Preventing Non-Technical Loss: Researching and Assessing Power Usage Behaviour of Commercial and Industrial Units (Loading Intelligence - @LI)

I. Background

Within the operation and business process of a power system, beside the portion of electricity used for power transmission and distribution, technical loss consists of:
+ Power usage of power plants and substations.
+ Power loss due to heating of electrical resistors and instruments.
+ Power loss due to interaction of high voltage systems with the environment.

There is another type of loss termed 'non-technical loss' (or 'commercial loss'), which consists of:
+ Bad debts and delayed payments.
+ Loss due to metering system errors.
+ Loss due to faulty design, installation, and operation of metering circuits.
+ Loss due to theft.

Loss reduction has always been a top business priority for all power companies in the world generally and in Vietnam particularly. Within the scope of this article, we focus only on the possibility of non-technical loss reduction through the use of a research and assessment tool on Commercial and Industrial (C&I) customers' power usage behavior. By detecting abnormality in customers' process of power usage, power companies will be able to devise appropriate measures in preventing and reducing power loss.

II. Databases and Algorithms

Currently, regarding EVN power network, the process of collecting data from meters are still primarily done manually at the frequency of 1, 2, or 3 times per month. Due to the small amount of collected data that are essentially unrelated, it is difficult to evaluate and detect abnormality in this metering system and in customers' power usage behavior. Usually, only by reading month electricity bill and only when there are obvious abnormal changes, can business staffs pose questions on load usage.

In 2009, ATS Co., Ltd. began a pilot telemetry project with electronic meters (A1700 and Landis+Gyr) of 30 major customers of Vinh Phuc Power Company (see "Putting into Operation Remote Metering System in Vinh Phuc Power Company", Electricity Magazine, October 2009). The project, using ATS product @IMIS, collected and stored a large amount of data with sufficiently small resolution (1 minute/time with Landis+Gyr and 10 minutes/time with A1700), creating an ample database for load research and assessment. The database used in this project was Plant Information (PI), manufactured by OSI Inc. (USDA), and was developed for the storage of different data types generated by continuous processes, such as the continuous operation of power plants, power system, oil refinery, and chemical plants. Metering data collected and stored by PI historical data management system was validated and corrected (Validation – Estimation - EroCorrection). This ensured the accuracy level required in inputs for a load behavior research program named Loading Intelligence (@LI) and developed by ATS Co., Ltd.

In researches on trends and behaviours of large databases, we often need to use statistical classification algorithms, such as Decision Tree, Support Vector Machine, Bayes Naïve, and Artificial Neural Network. These are collectively called the "Artificial Intelligence", assisting in the classification and detection of data abnormality. In addition, the research process on typical load shapes is also widely used, with the main objective being indentifying load shape. Power companies in Taiwan and Malaysia have also conducted research programs on power usage behaviour by classifying loads that have common characteristics, the data of which are collected from 1 to 2 times per month. Based on loads of the same types, the programs perform computations in order to discover which customers are questionable in terms of their power usage.

In the case of the @IMIS system ATS deployed at Vinh Phuc Power Company, customers equipped with metering data collection and management devices were industrial customers with high monthly power consumption, each of whose power usage also differed. Consequently, in order to improve accuracy, the computation solution was directed to each customer (load) separately. The algorithm served in studying the loading pattern of each customer at different times of the day and different days of the month. The computation results served to identify the time interval in which occurred abnormality in the customers' power usage pattern.

III. Algorithm Description

After studying and evaluating the strengths and weaknesses of different algorithms and by reviewing research results of several authors in the area, it was decided that "Bayes Naïve" was to be use as the core algorithm of the program.

The basic block diagram of the program's 7-step computational procedure is displayed in Figure 1.

1. Data was input into the database of @IMIS after being processed through V-E-E.
2. When it was required to observe a specific customer's power usage process, data was input for each day (the total amount of power consumed in different phases) with a resolution for a certain time. For instance, the resolution of 10 minutes was in used for the current computation.

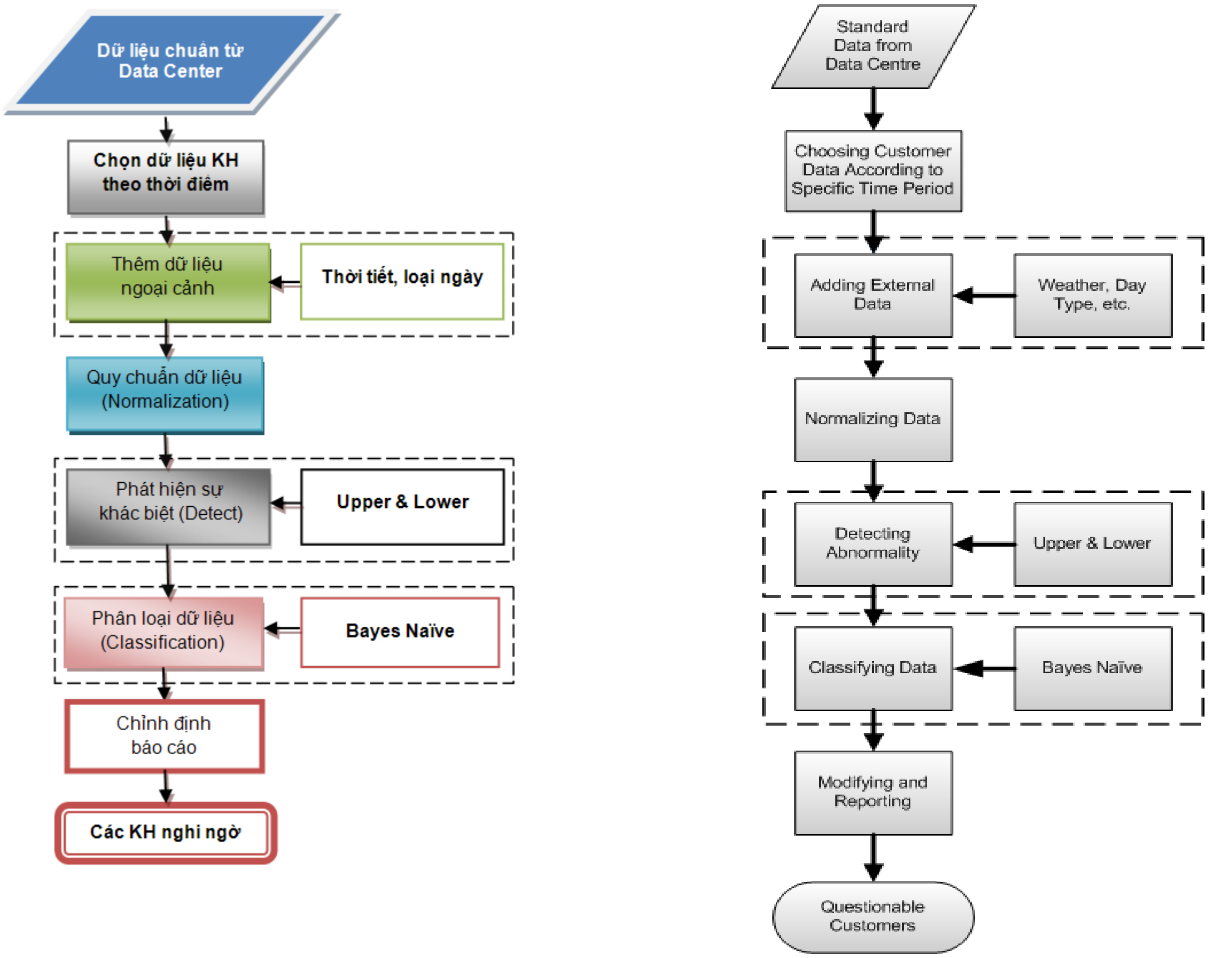

Figure 1: Basic Block Diagram of Computational Steps of Loading Intelligence-@LI

3. External data, such as the day of the week and weather conditions, was added. These data needed to reflect factors that influence how loads were classified. During classification, the below characteristics were often considered:
   - Type of load unit: industrial, commercial, or household.
   - Geographical area: urban or rural.
   - Voltage level: high, medium, or low
   - Season, or to be precise, weather: hot, cool, or cold
   - Day feature: working days, weekends, or holidays

   When computing, the first three characteristics was ignored as this research examined each consumption unit separately. Day feature was categoried into seven day types within a week.

4. The data normalization process was started by converting power consumption data into values without a dimensional unit and varying between 0 and 1 inclusively. This was done by dividing the value of each row by its maximum value; values of other rows were also processed in a similar fashion.

$$X_i = \frac{A_i}{A_{max}} = \frac{P_i \cdot \frac{10}{60}}{P_{max} \cdot \frac{10}{60}} = \frac{P_i}{P_{max}}$$

5. The next step was considered as the first step in filtering problematic data. In this research, day feature was divided into 7 types from Monday to Sunday. Corresponding to each day type was created 2 boundary lines, the Upper Boundary and the Lower Boundary

$$\text{Upper, } L_1 = \overline{X} + k.\sigma \qquad \overline{X} = \frac{1}{N}.\sum_{1}^{N} X_i \qquad \sigma = \sqrt{\frac{1}{N}.\sum_{1}^{N}(X_i - \overline{X})^2}$$

$$\text{Lower, } L_2 = \overline{X} - k.\sigma$$

In the computations, the coefficient k takes one of the following values, [1.3, 2, 4], corresponding to the accuracy of the data capturing process in separate areas. The smaller the value of k, the lower the amount of data captured but the higher the level of accuracy and vice versa.

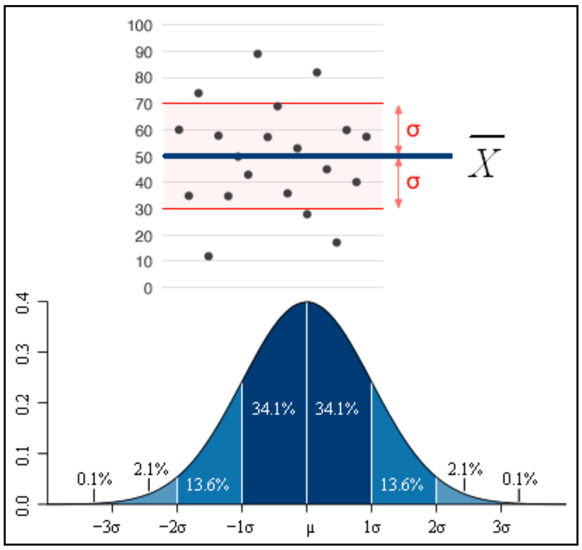

Figure 2: Value of k and its effect

Figure 2 illustrates the effects of k value on the accuracy and frequency of data of different values.
After data was classified by the two boundary lines Upper and Lower, the data was labeled. Label 0 was applied to data lying outside of the boundaries, and Label 1 was applied to data lying within the 2 boundaries. The results is shown in the below figure.

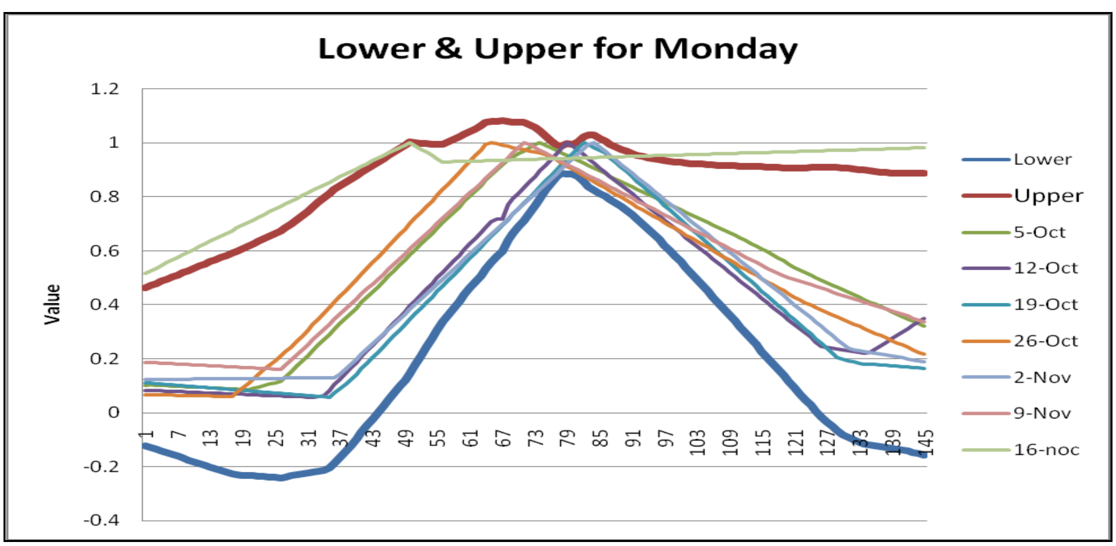

Figure 3: Upper and Lower Model Graph for Input Data

6. The next step involved combining the above analytical results with auxiliary data and using the statistical classification algorithm Bayes Naïve to detect problematic days.
   The input data for Bayes Naïve consisted of 3 types:
   - Day Feature Data (Date): the day of the week, from Monday to Sunday.
   - Weather Characteristic Data (Atmos): categorized into to 3 types of temperature - hot, cool, or cold.
   - Data: labeled 0 or 1, a result of step 5

$$\boxed{\begin{array}{c} P = P(1)*Date(1)*Atmos(1) \\ N = N(0)*Date(0)*Atmos(0) \end{array}}$$

   Where: P value represented good behavior and N abnormal behavior.
   When N>P, the subject was considered problematic, and vice versa.

7. After the data processing and computating steps of 1 to 6, the results was output in an Excel file, reporting days in which power consumption was questionable. The report was accompanied with computational parameters and necessary values required for supplying the researchers with useful information.

   In addition, the tool also had a Graphics Module which provided users with visual evidences of days in which load usage deviated out of the 2 boundary lines.

IV. Load Intelligence - @LI Results

The next section presents actual results obtained by @LI program in load research and analysis.

<u>*Computation Module:*</u>

Figure 4 displays the layout of the actual computation module.

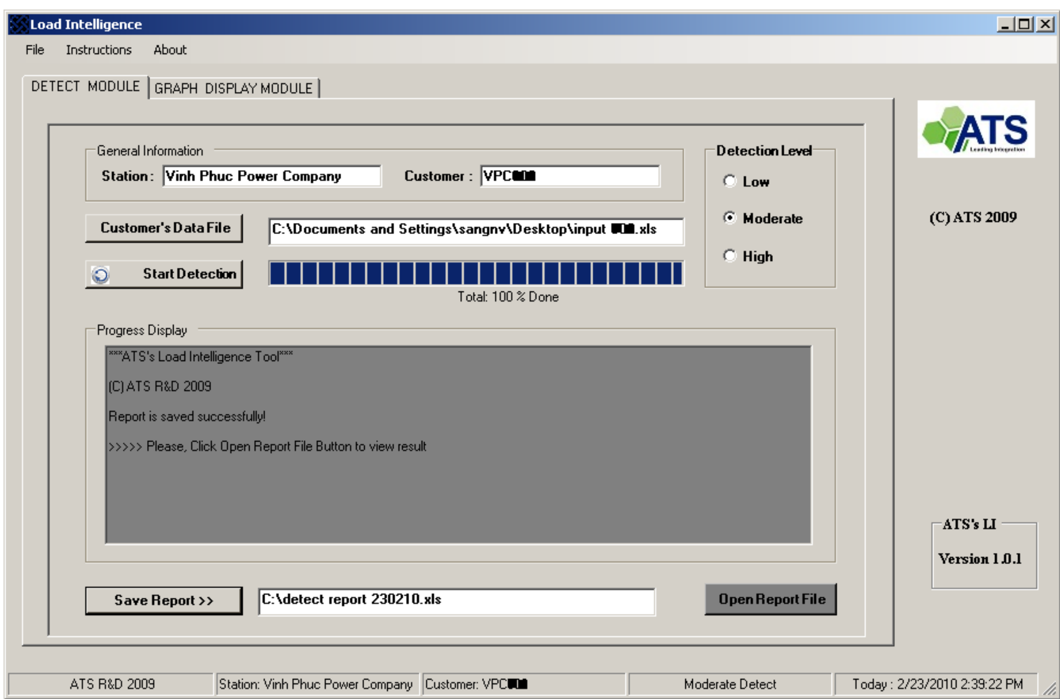

Figure 4: @LI computational module Layout

The input data was updated by the PI access tool of @IMIS in Excel (*.xls) files and was supplemented with Day Feature Data (Monday, Tuesday,…, Sunday) and Weather Characteristic Data (hot, cool, cold). Loading Intelligence then performed its computation process following the algorithm steps described above.

Graphic Display Module:

Figure 5 shows a graphic display of computation results.

There were 3 levels of data classifications corresponding to different Upper and Lower values, consistent with the accuracy level of different situations.

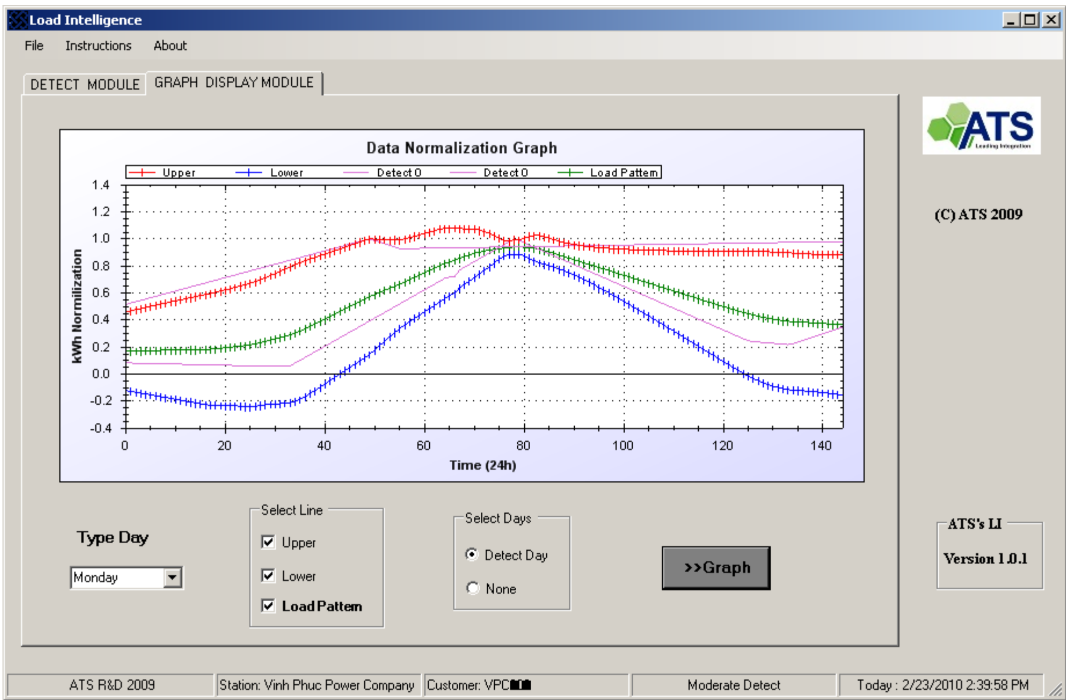

Figure 5: Graphic Display Module

Report-Generation Module:

Finally, @LI generated an analytical report in the form of an Excel (*.xls) file. The main content of the report included questionable days to be furthered examined by business staffs. Furthermore, data on weather condition, reliability level, and Bayes Naïve computation results (P and N) were also reported in order to further assist business staffs. The ability to pinpoint questionable power consumption period will become a useful supporting tool for the accurate evaluation of load or the metering system. The below Figure 6 illustrates a specific @LI Results Report.

Applied Techincal Systems Company — Load Intelligence Tool

R&D Department — (C) ATS 2009

DETECTION REPORT

Detection Time : 2/23/2010 2:45:53 PM

Station : Vinh Phuc Power Company

Customer : VPC[redacted]

Detection Level : High

Notice

- STATE is the state of weather, COLD, FRESH or HOT - depend the asmotphere's temperature

- C, S is the probability result to compare. Result C for Clear and S for suspicious case

- If S > C the Consumption is suspicious, and C > S for Clear Consumption Case)

<<Result List for Suspicious Consumption Day>>

STT	DATE	DAY	STATE	C	S	Days (Load)'s Status
001	13-Nov	Fri	Cold	0.0194	0.0213	Suspect
002	20-Nov	Fri	Cold	0.0194	0.0213	Suspect

Figure 6: Load Intelligence Report Module

V. Program Advantages

The program can utilize the advantages of @IMIS – a load data collecting, processing, storing, and retrieving integrated system – with a large volume of highly accurate low-resolution real-time data in creating input for @LI. It also uses the statistical classification algorithm Bayes Naïve, popular in the analysis of large amounts of data. @LI is suitable for computing with the large loads of industrial, commercial and household units at different voltage levels, independent of geographical areas and other load characteristics.

The program is simple to use as it operates in Microsoft® Excel environment where users can review intermediate results in graphic and tabulated formats. From Excel environment, users can switch to any other types of working environment in order to compile reports required by power management and loss reduction activities.

With computation results pointing out questionable time period, business staffs with their experiences can devise effective measures to timely resolve issues, reducing non-technical loss.

VI. Issues to be Considered

In order to achieve results of the highest accuracy, it is best that the historical database system be created right at the time when metering systems are installed. This is done to ensure the assumption that the metering systems are in perfect condition and customers have not exhibited any abnormal behaviours.

In time, other algorithms (Support Vector Machine, Artificial Neural Network, etc.) should be further developed, creating other options for computation method and allowing for result comparison, thus improving the accuracy level of computation results.

As a final note, the current computation results are only valid as a reference and a warning. They serves as a starting point for more detailed assessment plans in reducing non-technical loss.