

## WYKORZYSTANIE ANALIZY SKUPIEŃ DO IDENTYFIKACJI OZNACZONYCH WYNIKÓW POMIARU JAKOŚCI ENERGII ELEKTRYCZNEJ

Michał JASIŃSKI<sup>1</sup>, Tomasz SIKORSKI<sup>2</sup>, Dominika KACZOROWSKA<sup>3</sup>, Klaudiusz BORKOWSKI<sup>4</sup>

1. Politechnika Wrocławska, Wydział Elektryczny  
tel.: 71 320 20 22 e-mail: [michal.jasinski@pwr.edu.pl](mailto:michal.jasinski@pwr.edu.pl), [tomasz.sikorski@pwr.edu.pl](mailto:tomasz.sikorski@pwr.edu.pl),
2. Politechnika Wrocławska, Wydział Elektryczny  
tel.: 71 320 21 60 e-mail: [tomasz.sikorski@pwr.edu.pl](mailto:tomasz.sikorski@pwr.edu.pl),
3. Politechnika Wrocławska, Wydział Elektryczny  
tel.: 71 320 20 22 e-mail: [dominika.kaczorowska@pwr.edu.pl](mailto:dominika.kaczorowska@pwr.edu.pl)
4. KGHM Polska Miedź S.A.  
tel.: 76 748 11 69 e-mail: [klaudiusz.borkowski@kgm.com.pl](mailto:klaudiusz.borkowski@kgm.com.pl)

**Streszczenie:** W artykule zaprezentowano wykorzystanie eksploracji danych w analizie parametrów określających jakość energii elektrycznej (JEE) pod kątem identyfikacji danych podlegających regule oznaczania w rozumieniu normy PN EN 61000-4-30 [1]. Przedstawiono zastosowanie analizy skupień jako narzędzia umożliwiającego podział zagregowanych danych pomiarowych na grupy reprezentujące wyniki pomiarów wolne od zdarzeń napięciowych oraz wyniki pomiarów, w trakcie których wystąpiło zdarzenie napięciowe. Przebadano wrażliwość algorytmu k-średnich na identyfikację danych zawierających przerwy, zapady, wzrosty oraz szybkie zmiany napięcia. Za zbiór danych testowych wykorzystano synchroniczne pomiary przeprowadzone w sieci zakładów górniczych. Uzyskane wyniki pozwalają na określenie skuteczności wykorzystania analizy skupień do identyfikacji danych zagregowanych zawierających zdarzenia napięciowe.

**Słowa kluczowe:** jakość energii elektrycznej, analiza skupień, koncepcja oznaczania.

### 1. WSTĘP

Rozwój i informatyzacja systemu elektroenergetycznego wskazuje na ciągłą potrzebę poszukiwania narzędzi wspierających wykorzystanie potencjału dużych zbiorów danych (ang. big data). Jedną z możliwości jest wykorzystanie eksploracji danych (ang. data mining). Eksploracja danych (ED) w literaturze definiowana jest na różne sposoby m. in.

- ED to proces odkrywania ciekawych wzorców i wiedzy z dużych zbiorów danych [2],
- ED jest analizą (często dużych) obserwowalnych zbiorów danych w celu odnalezienia nieoczekiwanych związków i by podsumować dane w sposób zrozumiały i użyteczny dla użytkownika [3],
- ED jest wydobywaniem ukrytych, wcześniej nieznanymi i potencjalnych informacji z danych [4],
- ED to proces odkrywania znaczących nowych korelacji, wzorców i trendów poprzez przeszukiwanie dużych liczby danych przechowywanych w repozytoriach, przy użyciu technologii rozpoznawania wzorców, a także technik statystycznych i matematycznych [5].

Wskazywane w literaturze definicje zawierają wspólne komponenty, do których należą m. in.: wzorce, informacje, związki, które wynikają z danych zawartych w dużych zbiorach oraz interpretacja przydatna dla odbiorcy [6]. Do przykładów zastosowania narzędzi eksploracji danych we współczesnych sieciach elektroenergetycznych zaliczyć można następujące zagadnienia: klasyfikacja i lokalizacja zwarć, prognozowanie obciążenia, monitorowanie wyładowań niezupełnych w transformatorach, ocena stabilności systemu, klasyfikacja zaburzeń jakości energii elektrycznej [7] [8] [9] [10].

W niniejszej pracy zaprezentowano wyniki badań wykorzystania analizy skupień (AS) jako jednej z technik eksploracji danych do potrzeb identyfikacji wyników pomiarowych podlegających regule oznaczania. Jako źródła danych wykorzystano rzeczywiste wyniki pomiarów parametrów jakości energii elektrycznej, zarejestrowane z sieci kopalnianej. Badania przeprowadzono pod kątem skuteczności identyfikacji danych w porównaniu do klasycznego algorytmu oznaczania opisanego w normie PN EN 61000-4-30.

### 2. ANALIZA SKUPIEŃ

Analiza skupień (ang. cluster analysis, clustering) w polskiej literaturze znana jest również jako, klastrowanie, grupowanie oraz taksonomia [11]. Zasadniczym celem analizy skupień jest uzyskanie homogenicznych grup danych (skupień) [12]. Istnieje wiele różnych możliwości doboru warunków podziału danych na skupienia. Za najczęściej stosowane kryteria można przyjąć podobieństwo obiektów przypisanych do danej grupy lub odmiennosć obiektów jednej grupy od elementów grup pozostałych [13],[14]. Jednak ze względu na podejście do procesu grupowania można wyróżnić dwie zasadnicze metody podziału podanych na skupienia:

- hierarchiczne - realizowane są poprzez tworzenie hierarchii klasyfikacji – dla  $n$  obserwacji tworzy się hierarchię składającą się z  $k$  klas,  $k=1, 2, 3, \dots, n$ ,

- niehierarchiczne - zwane również kombinatorycznymi oparte są na przyporządkowaniu  $n$  obiektów do zadanej wcześniej  $k$  liczby skupień.

W przypadku niehierarchicznych metod analizy skupień za najpopularniejsze algorytmy można przyjąć [12]:

- algorytm  $k$ -średnich (ang.  $k$ -mean algorithm),
- algorytm EM (ang. algorithm EM – Expectation Maximization),
- FCM: algorytm  $k$ -średnich rozmytych (ang. Fuzzy c-Means).

W przypadku zastosowanego w prezentowanych wynikach badań algorytmu  $k$ -średnich istotnym elementem jest określenie najbliższych środków skupień (centroidów). Do realizacji tego celu możliwe jest zastosowanie wielu miar odległości elementów zbiorów danych. Do najpopularniejszych metod wybranych z [12], [15] należą:

- odległość Minkowskiego,
- odległość Mahalanobisa,
- dywergencja Bergmana,
- odległość kosinusowa,
- odległość potęgowa,
- odległość euklidesowa,
- kwadrat odległości euklidesowej,
- odległość miejska (Manhattan),
- odległość Czebyszewa.

W badanym przypadku wybrano odległość Czebyszewa wyznaczaną następująco:

$$OC(x, y) = \text{Maksimum} |x_i - y_i| \quad (1)$$

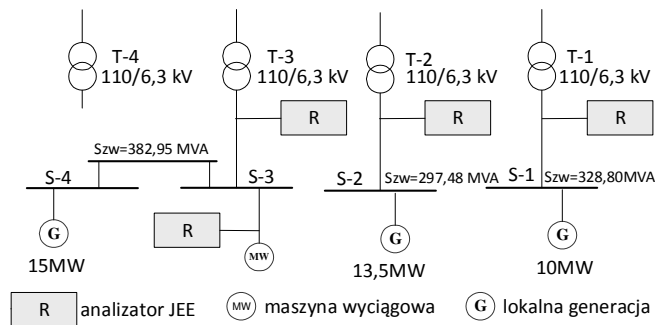
gdzie:  $OC(x, y)$  – odległość Czebyszewa między wektorami  $x$  i  $y$ ,  $x_i$ - wektor obserwacji (elementów) należących do skupienia  $x$ ,  $y_i$ - wektor obserwacji (elementów) należących do skupienia  $y$ .

Wybór odległości Czebyszewa podyktowany jest ukierunkowaniem na wskazanie maksymalnych różnic między grupami (skupieniami), które reprezentują różne uwarunkowania pracy sieci. Wybór metody poprzedzono badaniami wstępnymi dla różnych definicji odległości w środowisku Statistica. Wykonano analizy dla różnych kombinacji liczby końcowej skupień, odległości i sposobu doboru początkowych środków skupień.

### 3. OPIS OBIEKTU BADAŃ ORAZ ŹRÓDŁO DANYCH POMIAROWYCH

Wykorzystane w artykule dane pomiarowe JEE pochodzą z elektroenergetycznych sieci zakładów górniczych [16]. Badania przeprowadzono z wykorzystaniem przyrządów klasy A w okresie od 27.04.2017-28.06.2019 w czterech punktach pomiarowych średniego napięcia, w tym trzech polach transformatorów oraz jednym polu odpływowym znaczącego odbioru - maszyny wyciągowej (MW). Uproszony schemat badanego fragmentu sieci wraz lokalizacją rejestratorów został zaprezentowany na rysunku 1. W warunkach normalnej pracy sekcje 1 i 2 (S-1, S-2) pracują niezależnie natomiast sekcje 3 i 4 są sprzęgnięte. Badana sieć charakteryzuje się zmiennym obciążeniem pochodzącym przede wszystkim od maszyn wyciągowych, pomp odwadniania kopalń, wentylatorów, przenośników taśmowych oraz zmiennym poziomem generacji rozproszonej. Do lokalnych źródeł energii elektrycznej należą trzy bloki gazowo-parowe

o mocach odpowiednio 15 MW, 13,5 MW oraz 10 MW. Lokalizację źródeł generacji oraz miejsca instalacji rejestratorów przedstawiono na rysunku 1.



Rys. 1. Uproszczonego schematu badanej sieci z uwzględnieniem lokalizacji analizatorów JEE

### 4. WYNIKI IDENTYFIKACJI DANYCH ZAGREGOWANYCH ZAWIERAJĄCYCH ZDARZENIA NAPIĘCIOWE

Podczas pomiarów odnotowano w każdym z punktów pomiarowych zdarzenia napięciowe. Zgodnie z regułą oznaczania (ang. flagging concept) w rozumieniu normy PN EN 61000-4-30 [1], jeżeli podczas danego przedziału czasu agregacji wystąpi zdarzenie typu zapad napięcia, wzrost lub przerwa, to każdą zagregowaną wartość parametrów jakości energii elektrycznej zawierającą ten znacznik czasowy należy oznaczyć i nie uwzględniać w ocenie jakości energii. Dodatkowo na potrzeby badań wrażliwości algorytmu analizy skupień poszerzono wskazaną grupę zdarzeń warunkujących zastosowanie oznaczania o szybkie zmiany napięcia [6], [16].

Wskazana reguła oznaczania obejmuje czasy agregacji właściwe dla poszczególnych parametrów jakości energii (10 sekund, 10 minut, 2 godziny). W związku z tym, że większość parametrów jakości energii wykorzystuje agregację 10-minutową na potrzeby artykułu przyjęto, iż do celów identyfikacji danych podlegających regule oznaczania zastosowany zostanie zbiór parametrów 10-minutowych składający się z następujących parametrów:

- częstotliwość,
- wartość skuteczna napięcia,
- krótkookresowy wskaźnik migotania światła,
- asymetria napięcia,
- całkowity współczynnik zawartości harmonicznych w napięciu.

Przyjęcie wspólnego czasu agregacji 10-minut, pozwoliło uzyskać jednolity wymiar zbioru danych. Dla tak przygotowanego zbioru danych przeprowadzono w pierwszej kolejności analizę oznaczania wyników pomiaru w badanych punktach pomiarowych niezależnie z klasycznym podejściem zgodnym z przywołaną normą PN-EN 6100-4-30. Wyniki zawarto w tabelicy 1.

Tabelica 1. Wyniki oznaczania danych pomiarowych metodą klasyczną w poszczególnych punktach pomiarowych

| Punkt pomiarowy | Liczba danych 10-minutowych objętych oznaczeniem |
|-----------------|--|
| T-1             | 50   |
| T-2             | 82   |
| T-3             | 278  |
| MW              | 13   |

Przedstawiona analiza zdarzeń i wyznaczona na tej podstawie liczba danych, które w ujęciu klasycznym powinny zostać objęte regułą oznaczania jest podstawą do oceny skuteczności zaproponowanej w artykule metody oznaczania danych za pomocą analizy skupień.

W celu zastosowania analizy skupień do identyfikacji danych przeznaczonych do oznaczenia, zastosowano zadaną wejściową liczbę skupień równą 2, przez co oczekiwano podziału na dane oznaczone i nieoznaczone. Wybrano algorytm *k*-średnich wykorzystując odległość Czebyszewa ze względu na uzyskanie dużej wrażliwości na różnicę między danymi przypisywanymi do skupień. Za zbiór danych wejściowych wykorzystano omówiony wcześniej zbiór danych 10-minutowych, dla którego dokonano identyfikacji danych oznaczonych metodą klasyczną.

Na podstawie przeprowadzonej analizy skupień o zadanych parametrach (algorytm *k*-średnich, odległość Czebyszewa, zadana liczba skupień 2, zbiór danych zagregowanych 10-minutowych) uzyskano podział wszystkich danych pomiarowych na dwa skupienia, których charakterystyka jest następująca:

- skupienie 1: reprezentujące zbiór 10-minutowych danych pomiarowych nie zawierających danych, w trakcie których mogło nastąpić dynamiczne zdarzenie napięciowe,
- skupienie 2: reprezentujące zbiór 10-minutowych danych pomiarowych, których wartości świadczą o tym, że podczas czasu agregacji mogło wystąpić zdarzenie napięciowe

Dla porównania wyników oznaczania danych metodą klasyczną i proponowaną techniką wykorzystującą analizę skupień na rysunku 2 przedstawiono porównanie identyfikacji danych oznaczonych i nieoznaczonych z użyciem obu metod. Następnie oceniono skuteczność oznaczania z wykorzystaniem analizy skupień. Uzyskano to poprzez ocenę różnicy zdarzeń przyporządkowanych jako oznaczone w stosunku do wyników identyfikacji uzyskanych metodą klasyczną. Zestawienie ilościowe przedstawiono w tabelicy 2. Uzyskano średnią skuteczność identyfikacji danych oznaczonych z zastosowaniem proponowanej metody na poziomie 70%. Skuteczność ta dla pól transformatorowych była zdecydowanie wyższa – na poziomie około 84%. Otrzymana średnia skuteczność w znacznym stopniu została zanizowana przez niewrażliwość metody na identyfikację szybkich zmiany napięcia, które dominowały w punkcie pomiarowym maszyny wyciągowej.

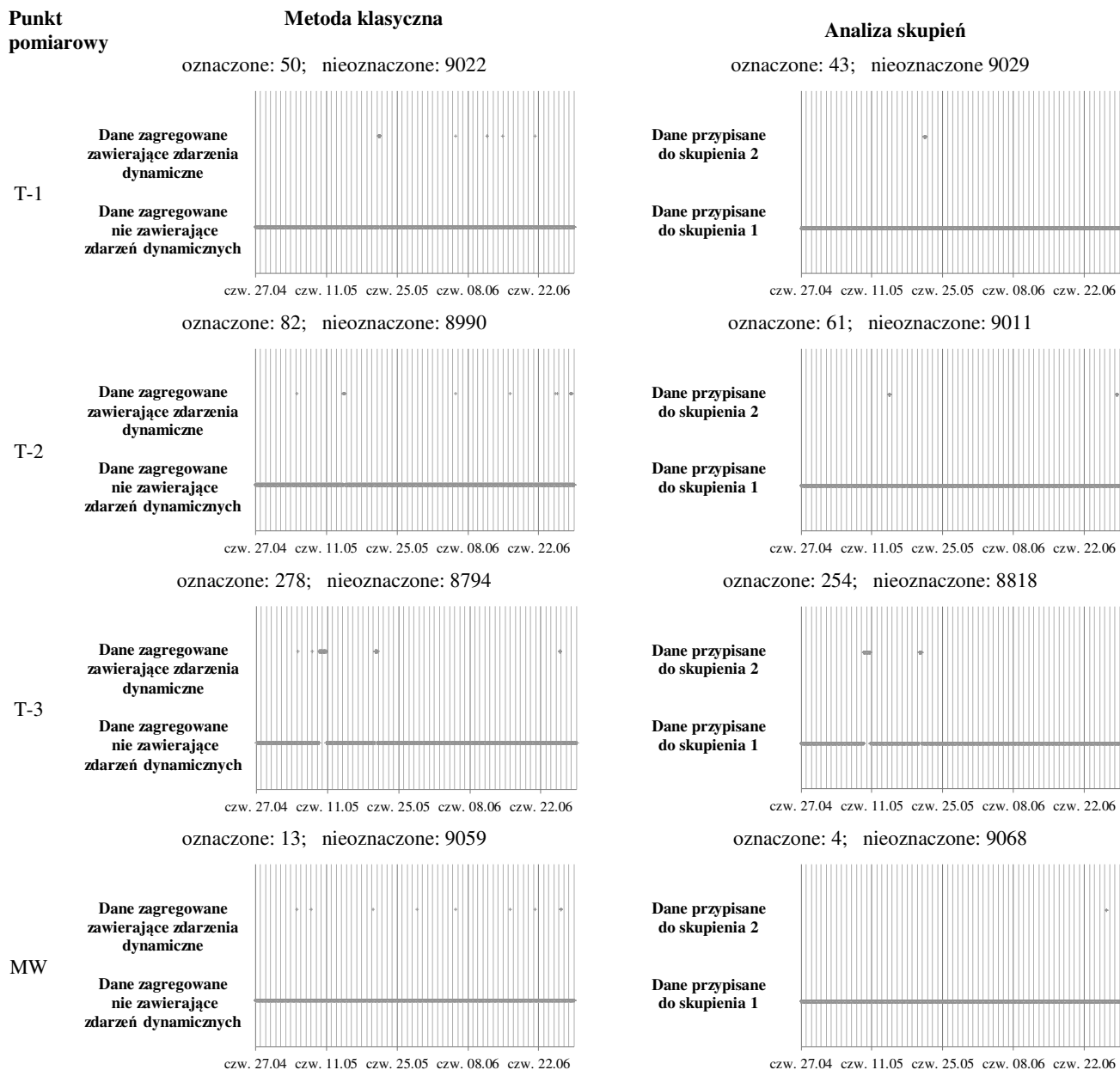
## 5. WNIOSKI

W artykule zaprezentowano możliwość wykorzystania analizy skupień do identyfikacji oznaczonych wyników pomiaru jakości energii elektrycznej. Wyniki uzyskane z zastosowaniem proponowanej metody porównano z wynikiem identyfikacji danych oznaczonych uzyskanych metodą klasyczną tj. zgodnie z koncepcją oznaczania przedstawioną w normie PN-EN 61000-4-30. Uzyskana średnią skuteczność metody na poziomie 70% dla rozszerzonego scenariusza identyfikacji zdarzeń tj. uwzględniającego zarówno zdarzenia typu zapad, wzrost lub przerwa, jak również szybkie zmiany napięcia.

Stwierdzono, iż zastosowana metoda analizy skupień nie jest wystarczająco wrażliwa do identyfikacji szybkich zmian napięcia. To skłania do dalszych badań metody nad możliwością zwiększenia wrażliwości identyfikacji oraz weryfikacji wyników na większej grupie danych pomiarowych z dłuższego okresu czasu.

## 6. BIBLIOGRAFIA

1. PN EN 6100-4-30: Kompatybilność elektromagnetyczna (EMC) -- Część 4-30: Metody badań i pomiarów-- Metody pomiaru jakości energii 2015.
2. Han J., Kamber M.: Data Mining: Concepts and Techniques, Waltham 2011.
3. Hand D., Mannila H., Smyth P.: Principles of data mining, Cambridge 2001.
4. Witten I. Frank E.: Data Mining: Practical Machine Learning Tools and Techniques, Burlington 2011.
5. Larose D.: Discovering knowledge in data. An Introduction to Data Mining, New Jersey 2005.
6. Jasiński M., Sikorski T., Borkowski K.: Clustering as a tool to support the assessment of power quality in electrical power networks with distributed generation in the mining industry, Electrical Power System Research, Nr 166, 2019, str. 52–60.
7. CIGRE, Brochure 292: Data mining techniques and applications in the power transmission field, 2006.
8. Morais J., Pires Y.: An overview of data mining techniques applied to power systems in Data Mining and Knowledge Discovery in Real Life Applications, I-Tech, 2009, str. 438.
9. Abdel-Galil T., Kamel M.: Youssef A., El-Saadany E., Salama M., Power quality disturbance classification using the inductive inference approach, IEEE Transactions Power Delivery, Nr 19, no. 4, 2004, str. 1812–1818.
10. Asheibi D., Stirling A., Perera A., Robinson S.: Power Quality Data Analysis Using Unsupervised Data Mining, 2004 Australas. Univ. Power Eng. Conf. AUPEC, 2004.
11. Larose D.: Odkrywanie wiedzy z danych. Wprowadzenie do eksploracji danych. Wydawnictwo Naukowe PWN, Warszawa 2006.
12. Wierzchoń S., Kłopotek M.: Algorithms Of Cluster Analysis, Institute of Computer Science Polish Academy of Sciences, Warszawa 2015.
13. Cabena P., Hadjinain P., Stadler R., Verhees J., Zanasi A.: Discovering data mining: from concept to implementation. New Jersey 1998.
14. Jain A., Murty M., Flynn P.: Data clustering: a review, ACM Comput. Surv., Nr. 31 (3), 1999, str. 264–323.
15. Statsoft Polska, StatSoft Electronic Statistic Textbook, 2016.
16. Jasiński M., Borkowski K., Sikorski T., Kostyla P.: Cluster Analysis for Long-Term Power Quality Data in Mining Electrical Power Network, 2018 Progress in Applied Electrical Engineering (PAEE), Kościelisko, 2018, str. 1–5.



Rys. 1. Porównanie wyników oznaczania danych metodą klasyczną i proponowaną techniką wykorzystującą analizę skupień

Tablica 2. Zestawienie wyników oznaczania danych metodą klasyczną oraz z wykorzystaniem analizy skupień

| Punkt pomiarowy | Łączna liczba danych | Dane oznaczone metodą klasyczną | Dane oznaczone wykorzystując analizę skupień | Skuteczność oznaczenia danych [%] |
|-----------------|----------------------|---------------------------------|--|-----------------------------------|
| T-1             | 9072                 | 50                              | 43   | 86,0%                             |
| T-2             | 9072                 | 82                              | 61   | 74,4%                             |
| T-3             | 9072                 | 278                             | 254  | 91,4%                             |
| MW              | 9072                 | 13                              | 4  | 30,8%                             |
| Średnia         |                      |                                 |  | 70,6 %                            |

## APPLICATION OF CLUSTER ANALYSIS TO IDENTIFICATION FLAGGED POWER QUALITY MEASUREMENTS

The article presents the use of data mining to power quality issue. The application of cluster analysis as a tool which lead to division into groups representing the measurement period for which aggregated data (within the meaning of PN EN 61000-4-30 standard) contain and do not contain aggregated voltage events is presented. The sensitivity of the K-means algorithm to the identification of data containing interruptions, dips, increases and rapid voltage changes was tested. Synchronous measurements carried out in the mining plant network were used for the test data set. The obtained results allow determining the effectiveness of using cluster analysis to identify aggregated data containing voltage events.

**Keywords:** power quality, cluster analysis, flagging concept.