JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

ANOMALY DETECTION BASED ON VERY SHORT-TERM LOAD FORECASTING CONSIDERING SIGMOIDAL THRESHOLD

Vipin Kumar Er Gurpreet Kaur
M.Tech Student, Guru Kashi University, Talwandi Sabo
Assistant Professor, Guru Kashi University, Talwandi Sabo
vipinahuja1008@gmail.com
ergurpreet88@ gmail.com

Abstract- In contemporary power systems, technology has made it possible to mine, store and analyze data even at the residential level. As data analysis at the residential level has many more complexities than commercial and industrial levels. The electrical load consumption temporal dataset of a residential consumer has distinct features, owning to their stochastic nature. The data distribution has salient features like it does not follow a normal distribution, which is an assumption for various popular anomaly detection techniques. Hence, increase complexities in anomaly detection. There is a wide range of appliances in a household and the higher rating appliances are used seldomly. These points are highly vulnerable to false detection as an anomaly**.** Therefore, this thesis presents a novel approach to detect anomalies, which is specific for residential customer's load profiles.

Keywords:VSTLF,Intrusion detectionSystem,loadprofiles

I. INTRODUCTION

The conservative electricity setup has four important sectors: generation, transmission, distribution, and consumption. Key features of every of the part are briefly examined in the context of the description of the different roles of Demand Side Management to upgrade the efficiency of operation and investment in the setup

Anomaly detection is the technique of locating unknown items or events in value sets, which unique from the norm. And anomaly detection is repeatedly applied to unknown data which is known as uncontrolled anomaly detection. Anomaly detection has 2 basic guesses Anomalies only occur very normally in the data.

An anomaly-based intrusion detection system (IDS) is any setup designed to locate and preserve malicious activity in a computer network CN. This is also sometimes called network behavior anomaly detection, and this is the type of online monitoring network behavior anomaly detection methods are produced to give a response. Very short-term load forecasting (VSTLF) gives load forecasts up to every day
ahead. Through the power industry, forecasts are typically used by utilities and grid mechanics for real-time scheduling of electricity-producing, load frequency control LFC, and demand response DR. The VSTLF are also necessary to business tasks of retailers, power marketers, and trading companies.

Very short term load forecasts VSTLF is frequently viewed as a minor problem of short-term load forecasting (STLF), vastly because pair can take weather forecasts as the excitation for the forecasting time. STLF has been extensively studied over the earlier many decades, as summarized by many reviews written. A recent upgrade on STLF was across the Global Energy Forecasting Competition 2014.

Several STLF models, like regression models and artificial neural networks (ANN), can be used for

VSTLF. Nevertheless, to attain high precision in the very short horizon, it should be analyzed that the changes between VSTLF and STLF in practice are two-fold. From the modeling position, VSTLF models can depend on delayed load as an independent variable in adding to others like weather and calendar changes that are commonly used in STLF. From the execution perspective, VSTLF needs the model to be estimated rather quickly to process the forecast in time. The short lead time also challenges the data collecting process. Although the smart grid technologies now a day have made it feasible to push current load information to the operation room, several power firms still do not have access to high-quality load values of the most current hour when forecasting the load of the next hour.

II.PROBEM DEFINING

The electrical load consumption temporal dataset of a residential consumer has distinct features, owning to their stochastic nature. The data distribution has salient features like:

which is an assumption for various popular anomaly detection techniques. Hence, increase complexities in anomaly detection.

- There is a wide range of appliances in a household and the higher rating appliances are used seldomly. These points are highly vulnerable to false detection as an anomaly.

These suspicious points should be analyzed about the density of these points and variance of the data

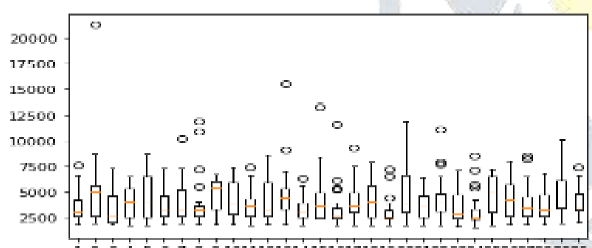


Fig 1. Box plot of a household for a month

distribution. Behavior analysis should be conducted and compared with these points. Due to the salient features of the data distribution of a residential consumer, special techniques for anomaly detection should be reviewed and new techniques should Proposed.

2.2 Objectives

To study and analyze the demand side management techniques for non- technical losses

To study and analyze the impact of non-technical losses on the electricity consumption profile of the end-user

To study and analyze various techniques to find anomality in the electricity consumption behavior of the customer.

To study and analyze impact of various thresholds for anomality detection.

III Research Methodology

Firstly, all the given sector d identified,

$$X_{td} = [x_{1d}, x_{2d}, x_{3d}, x_{4d}, \dots \dots \dots \dots \dots_{(k-1)d}, x_{kd}] \quad (1)$$

$$M_{id} = [m_{1d}, m_{2d}, m_{3d}, m_{4d}, \dots \dots \dots \dots m_{(k-1)}, m_{kd}] \quad (2)$$

$$Max_d = [mx_1, mx_2, mx_3, mx_4, \dots \dots \dots mx_{(k-1)}, mx_k] \quad (3)$$

$$Std_d = [std_1, std_2, std_3, std_4, \dots \dots \dots std_{(D-1)}, std_D] \quad (4)$$

Where,

X_{td} is the load points at *t* time period and *d* day,

M_{id} is the vector of suspicious points with the criteria, $1.5*Q_3$,

Q_3, is 3rd interquartile range *i.e.* 75% of data distribution,

Max_d is the vector of max points of each d day,

Std_d is the vector of standard deviation of each d day.

After identifying the four-vectors, the three types of distances should be calculated and distance vectors should be formed:

1) Distance of suspicious points from the mean load consumption of that day (ω_{Id}).

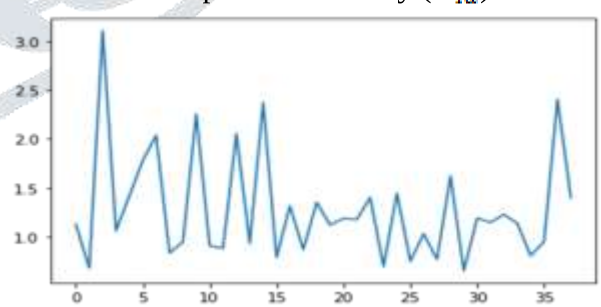


Fig. 2: Relative Distance of suspicious points from the mean of that day ($\boldsymbol{\omega_{Id}}$)

Where the ω_{id} is the distance between the suspicious points and the mean of that day, it will give the sense of the distance from the central consumption tendency. Which is been calculated using the following formula Where, abs is the absolute value of the difference between the mean of that day and the suspicious point. And to standardize it we will divide it by the means so that the magnitude doesn't affect with the higher values of the central consumption of different days.

2) Distance of suspicious points from the mean max load consumption that month (τ_{idh}).

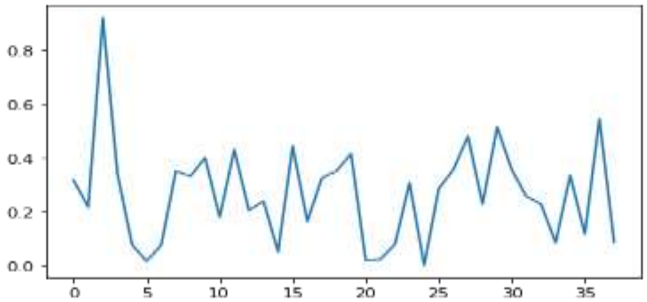


Fig. 3: Relative Distance of suspicious points from the max of that day (τ_{id})

There is a possibility that the customer has a high-power device such as mill, large motors, etc. which are operated very rarely. So, this will take of that factor that if a customer has high range device, it won't detect as abnormality.

Where the τ_{id} is the distance between the suspicious points and the mean of that day, it will give the sense of the distance from the central consumption tendency. Which is been calculated using the following formula

$$\tau_{id} = \frac{abs\ |m_{id} - \overline{Max_d}|}{\overline{Max_d}}$$

Where abs is the absolute value of the difference between the mean of maximum values of days for a month and that day's the suspicious point. And to standardize it we will divide it with the means maximum values of days for a month so that the magnitude doesn't affect with the higher values of the central consumption of different days.

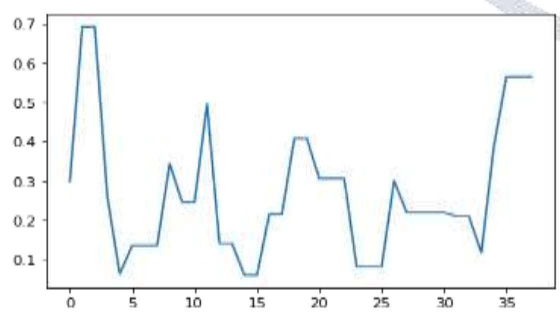


Fig. 4: Relative Distance of std of that day from the mean of monthly std ($\boldsymbol{\sigma_{Id}}$)

3) Distance of that day's standard deviation from the mean standard deviation of that month (σ_{Id}).

Where the σ_{id} is the distance between the suspicious point days' standard deviation and the mean standard deviation of the hole month it will give the sense of the distance from the central consumption tendency. Which is been calculated using the following formula.

$$\sigma_{id} = \frac{abs\ |std_d - \overline{std_d}|}{\overline{std_{td}}}$$

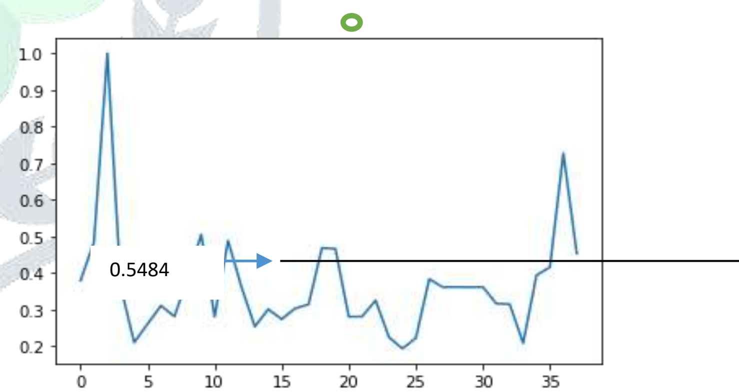
Where abs are the absolute value of the difference between the mean of maximum values of days for a month and that day's the suspicious point. And to standardize it we will divide it with the means maximum values of days for a month so that the magnitude doesn't affect with the higher value of the central consumption of different

These distances are on their own relative scale, for better comparison all the distances should be normalized between 0-1.

V. Result &Discussion

In the cleaning phase, the proposed anomality detection technique is utilized on the given dataset to find anomality. Fig. 5 shows the point distance (*D*) on the load consumption of household 1.

As it can be seen that only two points are identified as anomality, however factors sensitivity and threshold level can be changed by varying weights. Here with given weights the threshold is 0.5484. Any point above the threshold will consider as an anomality. Similarly, the anomality is removed from all the 200 households in the given dataset.



VI. CONCLUSION:

For an analysis of load profiles of residential customers, the pre-processing stage is very important for accurate results. But due to salient features of residential customer's load profiles, it requires dedicated techniques for data pre-processing. Therefore, for anomality detection, a novel methodology has been proposed and it has been seen that the proposed method has pointed out two anomalies from the data of household 1.

Whereas, quantile approach was suggesting 32 anomalies All 32 points are not anomalies, they could be the rarely operated high rated equipment. As the proposed method has successfully pointed out the anomalies only, as it considered the behavior of the customer also.

In the perspective of future scope, there could be a lot of scope for future researchers like including the likelihood of the customers to use appliances at the particular hour of the time. And there is a lot of scopes to include other features to contribute amorality detection.

REFERENCES

[1] J.Luo (VSTLF), "Real-time anomaly detection for very short-term load forecasting," *SGEPRI*, 2018.

[2] M. Yue, "Anomaly Detection Based on Long Short-Term Memory Neural Network," *IEEE*, 2017.

[3] A. lazaris, "An LSTM Framework For Modeling Network Traffic," *IEEE*, 2019.

[4] v. k. prasanna, "A short-term building cooling load prediction method," 2019.

[5] P. Zhao, "Advanced correlation-based anomaly detection method for predictive maintenance," *IEEE*, 2017.

[6] k. zhang, "Mid-term electricity load prediction using CNN and Bi-LSTM," *IEEE*, 2020.

[7] G. Strbac, "Demand side management: Benefits and challenges," *Energy Policy*, 2020.

[8] X. Wang, "Real-time prediction and anomaly detection of electrical load in a residential community," *IEEE*, 2019.