

18. Tarasov V.N., Bahareva N., Gorelov G.A. Mathematical model of heavy-tailed traffic distribution based on the H2/M/1 queuing system. *Infokommunikacionnye tehnologii*, 2014, vol. 12, no. 3, pp. 36–41. (In Russian.)
19. Downey A. Lognormal and Pareto distributions in the Internet. *Computer Communications*, 2005, vol. 28, no. 7, pp. 790–801.
20. Dang T.D., Sonkoly B., Molnar S. Fractal analysis and modeling of VoIP traffic. *11th International Telecommunications Network Strategy and Planning Symposium, NETWORKS 2004*, 2004, pp. 217–222.
21. Buranova M.A., Kartashevskij V.G. Analysis of the latency for a G/D/1 host with imprecise knowledge of traffic parameters. *Infokommunikacionnye tehnologii*, 2017, vol. 5, no. 1, pp. 24–33. (In Russian.)
22. Beshley M. et al. Increasing the efficiency of real-time content delivery by improving the technology of priority assignment and processing of IP traffic. *Smart Computing Review*, 2015, vol. 5, no. 2, pp. 1–13.
23. Golubintsev A.V., Mjasnikova A.I., Legkov K.E. Architectural principles of organizing automated control systems for infocommunication networks for special purposes. *Naukoemkie tehnologii v kosmicheskikh issledovaniyah Zemli*, 2015, vol. 7, no. 4, pp. 16–23. (In Russian.)
24. Ushanev K.V. Simulation models of the queuing system Pa/M/1, H2/M/1 and the study on their basis of the quality of service with a complex structure. *Sistemy upravlenija, svjazi i bezopasnosti*, 2015, no. 4, pp. 217–251. (In Russian.)
25. Tarasov V.N., Kartashevskij I.V. Approximation methods for input distributions for the G/G/1 system and analysis of the results. *Sistemy upravlenija i informatsionnye tehnologii*, 2015, no. 3.1, pp. 182–185. (In Russian.)
26. *Consent criteria*. URL: http://sernam.ru/book_tp.php?id=34 (accessed 26.06.2018). (In Russian.)
27. Crovella M., Lipsky L. Long-lasting transient, conditions in simulations with heavytailed workloads. *Proc 1997 Winter Simulation Conference*, 1997.

Received 06.02.2020

ТЕХНОЛОГИИ КОМПЬЮТЕРНЫХ СИСТЕМ И СЕТЕЙ ТРУДЫ УЧЕНЫХ КАФЕДРЫ ИНФОКОММУНИКАЦИЙ СЕВЕРО-КАВКАЗСКОГО ФЕДЕРАЛЬНОГО УНИВЕРСИТЕТА

УДК 004.724.4

ИСПОЛЬЗОВАНИЕ МЕТОДОВ РЕГРЕССИОННОГО АНАЛИЗА ДЛЯ ПОСТРОЕНИЯ ОПТИМАЛЬНОЙ МОДЕЛИ ЗАВИСИМОСТИ РАЗМЕРА ОЧЕРЕДИ ОТ ПОКАЗАТЕЛЯ ХЕРСТА ПРИ ПРЕОБРАЗОВАНИИ САМОПОДОБНОГО ВХОДНОГО ПОТОКА ПАКЕТОВ В ПОТОК, ИМЕЮЩИЙ ЭКСПОНЕНЦИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

*Линец Г.И., Воронкин Р.А., Говорова С.В., Мочалов В.П., Палканов И.С.
Северо-Кавказский федеральный университет, Ставрополь, РФ
E-mail: kbytw@mail.ru*

С использованием методов машинного обучения получена модель для предсказания размера очереди входного самоподобного потока пакетов, распределенного по закону Парето при его преобразовании в поток, имеющий экспоненциальное распределение. Поскольку величина потерь в общем случае не дает никаких сведений об эффективности использования буферного пространства в процессе преобразования самоподобного потока пакетов, для оценки качества исследуемых моделей введена метрика качества (штраф), являющаяся комплексным показателем. Данный критерий учитывает как потери пакетов в процессе функциональных преобразований, так и неэффективное использование буферного пространства узлов коммутации. Показано, что лучшими по исследуемому критерию являются модели, использующие методы изотонической регрессии и опорных векторов.

Ключевые слова: телекоммуникационная сеть, самоподобный трафик, показатель Херста, производительность, распределение Парето, потери пакетов, регрессионный анализ, метрики качества, функция штрафа, машинное обучение

Введение

Одной из проблем телекоммуникационных сетей является неэффективное использование узловых и канальных ресурсов при обслуживании самоподобного трафика. Важным параметром информационного обмена, влияющим на качество обслуживания пользователей, а также на производительность сети, служит вероятность потери сообщений за счет переполнения буферных запоминающих устройств (БЗУ) в узлах коммутации.

Основной причиной, ведущей к переполнению буфера, является наличие долговременной зависимости в сетевом трафике, обусловленной его самоподобием, вследствие чего суммарный кумулятивный эффект в широком диапазоне задержек может значительно отличаться от того, который наблюдается в кратковременно зависимом процессе [1]. Для устранения самоподобия сетевого трафика используются различные модели и устройства преобразования трафика [2], одной из них является асинхронная имитационная модель, описанная в [2], для которой существует программная реализация [3].

Важным показателем работы данной модели является размер очереди, используемый в процессе преобразования трафика. Поскольку из-за ограниченности ресурсов ЭВМ очередь не может иметь бесконечный размер, возникает задача предсказания размера очереди в зависимости от меры самоподобия входного трафика, в качестве которой выступает показатель Херста.

Постановка задачи

С использованием методов машинного обучения необходимо разработать модель для предсказания размера очереди в зависимости от показателя Херста на основании данных, полученных при выполнении преобразования входного самоподобного потока, распределенного по закону Парето в поток, имеющий экспоненциальное распределение.

Так как машинное обучение включает в себя множество методов, на начальном этапе для дальнейшего сравнения с более сложными моделями, построенными, в частности, с использованием методов глубокого обучения, целесообразно рассмотреть только методы парного регрессионного анализа, изотонической регрессии и метод опорных векторов.

Введем метрику качества (штраф), которая является комплексным показателем и учитывает как потери пакетов в процессе преобразования трафика, так и неэффективное использование буферного пространства. Проведем выбор наилучшей модели для осуществления предсказания размера очереди в зависимости от показателя Херста входного потока с помощью следующих метрик качества:

- коэффициент детерминации;
- среднеквадратичная ошибка регрессии;
- средняя абсолютная ошибка;
- величина штрафа.

Решение задачи

Имитационная модель, представленная в [2], обеспечивает преобразование входного потока пакетов, который является заведомо самоподобным, в заданный закон распределения, в частности в экспоненциальный. Объектом преобразования является одномерная плотность распределения интервалов времени между пакетами входного потока. С использованием разработанной модели проведены $1,1 \cdot 10^3$ испытаний и получены данные для статистического анализа.

Поскольку величина потерь в общем случае не дает никаких сведений об эффективности использования очереди в процессе преобразования трафика, для оценки качества полученной модели введем метрику качества – штраф, которая учитывает не только величину потерь, но и нерациональное использование буферной памяти.

Пусть y_i – истинное значение размера очереди в выборке; $\hat{y}_i$ – предсказанное значение размера очереди в выборке, соответствующее истинному значению y_i . Если $y_i > \hat{y}_i$, будем штрафовать обучаемую систему на величину $\beta_+(y_i - \hat{y}_i)$. Если $y_i \leq \hat{y}_i$, величина штрафа будет зависеть от величины разности $\Delta_i = y_i - \hat{y}_i$, при $\Delta_i \geq \Delta$ величина штрафа составит $\beta_-(\Delta_i - \Delta)$, и 0 – в противном случае. Проиллюстрируем сказанное примером (рисунок 1). Рассмотрим три случая, каждому из которых соответствуют истинные значения размеров очереди y_1 , y_2 и y_3 . Допустим, предсказанные значения размеров очереди в каждом из трех случаев совпадают, иными словами, $\hat{y}_1 = \hat{y}_2 = \hat{y}_3 = \hat{y}$. Тогда, в первом случае $y_1 > \hat{y}$ и величина штрафа определяется как $\beta_+(y_1 - \hat{y})$. Во втором случае $\hat{y} - \Delta \leq y_2 \leq \Delta$ и величина штрафа равна 0, предполагается, что

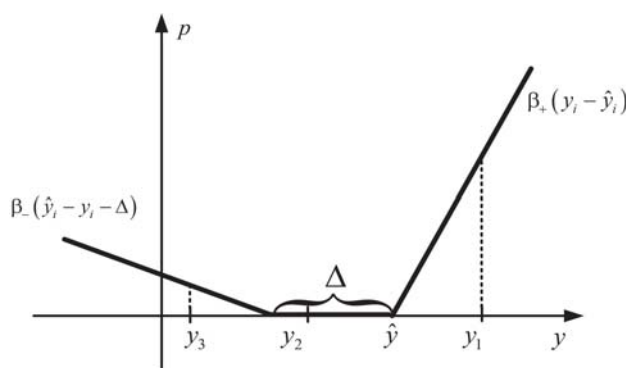


Рисунок 1. График зависимости величины штрафа от объема буфера

$\hat{y}$ – предпочтительный размер очереди для y_2 . В третьем случае $y_3 < \hat{y} - \Delta$, величина штрафа определяется из выражения $\beta_-(\hat{y} - y_3 - \Delta)$.

Таким образом, величина штрафа будет вычисляться из равенства:

$$p_i = \begin{cases} \beta_+(y_i - \hat{y}), & \text{если } y_i > \hat{y}, \\ \beta_+(\hat{y} - y_i - \Delta), & \text{если } \hat{y} - y_i > \Delta, \\ 0, & \text{в противном случае.} \end{cases}$$

Общий штраф для всех испытаний определяется как среднее арифметическое между штрафами по каждому испытанию $p = \frac{1}{n} \sum_{i=1}^n p_i$, где n – число испытаний. В процессе обучения модели необходимо обеспечить минимальное значение штрафа для всех испытаний, иными словами, $p \rightarrow \min$.

Представленная система штрафов предусматривает введение трех гиперпараметров: β_+ , β_- и Δ , где $\beta_+ > \beta_- > 0$ и $\Delta > 0$. Установим их значение следующим образом: $\beta_- = 0,3$, $\beta_+ = 1$, $\Delta = 50$.

Первоначальный анализ данных

На рисунке 2 показана диаграмма рассеивания зависимости размера очереди от показателя Херста. Из рисунка отчетливо видно, что существует определенная корреляция между показателем Херста и объемом буферной памяти. Уровень значимости в статистике является важным показателем, отражающим степень уверенности в точности, истинности полученных (прогнозируемых) данных. Большой объем проведенных испытаний позволяет не вычислять p – уровень значимости, поскольку, как показывает практика, в данном случае p – значение гораздо меньше 0,05 [4].

Предварительно сгруппируем испытания по значению показателя Херста. Выделим 30 групп для оценки разброса размера очереди. Постро-

им box-plot для каждой группы. Из рисунка 3 следует, что наибольшее количество выбросов сверху наблюдается для первых 10 групп, что соответствует показателю Херста, близкому к 0,5. Следовательно, при этих значениях показателя Херста возможно появление потерь из-за того, что требуемый объем буфера будет больше предсказанного. Для групп с 28 по 30 наблюдаются значительные выбросы снизу, что приводит к неэффективному использованию буферной памяти.

Регрессионный анализ

Машинное обучение – это подраздел искусственного интеллекта, в котором изучаются и исследуются алгоритмы, способные обучаться без прямого программирования. Линейная регрессия является типичным представителем алгоритмов машинного обучения [5]. Выделяют следующие задачи, решаемые машинным обучением: обучение с учителем, обучение без учителя, обучение с подкреплением, активное обучение, трансфер знаний и т. д.

Регрессия (как и классификация) относится к классу задач обучения с учителем, когда по заданному набору признаков наблюдаемого объекта необходимо спрогнозировать некоторую целевую переменную. Как правило, в задачах обучения с учителем опыт E представляется в виде множества пар признаков и целевых переменных: $D = \{(x_i, y_i)\}_i = 1 \dots n$. В случае линейной регрессии признаковое описание объекта – это действительный вектор $x \in R^m$, где R – множество действительных чисел, а целевая переменная – это скаляр $y \in R$. Самой простой мерой качества L для задачи регрессии является

$$L(y, \hat{y}) = (y - \hat{y})^2,$$

где $\hat{y}$ – это оценка реального значения целевой переменной [5; 6].

Восстановим зависимость, представленную на рисунке 2, используя регрессионный анализ, основу которого составляет метод наименьших квадратов (МНК), где в качестве уравнения регрессии берется функция $y = f(x)$ – такая, чтобы сумма квадратов разностей удовлетворяла условию $S = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min$. С использованием методов парного регрессионного анализа проведем статистический анализ данных, полученных при выполнении преобразования входного самоподобного потока, распределенного по закону Парето в поток, имеющий экспоненциальное распределение. Рассмотрим методы, наиболее широко используемые на практике, позволяющие

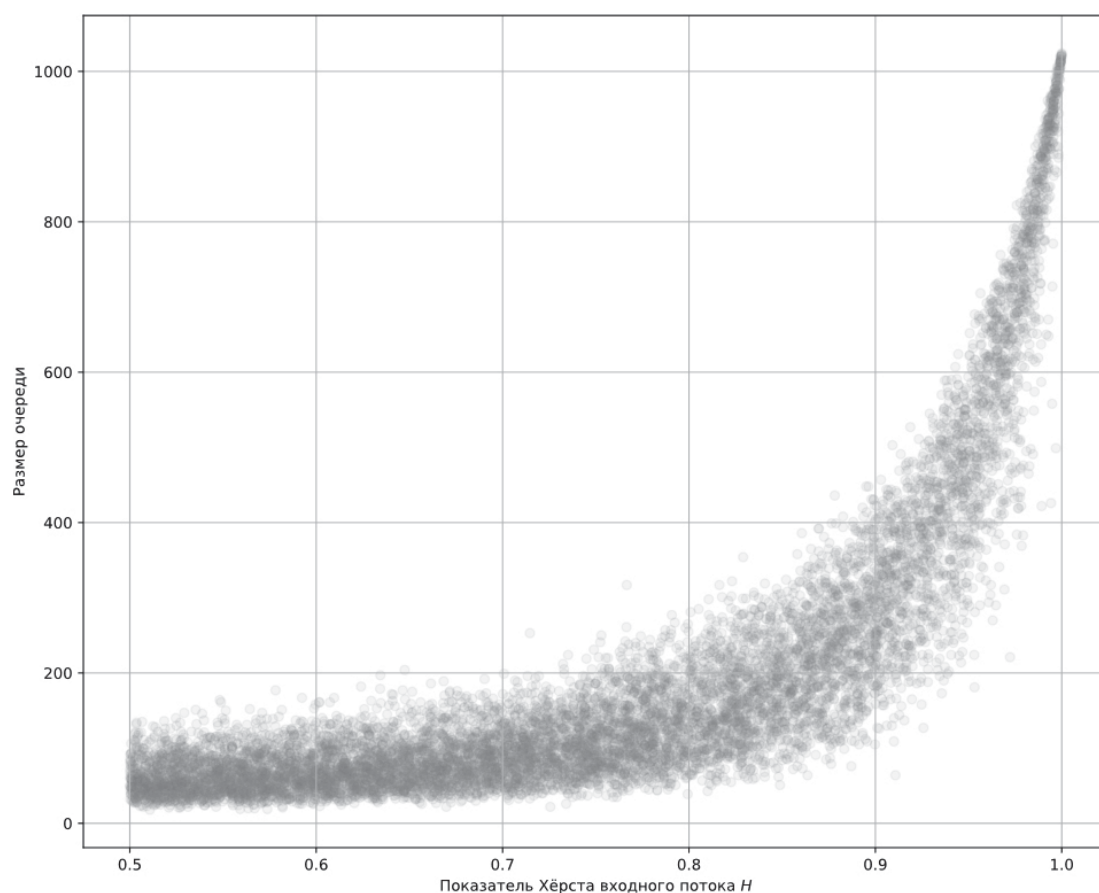


Рисунок 2. Диаграмма рассеивания зависимости размера очереди от показателя Херста

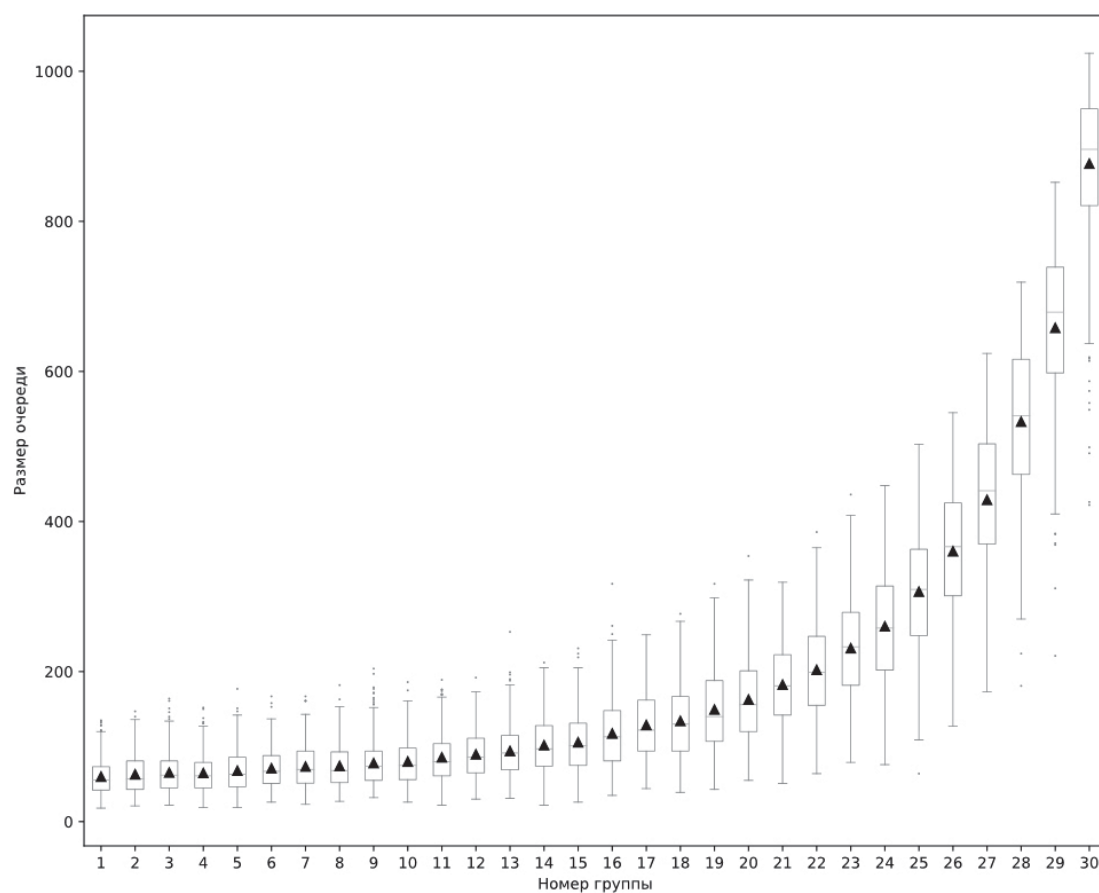


Рисунок 3. Вох-plot 30 групп анализируемых данных

Таблица 1. Метрики качества линейной регрессионной модели

Коэффициент детерминации R2	0,584792
Среднеквадратичная ошибка регрессии RMSE	130,908538
Средняя абсолютная ошибка MAE	96,808209
Величина штрафа p	55,710275

находить объем буфера для входного потока с заданным показателем Херста.

Линейный регрессионный анализ

В этом случае взаимосвязь между показателем Херста H и размером очереди $\hat{y}$ определяется согласно линейному уравнению $\hat{y} = b_0 + b_1 H$.

При помощи метода наименьших квадратов получим уравнение регрессии

$$\hat{y} = -611,182635 + 1077,442810H. \quad (1)$$

В таблице 1 приведены метрики качества полученной линейной регрессионной модели применительно к исходным данным. Полученное значение коэффициента детерминации говорит о том, что только около 58 % случаев изменения показателя Херста приводят к изменению размера очереди в рамках линейной модели.

Полученный результат является неудовлетворительным для практики, поэтому в простейшем случае имеет смысл рассмотреть другие способы, воспользовавшись методами линеаризации нелинейных зависимостей. В результате нелинейная зависимость может быть приведена к линейной, и далее можно использовать метод наименьших квадратов.

Гиперболическая регрессия

Для гиперболической регрессии взаимосвязь между H и $\hat{y}$ может быть описана как $\hat{y} = b_0 + \frac{b_1}{H}$. Линеаризация гиперболического уравнения достигается заменой $1/H$ на новую переменную, которую обозначим через z [4]. Тогда уравнение гиперболической регрессии примет вид $\hat{y} = b_0 + b_1 z$. Уравнение регрессии получим при помощи метода наименьших квадратов:

$$\hat{y} = \frac{875,438048 - 489,379751}{H}. \quad (2)$$

В таблице 2 приведены метрики качества полученной гиперболической регрессионной модели применительно к исходным данным. Полученное значение коэффициента детерминации говорит о том, что около 45 % случаев изменения показателя

Таблица 2. Метрики качества гиперболической регрессионной модели

Коэффициент детерминации R2	0,453263
Среднеквадратичная ошибка регрессии RMSE	150,218740
Средняя абсолютная ошибка MAE	110,511548
Величина штрафа p	63,841249

Таблица 3. Метрики качества измененной гиперболической модели

Коэффициент детерминации R2	0,591293
Среднеквадратичная ошибка регрессии RMSE	223,798030
Средняя абсолютная ошибка MAE	77,543626
Величина штрафа p	39,537133

Херста приводят к изменению размера очереди. Поскольку это гораздо хуже показателя линейной модели, имеет смысл рассмотреть модель гиперболической регрессии вида $\hat{y} = \frac{1}{b_0 + b_1 H}$.

При помощи метода наименьших квадратов получаем уравнение регрессии вида

$$\hat{y} = \frac{1}{0,039996 - 0,039720H}. \quad (3)$$

В таблице 3 отражены метрики качества измененной гиперболической регрессионной модели применительно к исходным данным.

Полученное значение коэффициента детерминации около 59 %, что несколько лучше линейной модели.

Степенная регрессия

В случае степенной регрессии взаимосвязь между H и $\hat{y}$ имеет вид $\hat{y} = b_0 H^{b_1}$. Данное уравнение является нелинейным по коэффициенту b_1 и относится к классу моделей регрессии, которые можно привести [4] к линейному виду $\ln y = \ln b_0 + b_1 \ln H$.

Показательная функция является внутренне линейной, поэтому оценки неизвестных параметров ее линеаризованной формы можно найти также при помощи метода наименьших квадратов. Уравнение регрессии имеет вид

$$\hat{y} = 401,143661H^{3,596636}. \quad (4)$$

В таблице 4 показаны метрики качества полученной степенной регрессионной модели применительно к исходным данным. Полученное значение 70 % гораздо лучше коэффициента детерминации линейной модели.

Таблица 4. Метрики качества степенной регрессионной модели

Коэффициент детерминации R2	0,699138
Среднеквадратичная ошибка регрессии RMSE	128,675573
Средняя абсолютная ошибка MAE	72,823850
Величина штрафа p	53,042530

Экспоненциальная регрессия

Для показательной регрессии взаимосвязь H и $\hat{y}$ имеет вид $\hat{y} = b_0 e^{b_1 H}$. Данное уравнение является нелинейным по коэффициенту b_1 и относится к классу моделей регрессии, которые, согласно [4], также приводятся к линейному виду $\ln \hat{y} = \ln b_0 + H \ln b_1$. Показательная функция является внутренне линейной, поэтому оценки неизвестных параметров ее линеаризованной формы можно найти при помощи классического метода наименьших квадратов. Уравнение регрессии имеет вид

$$\hat{y} = 2,926343e^{5,089127H}. \quad (5)$$

В таблице 5 приведены метрики качества экспоненциальной регрессионной модели применительно к тем же исходным данным. Полученное значение коэффициента детерминации говорит о том, что около 74 % случаев изменения показателя Херста приводят к изменению размера очереди в рамках экспоненциальной модели, что является наилучшим результатом при использовании методов парного регрессионного анализа. Такой же результат дает и анализ величины штрафа.

Проведем сравнение полученных результатов. Графики уравнений регрессии (1)–(5) представлены на рисунке 4. Очевидно, что экспоненциальная регрессия и степенная регрессия наиболее близко описывают зависимость между показателями Херста и объемом буфера.

Тривиальные модели парной регрессии, описанные выше, недостаточно хорошо описывают зависимость размера очереди от показателя Херста, поэтому произведем усложнение модели. Одним из возможных вариантов является изотоническая регрессия.

Изотоническая регрессия

В статистике изотоническая регрессия или монотонная регрессия – это метод подгонки линии свободной формы к последовательности наблюдений при следующих ограничениях: подобранная линия свободной формы должна быть неубывающей (или неувеличивающейся) на области определения и должна лежать как можно

Таблица 5. Метрики качества экспоненциальной регрессионной модели

Коэффициент детерминации R2	0,745779
Среднеквадратичная ошибка регрессии RMSE	112,443773
Средняя абсолютная ошибка MAE	65,199678
Величина штрафа p	46,768626

Таблица 6. Метрики качества изотонической регрессии

Коэффициент детерминации R2	0,928199
Среднеквадратичная ошибка регрессии RMSE	54,437567
Средняя абсолютная ошибка MAE	39,500659
Величина штрафа p	21,269167

ближе к наблюдениям [11]. В процессе построения изотонической кривой решается следующая задача [11]:

$$\sum_i w_i (y_i - \hat{y}_i)^2 \rightarrow \min,$$

где значение весового коэффициента $w_i > 0$. Это дает вектор, который состоит из неубывающих элементов, наиболее близких по среднеквадратической ошибке. На практике этот список элементов образует кусочно-линейную функцию.

Выполним обучение модели изотонической регрессии с использованием пакета scikit-learn языка программирования Python 3 [7]. Построим график, соответствующий модели, построенной с помощью изотонической регрессии, – см. рисунок 5.

В таблице 6 приведены метрики качества полученной регрессионной модели применительно к исходным данным. Полученное значение коэффициента детерминации говорит о том, что около 92 % случаев изменения показателя Херста приводят к изменению размера очереди в рамках данной модели, что гораздо лучше моделей, построенных на основе методов парной регрессии.

При этом величина штрафа для изотонической регрессии в два раза меньше соответствующей величины для парной регрессии.

Метод опорных векторов

Метод опорных векторов, или SVM (от англ. Support Vector Machines) – это линейный алгоритм, используемый в задачах классификации и регрессии (для задач регрессии он носит название SVR – Support Vector Regression). Основная идея метода заключается в построении гиперплоскости, разделяющей объекты выборки оптимальным способом [8–10].

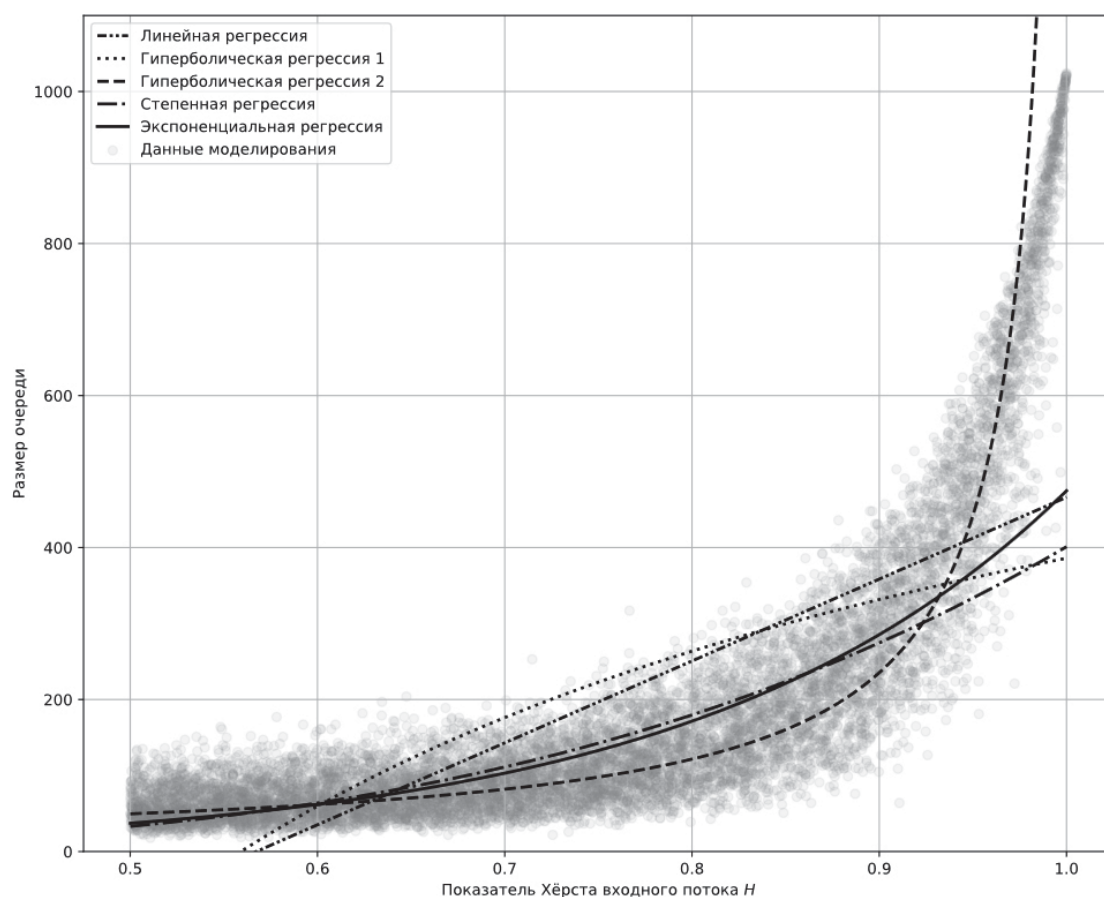


Рисунок 4. Сравнительный анализ результатов парного регрессионного анализа

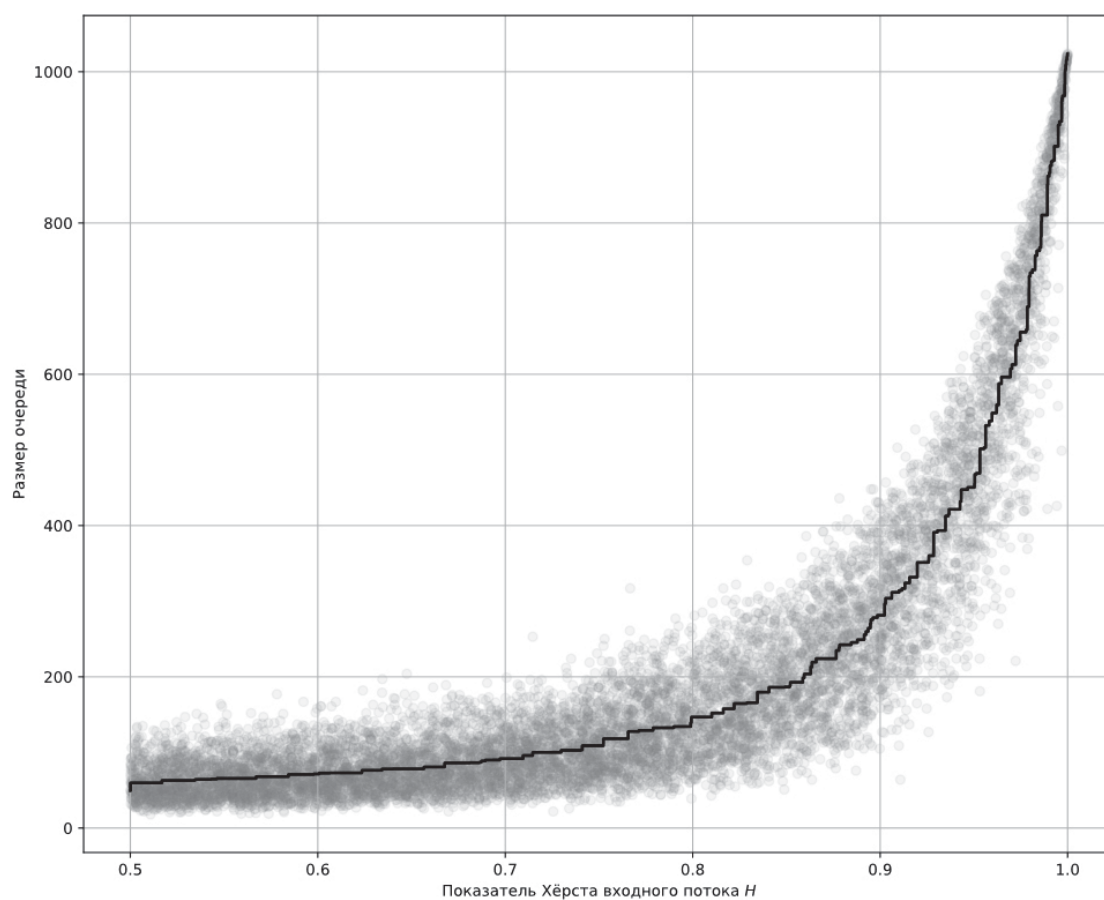


Рисунок 5. Построение изотонической кривой применительно к набору данных

Таблица 7. Метрики качества модели опорных векторов

Коэффициент детерминации R2	0,901167
Среднеквадратичная ошибка регрессии RMSE	63,868307
Средняя абсолютная ошибка MAE	52,506259
Величина штрафа p	18,374489

Таблица 8. Метрики качества модели опорных векторов в логарифмическом масштабе

Коэффициент детерминации R2	0,923960
Среднеквадратичная ошибка регрессии RMSE	56,021631
Средняя абсолютная ошибка MAE	39,583393
Величина штрафа p	21,724137

Таблица 9. Сравнительная характеристика методов регрессии при $0,5 < H < 1$

	Коэффициент детерминации R2	Среднеквадратичная ошибка регрессии RMSE	Средняя абсолютная ошибка MAE	Величина штрафа p
Линейная регрессия	0,584792	130,908538	96,808209	55,710275
Гиперболическая регрессия 1	0,453263	150,218740	110,511548	63,841249
Гиперболическая регрессия 2	0,591293	223,798030	77,543626	39,537133
Степенная регрессия	0,699138	128,675573	72,823850	53,042530
Экспоненциальная регрессия	0,745779	112,443773	65,199678	46,768626
Изотоническая регрессия	0,928199	54,437567	39,500659	21,269167
Метод опорных векторов SVR	0,901167	63,868307	52,506259	18,374489
Метод опорных векторов в логарифмическом масштабе	0,923960	56,021631	39,583393	21,724137

Метод опорных векторов максимизирует отступы объектов, что тесно связано с минимизацией вероятности переобучения. При этом он позволяет очень легко перейти к построению нелинейной разделяющей поверхности благодаря ядерному переходу [8; 9].

Выполним обучение модели на основе SVR. Нелинейный характер зависимости между показателем Херста и размером очереди говорит о необходимости выбора радиально-базисного ядра для модели SVR. Обучение данной модели произведено с использованием пакета scikit-learn языка программирования Python 3 [10]. На рисунке 6 представлен график зависимости между размером очереди и показателем Херста, соответствующий обученной модели SVR.

В таблице 7 приведены метрики качества полученной модели, использующей SVR. Полученное значение коэффициента детерминации – около 90 %, что немного хуже, чем у метода с использованием изотонической регрессии.

Однако величина штрафа у этого метода меньше, чем в случае применения изотонической регрессии. Исходя из характера зависимости между размером очереди QS и показателем Херста H , целесообразно при использовании SVR оценивать

не QS , а значение $\ln(QS + 1)$, переходя, таким образом, к спрямляющему пространству.

Выполним обучение модели на основе SVR с использованием пакета scikit-learn языка программирования Python 3 [10]. На рисунке 7 представлен график зависимости размера очереди и показателя Херста, соответствующий обученной модели методом опорных векторов с переходом в спрямляющее пространство.

В таблице 8 приведены метрики качества полученной модели, использующей SVR в логарифмическом масштабе, применительно к исходным данным.

Полученное значение коэффициента детерминации порядка 92 % практически совпадает с методом изотонической регрессии. Однако величина штрафа в этом случае меньше, чем для метода SVR. Таким образом, переход к спрямляющему пространству не приводит к улучшению качества обучения на основании значений введенной метрики качества, которой является штраф.

Сравнительный анализ моделей

Результаты исследования представлены в таблице 9 для оценки и выбора наилучшего мето-

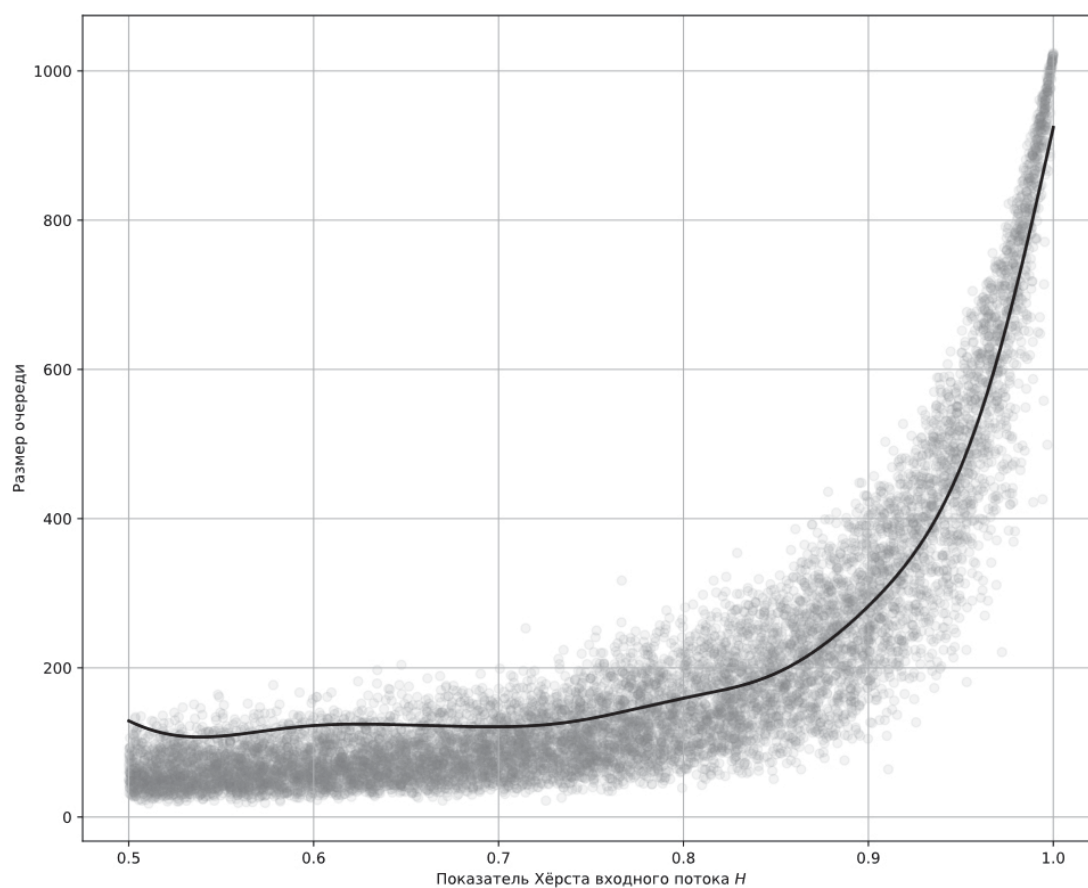


Рисунок 6. График, соответствующий модели, обученной методом опорных векторов

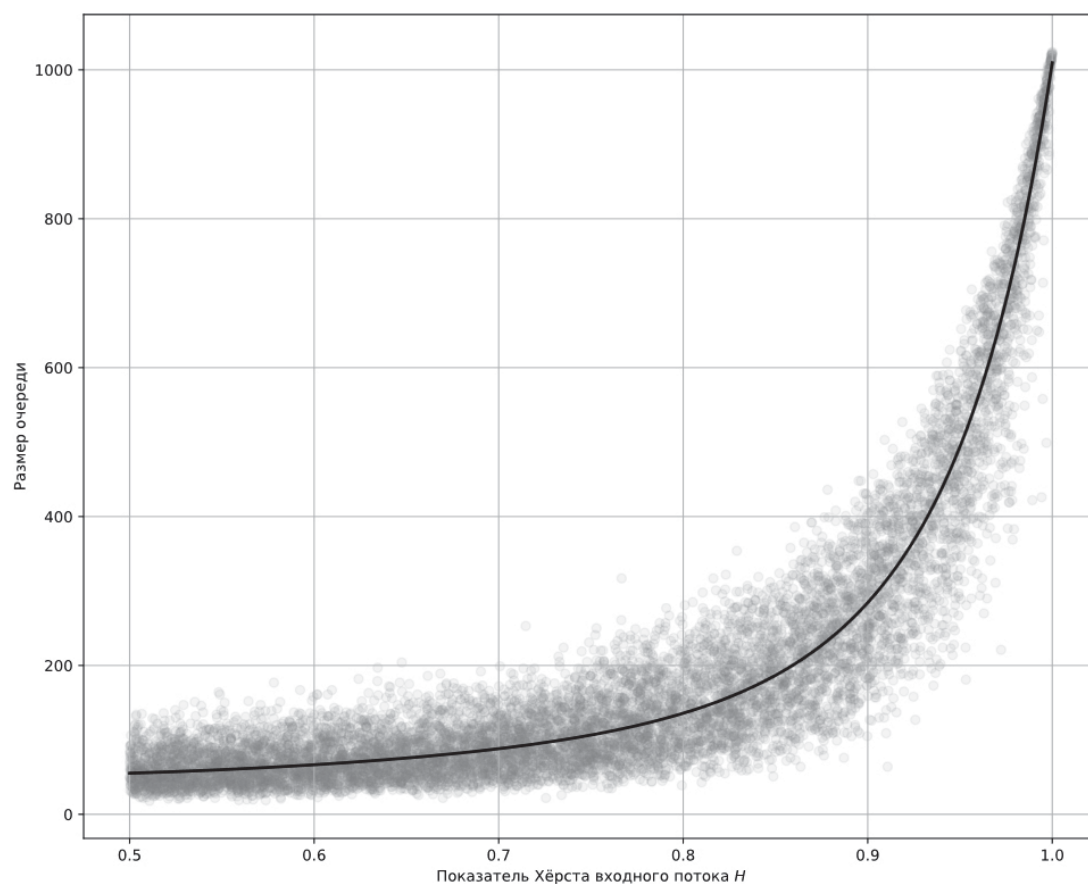


Рисунок 7. График, соответствующий модели, обученной методом опорных векторов, применительно к набору данных

да в интересах предсказания размера очереди по значению показателя Херста.

На основании данных сводной таблицы можно сделать вывод, что наилучшей предикативной способностью на основании введенной метрики качества является модель, построенная с использованием метода опорных векторов. В рамках данного исследования можно заключить, что усложнение SVR путем перехода в спрямляющее пространство не приводит к улучшению качества обучения.

Выводы

Таким образом, исследованы восемь моделей, позволяющие предсказать размер очереди при преобразовании входного потока, имеющего распределение Парето, в выходной поток с экспоненциальным распределением в зависимости от показателя Херста входного потока, построенные на основании методов машинного обучения.

Поскольку величина потерь в общем случае не дает никаких сведений об эффективности использования очереди в процессе преобразования трафика, для оценки качества полученной модели введен штраф, учитывающий не только величину потерь, но и нерациональное использование буферной памяти.

Каждая модель исследована с использованием следующих метрик качества: коэффициента детерминации, среднеквадратичной ошибки регрессии, средней абсолютной ошибки, величины штрафа.

Лучшими по выбранным метрикам качества являются модели, которые используют методы изотонической регрессии и опорных векторов.

Финансирование

Данные исследования выполнены при поддержке Российского фонда фундаментальных исследований (РФФИ), грант № 19-07-00856\20.

Литература

1. Шелухин О.И., Тенякшев А.М., Осин А.В. Фрактальные процессы в телекоммуникациях. М.: Радиотехника, 2003. 479 с.
2. Имитационная модель асинхронного преобразования самоподобного трафика в узлах коммутации с использованием очереди / Г.И. Линец [и др.] // Инфокоммуникационные технологии. 2019. Т. 17. № 3. С. 293–303.
3. Линец Г.И., Говорова С.В., Воронкин Р.А. Программа формирования набора данных для исследования статистических характеристик модели преобразования самоподобного трафика // Свидетельство о гос. регистрации программы для ЭВМ № 2019619275. Дата регистр. 15.07.2019.
4. Handbook of Mathematics; 6th ed. / I.N. Bronshtein [et al.]. Berlin: Springer, 2015. 1151 p. DOI: <https://doi.org/10.1007/978-3-662-46221-8>.
5. Базовые принципы машинного обучения на примере линейной регрессии. URL: <https://habr.com/ru/company/ods/blog/322076> (дата обращения: 01.04.2020).
6. Коэльо Л.П., Ричард В. Построение систем машинного обучения на языке Python / пер. с англ. А.А. Слинкина. М.: ДМК Пресс, 2016. 302 с.
7. Isotonic Regression. URL: <https://scikit-learn.org/stable/modules/isotonic.html> (дата обращения: 01.04.2020).
8. Шарден Б., Массарон Л., Боскетти А. Крупномасштабное машинное обучение вместе с Python / пер. с англ. А.В. Логунова. М.: ДМК Пресс, 2018. 358 с.
9. Рашка С. Python и машинное обучение / пер. с англ. А.В. Логунова. М.: ДМК Пресс, 2017. 418 с.
10. Support Vector Regression (SVR) using linear and non-linear kernels. URL: https://scikit-learn.org/stable/auto_examples/svm/plot_svm_regression.html?highlight=svr (дата обращения: 01.04.2020).
11. Westling T., Gilbert P., Carone M. Causal isotonic regression // arXiv:1810.03269. 2019. URL: <http://arxiv.org/abs/1810.03269> (дата обращения: 23.05.2020).

Получено 25.06.2020

Линец Геннадий Иванович, д.т.н., заведующий кафедрой инфокоммуникаций (ИК) Северо-Кавказского федерального университета (СКФУ). 355028, Российская Федерация, Ставропольский край, г. Ставрополь, пр. Кулакова, 2 (корп. 9). Тел. +7 865 295-69-97. E-mail: kbytw@mail.ru

Воронкин Роман Александрович, к.т.н., доцент кафедры ИК СКФУ. 355028, Российская Федерация, Ставропольский край, г. Ставрополь, пр. Кулакова, 2 (корп. 9). Тел. +7 865 295-69-97. E-mail: roman.voronkin@gmail.com

Говорова Светлана Владимировна, старший преподаватель кафедры ИК СКФУ. 355028, Российская Федерация, Ставропольский край, г. Ставрополь, пр. Кулакова, 2 (корп. 9). Тел. +7 865 295-69-97. E-mail: mitnik2@yandex.ru

Мочалов Валерий Петрович, д.т.н., профессор кафедры ИК СКФУ. 355028, Российская Федерация, Ставропольский край, г. Ставрополь, пр. Кулакова, 2 (корп. 9). Тел. +7 865 295-69-97. E-mail: mochalov.valery2015@yandex.ru

Палканов Илья Сергеевич, аспирант кафедры ИК СКФУ. 355028, Российская Федерация, Ставропольский край, г. Ставрополь, пр. Кулакова, 2 (корп. 9). Тел. +7 865 295-69-97. E-mail: ilya0693@yandex.ru

USING REGRESSION ANALYSIS METHODS FOR BUILDING AN OPTIMAL MODEL OF DEPENDENCY BETWEEN THE QUEUE SIZE AND HURST EXPONENT WHEN TRANSFORMING A SELF-SIMILAR INPUT PACKET FLOW INTO A FLOW WITH EXPONENTIAL DISTRIBUTION

*Linets G.I., Voronkin R.A., Govorova S.V., Mochalov V.P., Palkanov I.S.
North-Caucasus Federal University, Stavropol, Russian Federation
E-mail: kbytw@mail.ru*

Using machine learning methods, the model has been obtained for predicting the queue size of an input self-similar packet flow distributed according to the Pareto law when it is transformed into a flow with exponential distribution. Since the amount of losses in general case does not provide any information about the efficiency of using buffer space in the process of transforming a self-similar packet flow, a complex quality metric (penalty) was introduced to assess the quality of investigated models. This metric takes into account both packet loss during functional transformations and inefficient use of buffer space of switching nodes. It was shown that the models using isotonic regression and support vectors methods are the best by the considered metric.

Keywords: *telecommunications network, self-similar traffic, Hurst indicator, performance, Pareto distribution, packet loss, regression analysis, quality metrics, penalty function, machine learning*

DOI: 10.18469/ikt.2020.18.3.04

Linets Gennadiy Ivanovich, North-Caucasus Federal University, 2 (building 9), Kulakov Avenu, Stavropol, Stavropol Territory, 355028, Russian Federation; Head of Infocommunication Department, Doctor of Technical Science. Tel. +7 865 295-69-97. E-mail: kbytw@mail.ru

Voronkin Roman Alexandrovich, North-Caucasus Federal University, 2 (building 9), Kulakov Avenu, Stavropol, Stavropol Territory, 355028, Russian Federation; Associate Professor of Infocommunication Department, PhD in Technical Science. Tel. +7 865 295-69-97. E-mail: roman.voronkin@gmail.com

Govorova Svetlana Vladimirovna, North-Caucasus Federal University, 2 (building 9), Kulakov Avenu, Stavropol, Stavropol Territory, 355028, Russian Federation; Senior Teacher of Department of Infocommunication. Tel. +7 865 295-69-97. E-mail: mitnik2@yandex.ru

Mochalov Valeriy Petrovich, North-Caucasus Federal University, 2 (building 9), Kulakov Avenu, Stavropol, Stavropol Territory, 355028, Russian Federation; Professor of Infocommunication Department, Doctor of Technical Science. Tel. +7 865 295-69-97. E-mail: mochalov.valery2015@yandex.ru

Palkanov Ilya Sergeevich, North-Caucasus Federal University, 2 (building 9), Kulakov Avenu, Stavropol, Stavropol Territory, 355028, Russian Federation; PhD Student of Infocommunication Department. Tel. +7 865 295-69-97. E-mail: ilya0693@yandex.ru

References

1. Sheluhin O.I., Tenjakshev A.M., Osin A.V. *Fractal Processes in Telecommunications*. Moscow: Radiotekhnika, 2003, 479 p. (In Russian.)
2. Linets G.I. et al. Simulation model of asynchronous transformation of self-similar traffic in switching nodes using a queue. *Infokommunikacionnye tehnologii*, 2019, vol. 17, no. 3, pp. 293–303. (In Russian.)
3. Linets G.I., Govorova S.V., Voronkin R.A. A program for generating a dataset for studying the statistical characteristics of a self-similar traffic transformation model. *Svidetel'stvo o gos. registratsii programmy dlja EVM № 2019619275*, 2019. (In Russian.)
4. Bronshtein I.N. et al. *Handbook of Mathematics*. 6th ed. Berlin: Springer, 2015, 1151 p. DOI: <https://doi.org/10.1007/978-3-662-46221-8>.
5. *Basic principles of machine learning using linear regression as an example*. URL: <https://habr.com/ru/company/ods/blog/322076> (accessed 01.04.2020).
6. Koel'o L.P., Richard V. *Building Machine Learning Systems in Python*. Trans. from Eng. by A.A. Slinkin. Moscow: DMK Press, 2016, 302 p. (In Russian.)
7. *Isotonic Regression*. URL: <https://scikit-learn.org/stable/modules/isotonic.html> (accessed 01.04.2020).
8. Sharden B., Massaron L., Bosketti A. *Large-Scale Machine Learning With Python*. Trans. from Eng. by A.V. Logunov. Moscow: DMK Press, 2018, 358 p. (In Russian.)
9. Rashka S. *Python and Machine Learning*. Trans. from Eng. by A.V. Logunov. Moscow: DMK Press, 2017, 418 p. (In Russian.)
10. *Support Vector Regression (SVR) using linear and non-linear kernels*. URL: https://scikit-learn.org/stable/auto_examples/svm/plot_svm_regression.html?highlight=svr (accessed 01.04.2020).
11. Westling T., Gilbert P., Carone M. Causal isotonic regression. *arXiv:1810.03269*, 2019. URL: <http://arxiv.org/abs/1810.03269> (accessed 23.05.2020).

Received 25.06.2020

УДК 004.942

АНАЛИТИЧЕСКАЯ МОДЕЛЬ КОММУТАТОРА КОРПОРАТИВНОЙ ПРОГРАММНО-УПРАВЛЯЕМОЙ СЕТИ SDN

Мочалов В.П., Линец Г.И., Братченко Н.Ю.

Северо-Кавказский федеральный университет, Ставрополь, РФ

E-mail: n.b.20062@yandex.ru

Реализация практически неограниченных возможностей программно-управляемой сети вызывает необходимость дополнительного исследования ее инфраструктурного уровня, оценки телекоммуникационной составляющей. Поэтому целью исследования является разработка аналитической модели для анализа основных показателей качества сетевых коммутаторов и каналов связи. Для формирования математической модели коммутатора программно-управляемой сети, функционирующей в стационарном режиме и обслуживающей запросы случайного объема, предлагается использовать теорию системы массового обслуживания, а также аппарат полумарковских процессов. Модель коммутатора формально представлена в виде открытой сети массового обслуживания с пуассоновским потоком поступающих пакетов, отказами и блокировками. Первый этап обслуживания запросов включает в себя процессы записи сетевых пакетов в многолинейную буферную память коммутатора. В следующем центре осуществляется проверка соответствия адреса поступившего пакета записям в таблицах потоков. Третий однолинейный центр, реализующий процесс передачи выходящих из коммутатора пакетов, может быть описан системой с независимыми однолинейными системами массового обслуживания с n -местными буферами. Получены зависимости времени ожидания обслуживания и времени обслуживания поступающих сетевых пакетов от нагрузки, а также выражения для определения объема буферной памяти коммутатора при допустимой вероятности потери сообщений, показатели качества его передающей части.

Ключевые слова: таблицы потоков, OpenFlow коммутатор, модели массового обслуживания, контроллер, вероятностно-временные характеристики