

# Time Granularity Transformation of Time Series Data for Failure Prediction of Overhead Line

Yan Ma<sup>1</sup>, Wenbing Zhu<sup>1</sup>, Jinxia Yao<sup>1</sup>, Chao Gu<sup>1</sup>, Demeng Bai<sup>1</sup> and Kun Wang<sup>2</sup>

1 State Grid Shandong Electric Power Research Institute, Jinan, China 250002

2. Shandong Zhongshi Yitong Group Co. Ltd., Jinan, China 250002

E-mail: yanpony@126.com

**Abstract.** In this paper, we give an approach of transforming time series data with different time granularities into the same plane, which is the basis of further association analysis. We focus on the application of overhead line tripping. First all the relative state variables with line tripping are collected into our big data platform. We collect line account, line fault, lightning, power load and meteorological data. Second we respectively pre-process the five kinds of data to guarantee the integrality of data and simplicity of analysis. We use a representation way combining the aggregated representation and trend extraction methods, which considers both short term variation and long term trend of time sequence. Last we use extensive experiments to demonstrate that the proposed time granularity transformation approach not only lets multiple variables analysed on the same plane, but also has a high prediction accuracy and low running time no matter for SVM or logistic regression algorithm.

## 1. Introduction

At present the data generated from grid operation and equipment monitoring exponentially increase and the environment of electric big data forms. Different monitoring indicators are collected into separate business information systems at various intervals. In order to integrate the data from different systems including production management system (PMS), energy management system (EMS), online condition monitoring system and meteorological system [1], we build a big data platform of condition assessment for electric transmission and transformation equipment. The time interval of experimental data in PMS system depends on the period of specific experiments and maintenance plan. The voltage and current are stored every five minutes. The online condition monitoring system has different collecting intervals according to different monitoring devices and in general its data are collected every 24 hours. Meteorological system has two kinds of monitoring indicators: real-time state variables and predictable state variables, which also have different time granularities.

Overhead transmission lines are exposed in the wild all year round and usually suffer severe natural environment and serious disaster attack, which cause degrees of damage for safe operation and lead to the trip of transmission lines. In order to find the rules between line tripping and its relative state variables, a feature representation method of data in different time granularities is required to make conjoint analysis feasible. The relative data on line tripping are the fault description information including phase, incoming switch, outgoing switch, line fault position, line fault reason and whether reclosing successfully, account information including line ID, equipment code, line name, voltage class and maximum allowable electric current, lightning information, real-time meteorological



information, one hour power load information before line fault including current, active power and reactive power.

All the above information has different collecting time granularities, where loading data are every five minutes, real-time meteorological data are every 10 minutes and the time granularities of the other data are uncertain. In the paper we combine data characteristics of line tripping, and mainly research the uniform transformation method for time series data with different time granularities, which is fundamental for the future association analysis between line tripping and its relative state variables.

## 2. Related work

Time series data is a kind of high dimensional data related with time. Mining the valuable information and knowledge from massive time series data is one of the mainstream researches in data mining domain [2]. The studies of time series data mining include similarity measurement, classification, clustering, pattern discovery and association rules of time series data, etc. In the paper, we focus on the data pre-processing step of association rules [3].

Time granularity is regarded as temporary resolution. Time granularity has many levels. A single time granularity such as day, hour, second has absolute time length [4]. Each time interval can be measured by a time granularity with the same time length. If multiple attributes in the same time range have different time granularities, these attributes have different numbers of records. If we try to combine these attributes and analyze them, an approximate representation method of time series data with different time granularities is necessary. Since the structure of time series data is complicated, directly mining on the original data has a low efficiency and affects the accuracy and reliability of mining results [5]. Therefore, finding a proper representation method for multiple status data and meanwhile keeping the main information are the foundation of the further analysis.

At present, common representation methods consist of discrete wavelet transformation [6], piecewise linear representation [7], symbolic representation [8] and etc. Discrete wavelet transformation maps time domain to frequency domain through wavelet transformation and use a small amount of wavelet coefficients to represent original time series data. The method has a high degree of dimension reduction, but the inappropriate wavelet coefficients lose massive information. Piecewise linear representation uses  $k$  consecutively connected lines to represent a time sequence with length  $n$ . It could reduce data dimensions but the selection of  $k$  is critical. Symbolic representation first discretizes time sequence and then map each part of time series waveform to a symbol. It use research results of character string, but it is hard to discretize time series and define the meaning of symbols.

Before the further association analysis, we first preprocess time series data using the proposed transformation method. It combines aggregated representation and trend extraction methods, which absorbs the simplicity of aggregated representation and the refining of trend extraction. We analyze each state variable, which guarantees not to lose lots of information and also reduces the time complexity of data mining. The proposed method realizes the feasibility of jointly analyzing multiple attributes with different time granularities.

## 3. Time granularity transformation method

We first discuss the time range of different state variables. In order to cover the error of inconsistent clock, the time ranges of lightning and  $I$  data are set as 10 minutes before line fault and 2 minutes after line fault. The time range of power load information is set as one hour before line fault. Through data analysis, lightning, meteorological and power load data need doing time granularity transformation to have the same time granularity with line account data and fault data. Line account information and fault description data correspond to a device or a fault. We next separately process different attributes related to line tripping according to their characteristics.

### 3.1. Lightning data

If there is no lightning data in the defined time range before or after a fault, the attribute value is set to 0. If there are many lightning data before or after a fault, we just record the maximum, minimum and

average value, which uses three values in the same time interval to represent the time series information of lightning data.

### 3.2. Meteorological data

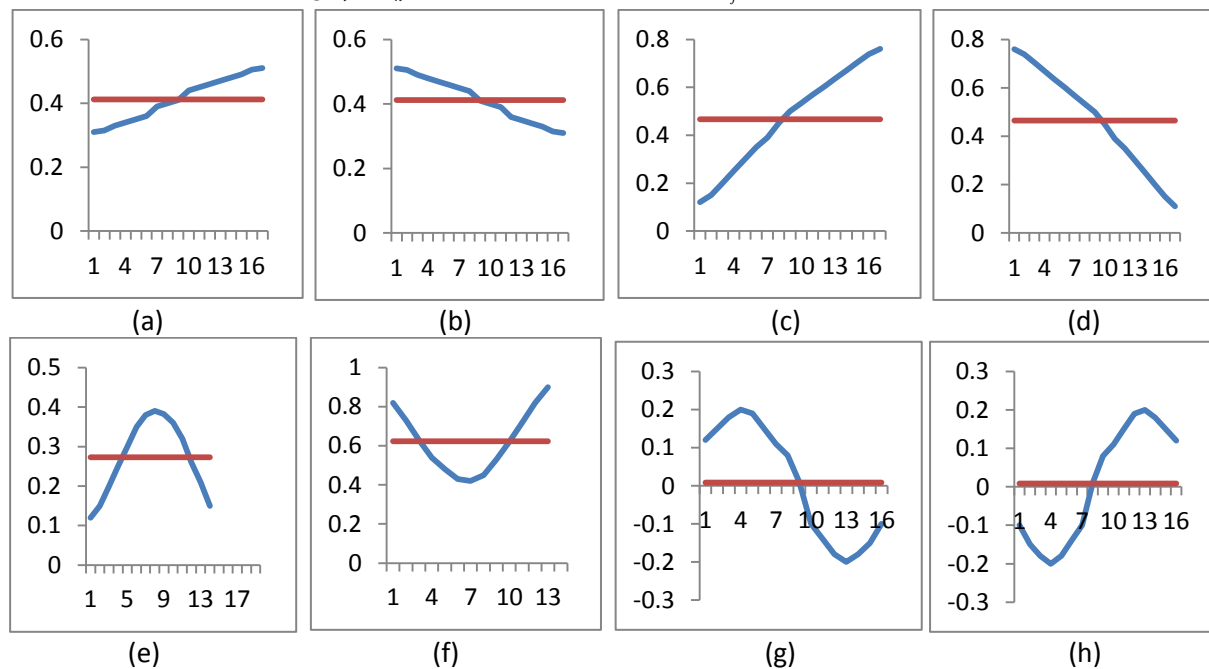
Meteorological data are provided by weather monitoring station. A line may cover two or more weather monitoring stations, and the real-time meteorological data are sent 10 minutes at a time. We choose the recent meteorological data from the weather monitoring station near to the fault and record its real-time data including wind direction, wind speed, gust wind direction, gust wind speed, amount of precipitation, relative humidity, temperature, air pressure, visibility, maximum wind direction, maximum wind speed and etc.

### 3.3. Power load data

Power load data include many information, thus we use a relatively complicated method to represent them. Power load data include three attributes: electric current, active power and reactive power. We first respectively compute their average values one hour before fault. Assuming the time sequence of electric current one hour before fault is denoted as  $I = \{i_1, i_2, \dots, i_n\}$ , its average value is computed as  $\bar{i} = \sum_{n=1}^n i/n$ . Meanwhile, we compute the variation factors of electric current, which include head variation factor and tail variation factor. Head variation factor and tail variation factor are represented as  $\Delta i_s = i_1 - \bar{i}$  and  $\Delta i_e = i_n - \bar{i}$ . The average value, head variation factor and tail variation factor of electric current as new state variables are added into line fault records.

In addition, in order to represent the long term trend of time series data, we introduce the conception of local extremums as the key point of trend variation ratio. Local minimum point is defined as follows. If the electric current  $i_j$  satisfies one of the following conditions:

1) For  $1 < j < n$ , there exist the subscripts  $r, t$  satisfying  $1 \leq r < j < t \leq n$ , and  $i_j$  is the minimum value among  $i_r, \dots, i_t$ ; 2) For  $j=1$ , there exists the subscript  $r$  satisfying  $j < r \leq n$ , and  $i_j$  is the minimum value among  $i_1, \dots, i_r$ ; 3) For  $j=n$ , there exists the subscript  $t$  satisfying  $1 \leq t < j$ , and  $i_j$  is the minimum value among  $i_t, \dots, i_n$ . Then the electric current  $i_j$  is defined as extreme minimal value.



**Figure 1.** The curvilinear trends of power load data

For power load information, we set eight kinds of curvilinear trends, which include ①slightly

ascending, ②slightly descending, ③obviously ascending, ④obviously descending, ⑤first descending then ascending, ⑥first ascending then descending, ⑦generally ascending, ⑧generally descending. Their concrete images are shown as figure 1. Taking current as examples, the unit of Y axis is ampere and the unit of X axis is the number of time intervals.

Their concrete methods of trend judgment are as follows.

1) We first determine whether there exist extremums during the time period. If there is no extremum, we could know that its trend is categorized to trend ①, trend ②, trend ③ or trend ④. If there is a extremum, its trend is trend ⑤ or trend ⑥. If the extremum is extreme maximum, it is trend ⑤; else if the extremum is extreme minimum, it is trend ⑥. If there are two extremums, the trend are categorized as trend ⑦ or trend ⑧.

2) If there are more than two extremums, we further divide time series data. After the segmentation, go to step 1)

3) Taking the electric current as an example, we compute its head variation factor and tail variation factor as  $\Delta_s = i_1 - \bar{i}$  and  $\Delta_e = i_n - \bar{i}$ . The minus between head variation factor and tail variation factor is  $\Delta = i_1 - i_n$ .

4) After determining the range of trend ① to trend ④ in step 1), we further give determination conditions. If  $0 < \Delta < \delta$  stands, it is slightly changing trend. If  $|\Delta| \geq \delta$  holds, it is obviously changing trend, where  $\delta$  is determined as 20 percent of the varying minus during a year for the state variable. If  $\Delta < 0$  holds, we know it is ascending trend; else it is descending trend.

5) After determining the range of trend ⑦ and trend ⑧, we next use  $\Delta$  as the judgment condition. If  $\Delta < 0$  stands, it is generally ascending trend; else it is generally descending trend.

After getting the trends of time series data, we use symbols to represent the eight kinds of trends, and add them into our data records.

#### 4. Experiments

In the section, we make extensive experiments to validate the performance of our proposed time granularity transformation method.

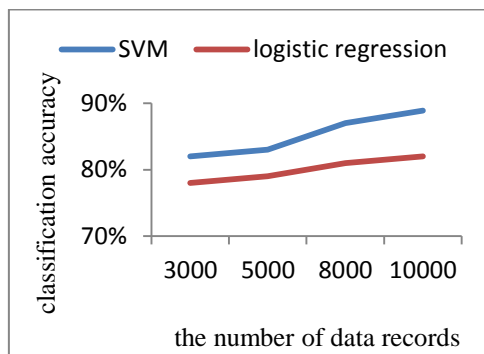
We first choose 10000 fault records from unplanned outage lines, which include fault line ID and fault happening time. Combing them with our given time range for each state variable, we could obtain all the relative state variables with line tripping through searching our big data platform. Then we preprocess the original data using the proposed transformation method. So far we get line fault data. Meanwhile, according to the aforementioned line ID, we choose a time interval when the line runs smoothly. In the similar way we preprocess and obtain line normal data.

We use SVM (support vector machine) and logistic regression algorithm to run the above data. The data pattern is organized as table 1, where the new generated variables are input attributes and the line state is goal attribute.

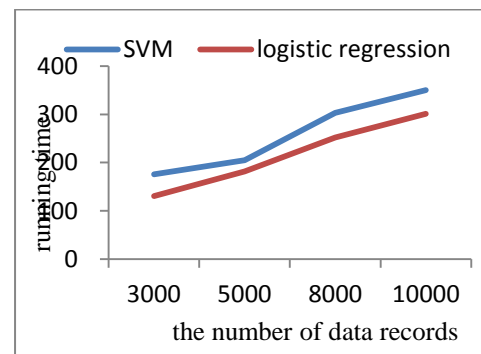
**Table 1.** Data pattern of line state prediction

Line accou nt varia bles	Real meteorolo gical variables	Lightn ing curren t maxi mum	Lightn ing curren t minim um	Lightn ing curren t averag e value	Curre nt maxi mum	Curre nt minim um	Curre nt aver age valu e	Curre nt head variati on factor	Curre nt tail variati on factor	Curre nt chang ing trend	...	active powe r head variati on factor	active powe r tail variati on factor	active power chang ing trend	Li ne sta te
--------------------------------------	---	---	---	--	----------------------------	----------------------------	---	--	--	--------------------------------------	-----	--	--	--	-----------------------

After running SVM and logistic regression algorithm according to the above data pattern, we get the prediction value of line state. Figure 2 and figure 3 shows their predication accuracy and running time. As seen from figure 2 we know the predication accuracy increase with the increasing number of records and SVM has a higher classification accuracy than logistic regression algorithm. In figure 3 their running times increase with the increasing number of data records as well and logistic regression algorithm has a shorter running time.



**Figure 2.** Classification accuracy with the number of data records.



**Figure 3.** Running time with the number of data records.

In all, no matter SVM or logistic regression algorithm, time granularity transformation method makes jointly analyzing multiple state variables feasible and largely reduces running time to an acceptable level. The most important advance is transformation method does not push down the classification accuracy.

## 5. Conclusions

In the paper, we propose a new transformation method for time series data with different time granularity. It totally considers the characteristics of overhead line and its concrete state variables. We aim for five kinds of business data: account data, fault data, lightning data, power load data and meteorological data. After the research, we could put all the state variables on the same plane, which lays a foundation for further joint analysis. After running SVM and logistic regression algorithm we validate that the proposed transformation method makes the classification accuracy and running time acceptable enough for failure prediction.

## Acknowledgement

This work was supported by the institute-level project in State Grid Shandong Electric Power Research Institute (2016-15). The authors are grateful for the anonymous reviewers who made constructive comments.

## References

- [1] Peng X, Deng D, Cheng S, Wen J, Li Z and Niu L 2015 *Proceedings of the CSEE*, **35** 503.
- [2] He X. Research on Key Issues in Time Series Data Mining 2014 *Ph.D. dissertation*.
- [3] Li H. Research on Feature Representation and Similarity Measure Methods in Time Series Data Mining 2012 *Ph.D. dissertation*.
- [4] Shao X, Hao W, Jin D, Wang Y 2011 *Journal of Computer Applications* **31**.
- [5] Xi W, Li P, Guo X, Xu A, Jiang Y, Zhang L and Wu Y 2014, *Power System and Clean Energy* **30**.
- [6] Akansu A, Haddad R 1992 *Academic Press*, ISBN 978-0-12-047141-6.
- [7] Yan C, Fang J, Wu L and Ma S 2013 *Journal of Information & Computational Science* **10** 2747.
- [8] Lin J, Keogh E, Wei L and Lonardi S 2007 *Data Mining & Knowledge Discovery* **15** 107.