一般社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report

Sensor Data Characteristics in the Wireless Train Track Unusual Notification Sensor Networks

Zhi Liu[+], Toshitaka Tsuda[+], Hiroshi Watanabe[*], Nagateru Iwasawa[#], Akio Hada[#]

[+]Global Information and Telecommunication Institute, Waseda University, Tokyo, 169-8555 Japan

[*]Department of Communications and Computer Engineering, Waseda University, Tokyo, 169-8555 Japan

[#]Signaling and Transport Information Technology Division, Railway Technical Resarch Institute, Tokyo 185-8540, Japan

Email: liuzhi@aoni.waseda.jp, tsuda-toshitaka@nifty.com, hiroshi.watanabe@waseda.jp, {iwasawa.nagateru.81, hada.akio.71}@rtri.or.jp

Abstract Unusual notification becomes an important and emergent research topic, especially in Japan, which is one of the countries most affected by natural disasters such as the earthquakes and tsunami. Wireless sensor networks (WSNs) are good options to help monitor the scene of interest (i.e. the scene we would like to monitor) and notify the unusual happening to the control center due to its efficiency, low cost, 24-hour sensing ability, etc. For WSNs, sensor data characteristics play an important role in the sensor nodes deployment, protocol design and the sensor data post-processing. Hence what characteristics the sensor data embed is an important issue. In this paper, we investigate the sensor data characteristics of the WSNs, where the WSNs are composed of Wi-Sun sensors for the train track monitoring. The sensor data are illustrated in the time domain and the frequency domain. We also calculate the correlations of the sensor data captured by different sensors in typical scenarios.

Keyword *wireless sensor networks, sensor data characteristics, unusual notification*

1. Introduction

Japan is one of the countries most affected by natural disasters, such as the earthquakes[1] and tsunami[2]. In some areas, heavy rain or other disasters may lead to the landslides[3]. These disasters may cause serious problems, e.g. they may destroy the nearby infrastructural facilities such as the train tracks. Hence how to monitor the scene of interest and detect the unusual in time to avoid unnecessary losses becomes a very important and emergent issue.

WSNs [1–3] have been used in many industrial and consumer applications such as the industrial processing monitoring and control [4], machine and building health monitoring [5] due to its efficiency, low-cost, 24-hour sensing ability, etc. The scene monitoring and event detection [6, 7] are typical applications of the WSNs and there are already systems existed to monitor the scene of interest and detect the unusual events.

For the WSNs, sensor data characteristics greatly affect the sensor nodes deployment, network protocols design and the sensor data processing. Specially, if the sensor data vary slightly across time, we can assign a simple threshold for the sensor nodes, but some adaptive threshold value are proper when the sensor data vary regularly. If sensor data generated by different sensor nodes is more or less the same, it is not necessary to tune every sensor node. But instead we could assign each sensor node with universal parameters, which could significantly reduce the deployment cost. The sensor data's variation also greatly affects the sensor data post-processing. The correlations among different sensors' captured data can decide whether it is necessary to adapt advanced sampling methods and whether partial data is sufficient for the scene monitoring. Hence what characteristics the sensor data embed is an important issue for the unusual notification WSNs. WSN differs from each other given the diverse sensor nodes types, the diverse sensor node deployments and monitoring environment, different applied protocols, etc. This brings us new challenges for the studies in the area of the WSNs since we need to investigate the sensor characteristics case by case.

[1] Japan Meteorological Agency | Earthquake Information http://www.jma.go.jp/en/quake

[2] Japan Meteorological Agency|Tsunami Information http://www.jma.go.jp/en/tsunami

[3] LANDSLIDE IN JAPAN' Homepage. http://www.tuat.ac.jp/ sabo/lj

This paper studies the sensor data characteristics of the WSNs deployed along the train tracks to monitor the 'unusual'. The sensor nodes used are the Wi-Sun sensors[4] developed by *National Institute of Information and Communications Technology* (NICT), Japan. The sensor nodes have two modes: normal and power-saving. Sensor nodes in normal mode monitor the nearby scene and broadcast the data periodically, and the data will eventually be forwarded to the control center via the sink nodes. Sensors in power-saving modes only broadcast the data when the captured scene incurs a value larger than the pre-assigned threshold value. If the sensor data captured is below the threshold, the sensor nodes in power-saving mode do not broadcast their captured data. Since normally there is no 'unusual' happening, this rare transmission mechanism greatly saves sensors' energy consumptions, where the sensor energy is critical in WSNs.

In this paper, we collect the sensor data generated by the Wi-Sun sensors for the train track monitoring and analyze the sensor data in the time domain and the frequency domain. We also compare different sensor nodes' data to investigate the correlations. Through the analysis, we find that the sensor data distribution has the shape of the normal distribution and the correlations between the sensor data captured by different sensor nodes are week. One reason of the week correlation is that the sensor data we currently have are not at the exact same time instances.

The rest of this paper is organized as follows: related work in the field of the WSNs and data distribution modeling are introduced in Section 2. Section 3 introduces the unusual notification sensor network system for the train tracks, the data distribution model and the way to calculate the sensor data correlation. We analyze the collected sensor data and show the sensor data characteristics in Section 4. Section 5 concludes this paper.

2. Related Work

This section introduces the WSNs and the sensor data modeling.

2.1 Wireless sensor networks

WSN [1–3] are mature topics in literature and it has been widely used in many industrial and consumer applications such as the industrial processing monitoring and control [4], machine and building health monitoring [5] due to its efficiency, low-cost, 24-hour sensing ability, etc. WSNs are more popular and necessary in countries like Japan, where the nature disasters happen quite often. By deploying the sensor nodes in the scene of interest and connecting the sensors using wireless such as Wi-Fi [8], WSNs can be utilized to help monitor the scene of interest and detect the unusual events. After collecting the data, the control center can then analyze the captured sensor data to judge whether there is unusual and disaster happening or not, hence the scene of interest can be monitored by the sensor nodes[5].

2.2 Sensor data modeling

Employing WSNs to monitor the scene of interest involves the sensor node deployment, sensor data collection, and sensor data post-processing. The sensor data characteristics greatly affect these components. But WSN differs from each other due to the diversity in terms of the sensor node types, WSN topologies, network protocols, etc. Hence each WSN should have its unique analysis. In literature, there are many work discussing how to extract knowledge from the captured data [9], in different aspects such as the clustering [10].

From the point of statistics, *Normal distribution* (or *Gaussian*) [11], *Log-normal distribution* [12], *Binomial distribution* [13] and the *Hypergeometric distribution* [14] are widely used distribution models. *Pearson product-moment correlation coefficient (PPMCC)* [15] is used to help calculate the correlations between two variables.

3. System Overview

In this section, we introduce the wireless sensor system discussed in this paper. The sensor data distribution models and the method to calculate the sensor data correlation are introduced in Section 3.2 and Section 3.3, respectively.

3.1. System overview

The system topology is shown in Fig. 1. In this train track monitoring system, there are two terminals (East and West). The train travels between the two terminals on the

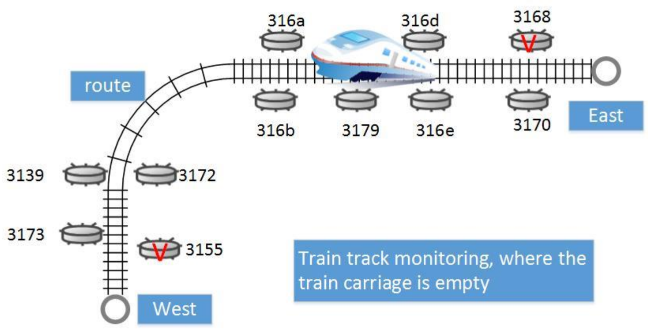


Figure 1: Illustration of WSNs system for the train track monitoring.

track as illustrated in the figure. The carriage of the train

[4] http://www.wi-sun.org/

[5] Local sensor data processing is also feasible if the sensor nodes have certain level computation capabilities, and this could reduce the network traffic.

is empty. The wireless sensors nodes are placed along the train tracks to help monitor whether there is unusual happening or not. The sensors nodes used are the Wi-Sun sensors and there are 11 sensors in Fig. 1. Each sensor detects the position in the x-direction, y-direction and z-direction [6], i.e. in the horizontal direction and vertical direction as shown in Fig. 2. Please note that each sensor has its own original point (0, 0, 0), and (0, 0, 0) is not set to be the original position. The coordination systems are not aligned. Wi-Fi is used for connecting the sensor nodes. The corresponding value generated by the sensor nodes can then be observed at the control center after the transmission over the WSNs.

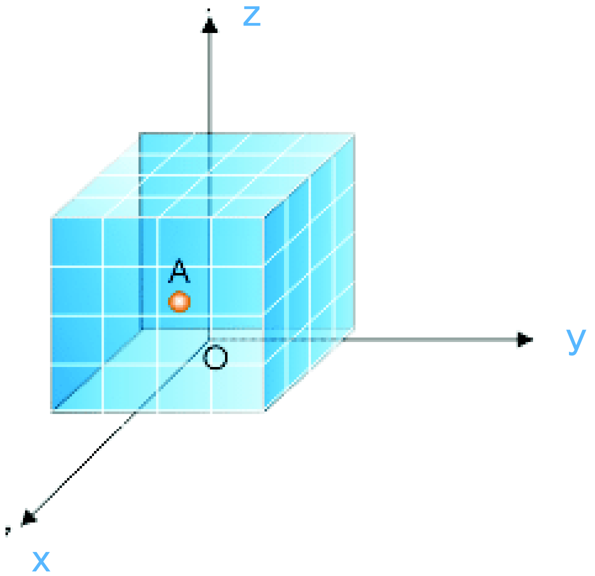


Figure 2：Illustration of the coordination system.

The sensors are running on two modes: normal and power-saving. Sensor nodes running on normal mode capture the environment and broadcast the sensed data to the data sink via the relay nodes periodically. Sensor nodes in power-saving mode also sense the nearby environment, but the sensed data is only delivered to the data sink when the captured data is above the predefined threshold. This is to say that the sensor nodes won’t turn on its antenna when the data is below the threshold. This could greatly save the sensor nodes’ energy consumption, where the energy consumption is critical in WSN. As illustrated in Fig. 1, sensor 3155 and sensor 3168 (marked by red ’V’) are the sensors in power-saving mode and all the other sensor nodes are in normal mode. The sensor nodes in normal mode send the data every one minute. Sensor nodes’ sampling period is 0.05*s*. But the sensors do not capture the scene for 60 seconds during one minute, e.g. a typical sensor in this system only captures the scene of the interest 1 second during one minute and there are 18 data captured during the one second period. But the ’work’ and ’sleep’ time instances are different for different sensors, therefore at each time instance there are data generated by some sensor nodes.

3.2 Distribution model

Normal distribution, *Log-normal distribution*, *Binomial distribution* and the *Hypergeometric distribution* are popular probability distribution models. According to the shape of the sensor data as shown in Section 4, we try the normal distribution in this paper, which is more fit at the first look.

Normal distribution is a very common continuous probability distribution. The normal distribution’s *probability density function* (PDF) could be expressed as follows:

$$f(x) = \frac{1}{\delta\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\delta^2}} \qquad (1)$$

In equation (1), x stands for the variable, μ is the mean of all the samples and the δ is the standard deviation.

3.3 Correlation coefficient

The correlations between different sensor nodes’ data is important for the WSNs. To calculate the correlation between the different sets of sensor data captured, we apply the PPMCC. PPMCC is a measure of the linear correlation (dependence) between two variables *X* and *Y*. The result of PPMCC is a value between +1 and −1 inclusive, where the 1 indicates a total positive correlation, 0 means there is no correlation and −1 represents the total negative correlation. The way to calculate PPMCC is as follows

$$\rho_{X,Y} = \frac{cov(X,Y)}{\delta_X \delta_Y} \qquad (2)$$

where *cov* is the covariance and can be calculated from the following equation

$$cov(X,Y) = E[(X-\mu_x)(Y-\mu_y)] \qquad (3)$$

the *E* is the expectation and μ_X (μ_Y) is the mean of *X* (*Y*). δ_X (δ_Y) is the standard deviation of variable *X* (*Y*).

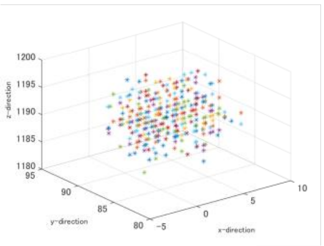
4. Formulation and Simulation results

In this section, we introduce how the captured sensor data looks like and the sensor data characteristics including

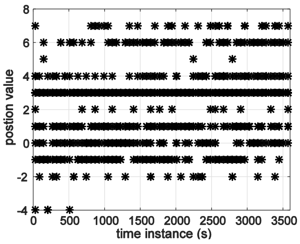
[6] The sensor data can be found in https://mtb-rims.m2mtestbed.jp/rims

the data distribution and different sensors' correlation.

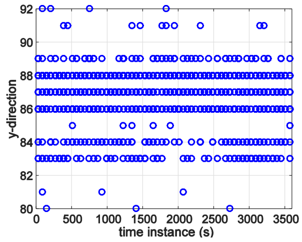
4.1 Sensor data illustration



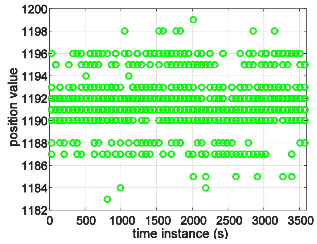
(a) 3D distribution.



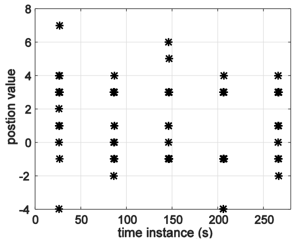
(b) x-direction data distribution



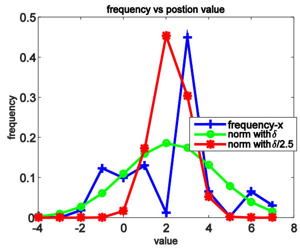
(c) y-direction data distribution.



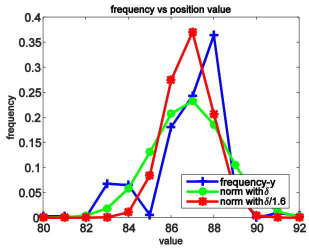
(d) z-direction data distribution.



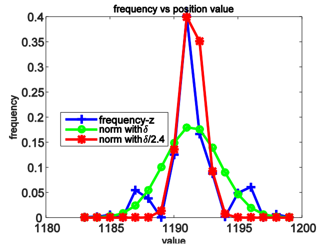
(e) x-direction data distribution. Small scale



(f) x-direction data frequency.



(g) y-direction data frequency.



(h) z-direction data frequency.

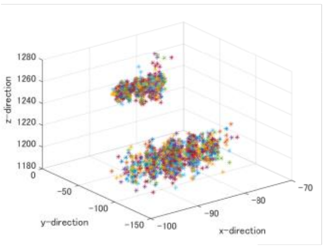
Fig. 3. The sensor node 316b' data distribution between 2015/03/27 10: 00 and 2015/03/27 11: 00.

As mentioned in Section 3, each sensor will detect the position value in the x-direction, y-direction and z-direction with the sampling period to be 0.05*s*. The sensor node in normal mode sends its captured data every one minute, and the sensor node in power saving mode sends the data only when the captured data is larger than the preassigned threshold. From the sensor data set, we can observe that, most of the time, the sensor data in the power-saving mode is unavailable at the control center, i.e. it has not been delivered.

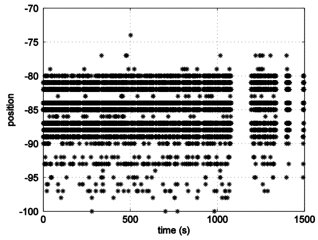
We notice that it is not guaranteed that all the sensor nodes broadcast data, and we choose to show the data between 2015/03/27 10: 00 and 2015/03/27 11: 00. During this time period, the control center has the sensor data coming from all the sensors. The sensor node 316b's data distribution is shown in Fig. 3. We plot all the points in the 3D space in Fig. 3 (a), we can observe that the data points are distributed in a limited space. To better illustrate the data distribution, we plot the position value in the x-direction, y-direction and z-direction across time in Fig. 3 (b), (c) and (d), respectively. From these three plots, we can find that the position value vary slightly across time. We mentioned that each sensor node does not work 60 seconds during one minute, and we illustrate this in Fig. 3 (e). Fig. 3 (e) plots the data distribution in the x-direction from time 0s to time 280s, and we can observe that the sensor data only exist at specific time instances. One explanation for this design is to save the critical battery life. Fig. 3 (f), (g) and (h) illustrate the x-direction, y-direction, and z-direction position value distribution in the frequency domain, respectively. From these figures, we can see the distribution has the similar shape with the normal distribution. To make the peak of the normal distribution and the frequency distribution match, we also choose smaller variance value for higher peaks and the results are shown using the red line in the figures. The δ is the standard variance value of the sensor data set and the standard variance value used for the red line is δ/2.5 in Fig. 3 (f) as written in the legend. We also use the similar way to show the data distribution of the sensor 3168 in Fig. 4, which is a sensor that works in the power-saving mode. From the figure we can observe the y-direction position value separate into two parts, which might be the reason the sensor nodes are activated to broadcast the data to the sink node. The x-direction value vary slight across time and z-direction position value vary a little big larger across time comparing with the x-direction sensor data.

After 1500 seconds, the sensor is not activated for broadcasting the data to the sink node. From the sensor data distribution shape, we can observe that x-direction position value has the shape of the normal distribution. y-direction and z-direction value has multiple peaks, and it has the shape of the normal distribution around each peak.

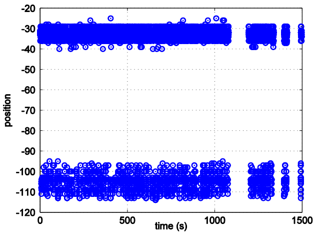
A warning message is sent at 2015/03/16 11: 23, indicating that the train needs to lower its speed before approaching the 'curve'. To show what happened during this period, we collect the data from the sensor 3173, 316*d*, 3168 and 3179 between 2015/03/16 11: 00 and 2015/03/16



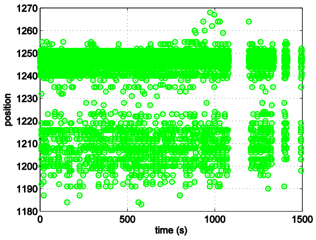
(a) 3D distribution.



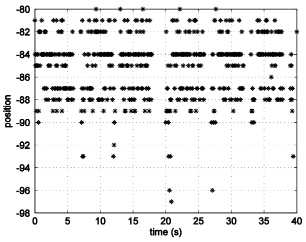
(b) x-direction data distribution.



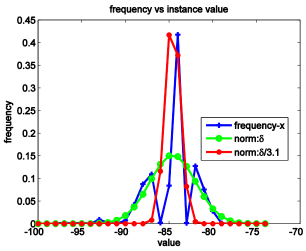
(c) y-direction data distribution.



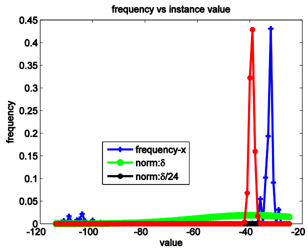
(d) z-direction data distribution.



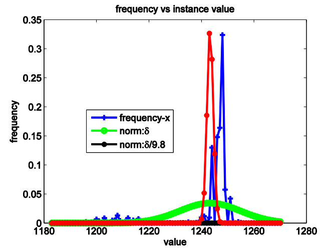
(e) x-direction data distribution in small scale



(f) x-direction data frequency.

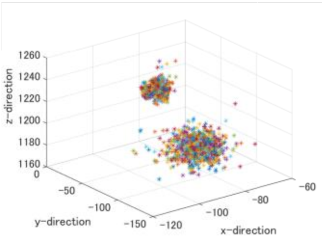


(g) y-direction data frequency.

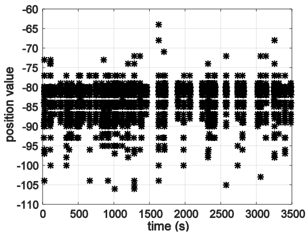


(h) z-direction data frequency.

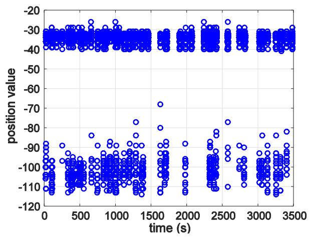
Fig. 4. The sensor node 3168' data distribution between 2015/03/27 10: 00 and 2015/03/27 11: 00.



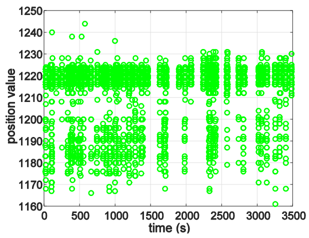
(a) 3D distribution.



(b) x-direction data distribution.

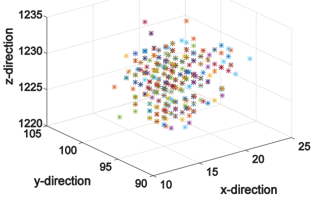


(c) y-direction data distribution.

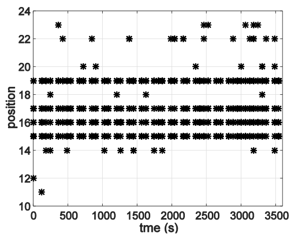


(d) z-direction data distribution.

Fig. 5. Illustration of the data distribution of sensor 3168 during emergency.

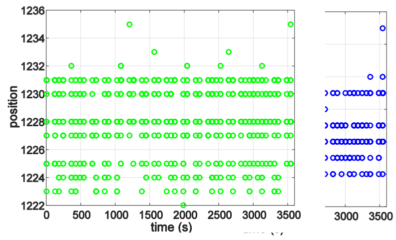


(a) 3D distribution.



(b) x-direction data distribution.

(c) y-direction data distribution.



(d) z-direction data distribution.

Fig. 6. Illustration of the data distribution of sensor 3179 during emergency.

12: 00. The power saving sensor's (ID 3168) data distribution is shown in Fig. 5. Fig. 5 (a) illustrates the overall sensor data distribution and we can observe that all the sensor data is distributed within a small space. Fig. 5 (b), (c) and (d) shows the position value in the corresponding direction across time, where the x-axis stands for the time. We could observe similar distribution pattern comparing with the Fig. 4. The y-direction value separate into two different parts. The other two directions' data distributions are more or less the same. At the time of the warning, the sensor data behaves similar with the data captured at the rest of the time. This indicates this sensor is not for this 'emergency' detection.

We also show the data generated by sensor 3179 from 2015/03/16 11 : 00 to 2015/03/16 12 : 00 and the corresponding data distribution is shown in Fig. 6, where we can observe similar pattern as Fig. 3. Sensor 316*d* and 3179 also have similar data distribution as 3179, due to the page limit, we omit them.

4.2 Correlation

We also calculate the sensor data correlations based on the method introduced in Section 3.3, and the results are shown in Table I. Given the time instances of different sensors are not exactly the same, we use the instance

TABLE I
CORRELATION COEFFICIENT.

ID	316b	3179	316e	3170	3173	3139	316a	316d	3168
3172	0.001, 0.009, -0.0023	-0.0787, 0.0121,0.0177	0.0167,0.0089, 0.0071	-0.03,NaN,-0.0753	-0.0001,-0.0118,0.0257	-0.001,0.0090, NA	0.0057, NaN, 0.0367	0.001, 0.0090,-0.0023	0.01,-0.0025,0.0238
316b		-0.0693,0.0352, 0.0348	-0.1, -0.0176, -0.0097	0.0576, NaN, 0.0230	0.0093, 0.0227,0.0933	0.0483, -0.0152, NA	0.0513, NaN, 0.0301	-0.0066, 0.0103, -0.0351	0.0182,0.0589,0.0091
3179			0.0843, 0.0080, 0.0115	0.0209, NaN, -0.0022	-0.0364, 0.0702, 0.0447	0.0024, -0.0250, NA	-0.0388, NaN, 0.0179	0.0870, 0.0660, 0.0447	-0.0396, -0.0908, -0.0694
316e				-0.0429, NaN, 0.0311	0.0406, 0.0421, 0.0566	-0.0899, -0.0083, NA	-0.0280, NaN, -0.001	0.1255, 0.0157, -0.0433	-0.01068,-0.0818,0.0815
3170					0.0226, NaN, -0.0050	0.0465, NaN, NA	-0.0132, NaN, 0.0134	-0.0440, NaN, -0.0159	0.0030, -Inf, 0.0226
3173						-0.0534, -0.0423, NA	0.0028, NaN, -0.0079	-0.0059, 0.0382, 0.0368	0.0538,0.0205,0.0024
3139							0.0563, NaN, 0.0370	-0.0754, 0.0373, 0.0214	0.0111,0.0452,0.0405
316a								-0.0756, NaN, 0.0060	-0.2330, -Inf, 0.2411
316d									-0.0317, 0.0474,0.0399

numbers instead of the time instance numbers except for the sensor 3168. For sensor 3168, we only use the sensor data samples at exactly the same time instances with other sensor nodes. From the results we can observe that the correlation is not that strong. One possible explanation is that the time instances are not exactly the same.

Please note that sensor node 3155 is omitted in this table, since it has too few samples. The sensor nodes 316*a*, 316*b*, 316*d* and 316*e* have relatively stronger correlations.

5. Conclusion

WSNs are widely used in many applications. The sensor data characteristics greatly affect the sensor nodes deployment, protocol design, post-processing, etc. In this paper, we study the sensor data characteristics in a wireless train track monitoring system, where the Wi-Sun sensor nodes are placed along the train tracks to help detect the unusual. From the results, we can observe that the sensor data distribution has the shape of the normal distribution, and the correlation between the sensor data captured by different sensor nodes is not strong. One possible reason is that the capturing time instances are different.

In the future, we will design the wireless sensor unusual notification system taking these observations into consideration.

6. Acknowledgment

The research results have been achieved by "Research and Development on Fundamental and Utilization Technologies for Social Big Data", the Commissioned Research of National Institute of Information and Communications Technology (NICT) , JAPAN.

Reference

[1] J. Yick, B. Mukherjee, and D. Ghosal, “Wireless sensor network survey,” *Computer networks*, vol. 52, no. 12, pp. 2292–2330, 2008.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “Wireless sensor networks: a survey,” *Computer networks*, vol. 38, no. 4, pp. 393–422, 2002.

[3] J. N. Al-Karaki and A. E. Kamal, “Routing techniques in wireless sensor networks: a survey,” *Wireless communications, IEEE*, vol. 11, no. 6, pp. 6–28, 2004.

[4] V. C. Gungor and G. P. Hancke, “Industrial wireless sensor networks: Challenges, design principles, and technical approaches,” *Industrial Electronics, IEEE Transactions on*, vol. 56, no. 10, pp. 4258–4265, 2009.

[5] G. J. Pottie and W. J. Kaiser, “Wireless integrated network sensors,” *Communications of the ACM*, vol. 43, no. 5, pp. 51–58, 2000.

[6] K. K. Khedo, R. Perseedoss, A. Mungur *et al.*, “A wireless sensor network air pollution monitoring system,” *arXiv preprint arXiv:1005.1737*, 2010.

[7] K. Lorincz, D. J. Malan, T. R. Fulford-Jones, A. Nawoj, A. Clavel, V. Shnayder, G. Mainland, M. Welsh, and S. Moulton, “Sensor networks for emergency response: challenges and opportunities,” *Pervasive Computing, IEEE*, vol. 3, no. 4, pp. 16–23, 2004.

[8] F. Ohrtman and K. Roeder, *Wi-Fi Handbook: Building 802.11 b Wireless Networks*. McGraw-Hill New York, NY, 2003, vol. 67.

[9] A. Mahmood, K. Shi, S. Khatoon, and M. Xiao, “Data mining techniques for wireless sensor networks: A survey,” *International Journal of Distributed Sensor Networks*, vol. 2013, 2013.

[10] J. Gama, P. P. Rodrigues, and L. Lopes, “Clustering distributed sensor data streams using local processing and reduced communication,” *Intelligent Data Analysis*, vol. 15, no. 1, pp. 3–28, 2011.

[11] Y. L. Tong, *The multivariate normal distribution*. Springer-Verlag New York, 1990, vol. 199017.

[12] M. Lynch, B. Walsh *et al.*, *Genetics and analysis of quantitative traits*. Sinauer Sunderland, 1998, vol. 1

[13] C. J. Krebs *et al.*, *Ecological methodology*. Benjamin/Cummings Menlo Park, California, 1999, vol. 620.

[14] D. G. Chapman, *Some properties of the hypergeometric distribution with applications to zoological sample censuses*. University of California Press, 1951, vol. 1, no. 7.

[15] A. Onwuegbuzie, L. Daniel, and N. Leech, “Pearson product-moment correlation coefficient,” *Encyclopedia of measurement and statistics*, pp. 751–756, 2007.