

УДК 004.62:621.311

МОДЕЛЬ ХРАНЕНИЯ И ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ РЕТРОСПЕКТИВНЫХ ДАННЫХ ПРИ РЕШЕНИИ ЗАДАЧИ КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ ЭЛЕКТРОПОТРЕБЛЕНИЯ

Р. Н. ХАМИТОВ¹, И. В. ЧЕРВЕНЧУК², А. С. ГРИЦАЙ³

¹apple_27@mail.ru, ²cherven-igor@yandex.ru, ³aleksandr.gritsay@gmail.com

ФГБОУ ВО «Омский государственный технический университет» (ОмГТУ)

Поступила в редакцию 14.10.2016

Аннотация. В настоящее время для субъекта оптового рынка электроэнергии – энергосбытовой компании – одним из основных бизнес-процессов является процесс решения задач краткосрочного прогнозирования электропотребления. При этом эксперту, как правило, приходится работать с большим массивом ретроспективных данных об электропотреблении и их выборками. Такие массивы данных имеют простую однородную структуру. При этом возникает проблема, связанная с объединением данных из различных источников и необходимо затратить длительное время на получение различных выборок данных, в связи с чем автоматизация предварительной обработки данных для последующего статистического анализа в условиях большого объема исходной информации об электропотреблении является актуальной.

Ключевые слова: анализ данных, краткосрочное прогнозирование электропотребления, информационное моделирование.

При построении прогностических моделей электропотребления [1, 2], субъект Оптового рынка электроэнергии и мощности (ОРЭМ), как правило, оперирует данными систем автоматизированного коммерческого учета (АСКУЭ). Такие данные могут быть представлены в формате XML 80020 и имеют простую однородную структуру. Пример таких данных приведен в табл. 1. Однако в настоящее время имеется большое количество предприятий-участников Оптового рынка электроэнергии с распределенной структурой. В этом случае объекты компании находятся на значительном территориальном удалении друг от друга. В этом случае целесообразно осуществлять прогноз для каждого объекта отдельно. Подобная ситуация возникает для энергосбытовой компании, которая может также являться субъектом оптового рынка электроэнергии. Часть потребителей энергосбытовой компании использует автоматизированные системы коммерческого учета электроэнергии (АСКУЭ), другая часть потребителей не имеет таких систем, что во многом усложняет процесс прогнозирования их электропотребления. В этом случае задача может быть решена с использова-

нием ценологического подхода [3]. С этой целью необходимо для всего объема имеющихся потребителей и их объектов сформировать ценологическую модель. В этой связи информационная система, при помощи которой будут реализовываться ранговые распределения, должна располагать возможностью осуществления операций суммирования/усреднения параметрических данных объектов, например данных об электропотреблении, метеорологических параметрах о погоде, освещенности, входящих или предполагаемых к включению в определенный ранг по заранее сформированным критериям. Поскольку в энергосбытовой компании количество таких объектов в биллинговой системе может быть достаточно большим (более 100000 объектов) процесс построения рангов усложняется и требует наличие производительного программно-аппаратного комплекса, а также соответствующим образом подготовленной структуры данных для осуществления возможности хранения динамически сформированных техноценозов [4]. С этой целью для реализации ранговой модели распределения, была разработана специальная структура хранения данных.

Таблица 1

Фрагмент данных субъекта ОРЭМ об электропотреблении

Дата Date	Час суток Hour	Температура, °C Temperature, °C	Электропотребление, МВт/ч Electricity load, MWh
01.01.2015	1:00AM	-12	718290,0
01.01.2015	2:00AM	-12	695724,0
01.01.2015	3:00AM	-12	678426,0

Для первичного хранения ретроспективных данных об электропотреблении, может быть с успехом использована модель данных, описываемая в терминах объектов и показателей [5]. В предлагаемой модели определяются объекты-носители свойств, описываемые посредством системы показателей. В общем случае предполагается, что объект описания является структурированным, состоящим из концептуально однородных элементов, иерархически связанных между собой. Система показателей так же частично или полностью структурирована, иерархична. Таким образом, выделяется две иерархии – иерархия объектов и иерархия показателей (свойств), используемых при описании ретроспективных данных об электропотреблении. Далее рассматриваемую модель будем называть моделью показатель-объект, основное назначение которой – обеспечить удобное хранение структурированных данных об электропотреблении объектов потребителей для последующего статистического анализа. Применение такого подхода при описании данных позволяет в естественной и компактной форме представить данные для хранения и использования [6], ввести характеристики компактности хранения данных [7].

На рис. 1 в нотации языка UML [8] показана структура данных системы, построенная на основе двух иерархий: иерархии объектов, и иерархии показателей, ориентированная на концептуально однородные данные, описываемые числовыми значениями.

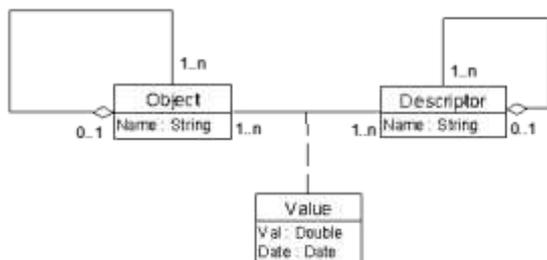


Рис. 1. Структура модели показатель-объект

Первая иерархия – иерархия объектов задается рефлексивной агрегацией класса Object(Объект), предполагающей отношение типа «часть-целое» между элементами структурированного объекта, например, цепочка «По-

требитель, объект потребления, узел учета, прибор учета». Вторая иерархия – система показателей, задается рефлексивной агрегацией класса Descriptor (Показатель), предполагающей структурированное описание показателя, например, электропотребление объекта. Существует принципиальное отличие иерархии объектов и показателей в том, что любой объект, независимо от его места в иерархии объектов, может иметь числовое значение, тогда как только полностью определенный показатель – элемент конца иерархии, имеет конкретные значения. Дерево показателей по некоторым свойствам схоже с деревом множественных атрибутов (MAT – деревом [9]).

Собственно значение сопоставляется не отдельно с каким-либо показателем и не с каким-либо определенным объектом, а с парой объект-показатель, поэтому класс Value (Значение) является классом-ассоциацией между объектами и показателями. Наличие атрибута Date (Дата), содержащего дату и время измерения электропотребления на данном объекте, позволяет впоследствии сортировать данные по датам, проводить анализ данных, построить язык манипулирования данными с широкими возможностями [10].

Иерархия является естественным и удобным средством поиска упорядоченных данных, основой для построения поискового иерархического меню, позволяющего эксперту легко обеспечить навигацию к конечным данным динамически сформированной ценологической модели.

Рассмотренная модель хранения данных легко отображается на реляционную модель данных. Реляционный профиль модели приводится на рис. 2.

Описание структуры объектов и показателей обеспечивается посредством таблиц соответственно TObject и TDescriptor, имеющих ссылки на себя. Каждая запись данных таблиц содержит ссылку на идентификатор записи более высокого уровня иерархии (ID_PreObj и ID_PreDes). Собственно значения хранятся в таблице значений TValue, которая в качестве первичного ключа содержит пару идентификатор объекта – идентификатор показателя. Записи таблицы значений кроме собственно значений могут содержать

дату/время измерения и служебные атрибуты, позволяющие обеспечить дополнительные возможности по предварительной обработке и манипулированию данными.

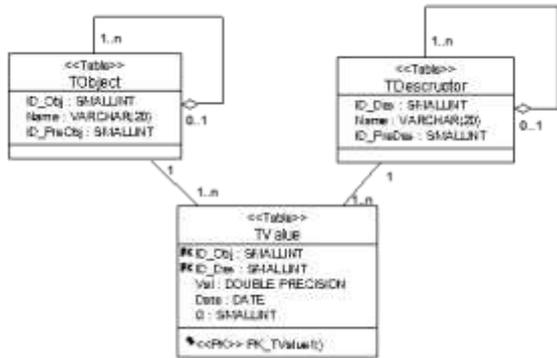


Рис. 2. Реляционный профиль модели показатель–объект

Иерархические структуры показателей (x) и объектов (y) с соответствующими значениями (z) и датами их измерений (d) задают четырехмерное информационно-поисковое пространство (ИП-пространство), обладающее рядом специфических свойств. Концепция ИП-пространства позволяет формализовать процедуры предварительной подготовки данных ценологическому анализу.

Использование предлагаемого аппарата, который назовем аппаратом исчисления выборок, позволяет формализовать задачу предварительной подготовки ценологических данных для построения рангового распределения. С этой целью используется временная привязка исходных данных, позволяющая накапливать данные о ценологической модели с возможностью учета ее динамики и изменения.

В нашем случае выборкой будем называть упорядоченное по времени множество значений для всех моментов времени некоторого показателя на некотором объекте. Выборка содержит данные. Данным будем называть пару <дата, значение>. Таким образом, выборка – это упорядоченное множество (отсортированная по возрастанию атрибута дата/время последовательность) данных. В данном контексте выборка может быть представлена таблицей из двух столбцов (табл. 2).

Таблица 2

Пример выборки

Date	Value
1.01.2015	12,1
1.02.2015	13,2
1.03.2015	11,7
1.04.2015	11,2

Введем операцию сложения выборок ($\overset{\cup}{+}$). Пусть у нас есть два объекта одного ранга, результаты наблюдения за объектом из каждого источника наблюдений образуют выборки v_1 и v_2 соответственно. Выборку v_3 – сумму выборок $v_1 \overset{\cup}{+} v_2$ можно найти, дополнив выборку v_1 данными из v_2 , имеющие даты которых нет в v_1 , если данные v_1 и v_2 имеют эквивалентные даты, то в v_3 заносится среднее по значению для выборок v_1 и v_2 .

В качестве эквивалентности дат можно использовать равенство или принадлежность к определенному периоду (неделе, декаде, месяцу и т.д.)

Очевидно, что операция сложения выборок коммутативна, идемпотентна, однако ассоциативность выполняется только в том случае, когда данные с эквивалентными данными встречаются не более одного раза. Данная операция над выборками определяется соответствующей операцией над данными.

Существует 0 операции – пустая выборка – v_0 , которая не содержит ни одного данного (или можно рассматривать, что все данные содержат значения NULL – пропуски). При этом для любой выборки v_1

$$v_0 \overset{\cup}{+} v_1 = v_0 \overset{\cup}{+} v_1 = v_1. \tag{1}$$

Операцию сложения выборок можно также применять для разных объектов и показателей, тогда данная операция будет иметь смысл обобщения.

Однако ввести операцию типа разности выборок не представляется возможным.

Таким образом, на множестве допустимых выборок V имеем алгебру A^+ с двуместной операцией сложения выборок и нульместной операцией – пустая выборка. Данная алгебра A^+ представляет собой идемпотентный коммутативный группоид с единицей и в таком виде мало удобна для использования.

Аналогично на множестве выборок V можно построить алгебру $A^{\wedge+}$. Для этого введем операцию $\overset{\wedge}{+}$. Выборку v_3 – сумму по пересечению выборок $v_1 \overset{\wedge}{+} v_2$ можно найти по правилу: если данные v_1 и v_2 имеют эквивалентные даты, то в v_3 заносится среднее по значению для выборок v_1 и v_2 , данные с датами встречающимися только в одной выборке не включаются в v_3 . Очевидно, что операция сложения выборок по пересечению также коммутативна и идемпотентна. Таким образом, алгебра $A^{\wedge+}$, содержащая операцию $\overset{\wedge}{+}$ и пустую выборку, является также идемпотентным коммутативным группоидом с единицей.

Для операций $\hat{+}$ и $\check{+}$ выполняется дистрибутивный закон

$$v1 \hat{+} (v2 \check{+} v3) = (v1 \hat{+} v2) \check{+} (v1 \hat{+} v3). \quad (2)$$

Дополним данное (т.е. пару <дата, значение>) дополнительным атрибутом (G), определенным на множестве целых чисел и содержащим число усреднения данного, такое данное будем называть расширенным данным. Расширенным данными будем называть тройку <дата, значение, число усреднения>. Число усреднения показывает, сколько данных участвовало в получении значения как среднего при эквивалентности дат, изначально для всех данных выборки степень усреднения равна единице. Таким образом, расширенная выборка – это упорядоченное множество (отсортированная по возрастанию даты/времени последовательность) данных. В данном контексте выборка может быть представлена таблицей из трех столбцов, как показано в табл. 3

Таблица 3

Пример расширенной выборки

Date	Value	G
1.01.2015	12,1	1
1.02.2015	13,2	2
1.03.2015	11,7	1
1.04.2015	11,2	1

Изначально для всех данных выборки степень усреднения равна 1.

На множестве расширенных выборок $V_{\check{+}}$ введем операцию сложения ($\check{+}$) – точнее операцию сложения по объединению расширенных выборок. Пусть у нас $v3 = v1 \check{+} v2$. Если для некоторых данных из $v1$ и $v2$ даты эквивалентны, то

$$\begin{cases} v3.g = v1.g + v1.g; \\ v3.val = \frac{v1.val \cdot v1.g + v2.val \cdot v2.g}{v1.g + v2.g} \end{cases} \quad (3)$$

Если даты не совпадают, т.е. для некоторого данного с датой/временем d одного операнда не существует данного с эквивалентной датой в другом операнде (или существует $NULL$ -данное), то значения переносятся в результат. Здесь число усреднения данного показывает, какое число значений участвовало в усреднении. В табл. 6 представлена сумма расширенных выборок из табл. 4 и 5.

Таблица 4

Выборка v1		
Date	Value	G
1.01.2015	10,2	1
1.03.2015	12,1	1
1.05.2015	13,4	2
1.06.2015	11,7	1
1.07.2015	11,2	1

Таблица 5

Выборка v2		
Date	Value	G
1.01.2015	10,8	1
1.02.2015	4,0	2
1.04.2015	7,0	1
1.05.2015	13,2	2
1.06.2015	11,5	1

Таблица 6

Выборка v3 = v1 $\check{+}$ v2		
Date	Value	G
1.01.2015	10,5	2
1.02.2015	4,0	2
1.03.2015	12,1	1
1.04.2015	7,0	1
1.05.2015	13,3	4
1.06.2015	11,6	2
1.07.2015	11,2	1

Сумма большого числа выборок будет стремиться к вектору математических ожиданий некоторого показателя по разным датам. Очевидно, что данная операция ассоциативна, коммутативна, однако строго не является идемпотентной.

Алгебра, образуемая операцией $\check{+}$ на $V_{\check{+}}$, является коммутативной полугруппой (по поводу нуля операции – пустой выборке – будет сказано далее).

На расширенных выборках условно можно задать операцию, противоположную суммированию – разность $\check{-}$ (точнее разность расширенных выборок по объединению).

Пусть $v3 = v1 \check{-} v2$. Если для некоторых данных из $v1$ и $v2$ даты эквивалентны, то

$$\begin{cases} v3.g = v1.g - v1.g; \\ v3.val = \frac{v1.val \cdot v1.g - v2.val \cdot v2.g}{v1.g - v2.g} \end{cases}, \quad (4)$$

если по некоторой дате/времени $d1$ данное вида $\langle d1, z1, g1 \rangle$ есть только в одной выборке то будем считать, что в другой выборке присутствует $NULL$ -данное вида $\langle d1, \frac{0}{0}, 0 \rangle$ и применим (4).

Использование неопределенности подобной $\frac{0}{0}$, в данном случае допустимо, поскольку при последующих вычислениях данная неопределенность снимается, так как и в (3) и в (4) значение данного ($v.val$) всегда умножается на степень усреднения ($v.g$), т.е.

$$v.val \cdot v.g = \frac{0}{0} \cdot 0 = 0.$$

Таким образом, для расширенных выборок $NULL$ -данные имеют четкую интерпретацию как

данные вида $\langle d1, \frac{0}{0}, 0 \rangle$ (где $\frac{0}{0}$ – *NULL*-значение), и результат суммы и разности некоторого данного с *NULL*-данными равняется этому данному. Дополним выборку $v1$, представленную в табл. 4, *NULL*-данными (табл. 7) и вычтем ее из выборки $v3$ представленной в табл. 3. В результате получим выборку $v2$, которая была ранее представлена в табл. 5, дополненную *NULL*-данными (табл. 8).

Таблица 7

Выборка $v1$, дополненная *NULL*-данными

Date	Value	G
1.01.2015	10,2	1
1.02.2015	$\frac{0}{0}$	0
1.03.2015	12,1	1
1.04.2015	$\frac{0}{0}$	0
1.05.2015	13,4	2
1.06.2015	11,7	1
1.07.2015	11,2	1

Таблица 8

Выборка $v2 = v3 - v1$

Date	Value	G
1.01.2015	10,8	1
1.02.2015	4,0	2
1.03.2015	$\frac{0}{0}$	0
1.04.2015	7,0	1
1.05.2015	13,2	2
1.06.2015	11,5	1
1.07.2015	$\frac{0}{0}$	0

Операции над расширенными выборками определяются соответствующими операциями над расширенными данными, нулями этих операций являются *NULL*-данные. Предполагается, что каждая выборка содержит необходимое количество *NULL*-данных (это количество зависит от данных в других выборках). *NULL*-выборка ($v0$) представлена необходимым количеством *NULL*-данных, при этом все *NULL*-выборки эквивалентны и содержательно являются пустыми. Данный подход позволяет выполнять операции сложения и разности, непосредственно используя только формулы (3) и (4) для всех соответствующих данных.

При применении операций $+$ и $-$ в произвольном порядке возможно появление фиктивных данных со степенью усреднения $g = 0$, которые при дальнейших операциях будут устранены. Фиктивные данные не участвуют во внешних операциях, а используются только при

манипуляциях выборками. *NULL*-данные являются фиктивными.

Для расширенных выборок справедливо тождество

$$v2 - v1 + v1 = v2. \tag{5}$$

При этом фиктивные данные, полученные при $- v1$, устраняются при $+ v1$.

Операция $-$ ассоциативна и ассоциативна с $+$.

Каждой выборке $v1$ можно сопоставить противоположный элемент $-v1 = v0 - v1$, при этом в силу (5) всегда будет выполняться: $v1 + v1 = v1 + (-v1) = v0$. Очевидно, что операция $-$ и взятие противоположного элемента по сути одна и та же операция.

Таким образом, на множестве расширенных выборок V_- имеем алгебру $A+/-$ с двуместной операцией сложения выборок, одноместной операцией взятия противоположного элемента и нульместной операцией – пустая выборка. Данная алгебра $A+/-$ является абелевой группой.

Для исходных выборок существует однозначное отображение $h^{-1}: V_- \rightarrow V_+$, при котором для всех данных (кроме неявно присутствующих фиктивных) в атрибут усреднения заносится 1 (единица).

Вычисления над выборками можно производить следующим образом: сначала посредством h^{-1} перейдем от исходных выборок к расширенным выборкам, используя операции над расширенными выборками, произведем вычисления, окончательный результат преобразуем посредством h в обычную выборку.

При выполнении сложных долговременных расчетов промежуточные результаты нужно хранить в расширенном формате, исходные данные и результаты в обычном.

Над подобными выборками могут быть введены операции сопоставления и совмещения, позволяющие представить весь спектр статистических вычислений, что может служить основой создания своеобразного языка манипулирования данными.

Привлекательность предлагаемого подхода заключается, прежде всего, в том, что он позволяет формализовать и автоматизировать процессы подготовки и предварительной обработки данных биллинговых систем энергосбытовых компаний, для составления требуемой модели данных, с целью последующего статистического анализа. Предложенная модель была с успехом

использована для автоматизации процессов предварительной обработки и хранения данных при решении задач краткосрочного прогнозирования электропотребления и легла в основу разработанной программы для краткосрочного прогнозирования электропотребления [11].

Предложенную модель можно также использовать для предварительной обработки, хранения и анализа первичной информации о внешних факторах, например, метеорологических, таких как температура, освещенность, что дает методологическую основу для построения более точных прогнозов электропотребления, учитывающих влияние внешних факторов.

СПИСОК ЛИТЕРАТУРЫ

1. **Грицай А. С., Тюньков Д. А.** Классификация методов краткосрочного прогнозирования электропотребления для субъектов ОРЭМ // Актуальные вопросы энергетики : материалы Всерос. науч. конф. студентов, магистрантов, аспирантов (Омск, 12 мая 2016). Омск: ОмГТУ, 2016. С 41–45. [A. S. Gritsay and D. A. Tyunkov, "Classification methods for predicting the short-term electricity OREM subjects," (in Russian), in *Proc. All-Russia Workshop on Actual energy issues*, Omsk, Russia, 2016, pp. 41-45.]
2. **Использование** нейронной сети для построения краткосрочного прогноза электропотребления ООО «Омская энергосбытовая компания» / В. И. Потапов [и др.] // Известия томского политехнического университета. Инжиниринг георесурсов. 2016. Т. 327, № 8. С.44–51. [V. I. Potapov, et al., "Using neural network to construct a short-term forecast of electricity consumption LLC "Omsk Energy Retail Company"," (in Russian), in *Izvestiya tomского politekhnicheskogo universiteta. Inzhiniring georesursov*, vol. 327, no. 8, pp. 44-51, 2016.]
3. **Исследование** технических систем типа электрический ценоз / Б. И. Кудрин [и др.] // деп. Информэнерго 25.11.85 г. № 2002-эн. М.: Изд-во МЭИ, 1985. 52 с. [B. I. Kudrin, et al., "Research of technical systems such as electric cenosis," (in Russian), deposited in Informenergo 11/25/1985 N 2002-эн. Moscow: Moscow Power Engineering Institute "MEI", 1985.]
4. **Гнатюк В. И.** Закон оптимального построения техноценозов. М.: Совместное издание ТГУ и Центра системных исследований, 2005. [V. I. Gnatyuk, *Law optimum construction technocenosis*, (in Russian). Moscow: TGU and Centr sistemnyh issledovaniy, 2005.]
5. **Червенчук И. В.** Применение иерархических структур для описания статистических данных // Динамика систем, механизмов, машин: материалы VII Междунар. науч.-техн. конф (Омск, 10–12 нояб. 2009). Омск: Изд-во ОмГТУ, 2009. Т. 1. С. 426–429. [I. V. Chervencuk, "The use of hierarchical structures for describing statistics," (in Russian), in *Proc. 7th Int. Workshop on Dynamics of Systems, Mechanisms and Machines*, Omsk, Russia, 2009, vol. 1, pp. 426-429.]
6. **Червенчук И. В.** Модель представления статистических данных // Природные и интеллектуальные ресурсы Сибири (СИБРЕСУРС-19-2013): материалы 19-й междунар. науч.-практ. конф. (Новокузнецк, 6–8 нояб. 2013). Томск: Изд-во «В-Спектр», 2013. С.187–191. [I. V. Chervencuk, "Model presentation of statistical data," (in Russian), in *Proc. 19th Int. Workshop on Natural and intellectual resources of Siberia (SIBRESURS-19-2013)*, Tomsk, Russia, 2013, pp. 187-191.]
7. **Червенчук И. В.** Характеристики эффективности структуры показателей // Омский научный вестник. 2002. № 20. С. 149–151. [I. V. Chervencuk, "Characteristics efficiency indicators of structure," (in Russian), in *Omskiy nauchnyy vestnik*, no. 20, pp. 149-151, 2002.]
8. **Буч Г., Рамбо Дж., Джекобсон И.** Язык UML Руководство пользователя / пер. с англ. Мухин Н. М.: ДМК Пресс, 2006. 496 с. [G. Booch, J. Rumbaugh and I. Jacobson, *The Unified Modeling Language User Guide (2nd Edition)*, (in USA) in Addison-Wesley Professional, 2005.]
9. **Вирт Н.** Алгоритмы + структуры данных = программы/ пер. с англ. М: Мир, 1985. 406 с. [N. Wirth, *Algorithms + Data Structures = Programs*, (in USA), in Prentice-Hall, 1976.]
10. **Червенчук И. В.** Язык манипулирования данными системы статистической обработки информации на основе исчисления выборок // Динамика систем, механизмов, машин: материалы IV Междунар. науч.-техн. конф., посвященной 60-летию ОмГТУ (Омск, нояб. 2002). Омск: Изд-во ОмГТУ, 2002. Т. 1. С. 245–247. [I. V. Chervencuk, "Language manipulate data processing system of statistical information on the basis of sample calculation," (in Russian), in *Proc. 4th Int. Workshop on Dynamics of Systems, Mechanisms and Machines*, Omsk, Russia, 2002, vol. 1, pp. 245-247.]
11. **Хамитов Р. Н., Грицай А. С., Тюньков Д. А.** Программа краткосрочного прогнозирования электропотребления для субъекта оптового рынка электроэнергии и мощности энергосбытовой компании с использованием аппарата искусственной нейронной сети. М.: ОФЭРНИО, 2016. Свидетельство о регистрации электронного ресурса № 22122 от 01 сентября 2016. [R. N. Khamitov, A. S. Gritsay and D. A. Tyunkov, "Program of short-term forecasting of electricity consumption using the artificial neural network for wholesale electricity and capacity market energy retail companies," (in Russian). Moscow: OFERNiO, 2016. The certificate of registration of an electronic resource N 22122 on September 1, 2016]

ОБ АВТОРАХ

ХАМИТОВ Рустам Нуриманович, проф. каф. Электрическая техника (ОмГТУ, 1984). Обл. иссл.: динамика электротехн. комплексов, включая подсистемы различной физической природы, управляемая виброударозащита крупногабаритных объектов, электромех. демпфирование.

ЧЕРВЕНЧУК Игорь Владимирович, доц. каф. Информатика и вычислительная техника (ОмГТУ, 1999). Обл. иссл.: анализ данных.

ГРИЦАЙ Александр Сергеевич, ст. препод. каф. Информатика и вычислительная техника (ОмГТУ, 2004). Обл. иссл.: анализ данных, прогнозирование электропотребления.

METADATA

Title: Data model for storage pretreatment retrospective data in electricity load field.

Authors: R.N. Khamitov¹, I.V. Chervenчук², A.S. Gritsay³

Affiliation:

Omsk State Technical University (OmSTU), Russia.

Email: ¹ apple_27@mail.ru, ² cherven-igor@yandex.ru, ³ aleksandr.gritsay@gmail.com

Language: Russian.

Source: Vestnik UGATU (scientific journal of Ufa State Aviation Technical University), vol. 20, no. 4 (74), pp. 125-131, 2016. ISSN 2225-2789 (Online), ISSN 1992-6502 (Print).

Abstract: At this time, for the subject of the wholesale electricity market, one of the main business process is the process of solving the problems of short-term forecasting of power consumption. In this case, the expert, as a rule, have to work with large sets of historical data on power consumption and their samples. On the one hand, these data sets have a simple uniform structure, on the other hand there is a problem with data fusion from multiple sources and for a long time for a variety of data samples, and therefore, the automation of data pre-processing for statistical analysis in a large the volume of initial information about the power consumption is one of the important requirements.

Key words: data mining, forecast electricity loads, modeling.

About authors:

KHAMITOV, Rustam Nurimanovich, Prof., Dept. of Electricity engendering (OmSU, 1984)

CHERVENCHUK, Igor Vladimirovich, Senior Lecturer of Computer Science Department (OmSU, 1999)

GRITSAY, ALEXANDER SERGEEVICH, Senior Lecturer of Computer Science Department (OmSU, 2004)