Data analysis framework for charging facility monitoring systems

Junghoon Lee¹, Jiwon Jung¹, Daejeon Kang², and Gyung-Leen Park¹ **

Dept. of Computer Science and Statistics, Jeju National University¹
Do-Dream Inc.²
Jeju-Do, Republic of Korea
{glpark,jhlee}@jejunu.ac.kr,jjy1809@naver.com,jjeonth@do-dream.co.kr

Abstract. Charging facilities, which are being constructed for the wide penetration of electric vehicles, generate massive amount of real-time status readings. The analysis of those streams provides a useful guideline for power provisioning and facility management. This paper builds a Hadoop-based framework which converts the raw stream data into manageable forms, filters necessary information fields, and creates preliminary statistics for the next-step analysis. Upon the collected records organized on the Linux file system, Pig scripts are implemented to obtain per-charger, per-driver, and per-day number of status reports for our raw data streams. The experiment finds a significant difference between respective electric vehicle entities according to the personal ownership, vehicle locations, and the like.

Keywords: electric vehicle, charging facility, tracking system, big data analysis, Hadoop Pig

1 Introduction

EVs (Electric Vehicles) have many eco-friendly benefits such as energy efficiency, low greenhouse gas emissions, and so on [1]. Hence, many countries are accelerating the penetration of EVs into our daily lives. The main effort is put on the construction of charging facilities over the community area to overcome their short driving range. In the EV-based transportation system, almost every element can be digitally activated and possibly monitored, creating a massive amount of data streams. Intelligently analyzed, those streams will give us a very useful guideline for power provisioning, facility management, and efficient reservation mechanisms for charging stations [2]. Jeju City, Republic of Korea, is also building a citywide charging infrastructure and accumulating the real-time monitoring data from charging facilities.

All newly built chargers are connected to the central data server via an appropriate communication channel and send periodic reports. As shown in Figure

** Following are results of a study on the “Leades INdustry-university Cooperation” Project, supported by the Ministry of Education, (MOE).
1, a single record consists of user ID, demand code, demand power amount, payment code, SOC (State-of-Charge) battery level, power amount, injected battery amount, retail price, battery status code, total battery capacity, battery voltage, battery electricity current, battery temperature, BMS (Battery Management System) version, time left, registration date, report date, and charger registration ID. It includes redundant and unessential fields and some of them are reserved just for the future use. Hence, it is necessary to filter the information really necessary. Here, Apache Hadoop allows us to manage a large amount of data stream and define what is really needed for a goal-specific analysis [3].

Fig. 1. Data stream organization

2 Basic stream analysis

To begin with, the status reports are stored in a text file format in Linux, as shown in Figure 1. For data processing in Hadoop family, it is much convenient to remove all spaces and represent the date type according to ISO8601 format. Then, we implement a Hadoop Pig script program to filter the necessary columns, namely, user ID, demand power amount, report data, and charger ID. Those fields provide basic statistics for the charger operation in our community. Figure 2 plots the number of records for each of 50 chargers during the test operation interval ranging from September 2014 to April 2015. The most actively used charger has 6,363 records, while 21 chargers creates less than 10 records.
Next, Figure 3 shows the number of records for each driver. Here, each driver is assigned a unique ID, but the graph does not display the ID, as it is too long. According to the figure, there exists a great difference between individuals. An EV driver has 3,609 records, while 11 out of 62 drivers have less than 10 records. They seem to either have no personal ownership or have their EVs charged in their homes rather than in public charging facilities. Meantime, 8 drivers have more than 1,000 records and can be considered intensive EV commuters.

Finally, Figure 4 plots the daily statistics for the charging status records. According to the penetration of EVs and chargers, the number of records increases. At the beginning, the monitoring system stopped working from time to time for field tests. However, during the last 30 days, almost every has hundreds of records, thanks to the stabilization of the monitoring system as well as the increased charging on community facilities.
3 Conclusions

As charging infrastructures keep generating a massive amount of real-time data streams, an efficient analysis framework is becoming more important. In this paper, we have presented the data organization of our charging facility tracking system, built a Hadoop-based data processing framework, and conducted a basic analysis. The result has found a significant difference between respective electric vehicle entities according to the personal ownership as well as vehicle locations. As future work, we are planning to develop more intelligent mining programs for detailed pattern recognition and demand prediction, possibly integrating a sophisticated statistical package [4].

References