

УДК 621.311.16.016.3:517.9

## КРАТКОСРОЧНОЕ ПРОГНОЗИРОВАНИЕ НАГРУЗКИ С ПОМОЩЬЮ ТЕОРИИ НАИМЕНЬШИХ КВАДРАТОВ ОПОРНЫХ ВЕКТОРОВ (LS-SVM)

Надтока И.И., Аль-Зихери Баласим М.

*ФГБОУ ВПО «Южно-Российский государственный политехнический университет (Новочеркасский политехнический институт) имени М.И. Платова», 346428, Ростовская обл., г. Новочеркасск, ул. Просвещения, 132, eps@eps.rstu.novoch.ru*

Краткосрочное прогнозирование суточных почасовых графиков электрической нагрузки является важной основой надежной и экономичной работы электроэнергетической системы. Точность прогноза электропотребления непосредственно влияет на качество диспетчерского управления и надежность электроснабжения. Таким образом, выбор подходящего метода прогнозирования нагрузки для повышения точности прогноза имеет важное практическое значение. В статье представлены результаты краткосрочного прогнозирования электропотребления в операционной зоне регионального диспетчерского управления с помощью регрессионной модели, построенной на основе теории опорных векторов (SVM). Используется модификация наименьших квадратов опорных векторов (LS-SVM). В прогнозной модели учитываются статистические и прогнозные данные температуры воздуха и естественной освещенности, влияющие на электропотребление. Показано, что существенное влияние на точность прогнозирования оказывают два параметра модели LS-SVM, выбираемые опытным путём.

Ключевые слова: прогнозирование электропотребления, метод опорных векторов, температура, освещенность.

## SHORT TERM LOAD FORECASTING BY USING LEAST SQUARES SUPPORT VECTOR MACHINE THEORY

Nadtoka I.I., Al-Zihery Balasim M.

*South-Russia State Technical University (NPI), 346428, Rostov region, c. Novocherkassk, st. Prosvesheniya, 132, eps@eps.rstu.novoch.ru*

Short-term forecasting of daily schedules hourly electrical load is an important basis for reliable and efficient operation of the electricity system. Accuracy of the forecast electricity consumption directly affects the quality of supervisory control and power supply reliability. Thus, the selection of an appropriate load forecasting method to improve prediction accuracy is of practical importance. The paper presents results short-term forecasting electricity consumption in the operational area of the regional supervisory control using a regression model based on the theory of support vector machine (SVM). Use a modification of the least squares support vector machine (LS-SVM). In the predictive model accounted for statistical and forecast data of air temperature and natural light, affecting the power consumption. It is shown that a significant effect on the accuracy of prediction is supported by two parameter model LS-SVM, chosen empirically.

Keywords: prediction of power consumption, the method of support vector machines, temperature, lighting.

Краткосрочное прогнозирование суточных почасовых графиков электрической нагрузки является важной основой надежной и экономичной работы электроэнергетической системы. Точность прогноза электропотребления непосредственно влияет на качество диспетчерского управления и надежность электроснабжения. Таким образом, выбор подходящего метода прогнозирования нагрузки для повышения точности прогноза имеет важное практическое значение.

Учет фактора температуры воздуха при прогнозировании электропотребления реализован в различных моделях [1; 2]. Влияние температуры воздуха на электропотребление имеет более определенный характер, чем облачность и освещенность. Зависимости ЭП от освещенности имеют существенно более неопределенный характер. На

рисунке 1 представлены примеры полученных зависимостей суточного ЭП от суточной интегральной освещенности для операционной зоны Ростовского РДУ в течение 2012 года. Коэффициенты детерминации как для линейной, так и нелинейной моделей свидетельствуют о слабых корреляционных связях. Разделение годовой выборки на сезоны года (рис. 2) приводит к повышению коэффициентов детерминации линейной и нелинейной моделей, причем, как это видно из рисунка, существенно увеличение коэффициента для нелинейной модели.

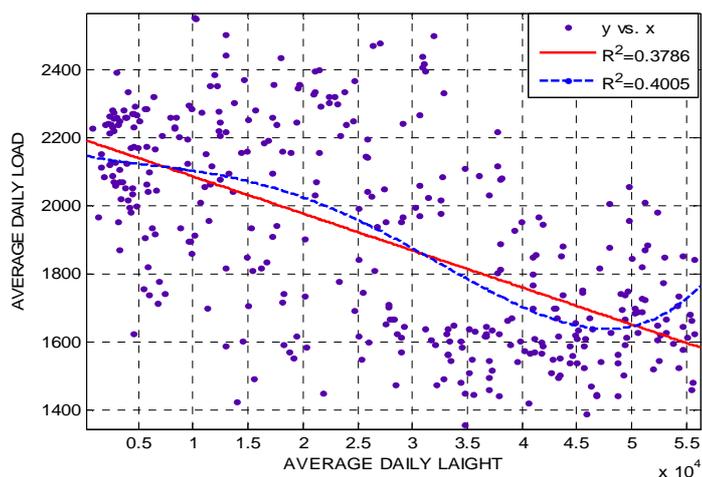


Рис. 1 – Зависимость суточного электропотребления от интегральной суточной освещенности на территории операционной зоны Ростовского РДУ в течение 2011 года

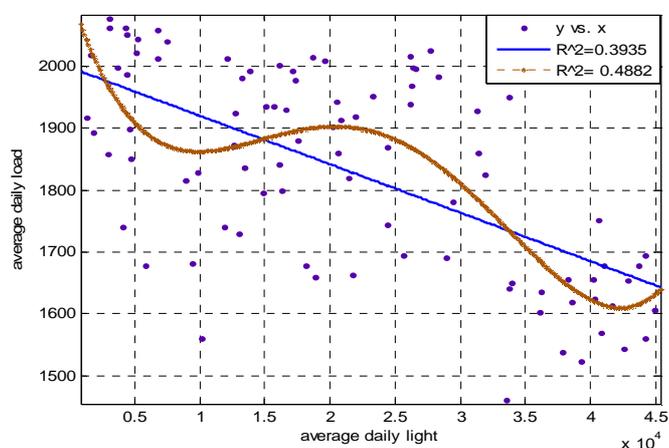


Рис. 2 – Зависимость суточного электропотребления от средней суточной освещенности на территории операционной зоны Ростовского РДУ в течение осени

Характер распределения точек суточного электропотребления в зависимости от освещенности показывает, что электропотребление имеет значительный разброс точек относительно кривых линейной и нелинейной регрессии.

Вид зависимостей, представленных на рисунках 1, 2, показывает, что связь между электропотреблением и освещенностью имеет нелинейный характер, что указывает на необходимость использования математических моделей для прогнозирования методов

электропотребления, которые позволяют учитывать нелинейный характер взаимосвязей электропотребления с влияющими на него факторами.

В статье приведены результаты моделирования и прогнозирования электропотребления на основе метода опорных векторов (SVM) [6; 7]. Метод опорных векторов представляет собой набор методов классификации и регрессии, который был разработан В.Н. Вапником в 1995 г. [7] и используется как для классификации, так и для построения регрессионных нелинейных моделей [4].

Модификация метода опорных векторов с использованием метода наименьших квадратов была впервые предложена J. Suykens [6] и является расширением стандартного метода SVM. В работе [5] метод LS SVM применен в задаче прогнозирования электрической нагрузки без учета метеофакторов (с учетом температуры воздуха). В [3] для прогнозирования суточных графиков электрической нагрузки метод опорных векторов применен в сочетании с генетическим алгоритмом и алгоритмом роя частиц. В данной работе указанный метод применяется для прогнозирования суточных часовых графиков электропотребления с учетом температуры воздуха и естественной освещенности.

Принцип минимизации цели в методе LS-SVM можно выразить как

$$\min J = \frac{1}{2} \|\omega\|^2 + \frac{1}{2} C \sum_{i=1}^l e_i^2 - \sum_{i=1}^l \alpha_i (\omega^T \varphi(x_i) + b + e_i - y_i) \quad (1)$$

Нелинейное выражение для модели прогнозирования будет иметь следующий вид:

$$y = \sum_{i=1}^l \alpha_i K(x_i, x) + b \quad i = 1, \dots, l, \quad (2)$$

где  $\alpha, b$  - линейные коэффициенты, которые можно получить, решив уравнение (4);  $K(x_i, x_j)$  - ядро-функция, которая учитывает нелинейное отображение из входного пространства к высокой размерности пространства признаков;  $e_i$  - ошибка,  $e \in R^{l \times 1}$  - ошибки вектора;  $\omega$  - вектор весов зазоров между кластерами (вектор коэффициентов разделения кластеров);  $C$  - параметр регуляризации, который контролирует ошибки;  $\alpha \in R^{l \times 1}$  - множители Лагранжа;  $y_i \in R$  - ожидаемое значение прогнозируемой суммы, общее число точек известных данных;  $\varphi(x)$  - нелинейное отображение входного пространства признаков в пространство более высокой размерности.

В качестве функции ядра часто используются следующие: линейные, полиномиальные, радиально-базисные, сигмовидные и Фурье-функции. Радиально-базисные функции имеют следующие преимущества: а) представления простых, даже для нескольких переменных входных; б) радиальная симметрия, гладкость, т.е. существуют производные

любого порядка; в) функция простого разбора, и поэтому она облегчит теоретический анализ.

В данной статье в качестве функции ядра в регрессионной модели используется радиально-базисная функция следующего вида:

$$K(x, x_i) = \exp(-\|x - x_i\|^2 / \sigma^2), \quad (3)$$

где:  $x$  -  $m$ -мерный входной вектор;  $\sigma$  - стандартизированный параметр, определяющий разброс функции вокруг центральной точки;  $x_i$  -  $m$ -мерный вектор координат центра рассеяния;  $\|x - x_i\|$  - норма вектора  $x - x_i$ , определяет расстояние между  $x$  и  $x_i$ . Радиальная базисная функция нелинейного преобразования обрабатывает нелинейные взаимосвязи между входной выборкой исходных данных в многомерном пространстве и выходной выборкой признаков в пространстве более высокой размерности. Коэффициент разброса (ширины) ядра  $\sigma$  отражает степень корреляции между обучающей выборкой входного пространства и вектором пространства признаков. При малых значениях  $\sigma$  ослабляются опорные векторы и усложняется процедура обучения (относительно сложные машины обучения), обобщающая способность не может быть гарантирована. При больших  $\sigma$  опорный вектор оказывает слишком сильное воздействие на регрессионную модель, и в результате трудно добиться достаточной точности прогнозирования.

Основные положения метода LS-SVM, применяемого для краткосрочного прогнозирования нагрузки [5; 6], изложены ниже.

Пусть обучающее множество задано как:

$$S = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \in R^d,$$

где  $d$  - число измерений выбранных входных переменных.

Выбираются положительные числа  $\sigma$  и  $C$ , и соответствующая функция ядра  $K(x, x')$ .

Далее решается задача оптимизации по методу множителей Лагранжа. Оптимальным решением будет вектор коэффициентов:

$$\alpha = (\alpha_1, \dots, \alpha_l)^T.$$

Построим функцию решения (формула (2)), где значения вектора коэффициентов  $b$  рассчитываются по формуле:

$$b = y_i - e - \sum_{i=1}^l (\alpha_i) K(x_j, x_i) \quad \alpha_j \in (0, C). \quad (4)$$

Выбор параметров  $\sigma$  и  $C$  в методе LS-SVM основывается на результатах опыта. В данной работе параметры функции ядра и параметры регуляризации выбираются как  $\sigma = 2 \cdot C = 30$ , соответственно. В работе [5] принимались значения  $\sigma = 7,27$  и  $C = 9079,14$ .

В данной работе матрица входных данных содержит статистические данные суточных часовых графиков нагрузки, средней освещенности за сутки, типа дня, максимальной, минимальная и средней температуры воздуха за сутки. Выход представляет значение суточного часового графика нагрузки в день прогнозирования.

Выборка исходных за год разбивается на четыре выборки по сезонам года: весна, лето, осень, зима. Данные за 2009-2011 годы рассматривались как обучающая выборка.

Вследствие различия величин исходных данных они нормализуются.

Нормализация типа дня (рабочий день и выходной), с понедельника по пятницу (рабочие сутки), принимается как 1, суббота и воскресенье принимаются как 0.

Нормализация температурных данных выполняется по выражению:

$$\theta'_{ij} = (\theta_{ij} - \theta_{j\min}) / (\theta_{j\max} - \theta_{j\min}), \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m, \quad (5)$$

где  $\theta_{ij}$  - фактическая температура  $^{\circ}\text{C}$ ;  $\theta_{j\min}$ ,  $\theta_{j\max}$ , минимальная, максимальная температура;  $\theta'_{ij}$  - нормализованная температура.

Нормализация данных освещенности выполняется аналогично:

$$E'_{ij} = (E_{ij} - E_{j\min}) / (E_{j\max} - E_{j\min}), \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m, \quad (6)$$

где  $E_{ij}$  - фактическая среднесуточная освещенность;  $E_{j\min}$ ,  $E_{j\max}$  - Минимальная и максимальная среднесуточная освещенности;  $E'_{ij}$  нормированная среднесуточная освещенность.

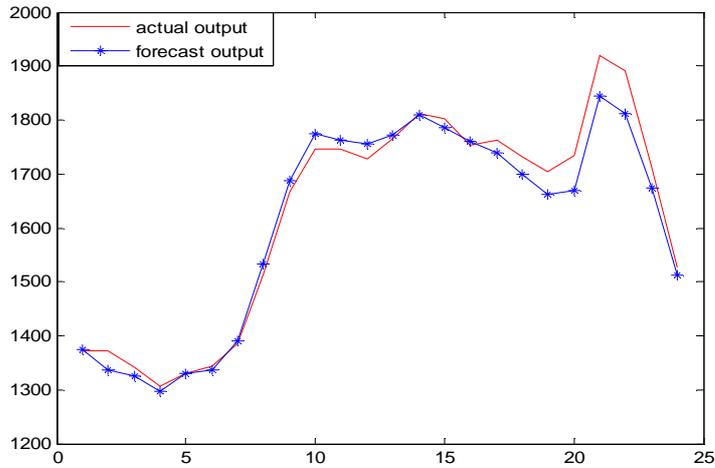
Данные нагрузки нормализуются логарифмированием

$$x'_{ij} = \log(x_{ij}), \quad (7)$$

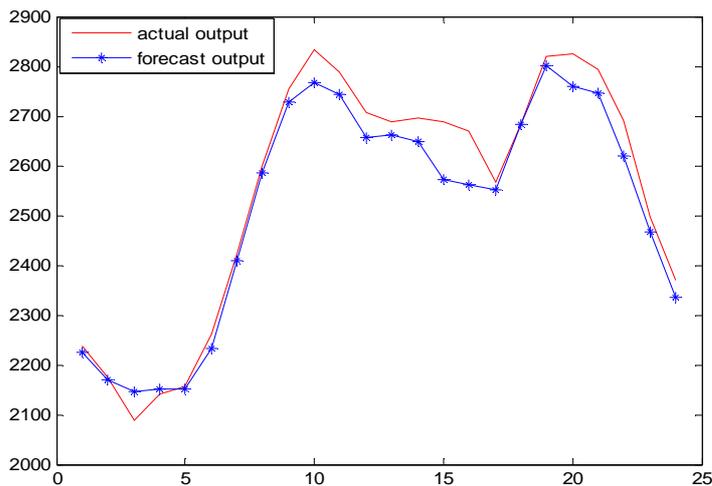
где  $x_{ij}$  - фактическая, а  $x'_{ij}$  - нормализованная нагрузки в  $j$ -й час  $i$ -х суток. При нормализации данных с помощью линейных преобразований типа (5, 6) погрешность прогнозирования увеличивается.

Средой моделирования является MALAB2011a. Исходные данные взяты из статистических показателей по Ростовской области за период с 2009 по 2012 г., выходные данные представляют данные нагрузки на день прогнозирования. Результаты прогнозирования и фактическая нагрузка на одни сутки в разные сезоны года показаны на рис. 3. В таблице 1 приведены данные по погрешностям прогнозирования в виде средней абсолютной ошибки (MAPE) в процентах.

$$MAPE = \frac{|P_{\text{прогн}} - P_{\text{факт}}|}{P_{\text{факт}}} \cdot 100\%$$



(a) лето 31/08/2011



(б) зима 26/01/2012

Рис. 3 – Фактические и прогнозные графики нагрузки в различные сезоны года

Таблица 1. Погрешности прогнозирования суточных графиков нагрузки по модели LS-SVM для разных сезонов года при  $C = 30$ ,  $\sigma = 2$ 

Время года	Лето	Зима	Осень	Весна
Погрешность (MAPE), %	1.4661	1.4541	1.1833	2.4341
Среднее значение погрешности (MAPE) 1.6344%				

Как и следовало ожидать, наибольшие погрешности получены для весны и осени, так как в указанные сезоны года имеет место наибольшая динамика длительности светового дня и погодных факторов.

Следует отметить также, что ошибка прогноза в начале сезона больше, чем ошибка в конце сезона. Это связано с тем, что величины освещенности в начале сезона близки по величине к освещенности в конце прошедшего сезона, и величины освещенности в конце сезона близки к освещенности в начале следующего сезона.

### Список литературы

1. Макоклюев Б.И. Анализ и планирование электропотребления. – М. : Энергоатомиздат, 2008. – 296 с.
2. Надтока И.И., Демура А.В., Губский С.О., Ваколюк А.Я., Горбачев В.В. Прогнозирование электропотребления с учетом температуры воздуха и естественной освещенности для региональных диспетчерских управлений. - Вестник СамГТУ. – 2012, № 1 (33). - С. 163-168.
3. Поляхов Н.Д., Приходько И.А., Ван Е. Прогнозирование электропотребления на основе метода опорных векторов с использованием эволюционных алгоритмов оптимизации // Современные проблемы науки и образования. – 2013. - № 2.
4. Christopner J.C. Burges A Tutorial on Support Vector Machines for Pattern Recognition // Journal of Data Mining and Knowledge Discovery, vol. 2, Springer, Netherlands, 1998. p. 2.
5. M. Božić , M. Stojanović , Z. Staji ć. Short-Term Electric Load Forecasting Using Least Square Support Vector Machines // Automatic Control and Robotics. Vol. 9, No. 1, 2010, pp. 141 – 150.
6. Suykens J. Least Squares Support Vector Machines , 2002 <http://www.esat.kuleuven.ac.be/sista/members/suykens.html>.
7. Vapnik, Vladimir N. and Cortes, Corinna, "Support-Vector Networks", Machine Learning, 20, 1995. <http://www.springerlink.com/content/k238jx04hm87j80g/>.

### Рецензенты:

Степанов В.П., д.т.н., профессор, заведующий кафедрой «Автоматизированные электроэнергетические системы» ФГБОУ ВПО «Самарский государственный технический университет», г. Самара.

Сычёв В.А., д. э. н., профессор, профессор кафедры «Производственный и инновационный менеджмент» ФГБОУ ВПО «Южно-Российский государственный политехнический университет» (НПИ), г. Новочеркасск.