Spatio-Temporal Mining for Power Load Forecasting in GIS-AMR Load Analysis Model

Heon Gyu Lee, Yonghoon Choi, Jin-ho Shin
Electronics and Telecommunications Research Institute
hg_lee@etri.re.kr

Yonghoon Choi
Electronics and Telecommunications Research Institute
ychoi@etri.re.kr

Jin-ho Shin
Korea Electric Power Research Institute
jinho@kepri.re.kr

doi: 10.4156/ijipm.vol2. issue1.7

Abstract

A spatio-temporal mining technique is used to predict power load patterns for a voltage transformer. It is applied from load data measured every thirty minutes and a GIS-AMR database collected by a transformer’s load measurement system over a wireless network. The proposed approach in this paper consists of three stages, (i) data preprocessing: noise or outlier is removed and the continuous attribute-valued features are transformed to new features (feature extraction and discretization), (ii) cluster analysis: SOMs (Self Organizing Maps) clustering is used to label the class and (iii) classification: we used and evaluated classification rules using spatio-temporal mining to build a suitable load forecasting model. In order to evaluate the result of classification, derived class labels from clustering and other features are used as input to build classification rules including time and spatial factors. Lastly, the result of our experiments is presented.

Keywords: Load Pattern Analysis, Spatio-temporal Mining, Load Forecasting

1. Introduction

Currently an automated methodology based on data mining techniques is presented for the classification of load patterns in load profiles. Load pattern discovery attempts to identify existing load patterns and recognize new load classification methods, employing methods from sciences such as statistical [1] and data mining techniques [2], [3].

In this paper, we wish to analyze the load on a voltage transformer’s system over time and space. A model for such an analysis can be taken from that of electrical load distribution network systems. A correctly analyzed model for forecasting power load is required to manage and plan a power distribution system. Load forecasting in the electrical supply industry helps decision making. Many variables can be considered. For instance, these include the cost of electricity, product, load switching, and upgrading a utility. In particular, it is important to forecast the for electricity suppliers, ISO Company, and financial agency for marketing concerned with electrical energy, transmission and distribution. Therefore, it is necessary to predict power load and the load pattern for efficient operation and planning in the electric industry. Power load data is spatio-temporal information which have timestamp and location information. However, previous researches such as clustering, classification and regression usually did not consider such temporal and spatial factors in load data or applied as static factors. Since the power load patterns have time-varying characteristic and vary in different localities, those give rise to the uninformative results if only traditional data mining is used. Therefore, if we consider time intervals under multiple time granularities and spatial information in order to forecast load patterns to load demand data analysis with spatio-temporal dimension, It is possible to discover useful and accurate load patterns during the given time interval.

A spatio-temporal pattern mining technique [4] is used to forecast load patterns for a voltage transformer. It is applied from load data measured every 30 minutes and a GIS-AMR database collected by a transformer’s load measurement system over a wireless network [5]. That is, the representative load pattern of the transformer is derived from load data checking the load on the transformer’s system by wireless. Next, the transformer’s load is predicted. The following steps are undertaken:
- Derive the transformer’s daily load patterns from load data measured every 30 minutes by a wireless load measuring system.
- Extract the feature and location data of the transformer from the GIS-AMR database.
- Create representative load patterns by clustering to form daily load patterns.
- Generate classification rules using spatio-temporal mining that considers both cyclic temporal patterns and spatial properties.
- Evaluate classification rules to build suitable prediction model.

2. Extraction of transformer load and spatio-temporal information

Load pattern data are derived every 30 minutes for the transformer. This is measured by a wireless load measuring system. Available power and time ID are extracted from the wireless load data. Temporal attributes of the time relation are used to extract time information of the transformer’s available load measured every 30 minutes by a wireless load check system. The customer information relation and voltage transformer information relation from a database of the GIS-AMR system are used to extract of the following data: transformer ID, load local property, total capability of each transformer, and sum of the transformer power load. The extracted power load record is restructured as a daily vector representation using the following formulation.

\[
V^{(t)} = \{V_0^{(t)}, \ldots, V_h^{(t)}, \ldots, V_{2330}^{(t)}\}
\]

(t: transformer, 0<h<2330, H=2330)

In data preprocessing phase, outlier detection method is applied to remove all outliers. Figure 2 shows the example of outlier detection.
In clustering phase, SOMs [6] is used to group the input data which is as shown in figure 3 and the optimal clusters are obtained. The use of clustering in this step detects the number of classes as an input of the classification model. In order to evaluate the performance of the clustering algorithm, adequacy measure (MIA: Mean Index Adequacy [2]) is applied. The purpose of adequacy measure is to obtain separated and compact clusters that make the load patterns well identified. Let’s suppose a set of $M$ load patterns separated in $k$ clusters with $k=1,\ldots,K$ and $K$ is the total number of clusters. Each cluster center is formed by a subset $C_{(k)}$ of load patterns, where $r_{(k)}$ is a pattern assigned to cluster $k$. MIA is defined as the average of the distances between each input vector assigned to the cluster and its center.

$$MIA = \frac{1}{K} \sum_{k=1}^{K} d^2(r_{(k)}, C_{(k)})$$  \hspace{1cm} (2)

The SOMs algorithm was used to generate class labels based on the MIA measure. For the data set of January, the obtained results are showed in Figure 3 and Figure 4. We selected the number of clusters as 21 (3 by 7) by considering the MIA. For achieving the best result of clustering, we set the options for SOMs as shown in Table 1. The number of cluster for other months is following: Feb.=20, Mar.=19, Apr.=23, May=21, Jun.=24, Jul.=22, Aug.=20, Sep.=19, Oct.=24, Nov.=23, Dec.=22.

<table>
<thead>
<tr>
<th>Table 1. Option for SOMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning rate decay</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Rough Estimation Phase</td>
</tr>
<tr>
<td>Tuning Phase</td>
</tr>
</tbody>
</table>

![Figure 3. Mean Index Adequacy for January](image-url)
Spatio-Temporal Mining for Power Load Forecasting in GIS-AMR Load Analysis Model
Heon Gyu Lee, Yonghoon Choi, Jin-ho Shin
International Journal of Information Processing and Management. Volume 2, Number 1, January 2011

21 clusters (January)

Figure 4. An Example of transformer’s representative load profiles

A transformer load pattern has 48 load points in a day. If we use that pattern in the mining process, performance is decreased by the presence of many feature values. If time and space are added to the model, then it is impossible to predict the load pattern, due to the size of input vector. In this paper, dimensionality reduction is performed by extracting load factors represented in a load pattern. The load factors are defined by an electrical expert, as the key features of load patterns. The load factors are defined by the following formula and include the four areas shown within Figure 3.

\[
\text{Early Morning (0h~7h): } LF_1 = \frac{1}{24} \sum_{i=1}^{24} P_{\text{avg of early_morning}}
\]

\[
\text{Morning (9h~12h): } LF_2 = \frac{1}{8} \sum_{i=8}^{16} P_{\text{avg of morning}}
\]

\[
\text{Afternoon (13h~17h): } LF_3 = \frac{1}{6} \sum_{i=17}^{23} P_{\text{avg of afternoon}}
\]

\[
\text{Night (19h~23h): } LF_4 = \frac{1}{6} \sum_{i=24}^{30} P_{\text{avg of night}}
\]

Temporal and spatial property, and extracted input factors are shown in Table 2. Continuous load factors, which are real numbers transformed to nominal values to perform discretization based on entropy [7], are applied for the spatio-temporal mining. Figure 6 is an example load analysis model generation. We retrieve information on the transformer’s location (Location ID) to obtain the spatial features of the transformer. Features extracted include: load location characteristic, light customer’s count, and power customer’s count.
In this chapter, we describe a process to generate classification rules to be added to temporal and spatial characteristics.

3.1. Time pattern [8] formulation for cyclic pattern discovering of load patterns

We apply a time pattern formulation that is able to represent periodic time to data in attributes of Table 2. Time pattern expression formulation is defined by a set of unit time concept. It is represented by a formulation derived for the user. For example, if the set of time unit is assumed to <month, week, day>, a meaning of time pattern <1,1,1> is “Monday of first week in January”, and <2,2,4> is “Thursday of second week in February”. The character “*” is defined by the entire time unit and includes the entire domain values to represent the cyclic pattern. For example, an expression of <1,*,3> is interpreted as “Wednesday of every week in January”, and <*,*,3> denotes “every Monday for each week in every month”. A time pattern that does not contain the “*” defined as the basic time pattern.

3.2. Generation of classification rules considering attribute of time and location

We extract the rules from the training data set of Figure 6, to generate the classification rules. A rule’s form R resembles the following.
$R: p_1 \land p_2 \land \ldots \land p_n \Rightarrow C$ (7)

$p_1 \land p_2 \land \ldots \land p_n$ is the attribute value on attributes of Figure 6, and $C$ is the class label called the representative load pattern (cluster).

The entire step of the algorithm is performed as in the following:

1. Total pattern satisfying the rule’s support is extracted for each of the basic time patterns. Rule count, the frequency pattern $p_1 \land p_2 \land \ldots \land p_n$ occurred and the class $C$ simultaneously from the relevant basic time pattern.

$$
support_i = \frac{\text{RuleCount}}{|D_i|}$$ (8)

2. All of the rules satisfied with a high level of correctness to the confidence rule are extracted. The rules are ordered in descending order of confidence.

3. To discover cycles, the entire extracted rules are updated to the time pattern including “*”, because they have the basic time pattern. Figure 7 shows the updating algorithm.

```
input Time pattern “*”, a set of rules with basic time pattern
output a set of rules with time pattern “*”
if time pattern “*” is updated for the first time then
    $R_k$ with “*” = $R_k$ with basic time pattern;
    update count = 1;
else
    for each $R \in R_k$ with basic time pattern - $R_k$ with “*”
        $R_k$.update count = 1;
    end
    for each $R \in R_k$ with basic time pattern $\cap R_k$ with “*”
        $R$.update count++;
    end
    $R_k = \{R$.update\}
end if
```

Figure 7. The update algorithm for the basic time pattern

Rules for each $i_{th}$ basic time pattern has an individual counter. If a step is in the initial update phase (lines 1–4) or a rule is first updated into basic time unit (5–7 line), then the counter is set to 1. Otherwise (lines 8–10), the current counter is incremented by 1. If the relevant time unit has the entire domain values, a pattern representation of the time unit is set to “*” to represent the periodicity. For example, if a set of the time units are defined to <month, week, day> for rule $R$, $R$: <2, *, 1> denotes rule $R$ instances are {<2,1,1>, <2,2,1>, <2,3,1>, <2,4,1>} within the basic time patterns. The total time patterns and classification rule are first generated. Then, rules that had the total time patterns contained within the basic time and the instance value of the test data for the test data set assigned to the representative load patterns are selected first.

4. Experiment and evaluation

Load data collected by the wireless load measurement system of the transformer is applied from the GIS-AMR data warehouse built by the Korean Electric Power Research Institute. Data from the Korean capital, Seoul, from January 2007 to December 2007 is used in this experiment. In the experiment, we evaluate the performance of classification rules and compare our classification method with Naïve Bayesian [9] and state-of-art classifiers. The widely known decision tree induction C4.5
Spatio-Temporal Mining for Power Load Forecasting in GIS-AMR Load Analysis Model
Heon Gyu Lee, Yonghoon Choi, Jin-ho Shin
International Journal of Information Processing and Management. Volume 2, Number 1, January 2011

[10]; a function-based classifier, SVM [11], a recently proposed classifier extending NB using long itemsets. The accuracy was obtained by using the methodology of stratified 10-fold cross-validation. The parameters of the four classification methods were set as follows. For our proposed algorithm, the minimum support was set to 0.4%, the minimum confidence to 70%. For the SVM, the soft margin allowed errors during training. We set 0.1 for the two-norm soft margin value. The Bayesian classifier and C4.5 parameters were default values. We tested the C4.5 tree method. In order to evaluate the classifiers' performance, an accuracy criterion was used. We would like to access how well the classifier can classify. For this purpose, the sensitivity and specificity measures were used and accuracy is defined as

\[
\text{Accuracy} = \frac{\text{Positive}}{\text{Positive} + \text{Negative}} + \frac{\text{spec.}}{\text{Positive} + \text{Negative}}
\]

(10)

\[
\text{sens.} = \frac{\text{True Positive}}{\text{Positive}}
\]

(11)

\[
\text{spec.} = \frac{\text{True Negative}}{\text{Negative}}
\]

(12)

The result is shown in Figure 9. Figure 10 also shows the mean of classifier error rate (RMSE: Root Mean Squared Error). As expected, our classification method was more accurate than the other classifiers.

Figure 8. Load presents state and density of section and transformer at GIS-AMR
Spatio-Temporal Mining for Power Load Forecasting in GIS-AMR Load Analysis Model
Heon Gyu Lee, Yonghoon Choi, Jin-ho Shin
International Journal of Information Processing and Management. Volume 2, Number 1, January 2011

![Classifier Accuracy](image1)

**Figure 9.** Comparison of classifier accuracy

![RMSE](image2)

**Figure 10.** Comparison of classifier error

Figure 11 is an example of a time pattern expression that includes periodicity “*”. In Figure 11, the upper graph (‘●’) is predicted in the 9th cluster as the Monday load pattern and is interpreted as “Monday of every week in April”, since these contain a time pattern <4,*,1>. The lower graph (‘▲’) contains the two weekend load time patterns: <3,*,6>, <4,*,7>, referring to respectively “Saturday of every week in March” and “Sunday of every week in April”.

- 64 -
It is very important to forecast accurate load for electric utilities in a competitive environment created by the electric industry deregulation. In this paper, the proposed main mining tasks include clustering method and spatio-temporal mining technique. Cluster analysis is used to define load pattern classes and representative load profiles for each class. Classification rules based on spatio-temporal mining use representative load profiles to construct a classifier able to assign different load patterns to the existing classes. The proposed classification method discovers electric load patterns in multiple time granularities and different localities. In the experiment, the applied SOMs and spatio-temporal classifier had been tested KEPRI GIS-AMR transformer data and discovered interest load patterns. As a result, spatio-temporal mining based classification (gave about 84%-92% of goodness of fit) outperformed the other classifiers.

5. Acknowledgments

This work was supported by the Postal Technology R&D program of MKE/IITA. [2006-X-001-02, Development of Real-time Postal Logistics System] and development of AMR system interfacing model on internet GIS environment project of the Korea Electric Power Research Institute (KEPRI).

6. References


