

УДК 681.518, 004.032.26

## АНАЛИЗ МЕТОДОВ ОБНАРУЖЕНИЯ НЕТЕХНИЧЕСКИХ ПОТЕРЬ ЭЛЕКТРОЭНЕРГИИ

Чухров Алексей Александрович

Аспирант;

ГБОУ ВО МО «Университет «Дубна»,

Институт системного анализа и управления;

141980, Московская обл., г. Дубна, ул. Университетская, 19;

e-mail: alexeych0@gmail.com.

*Применение технологий интеллектуального анализа данных в области обнаружения мошенничества привело к успехам в выявлении мошеннических транзакций в банковской сфере и электронной коммерции. В тоже время, несмотря на актуальность проблемы хищения электроэнергии, на данный момент, не существует эффективного решения обнаружения фактов безучетного потребления, используя имеющиеся в энергетических компаниях потоки данных. В этой статье представлено исследование проблемы хищения энергии, сравнительный анализ современных методов обнаружения источников нетехнических потерь.*

Ключевые слова: экономическая безопасность, энергетика, *Data mining*, прогнозирование электропотребления.

### ANALYSIS OF METHODS DETECTING NON-TECHNICAL LOSSES OF ELECTRICITY

Chukhrov Alexey

Postgraduate student;

Dubna State University,

Institute of system analysis and management;

141980, Dubna, Moscow reg., Universitetskaya str., 19;

e-mail alexeych0@gmail.com.

*The application of data mining in fraud detection has led to advances in detection of fraudulent transactions in banking and e-commerce. At the same time, despite the urgency of the problem of energy thefts, currently, there is no effective solution to detect unmetered consumption, using existing data streams in energy companies. This paper presents a study of the problem of energy thefts, analysis of the available data streams, opportunities and challenges of using these data to reveal the facts of energy thefts.*

Keywords: economic security, power utilities, *Data mining*, electricity consumption forecasting.

### Введение

Одним из главных показателей эффективности электроэнергетического комплекса является поддержание потерь при передаче и потреблении электроэнергии на минимальном уровне. Основным видом таких потерь являются нетехнические коммерческие потери, происходящие главным образом, от фактов хищений и безучетного потребления электроэнергии. Нетехнические потери наносят ущерб до 15% от общей годовой выручки энергокомпании [1].

Хищения в значительном большинстве случаев совершаются юридическими высоковольтными потребителями [1] (компании, производственный сектор, и пр.) путем различного рода мошеннических и коррупционных действий, направленных на искажение достоверной информации о количестве потребляемой электроэнергии. Также в связи с их объемом потребления, наносимый ущерб, во много раз превосходит ущерб от хищений в частном секторе.

На практике, искажения достоверной информации о количестве потребляемой энергии достигается злоумышленником, в основном, путем различного рода механических воздействий на прибор учета, программными средствами (прошивки), махинациями с погрешностью счетчика, подключением дополнительной мощности до прибора учета с помощью скрытой проводки, а также многими другими действиями [1].

Трудности обнаружения фактов занижения реального расхода потребляемой энергии, связаны с большими объемами поставляемой энергии, влиянием сезонного фактора на объемы потребления энергии, влиянием географического положения и общей экономической ситуации региона, большим количеством потребителей, наличием технических потерь, наличием дифференцированных нагрузок на электрические сети, погрешностями измерительных приборов, а также применением различных тарифных политик к потребителям.

На протяжении последних десяти лет в отрасли электроэнергетики активными темпами происходила модернизация технического оборудования, совершенствование измерительного процесса и биллинговых систем учета электроэнергии. Полученный в результате эффект заключается в появлении новых источников информации и данных больших размерностей (*Big Data*). Современные измерительные устройства [2] позволяют считывать данные об электропотреблении с частотой раз в 15 минут, что дает 96 замеров в день и почти 3000 в месяц [3]. Показания хранятся на самом устройстве, и передаются компании в конце каждого месяца. Эти, и другие потоки данных можно использовать для обнаружения источников нетехнических потерь.

Для эффективного решения задачи, методика должна отвечать современным требованиям к системам прогнозирования электропотребления [4]: учитывать факторы, влияющие на электропотребление, использовать дополнительную информацию, косвенно связанную с электропотреблением, предоставлять возможность моделирования различных сценариев нетехнических потерь и др.

В целом задачу обнаружения нетехнических потерь можно рассматривать как приложение задачи краткосрочного прогнозирования потребления электроэнергии, сопоставив реальные данные с полученными в результате прогноза можно найти искомые отклонения и аномалии. Результат зависит от точности и прогнозной модели. Далее рассмотрим существующие методы на основе прогнозирования потребления электроэнергии на основе статистической обработки данных биллинговых систем энергетических компаний и другой вспомогательной информации.

## ***Методы линейного регрессионного анализа***

Методы линейной регрессии используются для моделирования взаимосвязи электропотребления с другими факторами, такими как погода (температура, освещенность, влажность), класс потребителей, тип дня (выходной, рабочий).

Задача регрессионного анализа в общем случае решается в несколько шагов:

1. Предобработка исходных данных.
2. Определение вида уравнения регрессии.
3. Расчет коэффициентов уравнения регрессии.
4. Проверка адекватности полученной модели.

В работах [5], [6] описываются различные приложения линейных регрессионных методов для прогнозирования электропотребления. Методы линейного регрессионного анализа используются для краткосрочного прогнозирования электропотребления, когда необходимо составить прогноз в зависимости от одного и или нескольких влияющих факторов. Применение этих методов для долгосрочного прогнозирования неэффективно, т.к. предполагается неизменность коэффициентов значений регрессионной модели [7], появление новой информации требует корректировки модели.

Также, делается допущение о существовании линейной зависимости, между рассматриваемыми факторами и электропотреблением, хотя это не совсем так. Например, между погодными переменными и электропотреблением существуют нелинейные зависимости, что в итоге оказывает влияние на точность прогноза. Таким образом, линейные регрессионные модели рассматривают достаточно ограниченный набор факторов и не позволяют учитывать косвенные факторы, связанные нелинейными зависимостями с электропотреблением.

## Методы временных рядов (ARIMAX)

Собираемые данные об электропотреблении можно представить как последовательные измерения, совершаемые через равные промежутки времени (каждый месяц), что представляет собой модель временного ряда.

Методы анализа временных рядов основаны на предположении, что данные имеют некоторую внутреннюю структуру (тренд, сезонные компоненты, автокорреляции), каждое последующее значение временного ряда линейно зависит от предыдущего (авторегрессия).

Наиболее распространенными методами прогнозирования временных рядов являются модели *ARMA* (*AR* – авторегрессия, *MA* – скользящее среднее) [8].

Модель *ARMA* используются для стационарных процессов (не меняющих свои характеристики со временем), *ARIMA* является её расширением для нестационарных процессов. *ARMA* и *ARIMA* позволяют использовать в качестве входных параметров только время и электропотребление. Модель *ARIMAX*, описания нестационарных процессов с внешними параметрами, является наиболее подходящим инструментом для прогнозирования электропотребления среди моделей классических временных рядов. Использование дополнительных параметров позволяет учитывать влияние различных факторов. Модель прогнозирования *ARIMAX* имеет следующий вид:

$$y_t = c + \sum_{i=1}^p \alpha_i y_{t-i} + \sum_{i=1}^q \beta_i \varepsilon_{t-i} + \sum_{i=1}^r \theta_i z_{t-i} + \varepsilon_t$$

$p$ ,  $q$  – коэффициенты задающие порядок модели авторегрессии и скользящего среднего соответственно;  $y_{t-i}$  – предыдущие значения прогнозируемой переменной (электропотребления);  $\alpha_i$  – коэффициент авторегрессии;  $\varepsilon_{t-i}, \varepsilon_t$  – белый шум, последовательность независимых и одинаково распределенных случайных величин, показывает случайные колебания процесса ( $\varepsilon_t$  – остатки);  $\beta_i$  – коэффициент скользящего среднего;  $z_{t-i}$  – дополнительные параметры, регрессоры;  $r$  – их количество;  $\theta_i$  – степень их влияния, их веса;  $c$  – свободный коэффициент модели;

Модели авторегрессии скользящего среднего избавляют от необходимости нахождения коэффициентов регрессионных зависимостей. В качестве объясняющих переменных выступают прошлые значения самой зависимой переменной, а в качестве регрессионного остатка скользящие средние из элементов белого шума. Таким образом, к достоинствам методов прогнозирования этого класса относятся единообразие проектирования и анализа модели. Также сами природа данные потребления имеют структуру временного ряда.

Приложения модели *ARIMAX* для прогнозирования электропотребления рассматриваются в работах [9], [10]. Наиболее трудоемкой частью построения модели является определение начальных параметров. В [11] для их идентификации предлагается использование эволюционного алгоритма.

К минусам модели *ARIMAX* относятся перечисленные выше недостатки моделей линейной множественной регрессии: невозможность моделирования нелинейных зависимостей, и неизменность коэффициентов модели в процессе прогнозирования, что делает неэффективным её применение в задачах долгосрочного прогнозирования.

## Искусственные нейронные сети

Искусственные нейронные сети (ИНС) [12] – математическая модель биологической нейронной сети, основанная на применении методов искусственного интеллекта. ИНС – это по существу нелинейные функции, которые имеют способность аппроксимации нелинейных зависимостей. Выходы ИНС это линейные или нелинейные математические функции от её входов.

Задача прогнозирования сводится к построению ИНС, принимающая в качестве входных сигналов фактические значения временного ряда и значения внешних факторов, и описывающая их зависимости с будущим значением временного ряда.

Для построения ИНС необходимо решить следующие задачи: определить структуру ИНС, определить количество слоев нейронной сети, определить значения весов связей и значения порогов активации нейронов, определить вид функции активации нейронов, необходимо выбрать алгоритм обучения и выполнить процесс обучения.

Главным достоинством ИНС является то, что они позволяют моделировать сложные нелинейные зависимости. Также, при появлении новых данных не требуется перестраивать модель, сеть можно обучить дополнительно, что обеспечивает более высокую адаптивность.

Недостатком ИНС является сложность выбора параметров сети, сложность процесса обучения (возможны эффекты недообучения и переобучения), его ресурсоемкость. Поэтому обычно для подбора параметров, оптимизации процесса обучения используются различные дополнительные алгоритмы, такие как метод роя частиц [14], генетические алгоритмы [13]. Также, для повышения производительности, ИНС используют в комбинации с другими моделями, такими как вейвлет-преобразования [15] (вейвлет-нейронные сети), нечеткие множества [16] (нейро-нечеткие сети).

В целом, методы прогнозирования электропотребления, основанные на ИНС, отражают экспертное мнение. Результат прогнозирования сильно зависит от опыта эксперта, создающего ИНС, и проводящего обучение.

## Метод опорных векторов

Метод опорных векторов (*Support vector machine, SVM*) [17] – метод интеллектуального анализа данных используется как для классификации данных, так и для построения регрессионных нелинейных моделей.

В основу метода положена классификация, производимая за счет перевода исходных временных рядов в пространство более высокой размерности и поиска разделяющей гиперплоскости с максимальным зазором в этом пространстве. Далее, две параллельных гиперплоскости строятся по обеим сторонам гиперплоскости, разделяющей данные на классы. Алгоритм *SVM* работает в предположении, что чем больше расстояние между этими параллельными гиперплоскостями, тем меньше будет средняя ошибка классификатора. Задача прогнозирования решается таким образом, что на этапе обучения классификатора выявляются независимые переменные (внешние факторы), будущие значения которых определяют, в какой из определенных ранее подклассов попадет прогнозное значение.

В сравнении с методами ИНС, проблема выбора архитектуры ИНС здесь заменяется проблемой выбора наиболее подходящей функции ядра (*kernel trick*) [17], выполняющей нелинейное отображение исходных данных в пространство более высокой размерности. В качестве функции ядра часто используются следующие: линейные, полиномиальные, радиально-базисные функции (*RBF*), сигмоидные функции и Фурье-функции [17].

Использование метода опорных векторов для задачи краткосрочного прогнозирования электроэнергии представлено в работах [18-20].

Преимуществом метода опорных векторов является то, что параметры регрессионной модели определяются автоматически в процессе обучения классификатора.

К недостаткам относится отсутствие методологических рекомендаций по выбору функции ядра. Для метода опорных векторов характерны такие же сложности обучения, как и описанные выше для ИНС. Также т.к. задача прогнозирования связана с задачей построения классификатора, для получения точного прогноза метод нуждается в качественной подготовительной работе по выборке и оптимизации данных для обучения.

Метод опорных векторов получил наиболее широкое распространение для задач прогнозирования электропотребления, и часто используется для поиска аномалий в потоке транзакций. Другие методы классификации, применяются к решению этой задачи намного реже. К таким методам, например, относятся: метод главных компонент, метод *k*-средних, самоорганизующиеся карты Кохонена. Примеры их применения рассматриваются в работах [21] и [22] соответственно.

## Заключение

В обзоре рассмотрены наиболее часто используемые в настоящее время методы прогнозирования электропотребления с примерами научных работ, описывающих их применения.

Основная тенденция дальнейшего развития методов прогнозирования заключается в сочетании нескольких подходов и использовании дополнительных алгоритмов с целью повышения производительности и оптимизации методов. Такого рода методы получили название «гибридные алгоритмы», в работах [23], [24] описываются подходы, к прогнозированию электропотребления сочетающие в себе статистические методы, ИНС, методы экспертных оценок. Их развитие связано с попытками решить такие общие проблемы как: сложность идентификации модели, недостаточность методологической базы для выбора начальных параметров, сложность выборки экспериментальных данных, на основе которых следует рассчитывать регрессионные параметры (в статистических методах) или проводить процесс обучения (в методах интеллектуального анализа данных) и др.

Также, характеристикой методов является сложность моделирования электропотребления с учетом дополнительной информации и широкого набора внешних факторов. На практике, обычно ограничиваются выбором трех-четырёх влияющих факторов. Это связано с тем, что при рассмотрении большего их количества получается избыточно сложная модель, также становится сложно оценить влияние каждого фактора по отдельности.

Таким образом, проведенный анализ показывает, что в настоящее время не существует подхода к прогнозированию электропотребления, позволяющего смоделировать различные сценарии электропотребления, с учетом дополнительной информации, который позволял бы эффективно выявлять факты мошенничества и сами источники нетехнических потерь.

Конечный результат напрямую зависит от использования в исследовании всей доступной и косвенно связанной информации. Для эффективного обнаружения нетехнических потерь не следует ограничиваться применением одного метода, необходимо задействовать несколько подходов, учитывающих различные влияющие факторы, в таком случае проще принять достоверное и обоснованное решение.

## Список литературы

1. Красник В.В. 102 способа хищения электроэнергии. – М.: ЭНАС, 2011.
2. Фардиев И.Ш., Сафиуллин Д.Х. Об инновационном проекте «Умная сеть» // Энергетика Татарстана. 2010. – №3.
3. Z. Aung. Database systems for the smart grid // Smart Grid: Opportunities, Developments and Trends. – 2013.
4. Коморник С., Калечиц Е. Требования к системам прогнозирования энергопотребления // ЭнергоРынок. – 2008. – №3.
5. Hyde O., Hodnett P.F. An adaptable automated procedure for short-term electricity load forecasting // IEEE Transactions on Power Systems. – 1997. – Vol. 12. – Pp. 84-93.
6. Chen M.S., Van P. Nonparametric regression based short-term load forecasting // IEEE Transactions on Power Systems. – 1998. – Vol. 13. – Pp. 725-730.
7. Никифоров Г. В. Анализ устойчивости регрессионных моделей электропотребления // Промышленная энергетика. – 1999. – № 12. – С. 18-20.
8. Шувалова Е.Б. Теория статистики: Учебник. – М.: Финансы и статистика, 2004.
9. Desouky A.A. Hybrid adaptive techniques for electric-load forecast using ANN and ARIMA // IEE Proceedings of Generation, Transmission and Distribution. – 2000. – Vol. 4. – Pp. 213-217.
10. Cho M.Y., Hwang J. C., Chen C. S. Customer short-term load forecasting by using ARIMA transfer function model // Proceedings of the International Conference on Energy Management and Power Delivery, EMPD. – 1995. – Vol. 1. – Pp. 317-322.

11. Yang H.T., Huang C.M., Huang C.L. Identification of ARMAX model for short-term load forecasting: An evolutionary programming // *IEEE Transactions on Power systems*. – 1996. – Vol. 11. – Pp. 403-408.
12. Курбацкий В.Г. Прогнозирование электрической нагрузки с использованием искусственных нейронных сетей // *Электрика*. – 2006. – № 7.
13. Worawit T., Wanchai C. Substation short term load forecasting using neural network with genetic algorithm // *IEEE Conference on Computers, Communications*. – 2002. – Pp. 1787-1790.
14. Bashir Z.A., El-Hawary M.E. Short-Term Load Forecasting Using Artificial Neural Network Based on Particle Swarm Optimization Algorithm // *Canadian Conference on Electrical and Computer Engineering*. – 2007. – Pp. 272-275.
15. Huang C.M., Yang H.T. Evolving wavelet-based networks for short-term load forecasting, Generation, Transmission and Distribution // *IEE Proceedings*. – 2001. – Vol. 148. – Pp. 222-228.
16. Zhao Y., Zhang Y., Tang Y. Short-term load forecast based on neural network and fuzzy theory // *High Voltage Engineering*. – 2006. – Vol. 32. – Pp. 107-110.
17. Vapnik V., Cortes C. Support-vector networks // *Mach. Learn.* 1995. – Vol. 20. – Pp. 273-297.
18. Cherkassky V., Ma Y. Practical selection of SVM parameters and noise estimation for SVM regression // *Neural Netw.* – 2004. – Vol. 17. – Pp. 113-126.
19. Chen B.J. Load Forecasting Using Support Vector Machines // *A Study on EUNITE Competition 2001, National Taiwan University, Department of Computer Science and Information Engineering, National Taiwan University*. – 2001. – Pp. 20-26.
20. Надтока И.И. Краткосрочное прогнозирование электропотребления региона с учетом метеофакторов на основе метода опорных векторов и алгоритма роя частиц // *Изв. вузов. Электромеханика*. – 2014. – № 3.
21. Мозгалин А. В. Информационно-методическое обеспечение прогнозирования часовых объемов электропотребления при выходе предприятия на оптовый рынок электрической энергии // *Диссертация на соискание ученой степени кандидата технических наук*. – Москва, 2007.
22. Воронов И.В. Прогнозирование электропотребления промышленных предприятий с помощью искусственных нейронных сетей // *Диссертация на соискание ученой степени кандидата технических наук*. – Кемерово, 2010.
23. Евсеева А.В. Прогнозирование электропотребления при помощи гибридного алгоритма с применением технология CUDA // *Вестник ИГЭУ*. – 2011. – №3.
24. Лесниченко А. Ю. Метод среднесрочного прогнозирования электропотребления предприятий и организаций региона // *Электрика*. – 2010. – №8.