

СИСТЕМАТИЗАЦИЯ СТАТИСТИЧЕСКИХ МЕТОДОВ КЛАСТЕРИЗАЦИИ ВЕЙВЛЕТ-ПРЕОБРАЗОВАННЫХ ГРАФИКОВ ЭЛЕКТРИЧЕСКИХ НАГРУЗОК

А. VOLOSHKO, T. LUTCHYN

SYSTEMATIZATION STATISTICAL METHODS OF CLUSTERING WAVELET-CONVERT SCHEDULES OF ELECTRIC LOADS

Анотація. Пропонується систематизація статистичних методів кластеризації графіків електричних навантажень на основі вейвлет-коефіцієнтів. За допомогою вейвлет-аналізу можна не тільки виявити характерні частоти сигналу, але й отримати відомості про конкретні локальні координати, при яких проявляються ці частоти. В умовах нелінійності та аперіодичності вихідної інформації вейвлет-перетворені графіки електричних навантажень опрацьовують, застосовуючи відомі статистичні методи кластеризації.

Систематизація статистичних методів кластеризації перетворених даних оптимізує вибір найбільш ефективного методу вирішення. Залежно від поставлених задач визначається такий метод розрахунку, який забезпечить необхідний рівень інформативності результатних даних, застосовність розрахункових даних для подальшої обробки, високу швидкість і відновленість обчислень.

Ключові слова: вейвлет-коефіцієнти, метод кластеризації, електричне навантаження, інформативність результатів, систематизація.

Анотация. Предлагается систематизация статистических методов кластеризации графиков электрических нагрузок на основе вейвлет-коэффициентов. С помощью вейвлет-анализа можно не только выявить характерные частоты сигнала, но и получить сведения об определенных локальных координатах, при которых проявляются эти частоты. В условиях нелинейности и аперiodичности исходной информации вейвлет – преобразованные графики электрических нагрузок обрабатывают, применяя известные статистические методы кластеризации.

Систематизация статистических методов кластеризации преобразованных данных оптимизирует выбор наиболее эффективного метода решения. В зависимости от поставленных задач определяется тот метод расчета, который обеспечит необходимый уровень информативности результирующих данных, применимость расчетных значений для последующей обработки, высокую скорость и обратимость вычислений.

Ключевые слова: вейвлет-коэффициенты, метод кластеризации, электрическая нагрузка, информативность результатов, систематизация.

Annotation. Offered systematization of statistical methods for clusterization the graphs of electrical loads based on wavelet-coefficients. With the wavelet-analysis can not only reveal the characteristic frequency of the signal, but also obtain information about specific local coordinates at which these frequencies occur. In terms of nonlinearity and aperiodically initial information wavelet-transformed the graphs of electrical loads is treated by applying known statistical clustering methods.

Systematization of clustering methods optimizes the transformed data selection decisions. Depending on the goals we choose that method of calculation, which will provide the necessary level of information content of the resulting data, the applicability for further processing, speed and reversibility of computation.

Key words: wavelet-coefficients, method of clusterization, electrical load, informing results, systematization.

Введение

Одним из широко используемых способов обработки измерительной информации на сегодняшний день является кластеризация. Это связано с возможностью автоматического формирования результатов с учетом степени сжатия первоначальных данных. Методами кластерного анализа решается задача разбиения (классификации, кластеризации) множества объектов таким образом, чтобы все объекты, принадлежащие одному кластеру (классу, группе), были наиболее подобны друг другу по сравнению с объектами других кластеров [1].

Постановка задачи. В работе проводится сравнительный анализ различных методов кластеризации графиков электрических нагрузок на основе вектор-признаков, сформированных с помощью вейвлет-преобразования [2]. При использовании данного метода в вейвлет-спектре, сформированном на основе вейвлет-пакетов, формируются вектор-признаки из усредненных мощностей вейвлет-коэффициентов [3]. То есть, в качестве первичных признаков кластеризации выступает упорядоченная последовательность средних мощностей вейвлет-коэффициентов по субполосам разложения.

Вначале определим вид закона распределения вектор-признаков (последовательности средних мощностей его членов) для проведения сравнительного анализа методов кластеризации.

Рассчитаем математическое ожидание (среднее значение ряда) и разброс (среднее отклонение) по известным формулам для проверки закона распределения данных случайных величин:

$$P_{ci} = \frac{\sum P_{m,n}}{n}, \quad (1)$$

$$\delta_{ci} = \sqrt{\frac{\sum (\bar{P}_{m,n} - P_{ci})^2}{n-1}}, \quad (2)$$

где n – количество значений;

P_{ci} и δ_{ci} – среднее значение и среднее отклонение соответственно выборки данных ряда.

Результаты вычисления сведем в таблицу 1.

Таблица 1

Средние значения и средние отклонения для рассматриваемых суток

№ суток	P_{ci}	δ_{ci}
1	2943,848	14031,542
2	2274,796	8013,179
3	1914,709	6168,584
4	2603,249	13912,409
5	3040,636	14552,2
6	2921,272	14172,124
7	2865,889	13847,83

Величины из таблицы 1 будем принимать как базисные (для их расчета используются реальные графики электрических нагрузок в течение семи суток) в последующем сравнительном анализе.

В качестве методов кластеризации для анализа были выбраны следующие:

- кластеризация на основании максимумов плотности вероятности;
- кластеризация графиков плотности вероятности по касательным;
- кластеризация с учетом адаптивного коэффициента отклонения относительно общей выборки значений;
- кластеризация с учетом вариации полиномов.

Основным принципом применения всех кластерных методов является исследование графиков электрических нагрузок. В виде оценочных показателей используется разброс значений суточного потребления относительно средней величины. При описании каждого из предложенных методов необходимо установить пределы и осуществить кластеризацию.

Таблица 2

Алгоритм кластеризации на основании максимумов плотности вероятности

Закон распределения	• $f(\bar{P}_{m,n})_{\max} = \frac{1}{\sqrt{2 \cdot \pi} \cdot \delta_a}$
Расчет пределов	• $f(\bar{P}_{m,n})_{\max} = \frac{f(\bar{P}_{m,n})_{\max p}^* + f(\bar{P}_{m,n})_{\max e}^{**}}{2}$
Пределы	• $f(\bar{P}_{m,n})_{\max} \in [0; \bar{f}(\bar{P}_{m,n})_{\max}]$ • $f(\bar{P}_{m,n})_{\max} \in [\bar{f}(\bar{P}_{m,n})_{\max}; \infty]$

* $f(\bar{P}_{m,n})_{\max p}$ – максимальное значение плотности вероятности, соответствующее максимальному значению для выборки первой группы;

** $f(\bar{P}_{m,n})_{\max e}$ – максимальное значение плотности вероятности, соответствующее минимальному значению для выборки второй группы.

Кластеризация на основании максимумов плотности вероятности

Первый вид кластеризации, который осуществляется путем сравнения максимумов плотности вероятности, является одним из наиболее распространенных методов анализа. Исследуемая величина определяет вероятностные количественные значения, сконцентрированные в середине интервала. При этом чем меньше расчетная переменная δ_{ci} , тем равномернее численные значения исходных величин. Применительно к электроэнергетике равномерный график потребления свидетельствует о том, что нагрузка без провалов и пиков. Данный случай, в основном, соответствует выходным дням, когда электроэнергия потребляется оборудованием, не обеспечивающим основной технологический процесс. Характеристики, которые целесообразно использовать для поставленных задач, предварительно оцениваются. Так как кластеризация является многомерной статистической процедурой, позволяющей формировать группы из сравнительно однородных исходных данных, то в качестве ее алгоритма можно принять решение, наведенное в таблице 2.

Метод представлен в виде графического отображения (рис. 1) с необходимой проверкой.

На основании табличных данных (табл. 2) и рассчитанной функции плотности нормального закона распределения строится график (рис. 1).

Проверка данного способа кластеризации графиков электрических нагрузок, проведенная путем ввода контрольной выборки, показала его соответствие заданным пределам.

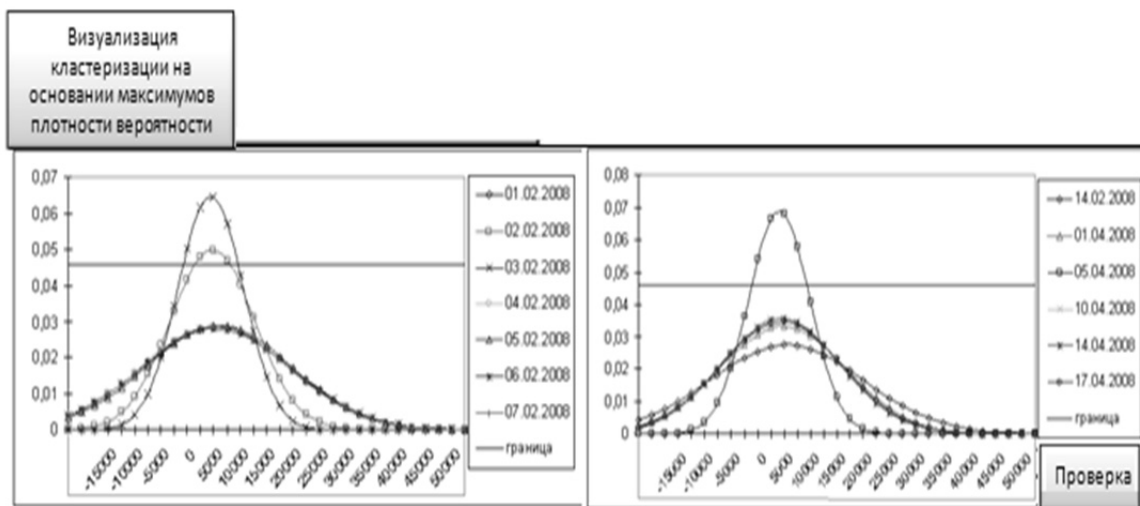


Рис. 1. Определение пределов и выполнения условий кластеризации плотности вероятности нормального закона распределения

Кластеризация графиков плотности вероятности по касательным

В результате исследования графиков плотности вероятности, распределенных по нормальному закону, был разработан способ определения пределов по касательным. Этот способ кластеризации основан на предыдущем методе, но он удобнее за счет автоматически создаваемых диапазонов, которые четко отображают границы разброса значений нагрузок для разных суток потребления. Алгоритм расчета аналогичен предыдущему методу, методика проведения кластеризации отличается (рис. 2).

Для проведения оптимального определения четких границ между группами предлагается использовать кластеры, установленные по касательным. Целесообразность применения приведенного метода заключается в определении параметра, по которому осуществляется кластеризация, с увеличением расстояния между кластерами, что приводит к снижению влияния величины потребления.

Кластеризация с учетом адаптивного коэффициента отклонения относительно общей выборки значений

Часто для оценки степени отклонения относительно среднего значения используется коэффициент вариации, который позволяет для групп объектов:

- сравнивать вариацию одного признака;
- выявить степень различия подобных признаков в разное время;
- сопоставить вариацию разных признаков в пределах одной группы.

Но при использовании данного коэффициента невозможно произвести четкую кластеризацию данных.

В результате математического анализа было выявлено, что если преобразовать схему определения выражения коэффициента вариации, то это приведет к возможности применения расчетных формул. В качестве нового параметра предлагается использовать адаптивный коэффициент отклонения относительно общей выборки значений (табл. 3).

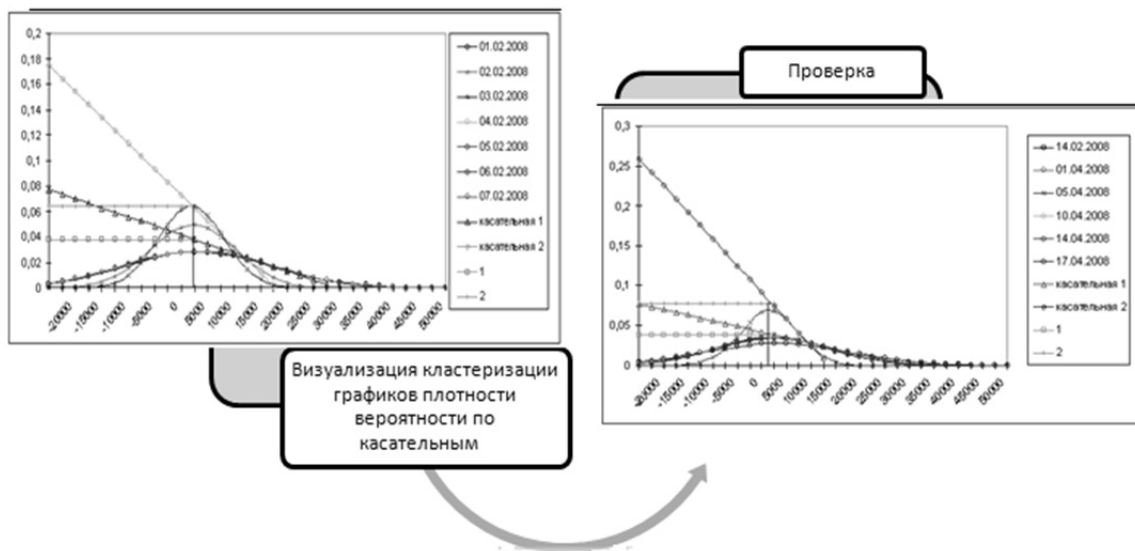


Рис. 2. Определение пределов и выполнения условий кластеризации плотности вероятности нормального закона распределения по касательным

Таблица 3

Алгоритм кластеризации с учетом адаптивного коэффициента отклонения относительно общей выборки значений

Закон распределения	<ul style="list-style-type: none"> $\beta = \frac{D_i^*}{\sum_{i,j=1}^{m,n} P_{m,n}}$
Расчет пределов	<ul style="list-style-type: none"> $\beta_{пред} = \frac{1}{2} \cdot \left(\frac{D_{e\max}^{**}}{\sum_{i,j=1}^{m,n} P_{m,n,e\max}^{***}} + \frac{D_{p\min}^{**}}{\sum_{i,j=1}^{m,n} P_{m,n,p\min}^{***}} \right)$
Пределы	<ul style="list-style-type: none"> $\beta \leq \beta_{пред}$ $\beta > \beta_{пред}$

* $D_i = \sum_{i,j=1}^{m,n} (P_{m,n} - \bar{P}_{ci})^2$ – дисперсия;

** $D_{e\max}$ и $D_{p\min}$ – дисперсия соответственно максимального потребления выходного дня и минимального потребления рабочего дня;

*** $\sum_{i,j=1}^{m,n} P_{m,n,e\max}$ и $\sum_{i,j=1}^{m,n} P_{m,n,p\min}$ – сумма средних значений мощности соответственно максимального потребления выходного дня и минимального потребления рабочего дня.

Результаты вычислений и проверку представим графически, проанализировав суточные ГЭН (рис. 3).

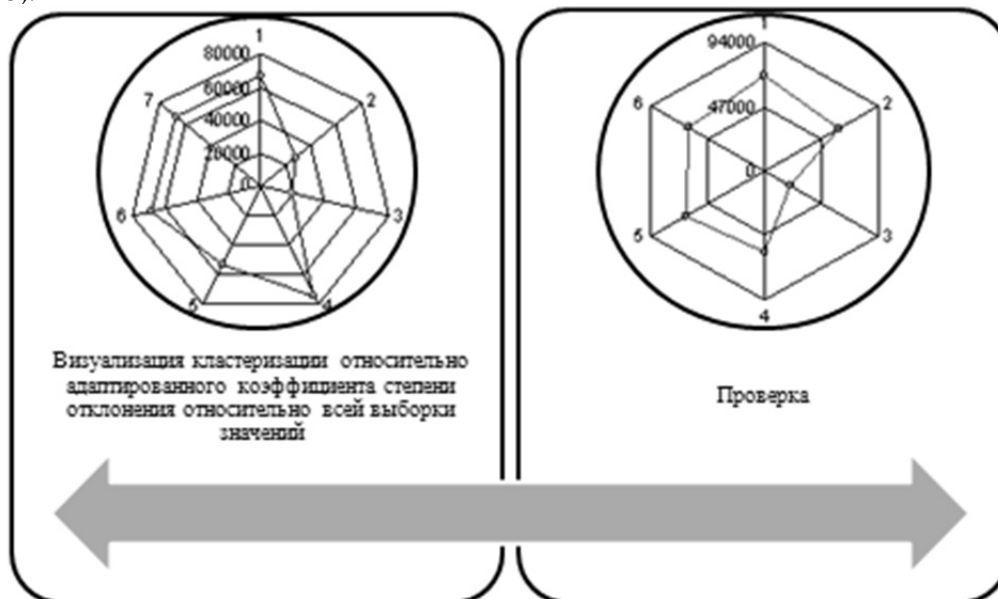


Рис. 3. Определение пределов и выполнения условий кластеризации величины адаптивного коэффициента степени отклонения относительно всей выборки значений

Как показали результаты проведения кластеризации, полученные группы соответствуют различным видам графиков потребления, а значит и есть решением поставленной задачи.

Кластеризация с учетом вариации полиномов

Одним из эффективных методов статистической кластеризации является комбинированный метод. Он часто используется при анализе двух классов объектов. Распространенность данного вида кластеризации обусловлена основными характеристиками метода, а именно:

- высокой точностью вычислений;
- высокой скоростью сходимости решений.

При сравнении рассматриваемых методов необходимо учитывать ряд показателей, а именно: быстроту установления пределов, простоту описания функций, оптимальный класс точности в соответствии с поставленными задачами. Первый метод из вышеприведенного перечня выбран в качестве сравнительного критерия, так как он достаточно хорошо разработан и широко применяется в вычислительной практике. Для данного способа кластеризации приведем методику решения в виде алгоритма (табл. 4).

В результате необходимо определить угол, вариация которого обеспечит проведение кластерного анализа. Вариация угла α – величины наклона характеристической прямой наглядно изображена на рис. 4.

Сформированные новые кластеры включают в себя ГЭН, вид которых соответствует выбранным ранее. Об этом свидетельствует четкое соблюдение расположения построенных прямых относительно предельной прямой.

Все вышеперечисленные методы представлены в виде сравнительного отображения (рис. 1) с необходимой проверкой. Данное графическое построение позволяет оценить не только степень расхождения, но и рационально выбрать метод кластеризации для решения задач планирования и управления электропотреблением.

В процессе сравнительного анализа при определении пределов кластеризации проводилось контрольное сравнение нормированных и текущих значений графиков с заданными допустимыми значениями.

Систематизация результатов методов кластеризации производится по скорости обработки исходных данных, информативности результирующих данных и обратимости исчислений к исходным значениям в виде табл. 5.

Таблиця 4

Алгоритм кластеризації з урахуванням варіації поліномів

Закон розподілення	<ul style="list-style-type: none"> • $P(\bar{P}) = A_0 + A_1 \cdot \bar{P}^r$
Расчет пределов	<ul style="list-style-type: none"> • Определяется по графику
Пределы	<ul style="list-style-type: none"> • $\alpha > \alpha_{ном}$ • $\alpha < \alpha_{ном}$

Визуализация кластеризации с использованием вариации угла

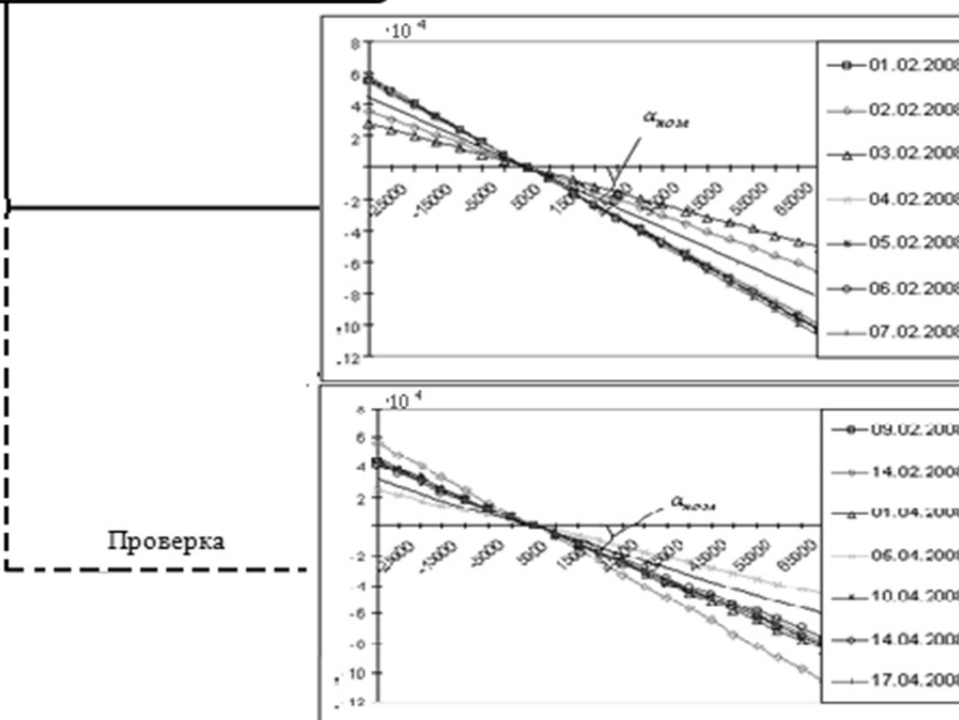


Рис. 4. Определение пределов и выполнения условий кластеризации с использованием вариации угла

Таблица 5

Систематизация статистических методов кластеризации вейвлет – преобразованных ГЭН с учетом их преимуществ и недостатков

Методы кластеризации	Быстрота обработки исходных данных	Информативность результирующих данных	Обратимость исчислений
Кластеризация на основании максимумов плотности вероятности	Быстрая сходимость, но необходимо задавать пределы	Определение группы	Восстановление степени отклонения
Кластеризация графиков плотности вероятности по касательным	Быстрая сходимость, но необходимо задавать пределы	Определение группы, взаимозаменяемость функций распределения	Восстановление степени отклонения
Методы кластеризации	Быстрота обработки исходных данных	Информативность результирующих данных	Обратимость исчислений
Кластеризация с учетом адаптивного коэффициента отклонения относительно общей выборки значений	Наиболее быстрый	Определение группы, вычисляется степень вариации данных	При нормировании среднего потребления можно установить степень отклонения
Кластеризация с учетом вариации полиномов	Средняя скорость сходимости и определения предела	Определение степени равномерности сигналов	Можно восстановить только точное среднее значение

В результате анализа табл. 5 следует отметить, что наиболее быстрый метод кластеризации – метод с учетом адаптивного коэффициента отклонения относительно общей выборки значений. Недостатком предложенного метода является низкий уровень преобразования результатов для дальнейшей обработки и анализа данных. В случае необходимости быстрого определения групп с точным восстановлением исходного значения степени отклонения наиболее целесообразно применять кластеризацию на основании максимумов плотности вероятности и кластеризацию графиков плотности вероятности по касательным. При этом следует учитывать, что корректность задания пределов непосредственно влияет на точность формирования кластеров. Если выборку исходных данных необходимо перепроверять на определение значительных погрешностей с целью их устранения, то при визуализации результатов кластеризации с учетом вариации полиномов отображается степень вариации используемых данных.

Сопоставим полученные результаты для разных предложенных методов кластеризации с существующим методом обработки данных. Для этого проведём автоматическую классификацию и группировку классов методом гиперсфер следующим образом. На первом этапе кластеризации за центр класса принимается k^1 , применяется первая выборка значений. Определяется для множества исходных выборок расстояние от k^1 к максимально удаленной группе.

$$d = \left[\sum_{i=1}^n (P_{m,ni} - P_{m,n1})^2 \right]^{1/2}. \quad (3)$$

Максимальное из этих расстояний принимается в качестве максимального радиуса R_{\max} гиперсферы. Все выборки $A_i \in A$ попадут в гиперсферу k^1 . Задаваясь радиусом $R_i = R_{\max} - \Delta R$, находятся выборки данных, для которых мера сходства удовлетворяет условию: $R(k^1, A_i) \leq R_i$. Для попавших в кластер выборок определяется центр тяжести, которым является вектор внутренних условий. В дальнейшем рассматривается множество $A \setminus k^1$. Описанная процедура

осуществляется для всех последующих значений R_i . Процесс кластеризации заканчивается на этапе, когда число выборок, входящих в класс, равно числу классифицируемых объектов.

Величина ΔR определяется согласно $\Delta R = R_{\max} / l$, где l – задаваемое число просмотров выборок.

Полученные результаты представим графически (рис. 5).



Рис. 5. Расстояние между определяемыми кластерами

Аналогично рассмотренным ранее методам кластеризации после осуществления кластеризации методом гиперсфер исходные выборки значений можно отнести к таким же кластерам.

Выводы

В результате проведенных исследований необходимо отметить следующее. Кластеризация графиков электрических нагрузок при их начальном вейвлет-преобразовании обоснована в условиях нелинейности и аперидичности исходной информации. В зависимости от поставленных задач выбирается наиболее эффективный метод расчета. При необходимости обеспечения высокого уровня информативности результирующих данных целесообразно применять кластеризацию с учетом полиномов, для обеспечения значительной скорости вычислений – кластеризацию с учетом адаптивного коэффициента отклонения относительно общей выборки значений, для применимости с последующей обработкой данных и обратимости вычислений – кластеризацию на основании максимумов плотности вероятности и кластеризацию графиков плотности вероятности по касательным.

Предложенные методы различаются алгоритмами решений, но приводят к одному и тому же результату. Для дополнительного обоснования истинности того или иного описанного метода кластеризации в качестве проверочного был выбран существующий метод гиперсфер. Детальный анализ показал, что во всех случаях кластеры включают одинаковый набор графиков электрических нагрузок.

Литература

1. Гайдышев И. Анализ и обработка данных: специальный справочник/ И. Гайдышев – СПб: Питер, 2001. – 752 с.
2. Волошко А. В. Метод формирования признаков классификации графиков электрических нагрузок на основе вейвлет – преобразования/ А. В. Волошко. - Промэлектро / Научн. журнал. - 2009. - № 1. - С. 39-43.
3. Дворников С. В. Метод распознавания на основе вейвлет – пакетов/ С. В. Дворников, А.М.Сауков. - Научное приборостроение, 2004. – т. 14, №1. – С. 57-65.
4. Prakhovnik A.V. Development of Techniques for Daily Load Prediction of Power Systems Using

ISSN 1813-5420

Енергетика: економіка, технології, екологія

Adaptive Methods of Forecasting/ Prakhovnik A.V. // Report of Delt University of Technology, Netherlands. – 1977. – 66 p.

5. Волошко А.В. Статистическая кластеризация информационных сигналов на основе вейвлет – преобразования/ А.В. Волошко, Т.Н. Лутчин / Енергетика: економіка, технології, екологія. – 2009. - №1. - С. 80-86.
-
-
-
-