International Journal of Advanced Smart Convergence Vol.9 No.4 173-178 (2020)
http://dx.doi.org/10.7236/IJASC.2020.9.4.173

IJASC 20-4-22

Deep Learning-Based Smart Meter Wattage Prediction Analysis Platform

Seonghoon Jang[1 †], Seung-Jung Shin[2]

[1†] PhD student, Department of IT Convergence , Hansei University, Korea
steezjang@gmail.com
[2]Professor, Department of ICT Convergence, Hansei University, Korea
expersin@hansei.ac.kr

Abstract

As the fourth industrial revolution, in which people, objects, and information are connected as one, various fields such as smart energy, smart cities, artificial intelligence, the Internet of Things, unmanned cars, and robot industries are becoming the mainstream, drawing attention to big data. Among them, Smart Grid is a technology that maximizes energy efficiency by converging information and communication technologies into the power grid to establish a smart grid that can know electricity usage, supply volume, and power line conditions. Smart meters are equient that monitors and communicates power usage.

We start with the goal of building a virtual smart grid and constructing a virtual environment in which real-time data is generated to accommodate large volumes of data that are small in capacity but regularly generated. A major role is given in creating a software/hardware architecture deployment environment suitable for the system for test operations. It is necessary to identify the advantages and disadvantages of the software according to the characteristics of the collected data and select sub-projects suitable for the purpose. The collected data was collected/loaded/processed/analyzed by the Hadoop ecosystem-based big data platform, and used to predict power demand through deep learning.

Key words: *Smart Meter, Deep-Learning, Hadoop Echo System, Bigdata*

1. INTRODUCTION

Big data is drawing attention as a key technology as various fields such as smart energy, smart city, artificial intelligence, the Internet of Things, unmanned cars, and robot industries have recently begun to reorganize the industry called the Fourth Industrial Revolution, which links people, things, and information.. Among the recent issues in the big data field, big data analysis is making great progress in each field, with real-time analysis of sensor data generated by Lee Se-dol and AlphaGo, 1GB per second of Google's unmanned vehicles, successful autonomous driving, and only big data analysis using real-time SNS(Social networking services) data in the U.S. presidential election predicting Trump's victory. In addition, artificial intelligence has been making progress in many areas due to the recent development of computer technology, which is based on the technology that enables rapid processing of accumulated data by applying machine learning to high-performance computers.

Manuscript Received: November. 23, 2020 / Revised: November. 28, 2020 / Accepted: December. 1, 2020
Corresponding Author: expersin@hansei.ac.kr
Tel: +82-10-3664-0666, Fax: +82-31-450-5172
Professor, Department of ICT Convergence, Hansei University, Korea

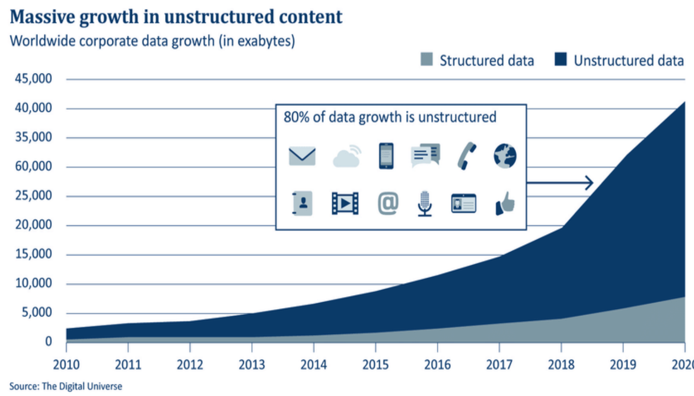


Figure 1. S/W importance

Data generated in the last two years accounts for 80% of the world's data, and the amount of data generated on Earth in the future is expected to grow to around 35,000 exabytes by 2020. bigdata analysis, which analyzes data from the past to the present and predicts my future by deriving meaning, is a concept that implies the format, processing speed, meaning of data rather than simply judging data by size, and collectively refers to the entire process of data collection, storage, search and analysis, not just statistical analysis. Hadoop software is at the center of the core technology of big data. Hadoop is a freeware Java software framework that supports distributed applications operating in large computer clusters capable of processing large amounts of data, a key technology that implements HDFS (Hadoop Distributed File System) and MapReduce1 that can replace Google's file system, a distributed processing system. Hadoop is formed by sub-projects around HDFS and MapReduce's core projects, which form the Hadoop eco system, which allows the system to be used to process big data into four layers (collection/loading/discovery/processing/analysis/application).

Processes them in installments and uses data to make predictions through machine learning.

2. Test environment domain

The following real-world simulation environment is established and the power volume prediction and consumption pattern analysis is performed through the data collection/loading/processing/discovery stages through big data solution.

- Create a smart meter log simulator because the actual 100 smart meters cannot be operated.
- The data collected by smart meters is a power generation log simulator that generates power consumption according to the data collection cycle based on the number of households / monthly power usage statistical data of KEPCO.

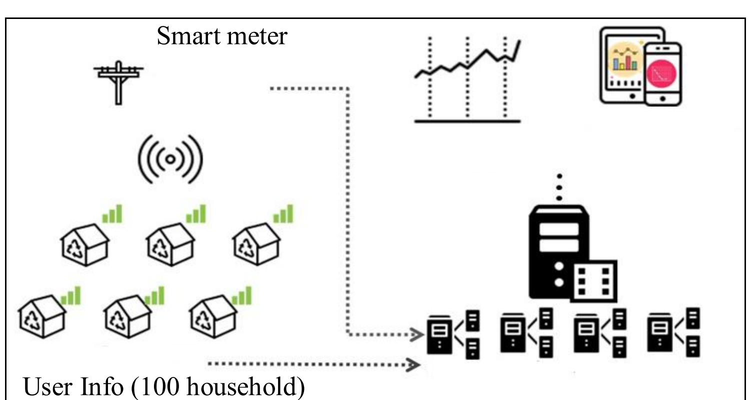


Figure 2. Test Environment

- Smart meter data is collected → loading → processing/discovery → analysis/application process and reconfiguration into easy-to-use datasets at each stage.
- Data mining such as exploration and analysis, and prediction of power consumption by applying machine learning techniques.

3. SYSTEM ARCHITECTURE

If a big data project is carried out to collect real-time, large-capacity data, dozens to hundreds of Hadoop cluster nodes should be configured for power-use homes, but it is practically impossible to configure such a big data environment in this project. Therefore, the goal was to build a big data pilot environment to the level where workstations applicable to the project could be utilized and to build a Hadoop ecosystem so that the pilot environment could cover all the key technologies and functions of big data.

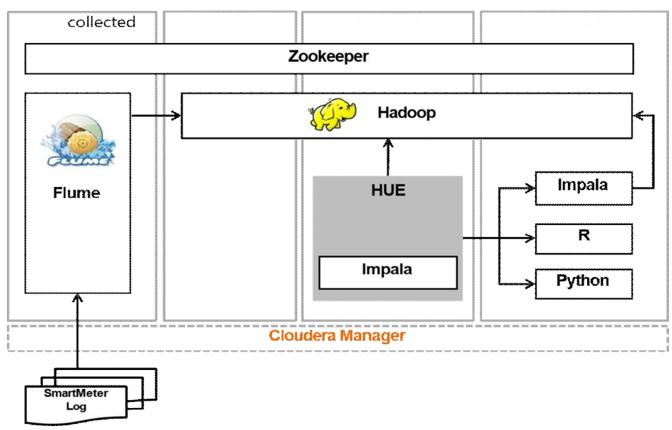


Figure 3. Structure of System

4. DATA ANALYSIS FORECAST

4.1 Current Status of Electricity Consumption for Housing

From 2014 to 2018, the data mart made of impala from Hadoop File System (HDFS), which stores log data that occurred every 15 minutes in smart meters, was used to analyze power consumption patterns of all households.

	1	2	3	4	5	6	7	8	9	10	11	12	AVG	MIN/MAX
2014	339	323	310	300	297	287	290	310	322	293	289	305	305	52
2015	352	335	322	311	308	298	302	322	335	304	301	317	317	54
2016	353	336	323	312	309	299	302	323	336	305	301	318	318	54
2017	338	322	309	299	296	286	290	309	321	292	289	304	305	52
2018	342	325	313	302	299	289	293	313	325	295	292	307	308	52

Figure 4. Monthly average annual power usage(KEPCO)

The table above is a monthly power consumption trend derived from log data generated by smart meters,

and the graph shows the average monthly power consumption for 2014-2018 years. According to the table, the demand for residential power tended to be high in the months corresponding to winter and summer, and both 2014-2018 showed more power consumption in winter. The difference between maximum and minimum power consumption by year is 52 kWh in 2014, 54 kWh in 2015, 54 kWh in 2016, 52 kWh in 2017 and 53 kWh in 2018, the maximum monthly power consumption by year and minimum.
The power consumption gap is around 50kWh. It also showed that the average amount of electricity in January was 339 kWh, the highest, and dropped to 287 kWh in June before rising again to 322 kWh in September. As such, the average power consumption during winter and summer is high.

MEMBER	1	2	3	4	5	6	7	8	9	10	11	12	AVG	MIN/MAX
1	229	218	210	202	200	194	196	210	218	198	196	206	206	35
2	321	305	294	284	281	272	275	294	305	277	274	289	289	49
3	356	339	327	315	312	302	306	326	339	308	305	321	321	55
4	375	357	344	332	328	318	322	344	357	324	321	338	338	58
5	394	375	361	349	345	334	338	361	375	341	337	355	355	60
6	436	414	399	385	381	369	373	398	414	376	372	392	392	67
7	431	410	395	381	377	365	369	395	410	373	368	388	388	66
8	515	491	472	456	451	437	442	471	490	446	440	464	465	79
9	475	451	433	419	414	401	405	433	451	410	404	426	427	75

Figure 5. Monthly average power usage by household member (KEPCO)

The following table compares the average monthly power consumption by the number of members of households in 2014-2018 and the graph shows the average power consumption by the number of households in 2014-2018: According to the tables and graphs, the higher the number of household members, the higher the power consumption, the more likely it is that the correlation between the number of household members and the amount of power consumption can be strong. In addition, the difference between the maximum and minimum power consumption by household members was the lowest at 35 kWh for single-person households, and the highest at 79 kWh for eight-person households, indicating that the greater the number of households, the greater the gap between maximum and minimum power consumption. Therefore, the current state of electricity consumption for residential use shows a significant increase in electricity consumption for residential use in winter and summer, and the higher the number of household members, the more (+) the correlation appears.

4.1 Daily power consumption forecast

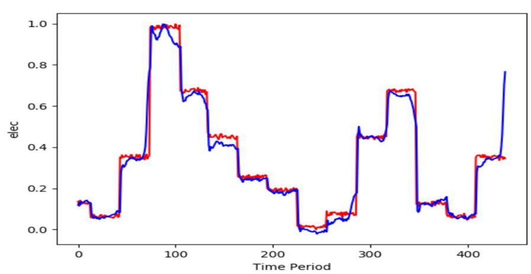
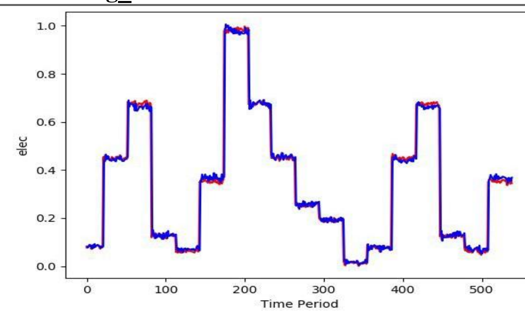
It predicted daily power demand using 2014-2018 power data. In addition, the 2014-2018 weather data (average temperature, lowest temperature, highest temperature, average dew point temperature, average local air pressure, average sea surface air pressure, total solar radiation, and average ground temperature) were used as variables for predicting power demand using the weather data opening portal. The following dataset is used to predict power demand.

Table 1. Weather information dataset

Data	avg temp	min temp	max temp	dew point temp	avg field press	min sea press	total daily dose	avg ground temp	power usage
20140101	5.1	0.4	9.7	-2.4	1007.4	1015.9	9.79	1.8	847701.4256
20140102	2.6	-2.2	9.1	-6.1	1013.1	1021.7	11.01	1.3	847868.9795
20140103	2.1	-3.4	9.6	-5.5	1009.3	1018	8.42	0.7	848362.2574
20140104	1	-2.7	6.5	-6.2	1011	1019.7	11.58	0.6	847253.1833
20140105	-0.8	-5.9	5.5	-6.3	1015.5	1024.3	10.97	-1.3	848006.2568

The dataset is daily data from 2014 to 2018, with a total of 1,826 rows and hyperparameters as follows.

Table 2. Comparison of predicted results

1st	2nd
Input_data_column_cnt : 9	**Input_data_column_cnt : 9**
output_data_columns_cnt : 1	**output_data_columns_cnt : 1**
seq_length : 365	**seq_length : 30**
rnn_cell_hidden_dim : 20	**rnn_cell_hidden_dim : 20**
forget_bias : 1.0	**forget_bias : 1.0**
num_stacked_layers : 1	**num_stacked_layers : 1**
keep_prob : 1	**keep_prob : 1**
epoch_num : 1000	**epoch_num : 1000**
learning_rate : 0.01	**learning_rate : 0.01**
	

LSTM's activation function used softsign, the optimization function used Adam Optimizer, and the loss function used the mean square error. The study set and test set were divided by the study set 0.7 and the test set 0.3. As a result of learning, the seq_length was lowered to 30 per month, and other conditions were set the same and proceeded again. As before, LSTM's activation function used softsign, the optimization function used Adam Optimizer, and the loss function used the mean square error. Similarly, the learning set and the test set were divided equally by the learning set of 0.7 and the test set of 0.

Better results were obtained when seq_length was set at 365 days, or 30 days a month, than at one year. Further reductions in seq_length did not change much in accuracy, so the model was decided at 30 days a month.

5. Conclusion

Designing a smart meter-based power demand forecasting platform, the company collected log data generated and implemented a series of processes, including loading, navigation and processing, and analysis. Log data generated through plums were collected and collected immediately by loading or processing the collected data into Hadoop. Hive, Impala and Spark were used to process and explore loaded bulk data quickly and effectively. Using the loaded data, various analyses and demand forecasts were made with R and Python.

The amount of data processed by the big data solutions designed above is as follows:

- 15-minute power generation data = 100 households * 15-minute cycle * 5-year = 17,280,000
- Power generation data per second = 100 households * 1 second cycle * 1 week = 60,480,000

We believe that while generating data from two different individuals and processing it into a solution, approximately 77 million data were collected loading, processing discovery, analysis-application processes through the solution, and that the role of the pilot project was sufficient.

Due to a hardware problem, real-time data was generated and tested for only 7 days. Based on 100 households, this data generates approximately 60 million data per week and over 1 billion data per year. This was a very limited sample of 100 households, but it was confirmed that a large amount of data had been collected. In fact, the number of domestic households in the 2017 Population and Housing Survey was 20168,000, and KEPCO designed the actual smart meters every 15 minutes because of the infinite data collection..

REFERENCES

[1] Chan-Ho Moon, Bo-Sung Kwon, Dong-Jin Bae, Kyung-Bin Song. "Load Forecasting Algorithm on Weekdays Using Solar Radiation Weight." Journal of the Korean Institute of Illuminating and Electrical Installation Engineers 34.6 (2020): 40-47.
DOI : 10.5207/JIEIE.2020.34.6.040

[2] Dohyun Kim, Ho Jin Jo, Myung Su Kim, Jae Hyung Roh, Jong-Bae Park. "Short-Term Load Forecasting Based on Deep Learning Model." The transactions of The Korean Institute of Electrical Engineers 68.9 (2019): 1094-1099.
DOI : 10.5370/KIEE.2019.68.9.1094

[3] Chi-Yeon Kim, Chae-Rin Kim, Dong-Keun Kim, Hyeong-Jin Choi, Si-Sam Park, Soo-Hwan Cho. "Scaled RMSE and Shewhart Control Chart-based Abnormal Reference Day Detection Method to Improve the Forecasting Accuracy of Community-level Power Demand." The transactions of The Korean Institute of Electrical Engineers 69.2 (2020): 245-257.
DOI : 10.5370/KIEE.2020.69.2.245

[4] Jo, Hyunsoo Lee. "Electricity Demand Forecasting Framework using Modified Attention-based LSTM." Journal of Korean Institute of Intelligent Systems 30.3 (2020): 242-250.
DOI : 10.5391/JKIIS.2020.30.3.242

[5] Dong-Ha Shin, Chang-Bok Kim. "A Study on Deep Learning Input Pattern for Summer Power Demand Prediction." The Journal of Korean Institute of Information Technology 14.11 (2016): 127-134.
DOI : 10.14801/jkiit.2016.14.11.127

[6] Ji-Won Lee, Hyung-Jun Kim, Mun-Kyeom Kim. "Design of Short-Term Load Forecasting based on ANN Using Bigdata." The transactions of The Korean Institute of Electrical Engineers 69.6 (2020): 792-799.
DOI : 10.5370/KIEE.2020.69.6.792