Electric Power Systems Research 166 (2019) 52–60

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Clustering as a tool to support the assessment of power quality in electrical power networks with distributed generation in the mining industry

Michał Jasiński[a], Tomasz Sikorski[a,*], Klaudiusz Borkowski[b]

[a] Department of Electrical Engineering Fundamentals, Faculty of Electrical Engineering, Wroclaw University of Science and Technology, Wroclaw, Poland
[b] KGHM Polska Miedź S.A., Polkowice, Poland

ARTICLE INFO

Keywords:
Data mining
Clustering
Distributed generation
Power quality
Mining industry

ABSTRACT

This paper presents a case study of using cluster analysis (CA) as one of the data mining (DM) techniques applied in the analysis of long-term power quality (PQ) data that is recorded in electrical power networks of the mining industry. The aim of the clustering is to highlight the impact of distributed generation (DG) on the level of power quality parameters. The carried out investigations concern the application of the K-mean clustering algorithm with Euclidean and Chebyshev distance with a different number of clusters for standardised and non-standardised data. The obtained results show the possibility to obtain automatic classification of data into distinguishable clusters that represent the period of time when local DG is active, switched-off or when a different power consumption level is denoted. It leads to the possibility of using CA as a suitable tool for assessing the impact of local generation on the working conditions of electrical power networks that depend on a DG contribution or power consumption. Additionally, the obtained results allow CA to be indicated as a proper method for the automatic identification of the PQ data which are affected by voltage events that can be treated as alternative way for present flagging concept.

1. Introduction

The quality of the supply is one of the prominent goal of present power system. The main issues are close to continuity of the supply and power quality. The power quality is currently defined as set of technical parameters mainly associated with low frequency disturbances of the supply voltage. For this kind of phenomena there are already standardised methods of measurements and assessment [1–3]. Fig. 1 presents typical division of power quality disturbances: voltage variations and voltage events.

The classical method of assessing power quality is based on choosing a representative period of time, normally one week, which corresponds to the normal working conditions of a network. However, the measurement data can be dependent on load changes or influence of generation as well as different configurations of the network. Due to these reasons, it is desirable to search for a comprehensive method to automatically classify long-term power quality data into clusters and to identify selected portions of the data that are characteristic for different network conditions [4]. One of the known power quality disturbances are voltage events like: voltage dips, rapid voltage changes (RVC), swells and interruptions. The detection and identification of voltage events are proposed by different methods e.g. Fourier transform, S-transform, Hilbert Huang transform, wavelet transform, support vector machine, neural network, fuzzy-expert system, neuro-fuzzy system, genetic algorithm [5]. Other new methods are: histogram of oriented gradients (HOG) [6], multi-resolution analysis (MRA) of S-transform [7], Fuzzy C-means clustering [8].

Presented in this paper investigations has two main goals. First is an automatic selection of the PQ data affected by the event. Second is automatic selection of the PQ into different groups which represent different working condition of observed electrical power network, especially related to impact of DG. In order to achieve mentioned goals proposed in this article approach is one of the selected data mining techniques known as clustering. Clustering is a data mining tool which allows to separate database of parameters in different point of view [9–14]. Usually the data selection is concentrated on identification of objects which are characterised by similar level of parameters. Other direction of the database clustering is not to reveal the objects but to identify different conditions which affect the objects. The aim of presented paper is close to concept of using the clustering for identification and comparison of different condition of local power system when impact of DG is considered. Several issue of proposed methodology can

* Corresponding author at: Wroclaw University of Science and Technology, Faculty of Electrcial Engineering, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland.
E-mail address: tomasz.sikorski@pwr.edu.pl (T. Sikorski).

https://doi.org/10.1016/j.epsr.2018.09.020
Received 4 April 2018; Received in revised form 19 July 2018; Accepted 26 September 2018

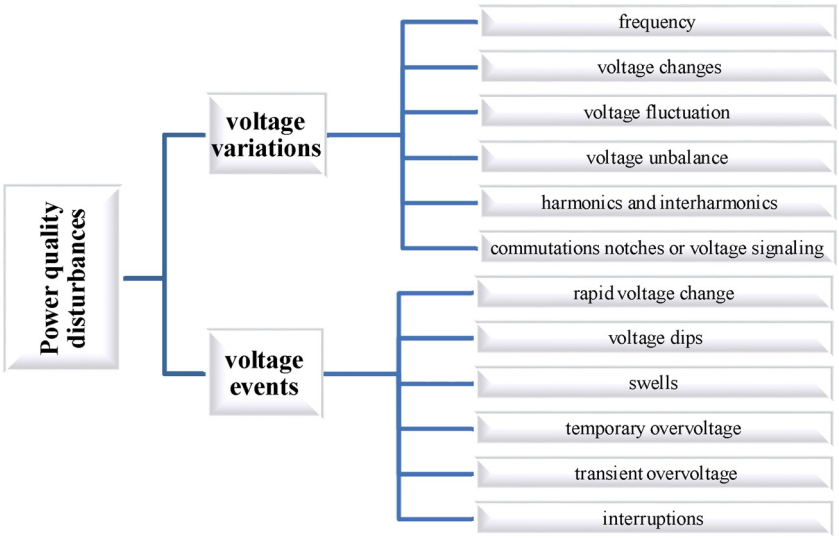

Fig. 1. Power quality disturbances.

be highlighted. Firstly, extraction of the PQ data affected by the event is proposed by using clustering with fixed number of clusters. Then, selection of the data related to different power system condition caused by different impact of DG is proposed by using clustering based on K-mean algorithm. Proposed clustering method is tested using different constructions of the database which take into account PQ data with the same and different aggregation time and PQ data extended by other parameters like total power. Used for trials PQ data comes from synchronized multipoint measurements performed in power system of mining industry which consists of DG. The aim of this efforts is to select the best database construction for the K-mean algorithm in order to obtain the best fitted selection of the PQ data to the condition of investigated power system. After data selection into clusters a comparative analysis is applied. In the paper, local stochastic analysis is applied for the data that is identified in particular clusters which exhibits impact of DG on local power system.

The proposed method is investigated using the real measurement of power quality recorded in the electrical power network of the copper mining industry. The power quality data comes from multipoint, time synchronized measurements that are system installed in selected power network substations on different voltage levels of 110 kV (HV), 6 kV (MV) and 0.4 kV (LV). The significant elements of the investigated power network are a combined heat and power plant with gas-steam turbines and also a welding machine connected to MV systems.

The obtained results allow cluster analysis to be used as a suitable method for classifying power quality data into groups. Presented classification assures the automatic division for cluster which represents different working condition e.g. when DG working and is switched-off. Then, the qualitative assessment of obtained clusters leads to a possibility of identifying the impact of different network conditions e.g. the distributed generation level on power quality.

2. Data mining techniques and clustering

Data mining can be seen to have many definitions. The most common are: “Data mining is a process of discovering interesting patterns and knowledge from large amounts of data” [10], “Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner” [11], or “Data mining is the extraction of implicit, previously unknown, and potentially useful information from data” [12]. Additionally, there are a lot of classifications of data mining techniques. Fig. 2 presents the classification of DM proposed by CIGRE in [9]. The presented paper is concentrated on clustering.

Clustering may be carried out using two methods: hierarchical and non-hierarchical [13,14]. Hierarchical methods constitute c classes of n observations. Non-hierarchical methods are based on assigning the observations to earlier known c clusters. Selected hierarchical methods are [14]:

- the single link method – based on the measurement of the smallest dissimilarity between two elements of a different cluster,
- the complete link method – based on the measurement of the biggest dissimilarity between two elements of a different cluster,
- the average link method – based on the measurement of mean dissimilarity between two elements of a different cluster,
- the Ward method – based on the measurement of dissimilarity between two clusters that are represented by the sum of squares of the deviation inside each cluster.

The main clustering non-hierarchical methods are based on the following algorithms:

- K-mean,
- K-median,
- Expectation-Maximization (EM),
- Self-Organizing Maps (SOM).

The data mining has many applications in power system. Selected applications related to power quality are presented in [15–19]. The system of power quality analysis based on data mining technologies, such as data cleaning, data fusion, cluster analysis, correlation analysis, is presented in [16] and shows that there are methods of assessing the influence of the environment on power quality parameters. A classification method based on the fast Independent component analysis algorithm (Fast-ICA) was presented in [17] in order to determine a global index for power quality. Classification of the events was made using the K-mean algorithm with Euclidean distance was described in [18]. Work [19] presents the application of clustering to power quality data, which includes events to achieve a classification of the events.

This article concerns the application of clustering for revealing the impact of distributed generation on the working conditions of local electrical power networks. In this paper authors suggest to use the non-hierarchical method based on the K-mean algorithm for the classification of power quality data.

In the case of clustering using the K-mean algorithm, the goal is to find the extremum of the objective function. The k-mean algorithm