ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 4, Issue 4, July 2015

State Recognition by Analyzing Time Series Data of Sensors

[1]Ganwen Jiang, [2]Hiroya Kamitomo

Abstract*— Wireless Sensor Network (WSN) play an important role on many areas of modern day living, such as the ability to measure and understanding environmental indicators. It has attracted attention for automatic state recognition by using many sensors. In future, big data analysis will become more and more important for state recognition and other field of computer science. In our study, many sensors such as human sensor and opening/closing sensor are used to count the number of people who exists in the office. Machine Learning (ML) is used to learn the pattern of people's flow, and predict the number of people by using learning result. In the study, a unique** **IoT Management technology** ***developed by our company is used to get data of each sensor. In addition, sensor's data is learned by Locally Weighted Regression (LWR) method. In this paper, Multivariate time series data of sensor is used to predict human count in the office. This paper presents an automatic state recognition system to predict human count by using Machine Learning.

***Index Terms*—Machine Learning, Time Series Data, State Recognition, Sensors.**

I. INTRODUCTION

The Internet has undergone severe changes since its first launch in the late 1960s as an outcome of the ARPANET with number of users about 20% of the world population. "7 trillion wireless devices serving 7 billion people in 2017". This vision reflects the increasing trend of introducing micro devices and tools in future i.e. IoT. In such ambient environment not only user become ubiquitous but also devices and their context become transparent and ubiquitous. With the miniaturization of devices, increase of computational power, and reduction of energy consumption, this trend will continue towards IoT [1]. At present, the Internet of Things (IoT) becomes more and more popular in the world. In the IoT paradigm, many of the objects that surround us will be on the network in one form or another. Radio Frequency Identification (RFID) and sensor network technologies will rise to meet this new challenge, in which information and communication systems are invisibly embedded in the environment around us. These results in the generation of enormous amounts of data which have to be stored processed and presented in a seamless [2]. Sensing data is direct reflection of the state. The data represents the state of human or things, such as "human exist or not exist", "temperature is high or low", and "door is opened or closed", and so on. However, how to use sensing data to do prediction or state recognition is a critical issue. Sensing data is "0" or "1", and it is not continuously, so it is difficult to use sensing data to predict the state. As to this problem, we propose a new viewpoint to predict state of human or things by using the Machine Learning (ML) method and wireless sensor network. Wireless sensor networks have become increasingly common in everyday applications due to decreasing technology costs and improved product performance. An ideal application for wireless sensor networks is a biomedical patient monitoring tool. Wireless patient monitoring systems improve quality of life for the subject by granting them more freedom to continue their daily routine, which would not be feasible if wired monitoring equipment were, used [3].

In our study, we develop a new IoT Management technology to access the sensor data conveniently, and use the Locally Weighted Regression (LWR) of Machine Learning (ML) and wireless sensor network to predict how many people in the office. In general, we can use cameras to count how many people in the room by image processing technology, but some people do not like to make his or her face (Personal information) public. Therefore, wireless sensor network system is better than camera system.

II. PROPOSED SYSTEM

A. Unique IoT Management Technology

In order to achieve easily to save and fetch sensor's data, the unique IoT Management technology had been developed by us shown in Fig. 1.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 4, Issue 4, July 2015

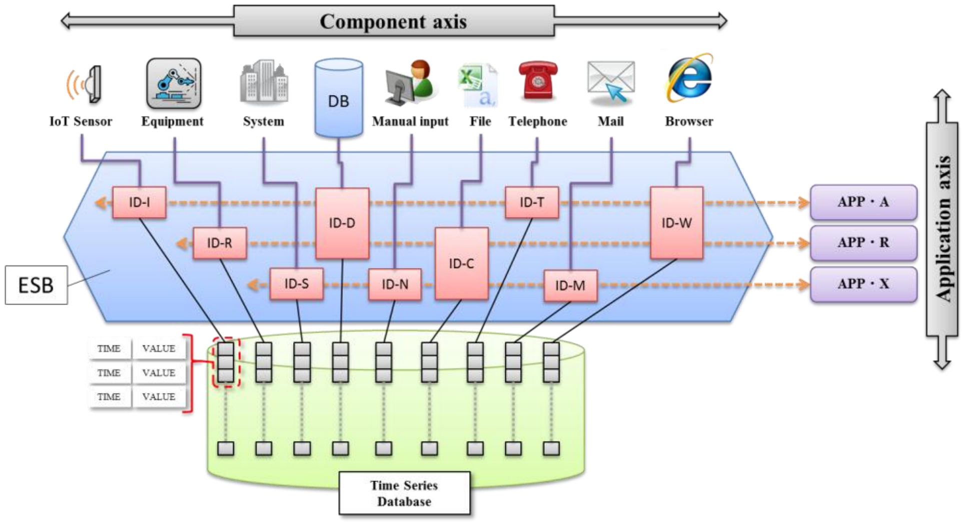


Fig. 1 Conceptual diagram of Enterprise Service Bus System

This technology is for IoT Management, and enormous amounts of data can be accessed by inherent ID easily. On component axis, each component is managed by inherent ID. Component is connected to application by using software bus (ESB: Enterprise Service Bus). The ESB is mechanism for circulating the data with a unique ID. It is possible to access the data by inherent ID for applications. Furthermore, the data is saved with inherent ID by using the time series database. Therefore, it can be used for flexible development in many fields.

B. Proposed Method

In our study, we use wireless sensor network and Machine Learning (ML) to predict how many people in the office. About wireless sensor network, we use three human sensors placed in different rooms, and two opening/closing sensors attached at different doors. We want to know the number of human by analyzing the people's flow (or people's density). Human sensor shows human exist or not exist, so the value is "0" or "1". The opening/closing sensor shows door is opened or closed, so the value is also "0" or "1". It is difficult to know the number of human by using this sensing data with "0"or "1". We choose the change of each sensor's data in a period of time as the feather to analyze people's flow (or people density). The flowchart of data processing is shown in Fig. 2.

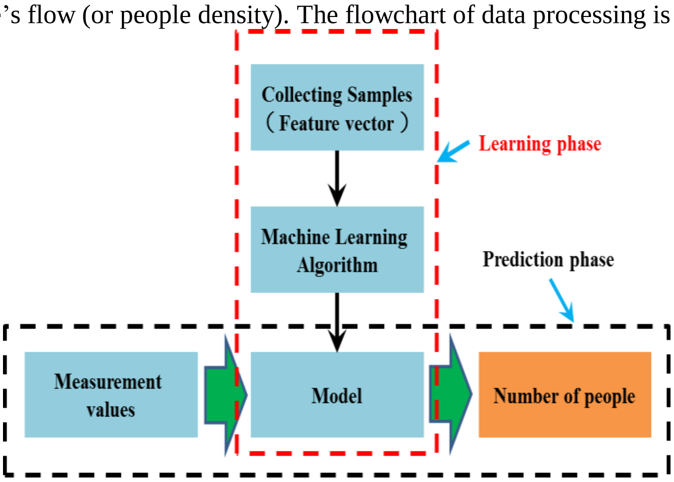


Fig. 2 Flowchart of data processing

In Fig. 2, there is two parts for data processing. One is learning phase, and another is prediction phase. In learning phase, the Locally Weighted Regression (LWR) algorithm is used. In addition, JSON data of sensors as the collecting samples are obtained by using our IoT Management technology, and it is shown in Fig. 3.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 4, Issue 4, July 2015

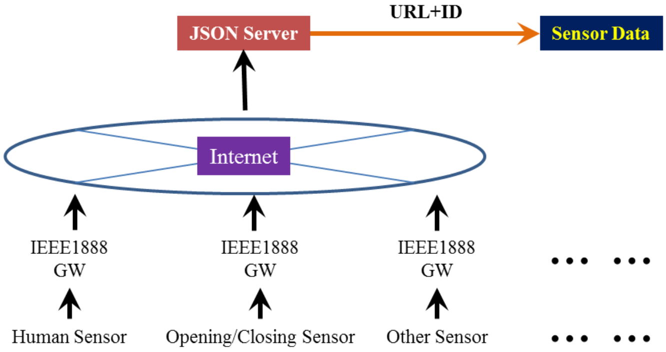


Fig. 3 Collection of Sensor Data

In Fig. 3, it shows JSON data of each sensor can be obtained by using "URL+ID". In addition, we use the Gateway technical developed by ourselves to change the protocol to IEEE1888. Furthermore, JSON data of "Key-Value" type can be got from JSON server by using inherent ID of each sensor. Because the learning data is dispersion, prediction result accuracy of the Linear Regression (LR) method would become lower. One of the problems with linear regression is that it tries to fit a constant line to your data once the model was created. Such behavior might be okay when your data follows linear pattern and does not have much noise. However, when the data set is not linear, linear regression tends to under fit the training data [4]. Therefore, we use LWR method to train samples and predict the number of people.

LWR method tries to alleviate this problem by assigning weights to your training data. Weights are bigger for the data points that are closer to the data you are trying to predict, thus local in the name. Another thing to note, LWR requires the entire data set every time you try to make a prediction making it much more computationally expensive compared to the simple linear regression. We have to do this because every time we try to make a prediction we are constructing a regression line that's local to the data point of our interest. LWR model is very similar to simple regression model, the only difference is that we are introducing a weight matrix W [5]. Once we have the weight matrix we can find the model parameters as follows:

$$\theta = (X^T WX)^{-1} X^T WY \tag{1}$$

To get the prediction we need to multiply betas with our inputs X_0.

$$h(x) = \theta^T X_0 \tag{2}$$

Kernel Smoothing [6] allows us to smooth out our point of interest (X_0) using nearby data points (X). The kernel is usually defined by a function D whose values are decreasing as the distance between X_0 and X increases:

$$D = \frac{\| X - X_0 \|}{h_\lambda (X_0)} \tag{3}$$

To construct matrix W we need to evaluate kernel for each training input (X) and the value we are trying to predict, X_0. The weight matrix is diagonal where all non-diagonal cells are 0. After running the kernel on our data set we will essentially create a weight matrix where the weights are decreasing as the distance between X_0 and X increases. This kernel property will make nearby data points more important when solving linear regression. I am going to use the Gaussian kernel [7] function for illustration purposes:

$$D = ae^{-\frac{\|X - X_0\|}{2c}} \tag{4}$$

Therefore, we calculate matrix W as follows:

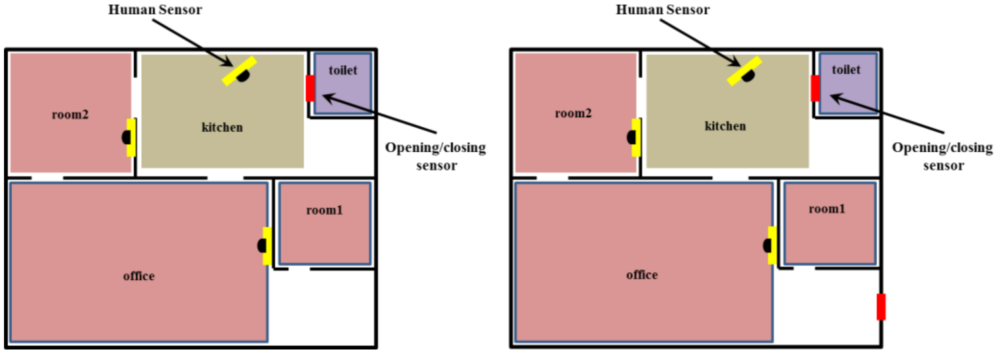
$$W = e^{-\frac{\|X - X_0\|}{2k^2}} \tag{5}$$

When matrix W is calculated, matrix θ can be calculated by equation (1). In addition, prediction result can be calculated by equation (2).

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 4, Issue 4, July 2015

III. EXPERIMENT ENVIRONMENT

In our study, we use multiple sensors to measure people's flow as the feature, and then learn this feature to automatically predict the number of people. As said before, because data is nonlinear, the LWR method of Machine Learning is used to carry out this study. There are two experiments with different number of sensors. In first experiment, we use three human sensors and one opening/closing sensor. In second experiment, we use three human sensors and two opening/closing sensors. The placement of sensors in experiment one and experiment two is shown in Fig .4(a) and Fig. 4(b).



(a) (b)

Fig. 4 Placement of sensors in experiment 1 and experiment 2 respectively

Human sensor is to make sure whether there are people in room or office or kitchen. If there is a person, value of human sensor will be changed from "0" to "1". And then, it will enter the waiting period in 6 minutes. In the waiting period, human sensor is with the same value "1". Then, if there is no person, the value of human sensor will become "0". Human sensor only can make sure human existence, but it cannot know the number of people. About opening/closing sensor, if the door where opening/closing sensor attached is opened or closing, sensor will make a response. The value of opening/closing sensor is also "0" or "1". In experiment 1, one opening/closing sensor is attached at the door of toilet. It can make sure how many times the door opened within a time period. Furthermore, we can use this data to predict how many people exist. In experiment 2, another opening/closing sensor is attached at the door of entrance. Therefore, we not only know how many times the door of toilet opened but also how many times the door of entrance opened. Furthermore, the flow of people can be estimated. And then, use this data to predict the number of people by the LWR method.

Human sensors and opening/closing sensors used are made by the EnOcean Company. They are shown in Fig. 5.

Fig. 5 Sensors made by the EnOcean Company

In Fig. 5, the human sensor is on the left, and the opening/closing sensor is on the right. The specification of each sensor is shown in Table. 1. The frequency of both are 928.35MHz, and data rate of both are 125kbps. The waiting period of human sensor is 6 minutes, and measurement distance is 4.5 meters.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 4, Issue 4, July 2015

Table. 1 Specification of each sensor

	Human Sensor	Opening/Closing Sensor
Measurement distance	4.5m	
waiting period	6min	Immediately
Weight	125g	28g
Transmission content	Digital signal	Digital signal
Frequency	928.35MHz	928.35MHz
Data rate	125kbps	125kbps

IV. EXPERIMENT RESULTS

In experiment 1 and experiment 2, we use reaction times in one hour of each sensor as the feature vector. Then, answer matrix is obtained by automatic collection system. In experiment 1, the number of learning data is 22899, and the number of learning data is 14710 in experiment 2. About this system, there is a button puts on the entrance. If people enter or come out, he has to push it to tell system he will enter or leave. In addition, the data will be transferred to server by using socket technology. When samples matrix and answer matrix are got, we will input them to our machine learning system. Then, we will predict the human count by the model calculated from machine learning. The result of difference between predicted human count and true value of human count is shown in Fig. 6. The horizontal axis represents time in second, and vertical axis represents human count. Green line shows true value of human count, and red line shows predicted human count.

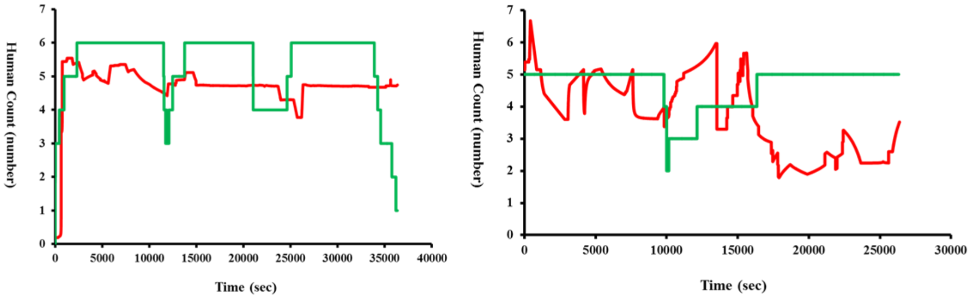


Fig. 6 Predicted human count by LWR method

In addition, average human count per hour is calculated and shown in Fig. 7. In Fig. 7, the green line shows average human count of true value, and red line shows average human count predicted. The horizontal axis represents time in hour, and vertical axis represents average human count.

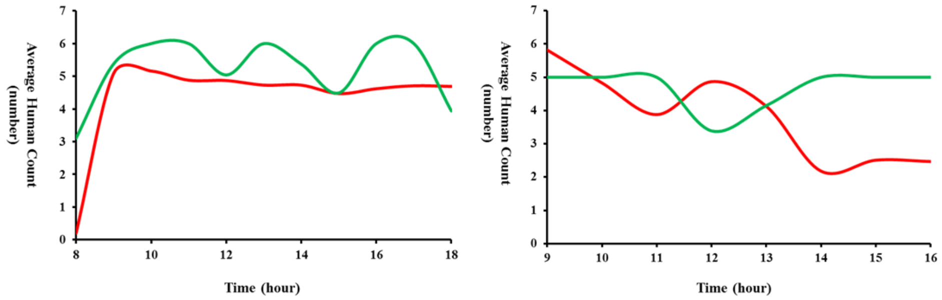


Fig. 7 Average predicted human count per hour

From Fig. 7, the similarity between predicted value and true value can be calculated. Results can be analyzed in two aspects. One is to calculate the correct rate. If the difference between predicted value and real value less than 1, we define it is correct. Then, we can calculate correct rate by the equation as follows. In equation (6), R is correct rate, CN is the number of correct value, and TN is total number of value.

$$R = \frac{CN}{TN} \quad (6)$$

Furthermore, we will calculate the Euclidean Distance between predicted value and real value, because it shows the

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 4, Issue 4, July 2015

similarity of two graphs. Equation (7) shows the method to calculate the Euclidean Distance [8]. D(a,b) is distance between graph a and graph b, a_i is a point of graph a, and b_i is a point of graph b.

$$D(a,b) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} \quad (7)$$

Results of correct rate and the Euclidean Distance are shown in Table. 2.

Table. 2 the similarity between predicted value and true value

	Correct Rate	Euclidean Distance
Experiment 1	0.54	4.07
Experiment 2	0.37	4.96

V. DISCUSSION

In our study, the purpose is to predict the number of people in the office by using Machine Learning algorithm. We want to achieve to analyze times series data of sensors accurately. Also, we want to make use of the times series data well in many fields by analyzing it. From the experiment results above, the effectiveness to predict human count by using LWR method can be verified. In experiment 1, correct rate is 54%, and the Euclidean Distance is 4.07 with real value. In experiment 2, correct rate is 37%, and the Euclidean Distance is 4.96 with the real value. From the experiment results, the correct rate is low, and the Euclidean Distance between predicted value and real value is also large. There are three peaks both in experiment 1 and experiment 2 shown in Fig. 8. The position of each peak in experiment 1 is different with it in experiment 2. In experiment 2, the shape of red graph is reverse with green graph, but there is not this case in experiment 1.

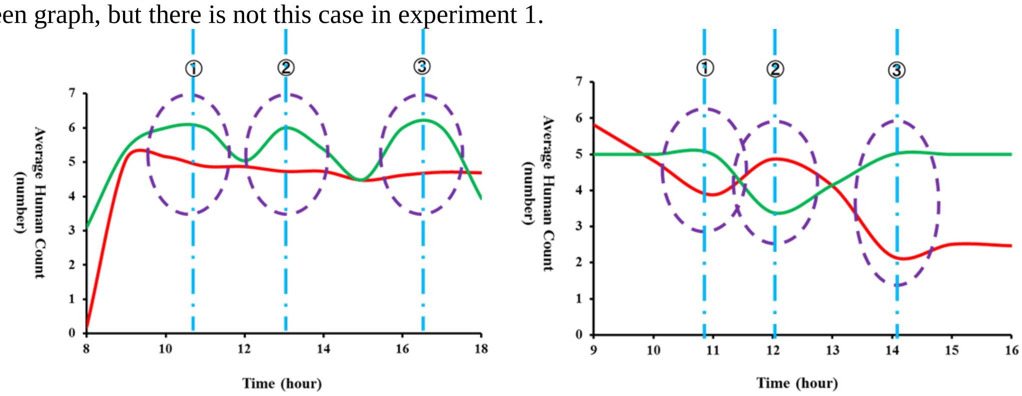


Fig. 8 Peak's change in each experiment

May be the value of opening/closing sensor at entrance affects results of experiment 2, because the opening/closing sensor of entrance is not used in the experiment 1. Entrance is a way contact indoors and outdoors, but we want to know the human count indoors, so the correlation is low between entrance and the human count indoors. Therefore, correct rate in experiment 2 is low. Also, there are other reasons. For example, in this system, people's flow is little. At 9:00, everyone comes to office, and there is scarcely movement between 9:00 and 12:00. From 12:00 to 13:00, everyone will have a lunch in outside, so there is any movement. From 13:00 to 18:00, there is scarcely movement also. If a person does not move, sensor will not react, so the value of sensor is "0" always. In addition, human count is small. There are only six people in our office. Therefore, we will carry out this experiment in bustling place, such as supermarket, train station, and so on.

Furthermore, the selection of feature vector is also very important. In our study, we choose the combination of human sensors and opening/closing sensors, they are discrete values. If there is no action, the value of sensor is "0" always. Because human sensor only knows whether a person exists in room, it is same when only one person exists and multiple persons exist. In future work, continuous value will be chosen, such as room temperature, humidity and power consumption, because these values will change continuously.

In future work, we will choose continuous value as the feature vectors, and experiments will be carried out in prosperous place, because people's flow is large. In addition, we want to try other machine learning method to do

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 4, Issue 4, July 2015

prediction. Future is the era of Big Data, it is important to make use of the big data, and machine learning will be a good approach.

VI. CONCLUSION

In this paper, about ten thousand samples are used to calculate the model, and then this model is used to predict the number of people, and the correct rate is 54% and 37% respectively. At first, it is effectiveness to use machine learning approach to do sensor's data processing. In the era of big data, machine learning is an effective method to analyze big data. In addition, sensor's data can be used to do prediction in many fields without camera, because other people cannot see your face from sensors, so it can effectively protect personal information. At present, sensor plays an important role in IoT system, and you can know the state of things fast and sensor can make control easily. Furthermore, we understand how to choose the feature vectors and it is very important to select good target for our system from the study. Finally, we will perform many researches to do big data processing by using other machine learning method, such as SVM, deep learning, k-means and so on.

REFERENCES

[1] Sachin Babar, Parikshit Mahalle, Antonietta Stango, Neeli Prasad and Ramjee Prasad, "Proposed Security Model and Threat Taxonomy for the Internet of Things (IoT)," Recent Trends in Network Security and Applications Communications in Computer and Information Science, vol. 89, pp. 420-429, July. 2010.

[2] Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic and Marimuthu Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," Future Generation Computer Systems, vol. 99, pp. 1645-1660, Sep. 1959.

[3] T. Ryan Burchfield and S. Venkatesan, "Accelerometer-based human abnormal movement detection in wireless sensor networks," HealthNet'07: Proceedings of the 1st ACM SIGMOBILE International Workshop on Systems and Networking Support for Healthcare and Assisted Living Environments, pp. 67-69, Oct. 2007.

[4] Zheng Yuan and Yuhong Yang, "Combining Linear Regression Models: When and How?," Journal of the American Statistical Association, vol. 100, pp. 1202-1214, Oct. 2007.

[5] Shiliang Su, Rui Xiao and Yuan Zhang, "Multi-scale analysis of spatially varying relationships between agricultural landscape patterns and urbanization using geographically weighted regression," Applied Geography, vol. 32, pp. 360-375, March. 2012.

[6] Abdelaati Daouia, Laurent Gardes and Stéphane Girard, "On kernel smoothing for extremal quantile regression," Bernoulli, vol. 19, pp. 2557-2589, Nov. 2013.

[7] Takamitsu Araki, Hideitsu Hino and Shotaro Akaho, "A Kernel Method to Extract Common Features Based on Mutual Information," Neural Information Processing Lecture Notes in Computer Science, vol. 8835, pp. 26-34, 2014.

[8] Lam Hong Lee, Chin Heng Wan, Rajprasad Rajkumar and Dino Isa, "An enhanced Support Vector Machine classification framework by using Euclidean distance function for text document categorization," Applied Intelligence, vol. 37, pp. 80-99, July. 2012.