

УДК 004.855.5

DOI: 10.25045/jpit.v11.i2.12

Шыхалиев Р.Г.Институт Информационных Технологий НАНА, Баку, Азербайджан
ramiz@science.az**ОБ ОДНОМ МЕТОДЕ ПРОГНОЗИРОВАНИЯ ТРАФИКА КОМПЬЮТЕРНЫХ СЕТЕЙ**

Поступила: 01.06.2020

Исправлена: 10.06.2020

Принята: 19.06.2020

Точное прогнозирование объема трафика компьютерных сетей (КС) как в долгосрочном, так и краткосрочном периоде времени играет очень важную роль в мониторинге, а также в эффективном управлении оптимальным использованием имеющихся сетевых ресурсов. Обычно более опытные сетевые администраторы интуитивно прогнозируют объем трафика КС, однако это совсем неприемлемо для администрирования современных, больших и сложных КС. Поэтому должны быть разработаны более точные методы прогнозирования объема трафика КС с использованием методов машинного обучения, которые помогут администраторам КС эффективно планировать и оптимально управлять использованием имеющихся сетевых ресурсов. В данной статье предлагается метод краткосрочного прогнозирования объема трафика КС на основе модели CART (Classification and Regression Trees). Суть метода заключается в том, что с помощью деревьев решений предыдущие множества состояний объема трафика классифицируются по множеству шаблонов состояний объема трафика и строится модель линейной регрессии, соответствующая каждому классу. Метод позволяет прогнозировать будущее состояние объема трафика КС путем кластеризации векторов текущих состояний трафика по наиболее подходящим предыдущим шаблонам и использования регрессии. Таким образом, задача краткосрочного прогнозирования объема трафика КС состоит в определении векторов текущих состояний объема трафика и модели регрессии для прогнозирования. Для оценки точности метода используется средняя абсолютная масштабированная ошибка MASE (Mean Absolute Scaled Error). Предложенный метод позволит прогнозировать объем трафика КС для коротких периодов времени, например, для периодов недель, дней, часов и секунд. Результаты краткосрочных прогнозов могут быть использованы для улучшения QoS (Quality of Service), предотвращения перегрузки каналов связи, оптимального управления имеющимися ресурсами КС и т.п.

Ключевые слова: компьютерные сети, сетевой трафик, прогнозирование трафика, модель CART, классификация, дерево решений.

Введение

За последнее десятилетие в компьютерных сетях (КС) трафик значительно вырос. Это связано с появлением новых сетевых технологий, сервисов и приложений, а также с увеличением числа пользователей и среднего объема интернет-трафика пользователей. Вместе с тем стремительно развиваются такие технологии, как интернет вещей (IoT), облачные вычисления, центры обработки данных и т.п. При этом имеющиеся в КС сетевые ресурсы используются совместно, а доступная пропускная способность сети меняется со временем. В таких условиях современные КС должны отвечать постоянно растущим потребностям в трафике и обеспечивать приемлемый уровень QoS (Quality of Service) для сетевых пользователей, приложений и сервисов. При этом эффективное управление пропускной способностью и производительностью сети, а также предотвращение перегрузок каналов связи имеют первостепенное значение. Однако в условиях непрерывного и чрезмерного увеличения объема и изменения характеристик трафика КС решение этих задач становится трудным. Для решения указанных задач необходимо

осуществлять постоянный и эффективный мониторинг трафика КС.

Одним из факторов, который влияет на эффективность мониторинга КС, можно считать точный прогноз трафика. Точный прогноз трафика КС позволит эффективно планировать и динамически распределять имеющиеся сетевые ресурсы (например, пропускную способность сети), предотвращать перегрузку каналов связи, обеспечивать приемлемый QoS для сетевых приложений, обеспечивать безопасность и т.п. Однако с ростом объема и сложности трафика КС задача прогнозирования усложняется и точность прогноза снижается. Выбор метода для прогнозирования трафика КС зависит от цели прогнозирования и характеристик трафика.

Прогнозирование трафика КС может быть разделено на две категории, такие как долгосрочное и краткосрочное прогнозирование [1]. Долгосрочное прогнозирование позволяет прогнозировать трафик КС на длительные периоды времени и оценивать будущие характеристики трафика, что позволит планировать КС более детально. Краткосрочный прогноз позволяет прогнозировать характеристики трафика КС на короткие периоды и решать проблемы, связанные с динамическим распределением сетевых ресурсов. Краткосрочное прогнозирование трафика КС может быть использовано для мониторинга КС в реальном времени. Результаты краткосрочных прогнозов могут быть использованы для улучшения QoS, контроля перегрузки каналов связи, оптимального управления ресурсами КС и т.п.

Предложенные в литературе методы по прогнозированию сетевого трафика в основном посвящены долгосрочному прогнозированию и основываются на традиционных методах прогнозирования. В данной работе рассматривается проблема краткосрочного прогнозирования трафика КС, а именно краткосрочного прогнозирования объема трафика КС. Потому, что объем сетевого трафика является одним из важных и фактических индикаторов активности сетей (сетевых пользователей, приложений, сервисов) и используется для мониторинга трафика КС. По анализу динамики изменения объема трафика, генерируемого сетевыми пользователями, приложениями и сервисами, можно определить шаблоны (профиль) их поведения. При этом задача администраторов КС заключается не только в реагировании на факт изменения объема трафика, но также в предвидении этих изменений путем прогнозирования краткосрочных изменений объема трафика. Например, если администратор точно спрогнозирует краткосрочное изменение объема трафика, то он соответствующим образом сконфигурирует КС, чтобы избежать неэффективного использования полосы пропускания, перегрузок каналов связи, заторов и т.п.

Известно, что краткосрочное изменение объема трафика, генерируемого пользователями, приложениями и сервисами КС, является нелинейным и нестационарным. Традиционные модели прогнозирования, такие как AR (Auto Regressive), MA (Moving Average), ARMA (Auto Regressive Moving Average), ARIMA (Auto Regressive Integrated Moving Average) и т.п., не позволяют получать точные краткосрочные прогнозы трафика КС из-за их неспособности аппроксимировать пики и быстрые колебания трафика КС. Поэтому появляется необходимость в разработке более точных и эффективных методов прогнозирования трафика КС.

Целью данной статьи является разработка точного и эффективного метода краткосрочного прогнозирования объема трафика КС. Для этого предлагается метод с использованием модели CART (Classification and Regression Trees) [2].

Обзор литературы

В литературе для прогнозирования различных характеристик сетевого трафика предложены различные подходы. Предлагаемые подходы основываются на методах статистического анализа данных, анализа временных рядов, интеллектуального анализа данных, нейронных сетей и т.п. Для прогнозирования передачи больших файлов в сетевом

трафике в работе [3] предлагается модель временных рядов ARMA (Autoregressive Moving Average). Авторы проводили эксперименты с наборами данных протокола передачи файлов (File Transfer Protocol – FTP). Метод ARMA был применен для прогнозирования файлов различными объемами на разных промежутках времени. В работе [4] для прогнозирования циклических и сезонных шаблонов трафиков битторрент-приложений была предложена модель временного ряда с использованием ARMA. Для оценки модели авторы использовали шесть наборов реальных данных, собранных в течение шести дней из трафиков битторрент, и использовали новый инструмент симулирования, названный простая студия сетевого прогнозирования (Simple Network Prediction Studio – SNPS). Оценка эффективности модели осуществлялась на основе метрики среднеквадратической ошибки (Mean Square Errors – MSE). Для прогнозирования высокоскоростного сетевого трафика в работе [5] была предложена модель временного ряда с использованием К-фактора ARMA. Для проверки модели авторы использовали набор данных, состоящих из данных различных сетевых приложений, таких как MPEG, VIDEO, JPEG и т.д. Эффективность прогнозирования модели К-фактора ARMA оценивалась с использованием средней абсолютной ошибки (Mean Absolute Error – MAE) и коэффициента отклонения (Error Ratio – ER).

В работе [6] для прогнозирования сетевого трафика используется модель линейных временных рядов ARIMA (Autoregressive Integrated Moving Average). Для представления временных рядов была использована функция коэффициента корреляции. Для оценки эффективности прогнозирования ARIMA авторы использовали метрику MAPE (Mean Absolute Percentage Error).

В работе [7] предлагается модель скорректированного авторегрессивного интегрированного скользящего среднего (Adjusted Autoregressive Integrated Moving Average – AARIMA), которая используется для моделирования сети и долгосрочного прогнозирования интернет-трафика. Для представления временных рядов авторы использовали метод Бокса-Дженкинса. Эффективность прогнозирования модели AARIMA оценена с использованием средней абсолютной ошибки (Mean Absolute Error – MAE).

В работе [8] предлагается модель нелинейных временных рядов GARCH (Generalized Auto Regressive Conditional Heteroskedasticity), которая используется для определения импульсивного характера интернет-трафика. Метод прогнозирования состоит из одного шага, где рекурсивное прогнозирование завершается, чтобы получить следующие прогнозные значения в подходящее время. Для измерения производительности метода прогнозирования используются метрики GARCH и ARCH (Auto Regressive Conditional Heteroskedasticity).

В работе [9] предложена модель нелинейного временного ряда на основе динамической билинейной рекуррентной нейронной сети (Dynamic Bilinear Recurrent Neural Network – D-BLRNN) для прогнозирования сетевого трафика. Были использованы реальные данные, собранные из ядра Bell в различных временных интервалах. Для измерения эффективности прогнозирования модели D-BLRNN была использована метрика среднеквадратической ошибки (Mean Square Error – MSE).

В работе [10] предложена нейронная сеть на основе многоуровневого восприятия (Multi-Layer Perception – MLP) для анализа и прогнозирования интернет-трафика IP-сетей. Для измерения эффективности прогнозирования нейронной сети была использована среднеквадратическая ошибка прогноза (Root Mean Squared Error of Prediction – RMSEP).

В работе [11] предложена модель нелинейных временных рядов на основе нейронной сети Эльмана (Elman Neural Network – ENN) для прогнозирования интернет-трафика. Для оценки прогнозирования ENN использовались такие метрики, как среднеквадратическая ошибка (Mean Square Error – MSE) и нормализованная среднеквадратическая ошибка

(Normalized Mean Square Error – NMSE).

В работе [12] предложена адаптивная система нейронного нечеткого вывода (Adaptive Neural Fuzzy Inference System – ANFIS) с временными рядами для моделирования и прогнозирования интернет-трафика. Авторы использовали данные, собранные из магистрального интернета по протоколу TCP/IP. Для оценки эффективности прогнозирования модели ANFIS были использованы такие метрики, как среднеквадратическая ошибка (Root Mean Square Error – RMSE) и средняя абсолютная относительная ошибка (Average Absolute Relative Error – AARE).

В работе [13] для прогнозирования долгосрочной зависимости и самоподобия сетевого трафика авторы объединили модели линейных временных рядов ARIMA и нелинейных рядов GARCH. Проверка полученной ARIMA/GARCH-модели была осуществлена с использованием набора данных Окленд, и эксперименты проводились в интервалах времени 1 мс, 10 мс и 100 мс. Для представления состояний временных рядов были использованы функции ACF (Autocorrelation Function) и PACF (Partial Autocorrelation Function), а модель ARIMA/GARCH использовалась для прогнозирования ошибок сетевого трафика в разных временных масштабах.

В работе [14] для краткосрочного прогнозирования сетевого трафика предлагается модель, являющаяся объединением моделей линейного временного ряда ARIMA и нелинейного временного ряда MLANN (Multilayer Artificial Neural Network). Для оценки эффективности прогнозирования полученной модели были использованы среднее абсолютное отношение (Mean Absolute Relation – MAR) и среднеквадратическая ошибка (RMSE).

В работе [15] для прогнозирования самоподобного сетевого трафика были использованы нейросетевая модель нелинейных временных рядов и модель линейного временного ряда ARIMA. Авторы использовали функцию автокорреляции для представления временных рядов с предсказанием одного шага и прогнозом множества шагов.

В работе [16] предложены комбинация модели линейного временного ряда ARIMA и нейросетевая модель нелинейных временных рядов для прогнозирования сетевого трафика. Для оценки эффективности прогнозирования предложенной модели была использована нормализованная среднеквадратическая ошибка (Normalized Mean-Square Error – NMSE).

В работе [17] предложена нелинейная декомпозированная модель временного ряда, разложенная по компонентам тренда, периода, мутации и случайности для долгосрочного прогнозирования сетевого трафика. Для оценки результатов симуляции модели используется функция автокорреляции. Авторы утверждают, что декомпозированная модель может получить более высокую точность при долгосрочном прогнозировании сетевых трафиков, чем модель линейного временного ряда ARIMA.

Метод краткосрочного прогнозирования сетевого трафика КС

Задача краткосрочного прогнозирования объема трафика КС заключается в прогнозировании объема будущего трафика на основе анализа объема предыдущего трафика в определенных коротких интервалах времени. При этом предполагается, что между текущими и предыдущими трафиками существует нелинейная зависимость. Для выбора метода решения задача краткосрочного прогнозирования объема трафика КС принимается как задача прогнозирования временных рядов. Модели временных рядов достаточно популярны для извлечения временных шаблонов сетевого трафика и прогнозирования [18, 19, 20]. Потому, что характеристика изменения объема трафика КС представляет регулярную последовательность временных рядов в течение определенного периода времени. В частности, прогнозирование временных рядов заключается в прогнозировании хронологически упорядоченных переменных.

Краткосрочное прогнозирование объема трафика КС подразумевает прогнозирование объема трафика КС для определенных коротких периодов времени (например, секунды,

минуты, недели), при этом в качестве входных данных используется временной ряд, состоящий из предыдущих наблюдений объема трафика, выбранного в таких же периодах времени (например, секунды, минуты, недели).

Анализ подходов к прогнозированию временных рядов, предложенных в литературе, показал, что методы прогнозирования временных рядов с использованием методов машинного обучения демонстрируют более высокую точность прогнозирования. CART является одним из методов машинного обучения, используемых для построения моделей прогнозирования, и состоит из деревьев классификации и регрессии [2]. При этом модели получаются путем рекурсивного разделения пространства данных и подбором простой модели прогнозирования в каждом разделе. Результаты разделения могут быть представлены в виде деревьев решений. Выход дерева классификации является дискретным значением, то есть классом, а выход дерева регрессии – непрерывным значением переменных.

Деревья решений имеют иерархическую структуру, которая состоит из узлов принятия решений, соединенных друг с другом ветвями, при этом ветви не должны образовывать цикл. Узлы принятия решений на основе конкретных условий оценивают значений определенных переменных для прогнозирования результирующего значения, то есть деревья решений выводят прогнозируемое значение, полученное в результате оценки входных атрибутов. В прогнозировании объема трафика КС конечные узлы в древовидной структуре представляют соответствующую модель линейной регрессии с полностью обученными коэффициентами, а ветви – связь атрибутов, которые приводят к этим целевым переменным.

Прогнозирование объема трафика КС может быть интерпретировано как модель, которая прогнозирует целевую переменную y (прогнозируемый объем трафика) на основе нескольких входных переменных x_i (предыдущие объемы трафика), которые принадлежат вектору состояния X . При этом каждый внутренний узел соответствует одной из входных переменных, а также имеются ветви, идущие к подузлам, соответствующим входным переменным. Каждый лист представляет значение целевой переменной, которая задается значениями входных переменных от корня до листа, а деревья решений имеют набор условий, которые разбивают обучающие образцы на более мелкие части.

Предполагается, что для модели $X_i = \{x_1, x_2, \dots, x_p\}$ является входом соответствующей зависимой переменной y_i . Набор $S_{tr} = \{X_1, X_2, \dots, X_{N_d}\}$ обозначает набор обучающих данных, который будет кластеризован, где размер набора обучающих данных равен N_d . При этом целью обучения дерева решений является кластеризация набора обучающих данных по различным шаблонам.

Формирование деревьев заключается в создании узлов принятия решений, а рост деревьев происходит при разделении узлов принятия решений. Причем должны быть установлены критерии разделения набора данных, а ветви перестают расти, когда разделение достигает листового узла. При этом каждый узел принятия решений реализует определенную функцию $f(S)$ для разделения соответствующего пространства данных S на два подмножества и для узла принятия решения m имеет следующий вид:

$$f_m(S_m) : x_i^j \geq \theta_{m0}, \quad (1)$$

где x_i^j обозначает i -й объект в пространстве данных S_m ; θ_{m0} – лучший порог разделения, который приводит к минимальной ошибке. Таким образом, наилучшее разбиение на узле решения делит входное пространство на два пространства: $L_m = \{S_m | x_i^j > \theta_{m0}\}$ и $R_m = \{S_m | x_i^j \leq \theta_{m0}\}$.

Для измерения достоверности разделения используется среднеквадратическая ошибка. Пусть, y_m представляет фактический объем трафика, а \hat{y}_m – его прогнозируемое

значение, тогда получим:

$$E_m = \frac{1}{N_m} \sum_t (y_{m,t} - \hat{y}_{m,t})^2, \quad (2)$$

где N_m – размер пространства данных на узле m ; прогнозируемое значение \hat{y}_m получается из весов: $\hat{y}_m = X_m \times w_m$. Аналогично, ошибка после разбиения пространства S_m на подмножества L_m и R_m может быть рассчитана по формуле:

$$E'_m = \frac{1}{N_{m,L}} \sum_t (y_{m,L,t} - \hat{y}_{m,L,t})^2 + \frac{1}{N_{m,B}} \sum_t (y_{m,B,t} - \hat{y}_{m,R,t})^2. \quad (3)$$

Разность уравнений 2 и 3 представляет уменьшение ошибки, при этом оптимальный признак разделения и порог θ_{m0} определяются путем расчета минимального уменьшения ошибки. Как только ветвь достигает узла, который удовлетворяет заданным условиям, она перестает расти и узел помечается как конечный узел. После этого пространство данных в листовом узле устанавливается методом линейной регрессии. Если число обучения меньше определенного процента от набора обучения, то узлы больше не разделяются. Это связано с тем, что любое решение, основанное на слишком малом числе экземпляров, может вызвать отклонения и привести к ошибкам. Как правило, отсечение частей дерева решений осуществляется после обучения. Для отсечения ребер необходимо использовать отдельный набор данных, поскольку применение набора данных, который рассматривался во время обучения, не позволит выявить каких-либо изменений.

Для прогнозирования объема трафика КС предварительно устанавливаются порог уменьшения ошибки θ_{er} и наименьшее количество векторов состояния, содержащихся в листовом узле θ_N . Дерево перестает расти тогда, когда выполняется одно из двух условий: ошибка уменьшения E_m меньше, чем θ_{er} , или количество векторов состояния содержащихся в листовом узле N_{leaf} не больше, чем θ_N .

Построенная с использованием обучающих данных модель CART используется для прогнозирования объема трафика КС. Для этого используется вектор состояния вновь поступающего трафика. При этом генерация прогноза y на основании независимых переменных X заключается в вычислении линейной регрессии независимых переменных, которые принадлежат шаблону трафика, соответствующего данному вектору состояний. Другими словами, набор обучающих данных классифицируется на листовые узлы с соответствующим пространством данных $\{X_i, y_i\}$. Модель регрессии $M_i: y_i \leftarrow f_i(X_i)$ встроена в листовую узел, показывает, что вектор X_i используется для определения y_i :

$$\begin{cases} M_1: y_1 \leftarrow f_1(X_1) \\ M_2: y_2 \leftarrow f_2(X_2) \\ \vdots \\ \vdots \\ M_n: y_n \leftarrow f_n(X_n) \end{cases} \quad (4)$$

Наконец, вновь поступающие векторы состояния классифицируются с использованием существующей модели CART. Следовательно, будущий объем краткосрочного трафика КС может быть получен с использованием соответствующего веса.

$$y_{new} = X_{new} \times w_m \quad . \quad (5)$$

Поскольку характеристика изменения объема трафика КС является своего рода временным рядом, то вектор состояния, вводимый в модель, может быть представлен в следующем виде:

$$X(t) = [V_t, V_{t-1}, \dots, V_{t-d}], \quad (6)$$

где d – соответствующее значение временных задержек; V_t – объем трафика в течение текущего интервала времени; V_{t-1} – объем трафика в предыдущем интервале времени. При этом в качестве y_t берется значение объема трафика V_{t+1} .

Выбор значения параметра d является очень важным для достижения точного прогнозирования, поскольку его значение влияет на точность прогноза и сложность модели. Например, при слишком больших значениях d полученная последовательность векторов состояний не будет отражать внезапных изменений объема трафика КС, а при слишком маленьких значениях d полученная последовательность векторов состояний не будет отражать тенденцию изменения состояния трафика КС. Поэтому при определении оптимального значения d необходимо учитывать точность и сложность модели. Обычно, чтобы уменьшить сложность модели прогнозирования, значение d выбирается меньше 10 [1]. По существу, шаблоны трафика КС могут не только демонстрировать изменение значений объема трафика, но также отражать состояние трафика (например, повышение, падение и перегруженность) в течение интервалов $d + 1$.

Краткосрочный прогноз объема трафика КС определяется следующим образом: сначала текущий вектор состояния трафика назначается определенному листовому узлу, а затем желаемое значение объема трафика вычисляется с помощью модели регрессии, соответствующей листовому узлу. Значение прогноза определяется с помощью модели регрессии $M_i: y_i \leftarrow f_i(X_i)$, соответствующей модели трафика. При этом модель линейной регрессии должна устанавливать связь между предыдущим и желаемым состояниями трафика, а веса, хранящиеся в листовых узлах, рассчитываются следующим образом:

$$w_i = (X_i^T \times X_i)^{-1} \times (X_i^T \times y_i) \quad . \quad (7)$$

Для оценки эффективности предложенного метода краткосрочного прогнозирования трафика КС используется средняя абсолютная масштабированная ошибка MASE (Mean Absolute Scaled Error) [21]. В отличие от традиционных показателей ошибок, MASE является масштабированной, которая учитывает градиент фактических значений. Чем меньше значение MASE, тем лучше предсказание. Значение MASE определяется следующим образом:

$$MASE = \frac{1}{K} \sum_{k=1}^K \left| \frac{\hat{V}_k - V_k}{\frac{1}{K-1} \sum_{k=2}^K |\hat{V}_k - \hat{V}_{k-1}|} \right|, \quad (8)$$

где K – общее количество ожидаемых объемов трафика в наборе тестовых данных; \hat{V}_k – фактический средний объем трафика в рассматриваемом интервале времени; V_k – значение прогнозирования, полученное с помощью предложенной модели прогнозирования.

Заключение

Точное краткосрочное прогнозирование объема трафика очень важно в мониторинге, а также в эффективном управлении оптимальным использованием имеющихся сетевых ресурсов КС. В данной работе была рассмотрена проблема краткосрочного прогнозирования объема трафика КС. Краткосрочное прогнозирование объема трафика КС подразумевает прогнозирование объема трафика КС для определенных коротких периодов

времени. Для выбора метода решения задача была сформулирована как задача прогнозирования временных рядов.

Анализ предложенных в литературе подходов к прогнозированию временных рядов показал, что прогнозирование временных рядов на основе методов машинного обучения демонстрирует более высокую точность. В данной статье для краткосрочного прогнозирования объема трафика КС была использована модель CART, которая является одним из методов машинного обучения.

Для построения модели CART сначала формируется пространство данных в виде векторов состояния, которые состоят из последовательности объемов трафика КС. В качестве обучающих данных используется пространство данных предыдущих состояний объема трафика КС. При построении модели CART сначала проводится кластеризация пространства данных в различных подмножествах. Для каждого подмножества применяется модель линейной регрессии, чтобы определить взаимосвязь между векторами предыдущих состояний объема трафика КС и желаемым значением впоследствии. После классификации и создания регрессии будущее состояние объема трафика КС может быть спрогнозировано путем присвоения текущего вектора состояния объема трафика оптимальному подмножеству и использования соответствующей модели регрессии. Для оценки точности метода используется средняя абсолютная масштабированная ошибка MASE.

Результаты краткосрочных прогнозов трафика КС могут быть использованы для оптимизации использования ресурсов КС, предотвращения заторов, перегрузки каналов связи, повышения QoS-показателей сети и т.п.

В будущем для повышения точности прогнозирования объема трафика КС мы намерены применять более современные методы машинного обучения и нейронные сети, в частности глубокие нейронные сети.

Литература

1. Zhani M.F., Elbiaze H. Analysis and Prediction of Real Network Traffic // Journal of networks, 2009, vol. 4, no. 9, pp. 855–865.
2. Breiman L., Friedman J.H., Olshen R.A., Stone, C.J. Classification and Regression Trees. Wadsworth International Group, Belmont CA, 1984, 368 p.
3. Hoong N.K., Hoong P.K., Tan I.K.T., Muthuvelu N.M., Seng L.C. Impact of Utilizing Forecasted Network Traffic for Data Transfers / IEEE 13th International Conference on Advanced Communication Technology, 2011, pp.1199–1204.
4. Hoong P.K., Tan K.T., Keong C.Y., BitTorrent Network Traffic Forecasting With ARIMA // International Journal of Computer Networks & Communications, 2012, vol.4, no.4, pp. 143–156.
5. Sadek N., Khotanzad A. Multi-scale High Speed Network Traffic Prediction Using K-Factor Gengendaue ARMA Model / IEEE International Conference on Communications, 2004, pp. 2148–2152.
6. Yu Y., Wang J., Song M., Song J. Network Traffic prediction and result analysis based seasonal and ARIMA and Correlation Coefficient / IEEE International Conference on Intelligent System Design and Engineering Application, 2010, vol.1, pp. 980–983.
7. El Hag H.M.A., Sharif S.M. An Adjusted ARIMA Model for Internet Traffic / IEEE AFRICON Conference, 2007, 6 p.
8. Anand N.C., Scoglio C.S., Natarajan B. GARCH Non-Linear Time Series Model for Traffic Modeling and Prediction / IEEE Network Operations and Management Symposium, 2008, pp. 694–697.
9. Park C., Woo D-M. Prediction of Network Traffic by Using Dynamic Bilinear Recurrent Neural Network / IEEE Fifth International Conference on Natural Computation, ICNC 2009, Tianjian, China, 14-16 August 2009, pp. 419–423.

10. Chabaa S., Zeroual A., Antari J. Identification and Prediction of Internet Traffic Using Artificial Neural Networks // Journal of Intelligent Learning Systems and Applications, 2010, vol. 2, no. 3, pp. 147–155.
11. Junsong W., Jiukun W., Maohua Z., Junjie W. Prediction of Internet Traffic Based on Elman Neural Network / IEEE Chinese Control and Decision Conference, 2009, pp. 1248–1252.
12. Chabaa S., Zeroual A., Antari J. ANFIS Method for Forecasting Internet Traffic Time Series / Mediterranean Microwave Symposium, 2009, pp. 1–4.
13. Zhou B., He D., Sun Z. Traffic predictable based on ARIMA/GARCH Model / IEEE 2006 2nd Conference on Next Generation Internet Design and Engineering, 2006, pp. 200–207.
14. Zeng D., Xu J., Gu J., Liu L., Xu G. Short Term Traffic Flow Prediction Using Hybrid ARIMA and ANN model / IEEE Workshop on Power Electronics and Intelligent Transportation System, 2008, pp. 621–625.
15. Rutka G., Network Traffic Prediction using ARIMA and Neural Networks Models // Elektrotehnika, 2008, vol. 84, no. 4, pp. 53–58.
16. Iqbal M. F., John L. K. Power and Performance Analysis of Network Traffic Prediction Techniques / IEEE International Symposium on Performance Analysis of Systems & Software, 2012, pp. 112–113.
17. Guang C., Jian G., Wei D. A Time series Decomposed Model of Network Traffic // Springer, 2005, pp. 338–345.
18. Hyndman R., Athanasopoulos G., Forecasting: principles and practice. OTexts, 2014. <https://books.google.com>
19. Theyazn H. H. Aldhyan, Manish R. Joshi Intelligent Time Series Model to Predict Bandwidth Utilization // International Journal of Computer Science and Applications, 2017, vol. 14, no. 2, pp. 130–141.
20. Jung S, Kim C., Chung Y. A Prediction Method of Network Traffic Using Time Series Models, ICCSA 2006, pp. 234–243.
21. Hyndman R. J., Koehler A. B. Another look at measures of forecast accuracy // International Journal of Forecasting, 2006, vol. 22, no. 4, pp. 679–688.

UOT 004.855.5

Şıxəliyev Ramiz H.

AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

ramiz@science.az

Kompüter şəbəkələrinin trafikinin proqnozlaşdırılmasının bir metodu haqqında

Kompüter şəbəkələrinin (KŞ) uzun və qısa müddətdə trafikinin həcmnin dəqiq proqnozlaşdırılması həm monitorinqdə, həm də mövcud şəbəkə resurslarının optimal istifadənin effektiv idarə edilməsində çox vacib rol oynayır. Adətən, daha təcrübəli şəbəkə administratorları KŞ-nin trafikinin həcmnin intuitiv şəkildə proqnozlaşdırırlar, lakin müasir, böyük və mürəkkəb KŞ-lərin idarəsi üçün bu tamamilə yolverilməzdir. Buna görə, KŞ administratorlarına mövcud şəbəkə resurslarından istifadəni səmərəli şəkildə planlaşdırmağa və optimal şəkildə idarə etməyə kömək edəcək, maşın təlimi metodlarından istifadə etməklə KŞ-nin trafikinin həcmnin proqnozlaşdırmaq üçün daha dəqiq metodlar işlənilməlidir. Bu məqalədə, KŞ-nin trafikinin həcmnin qısamüddətli proqnozu üçün, CART (*Classification and Regression Trees*) modelinə əsaslanan üsul təklif edilir. Metodun mahiyyəti ondan ibarətdir ki, qərar ağaclarının köməyi ilə əvvəlki trafik həcmnin vəziyyətləri çoxluğu trafik həcmnin vəziyyətləri şablonlarına görə klassifikasiya edilir və hər sinfə uyğun xətti reqressiya modeli qurulur. Metod, mövcud trafik həcmnin vəziyyətlər vektorlarını ən uyğun əvvəlki şablonlara görə qruplaşdıraraq və reqressiya tətbiq etməklə KŞ-nin gələcək trafik həcmnin vəziyyətini proqnozlaşdırmağa imkan verir. Beləliklə, KŞ-nin trafikinin həcmnin qısamüddətli proqnozunun məsələsi mövcud trafik həcmnin vəziyyətləri vektorlarını və proqnozlaşdırma üçün reqressiya modelini təyin etməkdir. Metodun düzgünlüyünü qiymətləndirmək üçün orta mütləq

miqyaslı səhv MASE (*Mean Absolute Scaled Error*) istifadə olunur. Təklif olunan metod KŞ-nin trafikinin həcmnin qısa müddət üçün, məsələn, həftələr, günlər, saatlar və saniyələr üçün proqnozlaşdırmağa imkan verir. Qısamüddətli proqnozların nəticələrindən QoS-u yaxşılaşdırmaq, rabitə kanallarının həddən artıq yüklənməsinin qarşısını almaq, KŞ-nin mövcud şəbəkə resurslarının optimal idarə etmək və s. üçün istifadə etmək olar.

Açar sözlər: *kompüter şəbəkələri, şəbəkə trafiki, trafikin proqnozu, CART modeli, klassifikasiya, qərar ağacı*

Ramiz H. Shikhaliyev

Institute of Information Technology of ANAS, Baku, Azerbaijan

ramiz@science.az

One method for computer networks traffic prediction

Accurate prediction of the traffic volume of computer networks (CN) for both the long and short term plays a crucial role in monitoring, as well as in the effective management of the optimal use of available network resources. Typically, more experienced network administrators intuitively predict the traffic volume of the CN, however this is completely unacceptable for the administration of modern, large and complex CNs. Therefore, more accurate methods for predicting the traffic volume of the CN should be developed using machine learning methods that will help the CN administrators effectively plan and optimally manage the use of available network resources. This article proposes a method for short-term CN traffic volume prediction based on the CART (Classification and Regression Trees) model. The essence of the method is that, using decision trees, the previous sets of traffic volume states are classified according to the set of traffic volume state patterns and a linear regression model is constructed corresponding to each class. The method allows predicting the future CN traffic volume state by clustering vectors of current traffic volume states according to the most suitable previous patterns and using regression. Thus, the task of CN traffic volume short-term prediction is to determine the vectors of the current traffic volume states and a regression model for forecasting. To assess the accuracy of the method, the MASE (Mean Absolute Scaled Error) is used. The proposed method allows predicting the traffic volume of the CN for short periods of time, for example, for weeks, days, hours and seconds. The results of short-term forecasts can be used to improve QoS (Quality of Service), prevent overload of communication channels, optimal control of the available CN resources, etc.

Keywords: *computer networks, network traffic, traffic prediction, CART model, classification, decision tree.*