Spatiotemporal Load-Analysis Model for Electric Power Distribution Facilities Using Consumer Meter-Reading Data

Jin-Ho Shin, Bong-Jae Yi, Young-Il Kim, Heon-Gyu Lee, and Keun Ho Ryu, *Member, IEEE*

***Abstract*—The load analysis for the distribution system and facilities has relied on measurement equipment. Moreover, load monitoring incurs huge costs in terms of installation and maintenance. This paper presents a new model to analyze wherein facilities load under a feeder every 15 min using meter-reading data that can be obtained from a power consumer every 15 min or a month even without setting up any measuring equipment. After the data warehouse is constructed by interfacing the legacy system required for the load calculation, the relationship between the distribution system and the power consumer is established. Once the load pattern is forecasted by applying a clustering and classification algorithm of temporal data-mining techniques for the power customer who is not involved in automatic meter reading, a single-line diagram per feeder is created, and power-flow calculation is executed. The calculation result is analyzed by using various temporal and spatial analysis methods, such as the Internet geographic information system, single-line diagram, and online analytical processing.**

***Index Terms*—Automatic meter reading (AMR), data mining, geographic information system (GIS), meter reading data, power load analysis, spatiotemporal.**

I. Introduction

SINCE today's power industry is in the process of shifting toward deregulation and a more competitive system, the efficient operation of the power system is becoming increasingly important. The concept of competition as introduced in the power market requires exploring more progressive and efficient system operation to expand social welfare and to reduce the electricity bill.

Even though massive investments are made in new distribution lines and in the operation of equipment, an imbalance in power-supply equipment, such as power outages due to excessive investment or a lack of investment, low-voltage areas, and unused power equipment are predicted depending on the area. This problem stems from the considerable difficulty in checking the load which changes every moment, checking the section with the maximum load of the distribution line, and checking the equipment/load information of the individual transformer. In addition, the current load analysis method uses the monthly or yearly maximum load of the transformer and almost the same correction coefficient [1].

According to the data on South Korea's distribution system and major equipment, there are 7 000 distribution lines with a total route length of 390 000 km, 7 600 000 poles, 1 700 000 transformers, and 120 000 switches. Monitoring the load of the section under the distribution line or individual transformers which are continually changing every moment would be very difficult due to the huge scale of equipment [2]. Although some switches and transformers are equipped with load measurement equipment, attaching this equipment to all facilities is practically impossible. Developing new technology for temporal and spatial load pattern analysis is required to replace the existing method of installing measurement equipment and showing the measured values.

In this paper, we propose a new load pattern analysis model for power facilities every 15 min using meter-reading data without attaching measurement equipment. First, we extract the customer, facility, meter reading, transformer, and feeder measurement data from legacy systems. Afterward, a load pattern analysis model data warehouse is constructed, and the relationship between the distribution facilities and the customer is then established. The load pattern for customers who are not involved in AMR is then predicted. A single-line diagram (SLD) for each distribution feeder is created, and the power flow is calculated.

Customers who are not involved in AMR are classified into high-voltage and low-voltage categories. Each 15-min load pattern is then predicted using AMR data for the high-voltage customers or using transformer wireless load monitoring data for the low-voltage customers. There are some techniques which are applied, such as clustering, classification and temporal data-mining algorithm [3], [4]. We also propose a data-mining framework for the non-AMR load pattern prediction and describe data

Manuscript received August 27, 2009; revised January 05, 2010, May 05, 2010, August 10, 2010; accepted September 20, 2010. Date of publication January 06, 2011; date of current version March 25, 2011. This work was supported by a grant (#R-2005-0-093) from the Power Industry Promotion Fund of the Ministry of Knowledge and Economy (MKE) of the Korean Government, in part by a grant (#07-KLSG-C02) from Cutting-Edge Urban Development—Korean Land Spatialization Research Project funded by the Ministry of Construction and Transportation of the Korean Government, in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF No. 2010-0001732: NRF No. 2009-0052728), and in part by the Development of Real-Time Postal Logistics Systemx(#2006-X-001-02) by the Postal Technology R&D program of MKE/IITA. Paper no. TPWRD-00652-2009.

J.-H. Shin, B.-J. Yi, and Y.-I. Kim are with the Korea Electric Power Research Institute (KEPRI), Daejeon 305-380, Korea (e-mail: jinho@kepco.co.kr; leeboja@kepco.co.kr; yikim@kepco.co.kr).

H.-G. Lee is with the Electronics and Telecommunications Research Institute (ETRI), Daejeon 305-700, Korea (e-mail: hg_lee@etri.re.kr).

K. H. Ryu is with the College of Electrical and Computer Engineering, Chungbuk National University, Cheongju 361-763, Korea (e-mail: khryu@dblab.chungbuk.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRD.2010.2091973

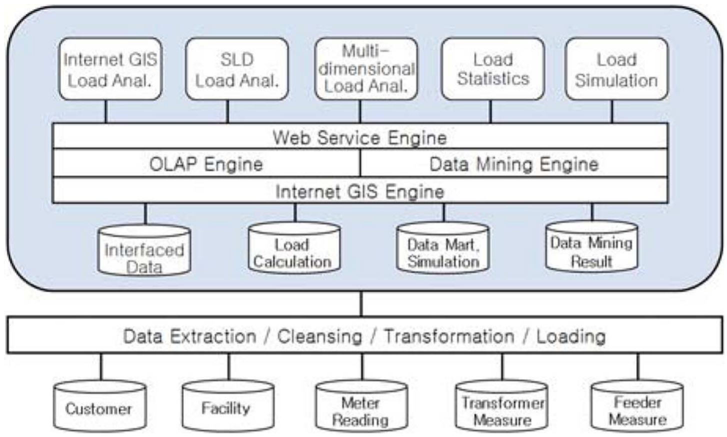

Fig. 1. Structure of the spatiotemporal load-analysis model.

preprocessing and outlier detection, the representative load pattern generation through clustering including the methodology to define the number of classes, the classification model, and the results after the performance evaluation of the classifiers. Our approach is able to represent the difference and diversity per day in the clustering process and to predict the load pattern of non-AMR customers in a classification model, avoiding the data distortion and reproducibility problems of the conventional load pattern research.

To enable users to perform an efficient, multifaceted analysis, we have developed spatial and temporal load pattern analysis models, such as an Internet GIS load analysis, a single-line diagram load analysis for each distribution feeder, and a multidimensional load analysis using online analytical processing techniques.

If the facilities under the distribution line can be analyzed over time and in detail as described before, it enables the ability to check exactly when and what section or equipment experiences how much overload, voltage drop, or power loss. Moreover, in the case of outage or fault, the load conditions can be analyzed at that time for the corresponding line. We can also increase or decrease the load at a specific time or simulate how the load changes over time by transferring it to other equipment; thus, enabling the enhancement of efficiency of facility operation or the optimum load system. We can also make a more accurate investment in facilities by checking the load of the existing line to determine when expansion is necessary for new equipment.

This load pattern analysis model makes it possible to improve the efficiency of facility operation and plan as follows:

1) generating optimal facility relocation based on load characteristics;
2) prioritizing maintenance target facilities;
3) creating basic and verifiable reports for the biannual plan;
4) creating analysis reports for the point of outage or failure time;
5) improving the calculation method of transformer compound load;
6) developing load simulation modules for distributed generation.

II. Model Overview

The overall structure of the temporal and spatial load analysis model is presented in Fig. 1. Data for load calculation include customer, facility, meter reading, transformer measure, and feeder measure. Meter-reading data are active power and reactive power information every 15 min for high-voltage customers. Facility data are GIS spatial information for the distribution system and facilities. Transformer measurement data of the wireless load monitoring system are the voltage and current information measured at the small number of critical transformers. Feeder measurement data are active power and reactive power information measured at the load out area of the substation feeder.

These data are transformed depending on the configuration and form of connection information data base and loaded after data cleansing is processed (e.g., data extraction from the operating system, deletion of rows without any data, and data input of analog data using the relationship between data).

In developing this model, applying information technology is essential. We use Internet GIS which can be used in the web environment considering user convenience. We also use the data warehouse as well as the online analytical-processing technique to connect the mass storage remote meter-reading data per 15 min of 135 000 high-voltage customers with contract power of more than 100 kW, which makes up 70% of the total distribution load, and to calculate and analyze the load every 15 min. Data mining is a technique that explores the previously unknown knowledge [3]; it is applied to the prediction of the transformer load pattern and customers who are not covered by AMR. We describe the detailed load calculation process and method in Section III.

III. Load Calculation Process and Method

Actual power consumers are divided by AMR and voltage in the section of the distribution system. Depending on the meter-reading frequency and method, customers who are actual power consumers are classified into AMR customers, and customers who are not covered by AMR and low-voltage customers as shown in Table I.

TABLE I
CUSTOMER BY METER READING

Code	Customer	Description
1	AMR customer	AMR, high voltage
2	Non-AMR customer	Non-AMR monthly meter reading, high voltage
3	Low-voltage customer	Monthly meter reading, low voltage

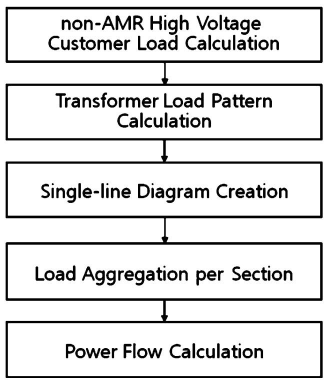

Fig. 2. Load calculation process.

Here, the high-voltage customer is defined as a customer using the code whose contract type is classified into a high voltage and power supply of more than 3.3 kV with the single-phase, two-wire system. There is no basic differences between the AMR customer and non-AMR customer, because the non-AMR customer is classified by not installing the communication modem according to location or intention of a customer. A low-voltage customer refers to the customer who is neither covered by AMR nor classified as a high-voltage customer. Customers who are in a parent-subsidiary relationship are excluded since their meters are read individually.

In the case of an AMR customer, power consumption is read every 15 min. For the non-AMR high-voltage customer and low-voltage customer, however, power consumption is read on a monthly basis; hence, there is the need to calculate the load pattern every 15 min to match the timing of remote meter-reading data. We develop a process that generates a load profile every 15 min by using the clustering and classification data-mining technique. Load patterns are created based on AMR data for non-AMR high-voltage customers or transformer wireless load monitoring system data for low-voltage customers. Once a 15 min-based load pattern is created for each customer, a single-line diagram is generated per distribution system section; each customer load supplying power to each section is then aggregated. Afterward, the apparent power, current, voltage, and loss for the line and section are obtained by calculating the power flow, and the statistical information is produced. The load calculation process described before is shown in Fig. 2.

A. Load Pattern Calculation of Non-AMR High-Voltage Customer and Transformer

The process wherein the representative load profile is created through cluster analysis using AMR load data for each

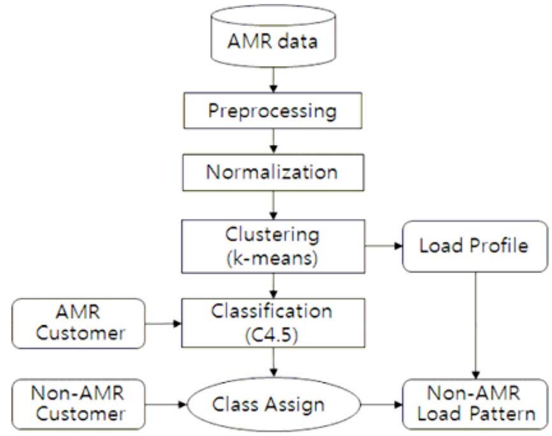

Fig. 3. Calculation process for the load pattern of the non-AMR high-voltage customer.

high-voltage customer or monthly load pattern for the non-AMR high-voltage customer is forecasted as shown in Fig. 3.

Basic information of the AMR high-voltage customer and load data every 15 min are extracted from the constructed load-analysis data warehouse. Since errors and outlier data in the collected data can cause serious performance deterioration, processing the data for data cleansing is essential. In case there are less than 96 data per day or less than 1 for the daily total of active power in raw AMR data, these data are excluded. The minimum AMR active power measured every 15 min is 0.08 kWh for a streetlight as contract power; if it is less than 1 kWh, it is considered not to be read. In addition, in data cleansing for the processing of the outlier self-organizing feature map (SOM), the clustering algorithm is applied [5]. In terms of the configuration matrix, if the data object included in a 10×10 cluster (100 clusters) is less than 1, it is considered outlier and consequently excluded.

If the daily power consumption vector is constructed with weekdays and holidays separated as in the existing load profile research [3], [4], analyzing the load pattern—which changes everyday—is difficult because the same loads on weekdays at the beginning or the end of the month are generated. Therefore, in this paper, the total monthly load every 15 min is reformulated for each customer as one vector as shown in

$$V^{(m)} = \sum_{i=1}^{30} \left\{ V_0^{(m)}, \ldots, V_h^{(m)}, \ldots, V_H^{(m)} \right\} \tag{1}$$

where V is the load vector, m is the customer, h is the interval $=$ $0, \ldots, 2345$ (every 15 minutes: 96 points), $i(\text{day} = 1, \ldots, 30.)$

If raw load data are used as is, clustering is generated according to the power consumption distribution. Therefore, clustering analysis should be performed after normalization. There are many clustering algorithms. We use the k-means algorithm due to a simple and fast in large database. Once a number of clusters are created as a result of the cluster analysis, model classification of each class is executed. Classification is used to enable the description tool to distinguish the objects of different classes and prediction of unknown class labels. In other words, each cluster is classified according to the AMR customer properties; when a non-AMR customer is inputted, the class label is predicted. Here, customer property information includes the contract type, contract power, power user type, industrial type, supply type, and monthly meter-reading value; they are used as input variables of the classifier. There are several types of classifiers: decision tree, Bayesian classifier, neural network, support vector machine (SVM), and rule-based classifier [5]–[8].

In this paper, the decision tree classifier is used considering its performance. The decision tree is made up of a set of nodes that classify the past realizations of the objective variable. Each classification is achieved by separation rules according to the numerical or categorical values of the explanatory variables. The classification rules of each node are derived from a mathematical process that minimizes the impurity of the resulting nodes, using the available learning set. The main advantage of the decision tree is the easy interpretability of the results and the supply of probability values without assuming normal distributions [7]. Non-AMR high-voltage customers' load prediction is performed by allocating the monthly pattern class using the C4.5 decision tree. This predicted pattern is the load pattern with normalization. Therefore, it should be reverted to the initial load capacity as follows:

$$C_k^{'} = T \times \frac{C_k}{\sum_{i=1}^{n} C_i} \tag{2}$$

where

- C' reverted active power;
- C normalized active power;
- k point of time, n: all point of time;
- T monthly reading value of a non-AMR customer.

Remote meter reading of low-voltage customers is carried out experimentally for only residential customers. Thus, that data cannot be used to create load patterns at the moment because there is not enough data and not for all types of customers, such as commercial, educational, and agricultural, etc. Therefore, in this paper, load patterns per transformer are created by using the transformer wireless load monitoring system data measured every 30 min and property information for low-voltage customers as supplied by the corresponding transformer. The creation of load pattern for the transformer is very similar to that one of non-AMR high-voltage customers; the representative load pattern is created by using the current data of the transformer wireless load monitoring system, and the load for the unmeasured transformer is predicted. Property information for the transformer and customer for the construction of the classification model includes the electric light number, power number, load area property, transformer capacity, customer's contract power, application of electricity, low/high voltage, and monthly power consumption.

B. SLD Creation and Power-Flow Calculation

An SLD is generated by formulating the sections in the extracted equipment diagram data and to calculate the total load per line and power flow. The section's power source side and load side are determined by searching the section where the distribution system loops and by comparing the existing equipment diagram, high-voltage system process diagram, and line diagram. The section loads are totaled by including the load in the preceding section in the case of the terminal supplying power directly to high-voltage customers from underground multi-circuit switches. The algorithm generating the SLD is the ternary tree recursive method; when there is no automatic load transfer switch (ALTS) or section information for the lower part and underground transformer, the switches are open after the root node is generated from substation circuit-breaker (CB) information, and the section at the branch point of the multicircuit switch is isolated; then, the ternary tree is generated while moving the pointer to the equipment on the upper-level power source side.

Once the SLD for each section is created, the load for each section is aggregated. If the loads supplied from the section to the AMR customer, non-AMR high-voltage customer, and transformer are aggregated, the section load per time zone is created, and power flow is calculated.

Power flow is calculated to obtain voltage, system loss, and power flow at the specific feeder and sections. In other words, line or transformer loss and the voltage phase angle of each bus are obtained by using the reactive power flow and power in the load power, reactive power, and distribution system [9]–[11]. Due to its operating characteristics, the distribution system has a radial structure. That is why tree-structured calculation is used. In this paper, the case with the main feeder is only explained by applying the Forward Sweeping method. Since the case with the lateral branch line is calculated by using the ternary tree method, the application is the same as the case of the main feeder. The following equation can be derived from the electrical equivalent model of the distribution system:

$$\begin{aligned} I(1) &= \frac{|V(1)|\angle\delta(1) - |V(2)|\angle\delta(2)}{R(1) + jX(1)} \\ P(2) - jQ(2) &= V^*(2)I(1). \end{aligned} \tag{3}$$

In other words, current (I) flowing in the number 1 bus and the number 2 bus is the same as the value obtained by dividing the difference in the two buses' voltages by the impedance value of the number 1 branch; the power passing the number 2 bus is the value obtained by multiplying the voltage of the number 2 bus by the current flowing into the number 2 bus. The voltage equation for the number 2 bus can be obtained by using (3)

$$|V(2)| = \left[\begin{array}{c} \left\{ \begin{array}{c} (P(2)R(1) + Q(2)X(1) - 0.5|V(1)|^2)^2 \\ -((R^2(1) + X^2(1))(P^2(2) + Q^2(2)) \end{array} \right\}^{1/2} \\ -(P(2)R(1) + Q(2)X(1) - 0.5|V(1)|^2) \end{array} \right]^{1/2} \tag{4}$$

According to (4), obtaining the voltage of the next available bus, the active and reactive power of the bus as well as the voltage of the preceding bus should be available. The voltage of the preceding bus can be easily obtained, since it is the voltage of the substation load out area. On the other hand, the power passing the corresponding bus can be obtained from (5). In other words, power flowing into the number 2 bus is the sum of the load for the number 2 bus $(PL(2), QL(2))$, the load for all of the next available buses $(PL(i))$, and power loss of all branches except the number 1 branch $(LP(i))$. Accordingly, to obtain the

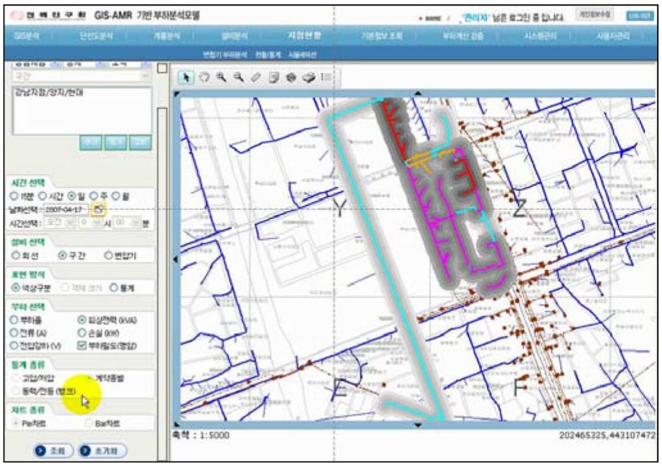

Fig. 4. Load analysis using Internet GIS.

Fig. 5. Load analysis using the SLD.

power, the loads of all buses and power loss of all branches except the number 1 branch should be available. Power loss can be obtained by using

$$LP(1) = \frac{R(1) \times [P^2(2) + Q^2(2)]}{|V(2)|^2}$$
$$LQ(1) = \frac{X(1) \times [P^2(2) + Q^2(2)]}{|V(2)|^2}. \quad (5)$$

According to (5), to obtain the power loss of branch, the voltage of $i+1$ bus should be available. If we go back to the first equation, the voltage of the $i+1$ bus is obtained from the power of the $i+1$ bus, which can be obtained from the loss of all branches, which, in turn, can be obtained from the voltages of all buses. Since it is a matter of circulation, the initial power loss value of all branches is first set to zero; the power of the $i+1$ bus is then obtained, and, finally, the voltage of the $i+1$ bus is derived. This process will continue until the last bus is reached. After the values for the last bus are obtained, the values are obtained from the beginning. If the difference in loss falls within the error range, the process is completed.

IV. Development of the Load-Analysis Model

Internet GIS load analysis is a module wherein facilities drawing as well as load conditions and statistics for the feeder, section, and transformer in various time zones can be analyzed on a web browser. As shown in Fig. 4, the daily maximum load for the three distribution lines is displayed by section with the power-load density and section apparent power with color differentiation. In spatial analysis, the selection of load time is possible only at a specific point of time; here, time refers to the fixed time, with the day, week, and month denoting the time of maximum load during the period. Through the temporal and spatial load analysis, the equipment diagram shows the combination of equipment, load, and demonstration methods. If a line is selected, checking the load conditions of neighboring lines is easy. If a section is selected, the load distribution of the section within the line can be analyzed. If a transformer is selected, the transformer load within the area can be analyzed depending on the colors or object size.

Load density uses three-step brightness. Facilities with a high level of load are shaded. Therefore, it is useful in checking the area with a high level of load when the reduced scale is adjusted. In addition, the high/low-voltage ratio and ratio by contract type for facilities load can be displayed by using the pie chart and bar chart on the facilities, thus enabling statistical and spatial analysis.

Although Internet GIS load analysis enables easier analysis of the load for the system distributed in an actual geographic space, checking the load of all sections for one distribution line is difficult. Therefore, we have developed an SLD per feeder in Active-X form so that the load in various time zones can be analyzed for both line and section through SLD on the web. Fig. 5 shows the SLD wherein one feeder is formulated with the ternary tree and load for the line and section (current, voltage drop, loss, power) is selectively demonstrated and analyzed by changing the time using the time-change button such as “15 minutes ago,” “15 minutes later,” “same time yesterday,” or “same time tomorrow.” Analyzing the load pattern every 15 min on the day, the monthly load pattern for the specific hour and minute as well as the monthly load pattern using the daily maximum load for the feeder and a specific section are also possible.

System and facilities load analysis is a module for performing multidimensional load analysis using the online analysis-processing technique. For the feeder, section, and transformer, the load pattern is analyzed using a table and a graph according to various time zones (15 min, hour, day, week, month, same time zone, latest) and load (load rate, apparent power, active power, reactive power, current, loss, and voltage drop). The analysis is done on a multidimensional basis, such as drill up/down wherein table and graphs are adjusted from the level of summary to the level of detail, pivoting is where time zones and the load axis are changed for analysis, slice/dice is where the subset of multidimensional array is created for analysis, and surfing is where the forms and conditions of reports are changed using the mouse. In case power outage or breakdown is inputted, the analysis is connected to Internet GIS load analysis, SLD load analysis, feeder analysis, and section analysis so that the load at the time of the outage or breakdown can be analyzed.

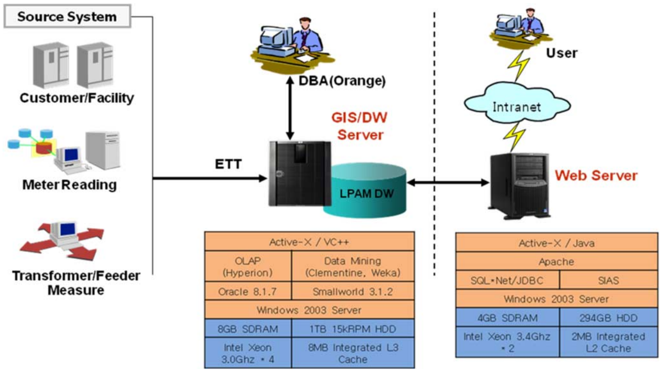

Fig. 6. Development environment.

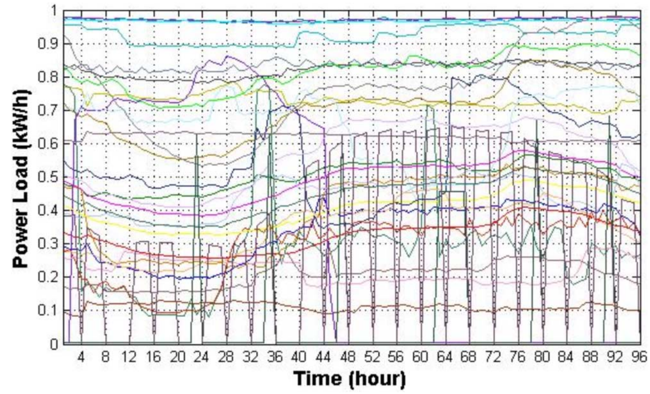

Fig. 7. AMR representative load pattern for January.

In the status/statistics for the feeder and section, the load ratio and apparent power are analyzed by high- and low-voltage and contract type (residential, commercial, educational, industrial, etc.); for the transformer, and analysis is carried out by power and light and contract type. In addition, the maximum/average/minimum values for the load ratio and apparent power and tables and graphs on the status/statistics at different time zones can also be analyzed multidimensionally. It also includes functions, such as load change simulation, where the load is increased or decreased for the system and transformer, load leveling simulation for the line, and load transfer simulation where the load is transferred to the other transformer.

V. Case Study on the Load Pattern Classification

We describe the determinaton of the number of clusters and the evaluation of classification model in the process of generating the load pattern for non-AMR high-voltage customers. In our experiments, we compare our load calculation with the existing measurement value. Also, experimental results show that our load pattern calculation results using data-mining techniques are statistically significant. We used real data from the Gangnam Branch Office of Korea Electric Power Corporation which supplies electricity to the Gangnam-gu and Seocho-gu areas of Seoul city. There are 303 distribution lines, 8 112 sections, 7 089 transformers (including 471 transformers measured by the wireless load monitoring system), 3 349 AMR customers, 792 non-AMR high-voltage customers, and 277 337 low-voltage customers. We experimented with the acquired real data from January to October 2007.

For the experiments, hardware was used in the GIS/DW server and web server, and software was used Smallworld for spatial analysis, Oracle for data warehouse, Hyperion for online analytical processing, and Clementine and Weka for data mining as shown in Fig. 6. The average elapsed time took a total of 45 h from extraction, transformation, transportation (ETT) to the power-flow calculation for processing monthly data of the source system. Daily data, however, took 1 h, 30 min using semiautomatic processing modules.

In cluster analysis, determining the optimum number of clusters k is very important. Thus, we used the reproducibility evaluation method to determine the number of clusters of the AMR representative load pattern. Specifically, we applied the method to utilize the data partitioning technique used in supervised learning, such as neural network, decision tree classification, and regression. Since the repetition of the same clustering method is enabled by data partitioning, a reproducibility evaluation can be performed. The following is the reproducibility evaluation procedure as follows.

TABLE II
Reproducibility Evaluation of the AMR Load Pattern for January

k Value	Number of Test Data	Number of Data out in Mainstream	Percentage of Data out in Mainstream
k=30	1553	41	2.64%
k=31	**1553**	**23**	**1.48%**
k=32	1553	23	1.48%
k=33	1553	114	7.34%

1) First, we partition the training data set into three parts. The ratio is 4:4:2. The larger two data set was used as the training set and the smaller one was used as a test set.
2) Second, run the k-means on two training data set to produce Rule1 and Rule2.
3) Third, apply the Rule1 and Rule2 on the test set and produce a confusion matrix to evaluate the result. If the selected number of clusters is optimal, the matrix will show a strong homologous characteristic.

As a result of the reproducibility evaluation of AMR data in January 2007, a cross-classified table is created as shown in Table II. The k value of 31 is selected as the lowest value of percentage of data deviating from the mainstream. Fig. 7 shows the representative load pattern of 31 clusters for AMR customers in January.

For the class allocation of the load pattern for non-AMR high-voltage customers, a classification model is created using the cluster label generated as a result of cluster analysis and customer property information. For the algorithm to create a model, decision tree (C4.5), Bayesian network, Naïve Bayesian, and SVM are used. In the evaluation of the classification model, the 10-fold cross validation method is applied. During the performance evaluation, the confusion matrix expressed as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) is used and calculated as shown

$$\mathrm{Accuracy} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}. \tag{6}$$

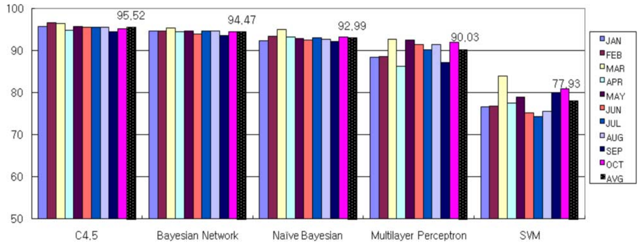

Fig. 8. Performance evaluation of the non-AMR load pattern classification model.

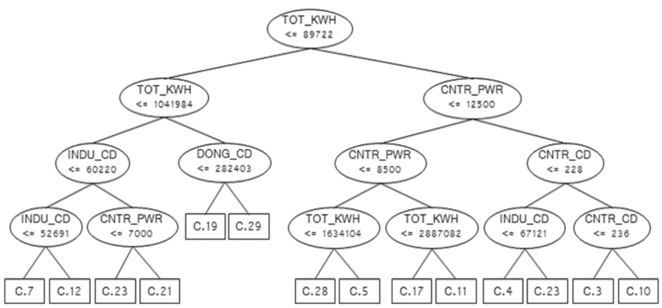

Fig. 9. Decision tree for AMR customers' clusters.

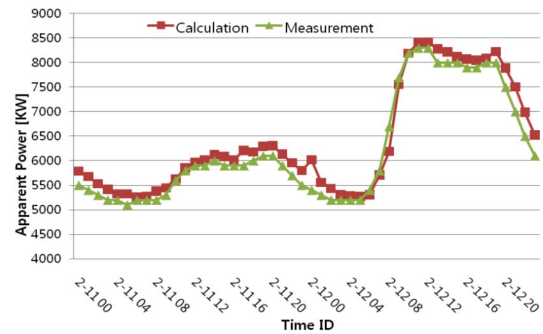

Fig. 10. Comparison of load calculation and measurement value.

As shown in Fig. 8, the accuracy ratio of the C4.5 algorithm is 95.52, the highest compared to other classification models. For the non-AMR load pattern forecast model, the C4.5 decision tree is selected.

Fig. 9 illustrates the partial result created by the decision tree classifier using customer property information for AMR clusters in January 2007. The customer information used here includes contract type code, contract power, power usage code, industrial class code, supply method code, area code, and a monthly meter-reading value. As shown in the figure, the monthly meter-reading information has become a critical factor in the classification standard. Likewise, when non-AMR customer information is inputted using a decision tree organized on a monthly basis, the cluster is determined using the specific customer's information. For example, from the inputted non-AMR customer information, if the monthly meter-reading value (TOT_KWH) is larger than 89 722 kWH, contract power (CNTR_PWR) is larger than 12 500 kW, contract type code (CNTR_CD) is the same or less than 228, and the industrial class code (INDU_CD) is larger than 67 121, then Cluster 23 is assigned.

To verify the power-flow calculation data for the feeder, it is compared with the feeder load data measured in the substation. Fig. 10 compares the two values for 48 h at intervals of 1 h from 00:00 to 23:00 on February 2007.

To measure the similarity of the two values, cosine similarity is calculated as shown

$$\cos(x, y) = \frac{x \cdot y}{\|x\| \|y\|}. \tag{7}$$

Here, $\bullet$ is the vector's dot product; it stands for $x \cdot y = \sum_k x_k y_k$, $\|x\|$ is the length of vector x denoting $\|x\| = \sqrt{\sum_k x_k^2}$. A cosine similarity that nears 1 means that the similarity is high. The actual calculation for the equation above gives a result of 0.9996, which means that the calculation and measurement values for the feeder load have high similarity. In addition, the average of the two values is 202 kW; the feeder load value based on the power-flow calculation has a high degree of accuracy if we assumed that substation measurement values are expressed in 100-kW units.

VI. Conclusion

In this paper, we have developed a model wherein the load of facilities under the distribution system line—which changes every moment—can be analyzed using meter-reading data periodically obtained from power consumers even without attaching measuring equipment to all large-scale power facilities. For the development of this model, data for the load calculation in the legacy system were connected, the load pattern for non-AMR customers was calculated by applying data mining, and power flow was calculated. The load for feeder, section, and transformer can be analyzed using Internet GIS, SLD, multidimensional tables and graphs, and simulation.

We have 135 000 high-voltage AMR customers and data which include all types of contracted power, such as residential, commercial, industrial, educational, and agricultural, etc. We used real data of the Gangnam Branch Office of Korea Electric Power Corporation (KEPCO) which also included all types of contract power. We anticipate that the number of clusters will be much more generated and the forecast accuracy will be improved if all AMR customers are applied. According to our experimental results, our approach is technically practical. The impact of computational burdens is enormous. Nevertheless, it can be solved with large database technology or a distributed system. Our study goal is not a real-time application, but a new load-analysis model. Despite being late one day, this load pattern-analysis model makes it possible to improve the efficiency of facility operation and plan.

Using the temporal and spatial analysis technique enables checking of the current situation and problems, illustrative demand analysis and load management, and checking of spatial-load distribution and density property, as well as the analysis of the relationship and effect between customers, facilities, load, and outages. We expect the model to contribute to accurate power-line load analysis, efficient facilities operation, and quality improvement of the power distribution system.

References

[1] C.-H. Park, J.-O. Kim, and J.-B. Lee, "A study on development and application of computer system for optimal distribution planning and operation." (Final Rep.). KEPRI Res. Rep. TR.98PS02.S2000.697, 2000.

[2] C.-H. Park, S.-M. Park, M.-S. Lee, and J.-E. Jo, "A study on the optimal load management of distribution transformer." (Final Rep.). KEPRI Res. Rep. TR.00PJ06.J2003.120, 2003.

[3] V. Figueiredo, F. Rodrigues, Z. Vale, and J. B. Gouveia, "An electric energy consumer characterization framework based on data mining techniques," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 596–602, May 2005.

[4] G. Chicco, R. Napoli, P. Postolache, M. Scutariu, and C. Toader, "Customer characterization options for improving the tariff offer," *IEEE Trans. Power Syst.*, vol. 18, no. 1, pp. 381–387, Feb. 2003.

[5] S. V. Verdu, M. O. Garcia, C. Senabre, A. G. Marin, and F. J. G. Franco, "Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1672–1682, Nov. 2006.

[6] G. Chicco, R. Napoli, F. Piglione, P. Postolache, and M. Scutariu, "Load pattern-based classification of electricity customers," *IEEE Trans. Power Syst.*, vol. 19, no. 2, pp. 1232–1239, May 2004.

[7] A. Ugedo, E. Lobato, J. Peco, and L. Rouco, "Decision trees applied to the management of voltage constraints in the spanish market," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 963–972, May 2005.

[8] D. Gerbec, S. Gasperic, I. Smon, and F. Gubina, "A methodology to classify distribution load profiles," in *Proc. IEEE/Power Eng. Soc. Transmission and Distribution Conf. Exhibit.: Asia Pacific*, Oct. 2002, vol. 2, pp. 848–851.

[9] D. Das, H. S. Nagi, and D. P. Kothari, "Novel method for solving radial distribution network," in *Proc. Inst. Elect. Eng., Gen., Transm., Distrib.*, Jul. 1994, vol. 141, no. 4, pp. 291–298.

[10] J. Liu, B. Pengxiang, Z. Yanqing, and W. Xiaomeng, "Power flow analysis on simplified feeder modeling," *IEEE Trans. Power Del.*, vol. 19, no. 1, pp. 279–287, Jan. 2004.

[11] H.-C. Wu and C.-N. Lu, "Data mining approach for spatial modeling in small area load forecast," *IEEE Trans. Power Syst.*, vol. 17, no. 2, pp. 516–521, May 2002.

Jin-Ho Shin received the B.S. degree in computer science from Hanbat University, Daejeon, Korea, in 1996 and the M.S. and Ph.D. degrees in computer science from Chungbuk National University, Cheongju, Korea, in 2004 and 2010, respectively.

Currently, he is a Senior Researcher with the Korea Electric Power Research Institute. His principal research areas include geographic information systems, pattern recognition, data mining, power load analysis and forecasting, and mobile computing.

Bong-Jae Yi received the B.S. degree in mathematics from Sung Kyun Kwan University, Seoul, Korea, in 1982.

In 1985, he joined Korea Electric Power Corporation as a Programmer and System Analyst. He is part of the principal technical staff of the Information Communications Group at the Korea Electric Power Research Institute. His principal research areas include geographical information systems, auto recognition, mobile, radio-frequency identification, and ubiquitous sensor network.

Young-Il Kim received the B.S. and M.S. degrees in computer engineering from Chungnam National University, Daejeon, Korea, in 1998 and 2000, respectively.

His employment experience has included the K4M Co, Electronics and Telecommunications Research Institute, and Korea Electric Power Research Institute. His principal research areas include power facility monitoring, mobile computing, geographic information systems, data mining, and Power IT.

Heon-Gyu Lee received the M.S. and Ph.D. degrees in computer science from Chungbuk National University, Cheongju, Korea, in 2005 and 2009, respectively.

He is a Senior Researcher with the Electronics and Telecommunications Research Institute, Daejeon, Korea. His principal research areas include data mining, bioinformatics, temporal databases, spatiotemporal databases, ubiquitous computing and stream data processing, and knowledge-base information retrieval.

Keun Ho Ryu (M'85) received the Ph.D. degree from Yonsei University, Seoul, Korea.

Currently, he is a Professor with Chungbuk National University and a Leader of Database and Bioinformatics Laboratory, Cheongju, Korea, in 1988. He served the Korean Army as ROTC. He was a Postdoctoral Researcher and Research Scientist at the University of Arizona, Tucson, and the Electronics and Telecommunications Research Institute, Daejeon, Korea. His research interests include temporal databases, the spatiotemporal database, temporal geographic information systems, ubiquitous computing and stream data processing, knowledge-base information retrieval, database security, data mining, bioinformatics, and biomedical.

Dr. Ryu has served on numerous program committees, including Demonstration Co-Chair of the International Conference on Very Large Data Bases, the PC committee member of APWeb, and the Advanced Information Networking and Applications. He has been a member of the Association for Computing Machinery since 1983.