IoT 환경에서의 효율적인 LSTM 구성

이종원[1] · 황철현[2] · 이성옥[3] · 송현옥[4] · 정회경[1]

[1]배재대학교 · [2]데이터마루(주) · [3]대우직업전문학교 · [4]다솜정보

Efficient LSTM Configuration in IoT Environment

Jongwon Lee[1] · Chulhyun Hwang[2] · Sungock Lee[3] · Hyunok Song[4] · Hoekyung Jung[1]

[1]PaiChai University · [2]DataMalu(co) · [3]DaeWoo Vocational Training Institute · [4]Dasommedia

E-mail : starjwon@naver.com / chhwang@einssnc.com / jesuissarah@hotmail.com / paperblue21@hanmail.net / hkjung@pcu.ac.kr

ABSTRACT

Internet of Things (IoT) data is collected in real time and is treated as highly reliable data because of its high precision. However, IoT data is not always highly reliable data. Because, data be often incomplete values for reasons such as sensor aging and failure, poor operating environment, and communication problems.

So, we propose the methodology for solve this problem. Our methodology implements multiple LSTM networks to individually process the data collected from the sensors and a single LSTM network that batches the input data into an array. And, we propose an efficient method for constructing LSTM in IoT environment.

키워드

Data Quality, IoT, Deep Learning, Recurrent Neural Network, LSTM

Ⅰ. Introduction

IoT technology helps to create a more sophisticated virtual world by recording the real world more closely. Therefore, IoT technology is the next generation tool that transforms most of our everyday life and industry[1]. IoT is defined as a global network with an infrastructure that has self-configuring capabilities [2].

Since the sensors connected to each other using various communication technologies form a network while interacting with each other, the data transmitted from each object must be reliable. However, the level of quality of IoT data is threatened due to external exposure or moving objects, the physically unprotected networks or local area networks, and the aging of the natural environment or objects.

In this paper, when LSTM is applied to quality problems such as missing data generation in IoT environment, the accuracy of prediction depends on the dimensionality of the input data. In the IoT environment, multiple data are collected at the same time, so it is possible to construct an individual LSTM network for each sensor or to integrate a large number of data into one LSTM network. In this paper, we try to show how the difference between the two methods affects the quality of IoT data.

Ⅱ. System Design

LSTM learns data input in time series (tn-1… t0) and predicts data of next time(t1). Assuming that the predicted value provides an accurate value above a certain level, the difference between the predicted value and the input data indicates the possibility of error data. In particular, if the input data has a missing value, it can be corrected to the predicted value calculated by the LSTM.

Figure 1 is a general model that uses LSTM to improve the quality of time series data. In Figure 1, t + 1 are the current time at which the missing value occurred and t2 ... t0 is the value of the previous data of the sensor.

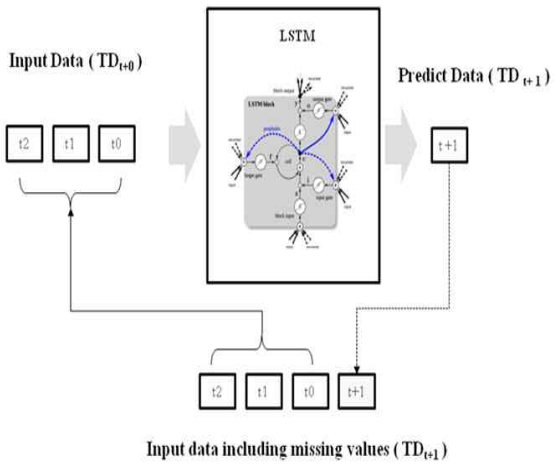

Figure 1. Missing Value Prediction Model Using LSTM

We looked at the most common model of using LSTM for data quality problems. However, in actual IoT environment, it is rare to collect only one sensor data. Most of the time, data is collected from many sensors at the same time.

In this case, the network is designed by determining whether the data input at the same time is processed as one LSTM input or each independent LSTM. In this process, it should not be chosen as expectation that the input of data at the same time without experiment or verification of the data environment will better describe the situation in which the data is generated. Therefore, in this paper, the difference of the prediction rate is verified through the experiment when the two methods are applied.

III. Conclusion

In the experiment using whole data, the individual LSTM construction method showed low error rate in all sensors. In some data experiments, 95 LSTM construction methods showed low error rate in 95 sensors. In both cases, it is suggested that the construction method of individual LSTM has higher predictive power than the method of inputting data at once. In particular, the error rate increases from 29% to 42% depending on the input method. This suggests that constructing and using LSTM by inputting collected data separately has better results in terms of long-term dependence.

The input data to be processed in one LSTM network is not only due to its ease of construction, but also to consider the effect of the data appearing at the same time. Experiments have shown that this method, however, reduces prediction accuracy compared to individual network conception methods. Therefore, we conclude that LSTM should be constructed separately for each number of time series data even in the environment where a large number of data is collected at the same time.

References

[1] B. Fekade, T. Maksymyuk, M. Kyryk, and M. H. Jo, "Probabilistic Recovery of Incomplete Sensed Data in IoT," *IEEE Internet of Things Journal*, Vol. 5, No. 4, pp. 2282- 2292, Jul. 2017.

[2] R. Minerva, A. Biru, and D. Rotondi, "Towards a Definition of the Internet of Things (IoT)," *IEEE Internet Initiative*, 2015.