The Electricity Journal 32 (2019) 106674

Contents lists available at ScienceDirect

The Electricity Journal

journal homepage: www.elsevier.com/locate/tej

Algorithmic analysis of intelligent electricity meter data for reduction of energy consumption and carbon emission

Archana Chaudhari*, Preeti Mulay

Symbiosis Institute of Technology (SIT), Symbiosis International (Deemed University), Pune, India

ARTICLE INFO

Keywords:
Microsoft azure
Gradational clustering algorithm
Intelligent meter
Data analytics

ABSTRACT

Intelligent Electricity Meters (IEMs) generate a considerable amount of household electricity usage data incrementally. Obviously, for the clustering task, it is better to incrementally update the new clustering results based on the old data rather than to recluster all the data from scratch. The gradational clustering is an essential way to accommodate the influx of new data seamlessly for accurate analysis. However, given the volume of IEM data and the number of data types involved makes the gradational clustering highly complex. Microsoft Azure provides the processing power necessary to handle gradational clustering analytics. The paper aim is to develop a Distributed Log-likelihood Based Gradational Clustering Algorithm on Microsoft Azure for analysis of IEM data. This research uses the real dataset of Irish households collected by IEMs and related socioeconomic data, including the geographic information, demographic data. It is visible from the study that algorithmic analysis helps the household customers to monitor and improvise electricity consumption patterns, Utility providers to reduce power outage and avoid capital expenses of building new plants. This research will be extremely useful for maintaining the environment by reducing pollution via carbon production by power plants.

1. Introduction

According to the World Resources Institute (WRI), (Chakrabarty, August 08, 2018), energy-related emissions make up more than two-thirds of India's overall emissions and represent more than three times the next largest source (the industry sector). Of the country's energy sector emissions, 77 percent are from electricity generation, making this a key target for reductions to meet India's climate, or Nationally Determined Contribution (NDC), and commitments(India, 2018). Although India's emissions are still comparatively low on a per-capita basis. The very size of the nation's population and the scope for them to increase the country's emissions of global concern. India's emissions from the energy sector are the fourth largest in the world (Navroz K. Dubash, August 22, 2018) (behind China, the United States, and the European Union), and rising.

One of the challenges faced by India's electricity sector is the capacity of the grid to accommodate renewable energy as the country invests in renewable energy to meet national targets. It will have a potential surplus, rather than a shortfall, in electricity, and yet it faces technical constraints in using that electricity. According to analyst firm Brookings, (Tongia, June 15, 2018) to avoid curtailing renewable energy generation, the country needs "a stronger grid, cheap storage, and the ability to shift load to match supply conditions."

In parallel with these challenges, in India today, utility companies still rely on manual electricity meters to track consumption at residential and industrial locations. Local utility companies send employees to read these meters each month and then generate monthly electricity bills. This system does not provide the utilities with insight into electricity consumption patterns, for example, based on time-of-day usage, nor does it enable consumers to understand their electricity consumption as part of the larger picture of electricity availability.

1.1. Motivation

The Intelligent Electricity Meters (IEM) combined with machine learning analytics; represent a key opportunity to address the above issues. IEM effectually communicate with utility companies for monitoring and management of electricity usage as well as with customers for observing electricity consumption in time intervals of one hour or less. In addition to saving the laborious process of manual meter readings, IEM enables utilities to monitor energy usage in real-time. Furthermore, they generate a considerable volume of incremental data. However, on an influx of new data, traditional clustering task re-cluster all of the data from scratch. The gradational clustering method is an essential way to solve the problem of clustering with dynamic data. Given the volume of IEM data and the number of data types involved,

* Corresponding author.
E-mail address: archana.chaudhari@sitpune.edu.in (A. Chaudhari).

https://doi.org/10.1016/j.tej.2019.106674

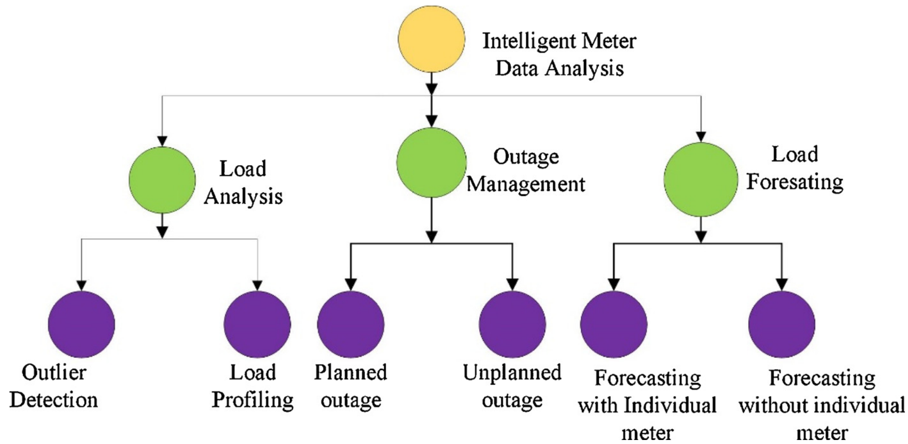

Fig. 1. Application of intelligent meter data analysis (Wang et al., 2018).

an incremental clustering method is highly complex. Microsoft Azure will provide the processing power necessary to handle incremental clustering analytics.

The objective of this paper is to perform IEM data analysis by Distributed Log Likelihood-based Gradational Clustering Algorithm (DL2GA) using Microsoft Azure. The proposed algorithm finds the hidden patterns from the extensive IEM data for better balancing between the reduction of electricity consumption and carbon emission.

The unique research contribution in this paper is as follows:

- The proposed Distributed Log Likelihood-based Gradational Clustering Algorithm accommodates the influx of new data seamlessly for learning, prediction, decision-making, etc.
- The Microsoft Azure-based Model provides superior computational resources for processing incremental IEM data (through virtual machines).
- The experimental evaluation and validation of the proposed methodology.

The rest of the paper has organized as follow; Section 2 addresses the problem of relevance and related literature. Section 3 describes the outline of the proposed system, subsequently proposed algorithm. The experimental result and implication are in Section 4. Section 5 provides a conclusion and an outlook on further application scenarios.

2. Literature survey

Various statistical, time series analysis, and mathematical modeling methods have been used to analyze residential electricity consumption data (ECD) to extract consumption patterns (Cominola et al., 2018). The unsupervised machine learning techniques specifically clustering, has used most widely to identify the pattern of electricity consumptions in the literature. According to the literature, the main clustering methods applied in the field of load profile are partition-based, hierarchical-based, grid-based, density-based, and model-based. Boudet et al. (2016) developed a clustering method to analyze the pattern of electricity consumers which support targeted demand-side management and efficient operation of the intelligent grid. The research finding from the paper (Ryu et al., 2016) used by power companies in identifying suitable customers for demand response strategies and improvement of load profile modeling. Numerous national and international project (DR-BOB, 2018; EDREAM, 2018) project are working on pattern recognition of electricity consumption to support demand-side management. The accuracy of the clustering algorithm influenced by the multidimensionality of electricity consumption data. Therefore, the various pre-clustering technique investigated in load profile clustering such as Wavelet Packet Decomposition (Khan et al., 2014), Discrete Fourier Transform (Zhong and Tam, 2015), piecewise linear regression and different levels of polynomial regression functions (Eichinger et al., 2015), Principle Component Analysis (Mehra et al., 2013). The similarity or distance measure used for the clustering also affect the pattern recognition result. To remove the dependency of the similarity and distance measure Mulay and Kulkarni (2013) developed a parameter-free Closeness Factor Based Algorithm(CFBA) to handle the influx of new data. It is further extended by TBCA (Threshold-based Clustering Algorithm) to analyze diabetic patients clinical parameters (Mulay, 2016) and diabetic Mellitus (Mulay et al., 2018). TBCA uses the threshold value 0 to 1 only. To further enhance the threshold range from -1 to +1 Correlation-Based Incremental Clustering Algorithm (CBICA) (Mulay and Shinde, 2017, Shinde and Mulay, 2017), a new variant of CFBA has taken shape. CBICA uses Pearson's coefficient of correlation similarity measures. The bibliometric analysis of incremental clustering (Chaudhari and Mulay, 2019, Chaudhari et al., 2019) shows the existing algorithms are static and require high computation time for pattern recognition. The primary applications of intelligent meter data analytics are classified into load analysis, demand response, load forecasting, customer characterization, Identifying Irregularities, and so on so forth, as shown in Fig. 1.

As the literature review shows, given the volume of data and the number of data types involved makes the intelligent meter data analytics are highly complex. To combine the power sources, plan the distribution of the electricity, computing load profiles, etc. requires dynamic systems that are distributed. Incremental learning of intelligent electricity meter data is not made explicit in literature, which is blurring their conceptual contours and constrains the efficacy of using the approaches in research and practice. In this paper, a distributed clustering algorithm applied to intelligent electricity meter consumer using Microsoft Azure for learning about the relationship between socioeconomic characteristic and electricity consumption pattern of the residential customer.

3. Proposed methodology

IEM collects electricity consumption as well as related socio-economic data, including the weather conditions at the meter installation sites, customer data, geographic information, demographic data, and sustainable data. However, given the volume of data and the number of data types involved makes gradational clustering highly complex. Microsoft Azure has provided the processing power to handle data analytics. The proposed methodology Distributed Log Likelihood-based Gradational Clustering Algorithm (DL2GA) using Microsoft Azure for Analysis of Intelligent Meters Data is presented in Fig. 2.

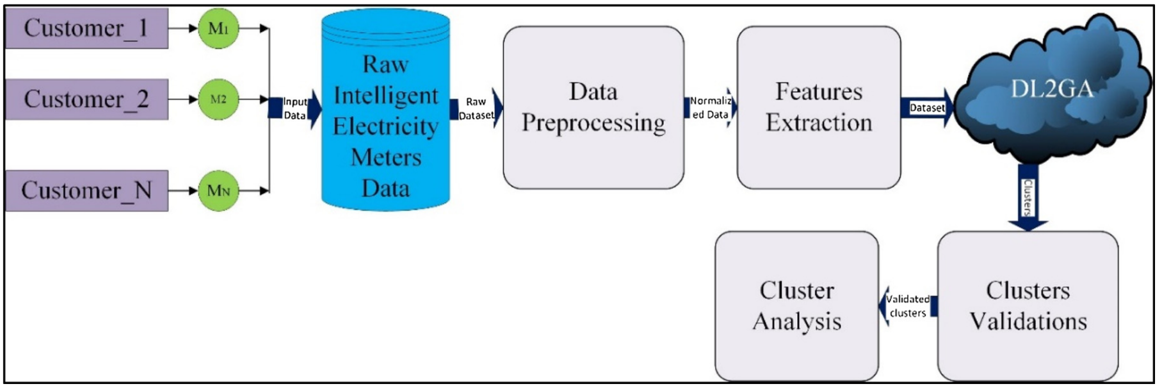

Fig. 2. Framework of the proposed approach.

3.1. Proposed system: DL2GA processing model on Microsoft Azure

Distributed Log Likelihood-based Gradational Clustering Algorithm (DL2GA) runs on the Microsoft Azure Cloud platform to process the distributed IEM data, accommodates the influx of new data, and form clusters dynamically. DL2GA uses a log-likelihood for pattern matching rather than distance measures (qualitative or quantitative). Due to that, DL2GA is ordered independence, parameter-free algorithm.

The DL2GA is working in two phases.

- Phase 1: Basic clusters are formed based on the available datasets.
- Phase 2: On arrival of an influx of new data recomputed the existing clusters or formed a new cluster automatically.

DL2GA distributed across the virtual machines of Microsoft Azure. The aim is to use the Azure services for faster processing and handle the data analysis complexity of dynamically growing IEM data. There are two ways to deliver an application on Microsoft Azure Cloud either by deploying pre-configured Virtual Machines (VM) (IaaS) or as fully functional SaaS solutions. Internally an application is composed of one or more web and/or worker roles, which resides in and execute the application logic on VMs. Web role providing a Graphical User Interface (GUI) in the form of a website, and worker role implementing the application logic. In other words, web role VMs and worker role VMs are Windows Servers with IIS installed and without IIS install respectively. Fig. 3 indicates the DL2GA processing model on Microsoft Azure. The proposed DL2GA processing model has implemented using the following steps:

- Collection of distributed intelligent meter data using IoT Hub service of Microsoft Azure
- Submit the dataset to the web role for processing request
- Every request causes the creation of a prediction job
- These datasets are serialized to the message by using azure services & then consume by successive worker role instances
- Worker role instances process & execute DL2GA clustering algorithm by using the virtual machine stream analytics, and machine learning services of Microsoft Azure
- Finally, save the result in terms of electricity consumption pattern (clusters) in the Azure file storage service
- Worker role confirms the completion of the task to the web role through output queue
- Electricity consumption patterns are retrieved from the azure storage

Algorithm: Manager Role DL2GA
Input: : Intelligent meters datasets from consumer
Output: A series of the cluster, which represents the consumption patterns
Method:

1. Apply pre-clustering techniques to normalize, remove noise from the dataset.
2. Phase I: Pre-Processed data has processed by computing the sum of attributes, log-likelihood, error, and closeness. Using these computed values, the basic clusters are formed
3. Phase II: On the influx of new data, automatically update the existing clusters or to form the new clusters. Update the cluster database. The clusters obtained from new data are compared with the basic clusters to check whether the same pattern is followed or not.

These roles are scaled out parallel by adding more computational cores (from 1 to 64) and perpendicular by using computational cores of different sizes (ranging from Extra Small (ES) to Extra Large (EL)) during the NF calculation process (Table 1) (Joshi and Mulay, 2018). Web role is responsible for user interaction and prepares workload; Worker role is a set of instances of Manager role (1 to 64) who perform parallel NF calculation

3.1.1. Advantages of DL2GA

- DL2GA does not require prior knowledge about the number of clusters.
- DL2GA learn from the new labeled or unlabelled instances of data, without discarding the previously acquired knowledge. This ability of DL2GA uses for knowledge refinement and knowledge building.

4. Experimental result

The main challenge to do research on Intelligent Electricity Meter data analysis in India is the dataset because IEM is being deployed in India on the basis of pilot projects by various utilities (Barua and Goswami, 2017). IEM data is not available in India. Due to this consideration, the authors used Ireland data for demo purpose.

4.1. Description of dataset

Dataset used in this research concern the electricity consumption of 6445 buildings recorded using intelligent meters installed by the Irish Commission for Energy Regulation(CER), Ireland (Archive, 2012) from 2009 to 2010. CER performed the IEM customer behavior trails for identifying the pattern of residential electricity consumption. IEM has distributed across the spatial locations of Ireland residents. The dataset set contains two types of information: electricity consumed at each building and additional building characteristics such as dwelling type, dwelling age, and to name a few. The IEM dataset comprised of 4232 m electricity consumption data of residential customer with a granularity of 30 min (one day 48 reading). After pre-processing (removal of noise and missing value) got 2995 m of residential electricity consumption data. Total electricity consumption dataset contains 2995*48* 365 instances. The overall electricity consumption pattern of the residential customers over the one year shown in Fig. 4

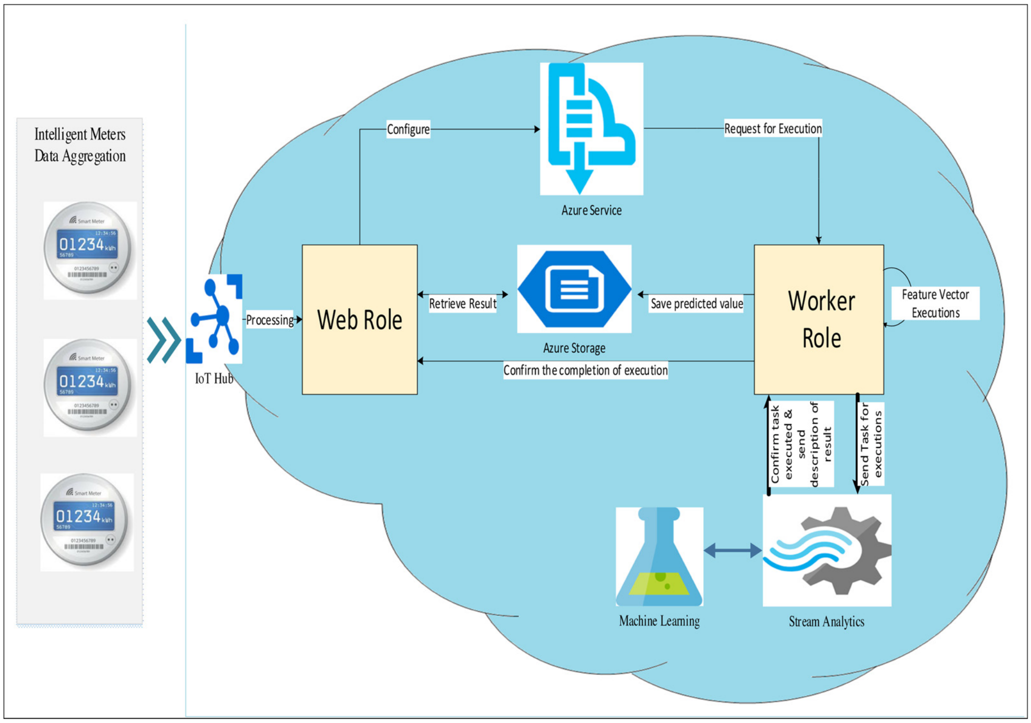

Fig. 3. DL2GA Processing Model (Microsoft Azure, 2018).

Table 1
Virtual machine sizes and their features for web and worker role instances (Mrozek et al., 2015; Azure, 2018).

VMs Type	VCPUs	RAM(GB)	Data Disks	Temporary Storage (GB)	Cost/Month
Extra Small (XS)	Shared Core	0.5	2	1	Unavailable
Small (S)	1	1.75	2	4	₹688.46
Medium (M)	2	7	8	14	₹12,392.25
Large (L)	4	16	8	32	₹18,490.03
Extra Large (XL)	8	32	16	64	₹36,980.06

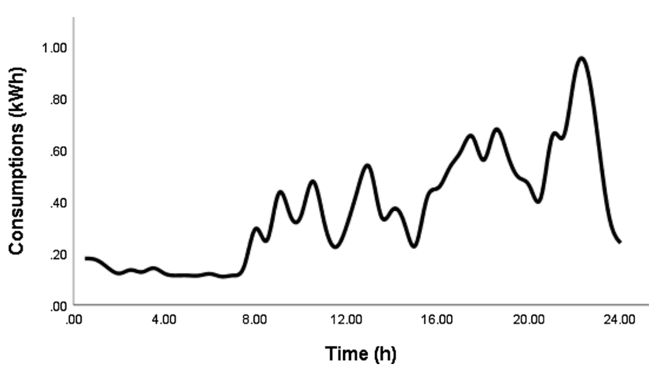

Fig. 4. Average electricity consumption of residential customers for the year 2010.

4.2. IEM data analysis using DL2GA system

In order to understand the consumption pattern of distributed IEM data DL2GA system used. As mention in Section 3.1, the DL2GA system uses log-likelihood value for automatically decide cluster member and a number of clusters. Fig. 5 depicts the clusters obtained from the DL2GA. Fig. 5a presents the implementation of the DL2GA system, which results in eight basic clusters formed. The influx of new data fit into the existing cluster can be seen in Fig. 5b. We observed from Fig. 5c and d that on the arrival of the influx of new data existing clusters re-computed. Table 2 depicts the python library used for the implementation of DL2GA.

4.3. Analysis of clustering result

Fig. 6 depicts the clusters obtained from Microsoft Azure-based DL2GA. There are eight clusters formed after the execution of DL2GA. Consumers have grouped based on electricity usage.

- Cluster1 constitutes 69.10% of the population, User contains by the cluster1 are ordinary consumers that consume average electricity. The peaks of electricity may appear at about 11.30 a.m.. Cluster1 shows that a majority of the residents leave home during the day. Also, observed their lunch and evening times.
- Cluster2 accounted for 12.10% of the population, and users in cluster1 are high demand consumers. Interpretation of cluster2, electricity consumption is lower during the morning, then stable consumption during the afternoon and peak at about 4.00 pm.
- Cluster3 constituted 15.00% of the population. The behavior of cluster3 is quite similar to the cluster2 in defiance of the fact that it has higher electricity consumption.
- Cluster4 accounted for 1.5% of the populations. It contains the highest demand consumer. The electricity consumption of the user of this cluster is not stable. Users consume as much as electricity as they need without considering the price. The peaks of electricity appear at about 7.00 pm.
- Cluster5 constitutes for 0.4% of the populations. Users in this cluster5 are low demand consumer that consumes the least electricity, daily consumption is very stable, and peak consumption point is 1.29 kW h per day.
- Cluster6 accounted for 0.6% of the populations. This cluster contains the consumer who consumes very low electricity. The patterns of consumption are similar to cluster5.
- Cluster7 accounted for 1% of the populations. The patterns of consumption show that peak occurs during the night. It indicates that

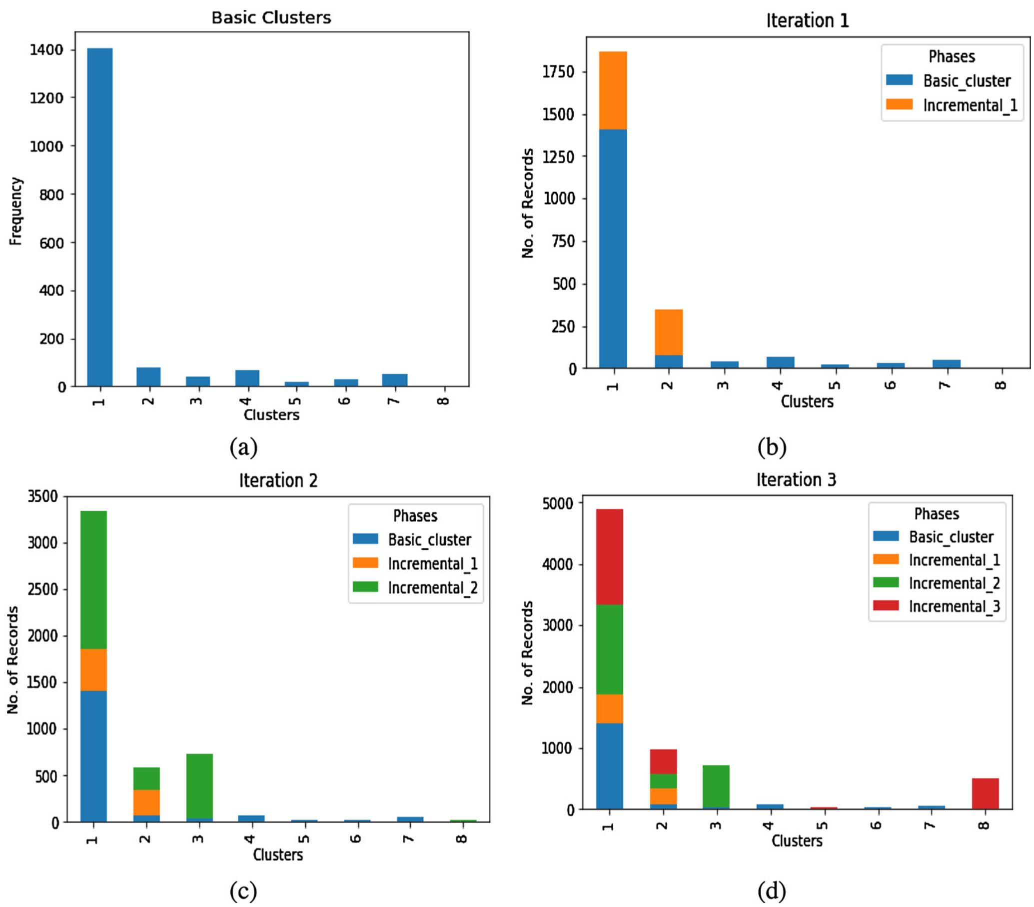

Fig. 5. DL2GA clustering result.

Table 2
Shows the python library used for the implementation of the proposed model.

Task	Python Library
Pre-processing	from sklearn.preprocessing import MinMaxScaler from sklearn.model_selection import train_test_split
DL2GA	import pandas as pd import numpy as np import math as np import azureml.dataprep as dprep from azure.storage.blob import BlockBlobService from io import StringIO
Visualization	import matplotlib.pyplot as plt
Validations	from sklearn.metrics import davies_bouldin_score from sklearn.metrics import silhouette_score

residents use electric heating most.

- Cluster8 constitutes 0.3% of the populations. The peaks of electricity may appear at about 10 a.m. after that daily consumption is very stable.

The clustering result (cluster1 to 8) and details of the pattern of consumption summarized in Fig. 7 for better illustration of the consumption behavior of different clusters. Cluster 5 consist of consumers whose electricity consumption is high, and cluster 6 consists of low demand consumers that consume stable and very less electricity.

The clustering results given by DL2GA algorithm indicates that there is a close relationship between residents' daily lives and the pattern of electricity consumption. A measurable valuation of the above observation can be made by validating the clustering results with the socioeconomic data. Fig. 8 shows the clustering results with four socioeconomic characteristics. Fig. 8a depicts the clusters according to social classes' low-class, middle class, and higher class observed in each cluster. Cluster4 contains a more significant number of high-class customers (spending limit, electricity consumption, no. of appliances, family size, etc. is more) as compared to others. Fig. 8b shows the usage of new technologies such as the internet/wi-fi routers, increased usage of handheld devices, etc. also depends on the age of residents (less the age more electricity consumption). Cluster2 internet usage is more because it contains higher young age group residents. Fig. 8c constitutes a number of appliances at household residents of each cluster. It is noted that less number of appliances in residents of cluster6. It is remarkable that the majority of consumers in all clusters use non-electric means of heating, as shown in Fig. 8d.

The Microsoft Azure-based DL2GA algorithm may not only decrease electricity consumption but also reduce the carbon emission in the meantime. For the reduction of electricity consumption and carbon emission requires a comprehensive understanding of electricity consumption pattern and how they relate to socioeconomic characteristics. The DL2GA ascertained the accurate electricity consumption patterns reflecting the lifestyles of the resident(s) could be deduced from the available socioeconomic data of the household, as shown in Fig. 8.

5. DL2GA validations

The DL2GA algorithm successfully implemented on virtual machines of Microsoft Azure. DL2GA runs on five virtual machines (compute units) out of that one unit of small size allocated for web role and remaining used for worker role in testing the scalability of the algorithm. The scalability can be determined based on execution time measurements.

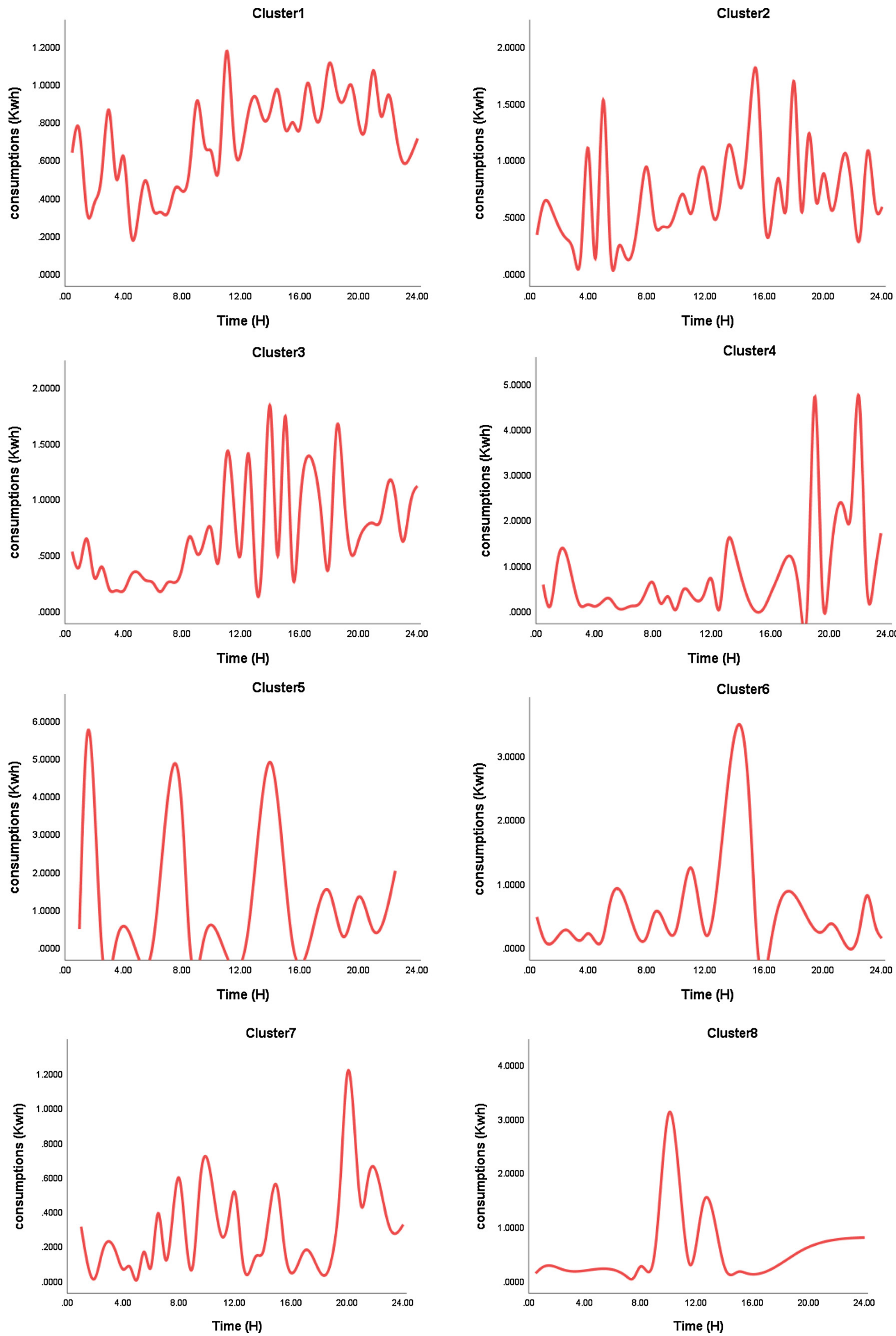

Fig. 6. Electricity consumption pattern of different cluster.

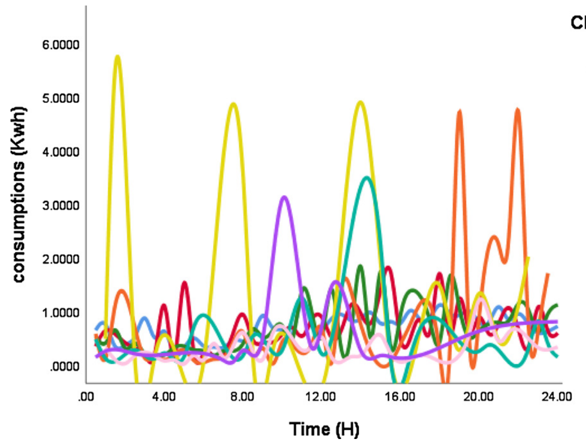

Fig. 7. Clustering result generated using the proposed algorithm.

5.1. Scalability

There are two types of scalability: horizontal scalability (number of worker role instances) and vertical scalability (size of worker role instances). All test performed on the Microsoft Azure cloud. Fig. 9 shows the total execution time required by the instances of the worker role. Fig. 10 depicts the speed of each instance of the worker role. It proved that total execution time for pattern recognition has exponentially reduced as a number of computing unit increases.

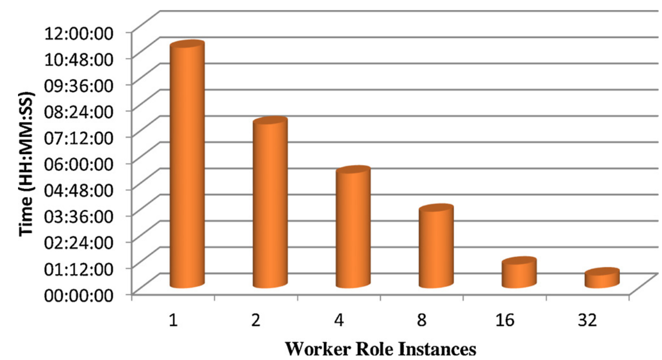

Fig. 9. Result of horizontal scalability.

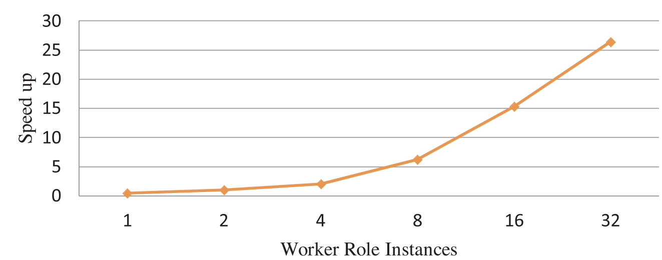

Fig. 10. Processing Speed of instances of worker role.

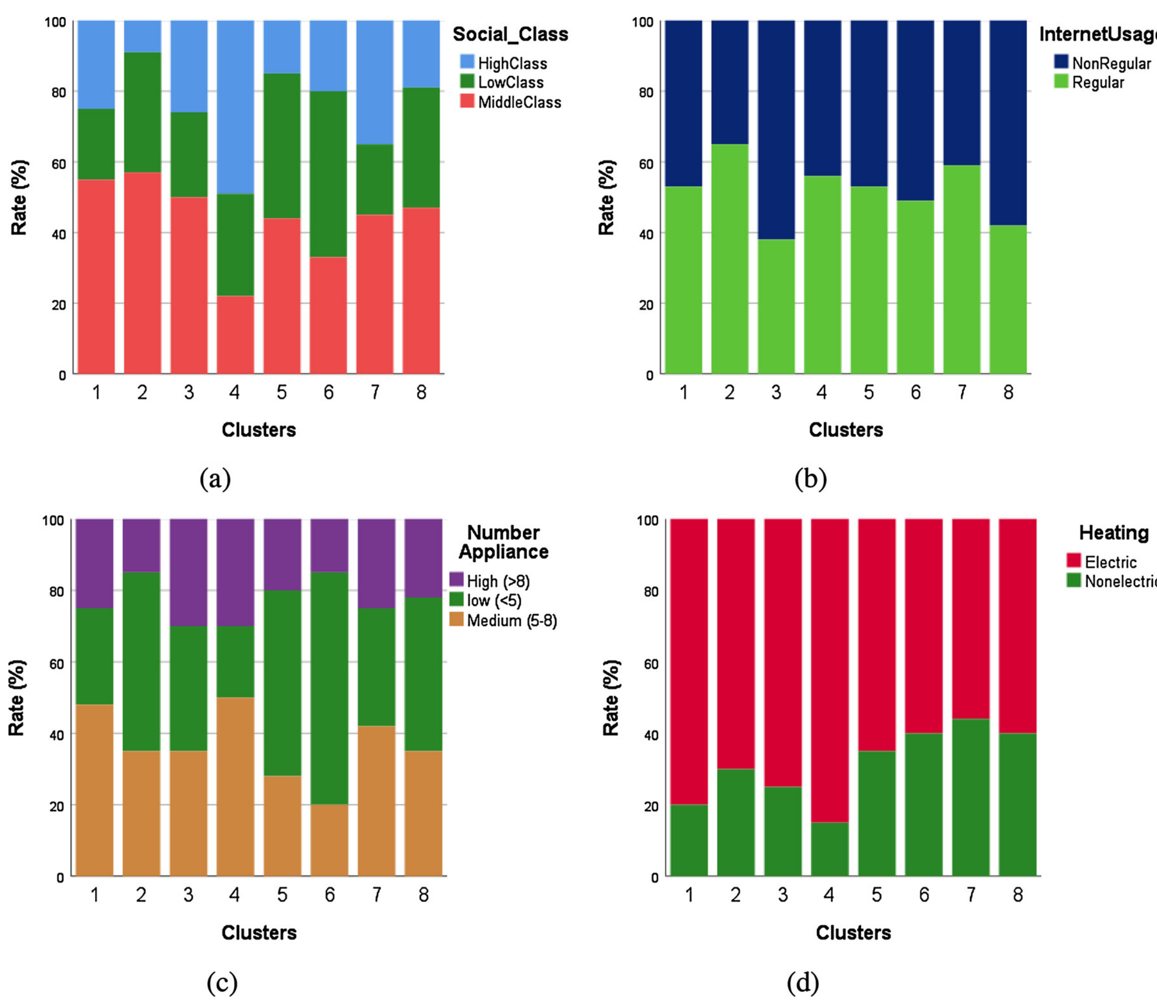

Fig. 8. Clustering results with four-socio-economic characteristics.

Fig. 11. Result of vertical scalability.

Table 3
Showing algorithmic and cluster-oriented comparative details.

Characteristics	DL2GA	DBSCAN
Cluster First Approach	√	
Converge Guaranteed	√	√
Data structure independent cluster analysis	√	√
Outlier Formations	√	√
Order Independence	√	
User input required		√
Incremental learning	√	
Time/day specific learning	√	
Seasonal electricity consumption learning	√	
Planned and unplanned outage computation	√	
Personification	√	
Cluster & attribute ranking during iterations	√	

Fig. 11 shows that execution of a single task on the dataset was performed faster on S-size instance and slower on XL-size instance. However, S-size instance can execute only one task, and XL-size instance can execute eight prediction task at the same time.

5.2. Cluster evaluations

The cluster generated by the proposed algorithm compared with the predefined clustering algorithm for the validation. Table 3 compares the DL2GA and DBSCAN algorithm with different characteristics. It is worthy to note that the DL2GA converges and may achieve better converge point compared to that with concurrent updates. DL2GA does not require any parameter for clustering.

6. Conclusions & outlook of the study

Intelligent Electricity Meters (IEMs) form the key infrastructure important for the growth of smart grids. The use of IEMs incurs benefits to the individual in various aspects such as environmental, social, and commercial, etc. IEMs collect consumption as well as related socio-economic data, including the weather conditions at the meter installation sites, geographic information, demographic data, etc. dynamically. However, given the volume of data and the number of data types involved makes data analysis highly complex, which requires dynamic systems. Microsoft Azure provides the processing power necessary to handle data analytics in a pay-as-you-go model. The developed DL2GA combines intelligent meters, cloud storage, and algorithmic data analysis to unlock efficiency in the national energy grid. Microsoft Azure-based DL2GA platform is scalable, can do manual or auto load balancing if one of the instances of Manager Role and Worker role goes down, concurrency of operations is achievable. It proved that total execution time for patterns recognition has exponentially reduced as a number of computing unit increases. The DL2GA not only assist decrease in electricity consumption but also benefit the reduction of the carbon emission. For the reduction of electricity consumption and carbon emission requires accurate hidden patterns of data and related socioeconomic characteristics of the people. The findings of this study will help:

- Household customers to monitor and improvise electricity consumption patterns.
- Utility providers to reduce power outage and avoid capital expenses of building new plants.
- Environment by reducing pollution via carbon production by power plants.

The future work of this research is to combine the proposed distributed system with some other machine learning technique (e.g., K-Nearest Neighbour, Naïve Bayes). KNN or NB can apply to analyze seasonality, day-night, and commercial/residential details, etc. by considering more characteristics of intelligent meter data. Intelligent meters data, including gas, water, solar, floating-solar, etc. can be analyzed using DL2GA in upcoming phases of research.

Acknowledgments

This research supported by "Microsoft" through AI for Earth project. Also, we would like to thank Sakal India Foundation for sponsoring this research. The authors acknowledge the Commission for Energy Regulation (CER) and the Irish Social Science Data Archive (ISSDA) for making the smart meter datasets of residential and SME customers available for research.

References

Chaudhari, Archana, Joshi, RahulRaghvendra, Mulay, Preeti, Kotecha, Ketan, Kulkarni, P., 2019. Bibliometric survey on incremental clustering algorithms. Library Philos. Pract. 2762 (e-journal).
Archive, I.S.S.D., 2012. Data From the Commission for Energy Regulation (CER)- Smart Metering Project. [Online]. Available: http://www.ucd.ie/issda/data/commissionforenergyregulationcer/ April 20, 2018]. .
Microsoft Azure, 2018. Azure Product, SDKs/Tools, Architeture. [Online]. Available: https://docs.microsoft.com/en-in/azure/ [Accessed December 14 2018]. .
BARUA, A.M., Goswami, P.K., 2017. Smart metering deployment scenarios in India and implementation using RF mesh network. In: 2017 IEEE International Conference on Smart Grid and Smart Cities (ICSGSC). 23-26 July 2017. pp. 243–247.
Boudet, H.S., Flora, J.A., Armel, K.C., 2016. Clustering household energy-saving behaviours by behavioural attribute. Energy Policy 92, 444–454.
Chakrabarty, S., 2018. By the Numbers: New Emissions Data Quantify India's Climate Challenge. August 08 [Online]. Available from:. World Resources Institute. https://www.wri.org/blog/2018/08/numbers-new-emissions-data-quantify-indias-climate-challenge.
Chaudhari, A., Mulay, P., 2019. A bibliometric survey on incremental clustering algorithm for electricity smart meter data analysis. Iran J. Comput. Sci.
Cominola, A., Spang, E.S., Giuliani, M., Castelletti, A., Lund, J.R., Loge, F.J., 2018. Segmentation analysis of residential water-electricity demand for customized demand-side management programs. J. Clean. Prod. 172, 1607–1619.
DR-BOB, 2018. The DR-BOB Collaborative Project. [Online]. Available: https://www.drbob.eu/ [Accessed 25 February 2019]. .
Mulay, Preeti, Shinde, K., 2017. CBICA: correlation based incremental clustering algorithm. IJCTA 10, 575–582.
EDREAM, 2018. The eDREAM. Project [Online]. Available: http://edream-h2020.eu/ [Accessed 25 February 2019]. .
Eichinger, F., Efros, P., Karnouskos, S., Böhm, K., 2015. A time-series compression technique and its application to the smart grid. VLDB J. 24, 193–2018.
INDIA, 2018. Climate Watch. Washington. [Online]. Available:. https://www.wri.org/our-work/project/climate-watch.
Shinde, Kaustubh, Mulay, P., 2017. CBICA: correlation based incremental clustering algorithm, a new approach. 2017 2nd International Conference for Convergence in Technology (I2CT).
Khan, J., Bhuiyan, S., Murphy, G., Williams, J., 2014. PMU data analysis in smart grid using WPD. In: IECON 2014 - 40th Annual Conference of the IEEE Industrial Electronics Society. 29 Oct.-1 Nov. 2014. pp. 2438–2444.
Mehra, R., Bhatt, N., Kazi, F., Singh, N.M., 2013. Analysis of PCA based compression and denoising of smart grid data under normal and fault conditions. In: 2013 IEEE International Conference on Electronics, Computing and Communication Technologies. 17-19 Jan. 2013. pp. 1–6.
Mrozek, D., Gosk, P., Małysiak-Mrozek, B., 2015. Scaling ab initio predictions of 3D protein structures in microsoft azure cloud. J. Grid Comput. 13, 561–585.
Mulay, P., 2016. Threshold computation to discover cluster structure, a new approach. Int. J. Electr. Comput. Eng. 6, 275–282 published by Elsevier.

Mulay, P., Joshi, R., Chaudhari, A., 2018. Mapping of six sigma to threshold based incremental clustering algorithm. IEEE Punecon, 30 Nov.-2 Dec. 2018 2018 Pune, India, India. IEEE 1–8.

Navroz K. Dubash, A.B., 2018. Guest post: India's emissions will double at most by 2030. Carbon Brief [Online]. August 22, Available from: https://www.carbonbrief.org/guest-post-indias-emissions-will-double-at-most-by-2030.

Mulay, Preeti, Kulkarni, P., 2013. Evolving Systems using incremental clustering approach. Evol. Syst. 4, 70–85.

Joshi, RahulRaghvendra, Mulay, D.P., 2018. Cloud4ICFBA: microsoft azure based distributed incremental closeness factor based clustering algorithm (DICFBA) for analysis of diabetes mellitus. J. Adv. Res. Dyn. Control Syst. 10, 2360–2368.

Ryu, S., Kim, H., Oh, D., No, J., 2016. Customer load pattern analysis using clustering techniques. KEPCO J. Electric Power Energy 2, 61–69.

Tongia, R., 2018. Embarrassment of riches? The rise of RE in India and steps to manage "surplus" electricity. Brookings [Online]. June 15, Available from: https://www.brookings.edu/blog/planetpolicy/2018/06/15/embarrassment-of-riches-the-rise-of-re-in-india-and-steps-to-manage-surplus-electricity/.

Wang, Y., Chen, Q., Hong, T., Kang, C., 2018. Review of smart meter data analytics: applications, methodologies, and challenges. IEEE Trans. Smart Grid 10, 3125–3148.

Zhong, S., Tam, K., 2015. Hierarchical classification of load profiles based on their characteristic attributes in frequency domain. IEEE Trans. Power Syst. 30, 2434–2441.

Archana Chaudhari received the Bachelor of Engineering (Computer Engineering) from North Maharashtra University, Jalgoan, India, and the Master of Engineering (Computer Engineering) from Savitribai Phule Pune University, Pune, India. She is currently working towards the PhD degree in the Symbiosis International (Deemed University), Pune, India. Her current research interest smart meter data analysis, outage management, machine learning, incremental learning, data analysis. Archana received "Sakal India Foundation's" Research Grant for the integrated research related to Machine Learning and Electricity Smart-meter data analysis and carbon emission study, in March 2019. She received Microsoft Azure Research Grant from May 2018-May 2020, through "AI for Earth" project.

Dr. Preeti Mulay is currently associated with the Symbiosis International (Deemed University), Pune, India. She has authored or coauthored more than 50 technical papers in journal and conference. Her areas of interest include machine learning, data mining, software engineering and knowledge augmentation. Dr. Preeti Mulay recognized as IEEE reviewer, Certified Springer Reviewer. She is reviewer of International Journal of Wireless and Mobile Computing (IJWMC), International Journal of Services Technology and Management (IJSTM)" Inderscience publications, IGI Global. She received "Microsoft Azure" research grant in the area of "Data Science" (CRM: 0795037) and "AI for Earth" for the research related to Diabetes and Smart-meter data analysis, 2018-19 and 2018-20 respectively. She Published research papers in ACM, Elsevier, Springer, InderScience and IEEE Xplore.