Big Data Analytics for Electric Vehicle Integration in Green Smart Cities

Boyang Li, Mithat C. Kisacikoglu, Chen Liu, Navjot Singh, and Melike Erol-Kantarci

The authors provide a comprehensive overview of the data analytics landscape on the EV integration to green smart cities. It serves as a roadmap to the future data analytics needs and solutions for EV integration to smart cities.

ABSTRACT

The huge amount of data generated by devices, vehicles, buildings, the power grid, and many other connected things, coupled with increased rates of data transmission, constitute the big data challenge. Among many areas associated with the Internet of Things, smart grid and electric vehilces have their share of this challenge by being both producers and consumers (ie., prosumers) of big data. Electric vehicls can significantly help smart cities to become greener by reducing emissions of the transportation sector and play an important role in green smart cities. In this article, we first survey the data analytics techniques used for handling the big data of smart grid and electric vehicles. The data generated by electric vehicles come from sources that vary from sensors to trip logs. Once this vast amount of data are analyzed using big data techniques, they can be used to develop policies for siting charging stations, developing smart charging algorithms, solving energy efficiency issues, evaluating the capacity of power distribution systems to handle extra charging loads, and finally, determining the market value for the services provided by electric vehicles (i.e., vehicle-to-grid opportunities). This article provides a comprehensive overview of the data analytics landscape on the electric vehicle integration to green smart cities. It serves as a roadmap to the future data analytics needs and solutions for electric vehicle integration to smart cities.

INTRODUCTION

According to International Data Corporation's (IDC's) visionary presentation on "The Digital Universe of Opportunities," the overall created and copied data volume worldwide was 4.4 zettabytes (ZB) in 2013. The volume of data is doubling every two years, and by 2020 the total volume will exceed 44 ZB (44 trillion GB). Besides the volume, the velocity of the data is growing as a result of the advances in communication technologies and the Internet of Things (IoT). Such enormous datasets with high velocity, veracity, and variety are expressed as the big data phenomenon.

Smart grid and electric vehicles (EVs) are among the main drivers of IoT, as they form a large connected network of things, such as vehicles, charging stations, smart meters, intelligent electronic devices (IEDs), and phasor measurement units (PMUs). They are also anticipated to be the drivers of green smart cities by enabling efficient integration of renewable energy and lower emissions. The green smart city vision anticipates almost all flat surfaces, including roads, covered by solar panels to maximize the utilization of solar energy [1]. EVs carry dozens of sensors that provide data including user driving behaviors, battery security via a battery management system (BMS), and grid charge management via charging stations. Drivers, as well, carry smart devices and wearables that contribute to the data generated on roads. With smart, autonomous, self-driving cars, those data will be continuously moving from cars to servers and cars to cars. In the case of EV grid integration (EVGI), their charging and discharging pattern is tightly coupled with the operation, security, and efficiency of the smart grid. In that sense, data analytics play a critical role in EVGI, green smart cities, and other green infrastructure as presented in [2]. In particular, charging planning and harmonization of EVs for selling power back to the grid (i.e., vehicle to grid, V2G) require fast and reliable data analytics techniques.

In this article, we provide a comprehensive survey of existing techniques, and provide a roadmap for future technologies in data analytics for EVGI applications in green smart cities. We start with a brief overview of smart grid and EVs to present the applications and potential challenges. We discuss the sources of big data generation in detail. Then we continue with a survey on data analytics tools that are used in this domain. The article aims to introduce the existing data analytics studies on EVs and smart grid. Hadoop-based cloud platforms, prediction methods, and decision support tools are among the surveyed data analytics studies. The article provides a requirement analysis for future data analytics tools and aims to serve as a roadmap for researchers in this area.

Figure 1 provides an overview of EV integration with V2G, G2V, power and information exchange, heterogeneous communication technologies, data flow, cloud integration, applications, and big data analytics tools. As shown in the figure, EV-EVSE-grid communication can be implemented by power line communications (PLC) and wireless networks. Note that EVSE is the EV supply equipment and is used interchangeably throughout this article. EVs can be charged and

Boyang Li is with Illinois Institiue of Technology; Mithat C. Kisacikoglu is with the University of Alabama; Chen Liu is with Clarkson University; Navjot Singh and Melike Erol-Kantarci are with the University of Ottawa.

Digital Object Identifier: 10.1109/MCOM.2017.1700133

Despite the advantages, there are still challenges for the widespread adoption of EVs. Limited driving range and associated driver range anxiety, long duration for charging, and non-ubiquity of charging stations are the critical barriers to the penetration of EVs.

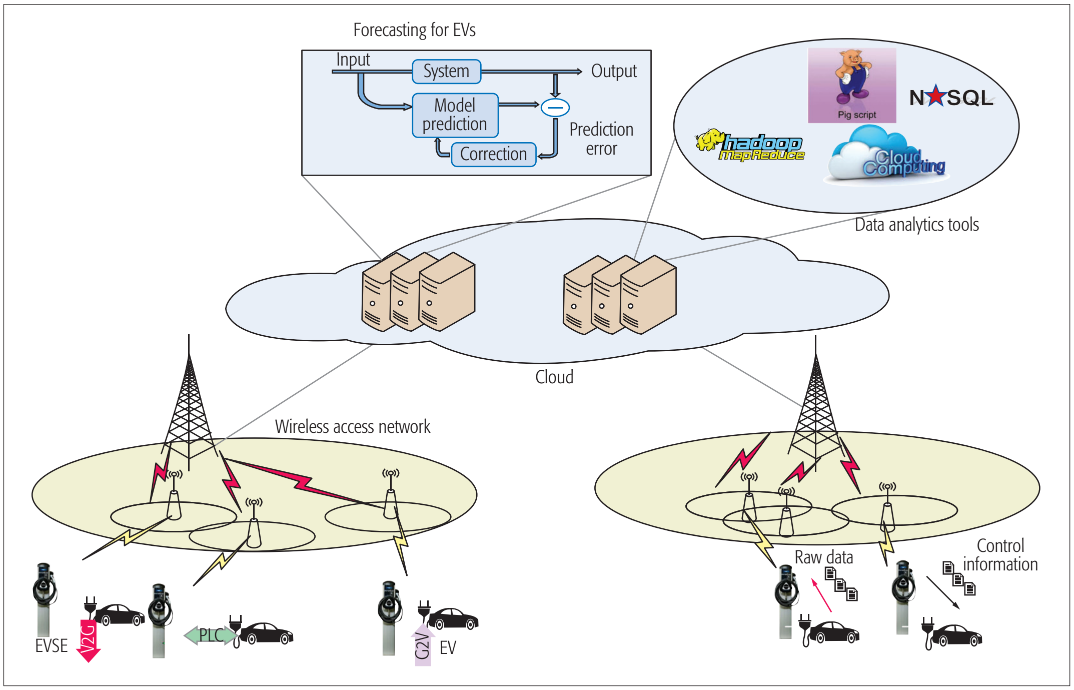

Figure 1. Implementation diagram of EVGI.

discharged in a coordinated fashion. The big data from EVs and all other entities are stored and processed over the cloud for various application purposes including optimized charging. Some widely used big data tools are also plotted in the figure.

The rest of the article is organized as follows. The following section provides an overview of the integration of EVs to the smart grid. Then we describe the sources of big data in EVs and G2V/V2G applications. Following that, we provide a detailed survey on the available big data approaches and platforms for processing EV data. We then discuss the open issues, requirements for data analytics tools tailored for EVs, and some future concepts that require more research. In the final section, we conclude the article.

An Overview of Smart Grid and Electric Vehicle Integration

Smart grid is the modernized electrical grid that integrates advanced sensing, communication, and control functionality for the purpose of enhanced efficiency, reliability, and security in the operation of the utility grid. EVs are either completely or partially powered by their onboard batteries, which are charged by the power grid. Plug-in EVs use less fossil fuels and emit much less CO_2 compared to conventional vehicles and therefore are incentivized by many governments to improve air quality and reduce greenhouse gas emissions. Despite the advantages, there are still challenges for the widespread adoption of EVs. Limited driving range and associated driver range anxiety, long duration for charging, and non-ubiquity of charging stations are the critical barriers to the penetration of EVs. In addition, with a growing EV market, the impact of EVs on the power grid is a matter of concern, especially at the distribution level. This may result in adverse effects such as peak loading, increased losses, voltage unbalance/deviations, and need for additional network reinforcements. Data management in EVGI plays a vital role for healthy integration of these new technology vehicles into the future green smart cities.

EV batteries are charged using onboard chargers and EVSE, also known as charging stations. EVSEs can be located at residential premises, parking lots of commercial buildings, and any roadside charging facility. The EV integration framework enables EVs to be controlled by the smart grid or aggregators via the communication between vehicles and the grid. Communication between EVs and the smart grid can be a mix of wireless and wired technologies including PLC, Zigbee, WiFi, LTE, and fifth generation (5G) wireless networks. A hardware description of EV grid interaction and the EVGI system with bidirectional power and communication architecture are shown in Fig. 2. The communication between an EV and the smart grid includes two concepts: basic signaling and high-level communication. Basic signaling refers to EV charging control methods utilizing the control pilot signal of the charging plug for basic G2V charging control, as shown in Fig. 2. This control is realized by modifying the duty cycle ratio of the control pilot signal, which is already available in all of the plug-in EVs in the market. Communication functions via the current control pilot pin have a fairly simple structure and cannot provide the bidirectional information required between EV and grid. On the other hand, V2G control is achieved via high-level communication that uses PLC superimposed on the control pilot signal. In the case of PLC, V2G communications is overseen by the EV communication controllers (EVCCs) and supply equipment communication controllers (SECCs). While the EVCC and SECC are primary actors, grid operators, charging aggregators, and electricity providers are the secondary actors of the charging communication system. EVs and EVSEs can also communicate through wireless networks for sharing data that is useful for trip planning, real-time pricing, and so on.

In G2V charging, the challenges mostly arise from the increased stress on distribution systems of the smart grid. Due to increased power consumption on the network during peak hours, off-peak hours are preferable for EV charging. In addition, if the night valley in the 24-hour electricity demand profile is filled with EV loads, the ramping up/down costs that occur in the morning/evening can be avoided. Meanwhile, charging during daytime, especially during peak hours of electricity demand, requires extra planning. Moreover, the distribution system suffers from overloading if several EV batteries are fed from the same transformer.

The charging management mainly relies on certain information being available to local (distributed) or global (centralized) controllers. Information exchanged between vehicle and controllers include user departure time, state of charge (SOC) of the battery, charging active/reactive power reference, and user-specific information such as charging preference, vehicle vendor, onboard charger power, and battery capacity. Compared to distributed control, centralized control achieves better utilization of EVs for grid support due to having more information and achieving optimum results. Therefore, central power optimization is one of the most explored analysis methods in EV charging networks to solve congestion related problems. On the other hand, the distributed strategy allows each EV to determine its own charging profile, which may not always result in an optimal aggregated charging regime. However, the distributed approach has gained more attraction in the literature because of its higher flexibility for the EV user, higher reliability, and easier field implementation [3]. In this case, data communication is much lower, and private information is mostly kept in vehicle.

Despite the large number of studies on EV charging, less has been explored on how this data will be attained, how it will be processed for larger penetration of EVs, and how data from other sources could be coupled with data from EVs to predict the behavior of a driver for charging or discharging the battery. In particular, data that are potentially available while the EV is on the move have been underutilized. Obviously, more data give more chance to derive useful insights, but decision making from big data of EVs requires finding the right information in near real time. In this context, data analytics techniques can increase the efficiency of EV data.

During V2G, when EVs act as distributed generators, data analytics become more of a concern and are more needed. When an EV is allowed to sell electricity, information on where the vehicle will be in the next time frame, how much energy will be left in the battery when it is reconnected, how much of this energy will be reserved for trading, load on the utility when the EV is plugged in, and similar information need to be available. Part of this data could be voluntarily made available by the driver, some could be predicted, and the rest could be collected from sensors. In any case, the amount and speed of data flow are quite big, and robust data analytics tools are needed to make effective and timely decisions.

In the following section, before we survey the data analytics tools, we analyze the potential data sources related to EVs. The data from EVs can be vehicle, driver, charging station, or even smart city related such as traffic condition on roads. The vehicle data can come from various sources such as batteries, onboard chargers, and trip logs. In addition, wearables on drivers contribute to the big data of cars. The utility grid power consumption data stream is also important to determine which charging/discharging scenario should be employed for the specific geographical location.

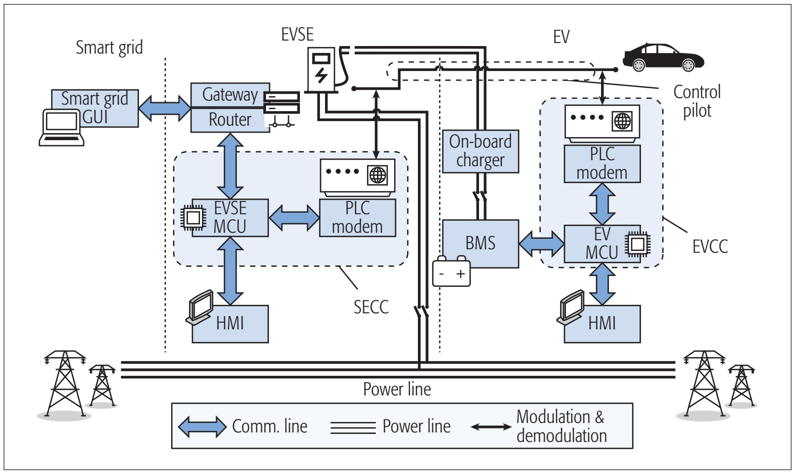

Figure 2. Overview of EV integration with V2G, G2V, heterogeneous communication technologies, data flow, cloud integration, applications, and big data analytics tools.

Big Data of Electric Vehicles

New generation autonomous self-driving cars, whether electric or not, are equipped with hundreds of sensors and surrounded by smart technologies. Furthermore, road infrastructure is also underway with large deployment of connected technologies (i.e., traffic lights, signs, and road cameras). The advances in wireless and vehicular communications enable these smart cars to be able to communicate with the infrastructure and other smart cars. Autonomous connected vehicles and their interaction with smart cities will increase the amount of data that is generated and shared. In addition, drivers carry a number of sensors on their smartphones and wearable devices. In general, IoT, and in particular the Internet of Vehicles (IoV) and Internet of Energy (also known as Energy Internet), benefit from cloud services [4]. Onboard and on-body devices have limited storage and processing capabilities. Meanwhile, their communication capability opens the door to accessing powerful cloud servers. The data from EVs, drivers, charging stations, and infrastructure constitute the big data of EVs, which requires data analytics tools running on the cloud.

Many automobile manufactures allow drivers to check the status of their EVs and remotely control their charging through mobile apps. These applications collect vehicle and trip data. EV data mostly come from onboard electronic control unit (ECUs) and battery management systems (BMSs). SOC of EV batteries is a key parameter for most charging and discharging decisions. BMS logs show SOC information and how an EV battery is performing. Malfunctioning battery cells, and heating and cooling details can be observed by these logs. Based on BMS logs, state of health

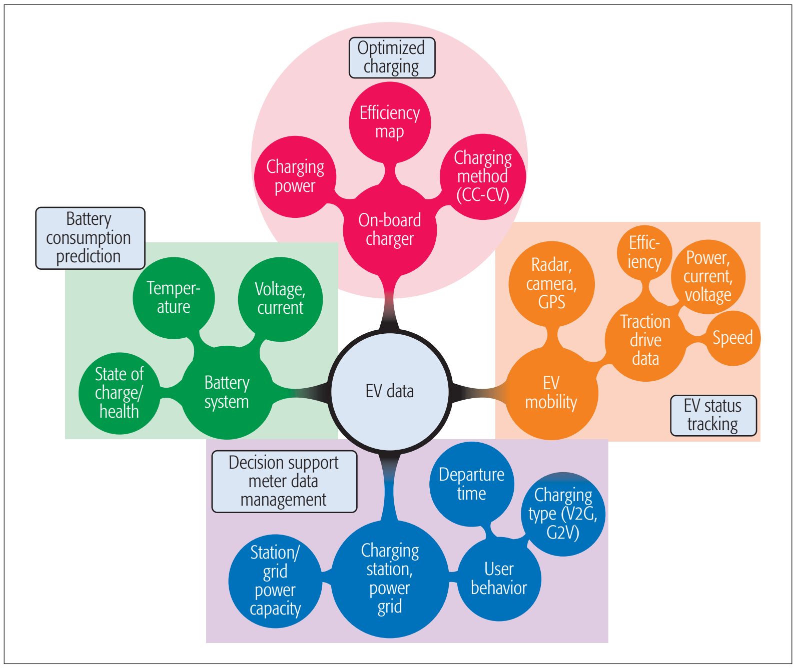

Figure 3. Summary of EV related data sources.

(SOH) information can be obtained, and the impact of V2G services on battery life can be accurately observed.

In addition to the data directly collected from EVs, drivers can voluntarily share information about their driving patterns and charging habits. Trip information including start and end times of journeys, connect and disconnect times of chargers, and the battery SOC can easily be collected. Advanced systems can record details like how much air conditioning is used, or how a driver accelerates or breaks. All these various kinds of data can be used for decision making through data analytics tools.

An important parameter for EV performance is the driving range. For a long time, market acceptance of EVs was low due to range anxiety, which is the worry that the EV battery will run out of power before the destination or a suitable charging point is reached. Big data is frequently used to estimate the driving range, which is an efficient way to diminish the range anxiety. In [5], the authors proposed a classification method related to driving range estimation. The data are classified as standard, historical, and real-time data. Those are defined as follows.

Standard Data: This includes the data obtained from official sources such as scheduled tours and activities from websites, the usual driving time to the destination according to Google Maps, or climatic conditions such as hurricane season or dry season.

Historical Data: It refers to the indirect data resulting from the feedback of other drivers' experience. For example, recent miles per gallon equivalent (MPGe) of a car can be used to predict the refueling stops on the road. Websites such as tripadvisor provide reviews from previous travelers who share similar trips. Yelp provides information on accommodation and food stops. These are examples of historical data.

Real-Time Data: This kind of data is closely related to emergency issues. Real-time traffic conditions are monitored by the GARMIN app, including examples of sudden rain or snow and unplanned road closures [5].

In [6], the authors proposed a framework to explore drivers' behavioral patterns and estimate the driving range. They have collected data from an EV in Taiwan over one year. They used the Growing Hierarchical Self-Organizing Maps (GHSOM) algorithm to categorize the driving pattern.

Besides range estimation, big data from EVs can be used by municipalities to make decisions on siting public charging stations. In this respect, the key factor is the evaluation of charging demand. Various kinds of data have been employed such as road traffic density, distribution of gas stations, and vehicle ownership. There are also several studies that use travel patterns of taxi fleets in order to derive optimal charging station siting. In [7], the authors proposed a way to site public EV charging stations using big-data-informed travel patterns of a taxi fleet. Using Beijing as a case study, they examined a large-scale data set containing 11,880 taxis for a month. Meanwhile, in [8] information from over 30,000 personal trip records in Seattle, Washington, gathered from the Puget Sound Regional Council's 2006 household travel survey, were used to determine public EV parking locations and durations. Regression methods have been used to predict parking demand variables, including total vehicle hours per zone, neighborhood and parked time per vehicle trip, and so on, as a function of site accessibility, local jobs, population densities, and trip attributes. As cities become smarter, such data will have vast volume, and mining them along with EV data will provide more opportunities for planning. A summary of the data related to EVGI explained in this section is presented in Fig 3. In the next section, we survey the data analytics tools that make use of the data described above.

Big Data Analytics Platforms and EV Integration

The data collected from EVs are various, and the volume is huge; therefore, traditional statistical ways to build a model may not work very well. Big data analytics have been useful for EV integration in a variety of ways such as optimized charging, battery management, and EV status tracking. In this section, we group the existing studies based on the platform used. The first subsection focuses on Hadoop-based techniques that allow parallel processing of EV big data. The second subsection presents a study that uses the Weka data mining tool.

Hadoop-Based Approaches

Optimized Charging: In [9], *Wei et al.* developed an optimized charging model, the multi-level feedback queue. Their model uses grid demand data, charging station data, EV battery data, user data, and data from a local distribution system. In order to handle the large amount of data from multiple sources, they proposed processing data in parallel using MapReduce over the Hadoop framework. The authors store their data using HBase, which is a NoSQL-based database used for big data storage on the cloud computing platform.

Scheme	Data source	Purpose	Platforms and tools
Optimized charging [9]	Demand information, charging station parameters, car battery data and user data	Optimizing the charging via job scheduling	Hadoop and HBase
Battery consumption prediction [10]	EVs data collected from Jeju Island testbed	Improving the accuracy of battery consumption model	Hadoop and R statistical package
Charging meter data management [11]	EVSE data collected from Jeju Island testbed	Improving the interoperability of heterogeneous chargers	Hadoop, Pig script, MySQL
EV status tracking [12]	Sensor data collected from EVs in Indianapolis	Extracting raw data and transforming into classified buckets	Hadoop and HBase
Weka-based decision support scheme [13]	New York City (NYC) demand data	Building decision support engine for power system operators	J48 and M5 algorithms from Weka platform

Table 1. Comparison of big data analytics tools used for EVs.

The presented approach is promising; however, its performance has not been evaluated with real data. Further studies are needed to show how charging can be optimized using Hadoop and big data from EVs.

Battery Consumption Prediction: In [10], the authors proposed spatio-temporal analysis of EV data using Hadoop and the R statistical package. The data were collected through a battery monitoring device that accumulated SOC records of EVs along major roads in Jeju Island, Korea. The authors used Pig scripts, which are high-level programming scripts designed for Hadoop, to filter the necessary fields from the raw data heap. Then they used the R package to conduct time series analysis and provided prediction of EVs' battery consumption.

Charging Meter Management: In [11], the authors used a similar framework to [10] to implement meter management over streaming EV data. They implemented a data analysis framework, which, after retrieving the temporal stream records from the Master Data Management Software (MDMS), used Hadoop Pig scripts to filter the raw data. Then the Hadoop Pig script results were converted to SQL commands to insert data to MySQL. At the final step, a neural network library was used to forecast future EV connections. Similar to [10], the authors used data collected from EVs in Jeju Island. The island aims for all its vehicles to be electric by 2030 as part of its becoming a carbon-fee city (http://spectrum.ieee.org/energywise/transportation/ efficiency/korean-island-plans-for-all-electric-vehicles-by-2030). The research in [10, 11] represent those efforts toward green cities and demonstrate how EV data analytics techniques can be used for this purpose.

EV Status Tracking: In [12], the authors addressed the unstructured nature of big EV data. They also used Hadoop and MapReduce. The authors first employed a preprocessing stage to remove inconsistencies and duplicates in the data to ensure optimum storage. Then the data was imported to HBase. In this study, raw EV data was extracted and transformed into classified buckets. The data were collected through the Think City project in Indianapolis. As a result, the authors observed more than 10 features for over 200 EVs. They tracked analytics on SOC, maximum vehicle speed, location, temperature, maximum voltage, and current of charging.

Weka-Based Approach

Decision Support Tool: In [13], Ranganathan *et al.* proposed using decision tree algorithms provided in the Weka data mining platform to analyze smart grid and EV data, and form a decision support tool for grid operators. The authors used NY Independent System Operators (NYISO) demand data that is publicly available. The proposed decision support system has two phases: data preprocessing and data classification. The data preprocessing stage removes irrelevant data and noise, while classification is used to reach a decision and is based on a decision tree with predefined rules. For classification, the authors use the J48 ad M5 algorithms, which are readily available in the Weka platform. Although the authors worked on large datasets from the power grid, these data sets are offline and their size is still manageable compared to the big data that will be flowing from millions of EVs. The proposed Weka-based decision support scheme needs to be further evaluated over streaming data from EVs.

A comparison of the surveyed big data analytics approaches is given in Table 1. The first three studies of this section [9–11] focus on big EV data in order to provide input for EV applications. The research in [12] studies EV data processing only and aims to structure the data for general EV applications. The final surveyed scheme, [13], is a decision support tool for power system operators. It is suitable for large datasets; however, when streaming big data from EVs are analyzed, Weka-Hadoop-based platforms may be considered. In addition, surveyed studies have focused on HBase, while there are other NoSQL databases such as Cassandra and MongoDB.

Requirement Analysis and Future Research Directions

Overview of the Landscape

The existing literature on applying big data tools on EV and smart grid data is limited. One of the major challenges is lack of publicly available real-world data. Several of the surveyed studies have worked on data collected in related projects; however, these are still not large-scale when compared to the anticipated higher penetration of EVs and increased data flow. The research in [14] generates synthesized EV data where EV characteristics have been superimposed on real

The value of big data analytic tools would be better evaluated as they transition into decisions for system operators. Nevertheless, there are many future opportunities to explore in this area.

traces of taxis in the San Francisco, California, area. This study provides a useful public dataset for EV integration; however, it is not suitable to evaluate big data methods as the size of data is still manageable with traditional data analytic techniques. In addition, in the real world, heterogeneous, unstructured EV data will be streaming from EVs in real time, and it is hard to mimic the challenges and evaluate the true performance of big data approaches with such static datasets. Electric vehicles and the new generation of autonomous vehicles can generate data on the order of several hundreds of gigabytes to thousands of gigabytes during mobility where the amount of the data depends on the variety of sensors used for autonomy (http://www.networkworld.com/article/3147892/internet/one-autonomous-car-will-use-4000-gb-of-dataday.html). During V2G/G2V operation of electric vehicles, the data generation rate and speed are expected to be lower compared to mobility, but the varied sensors collecting data from power grid, vehicle, charging station, and driver will need solutions from the big data domain.

Our article categorizes the surveyed approaches based on their platform selection. Most of the studies work on distributed Hadoop clusters and benefit from parallelization with MapReduce. They either use the data for a specific EV application such as charging, battery, or charger management, or store data in HBase to provide for future applications. There is also work on utilizing Weka, which is a well-known data mining platform.

The research on big data analytics for EVs is in its infancy. The performance of NoSQL databases such as Cassandra and MongoDB is unexplored. Furthermore, the analyzed data has not been transformed into decision making in many of the studies. The value of big data analytic tools would be better evaluated as they transition into decisions for system operators. Nevertheless, there are many future opportunities to explore in this area. In the following subsection, we focus on various applications in the EV and smart grid domain that can benefit from big data analytics.

Requirement Analysis and Applications for EVs

Analysis of grid integration of EVs includes different subsystems that operate in various time domains from microseconds to hours. These subsystems include transportation mobility and grid service requirements, EVSE and user behavior models, and onboard power electronics and battery system modeling for charging operation. In order for the utility to be spared the impact of the large number of EV connections and to utilize already available mass energy storage capacity in EVs, communication design of the EVGI framework and decision making through big data analytics play important roles.

Mobility needs of drivers can usually be captured with data tracking devices, which in turn helps to understand energy consumption profiles of drivers. Along with modeling the battery-to-wheel energy efficiency of different EV vendors, it may be possible to generate custom charging requirements for EVs. Data analytics is also important on the utility side requirements when controlling charging. The utility will eventually decide which services are needed by the EVs via analyzing its daily demand data stream. Data analytics is expected to become more important when EVs and intermittent renewable energy generators are integrated with daily demands of utility customers.

In addition to planning on the utility side, decision making tools need to account for user convenience such as each EV having satisfactory SOC at morning departure, as well as the emergency driving range that would be offered anytime. Besides the charging aspect, EVs can provide power to the grid through V2G applications. They can support ancillary services such as voltage support, reactive power compensation, active harmonic filtering, and power factor regulation, as well as load balancing, peak shaving, and renewable energy tracking. Additionally, in the case of a power outage, an EV can be used as an emergency backup source for the home, which is often called vehicle-to-home (V2H). An advanced charger can also provide vehicle-to-vehicle (V2V) charging to increase the charging availability of the EV, even when the EV is out of charge without a nearby charging station. This can be accomplished via wireless charging and by utilizing Uber-like social networking applications. All of these future applications of EVs will call for strong data analytics tools that fully integrate EV and smart grid data. As a result, data analytics techniques will need to work on more heterogenous and unstructured data from multiple sources flowing at higher speeds, while the decision making timeframe will need to be relatively smaller than for today's applications.

Future Directions

As mentioned above, the decision making timeframe will need to be reduced from minutes to seconds for integration with most smart grid applications. All of the surveyed approaches in this article consider Hadoop clusters in the cloud. However, the delay for accessing the cloud is a major concern for real-time applications. In this case, mobile edge computing using Hadoop-like parallelization can reduce the response time of decision making. This approach could parallelize computing tasks using MapReduce on the EVs. In fact, EVs have more computational power than other mobile devices such as smartphones or wearables; therefore, a group of EVs can be tasked with running data analytics in order to reduce latency.

On one hand, big data has enormous benefits in the economy, society, and the environment. On the other hand, there is concern about data security protection and privacy [15]. If data are excessively protected due to an individual's privacy, information would be significantly curtailed. As a result, much valuable information might be lost. Thus, there needs to be a balance between consumers' privacy and the benefit of sharing data. The utility-privacy trade-off has been explored in several studies, but there are open issues on how much uncertainty can be handled for EV integration to smart grid and green smart cities.

In summary, fast and effective data analytics approaches are required for real-time interaction of the EVs with the smart grid and smart city. Those approaches can benefit from the advances in mobile edge computing. However, security and privacy concerns escalate with distributed processing of EV data by other EVs. The nexus of processing capacity, delay, security, and privacy is an open issue that has yet to be addressed in the domain of big data analytics for EVs.

Conclusion

In this article, we review the state-of-the-art data analytics tools for electric vehicle integration to smart grids and thereon to green smart cities as well as the big data that are generated by cars and drivers. We first provide an overview of smart grid and electric vehicle integration. We present the challenges of EV integration, and discuss how these challenges can be addressed by data analytics. Then we discuss the sources of big data including EVs, drivers, EV batteries, chargers, and EVSEs. We provide a comprehensive survey on data analytics tools that are used in this domain. We conclude the article with a summary, a requirement analysis for data analytics tools for EV related applications, and finally, with a discussion of future directions.

Acknowledgment

Dr. Erol-Kantarci's research is supported by the U.S. National Science Foundation (NSF) under Grant Number CNS-1647135 and the Natural Sciences and Engineering Research Council of Canada (NSERC) under RGPIN-2017-03995.

References

[1] R. Northfield, "Greening the Smart City," *IEEE Engineering & Technology*, vol. 11, no. 5, June 2016, pp. 38–41.

[2] J. Wu *et al.*, "Big Data Meet Green Challenges: Big Data Toward Green Applications," *IEEE Systems J.*, vol. 10, no. 3, Sept. 2016, pp. 888–900.

[3] M. C. Kisacikoglu, F. Erden, and N. Erdogan, "Distributed Control of PEV Charging Based on Energy Demand Forecast," *IEEE Trans. Industrial Informatics*, vol. PP, no. 99, 2017, pp. 1–1.

[4] K. Wang *et al.*, "A Survey on Energy Internet: Architecture, Approach, and Emerging Technologies," *IEEE Systems J.*, vol. PP, no. 99, 2017, pp. 1–1.

[5] H. Rahimi-Eichi and M.-Y. Chow, "Big-Data Framework for Electric Vehicle Range Estimation," *Proc. Annual Conf. IEEE Industrial Electronics Society*, 2014, pp. 5628–34.

[6] C.-H. Lee and C.-H. Wu, "Collecting and Mining Big Data for Electric Vehicle Systems Using Battery Modeling Data," *Proc. 12th Int'l. Conf. Info. Technology — New Generations*, 2015, pp. 626–31.

[7] H. Cai *et al.*, "Siting Public Electric Vehicle Charging Stations in Beijing Using Big-Data Informed Travel Patterns of the Taxi Fleet," *Transportation Research Part D: Transport and Environment*, vol. 33, 2014, pp. 39–46.

[8] T. Chen, K. Kockelman, and M. Khan, "Locating Electric Vehicle Charging Stations: Parking-Based Assignment Method for Seattle, Washington," *Transportation Research Record: J. Transportation Research Board*, vol. 2385, 2014, pp. 28–36.

[9] W. Wei *et al.*, "Multi-Level Feeder Queue Optimization Charging Model of Electric Vehicle and its Implementation of MR Algorithm," *Int'l. J. u- and e-Service, Science and Technology*, vol. 9, no. 3, 2016, pp. 199–208.

[10] J. Lee *et al.*, "Spatio-Temporal Analysis of State-of-Charge Streams for Electric Vehicles," *Proc. 14th ACM Int'l. Conf. Info. Processing in Sensor Networks*, 2015, pp. 368–69.

[11] J. Lee and G.-L. Park, "Electric Vehicle Charger Management System for Interoperable Charging Facilities," *J. Teknologi*, vol. 78, no. 5–8, 2016.

[12] V. Bolly, J. Springer, and E. Dietz, "Using Open Source NOSQL Technologies in Designing Systems for Delivering Electric Vehicle Data Analytics," *Proc. 121st ASEE Annual Conf.*, Indianapolis, IN, 2014, pp. 24.1339.1–339.9.

[13] P. Ranganathan and K. Nygard, "Smart Grid Data Analytics for Decision Support," *Proc. IEEE Electrical Power and Energy Conf.*, 2011, pp. 315–21.

[14] H. Akhavan-Hejazi, H. Mohsenian-Rad, and A. Nejat, "Developing a Test Data Set for Electric Vehicle Applications in Smart Grid Research," *Proc. IEEE VTC-Fall 2014*, Vancouver, BC, Canada, 2014.

[15] P. D. Diamantoulakis, V. M. Kapinas, and G. K. Karagiannidis, "Big Data Analytics for Dynamic Energy Management in Smart Grids," *Big Data Research*, vol. 2, no. 3, Sept. 2015, pp. 94–101.

Biographies

Boyang Li obtained his B.E degree from Tianjin University in 2010 and his M.S. degree from Clarkson University in 2016, respectively. He is currently working toward a Ph.D. degree in computer science at Illinois Institute of Technology. His former research interests lie in big data and anomaly detection. His current research interests include HPC data analytics and job scheduling in large-scale systems.

Mithat C. Kisacikoglu received his Ph.D. degree from the University of Tennessee in 2013 in electrical engineering. He joined Hacettepe University, Ankara, Turkey, as an assistant professor in 2014. He then worked at the National Renewable Energy Laboratory, Golden, Colorado, as a research engineer between 2015 and 2016. He is currently an assistant professor in the Electrical and Computer Engineering Department at the University of Alabama, Tuscaloosa. His research interests include electric vehicles, grid integration, and power electronics.

Chen Liu currently is an assistant professor in the Department of Electrical and Computer Engineering at Clarkson University, Potsdam, New York. He received his M.S. in electrical engineering in 2002 from the University of California, Riverside, and his Ph.D. in Electrical and Computer Engineering in 2008 from University of California, Irvine, respectively. His research interests include computer architecture, the interaction between system software and microarchitecture, power-aware computing, and hardware acceleration techniques.

Navjot Singh is pursuing his M.A.Sc. degree in electrical and computer engineering at the University of Ottawa. He received his B.Tech. degree in computer science and engineering from Punjab Technical University, India, in 2016. His research interests include big data and wireless networks.

Melike Erol-Kantarci (melike.erolkantarci@uottawa.ca) is an assistant professor at the School of Electrical Engineering and Computer Science, University of Ottawa, and a courtesy assistant professor at Clarkson University. She received her Ph.D. and M.Sc. degrees in computer engineering from Istanbul Technical University in 2009 and 2004, respectively. Her main research interests are 5G and beyond wireless networks, smart grid, electric vehicles, and the Internet of Things.

Security and privacy concerns escalate with distributed processing of EV data by other EVs. The nexus of processing capacity, delay, security and privacy are open issues that are yet to be addressed in the domain of big data analytics for EVs.