Electrical Power and Energy Systems 78 (2016) 368–374

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

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

Regional electricity consumption analysis for consumers using data mining techniques and consumer meter reading data

Ravindra R. Rathod *, Rahul Dev Garg

Indian Institute of Technology Roorkee, Roorkee, India

ARTICLE INFO

Article history:
Received 29 May 2015
Received in revised form 2 June 2015
Accepted 26 November 2015

Keywords:
Data mining
Clustering
Association rules
Electricity consumption
Geographical features

ABSTRACT

Data Mining (DM) techniques are employed to discover electricity consumption pattern at regional level in a city and used to extract knowledge concerning to the electricity consumption with respect to atmospheric temperature and physical distance from geographic features like river, farm, ground and highway. In order to form the different clusters of temperature and consumers based on the basis of electricity consumption *K*-means clustering algorithm is applied. Association rule analysis is carried out to form association rules on electricity consumption to describe the result of physical distance between natural geographic objects and various regions. The work includes pre-processing of data, application of DM algorithms and the interpretation of the discovered knowledge. To validate the proposed work, real databases of around twenty thousand consumers from Sangli city are used.

Introduction

Data mining is used for extracting hidden predictive information from large databases and is used in almost every discipline of science and engineering applications. The principal data mining techniques are regression analysis, decision tree, classification and prediction, clustering, association rule analysis and combination of these techniques. Recently data mining techniques are used in electricity distribution system for electricity supply forecasting [1], designing tariff plans [2,3] and to carry out consumer classification [4–7]. The growth in population, improved living standards and dependency on electronic and electrical appliances in day to day life attracts data miners to analyse consumers and regional analysis on the basis of electricity consumption, revenue collection, theft and fraud detection and other dependent factors contributing on fluctuations in electricity demand and consumption [8].

The electricity demand is increasing every year in almost every sector [9]. Installing and starting new electricity generation plants is difficult due to environmental preservation awareness and pollution control policies of government. To overcome this problem, proper electricity consumption profiling can be used to find alternate ways for managing electricity loads in near future with existing electricity generation capacity. Electricity profiling can be carried out for individual consumer, on small area and at regional levels.

Different regions in a city have a different electricity consumption pattern with respect to locality and nearest man-made geographical features. Similarly, regions at different geographical locations may have similar or matching electricity consumption pattern in association with similar geographical features such as cities on the bank of rivers or at same altitude (hilly areas). The consumption pattern of region is highly influenced by geographical features and environmental conditions like atmospheric temperature, humidity and rainfall [10].

Related work

Data mining uses different approaches and builds different models depending upon the type of data and objectives. Broadly, data mining methods are classified as predictive and descriptive. Predictive methods are classification techniques and descriptive method covers association rules and clustering [11]. Depending on data types used for analysis, data mining methods are classified as web and text mining (internet application), multimedia mining (audio/video applications), spatial data mining (remote sensing and GIS application) and time series data mining (business and financial applications) [12].

Data mining techniques are applied to extract useful information from electricity consumption data of consumers. Data mining methods are used at different stages to generate useful information for preparing regional electricity consumption profile. Each region

* Corresponding author.
E-mail addresses: rrr.wce@gmail.com (R.R. Rathod), garg_fce@iitr.ac.in (R.D. Garg).

http://dx.doi.org/10.1016/j.ijepes.2015.11.110
0142-0615/

has few thousands of consumers where electricity is supplied in a hierarchical manner. Generated electricity is transmitted through transmission lines to electricity distribution network, substations, regional or area transformers, various electricity poles and at the end to consumer.

Data mining is used for short and long term forecasting of electricity demand, planning, fraud, theft and fault detection [13–16]. Many studies are carried out using consumers load profile analysis [3–5,10,17,18] for tariff planning and consumer characterization using house size and economic condition of consumer [19]. A seasonal electricity consumption analysis is carried out to identify environmental effect on consumers electricity consumption [2,7]. The load profile of individual consumer varies differently depending on income level, residence locality, residential building type, total number of persons living together and environmental factors [19,20]. Some studies were also carried out using sample data from city to study load profile of few consumers representing entire class or cluster they belong to [1,2,5,21,22].

In present study, the effect of various geographical objects nearer to the region and atmospheric temperature of the city on electricity consumption is studied using clustering and association rule analysis. To perform the analyses data of about 20,000 consumers from electricity distribution region (feeder) of Sangli city has been used. The consumers are divided into regions by government agency MahaDiscom (Maharashtra state electricity distribution company limited), which supplies electricity to all consumers. After grouping each consumer in individual regions, data cleaning is carried out to remove the consumers having very low annual electricity consumption of 150 units per annum (average monthly consumption less than 15 units; 1 unit of electricity equals to 1 kW-h consumption for 1 hr).

Geographic information system (GIS)

The term geographic information system (GIS) is defined by Chrisman [23] as "Geographic information system – Organized activity by which people measure and represent geographic phenomena then transform these representations into other forms while interacting with social structure. GIS system helps to manage spatially or geographically references (from earth) data by storing, transforming, accessing, editing and displaying for studying environmental and man-made (artificial) processes, detecting changes and trends for the possible planning and analysis purpose.

Commonly GIS is used for applications involving planning transportation network (Road, train) [24,25], municipal tax/revenue and sewer, water supply management [26], flood assessment modeling [27], urban growth and sprawl assessment [28], land and natural resources management system, planning information system [23].

Wang et al. [29] presented a model using GIS and spatial database using city type (major), highway (close) and population (>million). The presented model represents spatial features as attribute data to deduce rules for logistic industry in china. Such attribute reduction policy is very useful to minimize spatial data storage and processing cost while preserving spatial feature data in attribute form.

Data mining and GIS

Development and advances in remote sensing (RS), global positioning system (GPS) and GIS has generated massive amount of spatial and non-spatial. Such data is very helpful for various planning purposes. Data mining plays an important role in assessing massive GIS data for proper planning using spatial and non-spatial data [30,31].

Many researchers applied data mining techniques (spatial and non-spatial) for applications such as; car crash test in transportation using road data (width, roughness) [13] through GPS and GIS. An approach to build spatial database to perform spatial association rule mining is presented by Bogorny et al. [32] to show spatial relationship between spatial features (gas station, hospital, water bodies and street) with non-spatial feature population.

Some GIS applications includes electricity distribution network information system [33] used to manage entire electrical network, mapping electrical power distribution network [34,35], electricity loss, theft and fraud detection in power distribution network [34], managing electricity feeders for effective distribution of electricity to consumers [36], electricity distribution network load analysis and load pattern classification of consumers using clustering technique [17]. In electricity distribution system; spatial data mining has wide range of applications. For example, a GIS study gives us the high electricity consumption, high electricity losses or high revenue generating areas from a city or state. Such relationships between spatial features (area or zone) and non-spatial features (electricity consumption or revenue collection) are easy to discover with the help of GIS and data mining. Common spatial data mining tasks [31] are; spatial classification and prediction, spatial association rule mining, spatial cluster analysis and geovisualization.

Case study on a consumer's electricity consumption data from Sangli city

The study has been carried out on consumer's data from Sangli city (16.86° N, 74.57° E) located on the banks of river Krishna in the western part of Maharashtra state, India. The valley of the river Krishna offers many irrigation and agricultural advantages to Sangli city and district. Due to large scale irrigation, agriculture farming is the main business around Sangli district and nearby region. The physical location of Sangli city comes under Deccan plateau, nearly 120 km away from Arabian Sea and at 549 m height from mean sea level. Fig. 1 shows location map of Sangli city used for study.

Data selection, cleaning and pre-processing

The dynamic expansion of cities due to development in infrastructure, industries and residential growth increases electricity consumers which produce huge electricity consumption data. Study and analysis of consumers electricity consumption data is helpful for detecting consumption pattern at various levels such as individual consumer, zone and at regional level in electricity distribution network. To analyse huge data, data pre-processing operations are carried out in first step before applying data mining algorithm [11,12]. Data pre-processing includes preparation of data in desired form to work, which is clean and free from any noise. It is also used for reduction of actual huge data into summative workable data to avoid unnecessary processing of unwanted, meaningless data. Consumers monthly electricity consumption values are checked and inconsistent consumers are removed from study data.

Clustering

Clustering is an important technique applied to form groups or clusters of data, which represents a common property of entire elements within group [37–40]. Each element within cluster represents a class or common property among them. Clustering algorithms are used for various data types including numerical, categorical and multimedia [15,16]. In this proposed study work

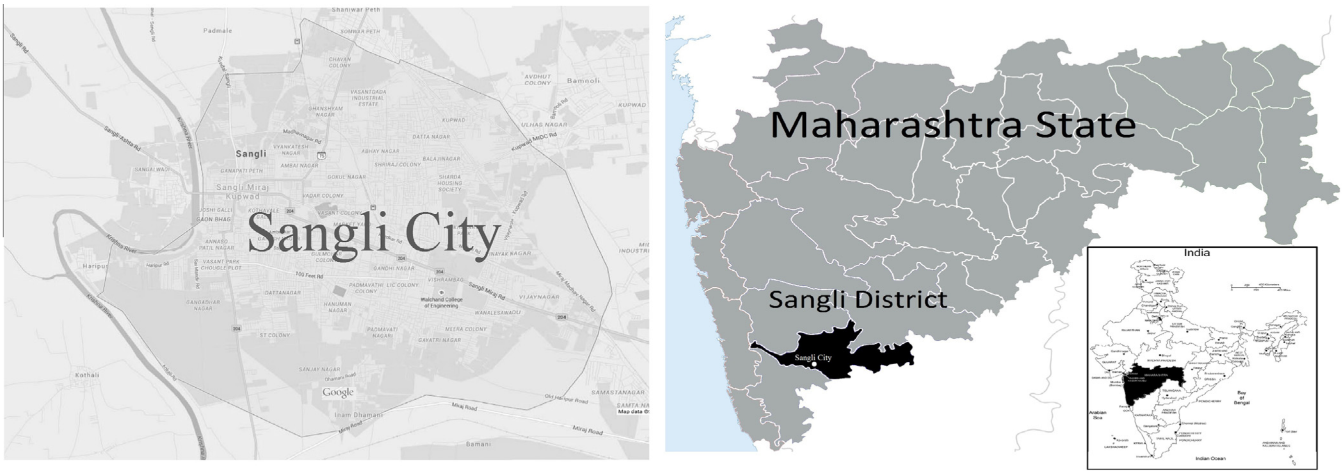

Fig. 1. Location map of Sangli city, Maharashtra State, India.

clustering algorithm is applied on consumer's monthly electricity consumption data and atmospheric temperature of study location. The detailed implementation is discussed as follows.

Consumers clustering

K-means clustering algorithm is used to form clusters of consumers using monthly electricity consumption data over year 2013. The results of implementation in Matlab are compared with WEKA software results to validate correctness and validity of implemented clustering algorithm. The *K*-means clustering algorithm using Euclidean distance is executed on consumer's data to form 4, 8, 16 and 32 clusters of consumers. It has been observed that 8 clusters represent useful information about consumers as the results are very close to WEKA results. Table 1 shows comparison of both results.

Atmospheric temperature clustering

Atmospheric temperature plays a substantial role in consumer behavior with respect to economic and health conditions. Atmospheric temperature of the Sangli city is amassed for the duration of year 2013 (www.accuweather.com/en/in/sangli/189290/month/) with daily minimum and maximum temperature. Clustering of monthly average temperature is carried out to form 3 and 4 clusters of months with similar weather conditions to examine the impact of weather on electricity consumption in the period. The obtained clusters of months based on atmospheric temperature using $K = \{3;4\}$ are $\{3;4;5\}$ and $\{1;3;3;5\}$ respectively. For further analysis, 3 clusters are used as it gives a proper classification of months over the year with respect to atmospheric temperature since the Indian weather has broadly 3 seasons). Table 2 shows

Table 1
Result comparison of Matlab program and WEKA software.

Cluster Id.	Consumer count		Consumers misclassification	
	K-means clustering	WEKA results	Count	% Difference
C	11,995	11,873	122	0.63
A	5866	5948	82	0.42
B	990	1028	38	0.19
F	165	170	5	0.02
E	88	88	0	0.00
G	47	47	0	0.00
H	20	20	0	0.00
D	3	3	0	0.00
Total	19,174		247	–

Table 2
Atmospheric temperature and its clustering result.

Month (Year 2013)	Atmospheric temperature (°C)		Cluster	
	Maximum	Minimum	Assigned cluster number in program	Assigned category to clusters
Jan	31.9	15.3	1	Low
Feb	32.4	17.3	1	Low
Mar	33.8	19.7	2	Medium
Apr	32.7	20.8	2	Medium
May	33.6	23.4	2	Medium
June	29.7	23.4	3	High
July	28.0	23.6	3	High
Aug	28.8	23.5	3	High
Sep	29.8	23.2	3	High
Oct	31.7	23.1	3	High
Nov	34.7	20.8	2	Medium
Dec	33.0	17.9	1	Low

average minimum and maximum atmospheric temperature for all months in the year 2013. *K*-means clustering algorithm for three clusters ($K = 3$) is executed and each month is attributed to its class (low, medium and high temperature).

Association rule generation on geographic features and electricity consumption

Association rule analysis or market basket analysis is widely used data mining technique for establishing relationships between different entities or phenomenon [41,42]. Association rules are used for marketing [43], crime and fraud detection [44], climate change [45,46] and disease analysis etc.

Association rules are of the "if-then" form often expressed as $X \rightarrow Y$; represents presence of Y when X is also present (X and Y is a set of entities). The quality of an association rule is mainly measured by its confidence and support. Rules satisfying user's threshold (support and confidence) are considered as good rules, representing facts within data entities. Association rule discovery is conducted out in two steps: step 1 generates frequent item sets and step 2 constructs association rules from frequent sets (dividing frequent sets into subsets).

Agrawal et al. [41] proposed 'Apriori' algorithm for finding an association rules using support and confidence measures. In market basket analysis; support of association rule is defined on itemsets which gives the frequency of item brought together in transactional data set. An itemset satisfying defined threshold for support is called as frequent itemset. Another measure, confidence is defined as the chance of the rule's left hand side when the transaction data also hold in the right hand side. Confidence measure is uni-directed and gives different values for the rules $X \rightarrow Y$ and $Y \rightarrow X$ [41]. Support and confidence of a rule are expressed by Eqs. (1) and (2).

$$\text{Support}(X) = \frac{\text{Support}(X)}{N}; \quad N = \text{Total number of transactions} \tag{1}$$

$$\text{Confidence}(X \rightarrow Y) = \frac{\text{Support}(X \cup Y)}{\text{Support}(X)} \tag{2}$$

Geographic features play a vital role in leaving style of human beings. Day to day life is highly affected by climatic conditions which depend on related geographic natural features. All historical and major cities are situated either on the bank of the river, at the bottom of mountains or at specific geographical condition to meet day to day needs for a better lifestyle.

Association rules are best to depict the relationship between dependent entities. The impact of nearest geographical features on electricity consumption is to be studied with the help of association rule building. For study purposes, geographical distance criteria are looked at to compare electricity consumption of six regions that have been used for clustering analysis. The geographic features selected for study are river, farm and open space (ground). Table 3 shows relative distance of all regions from selected geographical features.

Data mining model, analysis and results

Proposed data mining model works in two modules. In phase 1; relationship between atmospheric temperature and electricity consumption is established using electricity consumption data of one year and in phase 2; the effect of geographic objects on the electricity consumption of city at regional level.

The framework of proposed data mining model is simple and easy to carryout clustering and association rule analysis consisting two modules shown in Fig. 2. Clustering module uses K-means clustering algorithm; used to form clusters from consumer's monthly electricity consumption (units) data and average monthly temperature of city to get 3 and 8 clusters of data respectively (the number of clusters formed with respect to Sections 'Consumers clustering' and 'Atmospheric temperature clustering' respectively).

In second module association rules are generated using 'Apriori' algorithm [11] to generate association rules between (a) atmospheric temperatures and consumer's electricity consumption, and (b) consumers in electricity distribution region and geographical objects. Geographical objects are selected in such a way to check impact on electricity consumption by consumers.

Consumer electricity consumption data

Nearly twenty thousand consumers are used for study purpose. These consumers are clustered into 8 clusters using monthly electricity unit's consumption. It is found that nearly 62% consumers use less than 100 units per month while 30% consumers' consumption is ranging in between 100 and 200 units of electricity per month and only 20 consumers uses more than 3000–8000 units per month which is about 80% of total electricity use (few commercial offices and common buildings). Table 4 gives details of maximum and minimum monthly electricity consumption in each cluster.

It has been observed that each region have specific pattern of electricity consumption which represents regions locality and economic status. For all regions, electricity consumption pattern over

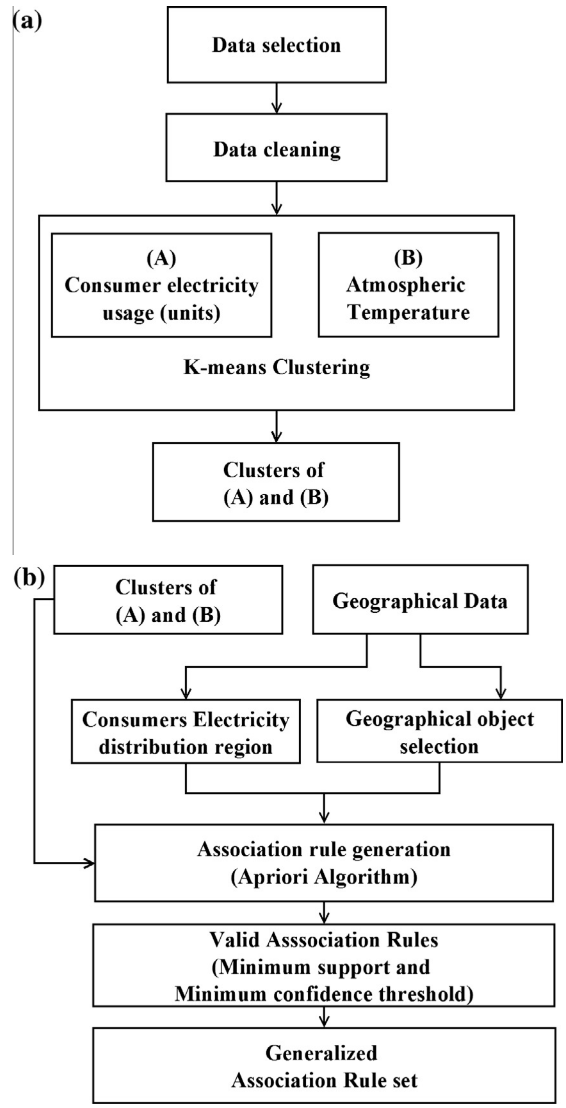

Fig. 2. Proposed data mining model. (a) Clustering module, (b) association rule module.

Table 3
Geographic features of all regions (feeder).

Feeder name	Feeder Id	Geographical features			Proximity to highway	Average annual electricity consumption	
		River	Ground	Farm		Units	Assigned category
Government colony	F-1	Yes	No	Yes	No	108.18	Low
Vijay nagar	F-2	No	Yes	No	Yes	190.97	High
Vishrambagh	F-3	No	Yes	No	Yes	139.82	Medium
Datta nagar	F-4	Yes	No	Yes	No	120.93	Low
Doordarshan	F-5	Yes	No	Yes	Yes	133.29	Medium
Industrial	F-6	No	No	No	No	132.97	Medium

the year is similar with variations in actual electricity consumption units. Figs. 3 and 4; show average monthly electricity consumption of all regions and electricity consumption clusters.

Atmospheric temperature

Sangli city has pleasant weather throughout the year. The minimum and maximum temperature ranges between 15 °C and 35 °C in year. Broadly we have clustered the months using temperature into three clusters. These clusters are categorized as low, medium and high temperature clusters. Table 2 shows assigned cluster and category to each month. Figs. 5 and 6 show average monthly temperature for the year 2013 and average temperature of assigned clusters.

Relationship between atmospheric temperature and electricity consumption

It has been observed that the electricity consumption is directly proportional to temperature. The electricity use increases as an increase in minimum temperature, because the maximum temperature is comparatively small for high temperature cluster than low and medium temperature clusters, but the lower limit temperature is comparatively larger than others. Normally minimum temperature is recorded at night time; it suggests that the electricity consumption increases during night time due to excessive consumption of electrical and electronic appliances for cooling, entertainment and lighting. It has been observed that the electricity consumption pattern is similar in all clusters. The electricity consumption variation in clusters (arranging in decreasing order: D, H, G, E, F, B, A and C) highlights the electricity consumption pattern of clusters, it is noticed that the electricity consumption varies drastically in response to temperature on high electricity using consumers than low. Figs. 7 and 8 represents the relationship between atmospheric temperature and electricity consumption for all clusters.

Association rule analysis

Matlab implementation of Apriori algorithm is executed on the data prepared using Table 3. With a minimum confidence of 30%; fourteen association rules are refined as follows.

(a) Geographical_feature (River, Near) → Electricity_consumptions (Low) Conf: (66.67%)
(b) Geographical_feature (River, Near) → Electricity_consumptions (Medium) Conf: (33.33%)
(c) Geographical_feature (Farm, Near) → Electricity_consumptions (Low) Conf: (66.67%)
(d) Geographical_feature (Farm, Near) → Electricity_consumptions (Medium) Conf: (33.33%)

Table 4
Average maximum and minimum electricity consumption of clusters.

Cluster Id.	Consumer count	Consumers (%)	Monthly electricity consumption (units)	
			Maximum	Minimum
A	5866	30.59	197.19	144.74
B	990	5.16	459.08	309.99
C	11,995	62.55	78.86	59.12
D	3	0.015	17215.66	12835.66
E	88	0.45	2217.68	1725.28
F	165	0.86	1023.81	777.98
G	47	0.24	3795.93	3148.85
H	20	0.10	7568	5333.55

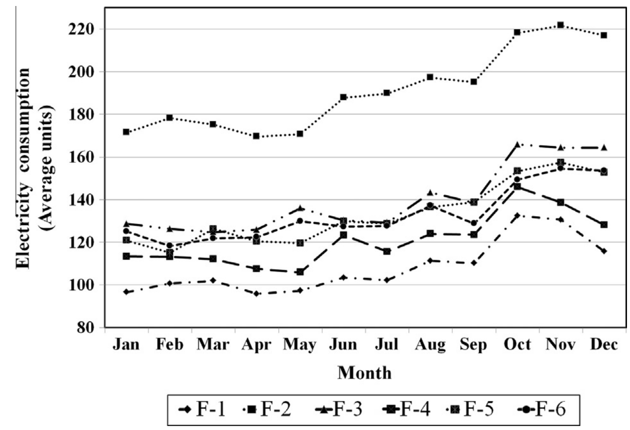

Fig. 3. Average monthly electricity consumption of all regions for year 2013.

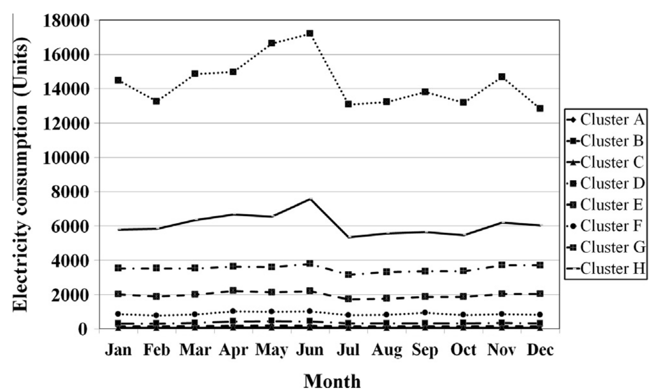

Fig. 4. Average monthly electricity consumption of all consumer clusters.

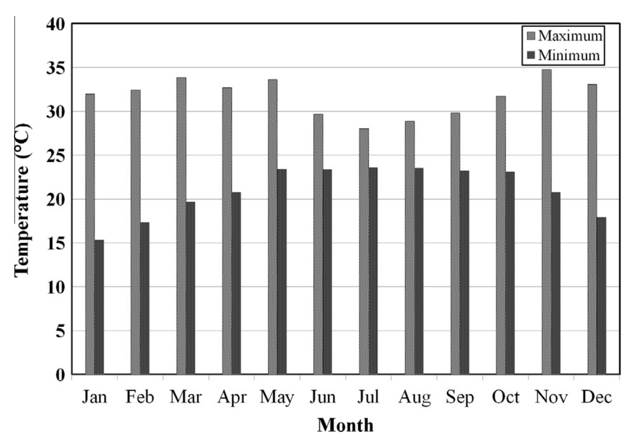

Fig. 5. Average monthly temperature for year 2013.

(e) Geographical_feature (Ground, Near) → Electricity_consumptions (High) Conf: (50.00%)
(f) Geographical_feature (Ground, Near) → Electricity_consumptions (Medium) Conf: (50.00%)
(g) Geographical_feature (Highway, Near) → Electricity_consumptions (High) Conf: (33.33%)
(h) Geographical_feature (Highway, Near) → Electricity_consumptions (Medium) Conf: (66.66%)
(i) Geographical_feature (River, Near) ∧ geographical_feature (Farm, Near) → Electricity_consumptions (Low) Conf: (66.67%)

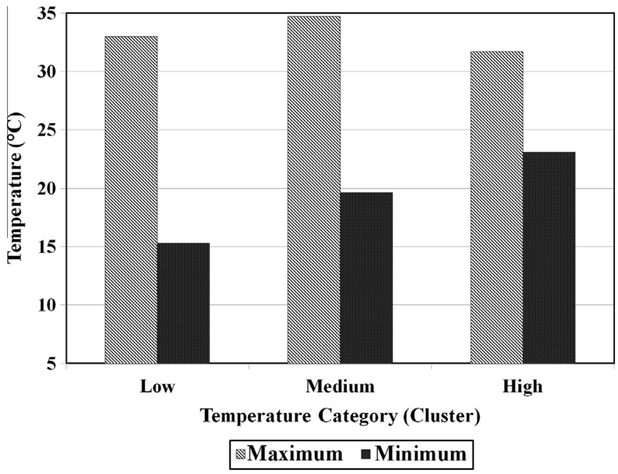

Fig. 6. Average temperature of assigned clusters.

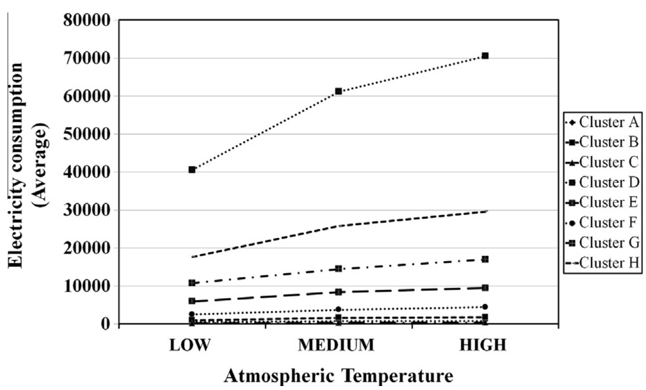

Fig. 7. Electricity consumption and temperature clusters.

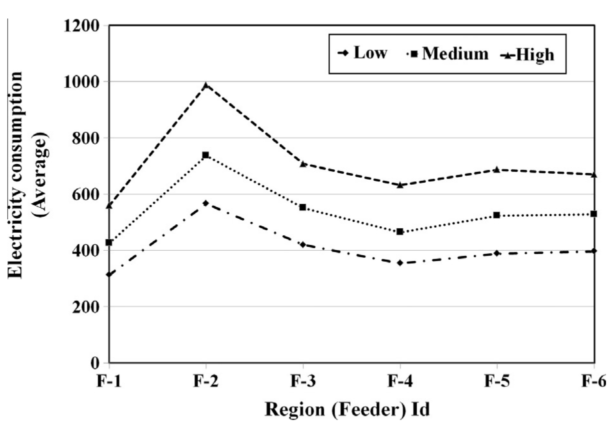

Fig. 8. Electricity consumption for all regions (feeders) with temperature (clusters).

(j) Geographical_feature (River, Near) ∧ geographical_feature (Farm, Near) → Electricity_consumptions (Medium) Conf: (33.33%)
(k) Geographical_feature (River, Near) ∧ geographical_feature (Highway, Near) → Electricity_consumptions (Medium) Conf: (100%)
(l) Geographical_feature (Farm, Near) ∧ geographical_feature (Highway, Near) → Electricity_consumptions (Medium) Conf: (100%)
(m) Geographical_feature (Ground, Near) ∧ geographical_feature (Highway, Near) → Electricity_consumptions (High) Conf: (100%)
(n) Geographical_feature (River, Near) ∧ geographical_feature (Farm, Near) ∧ geographical_feature (Highway, Near) → Electricity_consumptions (Medium) Conf: (100%)

Obtained association rules are simple and give information about electricity consumption patterns of regions. These association rules explain the effect of geographic proximity on electricity consumption. Lower consumption of electricity with 66.67% confidence is observed in nearby region of river or farm; alone or together whereas medium electricity consumption is in regions, where either ground or highway is near with a confidence in range 50–66.67%. The regions adjacent to ground and nearer to highway use highest electricity as compared with other regions having 100% confidence. It has been observed that highway is main cause for increase in electricity consumption whereas natural features river and farm decreases electricity consumption. Finally, generalized association rules are summarized in following five rules:

(a) **IF** (River **AND** Farm, Near) **THEN** Electricity consumption (Medium **OR** Low)
(b) **IF** (River **OR** Farm, Near) **THEN** Electricity consumption (Medium **OR** Low)
(c) **IF** (Ground **OR** Highway, Near) **THEN** Electricity consumption (High **OR** Medium)
(d) **IF** ((River **OR** Farm, Near) **AND** (Highway, Near) **THEN** Electricity consumption (Medium)
(e) **IF** (Ground **AND** Highway, Near) **THEN** Electricity consumption (High)

Conclusion and future scope

This data mining model is generalized and applicable to any area or city considering proximity to geographical and other man made features. The presented model is able to differentiate regions by their electricity consumption behavior using set of association rules. The innovative approach of this model is capability to handle large volume of data and performing area profile for planning residential area aiming electricity efficient living.

The present study focuses on examining consumer's electricity consumption within electricity distribution region and impact of spatial feature's proximity to consumer's location (region). Spatial features nearer to consumer's electricity distribution region are identified such as water body (river and pond), playing ground, farm (agricultural land) and highway. Any geographic feature dependencies on electricity consumption are reflected in generated association rules. The summarized association rule depicts the changes in consumption patterns for consumers residing nearer to highway or farm (high and low electricity consumption respectively).

The semantic knowledge about geographic features closer to consumer's location helps to predict their electricity consumption behavior. Moreover, such patterns are helpful for prediction of electricity usage in such regions and town planning.

A detailed study for individual consumers profiling is needed to characterize each consumer's electricity consumption pattern using spatial features such as location, type and size of residence (bungalow or apartment), vegetation surroundings and nearer to road. Some non-spatial features are also important to consider including income, number of peoples at residence and appliances.

Detailed consumer and regional profiling using GIS, GPS and remote sensing is helpful for prediction and demand analysis, theft and fraud detection, planning and growth in an electricity distribution network.

Acknowledgements

This work is carried out using consumer data from Sangli city provided by MahaDiscom. It is a public sector undertaking controlled by the Government of Maharashtra, India which is the second largest electricity distribution utility in the world after State Grid Corporation of China.

References

[1] Manonmani R, Renuka Devi M. Electricity forecasting using data mining techniques in Tamilnadu and other countries – a survey. Int J Emerg Trends Eng Develop 2012;6:295–302.
[2] Chen C, Hwang J, Huang C. Application of load survey systems to proper tariff design. Power Syst, IEEE Trans Power Syst 1997;12:1746–51. http://dx.doi.org/10.1109/59.627886.
[3] Chicco G, Napoli R, Postolache P, et al. Customer characterization options for improving the tariff offer. IEEE Trans Power Syst 2003;18:381–7. http://dx.doi.org/10.1109/TPWRS.2002.807085.
[4] Chicco G, Napoli R, Piglione F, et al. Load pattern-based classification of electricity customers. IEEE Trans Power Syst 2004;19:1232–9. http://dx.doi.org/10.1109/TPWRS.2004.826810.
[5] Figueiredo V, Rodrigues F, Vale Z, Gouveia JB. An electric energy consumer characterization framework based on data mining techniques. IEEE Trans Power Syst 2005;20:596–602. http://dx.doi.org/10.1109/TPWRS.2005.846234.
[6] Ferreira AMS, Cavalcante CAMT, Fontes CHO, Marambio JES. A new method for pattern recognition in load profiles to support decision-making in the management of the electric sector. Int J Electr Power Energy Syst 2013;53:824–31. http://dx.doi.org/10.1016/j.ijepes.2013.06.001.
[7] Benítez I, Quijano A, Díez J-L, Delgado I. Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers. Int J Electr Power Energy Syst 2014;55:437–48. http://dx.doi.org/10.1016/j.ijepes.2013.09.022.
[8] Zhang T, Siebers P-O, Aickelin U. A three-dimensional model of residential energy consumer archetypes for local energy policy design in the UK. Energy policy 2012;47:102–10. http://dx.doi.org/10.1016/j.enpol.2012.04.027.
[9] Navani JP, Sharma NK, Sapra S. Technical and non-technical losses in power system and its economic consequence in Indian economy. Int J Electron Comput Sci Eng 2012;1:757–61.
[10] Lin JK, Tso SK, Ho HK, et al. Study of climatic effects on peak load and regional similarity of load profiles following disturbances based on data mining. Int J Electr Power Energy Syst 2006;28:177–85. http://dx.doi.org/10.1016/j.ijepes.2005.11.014.
[11] Tintarev N, Masthoff J. Recommender systems handbook; 2011. http://dx.doi.org/10.1007/978-0-387-85820-3.
[12] Venkatadri M, Reddy L. A review on data mining from past to the future. Int J Comput Appl 2011;15:19–22. http://dx.doi.org/10.5120/1961-2623.
[13] Barai S. Data mining applications in transportation engineering. Transport 2003;18:216–23. http://dx.doi.org/10.1080/16483840.2003.10414100.
[14] Da Cunha C, Agard B, Kusiak a. Data mining for improvement of product quality. Int J Prod Res 2006;44:4027–41. http://dx.doi.org/10.1080/00207540600678904.
[15] Djeraba C. Data mining from multimedia. Int J Parallel Emergent Distrib Syst 2007;22:405–6. http://dx.doi.org/10.1080/17445760701207561.
[16] Su X. Data mining methods and models. Am Stat 2008;62. http://dx.doi.org/10.1198/tas.2008.s97. 91–91.
[17] Shin J-H, Yi B-J, Kim Y-I, et al. Spatiotemporal load-analysis model for electric power distribution facilities using consumer meter-reading data. IEEE Trans Power Deliv 2011;26:736–43.
[18] Wang Z, Bian S, Liu Y, Liu Z. The load characteristics classification and synthesis of substations in large area power grid. Int J Electr Power Energy Syst 2013;48:71–82. http://dx.doi.org/10.1016/j.ijepes.2012.11.032.
[19] Dzobo O, Alvehag K, Gaunt CT, Herman R. Multi-dimensional customer segmentation model for power system reliability-worth analysis. Int J Electr Power Energy Syst 2014;62:532–9. http://dx.doi.org/10.1016/j.ijepes.2014.04.066.
[20] Min B, Golden M. Electoral cycles in electricity losses in India. Energy Policy 2014;65:619–25. http://dx.doi.org/10.1016/j.enpol.2013.09.060.
[21] Mori H, Kosemura N, Toru K, Kazuyuki N. Data mining for short-term load forecasting. Power Eng Soc Winter Meet 2002;623–624. http://dx.doi.org/10.1109/PESW.2002.985075.
[22] Morais J, Pires Y, Claudomir C, Aldebaro K. An overview of data mining techniques applied to power systems. In: Ponce J, Adem K, editors. Data mining and knowledge discovery in real life applications; 2009. p. 438.
[23] Chrisman NR. Review paper what does 'GIS' mean ? Main 1999;3:175–86. http://dx.doi.org/10.1111/1467-9671.00014.
[24] Loukes D, McLaughlin J. GIS and transportation: Canadian perspective. J Surv Eng 1991;117:123–33.
[25] Lagunzad EL V. GIS applications for road network of the Philippines: a new technology in road management. J Eastern Asia Soc Transport Stud 2003;5:846–54.
[26] Venigalla M, Baik B. GIS-based engineering management service functions: taking GIS beyond mapping for municipal governments. J Comput Civ Eng 2007:331–42.
[27] Kulkarni AT, Mohanty J, Eldho TI, et al. A web GIS based integrated flood assessment modeling tool for coastal urban watersheds. Comput Geosci 2014;64:7–14. http://dx.doi.org/10.1016/j.cageo.2013.11.002.
[28] Shalaby AA, Ali RR, Gad A. Urban sprawl impact assessment on the agricultural land in Egypt using remote sensing and GIS: a case study, Qalubiya Governorate. J Land Use Sci 2012;7:261–73. http://dx.doi.org/10.1080/1747423X.2011.562928.
[29] Wang P, Ma L, Xi Y, Jin L. Research on logistics oriented spatial data mining techniques. In: Proceedings – international conference on management and service science, MASS 2009; 2009. p. 0–3.
[30] Li D, Wang S. Concepts, principles and applications of spatial data mining and knowledge discovery. In: Proceedings of international symposium on spatio-temporal modeling, spatial reasoning, analysis, data mining and data fusion (ISSTM 2005). Beijing, China; 2005. p. 1–13.
[31] Mennis J, Guo D. Spatial data mining and geographic knowledge discovery—an introduction. Comput Environ Urban Syst 2009;33:403–8. http://dx.doi.org/10.1016/j.compenvurbsys.2009.11.001.
[32] Bogorny V, Kuijpers B, Alvares LO. Reducing uninteresting spatial association rules in geographic databases using background knowledge: a summary of results. Int J Geogr Inform Sci 2008;22:361–86. http://dx.doi.org/10.1080/13658810701412991.
[33] Ajwaliya RJ, Udani PM. Power distribution information system using GIS – a case study for SAC-ISRO, Ahmedabad. J Geomatics 2013;7:107–11.
[34] Nawaz-ul-Huda S. GIS for power distribution network: a case study of Karachi, Pakistan. Malaysian J Soc Space 2012;1:74–82.
[35] Karampelas SLVVP, Ekonomou L. A power system simulation platform for planning and evaluating distributed generation systems based on GIS; 2013. p. 379–91. http://dx.doi.org/10.1007/s12667-013-0082-4.
[36] Salawudeen OS, Rashidat U. Electricity distribution engineering and geographic information system (DeGIS). Shape the change, 23 FIG congress, Munich Germany, October 8–13, 2006; 2006. Munich, Germany. p. 1–14.
[37] Leonard M, Wolfe B. Mining transactional and time series data. In: Nelson Gregory S, editor. Proceedings of the 30 annual SAS users group international (SUGI) Conference. Philadelphia; 2005. p. 1–26.
[38] Berkhin P. A survey of clustering data mining techniques. In: Kogan J, Nicholas C, Teboulle M, editors. Grouping multidimensional data. Springer; 2006. p. 25–71.
[39] Fu T. A review on time series data mining. Eng Appl Artif Intell 2011;24:164–81. http://dx.doi.org/10.1016/j.engappai.2010.09.007.
[40] Wang Z, Tu L, Guo Z, et al. Analysis of user behaviors by mining large network data sets. Future Gener Comput Syst 2014;37:429–37. http://dx.doi.org/10.1016/j.future.2014.02.015.
[41] Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD international conference on management of data – SIGMOD '93. New York (NY): ACM Press; USA; 1993. p. 207–16.
[42] Agrawal R, Srikant R. Fast algorithms for mining association rules. In: Proc. 20th int. conf. very large data bases, VLDB. Santiago de Chile, Chile; 1994. p. 1–32.
[43] Cil I. Expert Systems with Applications Consumption universes based supermarket layout through association rule mining and multidimensional scaling. Expert Syst Appl 2012;39:8611–25. http://dx.doi.org/10.1016/j.eswa.2012.01.192.
[44] Cheng C-W, Lin C-C, Leu S-S. Use of association rules to explore cause–effect relationships in occupational accidents in the Taiwan construction industry. Saf Sci 2010;48:436–44. http://dx.doi.org/10.1016/j.ssci.2009.12.005.
[45] Compieta P, Di Martino S. Exploratory spatio-temporal data mining and visualization. J Visual Languages Comput 2007. http://dx.doi.org/10.1016/j.jvlc.2007.02.006.
[46] Olaiya F, Adeyemo A. Application of data mining techniques in weather prediction and climate change studies. Int J Inform Eng Electron Bus 2012;1:51–9. http://dx.doi.org/10.5815/ijieeb.2012.01.07.