

Людмила Михайловна Погребняк, адъюнкт, Военный институт телекоммуникаций и информатизации, Киев, Украина.

Николай Иванович Науменко, доктор технических наук, профессор, профессор кафедры транспортных сетей, Военный институт телекоммуникаций и информатизации, Киев, Украина.

Liudmyla Pohrebniak, postgraduate student, Military institute of telecommunications and informatization, Kyiv, Ukraine.

Mykola Naumenko, doctor of technical sciences, professor, professor at the transport networks academic department, Military institute of telecommunications and informatization, Kyiv, Ukraine.

УДК 621.396

РУСЛАН АБДУРАМАНОВ
ЮРИЙ ХЛАПОНИН
АНТОН ХАДДАД

ПРИМЕНЕНИЕ МЕТОДОВ РЕГРЕССИОННОГО АНАЛИЗА ДЛЯ МОНИТОРИНГА И УПРАВЛЕНИЯ ТЕЛЕКОММУНИКАЦИОННЫМИ СИСТЕМАМИ

Телекоммуникационные системы являются важнейшей частью современной жизни человека. Абсолютно все аспекты деятельности современного человека попали в зависимость от эффективной работы телекоммуникационных сетей. Развитие телекоммуникационных систем и компьютерных сетей обуславливает необходимость создания и надежного функционирования большого набора инфокоммуникационных сервисов, обеспечивающих эффективную работу пользователя с разнородной информацией в телекоммуникационной сети. Исторически сложившаяся неоднородность, как телекоммуникационных систем, компьютерных сетей, сетевых информационных ресурсов, так и аудитории пользователей, которой данная информация адресована, осложняет объективный анализ и мониторинг телекоммуникационных архитектур и ресурсов. Поэтому, безусловно, актуально, что при эксплуатации телекоммуникационных систем и компьютерных сетей должен быть использован, достаточно широкий спектр современных и научно обоснованных технических и технологических решений их анализа и мониторинга. Этот факт и обуславливает важность решения задач мониторинга и управления телекоммуникационными системами. В последнее время стали популярными интеллектуальные системы принятия решений на базе систем обработки данных с применением технологий машинного обучения. В данной работе рассматривается одна из таких технологий, основанная на применении логистической регрессии. Используя данные, характеризующие работу телекоммуникационной сети, была построена модель, описывающая работу сети. В частности, доказана возможность использования логистической регрессии для прогноза вероятности возникновения неэффективной работы сетей на базе процессов, которые в них протекают. На базе модели, была построена интеллектуальная система принятия решения для мониторинга и управления состоянием работы сети в реальном времени.

Ключевые слова: телекоммуникационная система, интеллектуальная система принятия решений, логистическая регрессия, компьютерная сеть, протокол SNMP.

Постановка проблемы. Современный мир сложно представить без многочисленных сложных цифровых радиоэлектронных технических устройств, которые не только помогают

в организации бизнес процессов на крупных предприятиях, но и глубоко вошли в повседневную жизнь каждого человека. Несмотря на разнообразие существующих устройств, их функциональных возможностей и отличия друг от друга, они почти полностью теряют свою актуальность, если лишить их возможности связи друг с другом по средствам информационных технологий обмена данными. Несмотря на то, что на первый взгляд многие радиоэлектронные устройства кажутся самостоятельными, при более детальном рассмотрении они выполняют функции сбора, обработки и передачи информации. И даже если в устройстве предусмотрен весь вышеперечисленный функционал, они все равно нуждаются в обмене информацией друг с другом.

На текущий момент информационные технологии ушли далеко вперед, что делает возможным передачу информации между устройствами на высоком уровне. Однако из-за сложной многофункциональной природы данных устройств, ключевую роль в информационных технологиях играют вопросы поддержки высокой эффективности обмена информацией. Одной из основных характеристик, определяющих качество функционирования цифровых радиоэлектронных устройств, предназначенных исключительно для поддержки обмена данными между оставшимися устройствами, является надежность. Работоспособность и надежность телекоммуникационных систем и компьютерных сетей, состоящих из радиоэлектронных устройств обработки и передачи информации, в конечном итоге будут характеризоваться работоспособностью и надежностью самого неэффективного устройства. Поэтому, своевременное предупреждение путем выявления соответствующих неэффективных устройств является ключевым элементом в анализе и мониторинге подобных систем.

Анализ последних исследований и публикаций. Для решения задач анализа и мониторинга телекоммуникационных систем и компьютерных сетей, все чаще используют прикладные интеллектуальные технологии [1]. В основе таких технологий лежат так называемые технологии машинного обучения, на базе которых строятся системы принятия решений для интеллектуального управления сложными распределенными инфокоммуникационными сетями [2], [3]. Алгоритмы машинного обучения, в основе своей, используют идею анализа зависимости внутренних и внешних ситуаций (событий). Общая архитектура подобных интеллектуальных систем состоит из таких элементов:

- телекоммуникационная сеть;
- база данных характеристик работы телекоммуникационной системы;
- система обработки данных на базе технологий машинного обучения;
- система принятия решений;

Интеллектуальные системы могут существенно различаться по архитектуре и по функциям, выполняемым, но в них всегда в той или иной степени присутствуют указанные блоки [2].

Важно отметить, что главная архитектурная особенность, отличающая интеллектуальную систему управления (див. рис. 1.) от построенной по “традиционной” схеме, обусловлена подключением механизмов хранения и интеллектуальной обработки данных, характеризующих работу систем.

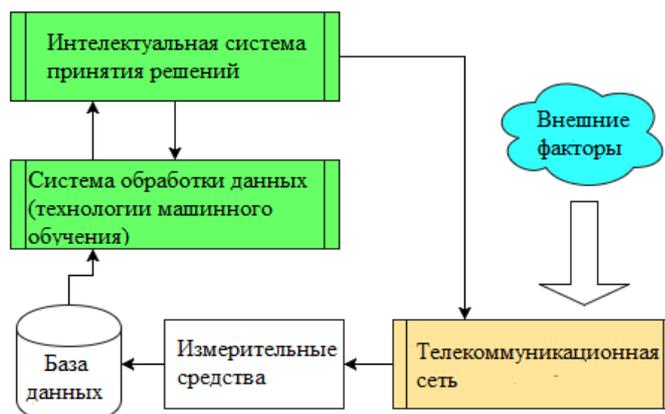


Рисунок 1 – Общая структура системы интеллектуального управления

Данная особенность интеллектуальных систем позволяет в реальном времени проводить мониторинг, анализировать и эффективно управлять телекоммуникационной сетью.

Основной целью исследования является построение оптимальной модели обработки данных, характеризующих работу телекоммуникационных систем, для оптимизации интеллектуальной системы принятия решений в процессе мониторинга и управления их состоянием в реальном времени.

Изложение основного материала исследований. Логистическая регрессия [3], [4], [7] применяется для предсказания вероятности возникновения некоторого события по значениям множества признаков. Для этого вводится так называемая зависимая переменная y , принимающая лишь одно из двух значений – как правило, это числа “0” (событие не произошло) и “1” (событие произошло), и множество независимых переменных (также называемых признаками, предикторами или регрессорами) – вещественных x_1, x_2, \dots, x_n , на основе значений которых требуется вычислить вероятность принятия того или иного значения зависимой переменной.

Делается предположение о том, что вероятность наступления события $y = 1$ равна:

$$P\{y = 1 | x\} = f(z),$$

где $z = \theta^T x = \theta_1 x_1 + \dots + \theta_n x_n$, x и θ векторы-столбцы значений независимых переменных x_1, x_2, \dots, x_n и параметров (коэффициентов регрессии) – вещественных чисел $\theta_1, \theta_2, \dots, \theta_n$, соответственно;

$f(z)$ – логистическая функция

$$f(z) = \frac{1}{1 + e^{-z}}.$$

Так как y принимает лишь значения “0” и “1”, то вероятность первого возможного значения равна

$$P\{y = 0 | x\} = 1 - f(z) = 1 - f(\theta^T x)$$

Для краткости функцию распределения y при заданном x можно записать в таком виде:

$$P\{y | x\} = f(\theta^T x)^y (1 - f(\theta^T x))^{1-y}, \quad y \in \{0, 1\}.$$

Подбор параметров. Для подбора параметров $\theta_1, \theta_2, \dots, \theta_n$ необходимо составить обучающую выборку, состоящую из наборов значений независимых переменных x и соответствующих им значений зависимой переменной y . Формально, это множество пар $(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})$, где $x^{(i)} \in \mathcal{R}^n$ – вектор значений независимых переменных, а $y^{(i)} \in \{0, 1\}$ – соответствующее им значение y . Каждая такая пара называется обучающим примером.

Обычно используется метод максимального правдоподобия [4]- [6], согласно которому выбираются параметры θ , максимизирующие значение функции правдоподобия на обучающей выборке:

$$\hat{\theta} = \arg \max_{\theta} L(\theta) = \arg \max_{\theta} \prod_{i=1}^m P\{y = y^{(i)} | x = x^{(i)}\}$$

Максимизация функции правдоподобия эквивалентна максимизации её логарифма:

$$\ln L(\theta) = \sum_{i=1}^m \log P\{y = y^{(i)} | x = x^{(i)}\} = \sum_{i=1}^m y^{(i)} \ln f(\theta^T x^{(i)}) + (1 - y^{(i)}) \ln(1 - f(\theta^T x^{(i)}))$$

Для максимизации этой функции может быть применён, например, метод градиентного спуска [5]. Он заключается в выполнении следующих итераций, начиная с некоторого начального значения параметров θ :

$$\theta := \theta + \alpha \nabla \ln L(\theta) = \theta + \alpha \sum_{i=1}^m (y^{(i)} - f(\theta^T x^{(i)})) x^{(i)}, \quad \alpha > 0.$$

Получение данных. В качестве примера рассматривалось использование протокола SNMP (Simple Network Management Protocol) для осуществления администратором сбора данных о функционировании компьютерной сети. Данные регистрировались с интервалом 1 мин. Для анализа были определены следующие характеристики:

- число потерянных пакетов;
- загрузка процессора;
- температура процессора;
- объем свободной памяти;
- напряжение шины.

Обработка данных. Все данные были проанализированы на наличие отсутствующих наблюдений и ошибок в значениях. Для визуального анализа, по каждому из параметров, были построены гистограммы плотности распределения значений.

В качестве зависимой переменной был взят показатель – число потерянных пакетов. Зная общее число пакетов, был получен относительный показатель потери пакетов. Далее, экспертно, была выбрана граница допустимого уровня потери пакетов ρ , с помощью которого данный показатель преобразовался в показатель с бинарными значениями (0, 1) по формуле:

$$y'_i = \begin{cases} 1, & \text{если } y_i > \rho \\ 0, & \text{в противном случае.} \end{cases}$$

В качестве независимых переменных, были взяты параметры: загрузка процессора, температура процессора, объем свободной памяти, напряжение шины. Все параметры были нормализованы для того, чтобы была возможность их соотнесения между собой в модели.

Выводы. 1. Анализ построенной модели показал, что есть значимая зависимость между переменной – удовлетворительный или неудовлетворительный процент потери пакетов и переменными: число потерянных пакетов, загрузка процессора, температура процессора, объем свободной памяти, напряжение шины.

2. Доказана возможность использования логистической регрессии для прогноза вероятности возникновения неэффективной работы сетей на базе процессов, протекающих в телекоммуникационных сетях.

3. Определены особенности статистического мониторинга телекоммуникационных сетей, а именно: нестационарность, периодичность (неравномерность загрузки каналов), нелинейное влияние характеристик работы сети на ее эффективность.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

- [1] Ю.І. Хлапонін, Г.Б. Жиров, та О.М. Нікітчін, “Застосування нейронних мереж в статистичній системі аналізу і моніторингу телекомунікаційних мереж”, *Технологічний аудит і резерви виробництва*, т. 5, № 2, с. 35-41, 2016.
- [2] И.С. Енюков, И.В. Ретинская, и А.К. Скуратов *Статистический анализ и мониторинг научно-образовательных интернет-сетей*. Москва, Российская Федерация: Финансы и статистика, 2004.
- [3] Ю.П. Недайбида, Ю.В. Котова, и Ю.И. Хлапонин, “Современные проблемы создания сложных информационно-управляющих систем реального времени”, *Захист інформації*, т. 14, № 4 (57), с. 50-55, 2012.
doi: 10.18372/2410-7840.14.3486.
- [4] С.А. Айвазян, В.М. Бухтштабер, И.С. Енюков, Л.Д. Мешалкин, *Прикладная статистика: классификация и снижение размерности*. Москва, Россия: Финансы и статистика, 1989.
- [5] D. Hosmer, and S. Lemeshow, *Applied Logistic Regression*, 2nd ed. New York, USA: Wiley, 2002.
doi: 10.1002/0471722146.

- [6] R. Koenker, and G. Bassett, "Regression Quantiles", *Econometrica*. vol. 46, no.1, pp. 33-50, 1978.
doi: 10.2307/1913643.
- [7] R. Koenker, K.F. Hallock, "Quantile Regression" , *Journal of Economic Perspectives*, vol. 15, no. 4, pp. 143-156, 2001.
doi: 10.1257/jep.15.4.143.

Статья поступила в редакцию 05 сентября 2017 года.

REFERENCE

- [1] Y.I. Khlaponin, H.B. Zhyrov, and O.M. Nikitchin, "Application of Neural Networks in the Statistical System for Analysis and Monitoring of Telecommunication Networks", *Technological Audit and Production Reserves*, vol. 5, no. 2, pp. 35-41, 2016.
- [2] I.S. Eniukov, I.V. Retinskaia, and A.K. Skuratov, *Statistical analysis and monitoring of scientific and educational Internet networks*. Moscow, Russia: Finance and Statistics, 2004.
- [3] Y.P. Nedaibida, Y.V. Kotova, and Y.I. Khlaponin, "Modern problems of creating complex real-time information management systems", *Ukrainian Information Security Research Journal*, vol. 14, no. 4 (57), pp. 50-55, 2012.
doi: 10.18372/2410-7840.14.3486.
- [4] S.A. Aivazian, V.M. Bukhtshtaber, I.S. Eniukov, and L.D. Meshalkin, *Applied statistics: classification and reduction of dimension*. Moscow, Russia: Finance and Statistics, 1989.
- [5] D. Hosmer, and S. Lemeshow, *Applied Logistic Regression*, 2nd ed. New York, USA: Wiley, 2002.
doi: 10.1002/0471722146.
- [6] R. Koenker, and G. Bassett, "Regression Quantiles", *Econometrica*. vol. 46, no.1, pp. 33-50, 1978.
doi: 10.2307/1913643.
- [7] R. Koenker, K.F. Hallock, "Quantile Regression" , *Journal of Economic Perspectives*, vol. 15, no. 4, pp. 143-156, 2001.
doi: 10.1257/jep.15.4.143.

РУСЛАН АБДУРАМАНОВ
ЮРІЙ ХЛАПОНІН
АНТОН ХАДДАД

ЗАСТОСУВАННЯ МЕТОДІВ РЕГРЕСІЙНОГО АНАЛІЗУ ДЛЯ МОНІТОРИНГУ ТА УПРАВЛІННЯ ТЕЛЕКОМУНІКАЦІЙНИМИ СИСТЕМАМИ

Телекомунікаційні системи є найважливішою частиною сучасного життя людини. Абсолютно всі аспекти діяльності сучасної людини потрапили в залежність від ефективної роботи телекомунікаційних мереж. Розвиток телекомунікаційних систем та комп'ютерних мереж обумовлює необхідність створення і надійного функціонування великого набору інфокомунікаційних сервісів, що забезпечують ефективну роботу користувача з різномірною інформацією в телекомунікаційній мережі. Історично склалася неоднорідність, як телекомунікаційних систем, комп'ютерних мереж, мережевих інформаційних ресурсів, так і аудиторії користувачів, якій ця інформація адресована, ускладнює об'єктивний аналіз і моніторинг телекомунікаційних архітектур і ресурсів. Тому, безумовно, актуально, що при експлуатації телекомунікаційних систем і комп'ютерних мереж повинен бути використаний, досить широкий спектр сучасних і науково обґрунтованих технічних і технологічних рішень їх аналізу та моніторингу. Цей факт і обумовлює важливість вирішення завдань моніторингу та управління телекомунікаційними системами. Останнім часом стали популярними інтелектуальні системи прийняття рішень на базі систем обробки даних із застосуванням технологій машинного навчання. У даній роботі розглядається одна з таких технологій -

логістична регресія. Використовуючи дані характеризують роботу телекомунікаційної мережі, була побудована модель, що описує роботу мережі. Зокрема, доведено можливість використання логістичної регресії для прогнозування імовірності виникнення неефективної роботи мереж на основі процесів, що в них відбуваються. На базі моделі, була побудована інтелектуальна система прийняття рішення, яка використовувалася для моніторингу та управління станом роботи мережі в реальному часі.

Ключові слова: телекомунікаційна система, інтелектуальна система прийняття рішень, логістична регресія, комп'ютерна мережа, протокол SNMP.

RUSLAN ABDURAMANOV
YURI KHLAPONIN
ANTON KHADDAD

USING OF REGRESSION ANALYSIS METHODS FOR MONITORING AND MANAGING TELECOMMUNICATION SYSTEMS

Telecommunication systems are a crucial part of modern human life. Absolutely all aspects of the activities of modern people have become dependent on the effective operation of telecommunications networks. The development of telecommunications systems and computer networks necessitates the creation and reliable operation of a large set of info communication services that ensure the effective operation of the user with heterogeneous information in the telecommunications network. The historically formed heterogeneity of both telecommunication systems, computer networks, network information resources, and the audience of users to whom this information is addressed complicates the objective analysis and monitoring of telecommunication architectures and resources. Therefore, it is certainly true that when operating telecommunication systems and computer networks, a fairly wide range of modern and scientifically sound technical and technological solutions for their analysis and monitoring should be used. Due to this fact, solving problems of the monitoring and managing telecommunications systems is of utmost importance. Recently, intelligent decision-making systems, based on data processing systems with the use of machine learning technologies, have become popular. In this paper, one such technology based on logistic regression is considered. Using real data of telecommunication network functioning, the model describing network functioning has been constructed. In particular, the possibility of using logistic regression to predict the probability of inefficient operation of networks based on the processes that occur in them has been proved. Based on the model, an intelligent decision-making system has been built, which is used to monitor and control the state of the network in real time.

Keywords: telecommunication system, intellectual decision-making system, logistic regression, computer network, SNMP protocol.

Руслан Айдерович Абдураманов, доцент кафедри кібернетической безпеки и компьютерной инженерии, Киевский национальный университет строительства и архитектуры, Киев, Украина.

E-mail: ara8131@gmail.com.

Юрий Иванович Хлапонин, доктор технических наук, старший научный сотрудник, заведующий кафедрой кібернетической безпеки и компьютерной инженерии, Киевский национальный университет строительства и архитектуры, Киев, Украина.

E-mail: khlaponin.yui@knuba.edu.ua.

Антон Георгович Хаддад, аспирант кафедри кібернетической безпеки и компьютерной инженерии, Киевский национальный университет строительства и архитектуры, Киев, Украина.

E-mail: symplicitystudio@gmail.com.

Руслан Айдерович Абдураманов, доцент кафедри кібернетической безпеки та комп'ютерної інженерії, Київський національний університет будівництва і архітектури, Київ, Україна.

Юрій Іванович Хлапонін, доктор технічних наук, старший науковий співробітник, завідувач кафедри кібернетичної безпеки та комп'ютерної інженерії, Київський національний університет будівництва і архітектури, Київ, Україна.

Антон Георгович Хаддад, аспірант кафедри кібернетичної безпеки та комп'ютерної інженерії, Київський національний університет будівництва і архітектури, Київ, Україна.

Ruslan Abduramanov, assistant professor at the cyber security and computer engineering academic department, Kyiv National University of Construction and Architecture, Kyiv, Ukraine.

Yurii Khlaponin, doctor of technical science, senior researcher, head at the cyber security and computer engineering academic department, Kyiv National University of Construction and Architecture, Kyiv, Ukraine.

Anton Khaddad, graduate student, the cyber security and computer engineering academic department, Kyiv National University of Construction and Architecture, Kyiv, Ukraine.