АЛГОРИТМ ОЦЕНКИ НАГРУЗКИ ИНТЕРНЕТ-СОЕДИНЕНИЯ НА ОСНОВЕ ЗАДАЧИ АППРОКСИМАЦИИ СЕГМЕНТНОЙ ФУНКЦИИ АЛГЕБРАИЧЕСКИМ ПОЛИНОМОМ ФИКСИРОВАННОЙ СТЕПЕНИ

И. Ю. Выгодчикова

Саратовский национальный исследовательский государственный университет им. Н. Г. Чернышевского, Россия
E-mail: pavlovni@info.sgu.ru, vyshkg@mrsu.ru

В статье представлен инструментарий, модель и метод оценки нагрузки интернет-соединения с использованием задачи аппроксимации сегментной функции алгебраическим полиномом фиксированной степени. Разработан алгоритм оценки нагрузки интернет-соединения на основе задачи аппроксимации сегментной функции алгебраическим полиномом фиксированной степени. Построена модель изменения уровня нагрузки по дням недели. Выполнена аппроксимация и оценка макета изменения нагрузки по дням недели. Обнаружена существенная зависимость уровня нагрузки интернет-сети от дня недели, получена качественная, с точки зрения корреляции с обследуемыми данными, модель процесса.

Целью статьи является построение модели и алгоритма нагрузки интернет-соединения с использованием задачи аппроксимации сегментной функции алгебраическим полиномом фиксированной степени.

THE ALGORITHM OF ESTIMATING THE LOAD OF INTERNET CONNECTION BASED AT PROBLEM OF APPROXIMATION SEGMENT FUNCTION BY ALGEBRAIC POLYNOMIAL OF FIXED DEGREE

I. Y. Vygodchikova

The article presents tools, model and method of estimating the Internet connection load using the problem of approximation a segment function by algebraic polynomial of fixed degree. An algorithm of estimating the Internet connection load is developed. Algorithm is based at problem of approximation the segment function by algebraic polynomial of fixed degree. A model of changing load levels by day of the week was built. The approximation and evaluation of the distribution of load changes by days of the week was carried out. The significant dependence of Internet network load level on the day of the week was found, and qualitative model of the process was obtained in terms of correlation with initial data.

The purpose of the article is building the model and algorithm of Internet connection load using the problem of approximation the segment function by an algebraic polynomial of fixed degree.

Математический метод. Пусть в узлах дискретной сетки $T=\{t_0<t_1<\ldots<t_N\}$ фиксируются сегменты суммарного объёма нагрузки интернет сети: $[y_{1,k}\,;y_{2,k}]$, $y_{2,k}\geq y_{1,k}$, $k=\overline{0,N}$, $N\geq n+1$. На основе этих данных требуется найти коэффициенты $A=(a_0,a_1,\ldots,a_n)\in R^{n+1}$ алгебраического полинома $p_n(A,t)=a_0+a_1t+\ldots+a_nt^n$, минимизирующие функцию:

$$\rho(A) = \max_{k=\overline{0,N}} f(A,k) \longrightarrow \min_{A \in R^{n+1}}, \tag{1}$$

где $f(A,k) = \max\{y_{2,k} - p_n(A,t_k); p_n(A,t_k) - y_{1,k}\}$.

Основным результатом математического исследования стал ***численный метод*** решения задачи, который обоснован следующим доказанным фактом.

Теорема 1. Вектор $A^* \in R^{n+1}$ является решением задачи (1) тогда и только тогда, когда удовлетворяет хотя бы одному из следующих условий:

(а) $\rho(A^*) = m$, где $m = 0.5 \max_{k=\overline{0,N}} (y_{2,k} - y_{1,k})$;

(б) для некоторого множества $\sigma = \{t_{j_0} < \ldots < t_{j_{n+1}}\} \subset T$ (называемого *базисом*) и $i = 0$ или $i = 1$, справедливы соотношения:

$$h_i(\sigma) = \begin{cases} y_{2,j_k} - p_n(A_i(\sigma), t_{j_k}), \textit{если } (k+i) - \textit{чётно}, \\ -y_{1,j_k} + p_n(A_i(\sigma), t_{j_k}), \textit{если } (k+i) - \textit{нечётно}, \end{cases} k = \overline{0, n+1},$$

$A^* = A_i(\sigma)$, $\rho(A^*) = h_i(\sigma)$. При этом $\rho^* = \rho(A^*)$.

Автором статьи доказано, что для того, чтобы $A^* \in R^{n+1}$ являлся единственным решением задачи (1), необходимо и достаточно, чтобы выполнялось условие (б) или условие (а), при том, что исходный ряд содержит не менее чем $n+1$ сегментов ширины $2m$.

Заметим, что задача (1) является обобщением известной задачи П.Л.Чебышёва [1]: $\max_{k=\overline{0,N}} |y_k - p_n(A,t_k)| \longrightarrow \min_{A \in R^{n+1}}$. Простые примеры показывают, что задача (1) не сводится к задаче Чебышёва, даже если взять $y_k = (y_{2,k} + y_{1,k})/2$, $k = \overline{0,N}$.

Алгоритм решения задачи. Алгоритм решения зад**ачи** (1) основан на итерационном решении систем линейных уравнений, по аналогии с алгоритмом Валле Пуссена, предназначенного для решения задачи в случае однозначности ряда. Введём следующее понятие. *Амплитудными* на базисе σ назовём функции $\varphi_{0,k}(\sigma) = \varphi_0(\sigma,k)$ и $\varphi_{1,k}(\sigma) = \varphi_1(\sigma,k)$, определяемые формулами:

$$\varphi_{1-i,k}(\sigma) = \begin{cases} i\, y_{2,j_k} + (1-i) y_{1,j_k}, k - \textit{чётно}, \\ i\, y_{1,j_k} + (1-i) y_{2,j_k}, k - \textit{нечётно}, \end{cases} \quad k = \overline{0,n+1}, \quad i = 0,1.$$

Аргумент (σ) для наглядности изложения алгоритма опустим.

На *первом* этапе алгоритма выбирается произвольно начальный базис и соответствующие ему *(n+2)* сегментов ряда.

Указанному многозначному ряду на базисе ставится в соответствие два однозначных ряда $\{\varphi_{0,k}\}$ и $\{\varphi_{1,k}\}$ таким образом, что при чётном k ряду $\{\varphi_{0,k}\}$ принадлежит верхнее значение $y_{2,k}$ сегмента, а ряду $\{\varphi_{1,k}\}$ принадлежит нижнее значение $y_{1,k}$, а при нечётном k наоборот. Для каждого из полученных однозначных рядов строится алгебраический полином с использованием известного метода (Чебышевской интерполяции) [1]. Обозначим коэффициенты полученных полиномов Чебышёва, аппроксимирующих ряды $\{\varphi_{0,k}\}$ и $\{\varphi_{1,k}\}$, через A_0 и A_1, соответственно. Вычисляем модуль разности между значениями полинома Чебышёва $p_n(A_0,t_k)$ и значениями $\{\varphi_{0,k}\}$, а также модуль разности между значениями полинома Чебышёва $p_n(A_1,t_k)$ и значениями $\{\varphi_{1,k}\}$ (по условию Чебышёва, каждое из этих отклонений одинаково для любого $k=0,\ldots,n+1$). Вычисляем максимальное из этих модулей, обозначим его $h_\beta = \left|\varphi_{\beta,j_0} - p_n(A_\beta,t_{j_0})\right|$, где $\beta \in \{0,1\}$, и полином Чебышёва с коэффициентами A_β будем использовать далее. Автором статьи доказано, что второй полином можно отбросить [2].

На *втором* этапе алгоритмической процедуры выполняется анализ, будет ли где A_β решением задачи (1) или же нужно продолжать поиск. Вычислим величину: $\varepsilon = \max\limits_{k\,\in\,\overline{0,N}} \max\{y_{2,k} - p_n(A_\beta,t_k); p_n(A_\beta,t_k) - y_{1,k}\}$. Выбираем индекс k_0 так, чтобы выполнялось следующее равенство: $\varepsilon = \max\{y_{2,k_0} - p_n(A_\beta,t_{k_0}); p_n(A_\beta,t_{k_0}) - y_{1,k_0}\}$. Если, при этом, выполняется равенство: $\varepsilon = y_{2,k_0} - p_n(A_\beta,t_{k_0})$, то положим *s=2*, иначе положим *s=1*. Если имеем $t_{k_0} \notin \sigma$, то выполняем следующее преобразование базиса: включаем в него узел t_{k_0} и исключаем из базиса узел согласно рекомендациям алгоритма Валле-Пуссена, применяемым для ряда $\{\varphi_{\beta,k}\}$ и значения границы сегмента

После этого возвращаемся к первому этапу алгоритмической процедуры с новым базисом. Если $t_{k_0} \in \sigma$, при этом $h_\beta = \varepsilon$, то A_β является единственным, искомым решением задачи (1) и алгоритм завершается. Если $t_{k_0} \in \sigma$, но $h_\beta < \varepsilon$, то переходим к третьему этапу.

На *третьем* этапе алгоритмической процедуры уже известно, что аппроксимирующий полином проходит через середины самых широких сегментов. Если таковых не менее чем *(n+1)*, решение будет единственным (алгебраическая интерполяция с использованием *(n+1)* середин). В противном случае, решений много, осуществляется итерационная процедура отыскания крайних точек, которых конечное число и которые позволяют описать всё множество решений. Алгоритм запрограммирован.

Для того чтобы зона допустимой нагрузки не была превышена, вводятся ограничения [2]:

$$\rho(A) = \max_{k\,=\overline{0,N}} f(A,k) \to \min_{A\,\in\,D}, \quad D = \{A \in R^{n+1} : y_{1,k} \le p_n(A,t_k) \le y_{2,k}\}. \qquad (2)$$

Для сравнения с построенной по критерию минимакса моделью использован стандартный метод наименьших квадратов (МНК):

$$z(A) = \sum_{k=0}^{N} (y_k - p_n(A,t_k))^2 \to \min_{A\in D}, \quad D = \{A \in R^{n+1} : y_{1,k} \le p_n(A,t_k) \le y_{2,k}\}, \qquad (3)$$

где $y_k = (y_{1,k} + y_{2,k})/2$ – середины сегмента $[y_{1,k}\,;\, y_{2,k}]$, $k=\overline{0,N}$.

Следует отметить, что в методе наименьших квадратов при подсчёте среднеквадратического отклонения все отклонения усредняются. В то же время существуют процессы, в которых большую значимость имеют крупных отклонения, которые могут приводить к существенным сбоям сети или простоям и значительной недозагрузке оборудования. В рассматриваемой модельной задаче предполагается, что значимость ошибки прогноза растёт по экспоненциальному закону в зависимости от величины отклонения показателя от аппроксимирующей функции.

В вычислительных экспериментах применялись задачи (2) и (3) с учётом следующих оценок качества прогноза: максимальная ошибка аппроксимации до дальней границы сегмента и сумма экспонент абсолютных ошибок аппроксимации по всем наблюдениям.

Вычислительный эксперимент. Выполним прогнозирование нагрузки канала интернет-сети с учётом одновременных соединений за период с 2 января по 15 февраля[1]. Можно заметить, что присутствует стабильность в распределении нагрузки по дням недели, если из рассмотрения удалить праздничные дни (начало января). Достаточно высокий коэффициент автокорреляции порядка 7 свидетельствует о наличии цикличности. Построим сегменты уровня телефонной нагрузки, взяв за каждый день минимум и максимум нагрузки (табл. 1).

Таблица 1

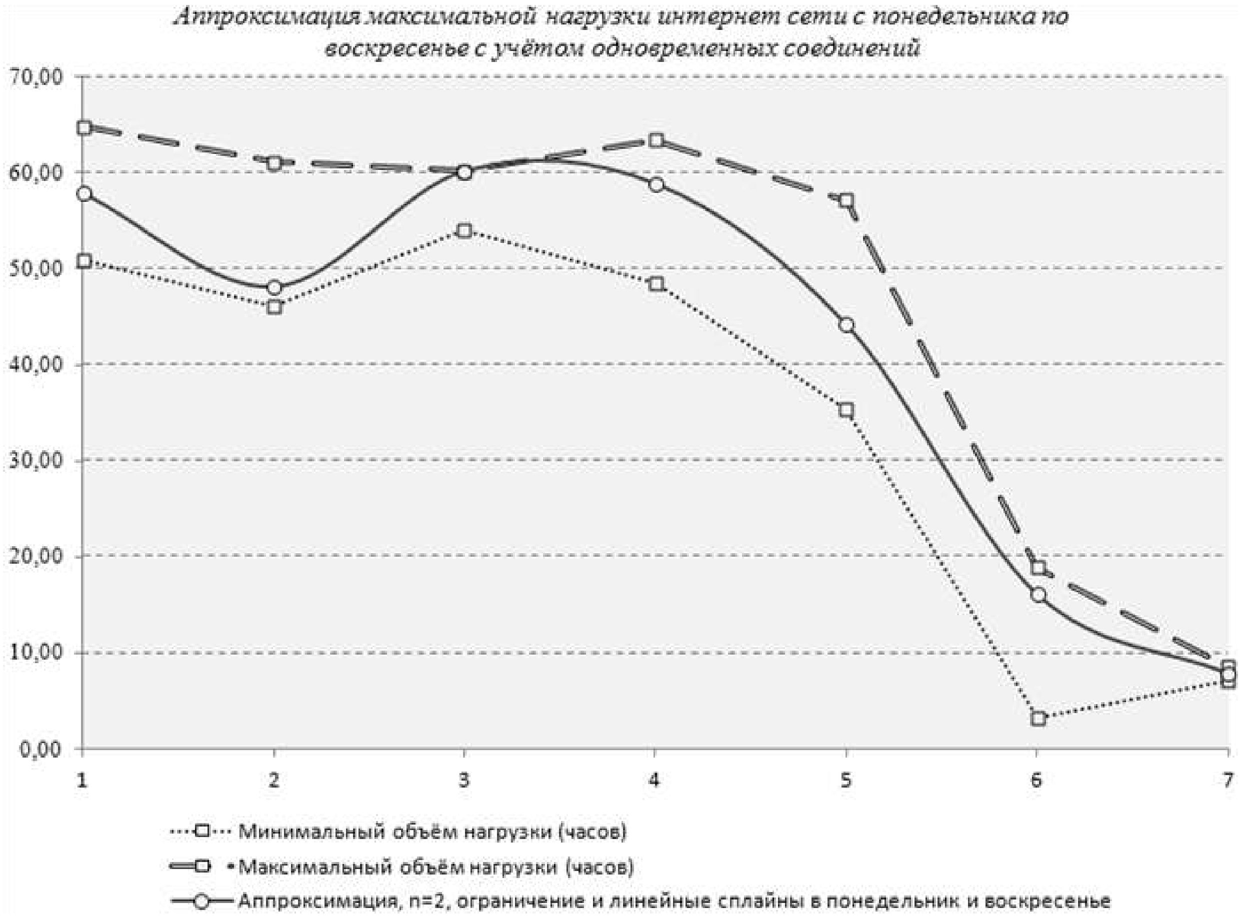
Анализ нагрузки интернет-соединения за каждую неделю (в часах)

1 неделя	2 неделя	3 неделя	4 неделя	5 неделя	6 неделя	Минимум	Максимум	День недели
58,99	50,9	64,81	58,9	58,13		48,5	63,4	пн
59,04	46,06		61,13	58,88		35,4	57,2	вт
55,08	60,2	53,97	58,92	59,58		3,2	18,9	ср
63,44	51,1	54,7	58,76	59,23	48,51	7,2	8,6	чт
35,38	57,17	50,17	53,61	55,53	54,83	50,9	64,8	пт
7,27	18,44	18,88	15,57	13,48	3,16	46,1	61,1	сб
8,56	7,91	7,52	7,29	7,18		54	60,2	вс

Нагрузка понедельника и воскресенья резко отличается от уровня нагрузки в другие дни недели. Поэтому для повышения точности прогноза целесообразно использовать полиномиальные сплайны.

Решены задачи (2) и (3) для сетки, содержащей узлы, соответствующие дням со вторника до субботы включительно, кривая достроена линейными сплайнами к средним значениям за понедельник и воскресенье. Ни рис. 1 и в табл. 2 представлено решение для задачи (1).

[1] Реальные данные по СГТУ (Саратов) за 2014 г.



Аппроксимация нагрузки

Таблица 2

Анализ нагрузки интернет-соединения за каждую неделю (в часах)

1 неделя	2 неделя	3 неделя	4 неделя	5 неделя	6 неделя	Аппроксимация	Максимальная ошибка для аппроксимации	День недели
58,99	50,9	64,81	58,9	58,13		57,85	6,96	пн
59,04	46,06		61,13	58,88		48,16	12,98	вт
55,08	60,2	53,97	58,92	59,58		60,2	6,23	ср
63,44	51,1	54,7	58,76	59,23	48,51	58,88	10,38	чт
35,38	57,17	50,17	53,61	55,53	54,83	44,19	12,98	пт
7,27	18,44	18,88	15,57	13,48	3,16	16,14	12,98	сб
8,56	7,91	7,52	7,29	7,18		7,87	0,69	вс

Сопоставление ошибок аппроксимации показало, что максимальная ошибка аппроксимации по минимаксу составила 12,98, а по МНК ошибка оказалась существенно выше и составила 15,98.

Заключение. Исследование показало, что в случае сильного влияния на поведение системы редких существенных колебаний внешней нагрузки применение предложенной модели аппроксимации, основанной на минимаксном критерии и использовании интервальных данных, позволит повысить эффективность управления ресурсами, требуемыми для надёжного и бесперебойного функционирования системы.

СПИСОК ЛИТЕРАТУРЫ

1. *Демьянов В. Ф. Малоземов В. Н.* Введение в минимакс / М. : Изд-во «Наука», 1972. 368 с.

2. *Выгодчикова И. Ю.* Минимаксный метод моделирования многозначных динамических рядов в экономике / С. : Саратовский социально-экономический институт (филиал) РЭУ им. Г. В. Плеханова, 2017. 116 с.