optimization of the system for classification accuracy and minimal power consumption through the switching on/off of sensors.

A. Model Selection

Despite advances in machine learning technology, there remain considerable challenges in the application of these tools to real-world systems. For example, large shifts in feature distribution can cause even a well-trained model to fail, regardless of the machine learning technology that is employed. In this section, we present a model selection scheme to address this problem.

The model selection problem is defined as follows: Given a set of trained models or a set of training data with a number of features, it is necessary to select the model or subset of features best suited to the classification of incoming testing data.

We begin with the collection of sensor data from a variety of parking lots at various times (with ground truth parking events labeled), for use in training a model for the newly deployed system. The fact that data collected from different spots at different times may differ from the data obtained at new parking lots makes it difficult to select a subset of data and features capable of maximizing detection accuracy.

The class of testing data is unknown at the time of prediction, which means that we need unsupervised methods to detect drifts in the predictions; however, it is not always clear what type of distribution these features obey. We adopted three tests to reveal shifts in feature distribution: the Kruskal-Wallis test [11], the Kolmogorov-Smirnov test [10], and Hellinger distance [8].

Kruskal-Wallis (KW) test : KW one-way analysis of variance by rank is a non-parametric method for testing the equality of population medians among groups. It is not based on any assumptions regarding similarities in population variables among the groups being compared. This test calculates the following statistic:

$$KW = (N - 1) \frac{\sum_{i=1}^g n_i (\bar{r}_i - \bar{r})^2}{\sum_{i=1}^g \sum_{j=1}^{n_i} (r_{ij} - \bar{r})^2} \quad (1)$$

where N is the total number of samples across classes, n_i is the number of samples of class i , r_{ij} is the rank of sample i in class j , \bar{r}_i is the average rank of samples in class i , and \bar{r} is the average of all ranks. Calculating the KW distance of feature distributions from the trained model and the testing data makes it possible to select a trained model with minimal KW distance.

Kolmogorov-Smirnov (KS) test : The KS test determines whether there is divergence between two underlying one-dimensional probability distributions. As with the KW test, KS makes no assumptions with regard to the distribution of data. The two-sample KS test calculates the following statistic:

$$KS = \max |S_{T1} - S_{T2}| \quad (2)$$

where S_T is the CDF of the samples during period T . A smaller KS means that feature distributions from the trained model and the testing data are more similar

Hellinger distance : Hellinger distance, also referred to as Bhattacharyya Distance, is a measure of distributional divergence. In [27], it was concluded that Hellinger distance offers a good compromise between linearity and resolution for linear ordination. The Hellinger distance has been used for clustering and ordination in species abundance data. The Hellinger distance calculates the following statistic:

$$H = \frac{1}{\sqrt{2}} \|\sqrt{P_{T1}} - \sqrt{P_{T2}}\|_2 \quad (3)$$

where P_T is the PDF of the samples during period T , and $\|\cdot\|_2$ is used to calculate the two-norm.

Score of features: Based on statistical tests used to quantify the amount of drift in the distributions, we propose six scoring schemes to enable an explicit estimate of the goodness-of-fit between the trained models/features and the testing data. Scoring is based on the principle that a model/feature is preferred if it i) has feature distributions similar to the testing data, and ii) transient and non-transient events have distributions sufficiently distinct that machine learning tools can guarantee a small degree of variance in classification error [25]. Therefore, we define the following scores as follows:

$$Score_KW_{stable}^f = \frac{1}{KW(P_{T1}^f, P_{T2}^f)} \quad (4)$$

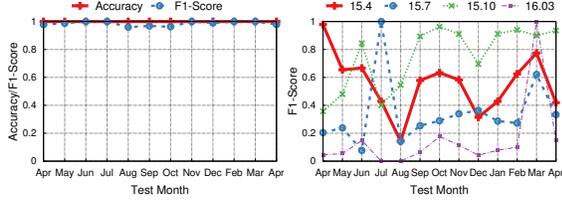
$$Score_KW_{comb}^f = \frac{KW(P_{T1|+}^f, P_{T1|-}^f) + KW(P_{T2|+}^f, P_{T2|-}^f)}{KW(P_{T1}^f, P_{T2}^f)} \quad (5)$$

where $Score_KW_{stable}^f$ is a score defined using the KW test, indicating that feature f presents similar distribution across the training data and testing data. $Score_KW_{comb}^f$ further evaluates feature f by checking the clarity of the distribution of transient events (i.e., $P_{T1|+}^f$) and non-transient events (i.e., $P_{T1|-}^f$).

Similarly, we define the following scores based on KS test and Hellinger distance:

$$Score_KS_{stable}^f = \frac{1}{KS(S_{T1}^f, S_{T2}^f)} \quad (6)$$

$$Score_KS_{comb}^f = \frac{KS(S_{T1|+}^f, S_{T1|-}^f) + KS(S_{T2|+}^f, S_{T2|-}^f)}{KS(S_{T1}^f, S_{T2}^f)} \quad (7)$$



(a) Cross-validation in one month (b) Cross-validation in one month and tests in all other months

Fig. 8: (a) Application of ten-fold cross validation illustrating the accuracy and F1-score when training and testing data were obtained in the same month; (b) Cross-validation in one month and testing in other months.

$$Score_H_{stable}^f = \frac{1}{H(P_{T1}^f, P_{T2}^f)} \quad (8)$$

$$Score_H_{comb}^f = \frac{H(P_{T1|+}^f, P_{T1|-}^f) + H(P_{T2|+}^f, P_{T2|-}^f)}{H(P_{T1}^f, P_{T2}^f)} \quad (9)$$

$Score_{stable}^f$ indicates that feature f is stable across two periods of time $T1$ and $T2$.

B. Failure Detection

We also need a method by which to determine whether feature distributions have changed sufficiently to necessitate the selection of another trained model.

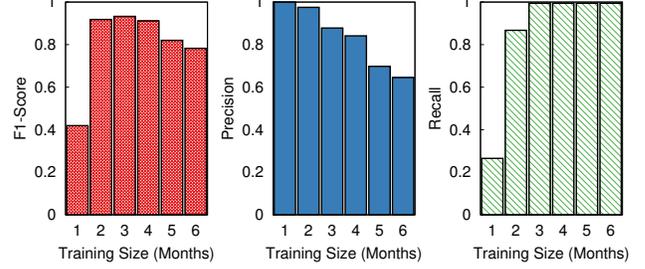
This can be achieved by periodically conducting the KW, KS, or Hellinger tests and checking whether the corresponding null hypothesis is rejected. The fact that the sensors perform sampling at 4Hz or higher means that it is easy obtain a sufficient number of samples to perform the tests. We determined that a given feature distribution could be expected to remain unchanged for weeks at a time and that only 18 hours are required to collect enough data to conduct statistically meaningful tests.

C. Sensor Selection

A large number of sensors can facilitate the detection of transient events; however, the number of sensors must be minimized to save power, extend the lifespan of sensors, and reduce management costs.

The proposed sensor selection scheme seeks an equable trade-off between higher accuracy and lower power consumption. Sensors are turned on and off according to observed environmental factors with the aim of achieving the following objective:

$$\text{Objective: } \min_I \sum_{i=1}^s I(i) \cdot [Pow(i) - C \cdot Prob(i|F)] \quad (10)$$



(a) F-1 Score (b) Precision (c) Recall

Fig. 9: Classification performance using training sets of various size. We increased the amount of training data to include data obtained from 2015.04 to 2015.09 and conducted tests using data from 2016.03.

where $I(i)$ is an indicator that sensor i should be turned on, $Pow(i)$ is the power consumption of sensor i obtained from Table I, $Prob(i|F)$ is a conditional probability representing the confidence that sensor i will work well under current environment factor F .

VI. EVALUATION

In this section, we analyze the data collected over a period of 13 months and evaluate the proposed schemes for model selection, failure detection, and sensor selection. Note that some tasks resulted in highly skewed data (i.e., many negative instances are created when a parking space is not occupied, but only a few positive instances were created when a car was moving into or out of the parking space). In these cases, the accuracy is not a representative indicator of detection performance. Thus, we conducted a more thorough analysis of the proposed system using accuracy, precision, recall, and F1-score, where $accuracy = \frac{TP+TN}{P+N}$, $precision = \frac{TP}{TP+FP}$, $recall = \frac{TP}{TP+FN}$, and $F1\text{-score}$ is the harmonic mean of them (i.e., $2 \cdot \frac{precision \cdot recall}{precision + recall}$), based on the number of positive samples P , negative samples N , true (false) positive inference TP (FN), and true (false) negative inferences TN (FN).

A. Model Selection

We applied ten-fold cross validation to data obtained in a given month in order to show that when training data and testing data are obtained at close to the same time, their feature distributions are similar and transient events can be detected accurately, as shown in Fig. 8(a). Our results are comparable to those reported for state-of-the-art commercial products [5], [2], with accuracy close to 1 and F1-scores ranging from 0.96 to 1. However, when training data and testing data were obtained at different times (e.g., separated by one or more months), the distributions were often significantly different, which greatly reduced the accuracy of the trained models.

As shown in Fig. 8(b), we selected data obtained in four months (2015.04, 2015.07, 2015.10, and 2016.03). Half of the samples from each month were used as training data. We

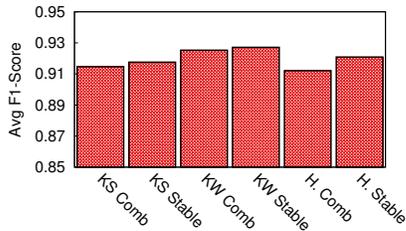


Fig. 10: Average F1-scores obtained using the model formulated using data collected in 2015.04 to predict transient events in the traces in 2015.05, 2015.06, and 2015.07. In this set of experiments, the characterization of the model is based on features with scores ranked in the top 40%.

then used four models to detect transient events in the other months. In cases where the testing month was the same as the training month, we used the other half of the samples for testing. Clearly, any temporal gap between training and testing undermined detection performance.

One approach to solving this problem would be to use a larger dataset. Thus, we expanded the training data from 1 month to 6 months (from 2015.04 to 2015.09), and tested the corresponding models using data from 2016.03, the results of which are presented in Fig. 9. This shows that the larger data set does not necessary yield a better F1-score because a change in the distribution of transient events in one month may be matched by the distribution of non-transient events in other months. Thus, a large dataset can actually reduce precision by introducing a larger number of false positive (i.e., treating non-transient events as transient events.) Moreover, increasing the amount of training data also increases the time required for data acquisition and model training.

As described in Sec. V-A, We addressed these problems by developing six scoring schemes to identify the models and/or features best suited to the current testing data. Fig. 10 plots the average F1-score obtained using the model formulated using data collected in 2015.04 to predict transient events in the traces in 2015.05, 2015.06, and 2015.07. In this set of experiments, we calculated the six scores for each feature related to the magnetic sensor, light sensor, and LoRa module. In each scoring scheme, the characterization of the model is based on top-40% features (features with scores ranked in the top 40%). The high F1-scores imply that the six scores can efficiently identify the drifts in distributions. As shown in Fig. V-A, all of the scoring methods performed similarly, which implies that KS, KW, and Hellinger distance are all effective methods by which to address drift in distribution.

Comparing with Fig. 8(b), Fig. 11 further demonstrates the effectiveness of the proposed scoring schemes in which features with scores ranked in the top 40% under the $Score_{KW_{stable}}$ scoring scheme were selected for classification. As shown in Fig. 11, regardless of the training and testing data, most of the F1-Scores were than 90%. Among them, the F1-Scores obtained under the model formulated using data collected in 2015.08 or 2015.12 to predict transient

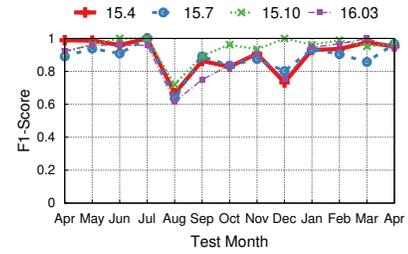
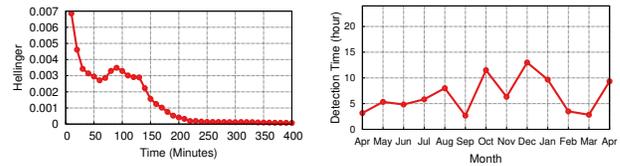


Fig. 11: Training based on data from one month with testing using data from other months. Unlike in Fig. 8, features with scores ranked in the top 40% under the $Score_{KW_{stable}}$ scoring scheme were selected for classification .



(a) Hellinger distance under various amount of data (b) Time to converge

Fig. 12: Hellinger distance decreased with an increase in the amount of data used. With sufficient data, the Hellinger distance converges, implying that the incoming data possesses the same distribution.

events in the traces in other months were lower than 90%. This is due to features with significant drift in the distributions with scores ranked in the top 40% under the KWstable scoring scheme.

B. Failure Detection

Fig. 12(a) calculates the Hellinger distance between the feature distribution obtained from a period of data for the entire month of April. The period, from the starting time of April to a specified end time, of used data is increasing from 0 to 400 minutes in April 2015. The Hellinger distance converges at 220 min, which implies that it takes 220 min to determine whether incoming data presents the same distribution as data from April 2015.

Fig. 12(a) shows the time required for the Hellinger distance to converge (i.e., $Hellinger \leq 0.001$) in various months. Based on the data collected in this study, it appears that feature distribution can remain unchanged for several weeks. Fig. 12(b) indicates that it would take 3 ~ 13 hours to determine whether incoming data has the same distribution. Once a change in distribution has been detected, the proposed model selection method can be used to identify a suitable model for the classification of incoming data.

C. Sensor Selection

Fig. 13 shows the confidence in the accuracy of classification events is (true positive and true negative), false

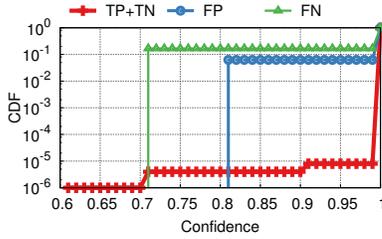


Fig. 13: CDF of confidence for correctly classified events (true positive and true negative, i.e., the red line), false positive events (blue line), and false negative events (green line).

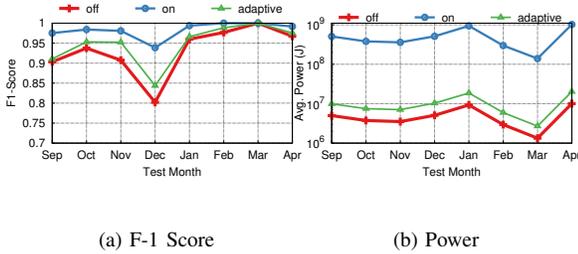


Fig. 14: Detection performance and power consumption with sensors switched on (red line), all sensors except magnetic sensor switched off (blue line), and the proposed adaptive sensor selection scheme (green line).

positive events, and false negative events. As shown, more than 9% of incorrectly classified events (i.e., FN and FP) achieved confidence values equal to or lower than 0.81. These confidence values make it possible to predict the probability that a prediction is correct. When confidence is lower than 0.81, additional sensors can be turned on.

Fig. 14 shows the accuracy and power consumption when all of the sensors are turned on, when all of the sensors except the magnetic sensor are off, and the results obtained when using the proposed adaptive sensor selection scheme. When magnetic sensors are used, the average F1-score is 0.93 and the power consumption is 5,114KJ per month. When all of the sensors are on, the F1-score increases to 0.98 and power consumption increases by 87 times to 447,790KJ per month. The adaptive sensor selection scheme is able to improve the F1-score to that achieved using only magnetic sensors, while maintaining power consumption at only 2% of the power consumed when all of the sensors are turned on.

VII. DISCUSSION

In this study, we sought only to differentiate between transient and non-transient events; i.e., cars entering parking spaces cannot be differentiated from cars exiting spaces. One potential solution would be to use duplicate magnetic sensors and exploit the spatial correlation between them in order to determine the direction in which cars are moving; i.e., arriving or departing. We established a prototype parking node with two magnetic sensors (i.e., magnetic sensors #1 and #2) on both sides of the board, separated by 95.5 millimeters. By

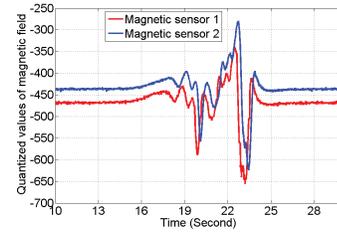


Fig. 15: Readings along the x axis collected from two magnetic sensors located on two sides of the board.

collecting readings through a serial link, we were able to sample magnetic fields using the two sensors at a sampling rate of 100 Hz. Fig. 15 shows the magnetic field projected on the x axis of two magnetic sensors. Magnetic sensor #1 is closer to the incoming vehicle; therefore, the magnetic field sensed by magnetic sensor #1 is affected by vehicles first. Locating peaks and/or dips in the magnetic fields makes it possible to calculate the delay between two adjacent peaks or dips in order to infer the direction in which vehicles are moving. We will continue conducting efforts to incorporate this design in our final system.

VIII. CONCLUSION

In this study, we collected data from a multi-slot parking area over a period of 13 months to elucidate the challenges and limitations of three sensing modules used to determine the occupancy of parking spaces. A review of interference patterns and long-term trends in the data led to the development of three adaptive machine learning schemes to deal with data of uncertain accuracy: model selection, failure detection and sensor selection. A series of experiments validated the accuracy of the adaptive schemes in the detection of transient events. The proposed system is able to improve accuracy by identifying the most appropriate model and features for training, detecting changes in the distribution of incoming data, and identifying the most reliable sensing modality. Our findings highlight the need for an adaptive machine learning scheme in the design of parking occupancy detection systems. In future work, we will continue the data collection process using other sensing modalities. We will also continue to improve the accuracy of the proposed system and perform field studies with actual prototypes in additional parking sites.

IX. ACKNOWLEDGEMENTS

We would like to thank Seng-Yong Lau for his contributions to the hardware design and prototype of the parking nodes, and Min-Jhih Huang and You-Zhi-Jia Hu for their assistance in the processing of the collected data. We are grateful to the Computer and Information Networking Center, National Taiwan University for the support of high-performance computing facilities. This work was supported in part by the Ministry of Science and Technology of Taiwan under Grants MOST 105-2633-E-002-001 and 105-2221-E-002-172, and NSFC No. 61572324.

REFERENCES

- [1] Accelerating sustainability: Demonstrating the benefits of transportation technology. <http://digitalenergysolutions.org/dotAsset/933052fc-0c81-43cf-a061-6f76a44459d6.pdf>.
- [2] ADEC Technologies. <http://www.adec-technologies.ch>.
- [3] AN218: Vehicle Detection Using AMR Sensor. http://www51.honeywell.com/aero/common/documents/myaerospacecatalog-documents/Defense_Brochures-documents/Magnetic_Literature_Application_notes-documents/AN218_Vehicle_Detection_Using_AMR_Sensors.pdf.
- [4] ARCT TM1276 LoRa Transceiver module. <http://www.arct.com.tw/web/>.
- [5] Fastprk. <http://www.fastprk.com/>.
- [6] Fastprk — Sigfox Partner Network. <https://partners.sigfox.com/products/fastprk>.
- [7] Fastprk — WorldSensing. <http://www.worldsensing.com/solutions/mobility/technology/fastprk-en.html>.
- [8] Hellinger distance. https://www.encyclopediaofmath.org/index.php/Hellinger_distance.
- [9] Honeywell HMC5883 3-Axis Digital Compass IC. <https://www.seeedstudio.com/wiki/images/4/42/HMC5883.pdf>.
- [10] Kolmogorov-Smirnov test. https://www.encyclopediaofmath.org/index.php/Kolmogorov%E2%80%93Smirnov_test.
- [11] Kruskal-Wallis H Test using SPSS Statistics. <https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php>.
- [12] The learning curve of smart parking. <http://www.nytimes.com/2012/12/23/technology/smart-parking-has-a-learning-curve-too.html>.
- [13] Periodic Reporting for period 1 - FASTPRK-2. http://cordis.europa.eu/result/rcn/186387_en.html.
- [14] SFpark project. <http://sfpark.org>.
- [15] Smart Parking Should Be High on List of Priorities for City Managers. <https://machinaresearch.com/report/smart-parking-should-be-high-on-list-of-priorities-for-city-managers/>.
- [16] Streetline and Cisco Introduce Camera-Based Smart Parking Sensing and Integrated Network Solution. <http://www.streetline.com/2014/05/streetline-cisco-introduce-camera-based-smart-parking-sensing-integrated-network-solution-2/>.
- [17] Streetline Inc. <http://www.streetline.com>.
- [18] Taoglas 915MHz Embedded Ceramic Patch Antenna. <https://www.taoglas.com/wp-content/uploads/2015/06/ISPC.91A.09.0092E.pdf>.
- [19] TAOS TSL2561 Light-to-Digital Converter. <https://cdn-shop.adafruit.com/datasheets/TSL2561.pdf>.
- [20] The project - FastPrk2. <http://www.fastprk2.eu/the-project/>.
- [21] TI MSP430 ultra-low-power Microcontrollers. http://www.ti.com/lscs/ti/microcontrollers_16-bit_32-bit/msp/overview.page.
- [22] Urbiotica. <http://www.urbiotica.com/en/>.
- [23] WorldSensing starts Fastprk-2 project. <http://www.fastprk.com/media-news/fp-news.html?wsdetail=%20284>.
- [24] K. Blumer, H. R. Halaseh, M. U. Ahsan, et al. Cost-effective single-camera multi-car parking monitoring and vacancy detection towards real-world parking statistics and real-time reporting. In *Proc. ICONIP 2012 - Volume Part V*, pages 506–515. Springer-Verlag, 2012.
- [25] D. Brain and G. I. Webb. On the effect of data set size on bias and variance in classification learning. In *Proc. Australian Knowledge Acquisition Workshop (AKAW-99)*, pages 117–128. The University of New South Wales, 1999.
- [26] M. Y. Idris, Y. Y. Leng, E. M. Tamil, et al. Car Park System: A Review of Smart Parking System and its Technology. *Information Technology Journal*, 8(2):101–113, 2009.
- [27] P. Legendre and E. D. Gallagher. Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129(2):271–280, 2001.
- [28] S. Mathur, T. Jin, N. Kasturirangan, et al. Parknet: Drive-by sensing of road-side parking statistics. In *Proc. MobiSys '10*, pages 123–136, New York, NY, USA, 2010. ACM.
- [29] S. Nawaz, C. Efstratiou, and C. Mascolo. Parksense: A smartphone based sensing system for on-street parking. In *Proc. MobiCom '13*, pages 75–86, New York, NY, USA, 2013. ACM.
- [30] San Francisco Municipal Transportation Agency. SFpark: Putting Theory Into Practice - Post-launch Implementation Summary and Lessons Learned. Technical report, August 2011.
- [31] H. Wang, W. Fan, P. S. Yu, et al. Mining concept-drifting data streams using ensemble classifiers. In *Proc. KDD '03*, pages 226–235, New York, NY, USA, 2003. ACM.